Calhoun: The NPS Institutional Archive

DSpace Repository

NPS Scholarship | Theses

2023-09

COCAINE SEIZURES AND CRIME: DATA ANALYTICS USING BIG DATA TOOLS

Martinez Galeano, Edwin A.

Monterey, CA; Naval Postgraduate School

https://hdl.handle.net/10945/72375

Copyright is reserved by the copyright owner.

Downloaded from NPS Archive: Calhoun

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

COCAINE SEIZURES AND CRIME:
DATA ANALYTICS USING BIG DATA TOOLS

by

Edwin A. Martinez Galeano

September 2023

Thesis Advisor: Anthony M. Canan
Co-Advisor: Richard D. Bergin IV

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2023	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE COCAINE SEIZURES AND CRIME: DATA ANALYTICS USING BIG DATA TOOLS	**5. FUNDING NUMBERS**
6. AUTHOR(S) Edwin A. Martinez Galeano	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Colombia's status as the largest cocaine producer in the world has prompted its government's strategies to combat drug trafficking. One of these strategies is to seize cocaine in the Colombian jurisdictional territory. The unintended consequences of this strategy on crime rates, particularly homicides, remain uncertain. Web scraping methods and big data tools were used to gather and construct a time series dataset on cocaine seizures from three distinct websites, while the homicides dataset was supplied by the Colombian Ministry of Defense (MDN). This study aims to investigate, from a quantitative standpoint, whether there is a link between cocaine seizures and homicides in the Colombian Pacific region, utilizing an exploratory data analysis (EDA) method and machine learning techniques. The study recognizes the constraints of the sample size and opts to reveal valuable insights through data analysis and modeling instead. Despite the constraints, two models were developed to partially explicate the significance of this correlation. The study's findings provide value for policymakers, military personnel, government officials, and academics, offering essential perspectives to devise improved policies and strategies to mitigate drug trafficking in the Colombian Pacific region without exacerbating homicide rates. Future research endeavors could consider expanding the sample size of the cocaine seizure time-series dataset to conduct a more robust correlation analysis.

14. SUBJECT TERMS drug trafficking, cocaine seizures, crime, homicides, web scrapping, big data, exploratory data analysis, machine learning, polynomial regressions, correlation	**15. NUMBER OF PAGES** 217
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

COCAINE SEIZURES AND CRIME:
DATA ANALYTICS USING BIG DATA TOOLS

Edwin A. Martinez Galeano
Capitan de Corbeta, Colombian National Navy
BEE, Universidad Catolica de Colombia, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
September 2023

Approved by: Anthony M. Canan
Advisor

Richard D. Bergin IV
Co-Advisor

Alex Bordetsky
Chair, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Colombia's status as the largest cocaine producer in the world has prompted its government's strategies to combat drug trafficking. One of these strategies is to seize cocaine in the Colombian jurisdictional territory. The unintended consequences of this strategy on crime rates, particularly homicides, remain uncertain. Web scraping methods and big data tools were used to gather and construct a time series dataset on cocaine seizures from three distinct websites, while the homicides dataset was supplied by the Colombian Ministry of Defense (MDN). This study aims to investigate, from a quantitative standpoint, whether there is a link between cocaine seizures and homicides in the Colombian Pacific region, utilizing an exploratory data analysis (EDA) method and machine learning techniques. The study recognizes the constraints of the sample size and opts to reveal valuable insights through data analysis and modeling instead. Despite the constraints, two models were developed to partially explicate the significance of this correlation. The study's findings provide value for policymakers, military personnel, government officials, and academics, offering essential perspectives to devise improved policies and strategies to mitigate drug trafficking in the Colombian Pacific region without exacerbating homicide rates. Future research endeavors could consider expanding the sample size of the cocaine seizure time-series dataset to conduct a more robust correlation analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1

A. BACKGROUND 1
B. STATEMENT OF THE PROBLEM 3
C. RELEVANCE OF STUDY FOR AUDIENCES 3
D. PURPOSE OF THE STUDY 4
E. RESEARCH QUESTION 4

II. REVIEW OF THE LITERATURE 5

A. THE COCAINE PROBLEM 5
B. THE COCAINE AND CRIME RELATION 6
C. IDENTIFICATION OF COCAINE AND CRIME FACTORS 9
D. COLOMBIAN PACIFIC REGION 12
E. DEFICIENCIES IN LITERATURE 14
F. PHILOSOPHICAL WORLDVIEW 15

III. METHODS 17

A. TYPE OF RESEARCH DESIGN 17
B. DATA COLLECTION AND PROCESSING INSTRUMENTS 17
C. INTERNET WEB BROWSER 18
1. Big Data Tools 18
2. Web Scraping Tools 19
3. Statistical Data Tools 34
D. DATA PROCESSING 34

IV. RESULTS 41

A. TIME-SERIES DATASETS 41
B. DATASETS TIME FRAMES 46
C. EXPLORATORY DATA ANALYSIS 47
D. FURTHER ANALYSIS 50
E. DSD+D+2-SUBSET 52
1. Graph Builder Fitting and Least Squares Analysis 52
2. Continuous Distribution Comparison 54
3. Profiling, Simulating, and Prediction Analysis 55
F. DSD+D+6-SUBSET 58
1. Graph Builder Fitting and Least Squares Analysis 58
2. Continuous Distribution Comparison 60

3. Profiling, Simulating, and Prediction Analysis 61

V. DISCUSSION 65

APPENDIX A. ARC NEWS WEBSITE LEGAL CONSTRAINTS 69

APPENDIX B. USCG NEWS WEBSITE LEGAL CONSTRAINTS 71

APPENDIX C. PONAL NEWS WEBSITE LEGAL CONSTRAINTS 73

APPENDIX D. ARC NEWS WEBSITE ANATOMY 75

APPENDIX E. USCG NEWS WEBSITE ANATOMY 77

APPENDIX F. WEB SCRAPING TOOLS ASSESSMENTS 79

APPENDIX G. OCTOPARSE TASKS DETAILS 123

A. TASK 1: ARC NEWS WEBSITE 123
B. TASK 2: PONAL NEWS WEBSITE 124
C. TASK 3: USCG NEWS WEBSITE 125

APPENDIX H. EXCEL FILE WITH SCRAPED DATA FOR COCAINE SEIZURES DATASET 127

APPENDIX I. HIVE SCRIPTS FOR COCAINE SEIZURES DATASET 129

A. HIVE SCRIPT FOR DATA SCRAPED FROM ARC NEWS WEBSITE 129
B. HIVE SCRIPT FOR DATA SCRAPED FROM PONAL NEWS WEBSITE 133
C. HIVE SCRIPT FOR DATA SCRAPED FROM USCG NEWS WEBSITE 137

APPENDIX J. EXCEL FILE WITH ETL DATA FROM HIVE FOR COCAINE SEIZURES DATASET 141

APPENDIX K. EXCEL FILES WITH FILTERS AND PROCESSES FOR COCAINE SEIZURES AND HOMICIDES DATASET 143

APPENDIX L. TIME SERIES DATASETS 145

APPENDIX M. FINDINGS ON POLYNOMIAL REGRESSIONS 147

A. FINDING 1 147
B. FINDING 2 148
C. FINDING 3 151
D. FINDING 4 153
E. FINDING 5 155
F. FINDING 6 157
G. FINDING 7 158

APPENDIX N. GRAPH BUILDER FITTING SUBSETS 161

APPENDIX O. DSD+S+2-SUBSET MODELS 163

APPENDIX P. DSD+S+6-SUBSET MODELS 165

SUPPLEMENTALS 167

APPENDIX H SUPPLEMENTALS 167
APPENDIX J SUPPLEMENTALS 167
APPENDIX K SUPPLEMENTALS 168
APPENDIX L SUPPLEMENTALS 168
APPENDIX M SUPPLEMENTALS 169
APPENDIX N SUPPLEMENTALS 171
APPENDIX O SUPPLEMENTALS 171
APPENDIX P SUPPLEMENTALS 172

LIST OF REFERENCES 175

INITIAL DISTRIBUTION LIST 191

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1. Cocaine factors associated to crime factors in literature review 10

Figure 2. Data collection and processing instruments 18

Figure 3. Ethical considerations and legal constraints of web scraping 24

Figure 4. Theory of Planned Behavior. Source: Ajzen (1991). 25

Figure 5. User's behavior flow when scraping open websites 26

Figure 6. Data processing 35

Figure 7. Filters and processing on the dataset for cocaine seizures in Excel 37

Figure 8. Filters and processing on the datasets for homicides in Excel 38

Figure 9. Time-series datasets description 41

Figure 10. Cocaine seizures in the Pacific 2012–2022 histogram 42

Figure 11. Homicides modalities 43

Figure 12. Homicides rate in the Colombian Pacific region 2012–2022 histogram 45

Figure 13. Alternative hypothesis from findings 2 and 4 50

Figure 14. Filtering, exploring, and modeling procedure in JMP 50

Figure 15. Standard least squares parameters 51

Figure 16. DSD+D+2-subset polynomial regression result 52

Figure 17. Studentized residuals plot for DSD+D+2-subset 53

Figure 18. DSD+D+2-subset without outliers' regression result and studentized residuals 53

Figure 19. DSD+D+2-subset without outlier (row 10) regression result and studentized residuals 54

Figure 20. DSD+D+2-subset without outlier (row 10) lognormal distribution 55

Figure 21. Model 1: Prediction profiler and simulator for Subset 1 56

Figure 22. Interaction profiles for Model 1 57
Figure 23. DSD+D+6-subset polynomial regression result 58
Figure 24. Studentized residuals plot for DSD+D+6-subset 59
Figure 25. DSD+D+6-subset without outliers' regression result and studentized residuals 60
Figure 26. DSD+D+6-subset without outliers Johnson Sb distribution 60
Figure 27. Model 2: Prediction profiler and simulator for Subset 2 61
Figure 28. Interaction profiles for Model 2 62
Figure 29. ARC website first scraping level 75
Figure 30. ARC website second scraping level with pagination 76
Figure 31. ARC website second scraping level 76
Figure 32. USCG website first scraping level 77
Figure 33. Second scraping level 78
Figure 34. Exclusive URL for ARC website 123
Figure 35. ARC semi-automatic field detection 123
Figure 36. Workflow for the ARC website 123
Figure 37. Exclusive URL for PONAL website 124
Figure 38. PONAL semi-automatic field detection 124
Figure 39. Workflow for the PONAL website 124
Figure 40. Exclusive URL for USCG website 125
Figure 41. USCG semi-automatic field detection 125
Figure 42. Workflow for the USCG website 125

LIST OF TABLES

Table 1. Identified commercial and open-source web scraping tools. 27

Table 2. Available commercial web scraping tools. 28

Table 3. Available open-source web scraping tools. 29

Table 4. Commercial and open-source web scraping tools for testing. 30

Table 5. Web scraping tools: Final scores and times assessment. 33

Table 6. Level of measurement for time-series datasets variables 45

Table 7. Records per year for time-series datasets 46

Table 8. Datasets time frames 47

Table 9. Findings on polynomial regressions 49

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AICc	Akaike Information Criterion corrected
AQIM	Al-Qaeda in the Islamist Maghreb
ARC	Colombian Navy
B+D+S	Beheading, dismemberment, and strangulation
BIC	Bayesian Information Criterion
CCENDU	Canadian Community Epidemiology Network on Drug Use
CFAA	Computer Fraud and Abuse Act
CLI	Command-Line Interface
D.A.W.N.	U.S. Federal Drug Abuse Warning Network
DOM	Document Object Model
DOS	U.S. Department of State
DSH	U.S. Department of Homeland Security
DUMA	Drug Use Monitoring in Australia
EDA	Exploratory Data Analysis
ELN	*Ejército de Liberación Nacional*
EMCDDA	European Monitoring Center for Drugs and Drug Addiction
ETL	Extract, transform, and load
FARC-EP	Colombian Revolutionary Armed Forces-People's Army
GUI	Graphic User Interface
H0	Null hypothesis
H1	Alternative hypothesis
HDFS	Hadoop Distributed File System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
KDD	Knowledge Discovery Database
MDN	Colombian Ministry of Defense
MINTIC	Ministry of Technology and Information in Colombia

NIDA	U.S. National Institute on Drug Abuse
NIH	U.S. National Institute of Health
NLP	Natural Language Processing
NPS	Naval Postgraduate School
ODHDN	Observatory for Human Rights and National Defense
OS	Operating System
PB	Peta Bytes
PNIS	Program for the Substitution of Illicit Crops
PONAL	Colombian Police
PPI	Protected Personal Information
p-value	Probability value
R^2	R-square value
RAM	Random Access Memory
RSME	Root Mean Square Error
SIMCI	Integrated Illicit Crop Monitoring System
SQL	Structured Query Language
SSD	Solid-State Disk
TPB	Theory of Planned Behavior
TRA	Theory of Reasoned Action
U.S.	United States
UK	United Kingdom
UN	United Nations
UNODC	United Nations Office on Drugs and Crime
URL	Uniform Resource Location
USCG	U.S. Coast Guard
VM	Virtual Machine
XBRL	eXtensible Business Reporting Language

EXECUTIVE SUMMARY

Cocaine consumption affects both the individual's health and their surroundings. The use of cocaine has stimulating and addictive properties that affects the brain's limbic system (Nestler, 2005), similarly or more intensely than alcohol (Nutt et al., 2007), with its detrimental impacts extending beyond personal health to societal welfare (Wood et al., 1996). The proliferation of cocaine use in North America and Europe from the 1970s onwards resulted in the creation of various organizations aimed at gathering and evaluating information about cocaine's influence on public health and society, from different perspectives.

One of the perspectives to consider involves the relationship between cocaine and crime, which is the specific focus of this study. After analyzing a comprehensive collection of academic researches that investigated the multifaceted relationships between these two variables, there were two studies that found associations between cocaine seizures and homicides. While an examination of cocaine seizures in Jamaica uncovered a reverse relationship between seizures amounts and homicide rates (UNODC, 2012), Miron (2001) identified a direct connection. Furthermore, Millán-Quijano (2020) asserted that drug trafficking leads to a rise in homicides in Colombian areas linked to the illegal trade. The results of these investigations present a paradoxical situation and raise questions about the potential variations in the relationship between cocaine seizures and homicide rates in diverse geographical contexts.

Given this context, this research centers on addressing the identified knowledge gap concerning the influence of cocaine seizures in the Pacific area on homicides in the Colombian Pacific region, driven by its unique geographical and social conditions. The Colombian Pacific region's isolation, challenging terrain, and limited law enforcement presence have made it a significant corridor for drug trafficking (Salazar, 2010) and illegal armed group activities (Aschner & Montero, 2021). The inhabitants of the region face underdevelopment, poverty, and socio-political challenges (Galvis-Aponte et al., 2017), propelled by illegal activities such as drug trafficking (Guzmán & Sánchez, 2021). The Colombian Pacific region is the primary global producer of coca bushes UNODC-

SIMCI (2022) and exhibits a higher homicide rates compared to the rest of the country (Álvarez & González, 2012).

To explore the relationship between cocaine seizures and homicides in the Colombian Pacific region, this thesis proposes a quantitative approach, employing a correlational design (Creswell, 2012). Although causal claims are not possible due to the limitations of the correlational design (Asamoah, 2014), the research seeks to provide insights between the variables. To achieve this purpose, data analysis and modeling processes imply the use of big data tools, advanced regression methods, and machine learning techniques.

To undertake the analysis and modeling of the variables, two datasets were necessary. The first dataset encompasses information concerning cocaine seizures in both the Pacific sea and land environments. To construct this dataset, big data tools were employed to extract news about cocaine seizures from three specific websites. A comprehensive procedure was carried out to select an appropriate web scraping tool, considering technical, ethical, and legal aspects. Most of the scraped data was derived from news articles published between 2012 and 2022, thereby defining the research's temporal scope. After scrapping, this data was stored in Hadoop and underwent processing in Hive, followed by a subsequent filtering procedure using Excel. On the other hand, the second dataset contains data about homicide rates in the Colombian Pacific region from 2012 to 2022. This dataset was provided by the Colombian Ministry of Defense (MDN) (MDN-ODHDN, 2023).

After exploring different combinations between the dependent and independent variables, two distinct alternative hypotheses were formulated based on the outcomes of regression analysis. The two findings suggest that large cocaine seizures in the Pacific area have an influence on the emergence of a high-intensity homicide rate in the Colombian Pacific region, becoming evident two and six days after the seizures, respectively.

A new filtering, exploring, and modeling procedure was done on these two alternative hypotheses to fit better models. Using the graph builder tool in JMP, a hands-

on exploration was conducted to redefine the cocaine seizures limits for both datasets derived from the alternative hypotheses. Enhancing the precision of these two models involved applying a least squares analysis, excluding outliers, to minimize disparities and uncover the optimal alignment between the variables. At the end two improved models were obtained and used for profiling, simulation, and prediction purposes.

For the first model, the statistical examination results in a high significant p-value <0.0001. Furthermore, the cubic fit model results demonstrate its effectiveness by accounting for 80% of the variability in homicide rates exceeding eight within the Colombian Pacific region. This variability is attributed to instances of cocaine seizures surpassing 750 kg, which occur in the Pacific area two days prior.

Regarding the second model, the statistical examination results in a high significant p-value <0.0011. Also, this cubic model applied demonstrates its effectiveness by accounting for 52% of the variability in homicide rates exceeding nine within the Colombian Pacific region. This variability is attributed to instances of cocaine seizures surpassing 545 kg, which occur in the Pacific area six days before.

These results suggest that large cocaine seizures in the Pacific area might have an influence in the homicide rates in the Colombian Pacific region. Hence, there is a need to balance anti-drug trafficking strategies to achieve immediate benefits while avoiding unintentional escalation of violence in Colombia. Persistent oversight of these strategies and vigilant scrutiny of homicide rates within their operational domains hold significant importance. Short-term successes should encourage coca growers to engage in sustained government programs for lasting results, requiring consistent commitment regardless of political changes.

Although the limitations of the models stem from difficulties in obtaining comprehensive international cocaine seizure data due to technical, ethical, and legal constraints, authorities struggling with cocaine trafficking issues should undertake future research on this correlation. In doing that, international collaboration is vital, but efforts must be carefully designed to avoid unintended consequences. Future research could expand sample sizes, analyze the impact of interdiction efforts on homicide rates in

bordering territories, or consider qualitative studies to validate findings and gain deeper insights from the real-world situation in the Colombian Pacific region.

References

Álvarez, C., & González, C. (2012). Análisis espacial de la violencia homicida en el pacífico colombiano [Spatial analysis of homicidal violence in the Colombian Pacific]. *Revista de Economía & Administración E-ISSN 2463–1035 ISSN 1794–7561*, *9*(1), Article 1. https://revistas.uao.edu.co/ojs/index.php/REYA/article/view/168

Asamoah, M. K. (2014). Re-examination of the limitations associated with correlational research. *Journal of Educational Research and Reviews*, *2*(4), 45–52.

Aschner, J. P., & Montero, J. C. (2021). Architectures, spaces, and territories of illicit drug trafficking in Colombia and Mexico. *Crime, Media, Culture*, *17*(3), 327–351. https://doi.org/10.1177/1741659020910212

Creswell, J. W. (2012). Educational research; planning, conducting, and evaluating quantitative and qualitative research. *Reference and Research Book News*, *26*(3). https://www.proquest.com/docview/869979977/citation/AEEE084E6C2F442EPQ/1

Galvis-Aponte, L. A., Moyano-Támara, L. M., & Alba-Fajardo, C. A. (2017). La persistencia de la pobreza y sus factores asociados [The persistence of poverty and its associated factors]. *Chapters*, 49–102.

Guzmán, J. P. S., & Sánchez, N. T. (2021). Desarrollo social, factor de mitigación de la criminalidad en el Pacífico colombiano [Social development, a mitigating factor for crime in the Colombian Pacific]. *Pensamiento Americano*, *14*(27), Article 27. https://doi.org/10.21803/penamer.14.27.351

Millán-Quijano, J. (2020). Internal cocaine trafficking and armed violence in Colombia. *Economic Inquiry*, *58*(2), 624–641. https://doi.org/10.1111/ecin.12771

Ministerio de Defensa Nacional-Observatorio de Derechos Humanos y Defensa Nacional. (2023). *Homicidios en Colombia 2012–2022* [dataset].

Miron, J. A. (2001). Violence, guns, and drugs: A cross-country analysis. *The Journal of Law & Economics*, *44*(S2), 615–633. https://doi.org/10.1086/340507

Nestler, E. J. (2005). The neurobiology of cocaine addiction. *Science & Practice Perspectives*, *3*(1), 4–10. https://doi.org/10.1151/spp05314

Nutt, D., King, L. A., Saulsbury, W., & Blakemore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse. *The Lancet*, *369*(9566), 1047–1053. https://doi.org/10.1016/S0140-6736(07)60464-4

Salazar, L. G. S. (2010). Corredores y territorios estratégicos del conflicto armado colombiano: Una prioridad por territorializar en la geopolítica de los actores armados / Strategic corridors and territories in the colombian armed conflict: A priority for territorializing for. *Perspectiva Geográfica*, 9–36. https://doi.org/10.19053/01233769.1729

United Nations Office on Drugs and Crime. (2012). *Transnational organized crime in Central America and the Caribbean: A Threat Assessment*. United Nations Office on Drugs and Crime. https://doi.org/10.18356/493ae18b-en

UNODC-Integrated Illicit Crop Monitoring System [SIMCI]. (2022). *Colombia: Monitoring of territories affected by illicit crops 2021* (p. 169). https://www.biesimci.org/fileadmin/2022/documentos/survey_of_territories_affected_by_illegal_crops_2021_english.pdf

Wood, A. J., Mendelson, J. H., & Mello, N. K. (1996). Drug therapy: Management of cocaine abuse and dependence. *The New England Journal of Medicine*, *334*(15), 965–972.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

My life has gained fresh energy during these two years at the Naval Postgraduate School, accompanied by a range of emotions. On one hand, I received somber news from home regarding my aunt's passing due to cancer. On the other side, I've been fortunate to experience this incredible country with my family and engage in its culture, particularly in such a peaceful and inspiring place like Monterey. Handling these emotions would have been a challenging endeavor without the support of my loving wife and amazing son. I am deeply grateful to Malu and Matias for their infinite love.

I want to thank God as well as my parents, Samuel and Alba, for granting me the opportunity to exist in this world; to my siblings, Erick and Lorena, along with other family members, for helping to mold my character to confront the challenges of life; to all the people who, in one way or another, have contributed to my professional development; and to the Colombian Navy and the U.S. Naval Mission in Colombia for having placed their trust in me to reach new frontiers of knowledge that I will use for the benefit of our institutions and countries.

I would like to extend my heartfelt thanks to everyone who supported me throughout my journey in both the administrative and academic aspects at NPS. I'm grateful to my U.S. sponsor LT Morgan White for her warm welcome and support; the International Graduate Program Office, especially Ms. Kathy Noyes for her unwavering dedication, and Ms. Martha Brumback for her invaluable assistance in enhancing my English language skills. My appreciation also goes out to the teams at Dudley Knox Library and the Graduate Writing Center for their outstanding support and commitment, with special recognition for Jasmine Mally, my writing coach. I am thankful for all the professors I had the privilege to interact with, and I want to highlight Prof. Mustafa Canan and Prof. Richard Bergin for their willingness and guidance as my thesis advisor and co-advisor.

I want to thank the United Nations Office on Drugs and Crime office in Colombia, especially CN (RA) Jorge Torres, for their support and interest in my thesis

topic. Also, I want to express my gratitude to the Human Rights and National Defense Observatory of the Ministry of Defense of Colombia for facilitating access to their databases for the development of my thesis.

Finally, I want to pour out my heartfelt appreciation to all those whose quiet support enriched my journey at NPS, in La Mesa, amidst the charm of Monterey, within the beauty of California, and across the vast expanse of the U.S. Your contributions have not only made our time immensely enjoyable and memorable, but have also played an instrumental part in turning the cherished dream of pursuing a master's degree at this esteemed institution into a reality I hold close to my heart. Thank you.

I. INTRODUCTION

A. BACKGROUND

Since 1997, Colombia has been the largest cocaine producer in the world (Rincón-Ruiz & Kallis, 2013). Recently, the U.S. Department of State estimated a global coca cultivation to be around 358,100 hectares, with Colombia as the main producer, at 65%, approximately (U.S. Department of State [DOS], 2021). The last United Nations' (UN) report on coca bush cultivation in Colombia estimated 204,000 hectares in 2021, up 43% from the year before (United Nations Office on Drugs and Crime [UNODC], 2023).

In 2016, the Colombian government and the Revolutionary Armed Forces-People's Army (FARC-EP) came to a consensus to resolve their internal conflict. As part of this agreement, both parties pledged to address problems associated with narcotics (Gobierno de la República de Colombia, 2016). Consequently, the Colombian government enacted the program for the substitution of illicit crops (PNIS), involving affected communities in collaborative plans (Ministerio del Interior de la República de Colombia, 2017). This approach is in line with the concept of establishing a viable government in the impacted territories, which is seen as a more effective long-term solution (Isacson, 2018).

Nevertheless, PNIS' report No. 24 recorded only 37,941 hectares of coca crops voluntarily eradicated as of July 2022 (UNODC, 2022a). The government is making slow progress in implementing the voluntary crop substitution strategy, yet new coca crops are proliferating (UNODC, 2022b). Furthermore, there has been a notable increase of 481% in the number of murders of social leaders who participated in substitution programs and actively opposed the expansion of illicit crops (Marín Llanes, 2022). This disparity calls attention to whether other strategies such as intensifying drug interdiction efforts, apprehending drug lords, blocking financial flows, or even implementing forced eradication, should be reinforced to diminish cocaine production. Although the government has implemented different law enforcement strategies, forced eradication might not be effective either due to the balloon effect (Dávalos & Morales, 2022).

Additionally, Castro (2017) claimed that the State was weak in most of the Colombian territory, leaving control to illegal armed groups. This seems to be particularly true in the Colombian Pacific region, where criminal activity is significantly elevated, with 29% of the municipalities experiencing a persistent high rate of homicides, particularly in the Valle department (Álvarez & González, 2012). The Colombian Pacific is a region isolated by the Western Ranges, with a complex river network and a wide coastline. Apart from its distinctive geographical conditions, the prevalence of high illiteracy rates, inadequate infrastructure, and limited formal job opportunities have fueled the increase in illegal activities, plunging the region into a cycle of poverty, violence, marginalization, and inequality (Galvis-Aponte et al., 2017) (Aschner & Montero, 2021).

Indeed, the Pacific Ocean is still considered a high flow volume cocaine smuggling route from Colombia to the U.S. (UNODC, 2023). Besides, the Colombian Pacific region holds the top position in terms of coca bush cultivation, making a substantial contribution of 44%, mainly in the Nariño department (UNODC-Integrated Illicit Crop Monitoring System [SIMCI], 2022). Consequently, this has led also to a rise in violence in Ecuador, as local gangs engage in fierce competition to secure their position within the criminal drug distribution network (Charles, 2022).

For now, the Colombian government has stated their intention of protecting coca bush producers and continuing the voluntary crop substitution process, while increasing intelligence capabilities and drug interdiction operations (Semana, 2022). In that sense, Guerrero Castro (2017) asserted on how the human and technical intelligence resources of the Colombian Navy (ARC) have helped increase the success of maritime drug interdiction operations. Also, Morselli and Petit (2007) revealed how intelligence gathering could be used to disrupt drug trafficking networks and understand their ability to reorganize after a seize-but-do-not-arrest strategy.

Although the unprecedent cocaine seizures during the last years in Colombia (LLC, 2023), Europe (LLC, 2022), and even West Africa (SyndiGate Media, 2022) has prevented the arrival of cocaine in the black market, it is crucial to recognize that these seizures could also result in a higher level of criminality. Disruptions to drug trafficking

networks by law enforcement may provoke retaliatory actions and collateral damage (LSE IDEAS, The London School of Economics and Political Science, 2014). While the increase in intelligence capabilities seems to be wise, interdiction operations in regions like the Colombian Pacific, where illegal armed groups maintain a firm grip on the cocaine trade, could potentially lead to a surge in homicides (Godoy, 2018).

In conclusion, the continuing presence of illegal armed groups, added to the unsatisfied needs of the population, have shaped the Colombian Pacific region as the ideal scenario for the production and distribution of cocaine. Also, the flourishing illegal cocaine business in this area has allowed authorities to seize more cocaine than any other region in Colombia. High level of seizures raises doubts about its potential correlation with the high annual homicide rates, suggesting a possible retaliatory response by drug trafficking networks following law enforcement disruptions. Therefore, this study has a limited scope, focusing solely on the examination of the issue in the Colombian Pacific region, due to its distinctive socio-geographical circumstances that separate it from other regions of the country.

B. STATEMENT OF THE PROBLEM

The problem is that it is not clear how cocaine seizures influence/correlate with criminality when the cocaine trafficking logistic chain is broken. If the correlation is positive, homicides rate would increase as authorities seize more cocaine. This might be a problem, because the government's actions to stop cocaine trafficking could inadvertently affect crime rate.

As an example, visualize what is happening with homicides in the Colombian Pacific after the ARC seizes 1 ton of cocaine on the sea. Also, picture what is happening when the police seize 100 kg of cocaine in a municipality. Are homicides increasing because of the action of the authorities in these areas?

C. RELEVANCE OF STUDY FOR AUDIENCES

Different audiences can benefit from this study, such as policymakers, military, government, and academics. Certainly, policymakers could use the link between cocaine

seizures and homicides to create better public policies. Likewise, military personnel could use it to plan strategies against cocaine trafficking while monitoring related crime. Additionally, government officials could implement policies and control the strategies, balancing the impact on homicides. Finally, academics and researchers could build upon this study, identify additional variables, generate new hypothesis, refer to the methodological process, or collaborate in the future. The collaboration can comprise from the verification process to the exchange of information for future studies in this field.

D. PURPOSE OF THE STUDY

The purpose of this study is to investigate if there is a correlation between cocaine seizures and homicides in the Colombian Pacific region. Studying this correlation can give insight into the effectiveness of the strategies against cocaine trafficking. Also, investigating this correlation could elicit new variables for shaping a more optimal set of policies in the future.

E. RESEARCH QUESTION

Do cocaine seizures in the Pacific cocaine trafficking flow influence/correlate with crime in the Colombian Pacific region?

II. REVIEW OF THE LITERATURE

A. THE COCAINE PROBLEM

Cocaine is an illicit drug known for its addictive nature and stimulating effects. Researchers have made significant progress in unraveling the process on the limbic system of the brain by which cocaine induces intoxication, cravings, and increased susceptibility to relapse (Nestler, 2005). Nutt, King, Saulsbury, and Blakemore (2007) classified cocaine as a class A drug, implying that the dependence, physical and social harm scores assigned by independent experts and psychiatrists were equal to or greater than those attributed to alcohol. The significant impact of cocaine extends beyond the individuals' health, negatively affecting the public health system and distorting the society (Wood, Mendelson, & Mello, 1996).

The use of cocaine has had rapid proliferation in North America and Europe since the 1970s. As a result, some institutions have emerged to evaluate and track the consequences of cocaine use on public health systems. During the 1980s, the U.S. Federal Drug Abuse Warning Network (D.A.W.N.) commenced its efforts to gather data related to cocaine, relying on statistics provided by medical facilities (National Institute on Drug Abuse, 1986). Later in the 1990s, the Canadian Community Epidemiology Network on Drug Use (CCENDU) was established to monitor drug use in major cities across Canada (Poulin, Fralick, Whunot, el-Guebaly, Kennedy, Bernstein, Boivin, & Rinehart, 1998).

Also, in the 1990s, the European Union created the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) to gather and analyze information about drugs, addiction, and their consequences (European Monitoring Centre for Drugs and Drug Addiction, 1993). In contrast to the conventional perspective, the UN founded the Office on Drugs and Crime, which plays a crucial role in gathering global data on drug cultivation, trafficking, use, and treatment (UNODC, 1997). Likewise, the Australian Institute of Criminology implemented the Drug Use Monitoring in Australia (DUMA) program in 1999. This initiative focused on collecting data on drug use from police

detainees throughout the country, aiming to establish connections between drug use and criminal activities (Australian Institute of Criminology, 1999).

B. THE COCAINE AND CRIME RELATION

Various studies have exposed the relationship between cocaine and crime. Desimone (2001) analyzed the correlation between cocaine prices and criminality in major cities across the U.S. during the 1980s and 1990s. The study elicited that higher cocaine prices corresponded to a reduction in demand and therefore crime. On this same path, Ferdinand, Blüm, and Verhulst (2001) found a strong association between delinquent behavior and drug use in male young adulthood. Another study in the U.S. found that the cocaine crack market emergence and the prevalence of firearms in African American communities could explain why the rate of homicides among black males was ten times higher than among white males (Evans, Garthwaite, & Moore, 2018). Also, Farber (2019) described how crack cocaine trade, dominated mostly by African Americans, generated a gang violence in U.S. during the 1980s and 1990s. These studies have shown a correlation between cocaine and crime in the U.S.

Other studies have shown the diverse dimensions of cocaine trafficking in Europe. Paoli, Greenfield, and Zoutendijk (2013) proposed a framework to analyze the challenges posed by cocaine trafficking in Belgium. This framework demonstrated significant associations of harm affecting individuals, institutions, and the environment. In contrast, Terenghi (2020) observed that the Italian cocaine market operates on trust-based interconnections among various entities such as families, ethnic groups, and work environments, reducing business risks and levels of violence when conflicts arise within the criminal network. This differs from the study conducted by Grassi and Sánchez-García (2021), which described a connection between cocaine and violence as portrayed in specific Italian rap lyrics, serving as a testimony to the experiences of marginalized communities.

Furthermore, cocaine use and its implications are examined in studies from the United Kingdom and Australia. In the United Kingdom (UK), a study on cocaine user-dealers concluded that these individuals found in the small trafficking a way to sustain

their consumption dependency instead of committing crimes to get money to buy drugs (Moyle & Coomber, 2015). Also in the UK, another research found that detainees testing positive for cocaine re-involved in minor crimes as prostitution and shoplifting (Pierce et al., 2015). Moving to Australia, an analysis of different data identified a correlation between the predominant increase in cocaine consumption with a rise in cocaine seizures, arrests, and related hospitalizations from 2003 to 2019 (Man et al., 2021).

Moreover, thorough investigations have shed light on the involvement of different states and regions in cocaine trafficking activities, emphasizing the reinforcement and expansion of distribution networks. For instance, Anastasijevic (2010) pointed out how cocaine trafficking was part of the organized crime in the Western Balkans countries during the first decade of this century, delaying their integration process into the European Union. Going south, Wehrey and Boukhars (2013) exposed the critical situation in Sahel and Sahara region after al-Qaeda in the Islamist Maghreb (AQIM) took control over Mali in 2012, organized crime and converted Mali in a cocaine hub for distribution towards Europe (p. 1-8). Besides, McCarthy-Jones (2018) advised about the emerging cocaine trafficking in the Indian Ocean region and the implications for Australia. In short, drug trafficking is no longer an isolated activity but a main component of organized crime (Reuter, 2014).

The dynamics of cocaine trafficking in Latin America reveal varying impacts of security policies and drug production on crime and violence. Ursin (2014) found that strong security policies reduced cocaine supply while crime remains on Brazilian streets. In contrast, Pfrimer and Motta (2021), based on cartography and literature review, demonstrated that there is a correlation between drug production and armed violence in Brazil. In Bolivia, since 2006, the presence of influential agricultural unions closely aligned with the government and their efforts to safeguard the interests of coca farmers in the Chapare region have effectively prevented the occurrence of high levels of violence that are commonly found in other drug-trafficking affected areas (Arias & Grisaffi, 2021). Nevertheless, in an attempt to gain control over cocaine trafficking routes, Brazilian criminal gangs have brought violence to the border regions of Bolivia (Ford, 2022).

Coca production and its associated challenges vary significantly across different countries in Latin America. For example, unlike Bolivia, where coca production is concentrated in the Yungas and Chapare regions, Peru has a minimum of 14 coca-growing valleys that are not well interconnected (Castillo Gallardo & Durand Guevara, 2008). On top of that, the main coca-growing areas in Peru were controlled by the insurgent group Sendero Luminoso, which deepened the society's criminal stigma towards coca growers (Durand-Ochoa, 2012; Morales, 2017). A similar situation but more complex was portrayed by Holmes, Amin Gutiérrez de Piñeres, and Curtin (2009) in Colombia, where guerrilla and paramilitary groups linked to cocaine trafficking plunged the country into a period of pervasive violence (p. 3-7).

The illegal cocaine trade serves as a concrete example of the resilience of a criminal network. Despite the dismantling of major Colombian cartels in the 1990s and the implementation of stringent security measures that significantly reduced cocaine production to historic lows in Colombia during the early 2000s (Thoumi, 2014), the trade persisted as smugglers adapted by fragmenting control into smaller trafficking groups (Echandía, 2013; Zuleta, 2022). These groups managed to sustain the business by assuming control of production and shifting the power dynamics to the Mexican cartels (Medel & Thoumi, 2014), who are now driving a new era of violence (Abi-Habib, Lopez, & Cegarra, 2022).

Also, some studies explained how force eradication policies such as aerial fumigation with glyphosates in Colombia failed because of the balloon effect (Rincón-Ruiz et al., 2022). The balloon effect refers to the phenomenon where the authorities' actions in eradicating coca crops in one area led to the displacement of crime and cultivation to other regions. Likewise, Toth and Mitchell (2018) argued how drug interdiction programs are ineffective as drug traffickers adapted by shifting their operations in time, space, and methodology to access consumer marketplaces.

Overall, the strategies employed by drug traffickers align with the dynamics explored by Duijn, Kashirin, and Sloot (2014) concerning the resilience of criminal cannabis networks after facing disruptions. This study revealed that despite the removal

of multiple actors, the impact on criminal activities was negligible, and surprisingly, the efficiency of the criminal network actually improved.

As per the UNODC, high volumes of cocaine flows persisted from Colombia, Peru, and Bolivia to Mexico, North America, Brazil, and Europe (UNODC, 2023) while moderate quantities have begun to make their way to Africa (SyndiGate Media, 2022), the Middle East (Matta Colorado, 2022a), and Oceania (Matta Colorado, 2022b). Additionally, the UNODC's findings between 2014 and 2019 indicated that while the number of cocaine seizures increased, cocaine production also rose, despite the notable efforts of global authorities (UNODC, 2021).

Other comprehensive studies have examined the associations between cocaine use and crime on a global scale. For example, Aziani (2020) provided evidence that fluctuations in the cocaine market value between 1998 and 2013 had a direct correlation with the intensity of violence in each of the 126 countries assessed. On the other hand, Jeynes (2022) carried out a meta-analysis of 75 studies that explored the potential association between drug consumption and academic and behavioral issues. The research findings suggested that drug use could be linked to challenges in academic performance and behavioral problems, as well as to violent behaviors towards others.

Perhaps, the study conducted by Miron (2001) is the most relevant research to the purpose of this thesis. His investigation highlighted violence as a prevalent way of resolving conflicts in the black market, holding that drug prohibition could increase it. This study also suggested a strong significance between cocaine base seizures and homicide rates between 1993 and 1996 in 66 countries, calling for further research using time-series data. In contrast, a case of study in Jamaica illustrated that between 2000 and 2011, as authorities seized less cocaine, the number of homicides increased (UNODC, 2012).

C. IDENTIFICATION OF COCAINE AND CRIME FACTORS

Upon reviewing the available literature concerning the issue addressed in this thesis, it was evident that various factors related to cocaine have an impact on crime, as illustrated in Figure 1.

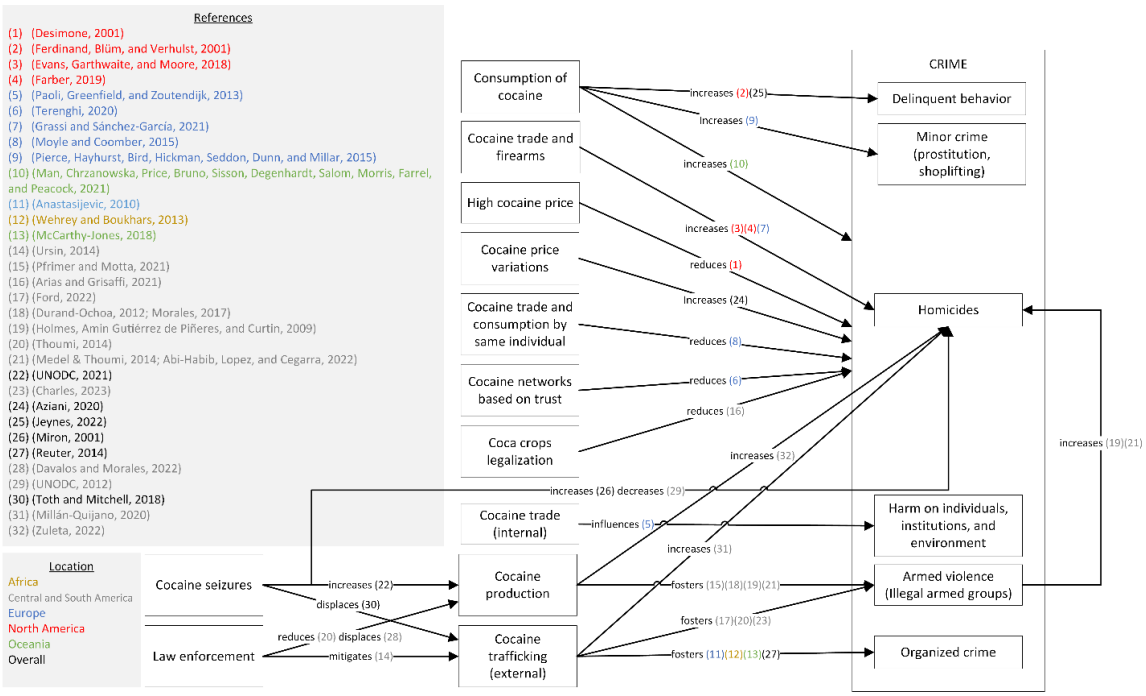


Figure 1. Cocaine factors associated to crime factors in literature review

The first cocaine factor, consumption, was found to have a positive relationship with delinquent behavior (Ferdinand et al., 2001; Jeynes, 2022), minor crimes (Pierce et al., 2015), and homicides (Man et al., 2021). The second factor, involving cocaine trade and firearms, showed a positive relationship with homicides (Evans et al., 2018; Farber, 2019; Grassi & Sánchez-García, 2021). Third, variations in prices were associated with an increase in criminality (Aziani, 2020), while high prices could reduce it (Desimone, 2001).

An intriguing fourth factor emerged when examining cases where the cocaine consumer also acted as a dealer, revealing a reduction in criminality within their area of influence (Moyle & Coomber, 2015). Similar findings were observed by Terenghi (2020) in Italy's cocaine networks based on trust, as well as in Bolivia where the government legalized coca crops (Arias & Grisaffi, 2021). Alternatively, Paoli et al. (2013) presented a framework that established the connection between cocaine trade and the resulting harm to individuals, institutions, and the environment.

From the perspective of cocaine trafficking, there was evidence highlighting how law enforcement efforts could mitigate cocaine trafficking (Ursin, 2014). Furthermore, there was compelling evidence to support the conclusion that it fostered organized crime (Anastasijevic, 2010; Wehrey & Boukhars, 2013; Reuter, 2014; McCarthy-Jones, 2018), and, in certain instances, led to the formation of illegal armed groups (Thoumi, 2014; Charles, 2022; Ford, 2022). This armed violence contributed to the escalation of homicides rates in Colombia and Mexico (Holmes et al., 2009; Medel & Thoumi, 2014; Abi-Habib et al., 2022).

However, evidence suggests that while certain law enforcement strategies may show initial success, they eventually worsen the situation. For instance, force eradication has successfully reduced cocaine production in the targeted areas (Thoumi, 2014). Unfortunately, these strategies have also led to the balloon effect, resulting in the displacement of criminal activities to other regions (Dávalos & Morales, 2022). Likewise, the seizure of cocaine displaces cocaine trafficking activities (Toth & Mitchell, 2018) and has been shown to incentivize an increase in production (UNODC, 2021). Also, it was shown how cocaine production contributed to the strengthening of illegal armed groups in countries like Colombia, Mexico, and Peru (Holmes et al., 2009; Durand-Ochoa, 2012; Medel & Thoumi, 2014; Morales, 2017; Pfrimer & Motta, 2021) and increased the homicide rates in Colombia (Zuleta, 2022).

Withal, studies on the correlation between cocaine seizures and homicide rates yielded contradictory results. While a research investigation on cocaine seizures in Jamaica revealed an inverse correlation between the amount of seizures and homicides rates (UNODC, 2012), Miron (2001) found a direct relationship. Also, Millán-Quijano (2020) claimed that drug trafficking increases homicides in Colombian territories associated to the illicit business. The outcomes of these studies are paradoxical and prompt inquiries about how the connection between cocaine seizures and homicides rate might vary across different regions.

D. COLOMBIAN PACIFIC REGION

The divergent outcomes of these studies serve as a compelling rationale for narrowing the focus of this research to the Colombian Pacific region. The Colombian Pacific is a geographically distinct region, enclosed by the formidable barrier of the Western Ranges, which contributes to its relative isolation from the rest of Colombia. This isolation has given rise to a unique and diverse landscape, characterized by a complex river network and a vast, pristine coastline along the Pacific Ocean.

The Colombian Pacific is known to be a significant corridor for drug trafficking, with cocaine often being transported through rivers to reach coastal areas for further distribution (Salazar, 2010). The vast and dense rainforests, as well as the limited presence of law enforcement in some remote areas (Castro, 2017), create opportunities for illegal armed groups to use the territories for their illicit operations (Aschner & Montero, 2021). The region's challenging geographical features continue to pose challenges for law enforcement and counter-narcotics efforts.

Moreover, the inhabitants of the Colombian Pacific region have remained trapped in a state of underdevelopment. The low-income levels, elevated illiteracy rates, lack of infrastructure, and prevalence of illicit economic activities in the region has indeed acted as a significant barrier to social advancement (Galvis-Aponte et al., 2017). As well, Guzmán and Sánchez (2021) described how the centralism of power and the counterculture fostered by drug trafficking have significantly contributed to pushing the population of the Colombian Pacific region into a state of illegality. These factors collectively created a poverty trap that impedes the progress and development of the region and its population. (Galvis-Aponte et al., 2017).

Due to these specific socio-geographical conditions, the Colombian Pacific region has emerged as the primary producer of coca bushes not only in Colombia but also globally. According to the UNODC-SIMCI (2022) report in 2021, the Pacific region held the highest position among all regions in Colombia with the largest area dedicated to coca crops. It contributed 44% of the national total, amounting to 83,266 hectares. Once more, from an international standpoint, Colombia retained its position as the top coca bush

producer in the world in 2021, with a total of 204,300 hectares out of 296,600 hectares, representing around 69% of the global production (UNODC, 2023).

Furthermore, according to Hinestroza, Sánchez, Aidar, and Palloni (2021) life expectancy on the Colombian Pacific region is significantly lower compared to the rest of the country. This disparity in life expectancy might be indicative of armed group confrontations and the prevailing criminal activity within the region. On this regard, Álvarez and González (2012) concluded that the Colombian Pacific region is characterized as a violence cluster, exhibiting a persistent high rate of homicides in around 29% of its municipalities in 2000, 2003, and from 2005 to 2010.

The definition of crime might be wide-ranging and ambiguous, but it encompasses any activity connected to organized crime. A comprehensive definition of organized crime, as Paoli and Vander Beken (2014) proposed, is a transnational, dynamic structure engaged in illegal and profitable enterprises that infiltrate various government and societal institutions. This illegal penetration leads to the generation of a wide array of other criminal activities. In that sense, the UNODC (2008) has identified different kinds of crime associated with the illegal drug business, including gang violence, money laundering, corruption, kidnapping, and human trafficking.

Though, for the scope of this thesis, the focus will be on the homicide rate in the Colombian Pacific region between 2012 and 2022. Although homicide is not an exclusive measure of crime, it serves as a reliable indicator, as it is less influenced by inconsistencies in law enforcement definitions (Fox, 2000). Additionally, the Observatory for Human Rights and National Defense (ODHDN), under the MDN, diligently monitored homicide counts, encompassing various categories that aid in facilitating investigative efforts. The homicide counts experience minimal underreporting and contain detailed information regarding the date, location, modality, and victim characteristics (Ministerio de Defensa Nacional [MDN]-Observatorio de Derechos Humanos y Defensa Nacional [ODHDN], 2023).

E. DEFICIENCIES IN LITERATURE

Literature evidenced some interesting relationships between law enforcement strategies and crime. For instance, research has revealed that the efforts of authorities can either lessen the impact of cocaine trafficking (Ursin, 2014), or lead to its relocation (Toth & Mitchell, 2018). Additionally, studies have shown that different kinds of law enforcement interventions can result in the displacement (Dávalos & Morales, 2022), reduction (Thoumi, 2014), or even escalation (UNODC, 2021) of cocaine production.

In turn, these shifts in cocaine production and trafficking have contributed to the growth of organized crime on a global scale (Anastasijevic, 2010; Wehrey & Boukhars, 2013; Reuter, 2014; McCarthy-Jones, 2018) , as well as an increase in armed violence within the countries where the drug is produced (Holmes et al., 2009; Durand-Ochoa, 2012; Medel & Thoumi, 2014; Morales, 2017; Pfrimer & Motta, 2021; Charles, 2022; Ford, 2022; Thoumi, 2014). The emergence of illegal armed groups has significantly escalated the homicide rates, particularly in countries like Colombia and Mexico (Holmes et al., 2009; Medel & Thoumi, 2014; Zuleta, 2022).

Certainly, these literary sources afford us a glimpse into how the actions of authorities, in line with government policies, disrupted the cocaine logistics, which in turn triggered a multitude of criminal activities. Notably, among these studies, only the investigations conducted by Miron (2001) and the UNODC (2012) specifically focused on the relationship between cocaine seizures (independent variable) and homicides (dependent variable), which are the main focus of the current thesis. Still, both studies arrived at contrasting findings.

The reasons for the disparities in conclusions could be attributed to various factors, such as differences in location, methodologies, data sources, sample sizes, time periods analyzed, and the specific contexts in which the studies were conducted. The disparities in findings underscore the importance of further research and analysis to gain a more comprehensive understanding of the dynamics between cocaine seizures and homicides.

For example, Miron's research in 2001 was broad, analyzing 66 countries to investigate the factors contributing to violence, considering drug prohibition and gun control in each nation. The study utilized cross-sectional data from the period between 1993 and 1996 to examine these relationships (Miron, 2001).

Conversely, the UNODC's study conducted in 2012 centered exclusively in Jamaica, relying on data from various U.S. agencies and governmental sources to estimate the levels of cocaine seizures and murders on the island between 2000 and 2011. This study provided historical insights into the return of Jamaican drug traffickers in the early 2000s and the successful measures implemented to curtail drug trafficking to Europe via air in 2002. Because of these actions, the study explained an upsurge in homicides on the streets of Jamaica (UNODC, 2012).

Hence, there is an important research gap in the existing literature when it comes to understanding how the cocaine seizures in the Pacific region influence the occurrence of homicides in the Colombian Pacific municipalities between 2012 and 2022. To fill this research gap, this study aims to use web scraping techniques to collect relevant news articles about cocaine seizures at sea and within the Colombian Pacific municipalities from available online sources. Furthermore, for data on homicides, the study will rely on the dataset made available by the ODHDN (MDN-ODHDN, 2023). Also, this study will implement Miron's suggestion of using a time-series dataset to explore the relationship between the variables of interest (Miron, 2001).

F. PHILOSOPHICAL WORLDVIEW

This is a correlation analysis following a postpositivism worldview (Phillips & Burbules, 2000 p. 1-26). The goal is to find whether the cocaine seizure law enforcement strategy is correlating with the homicides rate in the Colombian Pacific region. To find out or verify a priori assumptions of the relationships between the variables, it would be necessary to establish operational definitions for each variable, specify the measurement scale, and identify the instruments required to obtain the results.

Even though this study lacks an empirical nature, the act of filtering the information retrieved through web scraping and the homicide database available (MDN-

ODHDN, 2023) does correspond to a reductionist procedure. Besides, during the exploratory data analysis, maintaining objectivity is crucial to ensure a valid and reliable process that can be replicated consistently once the study is concluded. Overall, the probabilistic determination, reductionism, and theory verification implicit in this thesis are postpositivist assumptions that are particularly applicable to quantitative research methods (Creswell, 2014 p. 6-8).

III. METHODS

A. TYPE OF RESEARCH DESIGN

This thesis pertains to a quantitative study. According to Creswell (2014) a quantitative study should have a design that directs the investigation (p. 11-12). Furthermore, within the realm of quantitative studies, a correlational design, is employed to investigate the degree of association between two or more variables (Creswell, 2012). This quantitative study follows a correlational design to conduct an EDA of cocaine seizures and homicides datasets.

However, in a correlational design, controlling the independent variable is not feasible. As a consequence, it is not possible to establish a causal relationship but rather a correlational explanation (Cowls & Schroeder, 2015). Yet, the correlational design is known to have two inherent issues: directionality and third variable. The first problem refers to the possibility that the relationship between the variables may be inverse. The second problem suggests that there might be a third variable influencing and explaining the correlation observed between the variables under analysis (Asamoah, 2014).

Considering these problems, the research will involve applying various advanced regression techniques to develop a model that explains the association between cocaine seizures and homicides in the Colombian Pacific region. However, due to the correlational design's limitations, the study does not claim any causal between the variables but aims to elicit subtle insights through data analysis and modeling.

B. DATA COLLECTION AND PROCESSING INSTRUMENTS

This thesis comprises the use of two distinct datasets. The first dataset contains information on cocaine seizures in the Pacific influence area, including sea and land. Gathering this data would require web scraping from various websites to access news related to cocaine seizures resulting from interdictions by navies in the Pacific Sea over the past decade. On the other hand, the second dataset pertains to homicides in the municipalities of the Colombian Pacific region and has been provided by the ODHDN (MDN-ODHDN, 2023).

This study used specific hardware and software setup. The hardware consisted of a personal Lenovo IdeaPad Gaming 3 15ACH6 laptop featuring an AMD Ryzen 5 5600H processor, a Radeon Graphics 3301MHz processor, 8GB of Random Access Memory (RAM), and a 500GB Solid-State Disk (SSD). The software tools were divided into three groups: an internet web browser, big data tools, and statistical data tools. All these applications were running on the Microsoft Windows 11 Home operating system (OS). Figure 2 shows a comprehensive overview of the hardware and software versions used in the study.

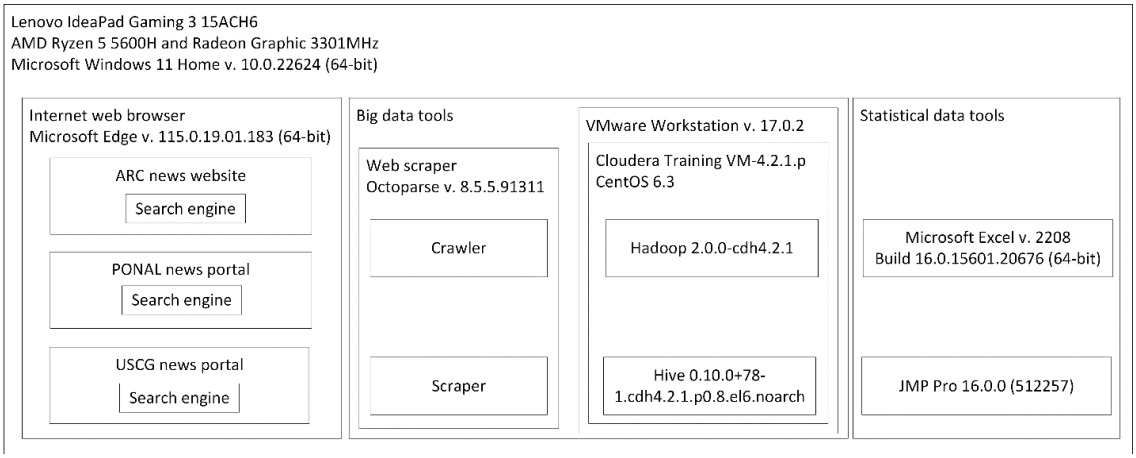


Figure 2. Data collection and processing instruments

C. INTERNET WEB BROWSER

In this study, the Internet web browser Microsoft Edge was utilized. Within Microsoft Edge, three specific URLs were accessed for the research purposes: (a) ARC news website (ARC, 2022b), (b) PONAL news website (PONAL, 2023a), and (c) USCG external news website (Bright Mountain Media, 2022). All search queries were conducted using the search engine present on each respective website.

1. Big Data Tools

This study made use of five big data tools. The initial one was Octoparse (Octopus Data Inc., 2023), employed as a web scraper to crawl websites and extract

information. The second software utilized was VMware Workstation (VMware, Inc., 2023), which served to virtualize the Cloudera Training virtual machine (VM). Within this virtual machine, the last three software pieces were installed, facilitating the operation of Hadoop and Hive applications on the CentOS operating system. The VMware Workstation used in the study was licensed by NPS, while the Cloudera Training VM was provided during the Big Data Management, Architecture, and Applications course at NPS.

Choosing Octoparse as the web scraping tool was a challenging endeavor. It involved a systematic process of exploration, testing, and decision-making. Throughout this process, various technical, legal, and ethical aspects were considered. Below is a comprehensive account of the process undertaken to make this selection.

2. Web Scraping Tools

a. Conceptual definition

Every day, the internet experiences a heightened level of dynamism because of the consistent and enthusiastic generation of websites and web pages. According to Bitkoska (2022) approximately 252,000 new websites were generated daily in 2022, while Verisign (2022) reported an estimate of 115,555 new websites per day during the second quarter of that year. Considering that each website likely contains one web page with an average size of 2MB (Kang, 2022), the Internet's growth could be projected to exceed 231GB per day.

This continuous influx of new data added to the Internet presents challenges in obtaining reliable and valuable information for analysis and decision-making. As pointed out by Van Meter (2020), the growing availability of data on the Internet may lead to a higher amount of disinformation, which can be misleading (Shabani & Sokhn, 2018; Shao et al., 2018; Zollo & Quattrociocchi, 2018; Di Sotto & Viviani, 2022). To tackle this issue, one potential solution involves the development of software designed to identify fake news on the Internet (Martens et al., 2018; Guess et al., 2019; Barua et al., 2019; Gaglani et al., 2020; Meddeb et al., 2022).

Another approach is to explore how users establish trust in websites, where factors such as service quality become significant in choosing retail websites, for example (Caruana and Ramaseshan, 2015). Likewise, for online news media, the quality of information plays a critical role in the selection process (Simanjuntak et al., 2022). By understanding the factors that influence user trust, it is feasible to enhance the user experience and improve the reliability of information amidst the vast and ever-growing volume of data on the Internet.

Having a trustable data source on the Internet is the initial and crucial step preceding the extraction and transformation of information. This idea has been present since the introduction of the Knowledge Discovery Database (KDD) model, which first phase involved carefully selecting data relevant to the topic of interest (Jaiswal and Patel, 2015). This principle remains relevant in the web scraping process proposed by Krotov and Silva (2018). In their approach, the first phase consisted in conducting a thorough analysis of the website's structure to ensure the data extracted is reliable and appropriate for the intended purpose.

Web scraping is consistently defined across various sources. The consensus among various researchers is that "web scraping is a technique used to extract data from websites and transform unstructured data into structured data" (Sirisuriya, 2015; Singrodia et al., 2019; Diouf et al., 2019). Zhao (2017) and Krotov and Silva (2018), similarly define web scraping as a method for gathering data and organizing it in a structured system to facilitate analysis. A web scraping tool comprises the use of a crawler and a scraper. A crawler is a program configured to navigate and download data from the Internet, while a scraper is a program to extract and transform the information of interest in structured data (Landers et al., 2016).

There are various web scraping techniques. In the study by Sirisuriya (2015) described some techniques from the fundamental "copy and paste" through more advanced methods. These included text wrapping, computer vision web page analysis, Hypertext Transfer Protocol (HTTP) programming, semantic annotation recognition, Hyper Text Markup Language (HTML) parsing, Document Object Model (DOM) parsing, vertical aggregation platforms, and web scraping software. Plus, Diouf et al.

(2019), included mimicry, weight measurement, differential, and machine learning approaches. On the other hand, Singrodia et al. (2019), named some other techniques such as syntactic web scraping, semantic web scraping, and computer vision web page analysis.

From all the techniques, the web scraping software was considered the simplest technique to implement (Sirisuriya, 2015). There are various options available in the market, including commercial and open-source. Diouf et al. (2019) classified these scrapers into two categories: ready-made tools and libraries for programming languages. The ready-made tools consist of web browser extensions and applications that users can readily employ for web scraping tasks. On the other hand, libraries developed for specific programming languages, such as Python (e.g., Beautiful Soup), Java (e.g., JSoup), and Node.js (e.g., Cheerio), can be configured for web scraping. Additionally, there are also frameworks, which serve as templates for creating new web scraping applications (Diouf et al., 2019).

b. Areas of application

Web scraping tools have gained popularity and are now being applied in various fields. For instance, Mendels et al. (2015) used Jsoup to retrieve data from different multilingual websites to build interpolated language models for low resources languages to improve speech recognition and keyword search technologies. Singrodia et al. (2019) mentioned data mining, research, marketing, company competition, and data combination as typical areas of web scraping usage.

Web scrapers have been used also in different contexts. Krotov and Tennyson (2018) used a web scraper developed in the "R" programming language along with the Rvest and eXtensible Business Reporting Language (XBRL) packages to extract financial data from different websites for educational purposes. Similarly, Mackey et al. (2020) mined information about COVID-19 from Instagram using a web scraper developed in Python to detect and prevent misinformation during the pandemic. Lastly, Ponmaniraj et al. (2020) scraped Uniform Resource Locations (URL) from seeded web pages to test their reliability and freshness aimed at cybersecurity implementations.

Another typical application for web scraping is data mining from on-line news websites serving various purposes. As a case in point, News One used web scraping to gather the latest news from over 100 national and international news sources (Sundaramoorthy et al., 2017). In another example, Ertam (2018) employed a web scraping machine learning approach to extract information from a Turkish news agency, creating a repository of data. Similarly, ProCircle used web scraping to extract promotions from news websites and social media (Junjoewong et al., 2018). Moreover, Maududie et al. (2018) and Sarr et al. (2018) independently applied web scraping to structure the information from Indian and Senegalese news websites, respectively. did a similar application over 15 Senegalese news websites. Also, Pande et al. (2022) used web scraping implementing Natural Language Processing (NLP) to clean the text from news gathered from websites to later detect their fakeness.

Numerous research studies have demonstrated the utilization of web scraping tools for gathering data on drug-related subjects. For example, Giommoni and Gundur (2018) used a web scraper to collect data from the website www.priceofweed.com, for their analysis of the cannabis market in the UK. A similar study conducted by Maybir and Chapman (2021) used a web scraping software, Scrape-Storm, to analyze the ecstasy's market trends. The EMCDDA also leveraged web scrapers with a machine learning approach to collect information from the internet for its projects eDrugTrends and CASSANDRA (Raubenheimer and Barratt, 2018). The U.S. National Institute of Health and National Institute on Drug Abuse (NIH/NIDA) supported the use of a similar technology in the eDarkTrends project to gather information from the dark net (Raubenheimer and Barratt, 2018). Plus, Li et al. (2019) went further building a web scraper in Python to characterize drug dealers on Instagram. Their research was extended to achieve an unsupervised machine learning approach model to scrape drug dealing comments from Instagram too (Shah et al., 2022).

Upon examining various domains where web scraping tools were applied, it becomes evident that one common practice is data mining digital news websites and extracting information from open sources to investigate illicit drug-related subjects. The successful application of web scraping techniques in previous studies, where data was

mined from digital news websites and gathered from open sources, serves as validation for its convenience as a valuable tool in collecting data on cocaine seizures for the current study.

c. Legal and ethical considerations

The use of web scraping tools can be limited by technical and legal factors. These might include the intricacy of the selected web scraping tool, legal restrictions set by website owners, and adherence to ethical principles during data collection from open sources on the Internet. An instance illustrating this challenge is attempting to use a web scraping tool on a website that explicitly forbids such activity.

Snell and Menaldo (2016) acknowledged the existence of a gray area in the legal domain. They identified five legal situations associated with web scraping: (a) copyright infringement, understood as an intellectual property violation as addressed in the Digital Millennium Copyright Act of 1998 (U.S. Copyright Office, 1998); (b) breach of contract, understood as the terms of use specified by the website owner (Dreyer & Stockton, 2013); (c) violation of the Computer Fraud and Abuse Act (CFAA), understood as the illegal use of the data scraped (Krotov & Silva, 2018), or the re-incidence in accessing to the website after the reception of a cease and desist letter (Snell & Menaldo, 2016); (d) trespass to chattels, understood as an unauthorized activity and consequent damage of the website's owner property (Goldman, 2013); and (e) hot news misappropriation, understood as the attribution of breaking news produced by other agencies (Snell & Menaldo, 2016).

There are some studies concerning the ethical field. Krotov and Silva (2018) identified three implications for the web scraping activity: (a) the unintended revelation of protected personal information (PPI) during the web scraping process; (b) the revelation of industrial secrets from organizations during the web scraping process (Ives & Krotov, 2006); and (c) the possibility of bypassing the commercial activity that supports the profit of the website owner during the web scraping activity (Hirschey, 2014). Figure 3 summarizes the ethical considerations and legal constraints.

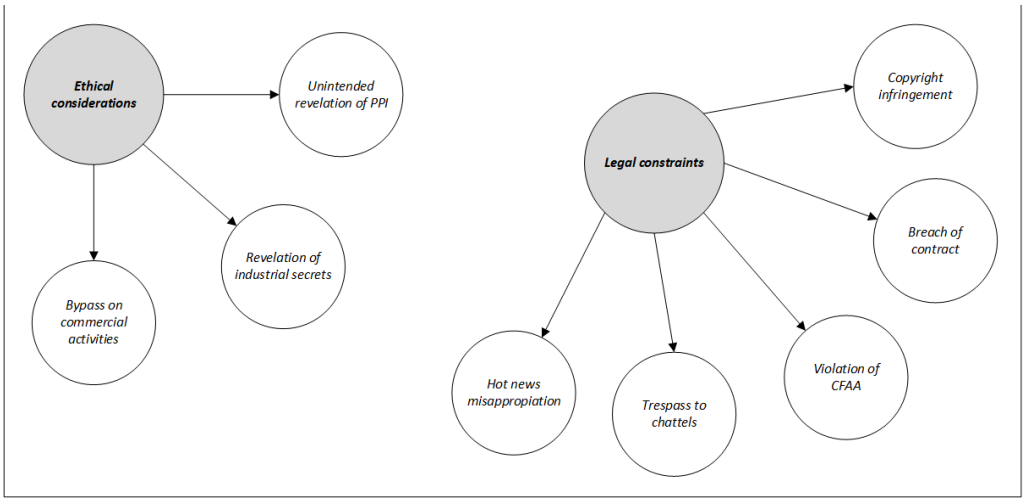


Figure 3. Ethical considerations and legal constraints of web scraping

Regarding the legal factors, Snell and Menaldo (2016) concluded that a website owner can improve the protection against web scraping activities in different ways: (a) addressing and forcing the recognition of the prohibition or limitation in using these technologies in the website's terms of service/use; (b) implementing technological tools to avoid the use of the website by a machine; (c) protecting access to the website; and (d) managing the website content use through registers, licenses, or authorizations. Regarding the technological tools to avoid the use of the website by a machine, Zhao (2017) identified three major procedures: (a) HTML headers identification; (b) Internet Protocol (IP) reputation analysis; and (c) pattern behavior analysis.

In contrast, Krotov and Silva (2018) proposed a survey for the web scraping user before starting the activity to be sure that: (a) web scraping is allowed; (b) there are not copyrights on the website data; (c) the project is for legal and genuine purposes; (d) the activity will not damage any website owner's property; (e) the extracted data will not reveal any personal private information nor any organizational secret; and (f) web scraping will not affect the organization's profit for extracting data instead of accessing the website content.

Although the legal and ethical suggestions about the use of web scraping tools, the absence of a well-defined legal and ethical framework could present a challenge for the web scraping user. Taking all these factors into account can impact an individual's

decision-making when it comes to using a web scraping tool for data collection. Some individuals may opt not to collect information due to the tool's complexity, concerns about potential legal violations, or adherence to ethical principles that discourage such actions.

This individual's dilemma might be related to the Theory of Planned Behavior (TPB) (Ajzen, 1991) and the Theory of Reasoned Action (TRA) (Ajzen & Fishbein, 1975). The beliefs and evaluations explored in the TRA appear to align with the individual's personal ethical considerations, while the subjective norm aligns with collective ethical considerations (Vallerand et al., 1992). Both factors play a significant role in shaping the individual's behavioral intentions as depicted in Figure 4.

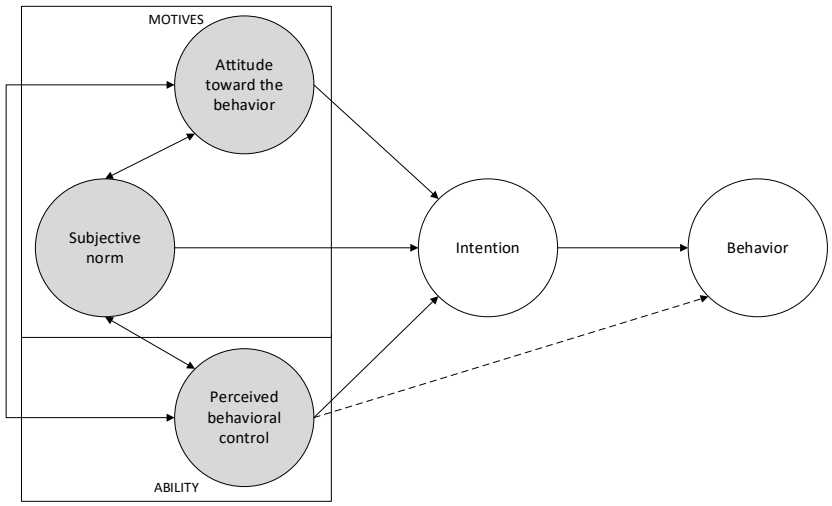


Figure 4. Theory of Planned Behavior. Source: Ajzen (1991).

Thus, the individual's ethical alignment with their organization might play a crucial role in determining their intended behavior. Despite the subjective norm influencing their behavior, the individual might choose to act based on their own beliefs, even if their attitude toward the behavior is negative. However, it is essential to note that mere alignment with ethical beliefs and external influences is not sufficient. The individual's ability to perform the behavior is equally vital, as it directly impacts the

likelihood of the intended behavior. Figure 5 depicts these new elements added to the TPB as a proposal to describe the user's behavior flow when scrapping open websites.

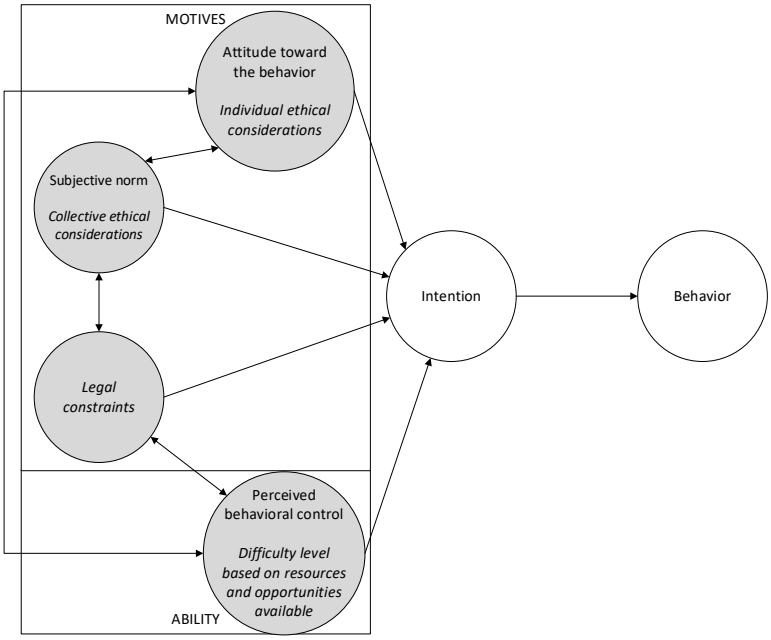


Figure 5. User's behavior flow when scraping open websites

Finally, formal collective and individual ethical guidelines will not be established for this thesis given that the open web encompasses websites accessible through standard browsers like Firefox, Chrome, or Edge. The websites targeted by this study must be in the open web domain without involving actions that bypass commercial activities, disclose PPI, or reveal trade secrets. Thus, web scraping the open web will be inherently regarded as both individually and collectively ethical by default.

d. Identification of web scraping tools

To identify the tools for web scraping used in previous researches, an exhaustive review of the literature related to this topic has been done. Table 1 shows a concise overview of reliable web scraping tools sourced from multiple academic papers

(Sirisuriya, 2015; Mendels et al., 2015; Milev, 2017; Zhao, 2017; Singrodia et al., 2019; Maybir and Chapman, 2021), including JMP[1] too.

Table 1. Identified commercial and open-source web scraping tools.

No.	Tool	Category		Trajectory	URL
		Commercial	Open-source		
1	80legs	X		2009-2022	https://80legs.com/
2	Beautiful Soup		X	2004-2022	https://www.crummy.com/software/BeautifulSoup/
3	Dexi.io	X		2015-2022	https://www.dexi.io/
4	Easy Web Extract	X		2010-2019	http://www.webextract.net/
5	Fminer	X		2009-2015	http://www.fminer.com/
6	Helium Scraper	X		2010-2022	https://www.heliumscraper.com/
7	Heritrix		X	2004-2022	https://heritrix.readthedocs.io/en/latest/
8	Import.io	X		2012-2022	https://www.import.io/
9	JMP	X		1989-2022	https://www.jmp.com/en_us/home.html
10	jsoup		X	2009-2022	https://jsoup.org/
11	Mozenda	X		2008-2022	https://www.mozenda.com/
12	Nutch		X	2004-2022	https://nutch.apache.org/
13	Octoparse	X		2016-2022	https://www.octoparse.com/
14	Out Wit Hub	X		2007-2020	https://www.outwit.com/
15	Parse Hub	X		2013-2022	https://www.parsehub.com/
16	Pyquery		X	2013-2020	https://pypi.org/project/pyquery/
17	Scrape Storm	X		2019-2021	https://www.scrapestorm.com/
18	Scrapinghub Zyte (Splash)[a]		X	2007-2022	https://www.zyte.com/
19	Screen-scraper	X		2002-2020	https://www.screen-scraper.com/
20	Selenium		X	2004-2022	https://www.selenium.dev/
21	Spinn3r	X		2005-2007	https://www.spinn3r.com/
22	urllib2		X	1990-2020	https://docs.python.org/2/library/urllib2.html
23	Visual Web Ripper Content Grabber[b] Sequentum Enterprise[b]	X		2014-2022	https://www.contentgrabber.com/
24	Visual Scraper	X		?[d]	http://www.visualscraper.com/
25	Web Content Extractor	X		2004-2022	https://www.newprosoft.com/
26	Web Scraper	X		2017-2022	https://webscraper.io/
27	Webhose.io Webz.io[c]	X		2016-2022	https://webz.io/
28	Web Sundew	X		2005-2022	https://websundew.io/
29	Zyte (Scrapy)		X	2008-2022	https://scrapy.org/

[a] Scrapinghub evolved to Splash, an open-source project maintained by Zyte.

[b] Visual Web Ripper evolved to Content Grabber, and this in turn evolved to a commercial low-code web data extraction software called Sequentum Enterprise.

[c] Webhose.io evolved to Webz.io, a commercial software to transform web data into structured data feeds.

[d] Visual Scraper website is down. It was not possible to determine software trajectory in the market.

[1] JMP is a computer program for statistical analysis.

Upon identifying 29 web scraping tools, a filtering process was initiated based on their respective track records. Apart from Visual Scraper, which was discarded due to the unavailability of its website, the final year of support served as a determining factor to assess the obsolescence of each tool. Table 2 shows the available commercial tools, introducing a new attribute (modality of service), while Table 3 shows the available open-source tools, incorporating a new attribute (programming languages). These attributes were crucial for proceeding to the next filter phase.

Table 2. Available commercial web scraping tools.

No.	Tool	Trajectory	Modality of service			
			Application	Extension	Data collection	Web Application
1	80legs[a]	2009-2022				X
2	Dexi.io[b]	2015-2022				X
3	Helium Scraper	2010-2022	X			
4	Import.io	2012-2022			X	
5	JMP	1989-2022	X			
6	Mozenda[b]	2008-2022				X
7	Octoparse	2016-2022	X			
8	Parse Hub	2013-2022	X			
9	Scrape Storm	2019-2022	X			
10	Sequentum Enterprise[c]	2014-2022	X			
11	Web Content Extractor	2004-2022		X		
12	Web Scraper	2017-2022	X			
13	Webz.io	2016-2022			X	
14	Web Sundew	2005-2022	X			

[a] 80legs requires credit card information to access the web application. This alternative was discarded.

[b] Dexi.io and Mozenda.io are similar web scraping web applications developed by Dexi. For this study, Dexi will be tested.

[c] The price for an annual subscription to the "Starter" version of the Sequentum Enterprise tool is $15,000. This price is out of budget; therefore, this alternative was discarded.

The available commercial web scraping tools were categorized into four modalities of service: (a) application, which refers to a desktop executable program compatible with operating systems like Windows 10/11; (b) extension, a small program running in the Chrome web browser (Google Developers, 2021); (c) data collection, a service where customers provide data requirements, and the company offers a quote for a customized solution; and (d) web application, an executable program accessible through a web browser and hosted on a remote server. Apart from 80 legs, Mozenda, and

Sequentum Enterprise (refer to Table 2 footnotes), the data collection options Import.io and Webz.io were excluded because a customized solution was unrealistic for the scope of this thesis.

Table 3. Available open-source web scraping tools.

No.	Tool	Trajectory	Programming language	
			Java	Python
1	Beautiful Soup	2004-2022		X
2	Heritrix	2004-2022	X	
3	jsoup	2009-2022	X	
4	Nutch	2004-2022	X	
5	Zyte (Splash)[a]	2007-2022	X	
6	Selenium	2004-2022	X	
7	Zyte (Scrapy)[a]	2008-2022		X

[a] Zyte is the company supporting the maintenance for Splash and Scrapy in two different programming languages.

The two programming languages used to collect data on the web with the identified open-source tools were Java and Python. The main difference between each open-source tool lies in how to manage a specific library, which still requires some programming knowledge in the Java or Python languages. Considering the author's coding skills in Python, only one Python web scrapping tool out of the two available was tested for this study. Scrapy (Zyte, 2022) was chosen over Beautiful Soup (Richardson, 2022) due to its better documentation, tutorials, and support on the Windows operating system.

Following an evaluation of the pricing information for each remaining commercial option listed in Table 2, Table 4 shows an overview of the final web scraping tools, both commercial and open-source options, selected for testing.

Table 4. Commercial and open-source web scraping tools for testing.

No.	Tool	Version	Price information	
			USD	Periodicity
1	Dexi.io	Intro	Quotation	Annual subscription
2	Helium Scraper	Basic	$ 99.00	One-time fee
3	JMP	Pro 16	Free license in NPS	One-time fee
4	Octoparse	Standard	$ 89.00	Monthly subscription
5	Parsehub	Standard	$ 189.00	Monthly subscription
6	Scrape Storm	Professional	$ 49.99	Monthly subscription
7	Web Content Extractor	Unique	$ 70.00	One-time fee
8	Web Scraper[a]	Professional	$ 100.00	Monthly subscription
9	Web Sundew	Standard	$ 399.00	One-time fee
10	Scrapy (Python)	2.7.1	Free	One-time fee

[a] Web Scraper is a commercial extension for Chrome web browser.

e. Websites to scrape

Following the procedure to comply with legal constraints when scraping open websites, three websites were chosen. All align with the characteristics of the open web domain. The first website was the ARC news portal (Armada de Colombia [ARC], 2022). The second website was the open United States Coast Guard (USCG) news website, supported by Bright Mountain Media (2022). The third website was the Colombian Police (PONAL) news portal (Policía Nacional de Colombia [PONAL], 2023a). Appendixes A, B, and C provide the detailed terms of use, copyright, privacy, and access policies for all websites, respectively.

For the purpose of testing, the ARC and USCG news websites were selected. Web scraping tools were utilized, using the websites' search engines, to find related news articles based on two search queries: (a) a search of the word *cocaína* in the ARC news portal (ARC, 2022); and (b) a search of the word *cocaine* in USCG (Bright Mountain Media, 2022). Appendixes D and E contain the anatomy of the ARC and USCG news websites at the first and second levels of scraping, respectively. For both cases, the scraping areas of interest to test were the titles, dates, and news content.

Both news websites presented different complexities during the scraping process. The ARC news website required a more complex crawling solution due to the presence of two-page iterations in both the first and second scraping levels. In contrast, the USCG news website case was simpler as it only involved one page iteration at the first scraping

level. To move from the first to the second level on the ARC website, it was necessary to extract the URLs of the titles to access and extract the corresponding news content.

The structure of both websites posed unique challenges when extracting the required information. For instance, each search webpage on the ARC news website had ten titles and dates, while the USCG search webpage had nine titles and dates. On the ARC news website, it was essential to extract dates in the first level of scraping because in certain cases, the date information was not accessible at the second scraping level.

Although the chosen web scraping tools were tested on both websites, variations in the results might occur due to variation in the testing conditions. Some of the web scraping tools were installed on virtual machines at the Naval Postgraduate School (NPS) servers, while others were installed on a personal laptop. As a result, minor discrepancies in time calculations arose due to network delays and disparities in computer processing capabilities.

f. Testing of web scraping tools

Before conducting the web scraping tool testing, it was crucial to establish the essential and additional features to evaluate the performance of each tool. The mandatory features are: (a) automatic scraping; (b) automatic crawling; and (c) deep scraping. Partial or complete failing of the mandatory features resulted in the tool's rejection. The optional features were: (a) level of complexity; (b) implementation time; (c) scraping time; (d) human behavior capabilities; (e) language fidelity; and (f) NPS network accessibility.

To establish a measurement, operationalization of the mandatory features is necessary. For example, for the automatic scraping feature, it means that the tool can systematically extract all titles, dates, and news content in a well-organized and logical manner from both websites. The automatic crawling feature implied that the tool could navigate to each web page indexed on the search page and effectively handle page iteration issues at the first scraping level. The scraping depth feature indicated that the tool could address the page iteration problem at the second scraping level and successfully extract the entire news content.

A similar process was undertaken for the optional features. The level of complexity was categorized into two groups: (a) code for open-source tools; and (b) graphic user interface (GUI) for commercial tools. Each category was further divided into three subcategories: easy, medium, and high. The easy subcategory scored three points when the total implementation time was 200 minutes or less. The medium subcategory scored two points when the total implementation time was between 200 and 400 minutes, inclusive. Lastly, the high subcategory scored one point when the total implementation time exceeded 400 minutes.

Moreover, the operationalization of total implementation and scraping times is required. Total implementation time encompassed the minutes spent attending tutorials and implementing the scraping solution for both news websites. On the other hand, scraping time referred to the minutes spent by the tool while scraping 100 web pages on the ARC news website and an additional 100 web pages on the USCG news website.

The human behavior capabilities operationalization involved three functions. Firstly, the assessment of IP address rotation, indicating the tool's ability to automatically change the IP for each new web page query, emulating different clients' actions on the server side, scored one point if present and zero if not. Secondly, the CAPTCHA solution was evaluated, scoring one point if the tool successfully solved CAPTCHA programs and zero if not. A sub-categorization of CAPTCHA solution was performed, with manual solution scoring one-point, partial solution (capable of solving image puzzles automatically) scoring two points, and automatic solution scoring three points. Lastly, tools equipped with a configurable fixed delay in milliseconds before scraping each web page received one point, while others scored zero.

Considering that the websites for this research were in Spanish or English, a language fidelity capability for each language was included. The Spanish language fidelity was divided into three subcategories: (a) low; (b) medium; and (c) high. A tool is classified in the low subcategory scored one point when the results showed HTML coding traces and weird codes to Spanish accents and/or punctuation. For medium subcategory, it scored two points when the results showed HTML coding traces or weird codes to Spanish accents and/or punctuation. Finally, for high subcategory, the tool

scored three points when the results were legible and sharp. The subcategorization for English was similar, with the exception that no accents were considered for the assessment.

The last capability, NPS network accessibility, is concerned with whether the web scraping tool is blocked by the NPS servers. If the tool was not blocked, it scored one point; otherwise, it scored zero. If the tool was not blocked, it received one point; otherwise, it received zero. This feature was crucial in determining the required resource for the thesis, whether a virtual machine or a personal laptop.

Table 5 summarizes the final scores and times for the 10 web scraping tools tested, while Appendix F contains detailed information about the implementation and assessment results for each web scraping tool.

Table 5. Web scraping tools: Final scores and times assessment.

No.	Tool	Mandatory features	Final score	Implementation time [min]	Scraping time [min]
1	Dexi.io	NO	10	454.73	22.15
2	Helium Scraper	NO	15	188.70	14.47
3	JMP	NO	6	30.00	83.33
4	**Octoparse**	**YES**	**14**	**183.30**	**35.50**
5	**Parsehub**	**YES**	**13**	**101.77**	**44.18**
6	**Scrape Storm**	**YES**	**11**	**357.30**	**25.42**
7	Web Content Extractor	NO	12	135.73	26.95
8	Web Scraper[a]	NO	10	324.77	85.06
9	Web Sundew	NO	9	450.00	16.98
10	Scrapy (Python)	NO	5	540.00	0.99

g. Selection of web scraping tool

After conducting the web scraping tool testing, certain limitations were identified. For commercial alternatives, a common limitation was the lack of an intuitive GUI, and in some cases, there was insufficient detailed tutorials and documentation to achieve successful web scraping results, especially when dealing with the ARC news website where customized commands were needed to address the page iteration problem at the second scraping level. Another minor limitation worth mentioning was the occurrence of semantic issues during the extraction process, particularly for the Spanish language.

On the other hand, for open-source web scraping tools, the primary limitation was the complexity in the implementation process, as it required certain knowledge in Java or Python programming languages to accomplish effective scraping of the websites.

In conclusion, among the ten web scraping tools tested, only three commercial tools (Octoparse, Parsehub, and Scrape Storm) fulfilled the mandatory criteria and were deemed suitable for use in this research. These three tools achieved final scores of 14, 13, and 11 points, respectively. Scrape Storm had the highest implementation time but the lowest scraping time on both websites, while Parsehub had the lowest implementation time but the highest scraping time. Octoparse appeared to be the most balanced option. However, both Octoparse and Scrape Storm websites were blocked by the NPS firewall. Thus, Octoparse installed on a personal laptop was chosen as the web scraping tool to collect cocaine seizures data for this study.

3. Statistical Data Tools

The last tools used in this study were the statistical data tools. Firstly, Microsoft Excel was used to parse, organize, and conduct preliminary statistic tests on datasets. Subsequently, JMP was employed for EDA. Wherein various polynomial regression techniques were applied to determine the final regression models.

D. DATA PROCESSING

Figure 6 outlines seven data processing phases for this study. The process began with web searches, followed by web scraping using Octoparse. After ingestion of structured data into Hadoop, extract, transform, and load (ETL) processes were implemented to store a preliminary dataset for cocaine seizures in the Hive warehouse. This draft file and the homicides' dataset were combined and processed in Excel using formulas and filters to create synchronized datasets. JMP (JMP Statistical Discovery LLC, 2023c) was then used for EDA to derive final regression models.

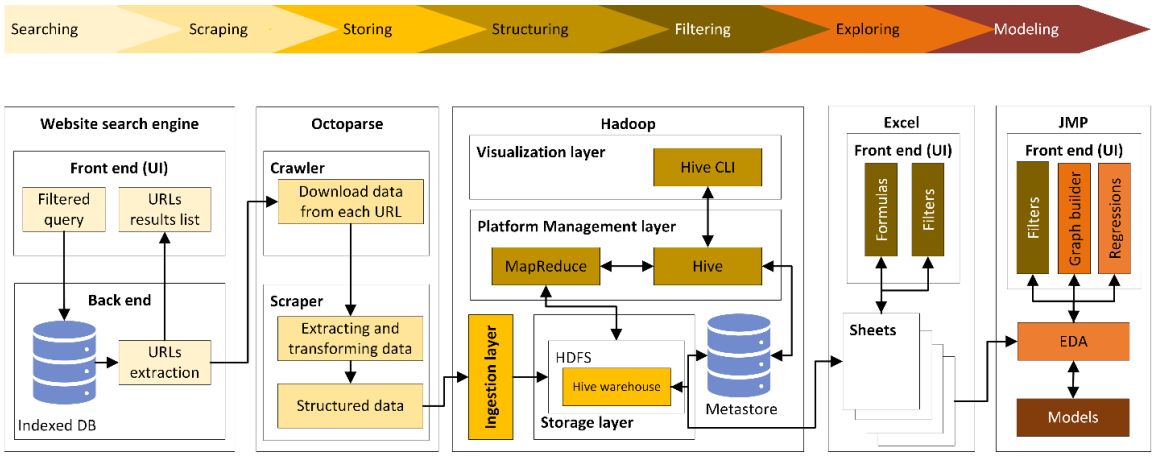


Figure 6. Data processing

Using the Microsoft Edge web browser, the searching phase started using the websites' search engines to find related news articles, based on three specified search queries: (a) a search of the word *cocaína* in the ARC news portal (ARC, 2022); (b) a search of the word *cocaína* in the PONAL news portal (PONAL, 2023a); and (b) a search of the word *cocaine* in USCG (Bright Mountain Media, 2022). Each search query yielded a list of URLs containing relevant news articles with the respective keywords. At the same time, each list of URLs was linked to an exclusive URL for individual reference and access.

The exclusive URL served as a requirement to generate a task in Octoparse. Upon task creation, Octoparse semi-automatically recognized the website fields (clicking) in both the initial and subsequent scraping levels. To achieve this, a workflow was implemented to crawl each URL from the list, followed by an ETL process, which converted the data from HTML into a table format with customized fields. This resulting table was then exported to an Excel file. For comprehensive information about the exclusive URL, automatic field detection, and the Octoparse workflows for each news website, please refer to Appendix G.

Once the structured data was ready, the next step involved ingesting and storing it into the Hadoop Distributed File System (HDFS). HDFS is a fundamental component within the Hadoop ecosystem, known for handling Big Data. This filesystem can store

files in the order of Peta Bytes (PB), working under the concept of the write-once/read-many-times pattern. One of its advantages is its compatibility with commodity hardware, leading to cost-effective implementations and allowing for a greater number of nodes to enhance redundancy in the system (White, 2015, p. 43-54). The data extracted from each website using Octoparse is enclosed in an Excel file located in Appendix H.

After storing the data in the HDFS, Hive was utilized to conduct another ETL process to obtain the desired information regarding cocaine seizures. To do that, Hive allows users to perform queries through a Command-Line Interface (CLI), a web interface, or a server infrastructure using Structured Query Language (SQL) alike syntax. The commands go to the Hive's execution engine to submit mapping and reducing jobs for the batch query processor known as MapReduce. MapReduce executes the jobs using the data available in the HDFS and the metadata, organized in tables, in a metastore denominated Hive warehouse (White, 2015, p. 471–480). The three Hive scripts used for ETL of cocaine seizures information can be found in Appendix I, while the draft cocaine seizure datasets are included in Appendix J.

Once the draft cocaine seizures data was brought into Excel, several filters and operations were implemented. The resulting outcome, as illustrated in Figure 7, represents the step-by-step procedure carried out on each cocaine seizures dataset, ultimately culminating in the creation of the Pacific cocaine seizures dataset spanning from 2012 to 2022, with daily records.

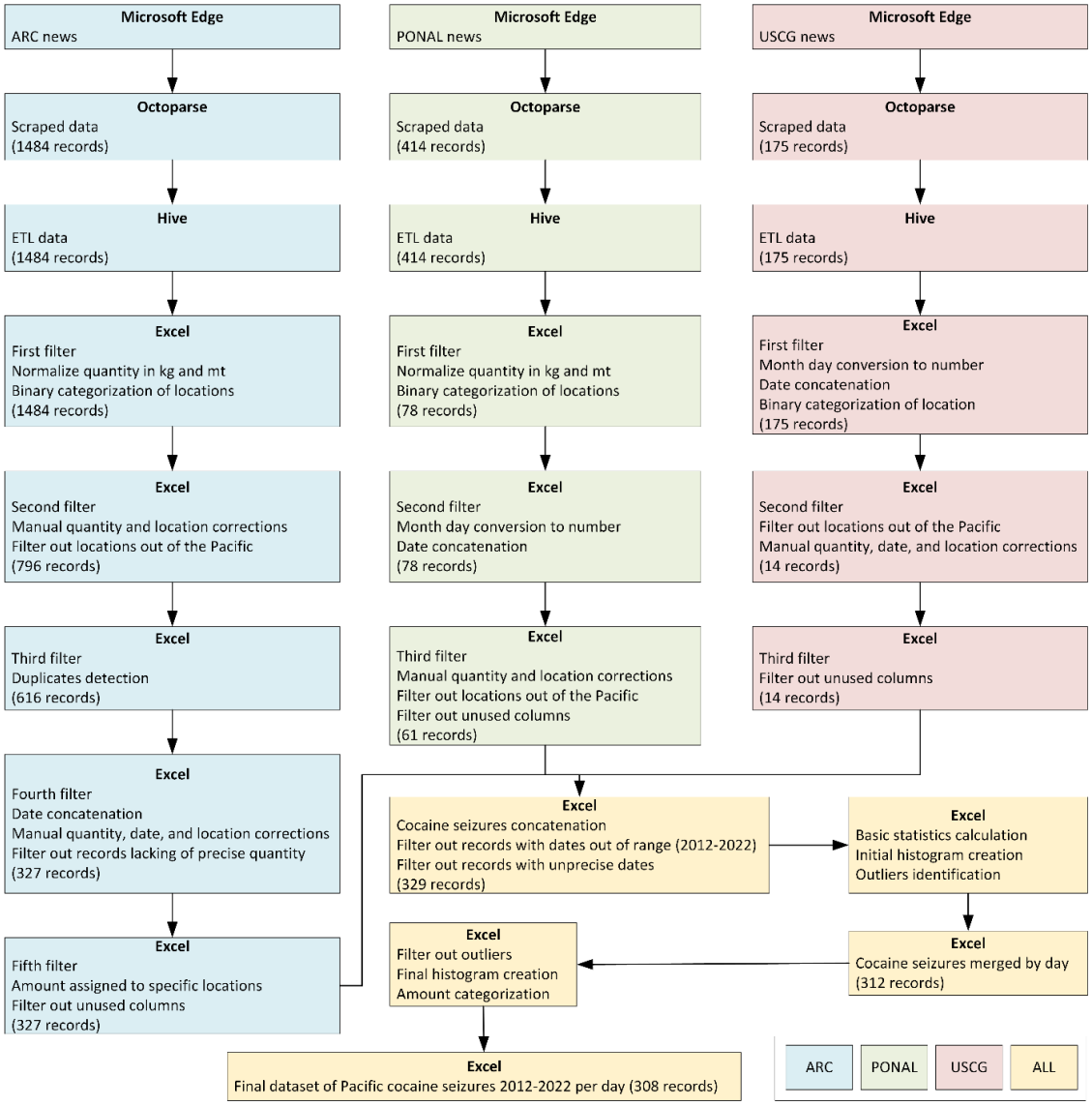


Figure 7. Filters and processing on the dataset for cocaine seizures in Excel

A detailed inspection of the day of the interdiction occurrence in the Pacific cocaine seizures dataset was of utmost importance, as it served as the starting point for assessing its influence on homicides in the Colombia Pacific region over the course of one to seven days following the event.

On the other hand, the dataset for homicides in Colombia underwent a more straightforward processing approach. Utilizing the Excel file provided by ODHDN (MDN-ODHDN, 2023), direct filtering and processing were carried out without the need

for employing big data tools. The resulting product, depicted in Figure 8, showcases the detailed sequence of operations performed on the dataset for homicides. This ultimately led to the development of the Colombian Pacific homicides' dataset, from 2012 to 2022, with daily records. The respective Excel workbooks containing the various filters and processes applied to both the cocaine seizures and homicides datasets can be found in Appendix K.

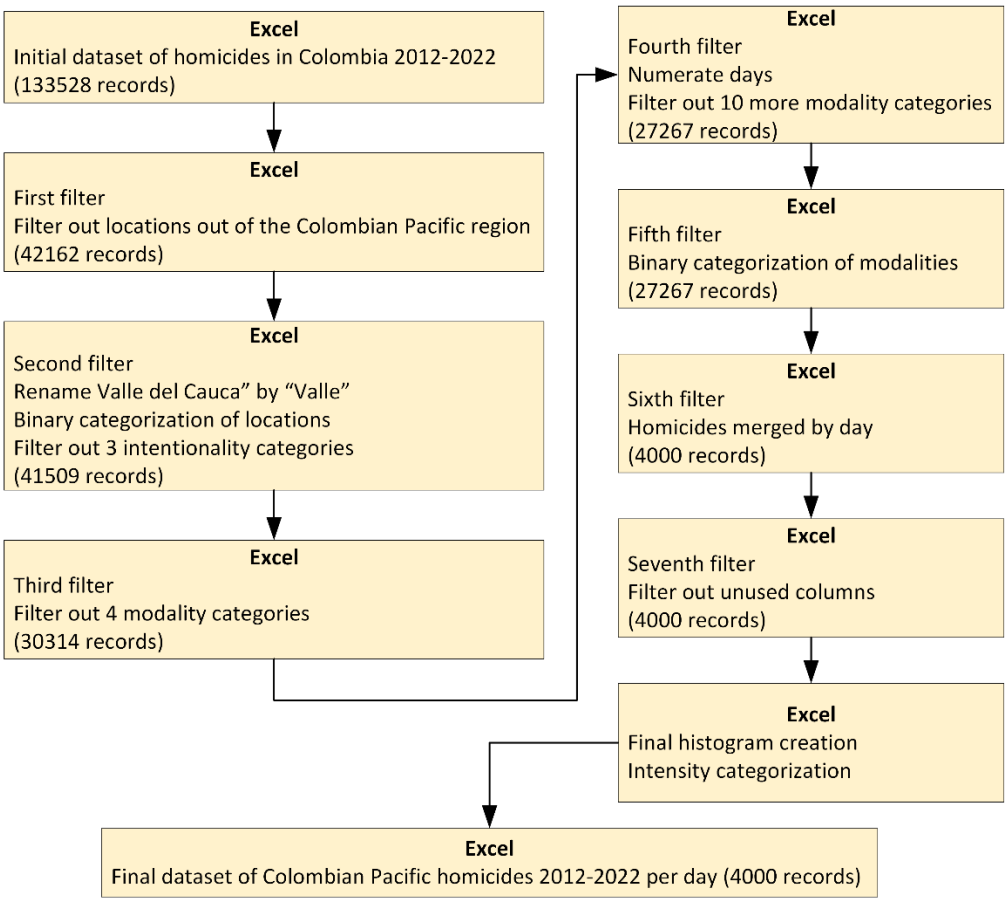


Figure 8. Filters and processing on the datasets for homicides in Excel

Although the initial number of records in the homicides' dataset was 4000, the resulting sample size for the time series data turned out to be smaller. For the purpose of this study, only seven time series datasets could be created, each containing a sample size of 308 records. This constraint was a direct result of the sample size limitations imposed

by the final cocaine seizures dataset. On top of that, during the EDA performed in JMP, this sample size was drastically reduced to less than 100 records due to the application of additional filters.

The limited size of this sample may not be sufficient to confidently assert a correlation between cocaine seizures and homicides in the Colombian Pacific region. To establish a reliable correlation, a sample size of around 250 records would be required (Schönbrodt & Perugini, 2013). However, the study conducted by MacCallum, Widaman, Zhang, and Hong (1999) concluded that specific rules of thumb for sample size were not of great importance when it comes to factor analysis. Instead, they suggested that even with smaller sample sizes, exploratory studies, such as the one presented here, can still be sufficient and meaningful in providing valuable insights and initial findings.

Taking this into consideration, an EDA was conducted in JMP using the seven available time series datasets. A digital version of these datasets can be found in Appendix L. Furthermore, Chapter IV of the present study gives a comprehensive overview of both the preliminary and final results of the EDA. This chapter provides descriptions of the analysis process, offering constructive comprehensions and preliminary outcomes of the relationships between cocaine seizures and homicides in the Colombian Pacific region.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

A. TIME-SERIES DATASETS

The data is contained in seven time-series sets, each containing 308 records with nine columns. The initial four columns pertain to cocaine seizure information, while the remaining five columns hold data on homicides. The variation across datasets lies in a single column, representing the day of homicides occurrence. In the first dataset, homicides are evaluated one day after the seizure; in the second dataset, two days after; and so on until the last dataset, which assesses homicides seven days after the seizure. The descriptions of datasets and columns can be found in Figure 9.

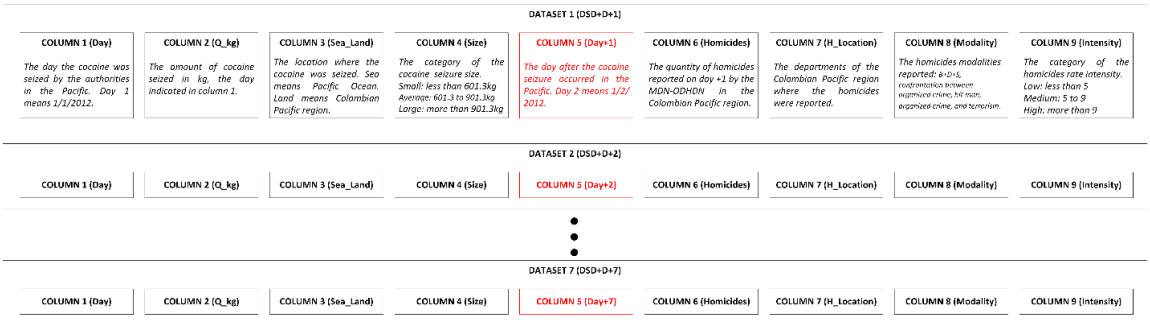


Figure 9. Time-series datasets description

The time and quantity of cocaine seizures are the crucial pieces of information found in the first two columns of the dataset. The first column of the dataset represents the date when authorities conducted operations that resulted in cocaine seizures in the Pacific. In this column, a value of "1" corresponds to 1st January 2012, and subsequent numbers represent consecutive dates following that day. The second column reflects the total quantity of cocaine seized in kilograms on each respective day mentioned in the first column.

The following two columns contain nominal data for location and the size of cocaine seized. The third column specifies the category for seizure location, using "Sea" to represent the Pacific Ocean and "Land" for any department in the Colombian Pacific

region. Column four embodies the seizures intervals; there are three intervals that represents the size of the cocaine seizures: "Small" represents any seizures that involves less than 601.3kg of cocaine, "Average" represents the interval for the seizures between 601.3kg and 901.3kg of cocaine, and "Large" for seizure quantities exceeding 901.3kg of cocaine. This categorization is shown in Figure 10 from the histogram related in Appendix K.

Median	Mean	Mode
600	736.23	600

n	# classes	MAX	min	Range	Bin width
308	10	3000	1.29	2999	300

Classes	min	MAX	Interval	Freq.	Cum.	Seizures
1	1	301	1.3-301.3	85	85	Small
2	301	601	301.3-601.3	74	159	
3	601	901	601.3-901.3	57	216	Average
4	901	1201	901.3-1201.3	34	250	
5	1201	1501	1201.3-1501.3	19	269	
6	1501	1801	1501.3-1801.3	17	286	
7	1801	2101	1801.3-2101.3	6	292	Large
8	2101	2401	2101.3-2401.3	10	302	
9	2401	2701	2401.3-2701.3	3	305	
10	2701	3000	2701.3-3000	3	308	

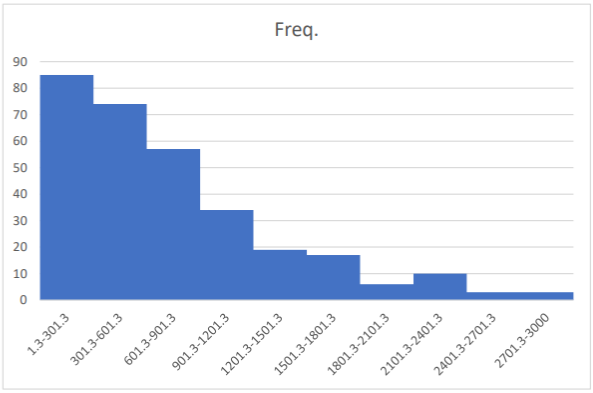


Figure 10. Cocaine seizures in the Pacific 2012–2022 histogram

The information related to homicides begins from column five, which is particularly crucial since it has a daily variation from one to seven days across all datasets. For example, in the first dataset, this column represents the day immediately following the cocaine seizure in the Pacific, while in the seventh dataset, it represents seven days after the seizure. In the first case, the primary objective is to assess the impact of cocaine seizures on the homicides rate in the Colombian Pacific region on the day right after the seizure. In the second scenario, the focus changes to evaluating this impact after a seven-day period following the seizure. For the intervening cases, the evaluation similarly shifts to corresponding days after the occurrence of the cocaine seizure.

Column six pertains to the rate of homicides, whereas column seven is related to the location of these incidents. The sixth column shows the overall count of homicides documented in the Colombian Pacific region for the specified date indicated in column five. Conversely, the seventh column specifies the departments where these homicides

were recorded. It is essential to note that the sixth column offers quantitative data for analysis, while column seven furnishes nominal data regarding the location.

Figure 11 shows column eight, which contains information about the homicide modalities. From the original dataset, seven modalities were recognized as closely associated with homicides resulting from illegal cocaine businesses in Colombia. The most prevalent modality on the list is the "hit man" with 26,471 instances, followed by organized crime-related homicides with 662 occurrences, organized crime confrontations with 210 cases, and terrorism with 109 incidents. The three remaining modalities of beheading, dismemberment, and strangulation (B+D+S), were merged into a single category, resulting in a total of 253 cases.

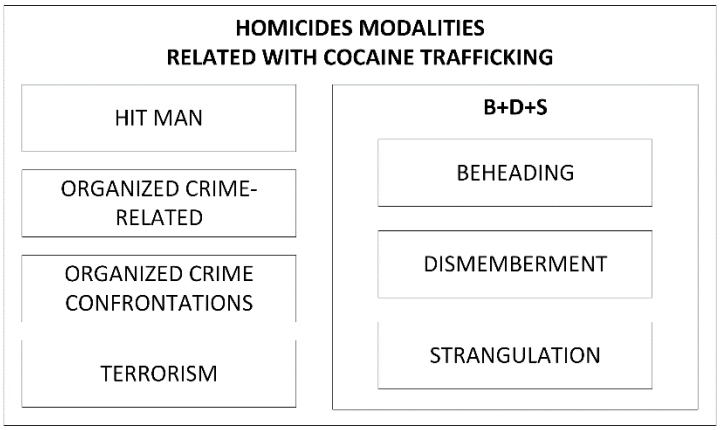


Figure 11. Homicides modalities

The "hit man" modality has been related to cocaine trafficking activities in Colombia. Originally, it was associated with cocaine business during the war declared by the drug traffickers against the Colombian government in the 1980s, where they offered a reward of $2,000 for each police officer killed (Pardo, 2000). Later, during the 1990s, paramilitary groups adopted the same modality to target anyone connected to Pablo Escobar (Polit Dueñas, 2013, p. 90). As indicated by the numbers, the "hit man" modality stands out as the most prevalent method of homicide in the Colombian Pacific region. Moreover, drug cartels in Mexico have also adopted and continued to utilize this modality (Durán-Martínez, 2015; Wainwright, 2016, p. 106, 109).

Organized crime-related homicides in Colombia mainly originate from drug trafficking. According to Mejía and Restrepo (2008), guerrillas and paramilitary groups became drug producers and traffickers, resulting in a significant increase in their financial resources, which, in turn, escalated the internal conflict in Colombia (Cotte Poveda, 2012). González Peña and Dorussen (2021) provided additional support to this observation by finding that the homicide rate did not increase in the municipalities where FARC ex-combatants demobilized. This implies that the funding for illegal armed groups comes from the drug trade, and when they exit this business, the propensity for homicide decreases in their influential areas.

Confrontations between organized crime refer to homicides that occur when two illegal groups engage in combat with each other. For example, during the COVID-19 pandemic, there were reports of clashes between the *Clan del Golfo* and the *Ejército de Liberación Nacional* (ELN) guerrillas as they struggled for control of territories to carry out illicit activities in *Chocó*, which were left open after the demobilization of the FARC (Gillies & Hume, 2022).

Regarding terrorism, this refers to the homicides caused by terrorist groups. Gonzalez, Sierra, and Fajardo (2019) affirmed that terrorism in Colombia has both political and narcotraffic objectives. For instance, this narcoterrorism included methods as bombings (Zapata, 2003) and selective assassinations of government officials advocating stringent anti-drug policies in Colombia, with the aim of instilling fear and terror among the population (Tarapues, 2012).

The merged category B+D+S includes the homicides due to beheading, dismemberment, and strangulation. Phillips (2015) supported the recognition of dismemberment as a brutal practice connected to the armed conflict in Colombia. Also, Fajardo Cely (2014) compiled documented cases where evidence indicated the utilization of diverse violent methods to torture and kill individuals during the Colombian armed conflict, such as beheading, dismemberment, decapitation, evisceration, incineration, castration, impaling, and the use of acid (p. 55). As mentioned earlier, the Colombian armed conflict is closely interlinked with the problem of drug trafficking.

Finally, in column nine, the severity of the homicide rate is sorted into three groups: "Low" for rates less than five, "Medium" for rates ranging between five and nine, and "High" for rates greater than nine. This classification was derived from the histogram detailed in Appendix K and it is shown in Figure 12.

Median	Mean	Mode
6	6.93	5

n	# classes	MAX	min	Range	Bin width
4000	13	26	1	25	2

Classes	min	MAX	Interval	Freq.	Cum.	Homicides
1	1	3	1-3	273	273	Low
2	3	5	3-5	737	1010	
3	5	7	5-7	1005	2015	Medium
4	7	9	7-9	883	2898	
5	9	11	9-11	535	3433	
6	11	13	11-13	303	3736	
7	13	15	13-15	142	3878	
8	15	17	15-17	67	3945	
9	17	19	17-19	26	3971	High
10	19	21	19-21	13	3984	
11	21	23	21-23	10	3994	
12	23	25	23-25	3	3997	
13	25	26	25-26	3	4000	

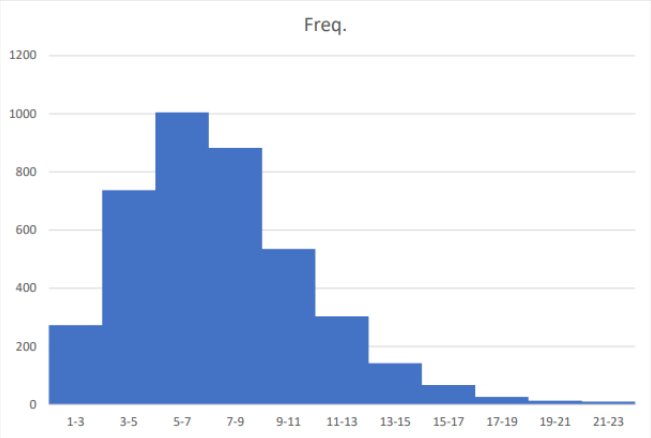


Figure 12. Homicides rate in the Colombian Pacific region 2012–2022 histogram

All the information regarding homicides was recorded based on data gathered by the MDN-ODHDN (2023). Table 6 summarizes the level of measurement corresponding to each variable.

Table 6. Level of measurement for time-series datasets variables

Column	Group	Classification	Variable	Level of measurement	Reference unit/category
1	Cocaine Seizures	Independent	Day	Interval	#
2			Q_kg	Ratio	kg
3			Sea_Land	Nominal	Sea or Land
4			Size	Ordinal	Small, Average, or Large
5	Homicides	Dependent	Day + n (n ∈ [1,7])	Interval	#
6			Homicides	Ratio	#
7			H_Location	Nominal	Cauca, Choco, Narino, or Valle
8			Modality	Nominal	Hit man, organized crime, organized crime confrontations, terrorism, or B+D+S
9			Intensity	Ordinal	Low, Medium, or High

B. DATASETS TIME FRAMES

After filtering the data, each time-series dataset contains only 308 records. Although the datasets still conserve records between 2012 and 2022, the distribution of records across individual years is uneven. This might be a limitation to the study, as data might not capture the trends between cocaine seizures and homicides in the Colombian Pacific region for some years. The distribution of records per year for the datasets is presented in Table 7.

Table 7. Records per year for time-series datasets

Year	Number of records
2012	3
2013	22
2014	17
2015	50
2016	54
2017	60
2018	53
2019	13
2020	12
2021	17
2022	7

Additionally, Table 8 illustrates the time spans covered by the datasets, showcasing the starting and concluding dates for each one. This delineation of initial and final dates, coupled with the record distribution showcased in Table 7, shows the consistency of the time intervals covered by the gathered and filtered data. Again, the scope of this study is constrained to the period between 2012 and 2022 as a result of the data obtained through web scraping from the ARC news website. The ARC website started the publication of seizure-related news in 2012, and any records preceding this timeframe are not accessible at its online platform.

Table 8. Datasets time frames

Reference	Day	Day+1	Day+2	Day+3	Day+4	Day+5	Day+6	Day+7
Initial day	264	265	266	267	268	269	270	271
Initial date (MM/ DD/YY)	9/20/12	9/21/12	9/22/12	9/23/12	9/24/12	9/25/12	9/26/12	9/27/12
Final day	4000	4001	4002	4003	4004	4005	4006	4007
Final date (MM/ DD/YY)	12/13/22	12/14/22	12/15/22	12/16/22	12/17/22	12/18/22	12/19/22	12/20/22

C. EXPLORATORY DATA ANALYSIS

Following with the process, these seven comprehensive datasets were imported into JMP to proceed with the analysis of correlations and trends between cocaine seizures and homicides in the Colombian Pacific region. By using graph builder functionality was used to compare all the independent and dependent variables (JMP Statistical Discovery LLC [JMP], 2023), identified in Table 6.

The resulting scatter graphs offer the option to select various types of elements (JMP Statistical Discovery LLC, 2023b). For exploration purposes, points and line of fit options were chosen. The chosen fitting approach was polynomial, encompassing linear, quadratic, or cubic degree regressions. Given the variables combinations and the polynomial degree options, a comprehensive exploration of regression models were tested for the best fit.

The metrics that were used to study statistical significance of the regression models are as follows. A 95% confidence interval was applied. The Root Mean Square Error (RMSE), R-square value (R^2), F test value, and probability value [p-value] for each of the regressors was extensively studied. Amon these metrics, only the p-value and R^2 were considered for the initial phase of the exploration.

Prior to assessing each regression, a null and an alternate hypotheses were identified. In this investigation, the null hypothesis (H0) is that the occurrences of cocaine seizures in the Pacific area do not have an impact on the homicide rate in the Colombian Pacific region; the alternative hypothesis (H1) is that the cocaine seizures in

the Pacific region do have an impact on the homicide rate in the Colombian Pacific region.

In this context, the p-value test indicates the degree of statistical significance of the regressors (independent variables). As a result, a significant p-value means to accept H1 and reject the null (Keller, 2012, p. 369). The p-value that is less than 0.01 indicates high significance, between 0.01 and 0.05 is significant, whereas other values indicate there is not significant relation between the independent and dependent variables, hence H1 is rejected (Keller, 2012, p. 371). R^2 represents the proportion of variability in the dependent variable that can be explained by changing the independent variable, with an interpretation of 1 denoting an absolute 100% correlation between the two variables (Keller, 2012, p. 139).

Following the application of the designated fitting parameters, information conforming to the chosen variable categories is extracted, creating new sub-datasets. Subsequently, JMP uses machine learning techniques to learn from the available data by recognizing patterns for each regression alternative. Upon executing all potential regressions, seven models with significant statistical evidence for H1 have been discerned. Table 9 provides an overview of these results, outlining the polynomial regression level, sample size, p-value, and R^2. Detailed information regarding these seven findings can be found in Appendix M.

Table 9. Findings on polynomial regressions

#	Finding	Degree	Sample size	p-value	R^2
1	Cocaine seizures on the Colombian Pacific Sea correlates with low-intensity rate of homicides in the Colombian Pacific Region a day after.	Linear	47/308	0.0277	0.103
2	Large cocaine seizures in the Colombian Pacific Region correlates and might be associated to the rate of homicides in the Colombian Pacific Region two days after.	Cubic	92/308	Low 0.0398 Med 0.0144 High <0.001	Low 0.279 Med 0.196 High 0.958
	Cocaine seizures in the Colombian Pacific Land correlates with high-intensity rate of homicides in the Colombian Pacific Region two days after.	Linear	25/308	0.0421	0.168
4	Large and average cocaine seizures in the Colombian Pacific Region correlates and might be associated to the high-intensity rate of homicides in the Colombian Pacific Region six days after.	Cubic	26/308	Avg 0.0491 Large 0.0099	Avg 0.606 Large 0.664
5	Large cocaine seizures in the Colombian Pacific Region correlates with the hit man homicides modality in the Colombian Pacific Region seven days after.	Cubic	82/308	0.0109	0.133
6	Cocaine seizures on the Colombian Pacific Sea correlates with high-intensity rate of homicides in the Colombian Pacific Region seven days after.	Cubic	18/308	0.0221	0.486
7	Large cocaine seizures in the Colombian Pacific Region correlates with low and high-intensity rate of homicides in the Colombian Pacific Region seven days after.	Quadratic	25/308	Low 0.0491 High 0.0099	Low 0.212 High 0.930

Figure 13 provides a preliminary overview of the alternative hypotheses derived from regression analysis. This analysis considered the correlation of the data across various categories of cocaine seizure sizes in the Pacific and their impact on the homicide intensity in the Colombian Pacific region. The bold arrows indicate a robust correlation, backed by high significance with p-values below 0.01. Meanwhile, the thin arrows represent significant correlations supported by p-values under 0.05. Notably, findings 2 and 4 suggest that significant large cocaine seizures in the Pacific have an influence on the emergence of a high-intensity homicide rate in the Colombian Pacific region, manifesting two and six days after the seizures.

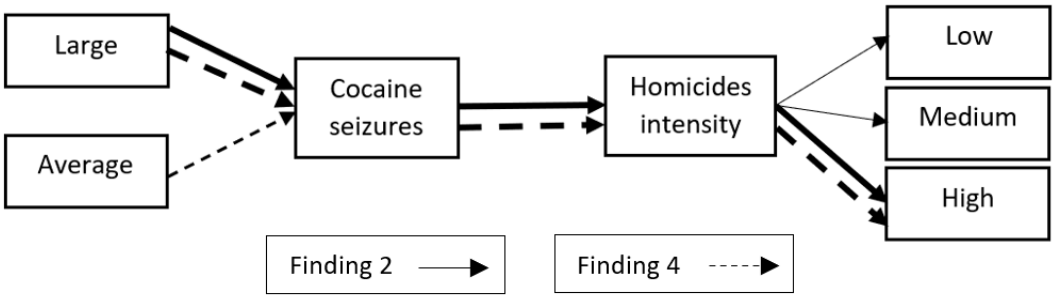


Figure 13. Alternative hypothesis from findings 2 and 4

D. FURTHER ANALYSIS

To further explain the relation between the independent and dependent variables, an improved approach was followed. The first step involves a fitting through the JMP graph builder fitting tools. This step allows to filter data and capture different behaviors that vary in different time intervals. The second step is to conduct a-least-squares regression analysis to examine, identify, and exclude (if necessary) data points with high impact on the model fit. The third step aimed at identifying the most suitable continuous distribution for the values of the independent variable, enabling the determination of distribution parameters. Lastly, a profiler is suggested for each model to simulate, predict, and visually illustrate how changes in the range of cocaine seizures and the classification of sea/land areas influence the homicide rate in the Colombian Pacific region. This procedure is shown in Figure 14.

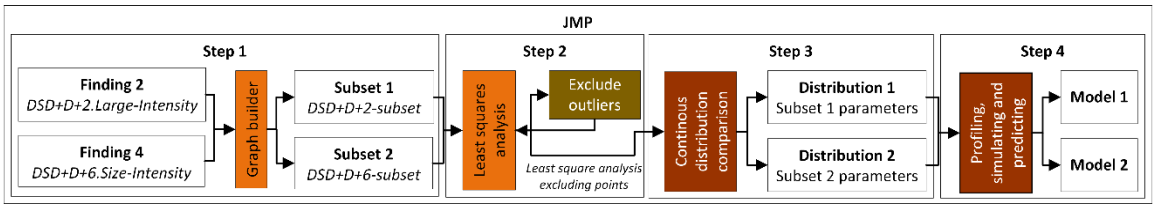


Figure 14. Filtering, exploring, and modeling procedure in JMP

Based on the outcomes derived from the *DSD+D+2.Large-Intensity* and the *DSD+D+6.Size-Intensity* datasets, which are contained in Appendix M, a graph builder

fitting technique was applied to create new datasets denominated *DSD+D+2-subset* and *DSD+D+6-subset*, respectively. This filtering procedure involved the elimination of categories and a hands-on exploration of data point boundaries to achieve a better p-value and R^2 value. These new datasets are provided in Appendix N.

To improve the accuracy of these two models, a least squares analysis was used to reduce discrepancies and discover the best match between the independent variable (cocaine seizures) and the dependent variable (homicides). Additionally, this method, along with an emphasis on the leverage effect, aids in examining influential datapoints and identifying instances of extreme values within the independent variable that could potentially impact the model's fit. After excluding the outliers, another least squares analysis was conducted, resulting in an improved model. Figure 15 shows the parameters specified to fit the model.

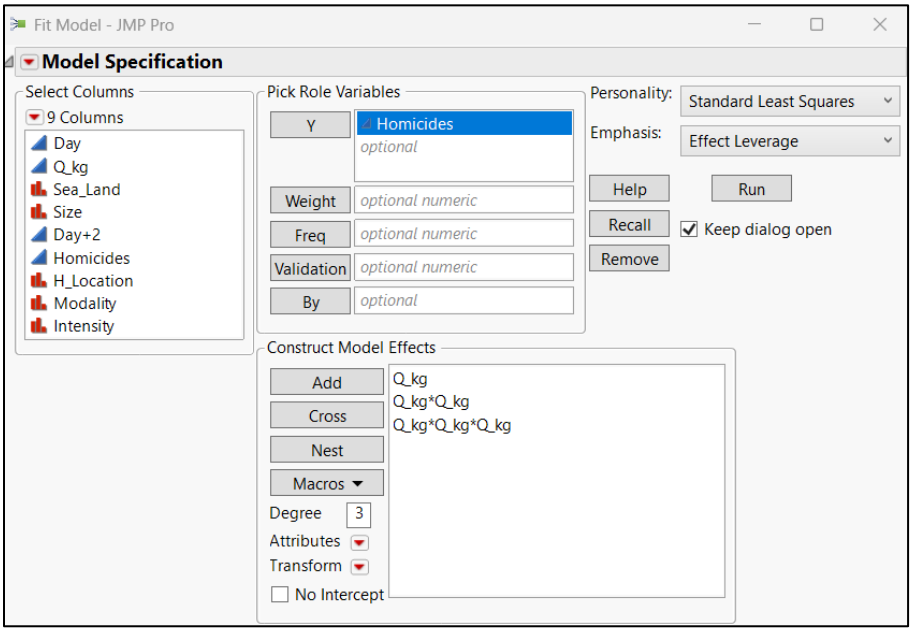


Figure 15. Standard least squares parameters

Although the least squares process employs identical parameters for both sub-sets of data, the outcomes diverge. For the continuous distribution comparison and the profiling processes, the parameters and corresponding results are different for each case.

E. DSD+D+2-SUBSET

1. Graph Builder Fitting and Least Squares Analysis

The filtering resulted in a subset, consisting of 23 data points out of the original 308 (approximately 7.46% of the total), reveals a correlation between cocaine seizures and the incidence of homicides in the Colombian Pacific region. The regression analysis yields a significant p-value of 0.0285. Furthermore, the applied cubic model effectively accounts for 37.2% of the variance in the homicide rates above eight within the Colombian Pacific region, attributed to cocaine seizures exceeding 750 kg occurring in the Pacific two days before. Also, the RMSE of the model suggests an average prediction deviation of around 2.81 homicides per day. Notably, all instances correspond to the "hit man" modality, while two align with the B+D+S modality, and one relates to confrontations between organized crime. On the other hand, at least 16 instances of cocaine seizures took place at sea. Figure 16 shows the *DSD+D+2-subset* polynomial regression result.

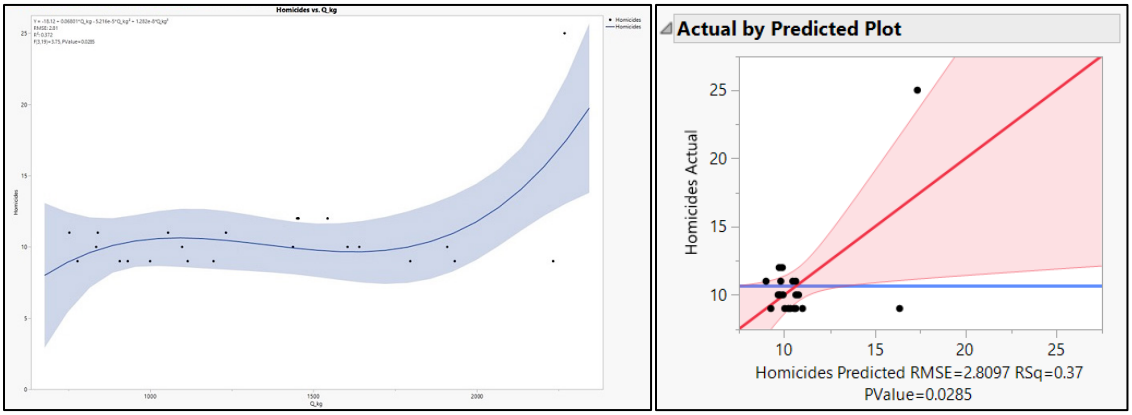


Figure 16. DSD+D+2-subset polynomial regression result

The actual by predicted plot, shown in Figure 16, indicates that majority of the data points are within the confidence interval. However, there are some outliers in the sample. To further analyze the outliers in the model, a row diagnosis of the studentized residuals was conducted. Figure 17 illustrates that 21 out of the 23 data points fall within the 95% confidence level's concurrent boundaries (red lines) and the individual limits

(green lines). The remaining two data points (specifically, rows 10 and 23 of the subset) lie beyond these boundaries.

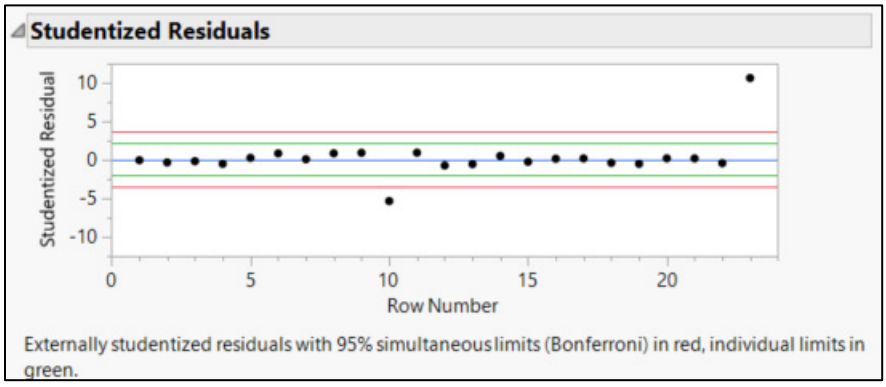


Figure 17. Studentized residuals plot for DSD+D+2-subset

After removing the outliers, another least squares model was conducted. The result of the polynomial regression is shown in Figure 18. In this scenario, the exclusion of outliers has a negative effect on the model's significance. Although the studentized residuals plot exhibits no data points exceeding the 95% confidence level, the actual by predicted plot suggests a lack of statistical significance, with a p-value surpassing 0.05 and lowered the R^2 below 0.3.

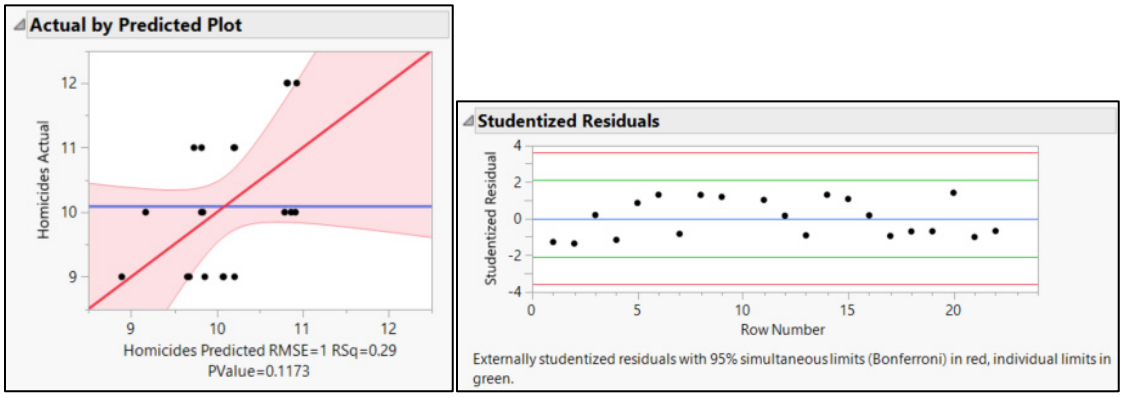


Figure 18. DSD+D+2-subset without outliers' regression result and studentized residuals

Through an iterative exploration, it was found that the model experiences significant improvement by excluding the outlier associated with row 10 and including the nominal variable Sea_Land. The statistical examination results in a high significant p-value <0.0001. Furthermore, the cubic fit model results demonstrate its effectiveness by accounting for 80% of the variability in homicide rates exceeding eight within the Colombian Pacific region. This variability is attributed to instances of cocaine seizures surpassing 750 kg, which occur in the Pacific two days prior. Also, the RMSE of the model suggests an average prediction deviation of around 1.78 homicides per day. The results of the polynomial regression are shown in Figure 19.

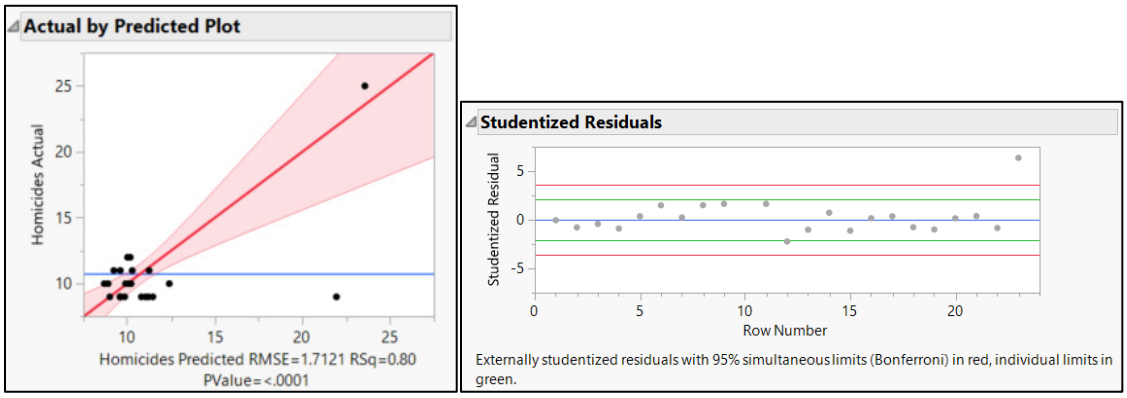


Figure 19. DSD+D+2-subset without outlier (row 10) regression result and studentized residuals

In this analysis, retaining the outlier associated with row 23, improves the model's significance. Consequently, the model shown in Figure 19 was chosen as "Subset 1" for the purposes of continuous distribution comparison.

2. Continuous Distribution Comparison

Considering the independent variable values of the Subset 1, corresponding to cocaine seizures, a new EDA took place. An exploration of possible distribution that can capture the cocaine seizures data was based on two criteria, which are the Akaike Information Criterion applicable to small samples (AICc) (Sugiura, 1978) and the Bayesian Information Criterion (BIC) (Schwarz, 1978) was conducted. Figure 20

illustrates the best fit with the lowest AICc and BIC values. For this case, the best continuous distribution was the lognormal.

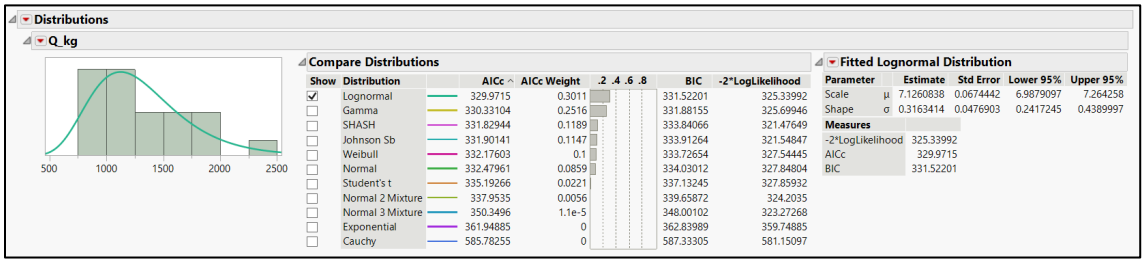


Figure 20. DSD+D+2-subset without outlier (row 10) lognormal distribution

The lognormal distribution is often used to model data exhibiting a positively skewed pattern, where the majority of values are concentrated on the lower end of the distribution (Crow & Shimizu, 1987). The parameters used to describe the fitted lognormal distribution were the location (μ=7.126) and scale (σ=0.316). Finding the best distribution was critical to identify the parameters for profiling, simulating, and prediction analysis.

3. Profiling, Simulating, and Prediction Analysis

JMP prediction profiler enables simulation and the visualization of interaction among the independent and dependent variables (homicides) in the Colombian Pacific region. This analysis was conducted by adjusting two variables: the quantity of seized cocaine (Q_kg) and the Sea_Land category. The lognormal distribution parameters for the former should remain consistent to accurately reflect the data in "Subset 1." Similarly, the occurrence probability derived from the information in "Subset 1" should be maintained for the latter. In addition, using the simulation tool allows the reproduction of diverse scenarios and prediction of varying outcomes. Figure 21 shows the prediction profiler and simulator created using data from "Subset 1," identified as "Model 1." As can be seen in the lower left quadrant in Figure 21, for larger seizures, Q_kg and Sea_Land variables interact.

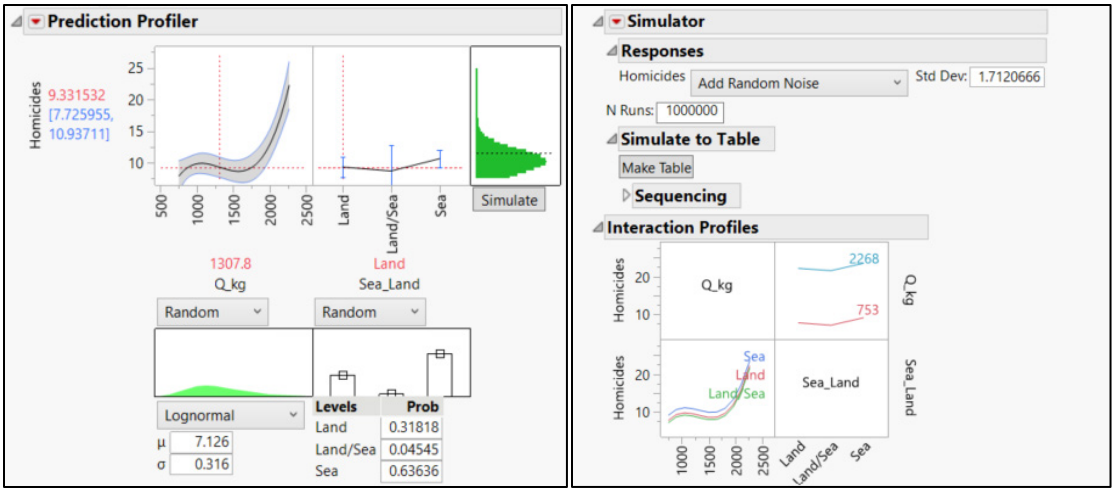


Figure 21. Model 1: Prediction profiler and simulator for Subset 1

In the simulator, a random noise can be configured and smeared to the simulation. Again, this standard deviation value matches with the RSME showcased in Figure 19. The number of runs correspond to additional parameters to perform a Monte Carlo simulation. All the parameters in the prediction profiler and the simulator can change to estimate different outcomes. Additionally, as the input parameters vary, different outcomes can be saved in a new table to create an output collection corresponding to the simulation results of different scenarios. Figure 22 shows the interaction profiles obtained with the parameters specified in Figure 21.

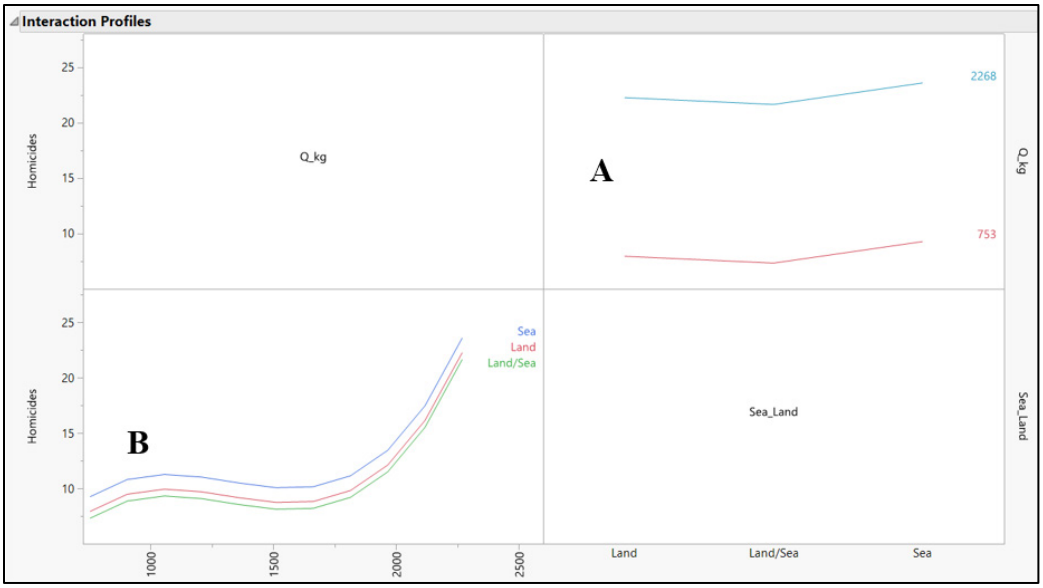


Figure 22. Interaction profiles for Model 1

In Figure 22, quadrants A and B show the outcomes for Model 1. In quadrant A is possible to interact with the Sea_Land categories within the specified range. For example, when Land category is selected at the lower limit (red line), it indicates that if 753 kg of cocaine were seized in Pacific land, there would be an estimated average of eight homicides in the Colombian Pacific region two days later. Following up with Land category until the upper limit (blue line) will depict the correlation between the cocaine seize in Pacific land and the homicides in the Colombian Pacific region two days later. This pattern is mirrored by the red curve shown in quadrant B of Figure 22. A similar scenario applies for Sea (blue curve) and Land/Sea (green curve) categories.

Regarding Model 1, the Sea category exerts a more significant influence on the rate of homicides. The graphical representation in quadrant B, indicated by the blue curve, shows an average of two more homicides than the red curve and three more homicides than the green curve. This suggests that when cocaine above 753 kg is seized in the Pacific sea, there would be an average of two additional homicides in the Colombian Pacific region two days later compared with the case of seizures in the Pacific land. All subsets and models discussed in this section are enclosed in Appendix O.

F. DSD+D+6-SUBSET

1. Graph Builder Fitting and Least Squares Analysis

The DSD+D+6-subset, comprising 31 samples out of a total of 308 (representing 10.06%), shows a significant correlation between cocaine seizures and homicides. This is a statistically significant result with a p-value of 0.011. This cubic fit model explains 33% of the variation in the rate of homicides above nine in the Colombian Pacific region due to cocaine seizures above 545 kg occurred in the Pacific six days before. The RMSE of the model suggests, on average, that the predictions are off by approximately 1.61 homicides per day. In this analysis, all instances correspond to the "hit man" modality, with four associated with B+D+S, two linked to organized crime confrontations, another two tied to organized crime, and two more connected to the terrorism modality. Conversely, 13 instances of cocaine seizures took place at sea, while the others occurred on land. Figure 23 shows the DSD+D+6-subset polynomial regression result.

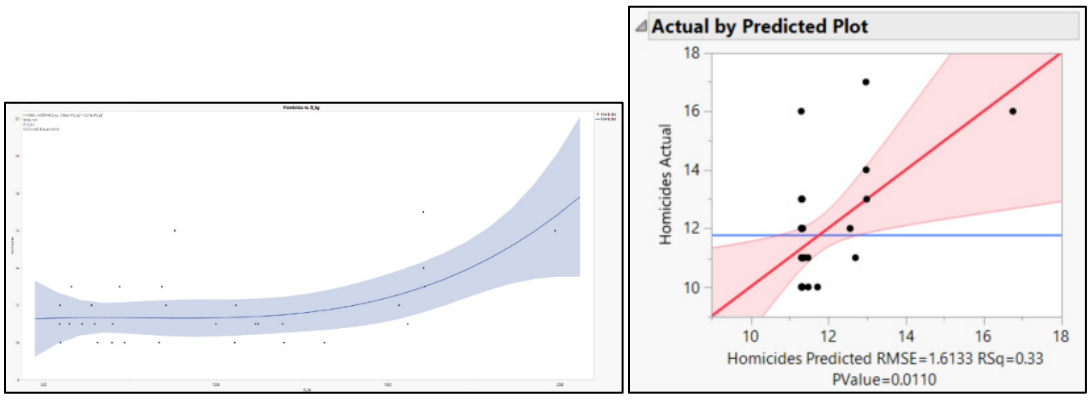


Figure 23. DSD+D+6-subset polynomial regression result

The actual by predicted plot shows the significance of the analysis as the confidence curve crosses the horizontal line. However, there are some outliers in the sample. To highlight the outliers in the model, a row diagnosis of the studentized residuals was applied. Figure 24 illustrates the fact that 29 out of the 31 data points fall within the 95% confidence level's concurrent boundaries (red lines) and the individual

limits (green lines). The remaining two data points (specifically, rows 13 and 17 of the subset) lie beyond these boundaries.

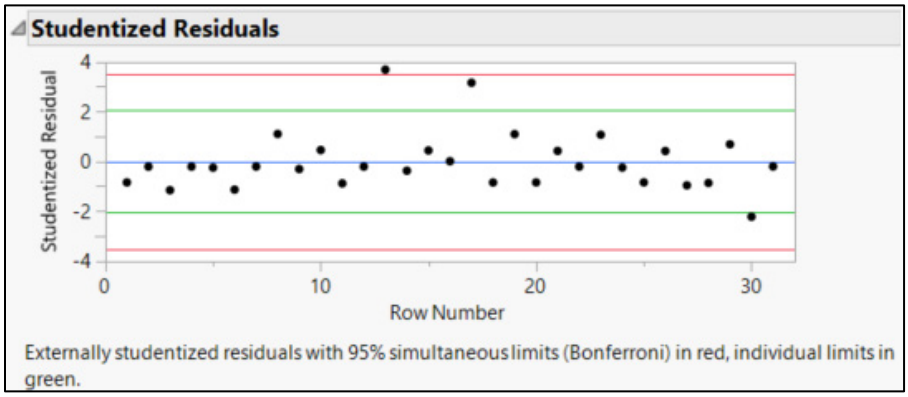


Figure 24. Studentized residuals plot for DSD+D+6-subset

Through an exploration analysis, it was found that the model improves significantly by excluding the outliers associated with rows 13 and 17, also including the nominal variable Sea_Land. The statistical examination results in a high significant p-value of less than 0.0011. Furthermore, the cubic model applied demonstrates its effectiveness by accounting for 52% of the variability in homicide rates exceeding nine within the Colombian Pacific region. This variability is attributed to instances of cocaine seizures surpassing 545 kg, which occur in the Pacific six days before. Also, the RMSE of the model suggests an average prediction deviation of around 1.06 homicides per day. The findings of the polynomial regression analysis are shown in Figure 25. This model was chosen as "Subset 2" for the purposes of continuous distribution comparison.

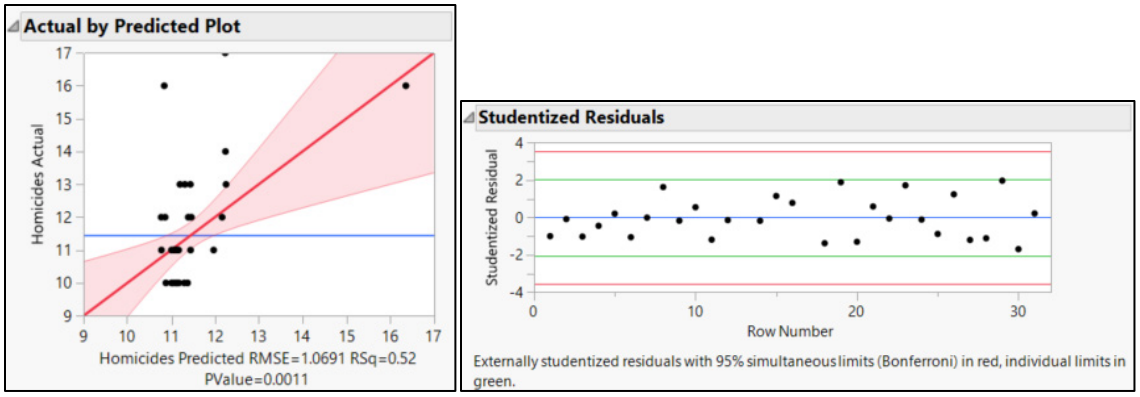


Figure 25. DSD+D+6-subset without outliers' regression result and studentized residuals

2. Continuous Distribution Comparison

Considering the independent variable values of the Subset 2, corresponding to cocaine seizures, another EDA was conducted. An exploration of possible distribution that can capture the cocaine seizures data was based on two criteria, which are the Akaike Information Criterion applicable to small samples [AICc] (Sugiura, 1978) and the Bayesian Information Criterion [BIC] (Schwarz, 1978) was conducted. Figure 26 illustrates the best fit distribution with the lowest AICc and BIC values. For this case, the best continuous distribution was the Johnson Sb (Slifker & Shapiro, 1980).

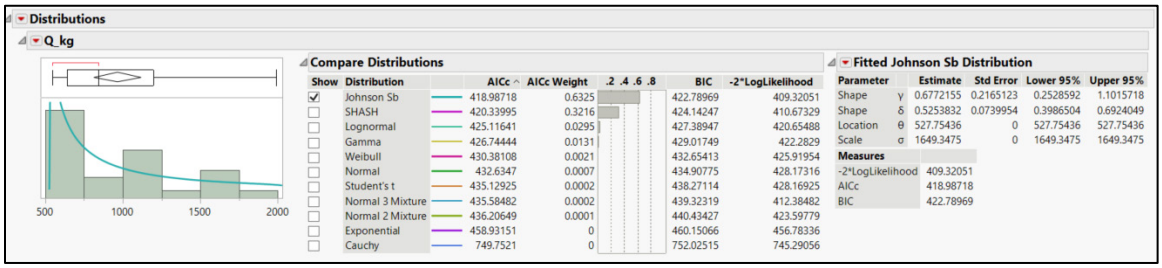


Figure 26. DSD+D+6-subset without outliers Johnson Sb distribution

The Johnson Sb distribution is useful when working with data that exhibit skewness and kurtosis (Jones, 2004). The parameters of the fit that are used to describe the fitted Johnson Sb distribution were the kurtois shape (γ=0.677), skewness shape

(δ=0.525), lower bound location (θ=527.75) and scale (σ=1649.34). Finding the best continuous distribution was necessary to identify the parameters for profiling, simulating, and predicting purposes.

3. Profiling, Simulating, and Prediction Analysis

JMP prediction profiler enables simulation and the visualization of interaction among the independent and dependent variables (homicides) in the Colombian Pacific region. This analysis was conducted by adjusting two variables: the quantity of seized cocaine (Q_kg) and the Sea_Land category. The Johnson Sb distribution parameters for the former should remain consistent to accurately reflect the data in "Subset 2." Similarly, the occurrence probability derived from the information in "Subset 2" should be maintained for the latter. However, employing a simulator allows the reproduction of diverse scenarios and prediction of varying outcomes. Figure 27 shows the prediction profiler and simulator created using data from "Subset 2," identified as "Model 2." As can be seen in the lower left quadrant in Figure 27, for larger seizures, Q_kg and Sea_Land variables interact.

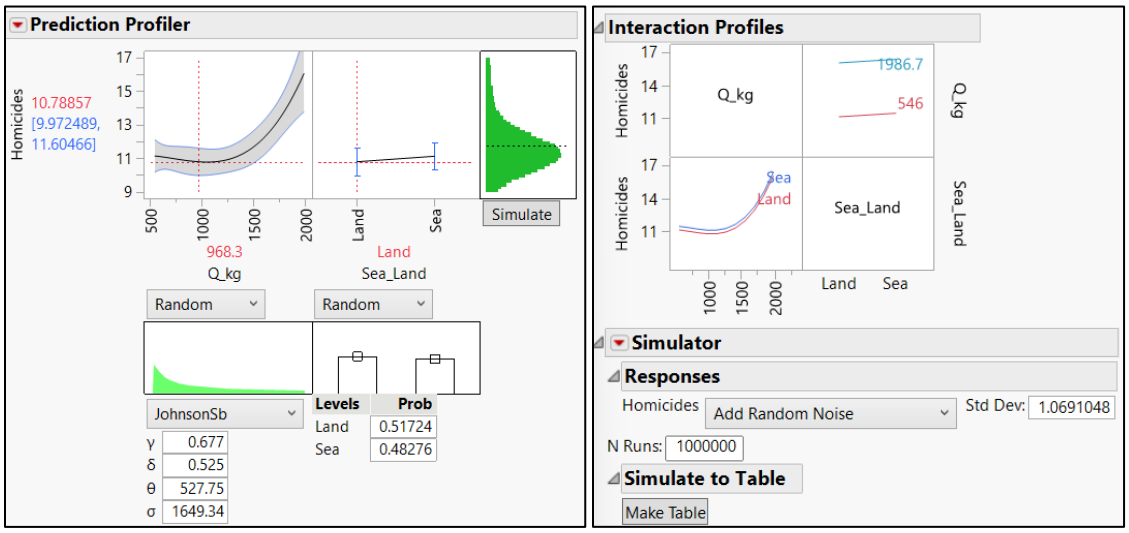


Figure 27. Model 2: Prediction profiler and simulator for Subset 2

Note also, that in the simulator responses' options, a random noise can be configured. Again, this standard deviation value matches with the RSME showcased in Figure 25. The number of runs correspond to additional parameters to perform a Monte Carlo simulation. All the parameters in the prediction profiler and the simulator can change to estimate different outcomes. Additionally, as the input parameters vary, different outcomes can be saved in a new table to create an output collection corresponding to the simulation results of different scenarios. Figure 28 shows the interaction profiles obtained with the parameters specified in Figure 27.

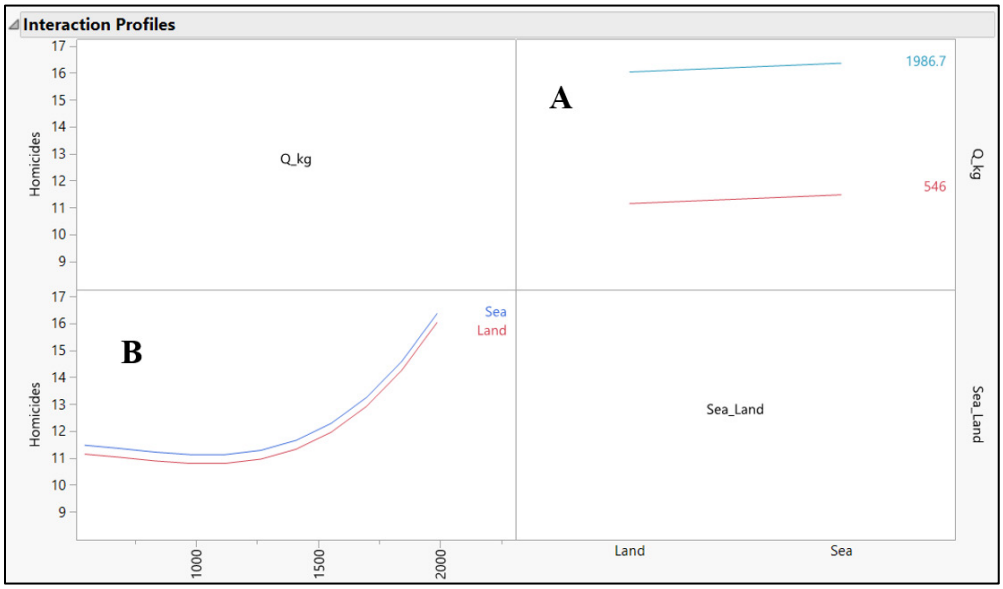


Figure 28. Interaction profiles for Model 2

In Figure 28, quadrants A and B show the outcomes for Model 2. In quadrant A is possible to interact with the Sea_Land categories within the specified range. Similarly, when Land category is selected at the lower limit (red line), it indicates that if 546 kg of cocaine were seized in Pacific land, there would be an estimated average of eleven homicides in the Colombian Pacific region six days later. Following up with Land category until the upper limit (blue line) will depict the correlation between the cocaine seize in Pacific land and the homicides in the Colombian Pacific region six days later.

This pattern is mirrored by the red curve shown in quadrant B of Figure 22. A similar scenario applies for Sea (blue curve) category.

Regarding Model 2, the Sea category exerts a more significant influence on the rate of homicides. The graphical representation in quadrant B, indicated by the blue curve, shows an average of one more homicide than the red curve. This suggests that when cocaine above 546 kg is seized in the Pacific sea, there would be an average of one additional homicides in the Colombian Pacific region six days later compared with the case of seizures in the Pacific land. All subsets and models discussed in this section are enclosed in Appendix P.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

Given the studies conducted by Miron (2001) and UNODC (2012), the collection and utilization of time series data led to results that were more closely aligned with Miron's findings. The findings of this study showed a connection between cocaine seizures and homicides. Model 1 captures the correlation between cocaine seizures exceeding 750 kg in the Pacific area (both sea and land) and the homicide rate (above eight) in the Colombian Pacific region with two days lag time. Similarly, Model 2 captures how cocaine seizures greater than 545 kg in the Pacific area impact the homicide rate (higher than nine) in the Colombian Pacific region six days later.

Although Model 1 might seem deceiving because it included an outlier data point to improve its significance, this situation improved in Model 2, as it provides more reliable outcomes concerning the relation among variables. The relation between the cubic fit models that underpin Model 2 more accurately captures the nature of the variables; this is primarily due to the presence of more data points. Also, Model 2 demonstrates a more logical progression of R^2 values. The explanation for the variation in the homicides rate (higher than nine) in the Colombian Pacific region due to cocaine above 545 kg seized in the Pacific area, with six days lag time, exhibited a gradual rise from 33% to 52% after excluding outliers.

Nevertheless, a critical drawback applicable to both models pertains to the sample size. While the sample size may not provide definitive conclusions about the relationship between cocaine seizures and homicides in the Colombian Pacific region, the EDA conducted in this thesis enables the identification of mathematical models that align with the available data. Accordingly, even though a linear correlation between the variables is not significant, the seizures of cocaine hydrochloride through both maritime and terrestrial means in the Pacific area appears to be a factor that influenced homicide rates in the Colombian Pacific region from 2013 to 2021 for Model 1, and from 2012 to 2022 for Model 2.

For this reason, it is crucial to strike a balance in implementing diverse strategies against drug trafficking to achieve short-term results while ensuring they do not foster violence within Colombian territory. To achieve this goal, it will be crucial to maintain continuous monitoring of the factors inherent in each strategy and to closely observe the homicide rate in the areas where these strategies are implemented. The positive outcomes yielded by law enforcement efforts in the short term should encourage coca growers to willingly participate in government programs. To ensure tangible and enduring results in the medium and long term, there must be an unwavering commitment to these programs, regardless of the government in power.

Another inherent constraint of the models relates to the data collection process. Taking into account the technical, legal, and ethical restrictions associated with this endeavor, gaining access to information regarding cocaine seizures from the relevant authorities of countries within the Pacific region, actively engaged in combating drug trafficking, has proven unattainable. In the absence of this broader international viewpoint, the analysis can only address the correlation between these variables within the confines of the Colombian context. Timely data collection efforts can resolve the lag time problem in the data and hence ameliorate both the time series analysis and regression analysis.

Although the Colombian government's efforts in combating drugs play a crucial role in addressing this issue, it is vital to maintain consistent support from the international community. Collaborative strategies involving numerous nations and institutions, as seen in the Naval Campaigns Orion (Ortega, 2022), must persist, but with careful consideration to avoid situations such as the balloon effect (Dávalos & Morales, 2022), or provoking retaliatory actions and unintended harm in the territories impacted by the strategies (LSE IDEAS, 2014).

A recommendation for future research is that it would be beneficial to increase the sample size by accessing databases from the authorities of various nations that conduct maritime and land interdiction in the Pacific region. Additionally, it could investigate the impact of these interdiction efforts on the homicide rate in the territories that border the Pacific Ocean. Moreover, a research group with greater resources could consider

obtaining official information from state entities' websites using web scraping techniques, following all the necessary legal and ethical procedures for such data collection.

Another viable research alternative is to utilize the existing data in the models to design a qualitative study. This study should encompass the validation of all the data available in each model to confirm whether the narrative of each case aligns with the reality implicitly developed in the results presented here. By conducting a thorough qualitative analysis, researchers might gain deeper insights into the accuracy and coherence of the findings and their alignment with real-world situations in the Colombian Pacific region.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. ARC NEWS WEBSITE LEGAL CONSTRAINTS

The terms of use, copyright, privacy, and access policies are not explicitly indicated on this website as consulted on January 14, 2023. In turn, there were two guidelines regarding the information security policy (ARC, 2022a). As of this date, any of the guidelines indicated any restrictions on the usage of robots, spiders, scripts, or automated tools to extract or collect data from the content.

As per legal regulations (Congreso de la República de Colombia, 2014), the information present in the ARC news portal is classified as public information regarding copyrights. This classification is due to the absence of any indications that would categorize it as classified or reserved (ARC, 2022a).

Concerning privacy matters, the author of the thesis recognized and agreed to abide by the policies established in accordance with the Habeas Data law and the regulation of managing information contained in personal databases (Congreso de la República de Colombia, 2008), as well as the law for the protection of personal data (Congreso de la República de Colombia, 2012).

Finally, this thesis does not disclose any information beyond what is publicly available from the ARC. Moreover, the data extracted from the ARC news portal is solely intended for personal academic purposes and is not being used for any commercial endeavors.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. USCG NEWS WEBSITE LEGAL CONSTRAINTS

The terms of use for this website were consulted by the thesis' author on January 14, 2023 (Bright Mountain Media, 2023b). As of this date, the website owner had not implemented any restrictions on the usage of robots, spiders, scripts, or automated tools to extract or collect data from the content. Furthermore, such restrictions were not among the 15 points outlined in their prohibited conduct guidelines.

Regarding copyrights, all the news that were gathered through scraping belonged to the USCG. Upon reviewing the privacy and security policy of the USCG news website, it stated that the information provided was intended as a public service (Department of Homeland Security [DSH], 2023).

With regards to privacy, the author of the thesis acknowledged and accepted the policies set forth by the website owner concerning the collection of PII, the use of cookies, and the storage of user information. Additionally, they acknowledged compliance with the California Consumer Privacy Act. (Bright Mountain Media, 2023a).

Ultimately, this thesis does not reveal any information beyond what is already provided by the USCG as a public service. Furthermore, the information extracted from the USCG is strictly intended for personal academic purposes and is not being utilized for any commercial use.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. PONAL NEWS WEBSITE LEGAL CONSTRAINTS

As of January 20, 2023, the terms of use stated on this website, as provided by the PONAL (2023b), do not contain restrictions on the utilization of robots, spiders, scripts, or automated tools for the purpose of extracting or collecting data from the content.

In terms of copyrights, the institution adheres to the open data policy established by the Ministry of Technology and Information in Colombia (Ministerio de Tecnologías de la Información y las Comunicaciones [MINTIC], 2021). This policy defines open data as public information presented in formats that enable its use and reuse under an open license, without any legal restrictions on its utilization.

Regarding privacy concerns, the author of the thesis acknowledges and accepts the disclosure of individual information collected by the PONAL in compliance with its privacy policy when visiting its news website (PONAL, 2015).

Finally, this thesis does not disclose any information beyond what is publicly available from the PONAL. Moreover, the data extracted from the PONAL news portal is solely intended for personal academic purposes and is not being used for any commercial endeavors.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. ARC NEWS WEBSITE ANATOMY

Figure 29. ARC website first scraping level

Figure 30. ARC website second scraping level with pagination

Figure 31. ARC website second scraping level

APPENDIX E. USCG NEWS WEBSITE ANATOMY

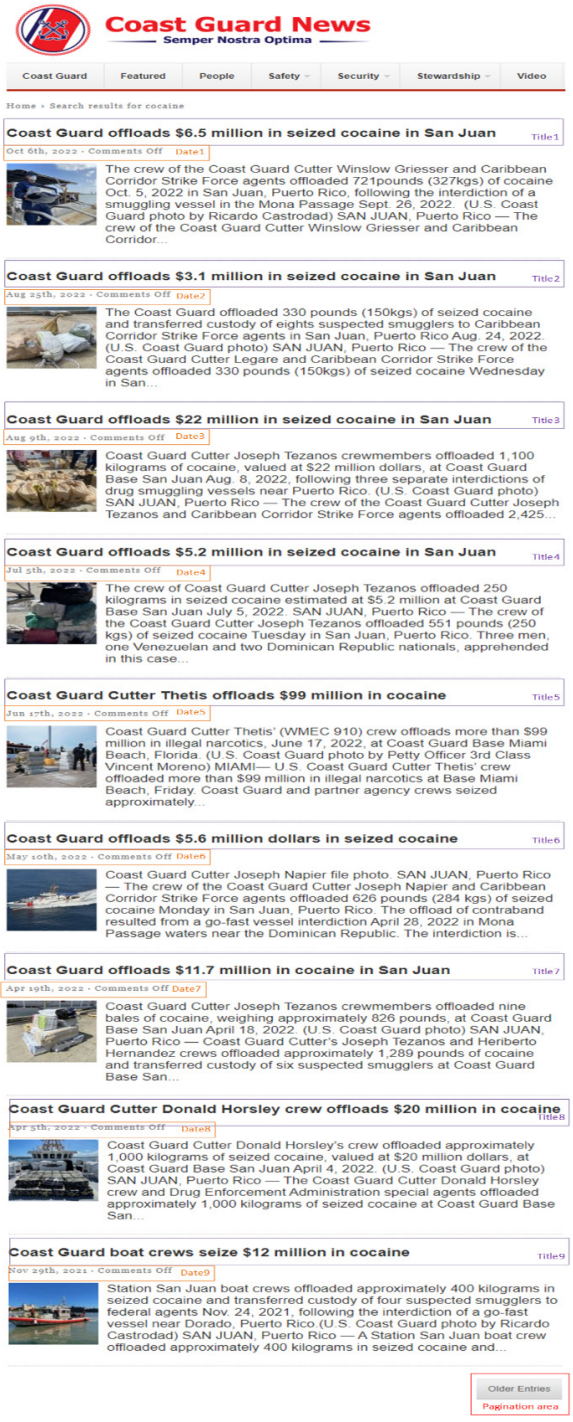


Figure 32. USCG website first scraping level

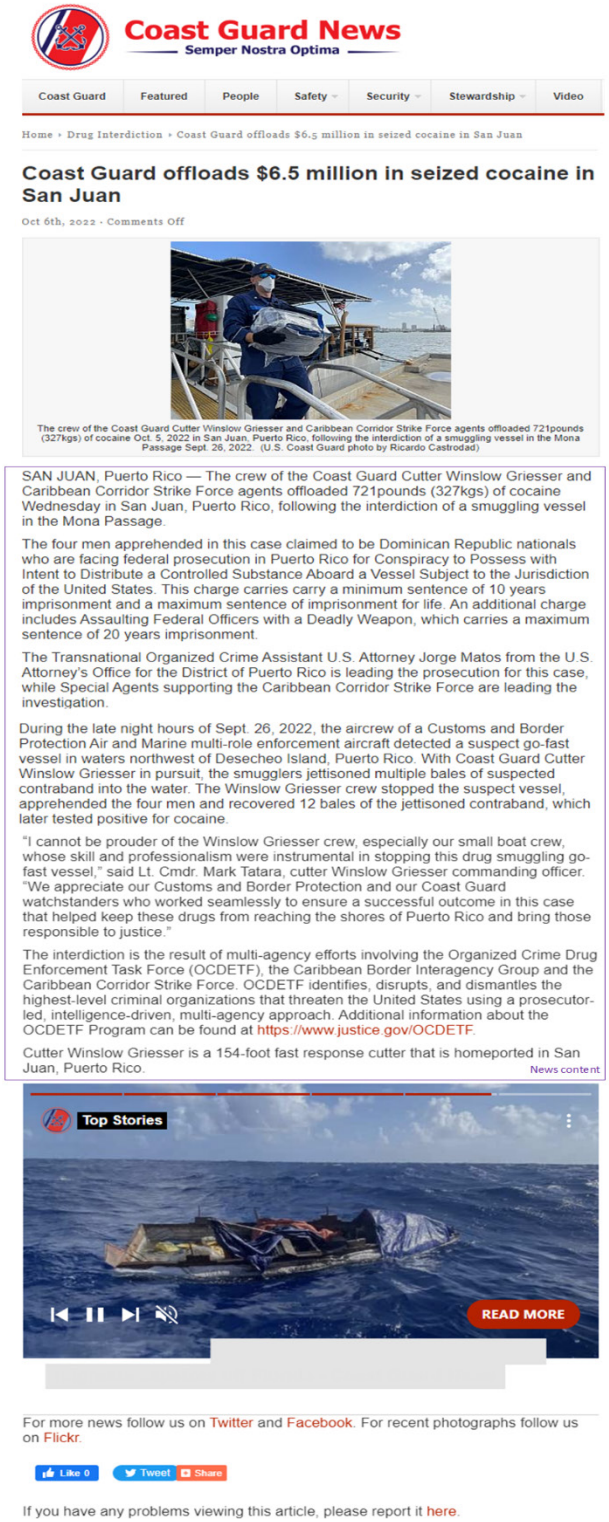
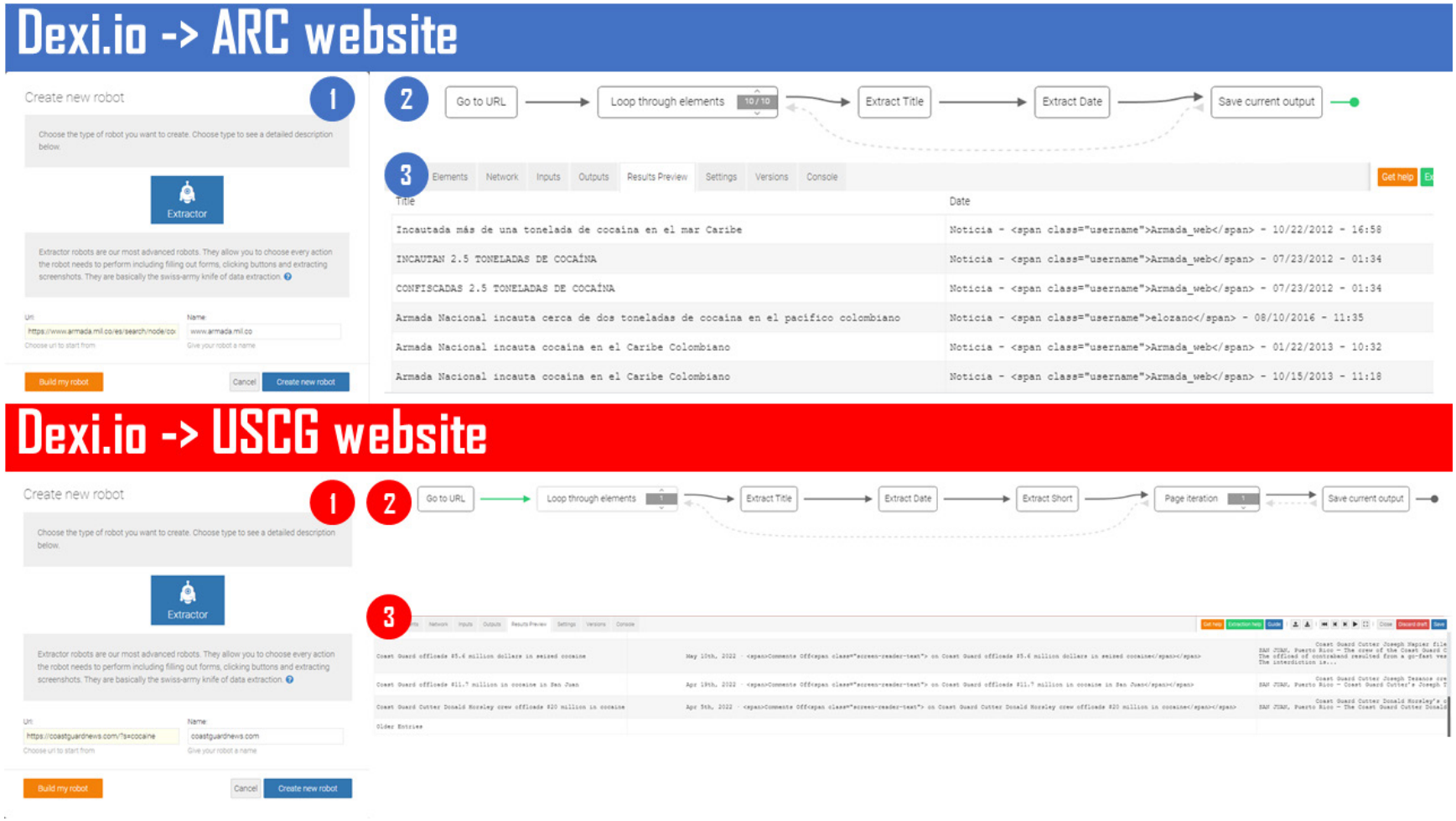


Figure 33. Second scraping level

APPENDIX F. WEB SCRAPING TOOLS ASSESSMENTS



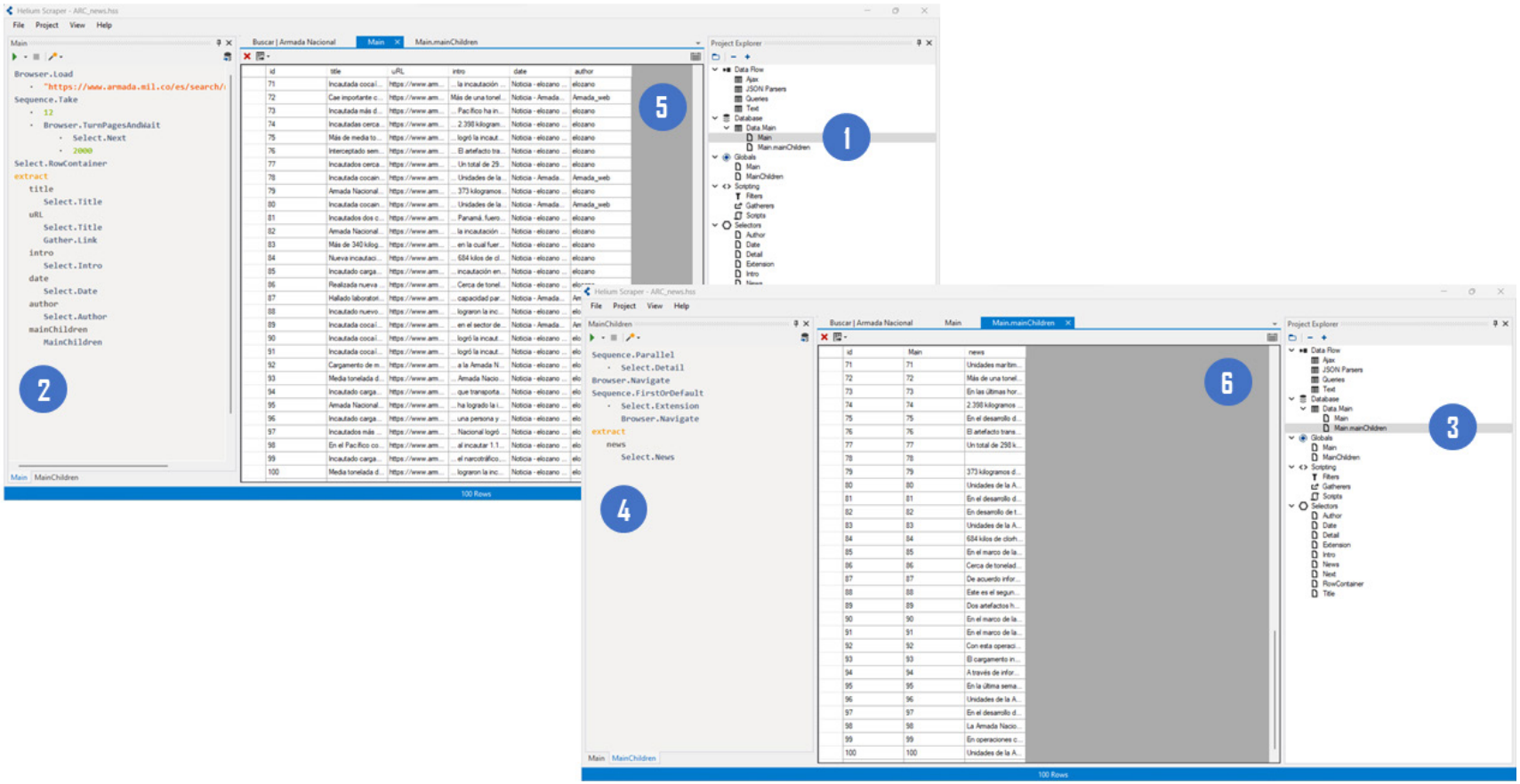
Dexi.io -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
YES	NO	NO	High	157.73	300.00	7.43	14.72	NO	YES	1000	Medium	Medium	NO
NO			1	454.73		22.15		0	3	1	2	2	1

Remarks:

1. During the implementation it was not possible to solve the page iteration either the crawling inside the title link to gather the news content data.
2. The GUI was difficult to understand. The implementation of the robot's steps was not easy using blocks.
3. The documentation and the interface did not relate a function about the IP rotation function.
4. It is shown in the preliminary results that the language fidelity for Spanish and English was not good because some HTML coding traces remained in the date field after the extraction.
5. The trial version did not allow to export data to Excel.

Helium Scraper -> ARC website



Helium Scraper -> ARC website results

id	title	uRL	date	author
1	Incautada más de una tonelada de cocaína en el mar Caribe	https://www.armada.mil.co/es/content/incautada-m%C3%A1s-de-u	Noticia - Armada_web - 10/22/2012 - 16:58	Armada_web
2	INCAUTAN 2.5 TONELADAS DE COCAÍNA	https://www.armada.mil.co/es/content/incautan-25-toneladas-de-c	Noticia - Armada_web - 07/23/2012 - 01:34	Armada_web
3	CONFISCADAS 2.5 TONELADAS DE COCAÍNA	https://www.armada.mil.co/es/content/confiscadas-25-toneladas-d	Noticia - Armada_web - 07/23/2012 - 01:34	Armada_web
4	Armada Nacional incauta cerca de dos toneladas de cocaína en el pacífico colomb	https://www.armada.mil.co/es/content/armada-nacional-incauta-cc	Noticia - elozano - 08/10/2016 - 11:35	elozano
5	Armada Nacional incauta cocaína en el Caribe Colombiano	https://www.armada.mil.co/es/content/armada-nacional-incauta-cc	Noticia - Armada_web - 01/22/2013 - 10:32	Armada_web
6	Armada Nacional incauta cocaína en el Caribe Colombiano	https://www.armada.mil.co/es/content/armada-nacional-incauta-cc	Noticia - Armada_web - 10/15/2013 - 11:18	Armada_web
7	Más de una tonelada de cocaína incautada en el Pacífico colombiano	https://www.armada.mil.co/es/content/mas-de-una-tonelada-de-cc	Noticia - elozano - 01/22/2016 - 14:13	elozano
8	Incautada cerca de media tonelada de cocaína a los autodenominados "Gaitanista	https://www.armada.mil.co/es/content/incautada-cerca-de-media-t	Noticia - elozano - 11/23/2016 - 10:51	elozano
9	Incautado cargamento de cocaína en Operación Naval Binacional	https://www.armada.mil.co/es/content/incautado-cargamento-de-c	Noticia - elozano - 12/22/2015 - 14:34	elozano
10	Más de dos toneladas de clorhidrato de cocaína incautadas al "Clan del Golfo"	https://www.armada.mil.co/es/content/mas-de-dos-toneladas-de-c	Noticia - elozano - 08/30/2016 - 10:08	elozano
11	Más de una tonelada de cocaína en el Pacífico colombiano	https://www.armada.mil.co/es/content/mas-de-una-tonelada-de-cc	Noticia - elozano - 04/06/2016 - 15:39	elozano
12	CONFISCAN 2.5 TONELADAS DE COCAÍNA EN DOS OPERACIONES	https://www.armada.mil.co/es/content/confiscan-25-toneladas-de-	Noticia - Armada_web - 07/23/2012 - 01:34	Armada_web
13	Incautadas 1.9 toneladas más de cocaína en el Chocó			
14	Fuerza Naval del Pacífico incauta más de 700 kilos de			
15	Incautamos más de 300 kilogramos de cocaína en el C			
16	Cerca de una tonelada de cocaína incautada en el Pac			
17	Incautados más de 2.200 kilos de cocaína en el mar C			
18	Armada Nacional incauta 850 kilos de cocaína en el pa			
19	Incautada más de una tonelada de cocaína de las Farc			
20	Incautada cocaína en un buque carbonero en la Guajir			
21	Se incautaron media tonelada de cocaína en el Caribe			
22	Decomisadas más de dos toneladas de cocaína y cuat			
23	Incautada más de una tonelada de cocaína en el mar			
24	Incautadas 1.3 Toneladas de clorhidrato de cocaína e			
25	Al menos una tonelada de cocaína incautada a las Far			
26	Cerca de 100 kilogramos de cocaína incautados en Ur			
27	Armada Nacional incautó más de 200 kilos de Clorhidr			
28	Incautado cargamento de cocaína perteneciente al "C			
29	Capturado guatemalteco, ecuatoriano y colombiano c			
30	Incautamos más de 800 kilos de cocaína en el océano			
31	Incautada cerca de una tonelada y media de clorhidra			
32	Incautadas cerca de 1,5 toneladas de cocaína en el pa			

Main | Main.mainChildren

Ready | Accessibility: Investigate

This program extracted data into two different tables corresponding to the first and second scraping levels. The program indexes the children table in relation to the main table.

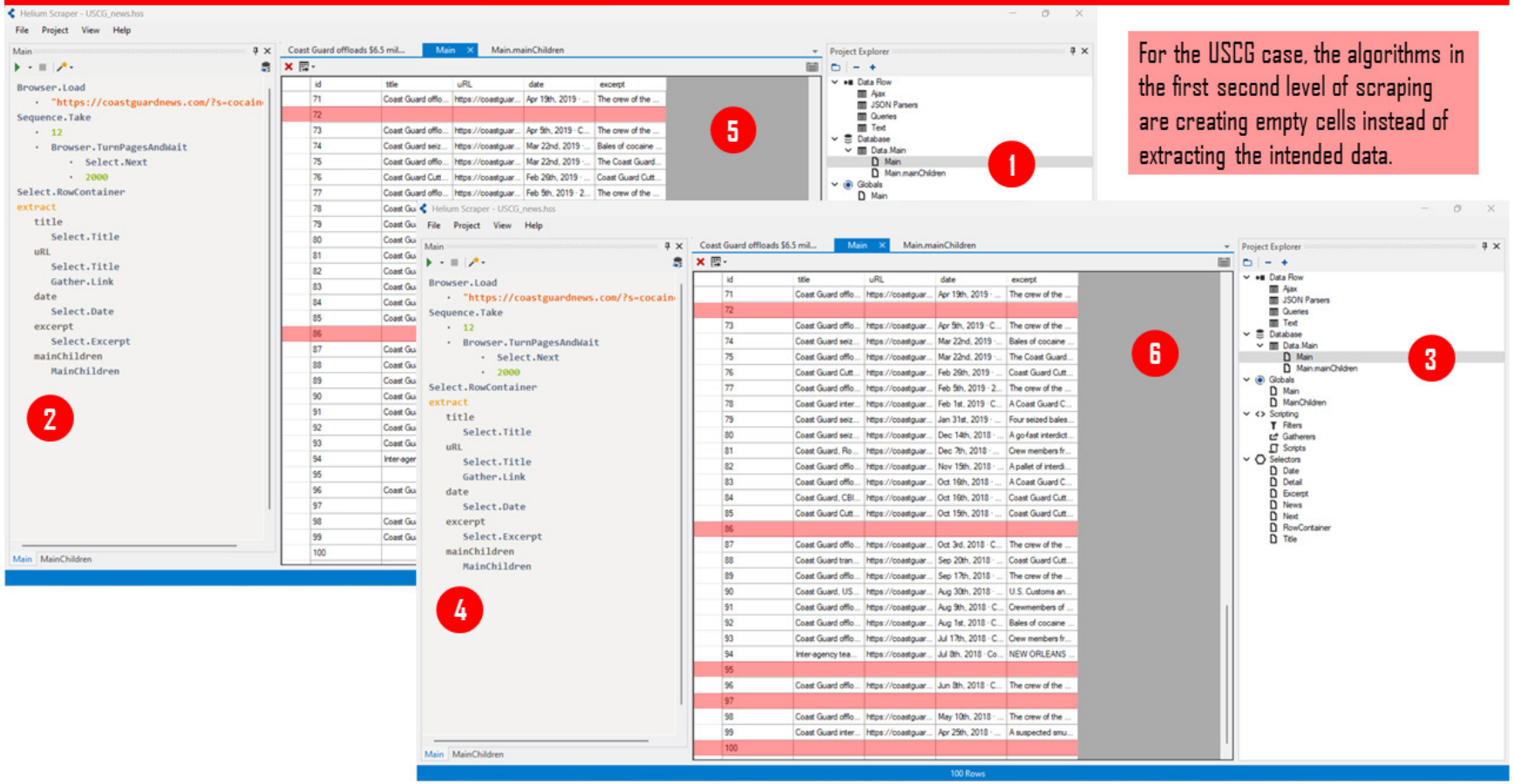
id	Main	news
1	1	Más de una tonelada de clorhidrato de cocaína fue incautada en las últimas horas en aguas del mar Caribe colombiano, en una operación combinada desarrollada por unidades de la Armada Nacional y una unidad aérea de la Armada Norteameri
2	2	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral Pacífico, revelaron fuentes oficiales. En el primer decomiso, unidades de las dos fuerzas colombianas hallaron 1,5 toneladas de cocaína escondi
3	3	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral Pacífico, revelaron fuentes oficiales. En el primer decomiso, unidades de las dos fuerzas colombianas hallaron 1,5 toneladas de cocaína escondi
4	4	En lo corrido del año, se han incautado más de 60 toneladas de clorhidrato de cocaína evitando que más de 1.800 millones de dólares ingresarán a las organizaciones narcotraficantes."Unidades de la Armada Nacional en el marco de la lucha fro
5	5	
6	6	En menos de una semana, la Armada Nacional ha incautado cerca de una tonelada de cocaína en operaciones realizadas en el Caribe y Pacífico colombiano.En las últimas horas, en desarrollo de operaciones de registro y búsqueda fueron incaut
7	7	En la última semana, la Institución Naval ha logrado la incautación de más de tres toneladas de cocaína.Unidades de la Armada Nacional en desarrollo de operaciones de control territorial lograron la incautación de más de una tonelada de clorh
8	8	En una operación que adelanta la Armada Nacional desde hace dos semanas, en zona rural de Tumaco, Nariño, fue hallado un nuevo depósito ilegal con 476 kilos de clorhidrato de cocaína que, al parecer pertenecerían a integrantes de los autod
9	9	En un área ubicada en la frontera entre Colombia y Ecuador, Unidades de la Armada Nacional en desarrollo de una operación combinada con la Armada del Ecuador, con apoyo de la Policía Nacional de Colombia, lograron interceptar una embarc
10	10	En la última semana, la Fuerza Naval del Pacífico ha incautado más de tres toneladas de clorhidrato de cocaína y capturado a alias 'Don Camilo' o 'Rocky', principal cabecilla de una organización señalada de traficar droga en esta zona del país.E
11	11	En los últimos días, la Fuerza Naval del Pacífico ha incautado más de dos toneladas de clorhidrato de cocaína, además de localizar y destruir un semisumergible.En las últimas horas, unidades de la Armada Nacional propinaron un nuevo golpe a la
12	12	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en Buenaventura y Antioquia.En el primer decomiso, unidades de las dos fuerzas hallaron el sábado 1,5 toneladas de cocaína escondidas en contenedores en
13	13	La droga estaba dentro de una caleta, lista para ser enviada al exterior. En el sitio también se encontró una planta eléctrica y equipos de comunicaciones. Según los primeros indicios, la cocaína pertenece a bandas de narcotraficantes de diferent
14	14	Unidades de la Flotilla de Superficie y Guardacostas de la Fuerza Naval del Pacífico, lograron impedir el tráfico de éste narcótico, cuando ubicaron dos lanchas tipo Go Fast, que al notar su presencia, emprendieron la huida lanzando al mar los bul
15	15	El cargamento está avaluado en más de once millones de dólares en el mercado ilegal internacional.Incautamos más de 300 kilogramos de cocaína en el Chocó En el desarrollo de una operación de registro y control militar, la Armada Nacional in
16	16	3.159 kilos de cocaína fueron incautados durante la última semana, por la Armada Nacional, en operaciones conjuntas y coordinadas con la Fuerza Pública y demás organismos del Estado.La más reciente operación se registró ayer, cuando unida
17	17	
18	18	En menos de una semana, la Fuerza Naval del Pacífico ha incautado 1.250 kilos de cocaína evitando que ingresaran a las estructuras narcotraficantes cerca de 38 millones de dólares.En un nuevo golpe a las organizaciones narcotraficantes que d
19	19	1.098 kilos de clorhidrato de cocaína incautados y la destrucción de dos campamentos para 30 personas, es el resultado hasta el momento.En las últimas horas, en el marco de la ofensiva que sostienen las tropas de la Armada Nacional contra la
20	20	En los últimos días, la Armada Nacional ha logrado la incautación de cerca de dos toneladas de alcaloides en el Caribe y Pacífico colombiano. Unidades de la Armada Nacional lograron en las últimas horas la incautación de 379 kilogramos de clor
21	21	505 kilogramos de clorhidrato de cocaína fueron incautados en una nueva operación conjunta y coordinada entre la Armada Nacional, la Fuerza Aérea Colombiana y el CTI. Una nueva operación contra el narcotráfico en el Caribe se llevó a cabo
22	22	Las incautaciones se desarrollaron en el estero San Antonio, donde las Unidades de Guardacostas interceptaron ambas embarcaciones; las cuales estaban adecuadas para transportar la droga en ingeniosas caletas bajo la cubierta. En total, fuerc
23	23	
24	24	Unidades de Superficie de la Armada Nacional desplegadas adelantando labores de control de tráfico marítimo en un área cercana a Cabo Manglares, en el departamento de Nariño, interceptaron dos embarcaciones tripuladas por seis extranjer
25	25	Unidades de la Armada Nacional en desarrollo de operaciones de control territorial en el departamento de Nariño, lograron la incautación de 1.123 kilos de clorhidrato de cocaína ocultos en zona de manglar, en área rural del municipio de Tumac
26	26	Unidades de la Armada Nacional y de la Policía en el municipio de Turbo, incautaron 98.6 kilogramos de cocaína que pretendían ser enviados hacia Europa, en un cilindro pegado debajo del casco del buque mercante de nombre "Santa María" y b
27	27	Tropas de la Armada Nacional incautaron más de 200 kilos de clorhidrato de cocaína que estaban escondidos en un depósito ilegal que fue hallado en una zona rural de la costa Pacífica nariñense.La operación fue realizada por tropas de la Fuerz
28	28	En el último mes, la Fuerza Naval del Pacífico ha incautado cerca de tres toneladas de alcaloides, entre clorhidrato de cocaína y marihuana "creepy" al "Clan Úsuga"En las últimas horas, unidades de la Armada Nacional propinaron un nuevo golpe
29	29	En lo que va corrido del año, la Armada Nacional ha logrado incautar más de 40 toneladas de clorhidrato de cocaína en todo el Pacífico colombiano.En las últimas horas, después de una persecución marítima por parte de unidades de la Armada
30	30	• La ofensiva que sostiene la Armada Nacional en contra de las organizaciones narcotraficantes en el territorio colombiano, permitió en los últimos días la incautación de 2.233 kilogramos de cocaína.• La Armada Nacional lograron asestar golpe
31	31	En las últimas horas, unidades de la Armada Nacional que estaban desplegadas adelantando labores de control de tráfico marítimo en un área cercana a Cabo Manglares, en el departamento de Nariño, interceptaron dos embarcaciones ilegales t
32	32	Más de 840 millones de dólares ha evitado la Armada Nacional que ingresen a las estructuras del narcotráfico con la incautación de cerca de 28 toneladas de clorhidrato de cocaína, en lo corrido del año.En las últimas horas, la Armada Nacional

Main | Main.mainChildren

Ready | Accessibility: Investigate | Display Settings | 100%

The algorithm in the second level of scraping is creating empty cells instead of extracting the news content.

Helium Scraper -> USCG website



For the USCG case, the algorithms in the first second level of scraping are creating empty cells instead of extracting the intended data.

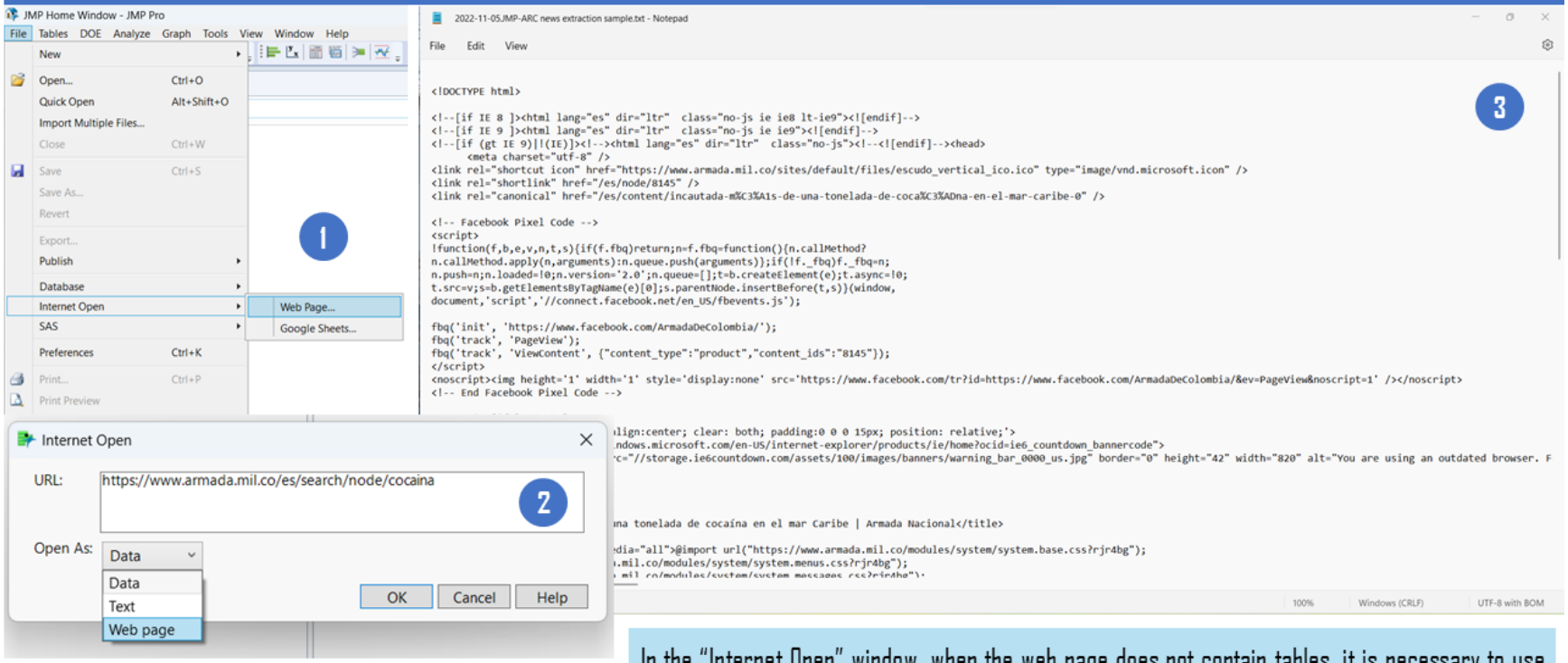
Helium Scraper -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
PARTIAL	NO	PARTIAL	Easy	38.70	150.00	6.93	7.53	YES	YES	2000	High	High	NO
NO			3	188.70		14.47		1	3	1	3	3	1

Remarks:

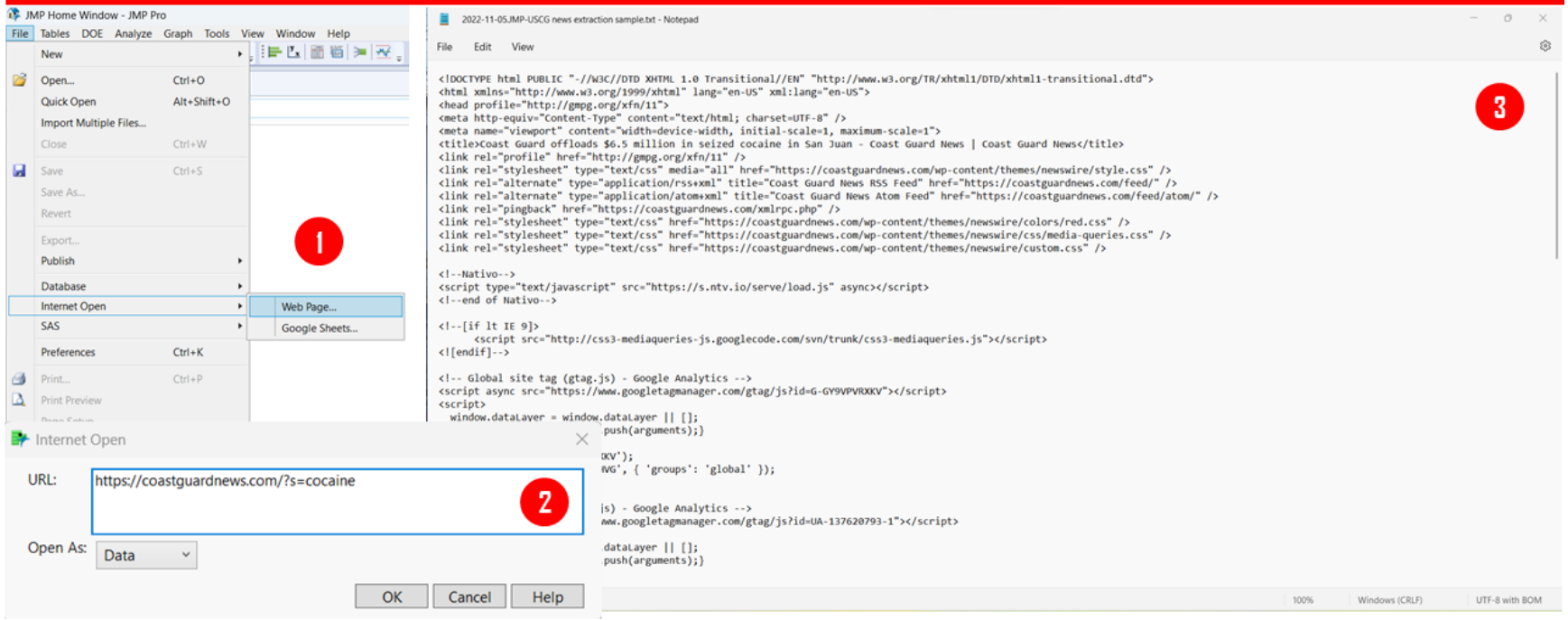
1. This program requires the creation of the main and child programs for the different levels of scrpaing.
2. The automatic scraping is partially done because in the USCG website is not extracting data properly.
3. The scraping depth is partially done because in the ARC and USCG websites is not extracting data properly.
4. The documentation and the interface did not relate a function about the IP rotation function.
5. It is shown in the preliminary results that the language fidelity for Spanish and English was not good because some HTML coding traces remained in the date field after the extraction.

JMP -> ARC website



In the "Internet Open" window, when the web page does not contain tables, it is necessary to use the "Web page" option to load the page source. This scraping should be done manually for each web page.

JMP -> USCG website



In the "Internet Open" window, when the web page does not contain tables, it is necessary to use the "Web page" option to load the page source. This scraping should be done manually for each web page.

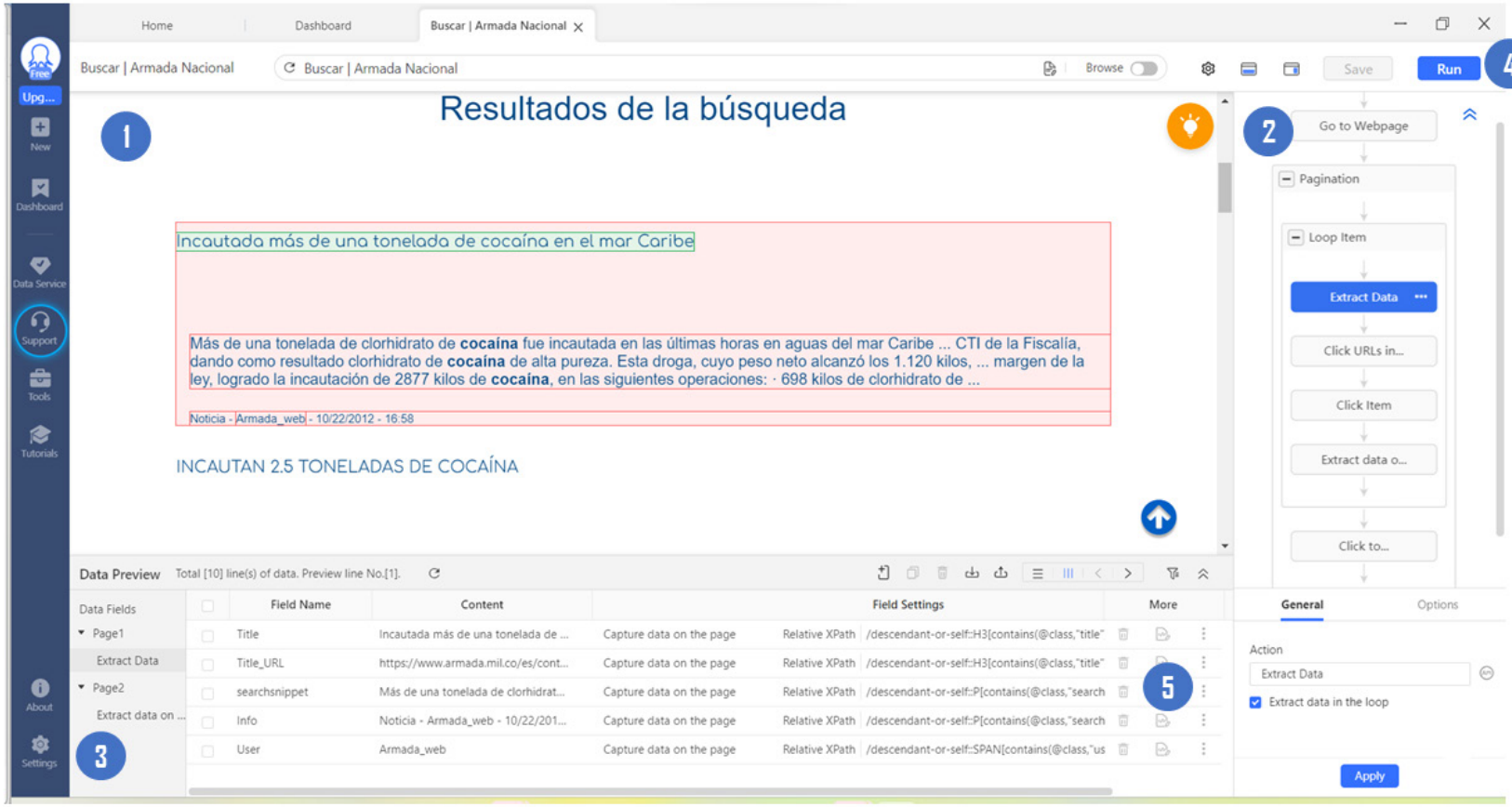
JMP -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
NO	NO	NO	Easy	15.00	15.00	50.00	33.33	NO	NO	NO	Low	Low	NO
NO			3	30.00		83.33		0	0	0	1	1	1

Remarks:

1. Due to the manual and individual extraction of page source for each web page, this program cannot do automatic scraping, automatic crawler nor scraping depth.
2. The scraping procedure is manual, therefore there was not any human behavior feature.
3. The language fidelity is low in both languages considering the only possible extraction is HTML (source page).

Octoparse -> ARC website



Octoparse -> ARC website results

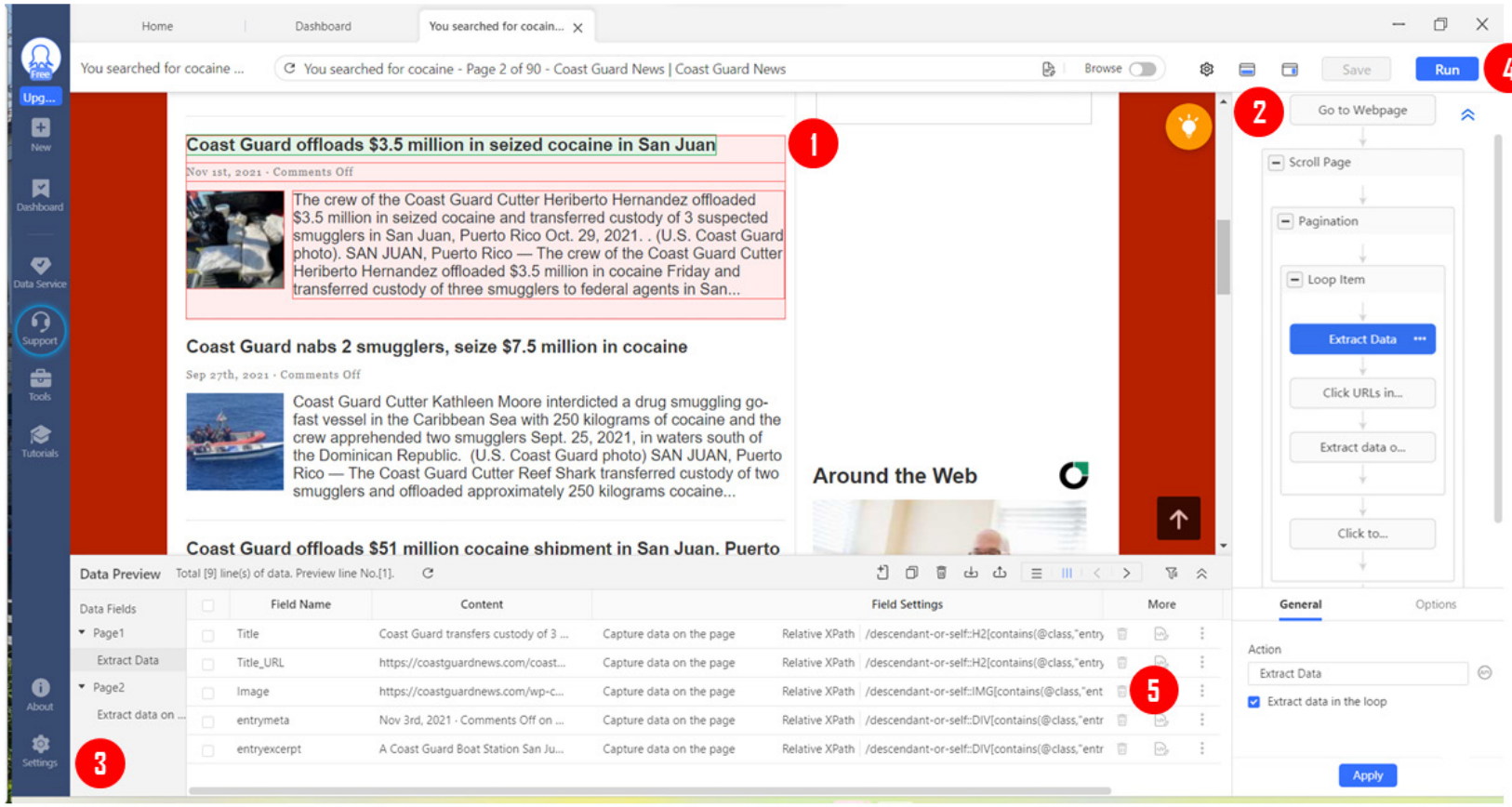
	A	B	D	E	F
1	Title	Title_URL	Info	User	Text
2	Incautada más de una tonelada de cocaína en el mar	https://www.armada.mil.co/es/c	Noticia - Armada_web - 10/22/2012 - 16:5	Armada_web	Más de una tonelada de clorhidrato de coc
3	INCAUTAN 2.5 TONELADAS DE COCAÍNA	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:3	Armada_web	La Armada y la Policía confiscaron 2,5 tone
4	CONFISCADAS 2.5 TONELADAS DE COCAÍNA	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:3	Armada_web	La Armada y la Policía confiscaron 2,5 tone
5	Armada Nacional incauta cerca de dos toneladas de c	https://www.armada.mil.co/es/c	Noticia - elozano - 08/10/2016 - 11:35	elozano	En lo corrido del año, se han incautado má
6	Armada Nacional incauta cocaína en el Caribe Colom	https://www.armada.mil.co/es/c	Noticia - Armada_web - 01/22/2013 - 10:3	Armada_web	En menos de una semana, la Armada Naci
7	Armada Nacional incauta cocaína en el Caribe Colom	https://www.armada.mil.co/es/c	Noticia - Armada_web - 10/15/2013 - 11:1	Armada_web	En menos de una semana, la Armada Naci
8	Más de una tonelada de cocaína incautada en el Paci	https://www.armada.mil.co/es/c	Noticia - elozano - 01/22/2016 - 14:13	elozano	En la última semana, la Institución Naval
9	Incautada cerca de media tonelada de cocaína a los	https://www.armada.mil.co/es/c	Noticia - elozano - 11/23/2016 - 10:51	elozano	En una operación que adelanta la Armada
10	Incautado cargamento de cocaína en Operación Nava	https://www.armada.mil.co/es/c	Noticia - elozano - 12/22/2015 - 14:34	elozano	En un área ubicada en la frontera entre Co
11	Más de dos toneladas de clorhidrato de cocaína inca	https://www.armada.mil.co/es/c	Noticia - elozano - 08/30/2016 - 10:08	elozano	En la última semana, la Fuerza Naval del P
12	Más de una tonelada de cocaína en el Pacífico colom	https://www.armada.mil.co/es/c	Noticia - elozano - 04/06/2016 - 15:39	elozano	En los últimos días, la Fuerza Naval del Pa
13	Fuerza Naval del Pacífico incauta más de 700 kilos de	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:2	Armada_web	Unidades de la Flotilla de Superficie y Gua
14	Incautamos más de 300 kilogramos de cocaína en el	https://www.armada.mil.co/es/c	Noticia - elozano - 09/27/2015 - 11:02	elozano	El cargamento está avaluado en más de o
15	Cerca de una tonelada de cocaína incautada en el Pa	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:2	Armada_web	3.159 kilos de cocaína fueron incautados c
16	CONFISCAN 2.5 TONELADAS DE COCAÍNA EN DOS OF	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:3	Armada_web	La Armada y la Policía confiscaron 2,5 tone
17	Incautadas 1.9 toneladas más de cocaína en el Choc	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:3	Armada_web	La droga estaba dentro de una caleta, lista
18	Incautados más de 2.200 kilos de cocaína en el mar (	https://www.armada.mil.co/es/c	Noticia - Armada_web - 08/24/2012 - 07:3	Armada_web	En menos de una semana, han sido incaut
19	Armada Nacional incauta 850 kilos de cocaína en el g	https://www.armada.mil.co/es/c	Noticia - elozano - 05/02/2016 - 14:22	elozano	En menos de una semana, la Fuerza Naval
20	Incautada más de una tonelada de cocaína de las Far	https://www.armada.mil.co/es/c	Noticia - elozano - 11/13/2014 - 16:36	elozano	1.098 kilos de clorhidrato de cocaína incau
21	Incautada cocaína en un buque carbonero en la Guaji	https://www.armada.mil.co/es/c	Noticia - elozano - 02/22/2016 - 15:35	elozano	En los últimos días, la Armada Nacional ha
22	Se incautaron media tonelada de cocaína en el Caribe	https://www.armada.mil.co/es/c	Noticia - elozano - 07/31/2014 - 17:17	elozano	505 kilogramos de clorhidrato de cocaína
23	Decomisadas más de dos toneladas de cocaína y cua	https://www.armada.mil.co/es/c	Noticia - Armada_web - 07/23/2012 - 01:3	Armada_web	Las incautaciones se desarrollaron en el e
24	Incautada más de una tonelada de cocaína en el mar	https://www.armada.mil.co/es/c	Noticia - Armada_web - 09/24/2012 - 10:2	Armada_web	Ayer unidades de la Armada Nacional, inca
25	Incautadas 1.3 Toneladas de clorhidrato de cocaína e	https://www.armada.mil.co/es/c	Noticia - elozano - 09/26/2017 - 16:36	elozano	Unidades de Superficie de la Armada Naci
26	Al menos una tonelada de cocaína incautada a las Fa	https://www.armada.mil.co/es/c	Noticia - elozano - 12/08/2014 - 15:58	elozano	Unidades de la Armada Nacional en desar
27	Cerca de 100 kilogramos de cocaína incautados en U	https://www.armada.mil.co/es/c	Noticia - elozano - 04/13/2014 - 11:10	elozano	Unidades de la Armada Nacional y de la Po
28	Armada Nacional incautó más de 200 kilos de Clorhi	https://www.armada.mil.co/es/c	Noticia - elozano - 10/03/2016 - 14:23	elozano	Tropas de la Armada Nacional incautaron
29	Incautado cargamento de cocaína perteneciente al "(	https://www.armada.mil.co/es/c	Noticia - elozano - 03/11/2016 - 09:50	elozano	En el último mes, la Fuerza Naval del Pacíf
30	Capturado guatemalteco, ecuatoriano y colombiano c	https://www.armada.mil.co/es/c	Noticia - elozano - 05/08/2017 - 16:22	elozano	En lo que va corrido del año, la Armada Na
31	Incautamos más de 800 kilos de cocaína en el océan	https://www.armada.mil.co/es/c	Noticia - elozano - 01/26/2017 - 04:34	elozano	• La ofensiva que sostiene la Armada Naci
32	Incautada cerca de una tonelada y media de clorhidr	https://www.armada.mil.co/es/c	Noticia - elozano - 09/28/2017 - 09:17	elozano	En las últimas horas, unidades de la Arma
33	Incautadas cerca de 1,5 toneladas de cocaína en el p	https://www.armada.mil.co/es/c	Noticia - elozano - 05/10/2016 - 07:42	elozano	Más de 840 millones de dólares ha evitad
34	Hallada cocaína que pretendía ser enviada de forma	https://www.armada.mil.co/es/c	Noticia - elozano - 11/16/2015 - 08:06	elozano	El hallazgo fue realizado por buzos de la A
35	Armada Nacional y Ejército incautaron al ELN más de	https://www.armada.mil.co/es/c	Noticia - elozano - 05/26/2016 - 08:09	elozano	El alcaloide había sido mezclado con comb
36	Cae cargamento con más de media tonelada de coca	https://www.armada.mil.co/es/c	Noticia - elozano - 05/04/2016 - 15:50	elozano	En el 2.016 se han incautado más de 18 to
37	Interceptada motonave cargada con más de media to	https://www.armada.mil.co/es/c	Noticia - elozano - 03/16/2015 - 10:37	elozano	Con las incautaciones de clorhidrato de co
38	144 kilos de cocaína han sido incautados en promedi	https://www.armada.mil.co/es/c	Noticia - elozano - 10/04/2015 - 15:37	elozano	El balance en la lucha contra el narcotráfic
39	MÁS DE 2,6 TONELADAS DE COCAÍNA INCAUTADAS F	https://www.armada.mil.co/es/c	Noticia - elozano - 11/14/2016 - 09:14	elozano	La primera incautación se dio en el área g
40	Incautada cocaína en el golfo de Urabá	https://www.armada.mil.co/es/c	Noticia - Armada_web - 03/13/2013 - 14:5	Armada_web	Unidades de la Armada Nacional incautard
41	Más de dos toneladas de cocaína incautadas en el m	https://www.armada.mil.co/es/c	Noticia - Armada_web - 09/26/2012 - 15:5	Armada_web	2.345 kilos de cocaína fueron incautados
42	Incautada cocaína en un buque en Cartagena	https://www.armada.mil.co/es/c	Noticia - elozano - 04/20/2016 - 16:09	elozano	La Armada Nacional y la Policía Nacional i

This program extracted data into a unique table including the first and second scraping levels.

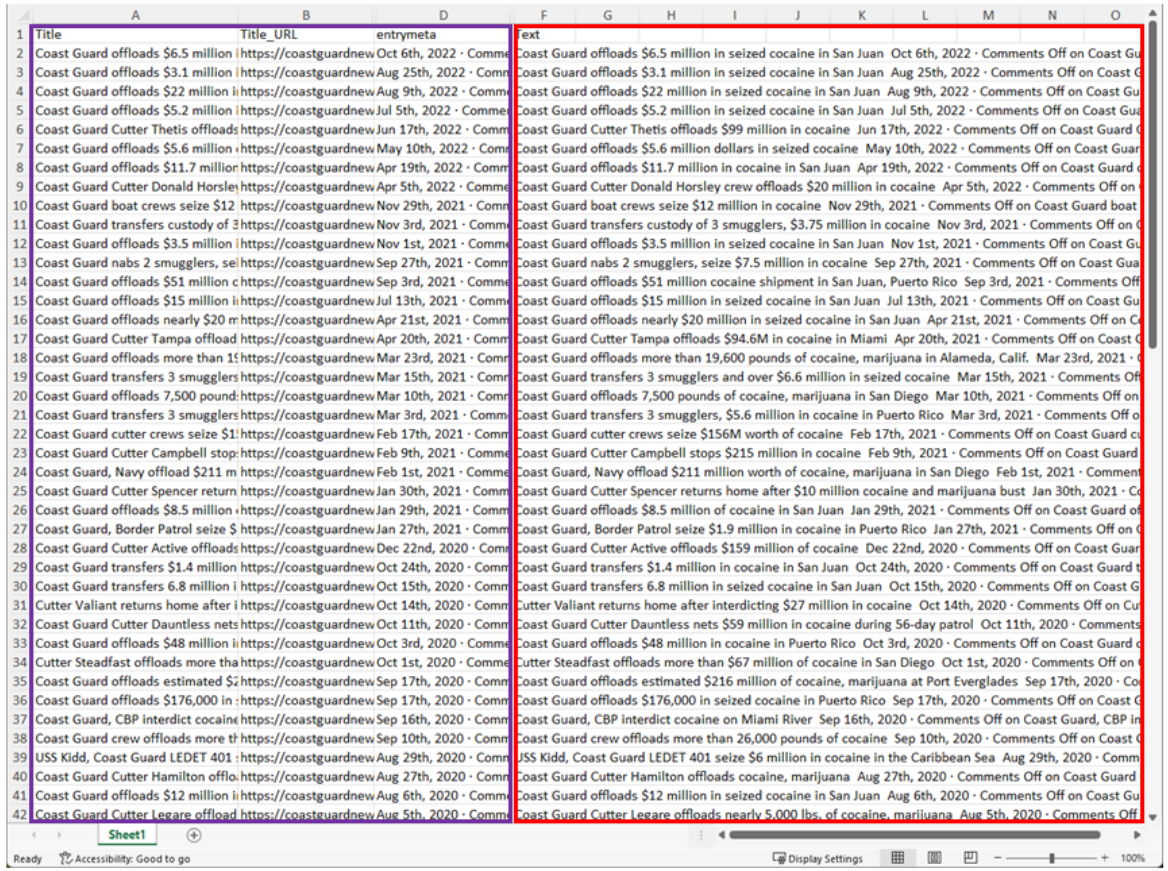
First scraping level

Second scraping level

Octoparse -> USCG website



Octoparse -> USCG website results



	A	B	D	F
1	Title	Title_URL	entrymeta	Text
2	Coast Guard offloads $6.5 million i	https://coastguardnew	Oct 6th, 2022 · Comme	Coast Guard offloads $6.5 million in seized cocaine in San Juan Oct 6th, 2022 · Comments Off on Coast Gu
3	Coast Guard offloads $3.1 million i	https://coastguardnew	Aug 25th, 2022 · Comn	Coast Guard offloads $3.1 million in seized cocaine in San Juan Aug 25th, 2022 · Comments Off on Coast G
4	Coast Guard offloads $22 million i	https://coastguardnew	Aug 9th, 2022 · Comme	Coast Guard offloads $22 million in seized cocaine in San Juan Aug 9th, 2022 · Comments Off on Coast Gu
5	Coast Guard offloads $5.2 million i	https://coastguardnew	Jul 5th, 2022 · Comme	Coast Guard offloads $5.2 million in seized cocaine in San Juan Jul 5th, 2022 · Comments Off on Coast Gua
6	Coast Guard Cutter Thetis offloads	https://coastguardnew	Jun 17th, 2022 · Comm	Coast Guard Cutter Thetis offloads $99 million in cocaine Jun 17th, 2022 · Comments Off on Coast Guard C
7	Coast Guard offloads $5.6 million	https://coastguardnew	May 10th, 2022 · Com	Coast Guard offloads $5.6 million dollars in seized cocaine May 10th, 2022 · Comments Off on Coast Guar
8	Coast Guard offloads $11.7 million	https://coastguardnew	Apr 19th, 2022 · Comm	Coast Guard offloads $11.7 million in cocaine in San Juan Apr 19th, 2022 · Comments Off on Coast Guard c
9	Coast Guard Cutter Donald Horsley	https://coastguardnew	Apr 5th, 2022 · Comme	Coast Guard Cutter Donald Horsley crew offloads $20 million in cocaine Apr 5th, 2022 · Comments Off on
10	Coast Guard boat crews seize $12	https://coastguardnew	Nov 29th, 2021 · Comn	Coast Guard boat crews seize $12 million in cocaine Nov 29th, 2021 · Comments Off on Coast Guard boat
11	Coast Guard transfers custody of 3	https://coastguardnew	Nov 3rd, 2021 · Comme	Coast Guard transfers custody of 3 smugglers, $3.75 million in cocaine Nov 3rd, 2021 · Comments Off on C
12	Coast Guard offloads $3.5 million i	https://coastguardnew	Nov 1st, 2021 · Comme	Coast Guard offloads $3.5 million in seized cocaine in San Juan Nov 1st, 2021 · Comments Off on Coast Gu
13	Coast Guard nabs 2 smugglers, sei	https://coastguardnew	Sep 27th, 2021 · Comn	Coast Guard nabs 2 smugglers, seize $7.5 million in cocaine Sep 27th, 2021 · Comments Off on Coast Gua
14	Coast Guard offloads $51 million c	https://coastguardnew	Sep 3rd, 2021 · Comme	Coast Guard offloads $51 million cocaine shipment in San Juan, Puerto Rico Sep 3rd, 2021 · Comments Off
15	Coast Guard offloads $15 million i	https://coastguardnew	Jul 13th, 2021 · Comme	Coast Guard offloads $15 million in seized cocaine in San Juan Jul 13th, 2021 · Comments Off on Coast Gu
16	Coast Guard offloads nearly $20 m	https://coastguardnew	Apr 21st, 2021 · Comm	Coast Guard offloads nearly $20 million in seized cocaine in San Juan Apr 21st, 2021 · Comments Off on C
17	Coast Guard Cutter Tampa offload	https://coastguardnew	Apr 20th, 2021 · Comm	Coast Guard Cutter Tampa offloads $94.6M in cocaine in Miami Apr 20th, 2021 · Comments Off on Coast G
18	Coast Guard offloads more than 19	https://coastguardnew	Mar 23rd, 2021 · Comn	Coast Guard offloads more than 19,600 pounds of cocaine, marijuana in Alameda, Calif. Mar 23rd, 2021 ·
19	Coast Guard transfers 3 smugglers	https://coastguardnew	Mar 15th, 2021 · Comn	Coast Guard transfers 3 smugglers and over $6.6 million in seized cocaine Mar 15th, 2021 · Comments Off
20	Coast Guard offloads 7,500 pound	https://coastguardnew	Mar 10th, 2021 · Comn	Coast Guard offloads 7,500 pounds of cocaine, marijuana in San Diego Mar 10th, 2021 · Comments Off on
21	Coast Guard transfers 3 smugglers	https://coastguardnew	Mar 3rd, 2021 · Comm	Coast Guard transfers 3 smugglers, $5.6 million in cocaine in Puerto Rico Mar 3rd, 2021 · Comments Off o
22	Coast Guard cutter crews seize $15	https://coastguardnew	Feb 17th, 2021 · Comn	Coast Guard cutter crews seize $156M worth of cocaine Feb 17th, 2021 · Comments Off on Coast Guard cu
23	Coast Guard Cutter Campbell stop	https://coastguardnew	Feb 9th, 2021 · Comme	Coast Guard Cutter Campbell stops $215 million in cocaine Feb 9th, 2021 · Comments Off on Coast Guard
24	Coast Guard, Navy offload $211 m	https://coastguardnew	Feb 1st, 2021 · Comme	Coast Guard, Navy offload $211 million worth of cocaine, marijuana in San Diego Feb 1st, 2021 · Comment
25	Coast Guard Cutter Spencer return	https://coastguardnew	Jan 30th, 2021 · Comm	Coast Guard Cutter Spencer returns home after $10 million cocaine and marijuana bust Jan 30th, 2021 · Co
26	Coast Guard offloads $8.5 million	https://coastguardnew	Jan 29th, 2021 · Comm	Coast Guard offloads $8.5 million of cocaine in San Juan Jan 29th, 2021 · Comments Off on Coast Guard of
27	Coast Guard, Border Patrol seize $	https://coastguardnew	Jan 27th, 2021 · Comm	Coast Guard, Border Patrol seize $1.9 million in cocaine in Puerto Rico Jan 27th, 2021 · Comments Off on C
28	Coast Guard Cutter Active offloads	https://coastguardnew	Dec 22nd, 2020 · Com	Coast Guard Cutter Active offloads $159 million of cocaine Dec 22nd, 2020 · Comments Off on Coast Guar
29	Coast Guard transfers $1.4 million	https://coastguardnew	Oct 24th, 2020 · Comm	Coast Guard transfers $1.4 million in cocaine in San Juan Oct 24th, 2020 · Comments Off on Coast Guard t
30	Coast Guard transfers 6.8 million i	https://coastguardnew	Oct 15th, 2020 · Comm	Coast Guard transfers 6.8 million in seized cocaine in San Juan Oct 15th, 2020 · Comments Off on Coast G
31	Cutter Valiant returns home after i	https://coastguardnew	Oct 14th, 2020 · Comn	Cutter Valiant returns home after interdicting $27 million in cocaine Oct 14th, 2020 · Comments Off on Cu
32	Coast Guard Cutter Dauntless nets	https://coastguardnew	Oct 11th, 2020 · Comn	Coast Guard Cutter Dauntless nets $59 million in cocaine during 56-day patrol Oct 11th, 2020 · Comments
33	Coast Guard offloads $48 million i	https://coastguardnew	Oct 3rd, 2020 · Comme	Coast Guard offloads $48 million in cocaine in Puerto Rico Oct 3rd, 2020 · Comments Off on Coast Guard c
34	Cutter Steadfast offloads more tha	https://coastguardnew	Oct 1st, 2020 · Comme	Cutter Steadfast offloads more than $67 million of cocaine in San Diego Oct 1st, 2020 · Comments Off on
35	Coast Guard offloads estimated $2	https://coastguardnew	Sep 17th, 2020 · Comn	Coast Guard offloads estimated $216 million of cocaine, marijuana at Port Everglades Sep 17th, 2020 · Co
36	Coast Guard offloads $176,000 in	https://coastguardnew	Sep 17th, 2020 · Comn	Coast Guard offloads $176,000 in seized cocaine in Puerto Rico Sep 17th, 2020 · Comments Off on Coast G
37	Coast Guard, CBP interdict cocaine	https://coastguardnew	Sep 16th, 2020 · Comn	Coast Guard, CBP interdict cocaine on Miami River Sep 16th, 2020 · Comments Off on Coast Guard, CBP in
38	Coast Guard crew offloads more th	https://coastguardnew	Sep 10th, 2020 · Comn	Coast Guard crew offloads more than 26,000 pounds of cocaine Sep 10th, 2020 · Comments Off on Coast G
39	USS Kidd, Coast Guard LEDET 401	https://coastguardnew	Aug 29th, 2020 · Comn	USS Kidd, Coast Guard LEDET 401 seize $6 million in cocaine in the Caribbean Sea Aug 29th, 2020 · Comm
40	Coast Guard Cutter Hamilton offlo	https://coastguardnew	Aug 27th, 2020 · Comn	Coast Guard Cutter Hamilton offloads cocaine, marijuana Aug 27th, 2020 · Comments Off on Coast Guard
41	Coast Guard offloads $12 million i	https://coastguardnew	Aug 6th, 2020 · Comme	Coast Guard offloads $12 million in seized cocaine in San Juan Aug 6th, 2020 · Comments Off on Coast Gu
42	Coast Guard Cutter Legare offload	https://coastguardnew	Aug 5th, 2020 · Comme	Coast Guard Cutter Legare offloads nearly 5,000 lbs. of cocaine, marijuana Aug 5th, 2020 · Comments Off

Sheet1

Ready Accessibility: Good to go Display Settings 100%

This program extracted data into a unique table including the first and second scraping levels.

First scraping level

Second scraping level

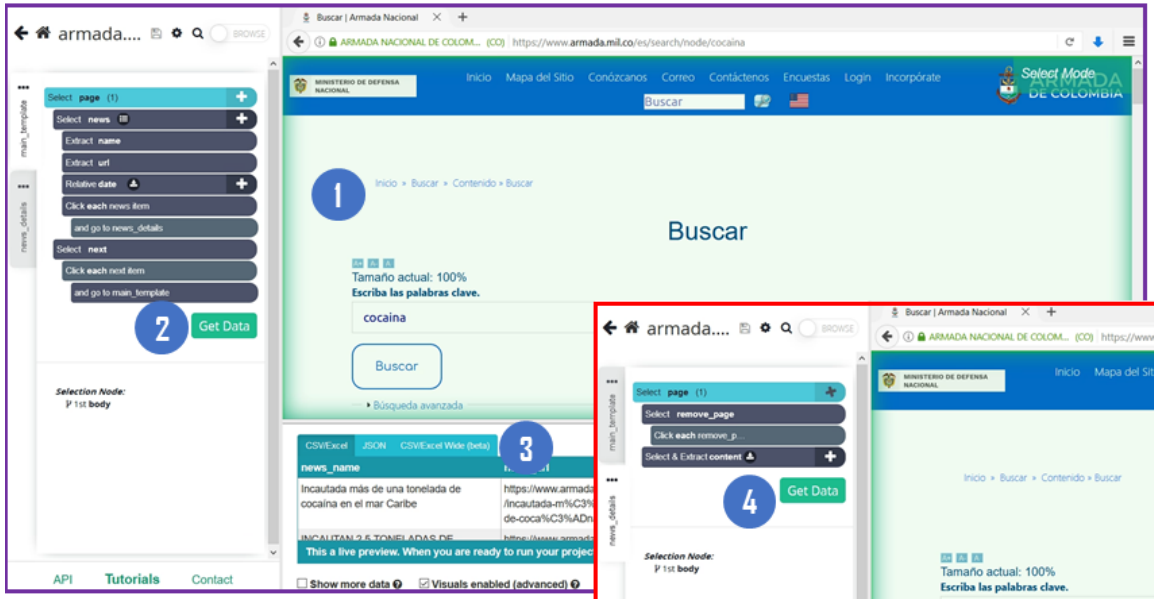
Octoparse -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
YES	YES	YES	Easy	63.30	120.00	6.18	29.32	YES	YES	YES	High	High	YES
YES			3	183.30		35.50		1	1	1	3	3	0

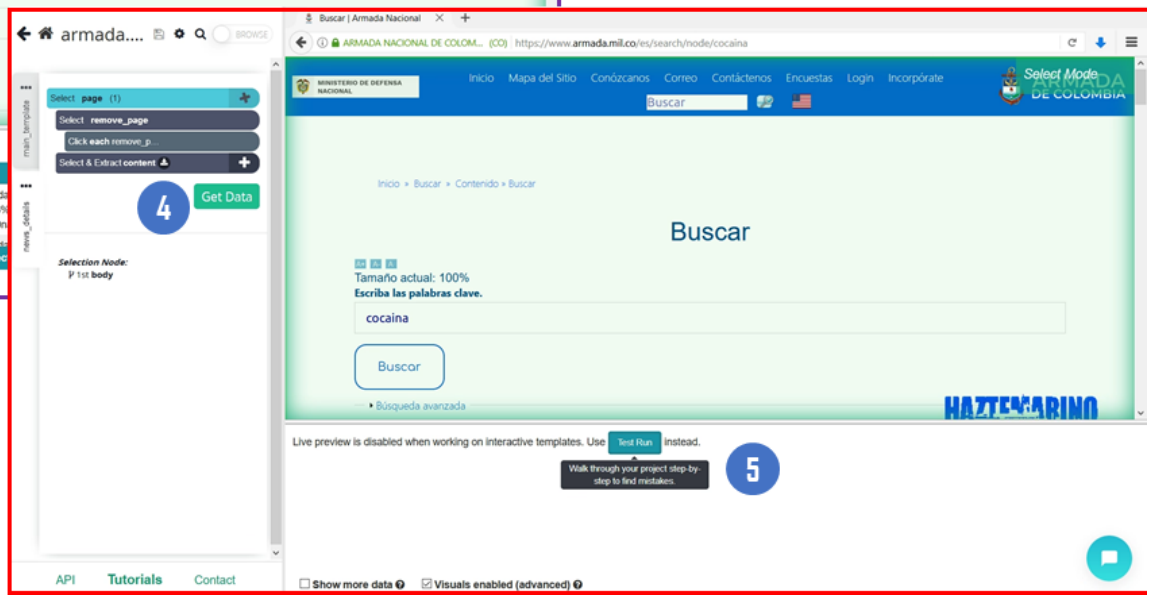
Remarks:

1. Octoparse uses the philosophy "Point-Click-Extract" to create the scraping and crawling tasks.
2. A manual Xpath correction should be done to achieve a proper scraping depth.
3. This program automatically detected the duplicates and provided an optional elimination of them.

Parsehub -> ARC website



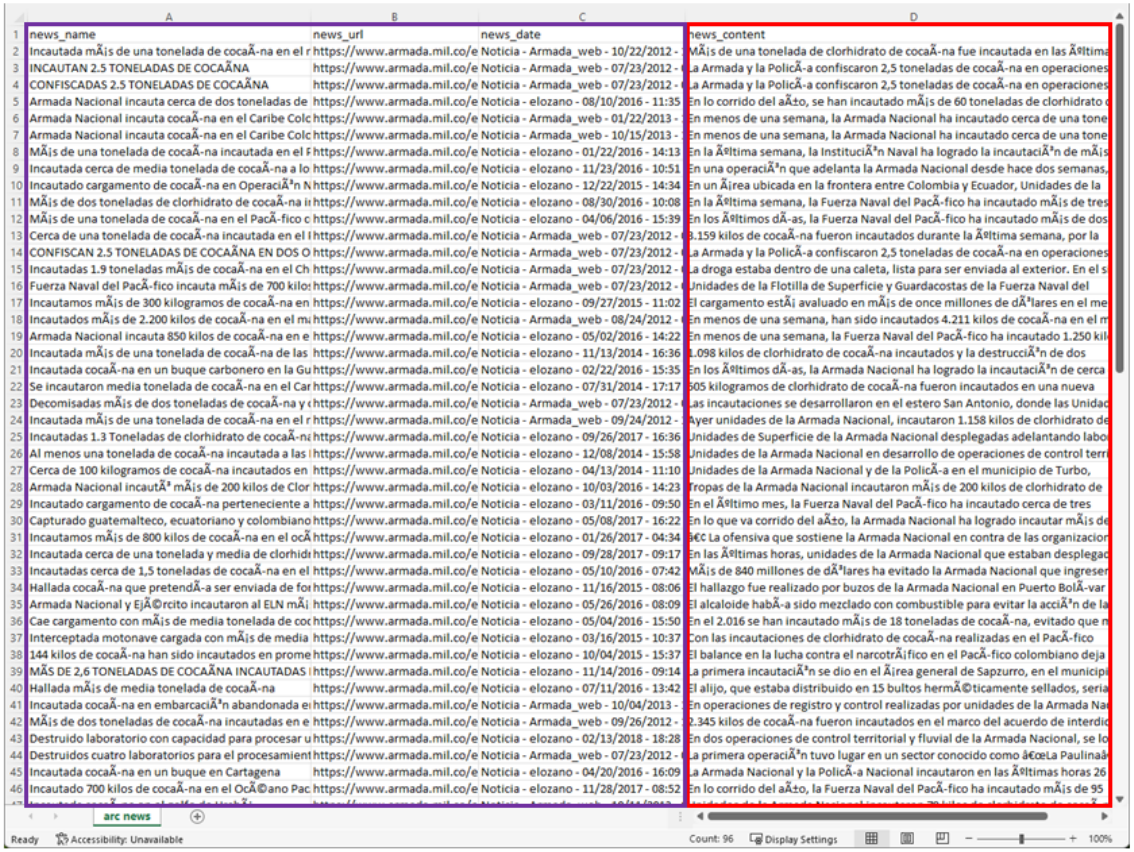
This program required a different algorithm for each scraping level.



First scraping level program

Second scraping level program

Parsehub -> ARC website results



	A	B	C	D
1	news_name	news_url	news_date	news_content
2	Incautada mÃ¡s de una tonelada de cocaÃ-na en el r	https://www.armada.mil.co/e	Noticia - Armada_web - 10/22/2012 -	MÃ¡s de una tonelada de clorhidrato de cocaÃ-na fue incautada en las Ã°ltima
3	INCAUTAN 2.5 TONELADAS DE COCAÃNA	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	La Armada y la PolicÃ-a confiscaron 2,5 toneladas de cocaÃ-na en operaciones
4	CONFISCADAS 2.5 TONELADAS DE COCAÃNA	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	La Armada y la PolicÃ-a confiscaron 2,5 toneladas de cocaÃ-na en operaciones
5	Armada Nacional incauta cerca de dos toneladas de	https://www.armada.mil.co/e	Noticia - elozano - 08/10/2016 - 11:35	En lo corrido del aÃ±o, se han incautado mÃ¡s de 60 toneladas de clorhidrato d
6	Armada Nacional incauta cocaÃ-na en el Caribe Colc	https://www.armada.mil.co/e	Noticia - Armada_web - 01/22/2013 -	En menos de una semana, la Armada Nacional ha incautado cerca de una tone
7	Armada Nacional incauta cocaÃ-na en el Caribe Colc	https://www.armada.mil.co/e	Noticia - Armada_web - 10/15/2013 -	En menos de una semana, la Armada Nacional ha incautado cerca de una tone
8	MÃ¡s de una tonelada de cocaÃ-na incautada en el F	https://www.armada.mil.co/e	Noticia - elozano - 01/22/2016 - 14:13	En la Ã°ltima semana, la InstituciÃ³n Naval ha logrado la incautaciÃ³n de mÃ¡s
9	Incautada cerca de media tonelada de cocaÃ-na a lo	https://www.armada.mil.co/e	Noticia - elozano - 11/23/2016 - 10:51	En una operaciÃ³n que adelanta la Armada Nacional desde hace dos semanas,
10	Incautado cargamento de cocaÃ-na en OperaciÃ³n N	https://www.armada.mil.co/e	Noticia - elozano - 12/22/2015 - 14:34	En un Ã¡rea ubicada en la frontera entre Colombia y Ecuador, Unidades de la
11	MÃ¡s de dos toneladas de clorhidrato de cocaÃ-na in	https://www.armada.mil.co/e	Noticia - elozano - 08/30/2016 - 10:08	En la Ã°ltima semana, la Fuerza Naval del PacÃ-fico ha incautado mÃ¡s de tres
12	MÃ¡s de una tonelada de cocaÃ-na en el PacÃ-fico c	https://www.armada.mil.co/e	Noticia - elozano - 04/06/2016 - 15:39	En los Ã°ltimos dÃ-as, la Fuerza Naval del PacÃ-fico ha incautado mÃ¡s de dos
13	Cerca de una tonelada de cocaÃ-na incautada en el I	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	8.159 kilos de cocaÃ-na fueron incautados durante la Ã°ltima semana, por la
14	CONFISCAN 2.5 TONELADAS DE COCAÃNA EN DOS O	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	La Armada y la PolicÃ-a confiscaron 2,5 toneladas de cocaÃ-na en operaciones
15	Incautadas 1.9 toneladas mÃ¡s de cocaÃ-na en el Ch	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	La droga estaba dentro de una caleta, lista para ser enviada al exterior. En el s
16	Fuerza Naval del PacÃ-fico incauta mÃ¡s de 700 kilo	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	Unidades de la Flotilla de Superficie y Guardacostas de la Fuerza Naval del
17	Incautamos mÃ¡s de 300 kilogramos de cocaÃ-na en	https://www.armada.mil.co/e	Noticia - elozano - 09/27/2015 - 11:02	El cargamento estÃ¡ avaluado en mÃ¡s de once millones de dÃ³lares en el me
18	Incautados mÃ¡s de 2.200 kilos de cocaÃ-na en el mi	https://www.armada.mil.co/e	Noticia - Armada_web - 08/24/2012 -	En menos de una semana, han sido incautados 4.211 kilos de cocaÃ-na en el m
19	Armada Nacional incauta 850 kilos de cocaÃ-na en e	https://www.armada.mil.co/e	Noticia - elozano - 05/02/2016 - 14:22	En menos de una semana, la Fuerza Naval del PacÃ-fico ha incautado 1.250 kil
20	Incautada mÃ¡s de una tonelada de cocaÃ-na de las	https://www.armada.mil.co/e	Noticia - elozano - 11/13/2014 - 16:36	1.098 kilos de clorhidrato de cocaÃ-na incautados y la destrucciÃ³n de dos
21	Incautada cocaÃ-na en un buque carbonero en la Gu	https://www.armada.mil.co/e	Noticia - elozano - 02/22/2016 - 15:35	En los Ã°ltimos dÃ-as, la Armada Nacional ha logrado la incautaciÃ³n de cerca
22	Se incautaron media tonelada de cocaÃ-na en el Car	https://www.armada.mil.co/e	Noticia - elozano - 07/31/2014 - 17:17	505 kilogramos de clorhidrato de cocaÃ-na fueron incautados en una nueva
23	Decomisadas mÃ¡s de dos toneladas de cocaÃ-na y (	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	Las incautaciones se desarrollaron en el estero San Antonio, donde las Unidac
24	Incautada mÃ¡s de una tonelada de cocaÃ-na en el r	https://www.armada.mil.co/e	Noticia - Armada_web - 09/24/2012 -	Ayer unidades de la Armada Nacional, incautaron 1.158 kilos de clorhidrato de
25	Incautadas 1.3 Toneladas de clorhidrato de cocaÃ-na	https://www.armada.mil.co/e	Noticia - elozano - 09/26/2017 - 16:36	Unidades de Superficie de la Armada Nacional desplegadas adelantando labo
26	Al menos una tonelada de cocaÃ-na incautada a las	https://www.armada.mil.co/e	Noticia - elozano - 12/08/2014 - 15:58	Unidades de la Armada Nacional en desarrollo de operaciones de control terr
27	Cerca de 100 kilogramos de cocaÃ-na incautados en	https://www.armada.mil.co/e	Noticia - elozano - 04/13/2014 - 11:10	Unidades de la Armada Nacional y de la PolicÃ-a en el municipio de Turbo,
28	Armada Nacional incautÃ³ mÃ¡s de 200 kilos de Clor	https://www.armada.mil.co/e	Noticia - elozano - 10/03/2016 - 14:23	Tropas de la Armada Nacional incautaron mÃ¡s de 200 kilos de clorhidrato de
29	Incautado cargamento de cocaÃ-na perteneciente a	https://www.armada.mil.co/e	Noticia - elozano - 03/11/2016 - 09:50	En el Ã°ltimo mes, la Fuerza Naval del PacÃ-fico ha incautado cerca de tres
30	Capturado guatemalteco, ecuatoriano y colombiano	https://www.armada.mil.co/e	Noticia - elozano - 05/08/2017 - 16:22	En lo que va corrido del aÃ±o, la Armada Nacional ha logrado incautar mÃ¡s de
31	Incautamos mÃ¡s de 800 kilos de cocaÃ-na en el ocÃ	https://www.armada.mil.co/e	Noticia - elozano - 01/26/2017 - 04:34	â€¢ La ofensiva que sostiene la Armada Nacional en contra de las organizacior
32	Incautada cerca de una tonelada y media de clorhidr	https://www.armada.mil.co/e	Noticia - elozano - 09/28/2017 - 09:17	En las Ã°ltimas horas, unidades de la Armada Nacional que estaban desplegad
33	Incautadas cerca de 1,5 toneladas de cocaÃ-na en el	https://www.armada.mil.co/e	Noticia - elozano - 05/10/2016 - 07:42	MÃ¡s de 840 millones de dÃ³lares ha evitado la Armada Nacional que ingresen
34	Hallada cocaÃ-na que pretendÃ-a ser enviada de fo	https://www.armada.mil.co/e	Noticia - elozano - 11/16/2015 - 08:06	El hallazgo fue realizado por buzos de la Armada Nacional en Puerto BolÃ-var
35	Armada Nacional y EjÃ©rcito incautaron al ELN mÃ¡	https://www.armada.mil.co/e	Noticia - elozano - 05/26/2016 - 08:09	El alcaloide habÃ-a sido mezclado con combustible para evitar la acciÃ³n de la
36	Cae cargamento con mÃ¡s de media tonelada de coc	https://www.armada.mil.co/e	Noticia - elozano - 05/04/2016 - 15:50	En el 2.016 se han incautado mÃ¡s de 18 toneladas de cocaÃ-na, evitado que n
37	Interceptada motonave cargada con mÃ¡s de media	https://www.armada.mil.co/e	Noticia - elozano - 03/16/2015 - 10:37	Con las incautaciones de clorhidrato de cocaÃ-na realizadas en el PacÃ-fico
38	144 kilos de cocaÃ-na han sido incautados en prome	https://www.armada.mil.co/e	Noticia - elozano - 10/04/2015 - 15:37	El balance en la lucha contra el narcotrÃ¡fico en el PacÃ-fico colombiano deja
39	MÃS DE 2,6 TONELADAS DE COCAÃNA INCAUTADAS	https://www.armada.mil.co/e	Noticia - elozano - 11/14/2016 - 09:14	La primera incautaciÃ³n se dio en el Ã¡rea general de Sapzurro, en el municipi
40	Hallada mÃ¡s de media tonelada de cocaÃ-na	https://www.armada.mil.co/e	Noticia - elozano - 07/11/2016 - 13:42	El alijo, que estaba distribuido en 15 bultos hermÃ©ticamente sellados, seria
41	Incautada cocaÃ-na en embarcaciÃ³n abandonada e	https://www.armada.mil.co/e	Noticia - Armada_web - 10/04/2013 -	En operaciones de registro y control realizadas por unidades de la Armada Na
42	MÃ¡s de dos toneladas de cocaÃ-na incautadas en e	https://www.armada.mil.co/e	Noticia - Armada_web - 09/26/2012 -	2.345 kilos de cocaÃ-na fueron incautados en el marco del acuerdo de interdic
43	Destruido laboratorio con capacidad para procesar u	https://www.armada.mil.co/e	Noticia - elozano - 02/13/2018 - 18:28	En dos operaciones de control territorial y fluvial de la Armada Nacional, se lo
44	Destruidos cuatro laboratorios para el procesamient	https://www.armada.mil.co/e	Noticia - Armada_web - 07/23/2012 -	La primera operaciÃ³n tuvo lugar en un sector conocido como â€œLa Paulina
45	Incautada cocaÃ-na en un buque en Cartagena	https://www.armada.mil.co/e	Noticia - elozano - 04/20/2016 - 16:09	La Armada Nacional y la PolicÃ-a Nacional incautaron en las Ã°ltimas horas 26
46	Incautado 700 kilos de cocaÃ-na en el OcÃ©ano Pac	https://www.armada.mil.co/e	Noticia - elozano - 11/28/2017 - 08:52	En lo corrido del aÃ±o, la Fuerza Naval del PacÃ-fico ha incautado mÃ¡s de 95

arc news

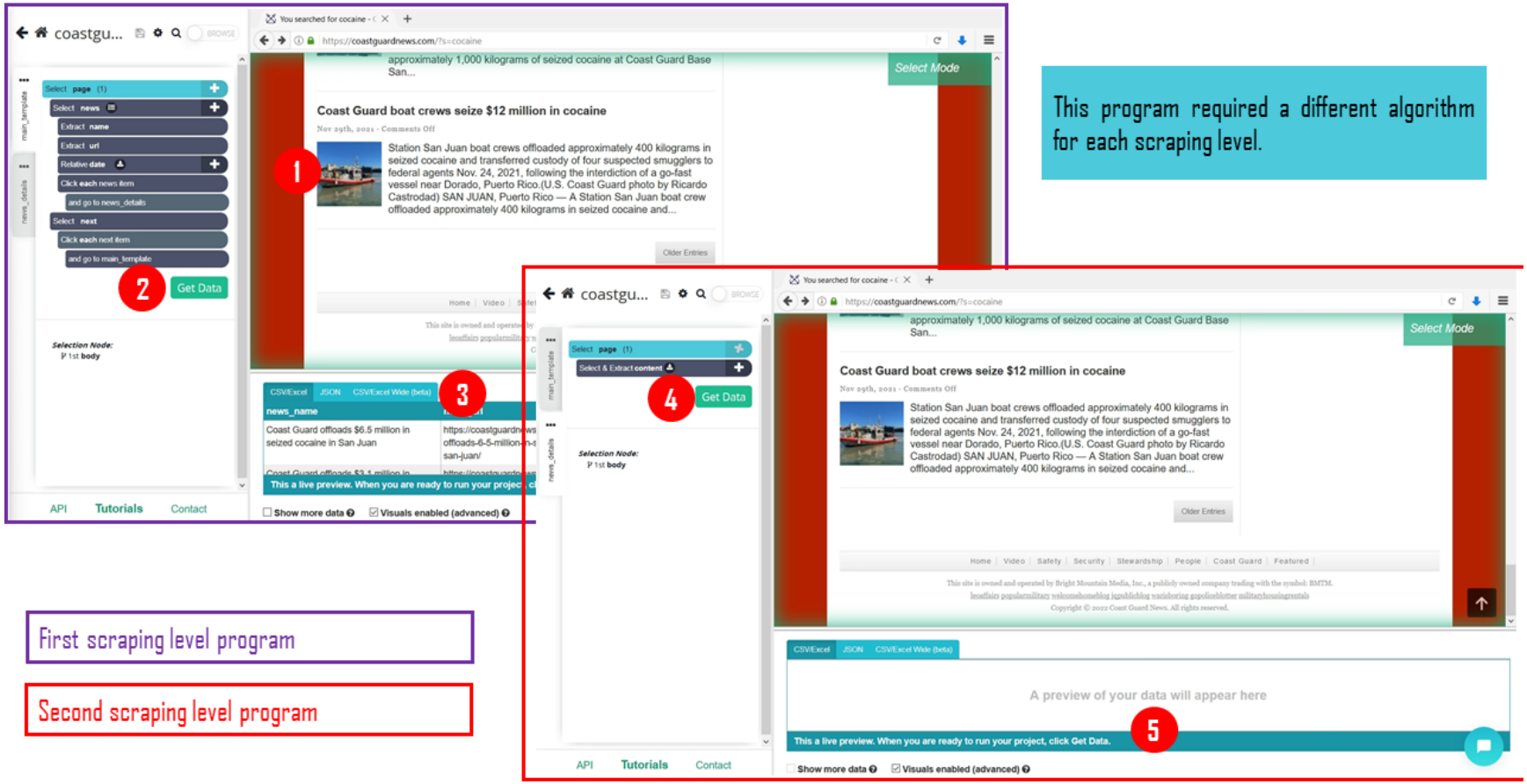
Ready | Accessibility: Unavailable | Count: 96 | Display Settings | 100%

This program extracted data into a unique table including the first and second scraping levels. However, some weird symbols are presented in the extracted data instead of Spanish accents.

First scraping level

Second scraping level

Parsehub -> USCG website

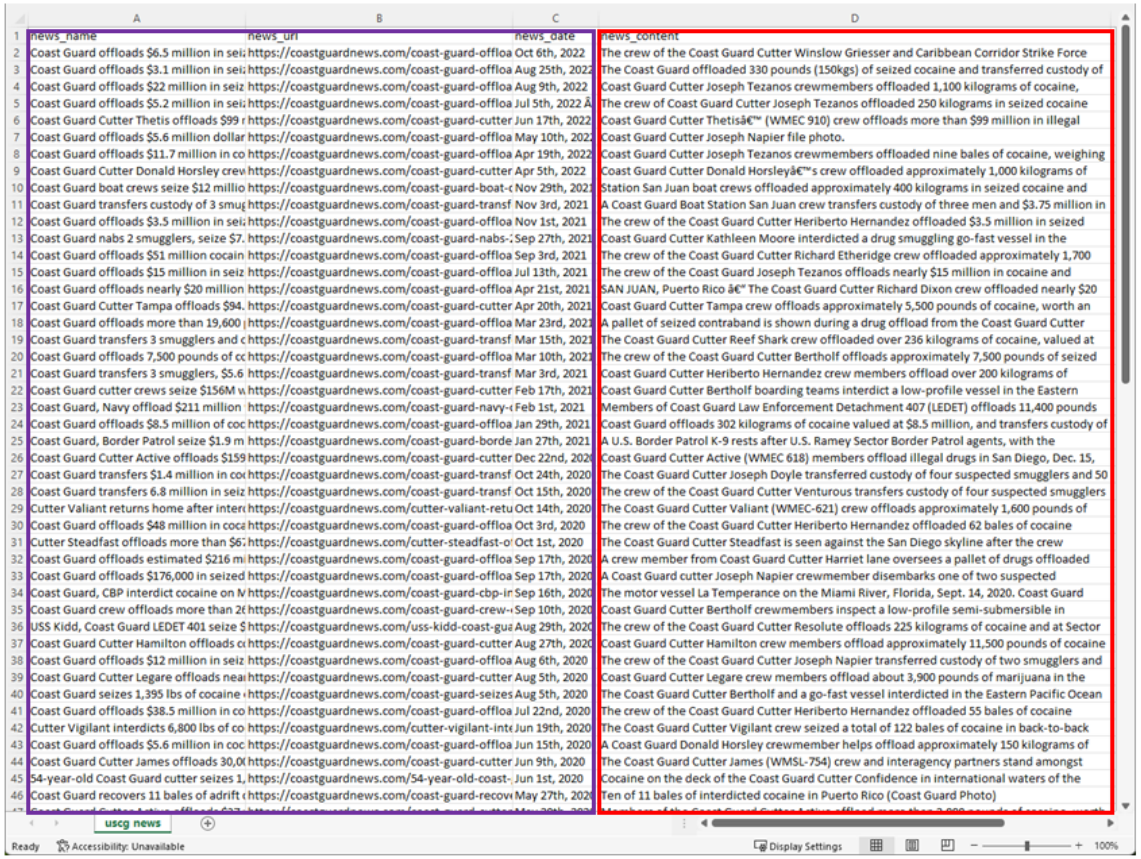


This program required a different algorithm for each scraping level.

First scraping level program

Second scraping level program

Parsehub -> USCG website results



news_name	news_url	news_date	news_content
Coast Guard offloads $6.5 million in sei	https://coastguardnews.com/coast-guard-offloa	Oct 6th, 2022	The crew of the Coast Guard Cutter Winslow Griesser and Caribbean Corridor Strike Force
Coast Guard offloads $3.1 million in sei	https://coastguardnews.com/coast-guard-offloa	Aug 25th, 2022	The Coast Guard offloaded 330 pounds (150kgs) of seized cocaine and transferred custody of
Coast Guard offloads $22 million in seiz	https://coastguardnews.com/coast-guard-offloa	Aug 9th, 2022	Coast Guard Cutter Joseph Tezanos crewmembers offloaded 1,100 kilograms of cocaine,
Coast Guard offloads $5.2 million in sei	https://coastguardnews.com/coast-guard-offloa	Jul 5th, 2022	The crew of Coast Guard Cutter Joseph Tezanos offloaded 250 kilograms in seized cocaine
Coast Guard Cutter Thetis offloads $99 r	https://coastguardnews.com/coast-guard-cutter	Jun 17th, 2022	Coast Guard Cutter Thetisâ€™ (WMEC 910) crew offloads more than $99 million in illegal
Coast Guard offloads $5.6 million dollar	https://coastguardnews.com/coast-guard-offloa	May 10th, 2022	Coast Guard Cutter Joseph Napier file photo.
Coast Guard offloads $11.7 million in co	https://coastguardnews.com/coast-guard-offloa	Apr 19th, 2022	Coast Guard Cutter Joseph Tezanos crewmembers offloaded nine bales of cocaine, weighing
Coast Guard Cutter Donald Horsley crev	https://coastguardnews.com/coast-guard-cutter	Apr 5th, 2022	Coast Guard Cutter Donald Horsleyâ€™s crew offloaded approximately 1,000 kilograms of
Coast Guard boat crews seize $12 millio	https://coastguardnews.com/coast-guard-boat-c	Nov 29th, 2021	Station San Juan boat crews offloaded approximately 400 kilograms in seized cocaine and
Coast Guard transfers custody of 3 smu	https://coastguardnews.com/coast-guard-transf	Nov 3rd, 2021	A Coast Guard Boat Station San Juan crew transfers custody of three men and $3.75 million in
Coast Guard offloads $3.5 million in sei	https://coastguardnews.com/coast-guard-offloa	Nov 1st, 2021	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded $3.5 million in seized
Coast Guard nabs 2 smugglers, seize $7.	https://coastguardnews.com/coast-guard-nabs-2	Sep 27th, 2021	Coast Guard Cutter Kathleen Moore interdicted a drug smuggling go-fast vessel in the
Coast Guard offloads $51 million cocain	https://coastguardnews.com/coast-guard-offloa	Sep 3rd, 2021	The crew of the Coast Guard Cutter Richard Etheridge crew offloaded approximately 1,700
Coast Guard offloads $15 million in seiz	https://coastguardnews.com/coast-guard-offloa	Jul 13th, 2021	The crew of the Coast Guard Joseph Tezanos offloads nearly $15 million in cocaine and
Coast Guard offloads nearly $20 million	https://coastguardnews.com/coast-guard-offloa	Apr 21st, 2021	SAN JUAN, Puerto Rico â€“ The Coast Guard Cutter Richard Dixon crew offloaded nearly $20
Coast Guard Cutter Tampa offloads $94.	https://coastguardnews.com/coast-guard-cutter	Apr 20th, 2021	Coast Guard Cutter Tampa crew offloads approximately 5,500 pounds of cocaine, worth an
Coast Guard offloads more than 19,600	https://coastguardnews.com/coast-guard-offloa	Mar 23rd, 2021	A pallet of seized contraband is shown during a drug offload from the Coast Guard Cutter
Coast Guard transfers 3 smugglers and c	https://coastguardnews.com/coast-guard-transf	Mar 15th, 2021	The Coast Guard Cutter Reef Shark crew offloaded over 236 kilograms of cocaine, valued at
Coast Guard offloads 7,500 pounds of cc	https://coastguardnews.com/coast-guard-offloa	Mar 10th, 2021	The crew of the Coast Guard Cutter Bertholf offloads approximately 7,500 pounds of seized
Coast Guard transfers 3 smugglers, $5.6	https://coastguardnews.com/coast-guard-transf	Mar 3rd, 2021	Coast Guard Cutter Heriberto Hernandez crew members offload over 200 kilograms of
Coast Guard cutter crews seize $156M w	https://coastguardnews.com/coast-guard-cutter	Feb 17th, 2021	Coast Guard Cutter Bertholf boarding teams interdict a low-profile vessel in the Eastern
Coast Guard, Navy offload $211 million	https://coastguardnews.com/coast-guard-navy-c	Feb 1st, 2021	Members of Coast Guard Law Enforcement Detachment 407 (LEDET) offloads 11,400 pounds
Coast Guard offloads $8.5 million of coc	https://coastguardnews.com/coast-guard-offloa	Jan 29th, 2021	Coast Guard offloads 302 kilograms of cocaine valued at $8.5 million, and transfers custody of
Coast Guard, Border Patrol seize $1.9 m	https://coastguardnews.com/coast-guard-borde	Jan 27th, 2021	A U.S. Border Patrol K-9 rests after U.S. Ramey Sector Border Patrol agents, with the
Coast Guard Cutter Active offloads $159	https://coastguardnews.com/coast-guard-cutter	Dec 22nd, 2020	Coast Guard Cutter Active (WMEC 618) members offload illegal drugs in San Diego, Dec. 15,
Coast Guard transfers $1.4 million in co	https://coastguardnews.com/coast-guard-transf	Oct 24th, 2020	The Coast Guard Cutter Joseph Doyle transferred custody of four suspected smugglers and 50
Coast Guard transfers 6.8 million in seiz	https://coastguardnews.com/coast-guard-transf	Oct 15th, 2020	The crew of the Coast Guard Cutter Venturous transfers custody of four suspected smugglers
Cutter Valiant returns home after inter	https://coastguardnews.com/cutter-valiant-retu	Oct 14th, 2020	The Coast Guard Cutter Valiant (WMEC-621) crew offloads approximately 1,600 pounds of
Coast Guard offloads $48 million in coc	https://coastguardnews.com/coast-guard-offloa	Oct 3rd, 2020	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded 62 bales of cocaine
Cutter Steadfast offloads more than $6	https://coastguardnews.com/cutter-steadfast-o	Oct 1st, 2020	The Coast Guard Cutter Steadfast is seen against the San Diego skyline after the crew
Coast Guard offloads estimated $216 m	https://coastguardnews.com/coast-guard-offloa	Sep 17th, 2020	A crew member from Coast Guard Cutter Harriet lane oversees a pallet of drugs offloaded
Coast Guard offloads $176,000 in seized	https://coastguardnews.com/coast-guard-offloa	Sep 17th, 2020	A Coast Guard cutter Joseph Napier crewmember disembarks one of two suspected
Coast Guard, CBP interdict cocaine on M	https://coastguardnews.com/coast-guard-cbp-in	Sep 16th, 2020	The motor vessel La Temperance on the Miami River, Florida, Sept. 14, 2020. Coast Guard
Coast Guard crew offloads more than 2(	https://coastguardnews.com/coast-guard-crew-o	Sep 10th, 2020	Coast Guard Cutter Bertholf crewmembers inspect a low-profile semi-submersible in
USS Kidd, Coast Guard LEDET 401 seize $	https://coastguardnews.com/uss-kidd-coast-gua	Aug 29th, 2020	The crew of the Coast Guard Cutter Resolute offloads 225 kilograms of cocaine and at Sector
Coast Guard Cutter Hamilton offloads c	https://coastguardnews.com/coast-guard-cutter	Aug 27th, 2020	Coast Guard Cutter Hamilton crew members offload approximately 11,500 pounds of cocaine
Coast Guard offloads $12 million in seiz	https://coastguardnews.com/coast-guard-offloa	Aug 6th, 2020	The crew of the Coast Guard Cutter Joseph Napier transferred custody of two smugglers and
Coast Guard Cutter Legare offloads neai	https://coastguardnews.com/coast-guard-cutter	Aug 5th, 2020	Coast Guard Cutter Legare crew members offload about 3,900 pounds of marijuana in the
Coast Guard seizes 1,395 lbs of cocaine	https://coastguardnews.com/coast-guard-seizes	Aug 5th, 2020	The Coast Guard Cutter Bertholf and a go-fast vessel interdicted in the Eastern Pacific Ocean
Coast Guard offloads $38.5 million in co	https://coastguardnews.com/coast-guard-offloa	Jul 22nd, 2020	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded 55 bales of cocaine
Cutter Vigilant interdicts 6,800 lbs of co	https://coastguardnews.com/cutter-vigilant-int	Jun 19th, 2020	The Coast Guard Cutter Vigilant crew seized a total of 122 bales of cocaine in back-to-back
Coast Guard offloads $5.6 million in coc	https://coastguardnews.com/coast-guard-offloa	Jun 15th, 2020	A Coast Guard Donald Horsley crewmember helps offload approximately 150 kilograms of
Coast Guard Cutter James offloads 30,0	https://coastguardnews.com/coast-guard-cutter	Jun 9th, 2020	The Coast Guard Cutter James (WMSL-754) crew and interagency partners stand amongst
54-year-old Coast Guard cutter seizes 1,	https://coastguardnews.com/54-year-old-coast-	Jun 1st, 2020	Cocaine on the deck of the Coast Guard Cutter Confidence in international waters of the
Coast Guard recovers 11 bales of adrift	https://coastguardnews.com/coast-guard-recov	May 27th, 2020	Ten of 11 bales of interdicted cocaine in Puerto Rico (Coast Guard Photo)

This program extracted data into a unique table including the first and second scraping levels.

First scraping level

Second scraping level

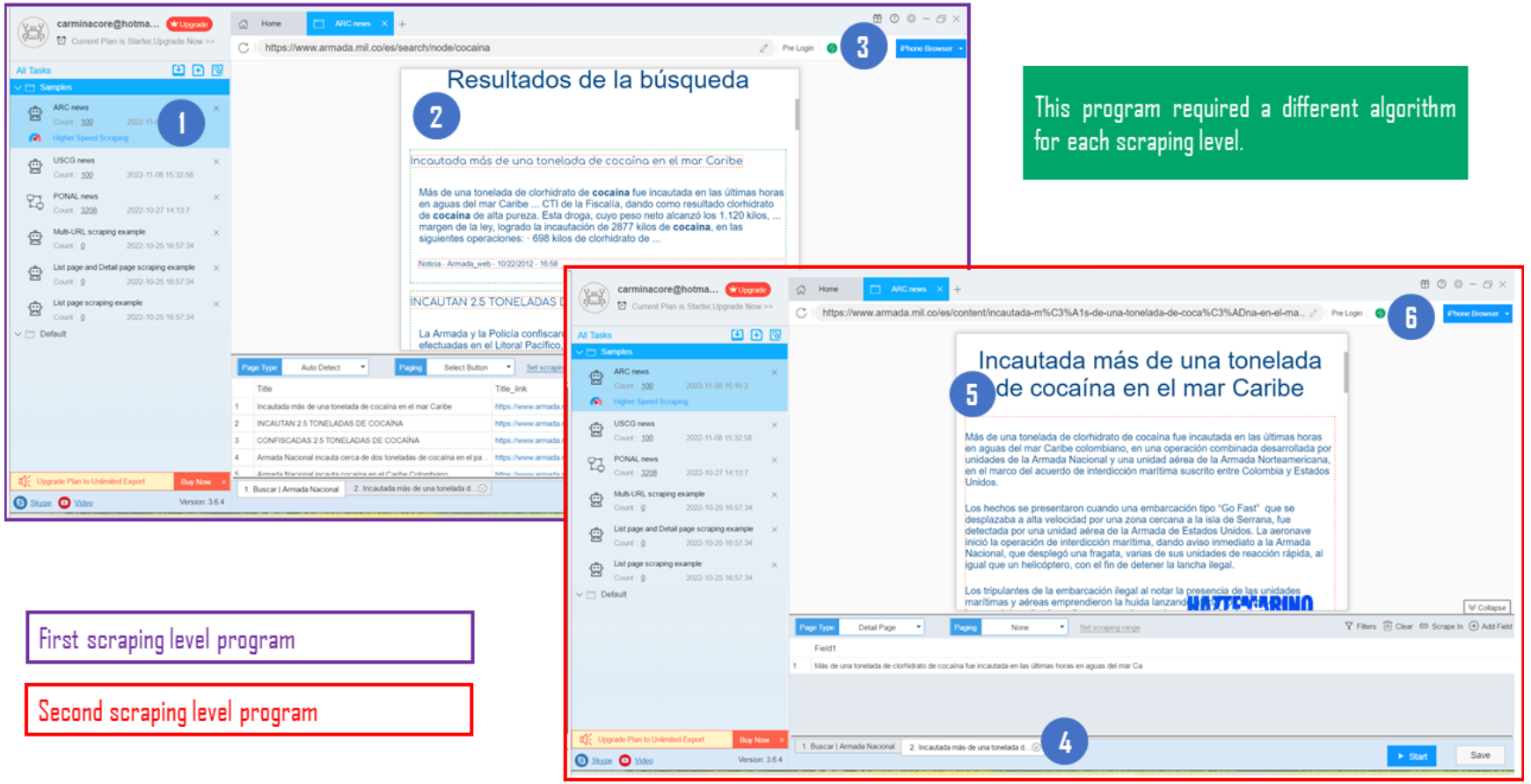
Parsehub -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
YES	YES	YES	Easy	61.77	40	21.60	22.58	YES	PARTIAL	5000	Medium	High	NO
YES			3	101.77		44.18		1	2	1	2	3	1

Remarks:

1. Parsehub uses the philosophy "Select-Correct-Command" to create the scraping and crawling tasks.
2. This program can solve CAPTCHAs based on images automatically.
3. The Spanish fidelity for this program is medium because it presents the accents with symbols when results are obtained in Excel.

Scrape Storm -> ARC website

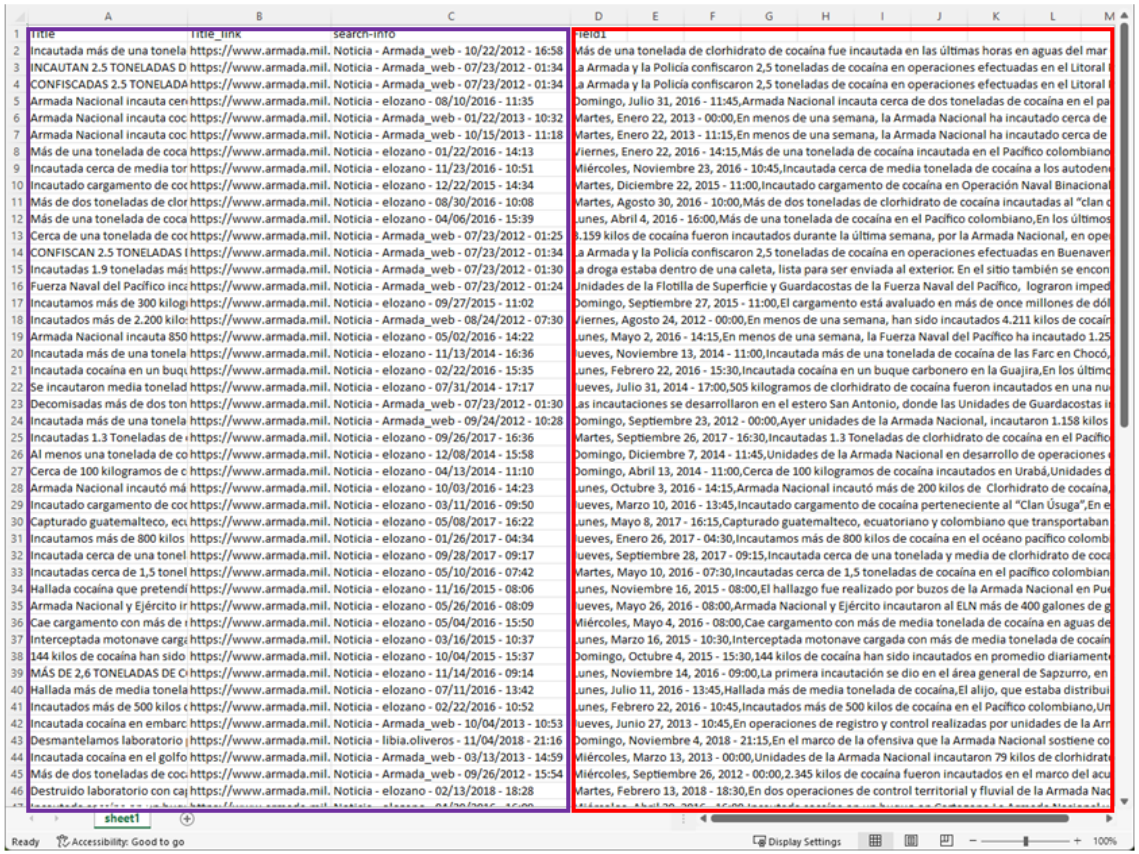


This program required a different algorithm for each scraping level.

First scraping level program

Second scraping level program

Scrape Storm -> ARC website results



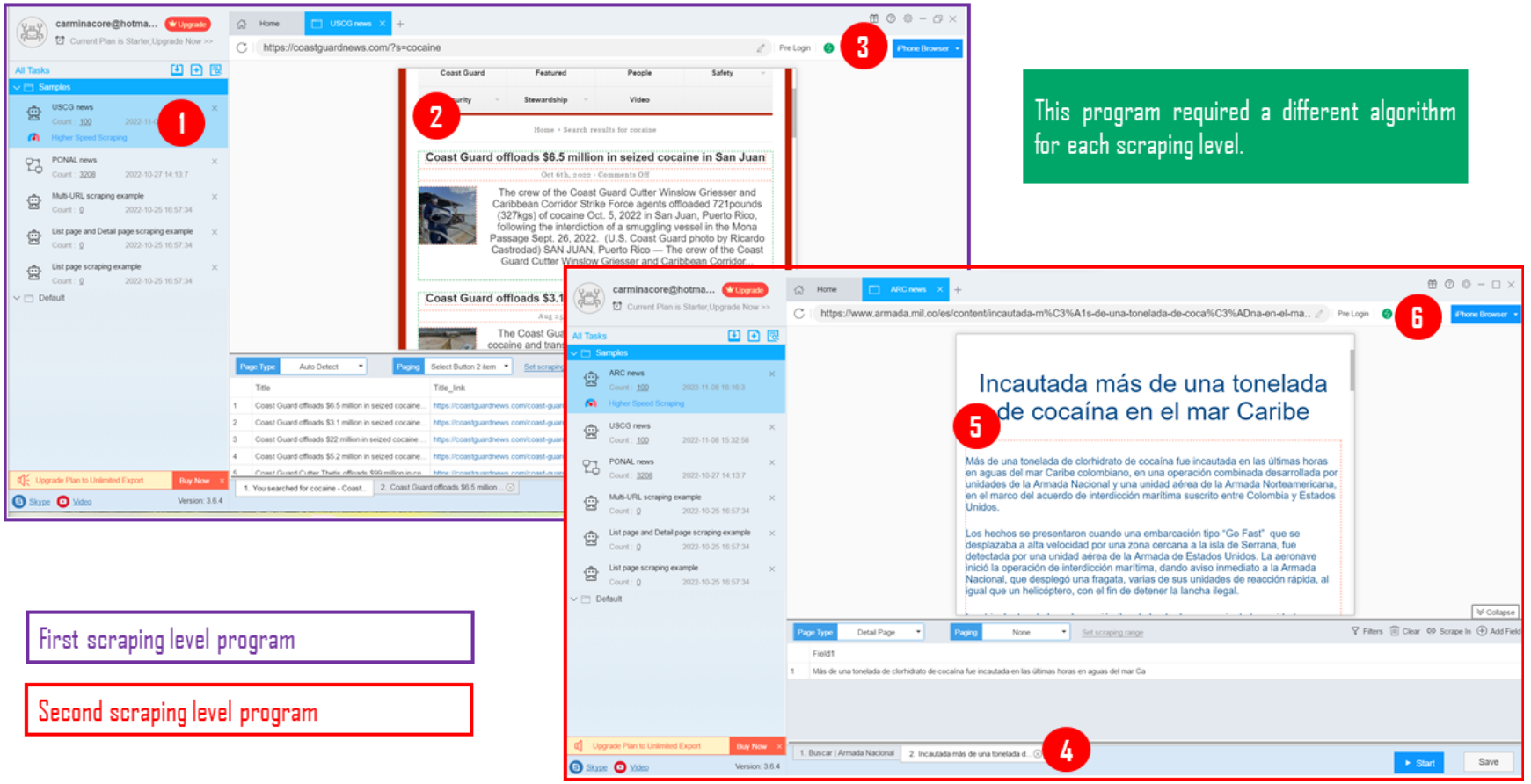
Title	Title_link	search-info	Field1
Incautada más de una tonela	https://www.armada.mil.	Noticia - Armada_web - 10/22/2012 - 16:58	Más de una tonelada de clorhidrato de cocaína fue incautada en las últimas horas en aguas del mar
INCAUTAN 2.5 TONELADAS D	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:34	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral
CONFISCADAS 2.5 TONELADA	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:34	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral
Armada Nacional incauta cer	https://www.armada.mil.	Noticia - elozano - 08/10/2016 - 11:35	Domingo, Julio 31, 2016 - 11:45,Armada Nacional incauta cerca de dos toneladas de cocaína en el pa
Armada Nacional incauta coc	https://www.armada.mil.	Noticia - Armada_web - 01/22/2013 - 10:32	Martes, Enero 22, 2013 - 00:00,En menos de una semana, la Armada Nacional ha incautado cerca de
Armada Nacional incauta coc	https://www.armada.mil.	Noticia - Armada_web - 10/15/2013 - 11:18	Martes, Enero 22, 2013 - 11:15,En menos de una semana, la Armada Nacional ha incautado cerca de
Más de una tonelada de coca	https://www.armada.mil.	Noticia - elozano - 01/22/2016 - 14:13	Viernes, Enero 22, 2016 - 14:15,Más de una tonelada de cocaína incautada en el Pacífico colombiano
Incautada cerca de media tor	https://www.armada.mil.	Noticia - elozano - 11/23/2016 - 10:51	Miércoles, Noviembre 23, 2016 - 10:45,Incautada cerca de media tonelada de cocaína a los autodef
Incautado cargamento de coc	https://www.armada.mil.	Noticia - elozano - 12/22/2015 - 14:34	Martes, Diciembre 22, 2015 - 11:00,Incautado cargamento de cocaína en Operación Naval Binacional
Más de dos toneladas de clor	https://www.armada.mil.	Noticia - elozano - 08/30/2016 - 10:08	Martes, Agosto 30, 2016 - 10:00,Más de dos toneladas de clorhidrato de cocaína incautadas al "clan d
Más de una tonelada de coca	https://www.armada.mil.	Noticia - elozano - 04/06/2016 - 15:39	Lunes, Abril 4, 2016 - 16:00,Más de una tonelada de cocaína en el Pacífico colombiano,En los últimos
Cerca de una tonelada de coc	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:25	3.159 kilos de cocaína fueron incautados durante la última semana, por la Armada Nacional, en ope
CONFISCAN 2.5 TONELADAS I	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:34	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en Buenaven
Incautadas 1.9 toneladas más	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:30	La droga estaba dentro de una caleta, lista para ser enviada al exterior. En el sitio también se encon
Fuerza Naval del Pacífico inca	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:24	Unidades de la Flotilla de Superficie y Guardacostas de la Fuerza Naval del Pacífico, lograron imped
Incautamos más de 300 kilogi	https://www.armada.mil.	Noticia - elozano - 09/27/2015 - 11:02	Domingo, Septiembre 27, 2015 - 11:00,El cargamento está avaluado en más de once millones de dól
Incautados más de 2.200 kilo	https://www.armada.mil.	Noticia - Armada_web - 08/24/2012 - 07:30	Viernes, Agosto 24, 2012 - 00:00,En menos de una semana, han sido incautados 4.211 kilos de cocaín
Armada Nacional incauta 850	https://www.armada.mil.	Noticia - elozano - 05/02/2016 - 14:22	Lunes, Mayo 2, 2016 - 14:15,En menos de una semana, la Fuerza Naval del Pacífico ha incautado 1.25
Incautada más de una tonela	https://www.armada.mil.	Noticia - elozano - 11/13/2014 - 16:36	Jueves, Noviembre 13, 2014 - 11:00,Incautada más de una tonelada de cocaína de las Farc en Chocó,
Incautada cocaína en un buq	https://www.armada.mil.	Noticia - elozano - 02/22/2016 - 15:35	Lunes, Febrero 22, 2016 - 15:30,Incautada cocaína en un buque carbonero en la Guajira,En los últimc
Se incautaron media tonelad	https://www.armada.mil.	Noticia - elozano - 07/31/2014 - 17:17	Jueves, Julio 31, 2014 - 17:00,505 kilogramos de clorhidrato de cocaína fueron incautados en una nu
Decomisadas más de dos ton	https://www.armada.mil.	Noticia - Armada_web - 07/23/2012 - 01:30	Las incautaciones se desarrollaron en el estero San Antonio, donde las Unidades de Guardacostas i
Incautada más de una tonela	https://www.armada.mil.	Noticia - Armada_web - 09/24/2012 - 10:28	Domingo, Septiembre 23, 2012 - 00:00,Ayer unidades de la Armada Nacional, incautaron 1.158 kilos
Incautadas 1.3 Toneladas de	https://www.armada.mil.	Noticia - elozano - 09/26/2017 - 16:36	Martes, Septiembre 26, 2017 - 16:30,Incautadas 1.3 Toneladas de clorhidrato de cocaína en el Pacífic
Al menos una tonelada de co	https://www.armada.mil.	Noticia - elozano - 12/08/2014 - 15:58	Domingo, Diciembre 7, 2014 - 11:45,Unidades de la Armada Nacional en desarrollo de operaciones
Cerca de 100 kilogramos de c	https://www.armada.mil.	Noticia - elozano - 04/13/2014 - 11:10	Domingo, Abril 13, 2014 - 11:00,Cerca de 100 kilogramos de cocaína incautados en Urabá,Unidades d
Armada Nacional incautó má	https://www.armada.mil.	Noticia - elozano - 10/03/2016 - 14:23	Lunes, Octubre 3, 2016 - 14:15,Armada Nacional incautó más de 200 kilos de Clorhidrato de cocaína,
Incautado cargamento de coc	https://www.armada.mil.	Noticia - elozano - 03/11/2016 - 09:50	Jueves, Marzo 10, 2016 - 13:45,Incautado cargamento de cocaína perteneciente al "Clan Úsuga",En e
Capturado guatemalteco, ec	https://www.armada.mil.	Noticia - elozano - 05/08/2017 - 16:22	Lunes, Mayo 8, 2017 - 16:15,Capturado guatemalteco, ecuatoriano y colombiano que transportaban
Incautamos más de 800 kilos	https://www.armada.mil.	Noticia - elozano - 01/26/2017 - 04:34	Jueves, Enero 26, 2017 - 04:30,Incautamos más de 800 kilos de cocaína en el océano pacífico colomb
Incautada cerca de una tonel	https://www.armada.mil.	Noticia - elozano - 09/28/2017 - 09:17	Jueves, Septiembre 28, 2017 - 09:15,Incautada cerca de una tonelada y media de clorhidrato de coca
Incautadas cerca de 1,5 tonel	https://www.armada.mil.	Noticia - elozano - 05/10/2016 - 07:42	Martes, Mayo 10, 2016 - 07:30,Incautadas cerca de 1,5 toneladas de cocaína en el pacífico colombian
Hallada cocaína que pretendí	https://www.armada.mil.	Noticia - elozano - 11/16/2015 - 08:06	Lunes, Noviembre 16, 2015 - 08:00,El hallazgo fue realizado por buzos de la Armada Nacional en Pue
Armada Nacional y Ejército ir	https://www.armada.mil.	Noticia - elozano - 05/26/2016 - 08:09	Jueves, Mayo 26, 2016 - 08:00,Armada Nacional y Ejército incautaron al ELN más de 400 galones de g
Cae cargamento con más de t	https://www.armada.mil.	Noticia - elozano - 05/04/2016 - 15:50	Miércoles, Mayo 4, 2016 - 08:00,Cae cargamento con más de media tonelada de cocaína en aguas de
Interceptada motonave carg	https://www.armada.mil.	Noticia - elozano - 03/16/2015 - 10:37	Lunes, Marzo 16, 2015 - 10:30,Interceptada motonave cargada con más de media tonelada de cocaín
144 kilos de cocaína han sido	https://www.armada.mil.	Noticia - elozano - 10/04/2015 - 15:37	Domingo, Octubre 4, 2015 - 15:30,144 kilos de cocaína han sido incautados en promedio diariament
MÁS DE 2,6 TONELADAS DE C	https://www.armada.mil.	Noticia - elozano - 11/14/2016 - 09:14	Lunes, Noviembre 14, 2016 - 09:00,La primera incautación se dio en el área general de Sapzurro, en
Hallada más de media tonela	https://www.armada.mil.	Noticia - elozano - 07/11/2016 - 13:42	Lunes, Julio 11, 2016 - 13:45,Hallada más de media tonelada de cocaína,El alijo, que estaba distribui
Incautados más de 500 kilos c	https://www.armada.mil.	Noticia - elozano - 02/22/2016 - 10:52	Lunes, Febrero 22, 2016 - 10:45,Incautados más de 500 kilos de cocaína en el Pacífico colombiano,Un
Incautada cocaína en embarc	https://www.armada.mil.	Noticia - Armada_web - 10/04/2013 - 10:53	Jueves, Junio 27, 2013 - 10:45,En operaciones de registro y control realizadas por unidades de la Arr
Desmantelamos laboratorio j	https://www.armada.mil.	Noticia - libia.oliveros - 11/04/2018 - 21:16	Domingo, Noviembre 4, 2018 - 21:15,En el marco de la ofensiva que la Armada Nacional sostiene co
Incautada cocaína en el golfo	https://www.armada.mil.	Noticia - Armada_web - 03/13/2013 - 14:59	Miércoles, Marzo 13, 2013 - 00:00,Unidades de la Armada Nacional incautaron 79 kilos de clorhidrat
Más de dos toneladas de coc	https://www.armada.mil.	Noticia - Armada_web - 09/26/2012 - 15:54	Miércoles, Septiembre 26, 2012 - 00:00,2.345 kilos de cocaína fueron incautados en el marco del acu
Destruido laboratorio con ca	https://www.armada.mil.	Noticia - elozano - 02/13/2018 - 18:28	Martes, Febrero 13, 2018 - 18:30,En dos operaciones de control territorial y fluvial de la Armada Nac

This program extracted data into a unique table including the first and second scraping levels.

First scraping level

Second scraping level

Scrape Storm -> USCG website

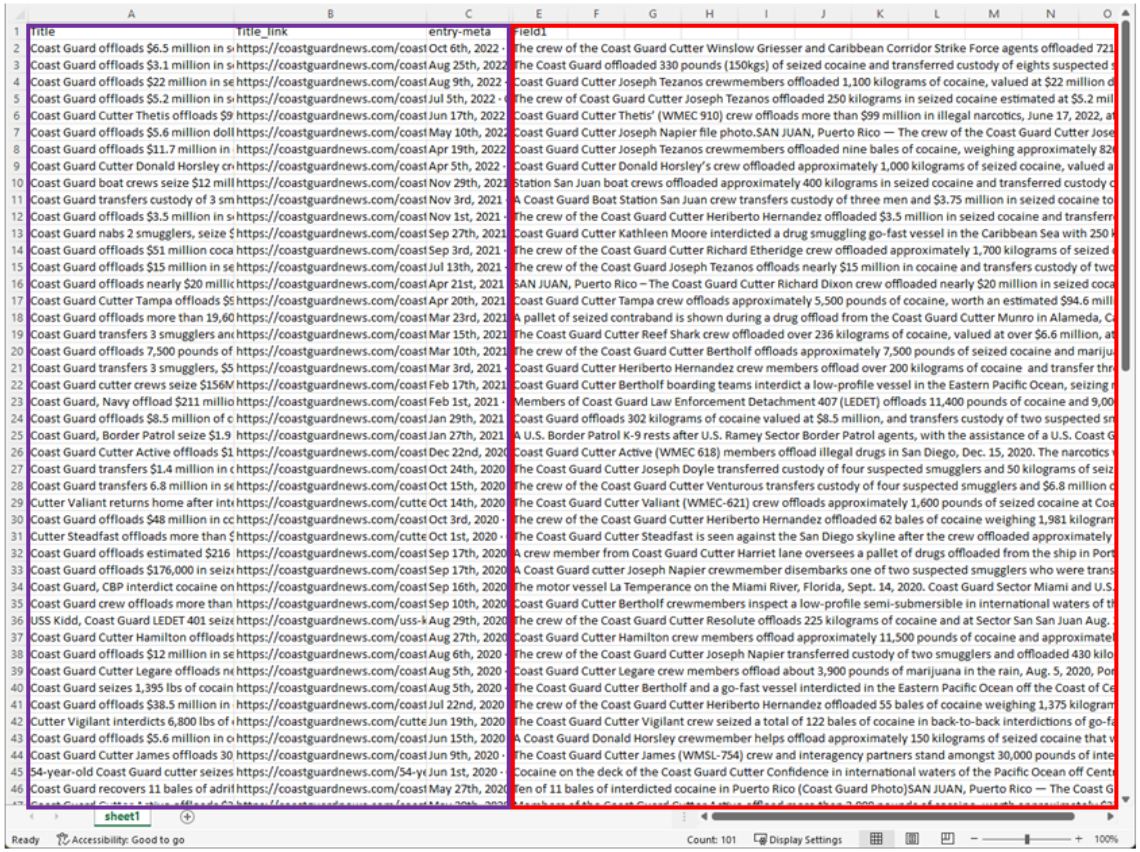


This program required a different algorithm for each scraping level.

First scraping level program

Second scraping level program

Scrape Storm -> USCG website results



Title	Title_link	entry-meta	Field1
Coast Guard offloads $6.5 million in s	https://coastguardnews.com/coast	Oct 6th, 2022 ·	The crew of the Coast Guard Cutter Winslow Griesser and Caribbean Corridor Strike Force agents offloaded 721
Coast Guard offloads $3.1 million in s	https://coastguardnews.com/coast	Aug 25th, 2022	The Coast Guard offloaded 330 pounds (150kgs) of seized cocaine and transferred custody of eights suspected s
Coast Guard offloads $22 million in se	https://coastguardnews.com/coast	Aug 9th, 2022 ·	Coast Guard Cutter Joseph Tezanos crewmembers offloaded 1,100 kilograms of cocaine, valued at $22 million d
Coast Guard offloads $5.2 million in s	https://coastguardnews.com/coast	Jul 5th, 2022 ·	The crew of Coast Guard Cutter Joseph Tezanos offloaded 250 kilograms in seized cocaine estimated at $5.2 mil
Coast Guard Cutter Thetis offloads $9	https://coastguardnews.com/coast	Jun 17th, 2022	Coast Guard Cutter Thetis' (WMEC 910) crew offloads more than $99 million in illegal narcotics, June 17, 2022, at
Coast Guard offloads $5.6 million doll	https://coastguardnews.com/coast	May 10th, 2022	Coast Guard Cutter Joseph Napier file photo.SAN JUAN, Puerto Rico — The crew of the Coast Guard Cutter Jose
Coast Guard offloads $11.7 million in	https://coastguardnews.com/coast	Apr 19th, 2022	Coast Guard Cutter Joseph Tezanos crewmembers offloaded nine bales of cocaine, weighing approximately 82
Coast Guard Cutter Donald Horsley cr	https://coastguardnews.com/coast	Apr 5th, 2022 ·	Coast Guard Cutter Donald Horsley's crew offloaded approximately 1,000 kilograms of seized cocaine, valued a
Coast Guard boat crews seize $12 mill	https://coastguardnews.com/coast	Nov 29th, 2021	Station San Juan boat crews offloaded approximately 400 kilograms in seized cocaine and transferred custody c
Coast Guard transfers custody of 3 sm	https://coastguardnews.com/coast	Nov 3rd, 2021	A Coast Guard Boat Station San Juan crew transfers custody of three men and $3.75 million in seized cocaine to
Coast Guard offloads $3.5 million in s	https://coastguardnews.com/coast	Nov 1st, 2021 ·	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded $3.5 million in seized cocaine and transfer
Coast Guard nabs 2 smugglers, seize $	https://coastguardnews.com/coast	Sep 27th, 2021	Coast Guard Cutter Kathleen Moore interdicted a drug smuggling go-fast vessel in the Caribbean Sea with 250 k
Coast Guard offloads $51 million coca	https://coastguardnews.com/coast	Sep 3rd, 2021 ·	The crew of the Coast Guard Cutter Richard Etheridge crew offloaded approximately 1,700 kilograms of seized c
Coast Guard offloads $15 million in se	https://coastguardnews.com/coast	Jul 13th, 2021 ·	The crew of the Coast Guard Joseph Tezanos offloads nearly $15 million in cocaine and transfers custody of two
Coast Guard offloads nearly $20 millic	https://coastguardnews.com/coast	Apr 21st, 2021	SAN JUAN, Puerto Rico – The Coast Guard Cutter Richard Dixon crew offloaded nearly $20 million in seized coca
Coast Guard Cutter Tampa offloads $5	https://coastguardnews.com/coast	Apr 20th, 2021	Coast Guard Cutter Tampa crew offloads approximately 5,500 pounds of cocaine, worth an estimated $94.6 mill
Coast Guard offloads more than 19,60	https://coastguardnews.com/coast	Mar 23rd, 2021	A pallet of seized contraband is shown during a drug offload from the Coast Guard Cutter Munro in Alameda, C
Coast Guard transfers 3 smugglers an	https://coastguardnews.com/coast	Mar 15th, 2021	The Coast Guard Cutter Reef Shark crew offloaded over 236 kilograms of cocaine, valued at over $6.6 million, at
Coast Guard offloads 7,500 pounds of	https://coastguardnews.com/coast	Mar 10th, 2021	The crew of the Coast Guard Cutter Bertholf offloads approximately 7,500 pounds of seized cocaine and mariju
Coast Guard transfers 3 smugglers, $5	https://coastguardnews.com/coast	Mar 3rd, 2021 ·	Coast Guard Cutter Heriberto Hernandez crew members offload over 200 kilograms of cocaine and transfer thr
Coast Guard cutter crews seize $156M	https://coastguardnews.com/coast	Feb 17th, 2021	Coast Guard Cutter Bertholf boarding teams interdict a low-profile vessel in the Eastern Pacific Ocean, seizing r
Coast Guard, Navy offload $211 millio	https://coastguardnews.com/coast	Feb 1st, 2021 ·	Members of Coast Guard Law Enforcement Detachment 407 (LEDET) offloads 11,400 pounds of cocaine and 9,00
Coast Guard offloads $8.5 million of c	https://coastguardnews.com/coast	Jan 29th, 2021	Coast Guard offloads 302 kilograms of cocaine valued at $8.5 million, and transfers custody of two suspected sm
Coast Guard, Border Patrol seize $1.9	https://coastguardnews.com/coast	Jan 27th, 2021	A U.S. Border Patrol K-9 rests after U.S. Ramey Sector Border Patrol agents, with the assistance of a U.S. Coast G
Coast Guard Cutter Active offloads $1	https://coastguardnews.com/coast	Dec 22nd, 2020	Coast Guard Cutter Active (WMEC 618) members offload illegal drugs in San Diego, Dec. 15, 2020. The narcotics w
Coast Guard transfers $1.4 million in c	https://coastguardnews.com/coast	Oct 24th, 2020	The Coast Guard Cutter Joseph Doyle transferred custody of four suspected smugglers and 50 kilograms of seiz
Coast Guard transfers 6.8 million in se	https://coastguardnews.com/coast	Oct 15th, 2020	The crew of the Coast Guard Cutter Venturous transfers custody of four suspected smugglers and $6.8 million c
Cutter Valiant returns home after int	https://coastguardnews.com/cutte	Oct 14th, 2020	The Coast Guard Cutter Valiant (WMEC-621) crew offloads approximately 1,600 pounds of seized cocaine at Coa
Coast Guard offloads $48 million in cc	https://coastguardnews.com/coast	Oct 3rd, 2020 ·	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded 62 bales of cocaine weighing 1,981 kilogram
Cutter Steadfast offloads more than $	https://coastguardnews.com/cutte	Oct 1st, 2020 ·	The Coast Guard Cutter Steadfast is seen against the San Diego skyline after the crew offloaded approximately
Coast Guard offloads estimated $216	https://coastguardnews.com/coast	Sep 17th, 2020	A crew member from Coast Guard Cutter Harriet lane oversees a pallet of drugs offloaded from the ship in Port
Coast Guard offloads $176,000 in seiz	https://coastguardnews.com/coast	Sep 17th, 2020	A Coast Guard cutter Joseph Napier crewmember disembarks one of two suspected smugglers who were trans
Coast Guard, CBP interdict cocaine on	https://coastguardnews.com/coast	Sep 16th, 2020	The motor vessel La Temperance on the Miami River, Florida, Sept. 14, 2020. Coast Guard Sector Miami and U.S.
Coast Guard crew offloads more than	https://coastguardnews.com/coast	Sep 10th, 2020	Coast Guard Cutter Bertholf crewmembers inspect a low-profile semi-submersible in international waters of th
USS Kidd, Coast Guard LEDET 401 seize	https://coastguardnews.com/uss-k	Aug 29th, 2020	The crew of the Coast Guard Cutter Resolute offloads 225 kilograms of cocaine and at Sector San Juan Aug.
Coast Guard Cutter Hamilton offloads	https://coastguardnews.com/coast	Aug 27th, 2020	Coast Guard Cutter Hamilton crew members offload approximately 11,500 pounds of cocaine and approximatel
Coast Guard offloads $12 million in se	https://coastguardnews.com/coast	Aug 6th, 2020 ·	The crew of the Coast Guard Cutter Joseph Napier transferred custody of two smugglers and offloaded 430 kilo
Coast Guard Cutter Legare offloads n	https://coastguardnews.com/coast	Aug 5th, 2020 ·	Coast Guard Cutter Legare crew members offload about 3,900 pounds of marijuana in the rain, Aug. 5, 2020, Por
Coast Guard seizes 1,395 lbs of cocain	https://coastguardnews.com/coast	Aug 5th, 2020 ·	The Coast Guard Cutter Bertholf and a go-fast vessel interdicted in the Eastern Pacific Ocean off the Coast of Ce
Coast Guard offloads $38.5 million in	https://coastguardnews.com/coast	Jul 22nd, 2020	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded 55 bales of cocaine weighing 1,375 kilogram
Cutter Vigilant interdicts 6,800 lbs of	https://coastguardnews.com/cutte	Jun 19th, 2020	The Coast Guard Cutter Vigilant crew seized a total of 122 bales of cocaine in back-to-back interdictions of go-fa
Coast Guard offloads $5.6 million in c	https://coastguardnews.com/coast	Jun 15th, 2020	A Coast Guard Donald Horsley crewmember helps offload approximately 150 kilograms of seized cocaine that w
Coast Guard Cutter James offloads 30	https://coastguardnews.com/coast	Jun 9th, 2020 ·	The Coast Guard Cutter James (WMSL-754) crew and interagency partners stand amongst 30,000 pounds of inte
54-year-old Coast Guard cutter seizes	https://coastguardnews.com/54-ye	Jun 1st, 2020 ·	Cocaine on the deck of the Coast Guard Cutter Confidence in international waters of the Pacific Ocean off Centr
Coast Guard recovers 11 bales of adrif	https://coastguardnews.com/coast	May 27th, 2020	Ten of 11 bales of interdicted cocaine in Puerto Rico (Coast Guard Photo)SAN JUAN, Puerto Rico — The Coast G

This program extracted data into a unique table including the first and second scraping levels.

First scraping level

Second scraping level

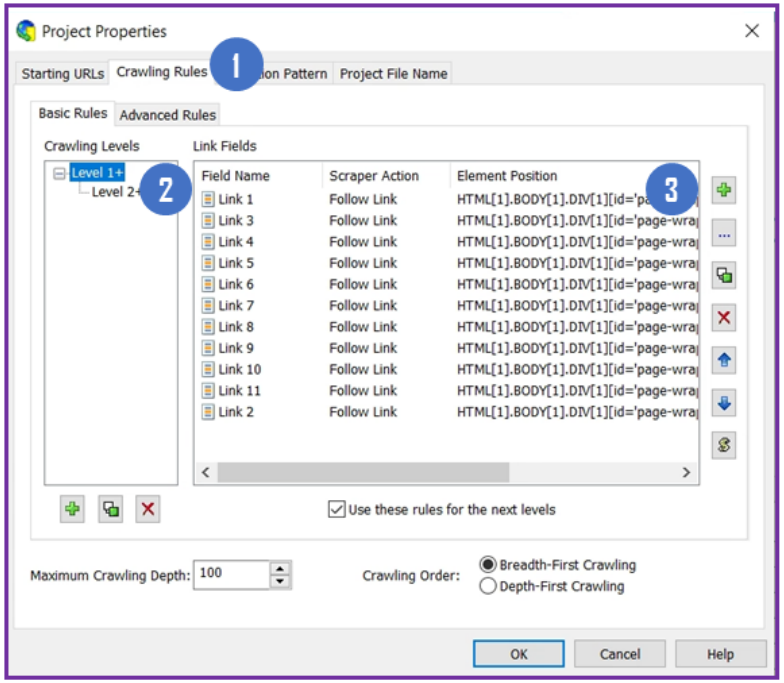
Scrape Storm -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
YES	YES	YES	Medium	107.30	250.00	15.57	9.85	YES	MANUAL	2000	High	High	YES
YES			2	357.30		25.42		1	1	1	3	3	0

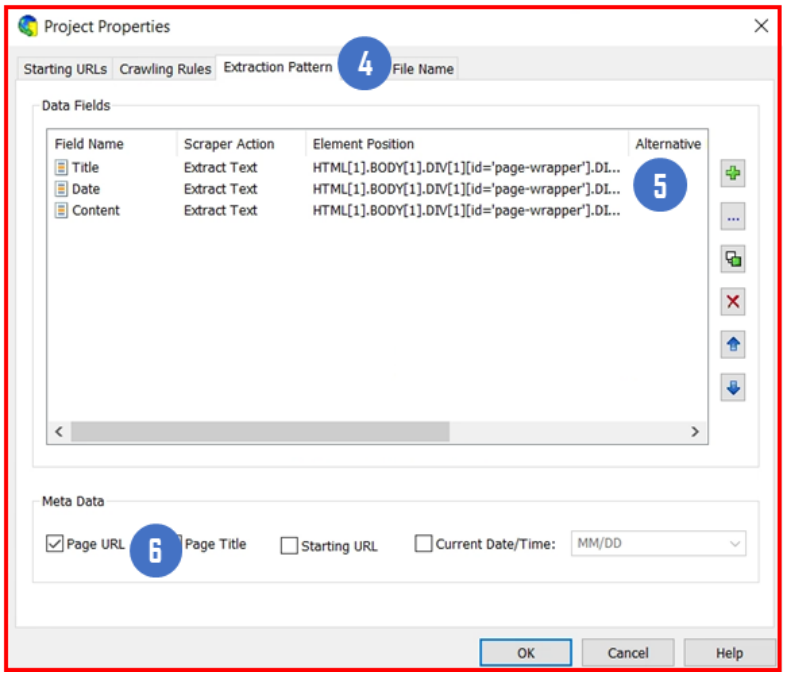
Remarks:

1. Scrape Storm uses two modes. In Smart mode the program creates the algorithm for each level of scraping automatically. In Flowchart mode the program allows the user to create an algorithm using graphical interface manually.
2. This program stops the scraping to allow the user to solve CAPTCHAs manually.

Web Content Extractor -> ARC website



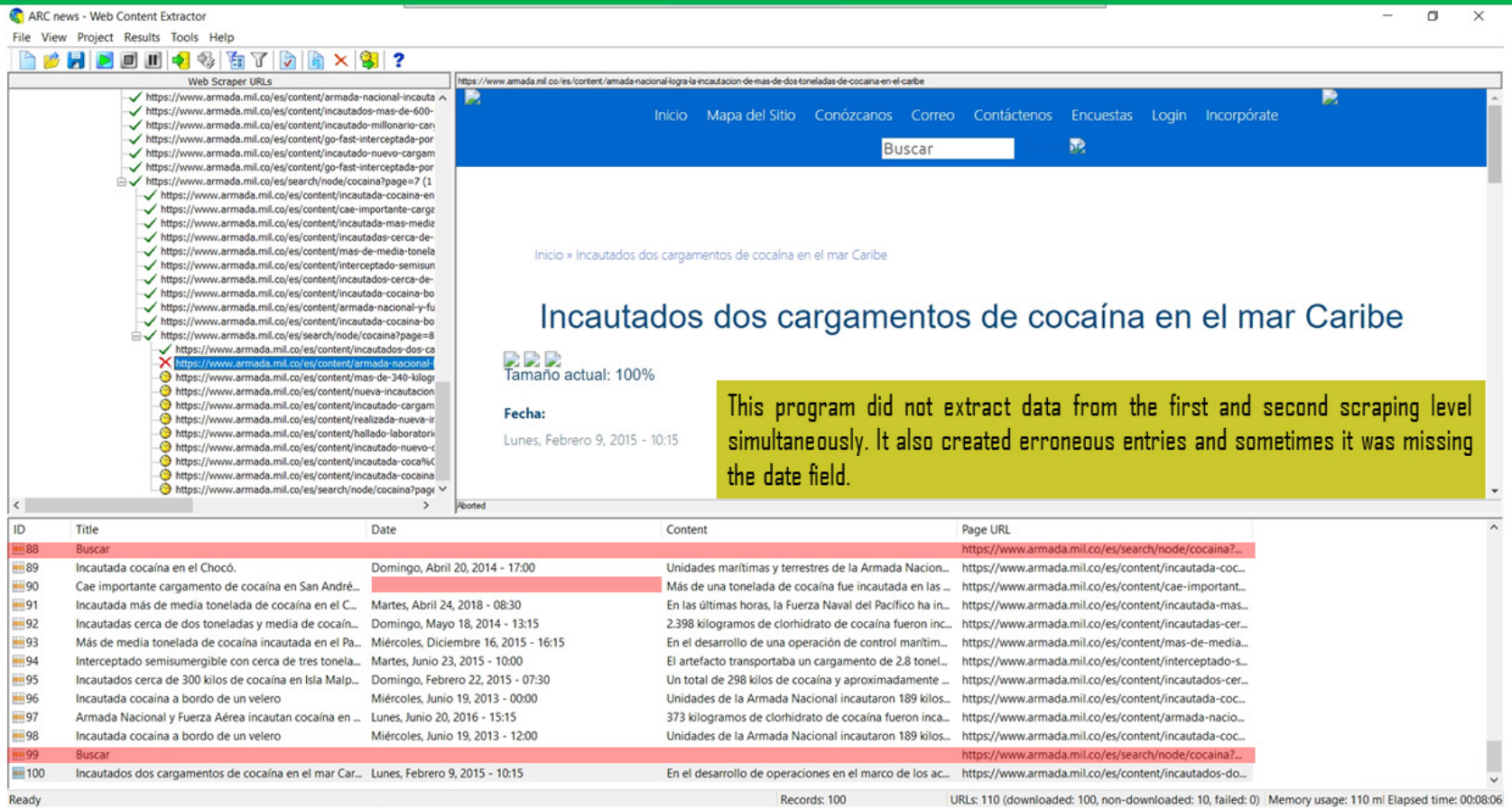
This program required a different algorithm for scraping and crawling.



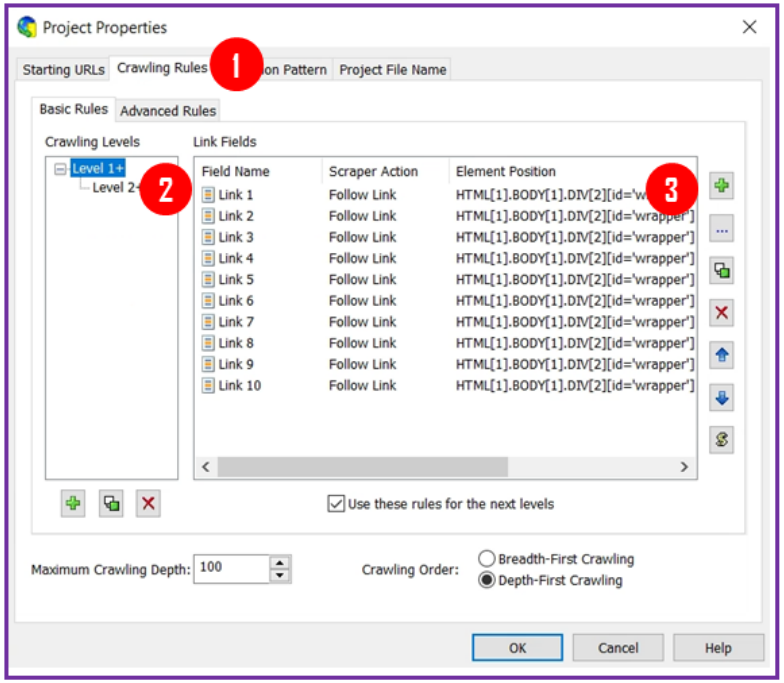
Crawling rules program

Scraping rules program

Web Content Extractor -> ARC website results



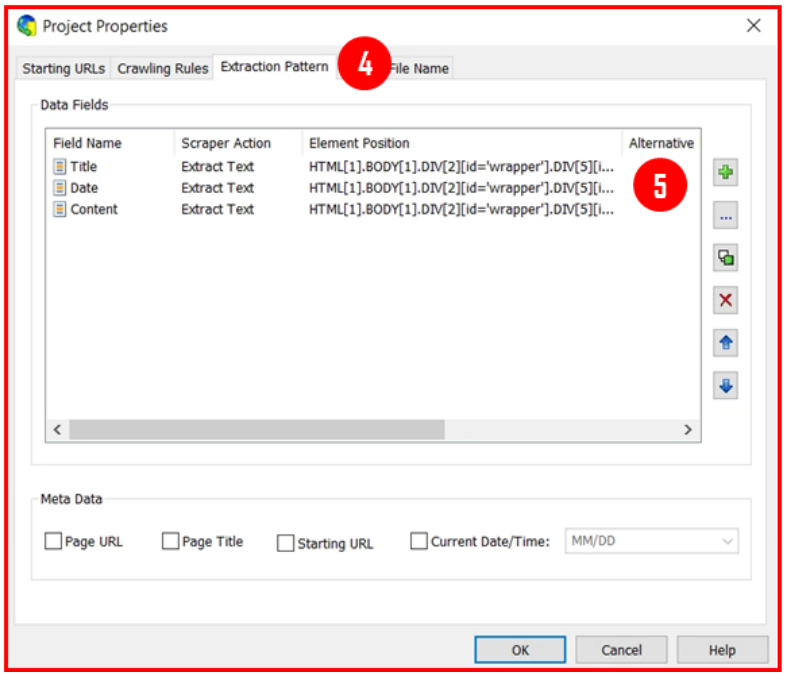
Web Content Extractor -> USCG website



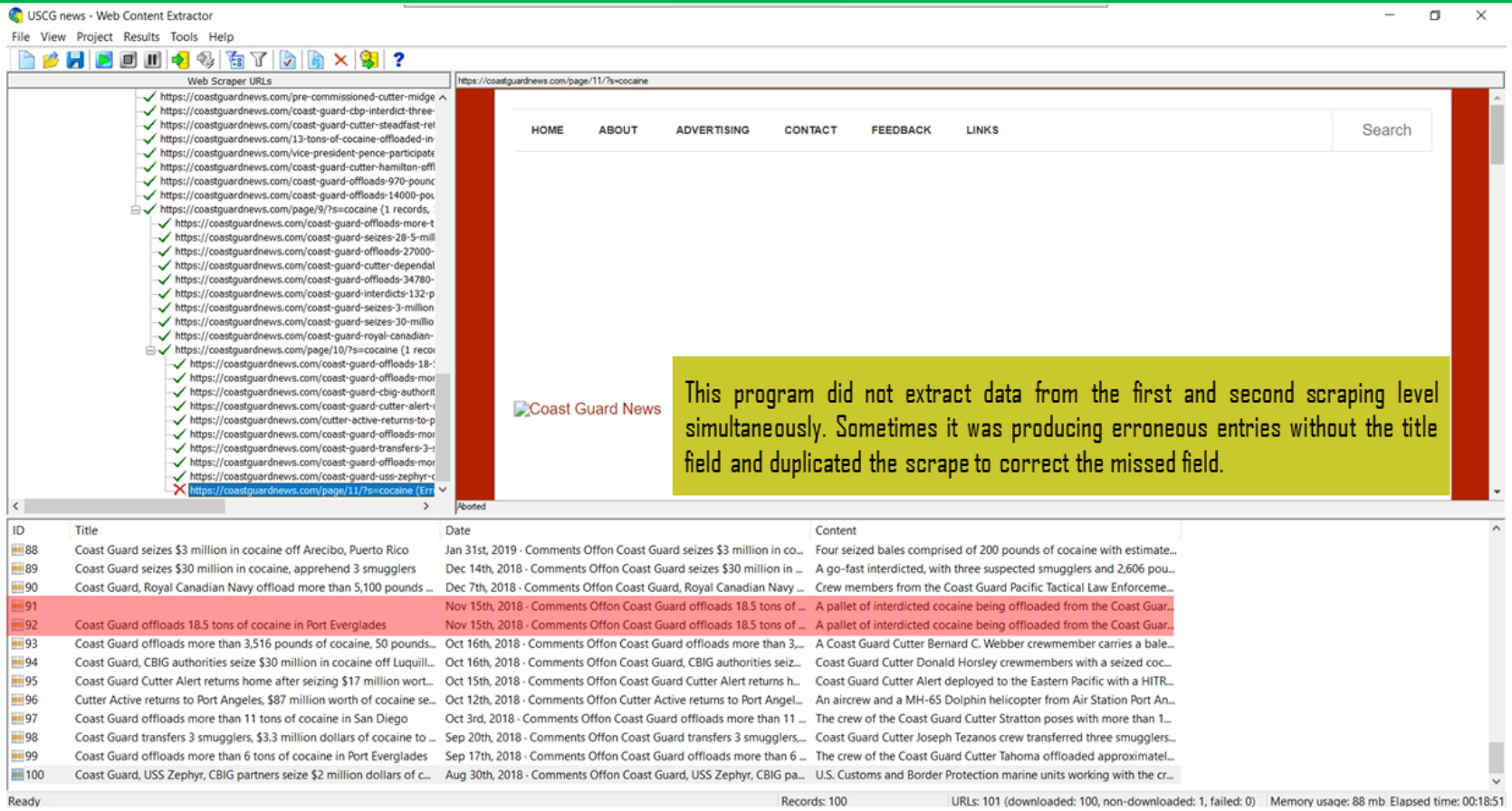
Crawling rules program

Scraping rules program

This program required a different algorithm for scraping and crawling.



Web Content Extractor -> USCG website results



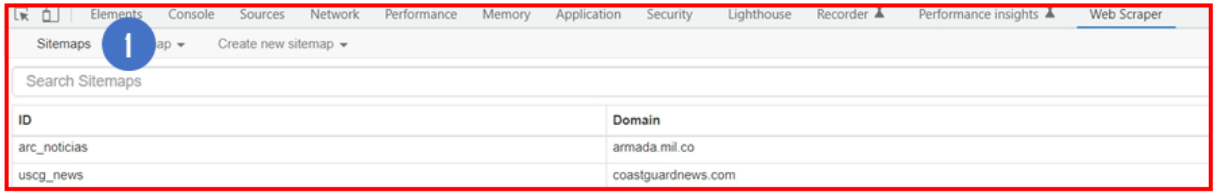
Web Content Extractor -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
PARTIAL	PARTIAL	PARTIAL	Easy	15.73	120	8.1	18.85	NO	MANUAL	2000	High	High	NO
NO			3	135.73		26.95		0	1	1	3	3	1

Remarks:

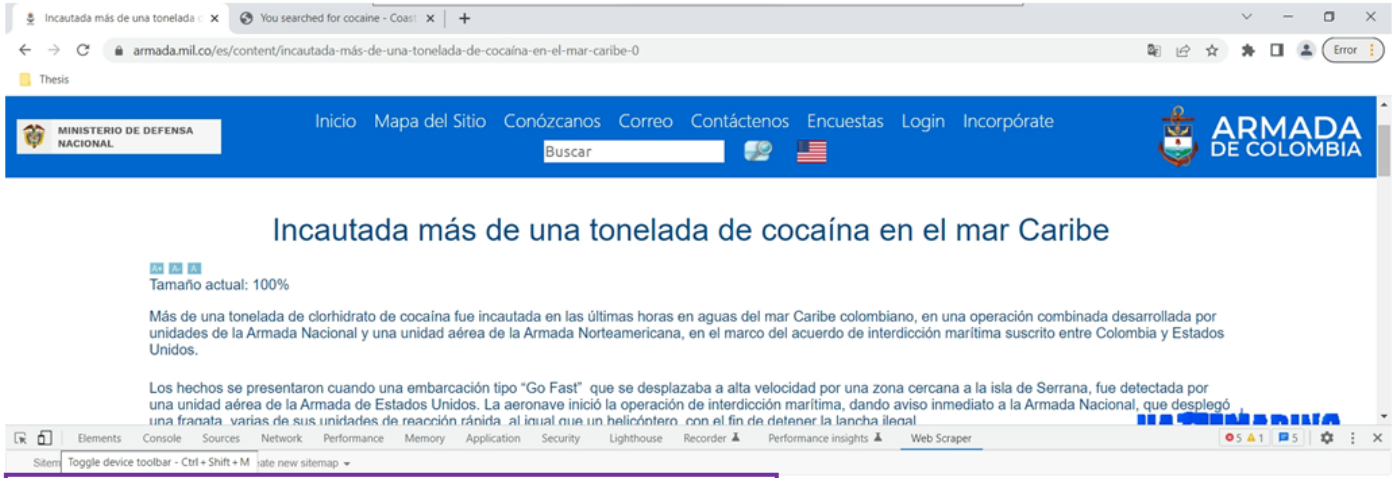
1. Web Content Extractor uses a methodology to create the crawling rules first and after, the extraction rules. This seems to affect the capability to extract data simultaneously from the first and second scraping levels.
2. This program creates erroneous entries in the scraping results for both websites.
3. The crawling rules seem to force the program to gather wrong fields in the websites.

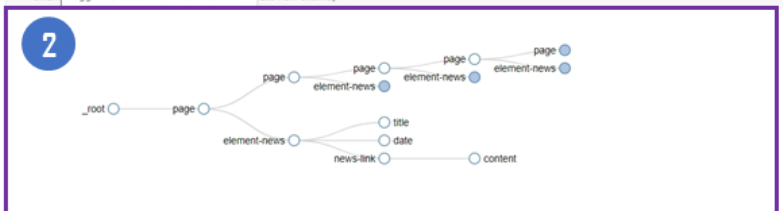
Web Scraper (Chrome extension) -> ARC website



Sitemaps creation

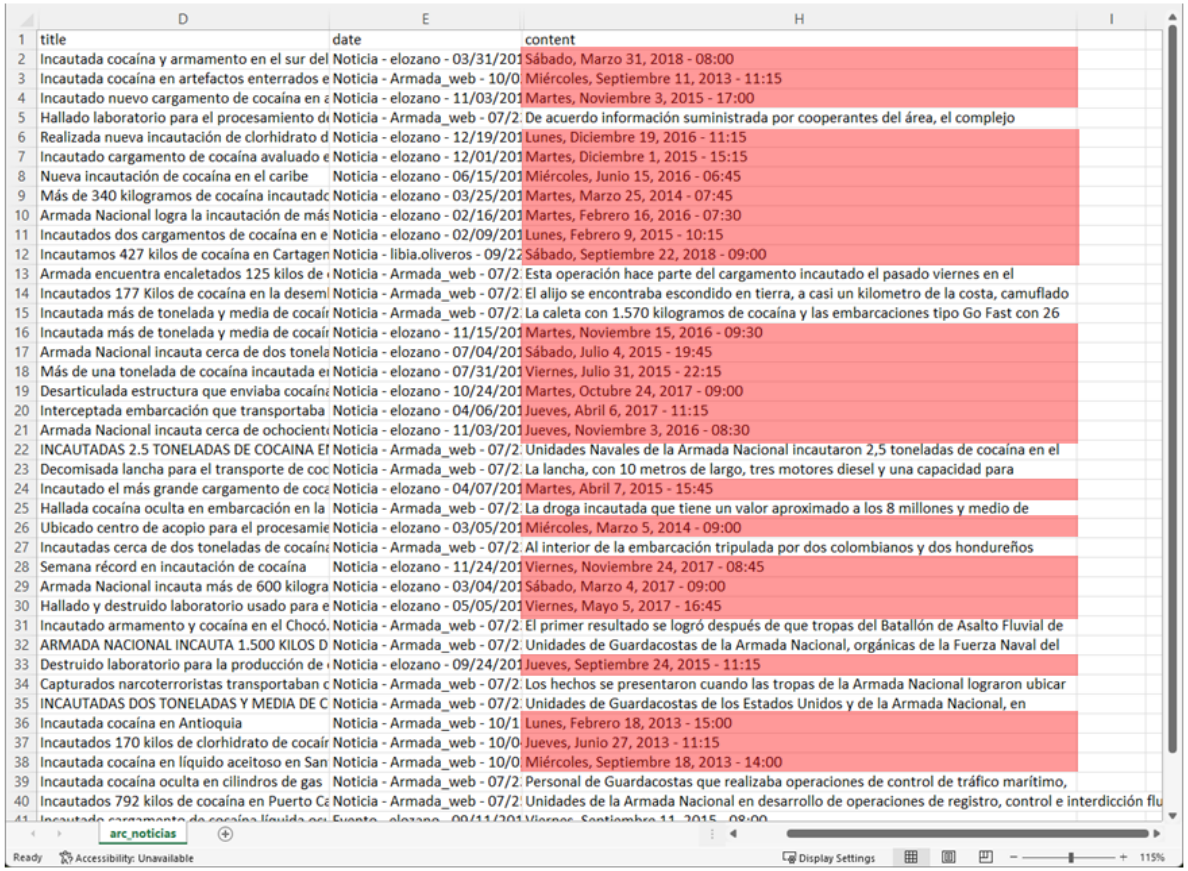
Selector graph





This program required create sitemaps for each website before scraping. Later, it was necessary to navigate through the web pages configuring selectors to program the scraping algorithm.

Web Scraper (Chrome extension) -> ARC website results



	D	E	H	I
1	title	date	content	
2	Incautada cocaína y armamento en el sur del	Noticia - elozano - 03/31/201	Sábado, Marzo 31, 2018 - 08:00	
3	Incautada cocaína en artefactos enterrados e	Noticia - Armada_web - 10/0	Miércoles, Septiembre 11, 2013 - 11:15	
4	Incautado nuevo cargamento de cocaína en a	Noticia - elozano - 11/03/201	Martes, Noviembre 3, 2015 - 17:00	
5	Hallado laboratorio para el procesamiento d	Noticia - Armada_web - 07/2	De acuerdo información suministrada por cooperantes del área, el complejo	
6	Realizada nueva incautación de clorhidrato d	Noticia - elozano - 12/19/201	Lunes, Diciembre 19, 2016 - 11:15	
7	Incautado cargamento de cocaína avaluado e	Noticia - elozano - 12/01/201	Martes, Diciembre 1, 2015 - 15:15	
8	Nueva incautación de cocaína en el caribe	Noticia - elozano - 06/15/201	Miércoles, Junio 15, 2016 - 06:45	
9	Más de 340 kilogramos de cocaína incautad	Noticia - elozano - 03/25/201	Martes, Marzo 25, 2014 - 07:45	
10	Armada Nacional logra la incautación de más	Noticia - elozano - 02/16/201	Martes, Febrero 16, 2016 - 07:30	
11	Incautados dos cargamentos de cocaína en e	Noticia - elozano - 02/09/201	Lunes, Febrero 9, 2015 - 10:15	
12	Incautamos 427 kilos de cocaína en Cartagen	Noticia - libia.oliveros - 09/22	Sábado, Septiembre 22, 2018 - 09:00	
13	Armada encuentra encaletados 125 kilos de	Noticia - Armada_web - 07/2	Esta operación hace parte del cargamento incautado el pasado viernes en el	
14	Incautados 177 Kilos de cocaína en la desem	Noticia - Armada_web - 07/2	El alijo se encontraba escondido en tierra, a casi un kilometro de la costa, camuflado	
15	Incautada más de tonelada y media de cocaí	Noticia - Armada_web - 07/2	La caleta con 1.570 kilogramos de cocaína y las embarcaciones tipo Go Fast con 26	
16	Incautada más de tonelada y media de cocaí	Noticia - elozano - 11/15/201	Martes, Noviembre 15, 2016 - 09:30	
17	Armada Nacional incauta cerca de dos tonela	Noticia - elozano - 07/04/201	Sábado, Julio 4, 2015 - 19:45	
18	Más de una tonelada de cocaína incautada e	Noticia - elozano - 07/31/201	Viernes, Julio 31, 2015 - 22:15	
19	Desarticulada estructura que enviaba cocaín	Noticia - elozano - 10/24/201	Martes, Octubre 24, 2017 - 09:00	
20	Interceptada embarcación que transportaba	Noticia - elozano - 04/06/201	Jueves, Abril 6, 2017 - 11:15	
21	Armada Nacional incauta cerca de ochocient	Noticia - elozano - 11/03/201	Jueves, Noviembre 3, 2016 - 08:30	
22	INCAUTADAS 2.5 TONELADAS DE COCAINA E	Noticia - Armada_web - 07/2	Unidades Navales de la Armada Nacional incautaron 2,5 toneladas de cocaína en el	
23	Decomisada lancha para el transporte de coc	Noticia - Armada_web - 07/2	La lancha, con 10 metros de largo, tres motores diesel y una capacidad para	
24	Incautado el más grande cargamento de coca	Noticia - elozano - 04/07/201	Martes, Abril 7, 2015 - 15:45	
25	Hallada cocaína oculta en embarcación en la	Noticia - Armada_web - 07/2	La droga incautada que tiene un valor aproximado a los 8 millones y medio de	
26	Ubicado centro de acopio para el procesamie	Noticia - elozano - 03/05/201	Miércoles, Marzo 5, 2014 - 09:00	
27	Incautadas cerca de dos toneladas de cocaín	Noticia - Armada_web - 07/2	Al interior de la embarcación tripulada por dos colombianos y dos hondureños	
28	Semana récord en incautación de cocaína	Noticia - elozano - 11/24/201	Viernes, Noviembre 24, 2017 - 08:45	
29	Armada Nacional incauta más de 600 kilogra	Noticia - elozano - 03/04/201	Sábado, Marzo 4, 2017 - 09:00	
30	Hallado y destruido laboratorio usado para e	Noticia - elozano - 05/05/201	Viernes, Mayo 5, 2017 - 16:45	
31	Incautado armamento y cocaína en el Chocó.	Noticia - Armada_web - 07/2	El primer resultado se logró después de que tropas del Batallón de Asalto Fluvial de	
32	ARMADA NACIONAL INCAUTA 1.500 KILOS D	Noticia - Armada_web - 07/2	Unidades de Guardacostas de la Armada Nacional, orgánicas de la Fuerza Naval del	
33	Destruido laboratorio para la producción de	Noticia - elozano - 09/24/201	Jueves, Septiembre 24, 2015 - 11:15	
34	Capturados narcoterroristas transportaban c	Noticia - Armada_web - 07/2	Los hechos se presentaron cuando las tropas de la Armada Nacional lograron ubicar	
35	INCAUTADAS DOS TONELADAS Y MEDIA DE C	Noticia - Armada_web - 07/2	Unidades de Guardacostas de los Estados Unidos y de la Armada Nacional, en	
36	Incautada cocaína en Antioquia	Noticia - Armada_web - 10/1	Lunes, Febrero 18, 2013 - 15:00	
37	Incautados 170 kilos de clorhidrato de cocaí	Noticia - Armada_web - 10/0	Jueves, Junio 27, 2013 - 11:15	
38	Incautada cocaína en líquido aceitoso en San	Noticia - Armada_web - 10/0	Miércoles, Septiembre 18, 2013 - 14:00	
39	Incautada cocaína oculta en cilindros de gas	Noticia - Armada_web - 07/2	Personal de Guardacostas que realizaba operaciones de control de tráfico marítimo,	
40	Incautados 792 kilos de cocaína en Puerto Ca	Noticia - Armada_web - 07/2	Unidades de la Armada Nacional en desarrollo de operaciones de registro, control e interdicción flu	

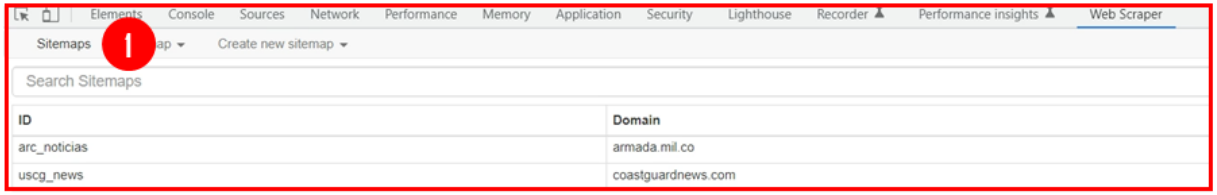
arc_noticias

Ready Accessibility: Unavailable Display Settings 115%

This program extracted data just in the first scraping level.

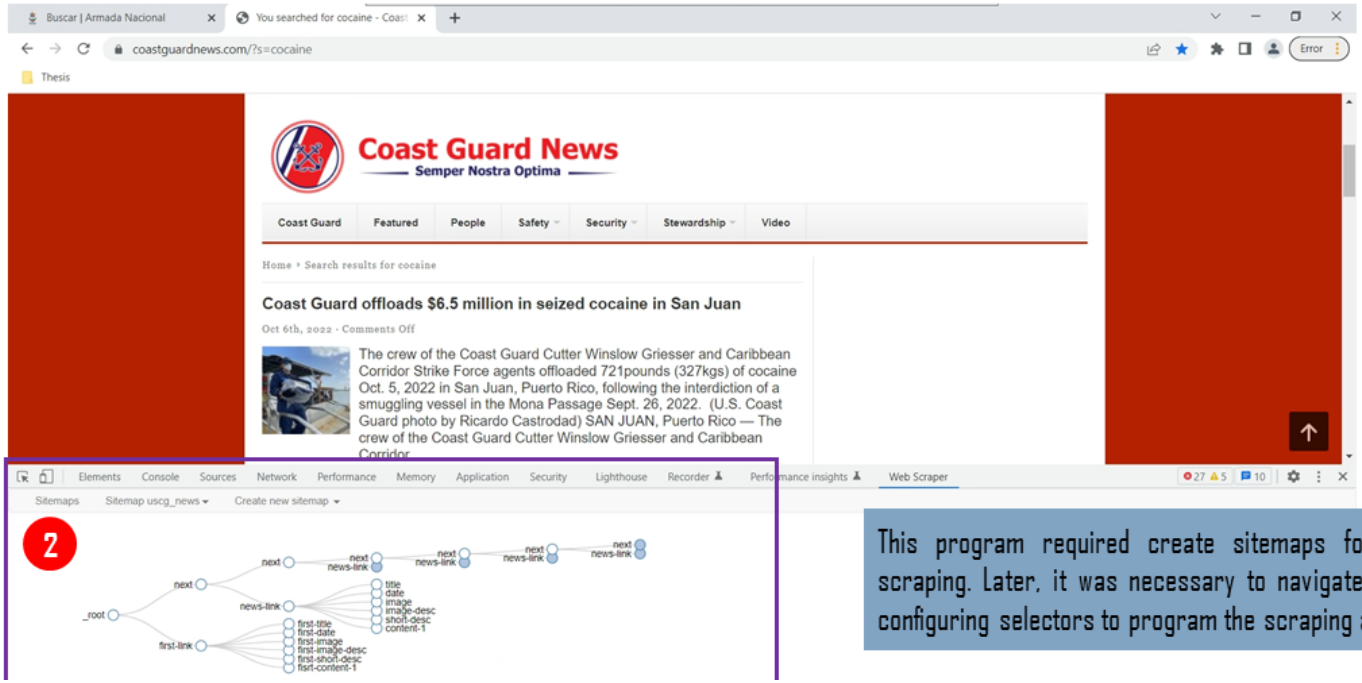
Most of the time is scraping dates in the content field.

Web Scraper (Chrome extension) -> USCG website



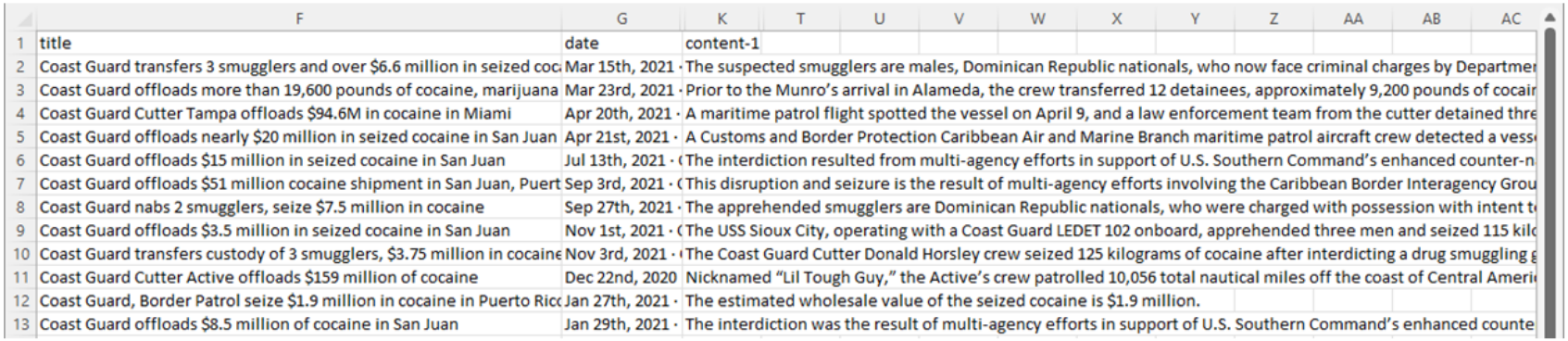
Sitemaps creation

Selector graph



This program required create sitemaps for each website before scraping. Later, it was necessary to navigate through the web pages configuring selectors to program the scraping algorithm.

Web Scraper (Chrome extension) -> USCG website results



	F	G	K	T	U	V	W	X	Y	Z	AA	AB	AC
1	title	date	content-1										
2	Coast Guard transfers 3 smugglers and over $6.6 million in seized coc	Mar 15th, 2021 ·	The suspected smugglers are males, Dominican Republic nationals, who now face criminal charges by Departmer										
3	Coast Guard offloads more than 19,600 pounds of cocaine, marijuana	Mar 23rd, 2021 ·	Prior to the Munro's arrival in Alameda, the crew transferred 12 detainees, approximately 9,200 pounds of cocair										
4	Coast Guard Cutter Tampa offloads $94.6M in cocaine in Miami	Apr 20th, 2021 ·	A maritime patrol flight spotted the vessel on April 9, and a law enforcement team from the cutter detained thre										
5	Coast Guard offloads nearly $20 million in seized cocaine in San Juan	Apr 21st, 2021 ·	A Customs and Border Protection Caribbean Air and Marine Branch maritime patrol aircraft crew detected a vess										
6	Coast Guard offloads $15 million in seized cocaine in San Juan	Jul 13th, 2021 · (	The interdiction resulted from multi-agency efforts in support of U.S. Southern Command's enhanced counter-n										
7	Coast Guard offloads $51 million cocaine shipment in San Juan, Puert	Sep 3rd, 2021 · (	This disruption and seizure is the result of multi-agency efforts involving the Caribbean Border Interagency Grou										
8	Coast Guard nabs 2 smugglers, seize $7.5 million in cocaine	Sep 27th, 2021 ·	The apprehended smugglers are Dominican Republic nationals, who were charged with possession with intent t										
9	Coast Guard offloads $3.5 million in seized cocaine in San Juan	Nov 1st, 2021 · (	The USS Sioux City, operating with a Coast Guard LEDET 102 onboard, apprehended three men and seized 115 kilo										
10	Coast Guard transfers custody of 3 smugglers, $3.75 million in cocaine	Nov 3rd, 2021 · (	The Coast Guard Cutter Donald Horsley crew seized 125 kilograms of cocaine after interdicting a drug smuggling g										
11	Coast Guard Cutter Active offloads $159 million of cocaine	Dec 22nd, 2020	Nicknamed "Lil Tough Guy," the Active's crew patrolled 10,056 total nautical miles off the coast of Central Ameri										
12	Coast Guard, Border Patrol seize $1.9 million in cocaine in Puerto Ric	Jan 27th, 2021 ·	The estimated wholesale value of the seized cocaine is $1.9 million.										
13	Coast Guard offloads $8.5 million of cocaine in San Juan	Jan 29th, 2021 ·	The interdiction was the result of multi-agency efforts in support of U.S. Southern Command's enhanced counte										

This program extracted data just in the first scraping level.

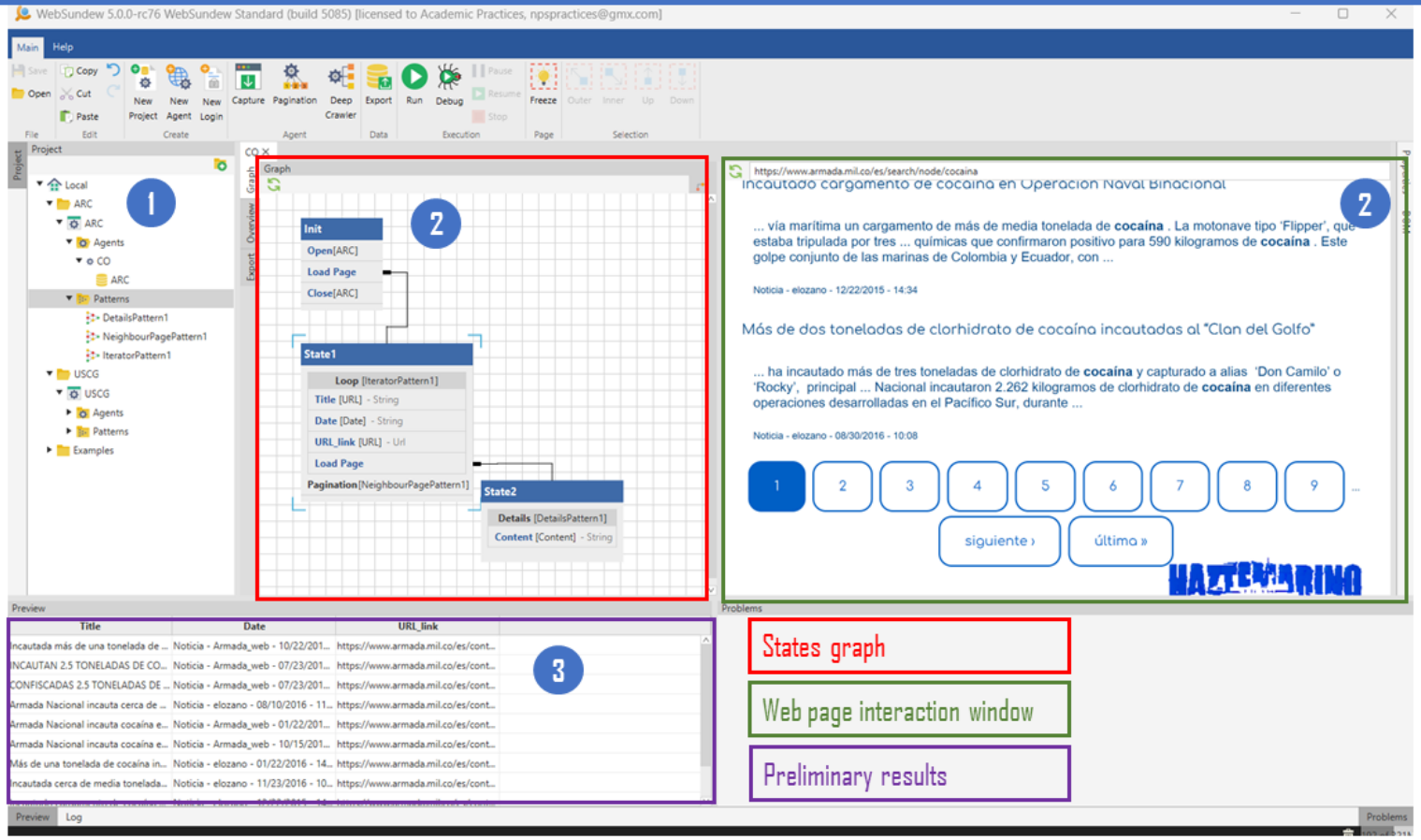
Web Scraper (Chrome extension) -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
PARTIAL	PARTIAL	PARTIAL	Medium	24.77	300.00	30.90	54.16	NO	NO	2000	High	High	NO
NO			2	324.77		85.06		0	0	1	3	3	1

Remarks:

1. This program extracted data erroneously if within the web page there is a link to extend the news content.
2. Because pagination is a beta feature, sometimes is crawling to random pages without following a logic order.
3. Content extension is not working properly in the second scraping level.

Web Sundew -> ARC website



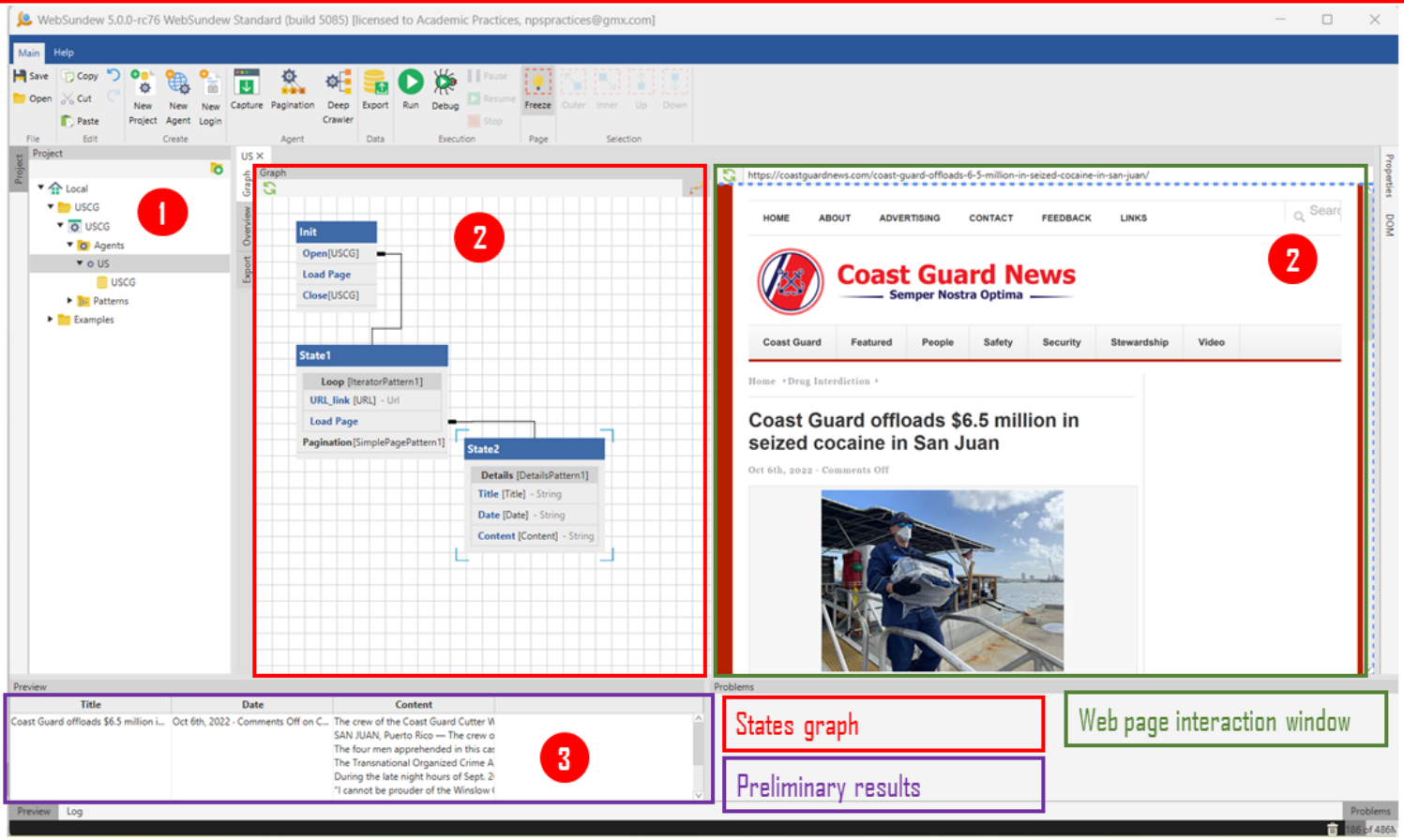
Web Sundew -> ARC website results

	A	B	D
1	Title	Date	Content
2	Incautada más de una tonelada de cocaína en	Noticia - Armada_web - 10/22/2012	Más de una tonelada de clorhidrato de cocaína fue incautada en las últimas horas en aguas del mar Ca
3	INCAUTAN 2.5 TONELADAS DE COCAÍNA	Noticia - Armada_web - 07/23/2012	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral Pac
4	CONFISCADAS 2.5 TONELADAS DE COCAÍNA	Noticia - Armada_web - 07/23/2012	La Armada y la Policía confiscaron 2,5 toneladas de cocaína en operaciones efectuadas en el Litoral Pac
5	Armada Nacional incauta cerca de dos tonelad	Noticia - elozano - 08/10/2016 - 11:3	Domingo, Julio 31, 2016 - 11:45
6	Armada Nacional incauta cocaína en el Caribe	Noticia - Armada_web - 01/22/2013	Martes, Enero 22, 2013 - 00:00
7	Armada Nacional incauta cocaína en el Caribe	Noticia - Armada_web - 10/15/2013	Martes, Enero 22, 2013 - 11:15
8	Más de una tonelada de cocaína incautada en	Noticia - elozano - 01/22/2016 - 14:1	Viernes, Enero 22, 2016 - 14:15
9	Incautada cerca de media tonelada de cocaína	Noticia - elozano - 11/23/2016 - 10:5	Miércoles, Noviembre 23, 2016 - 10:45
10	Incautado cargamento de cocaína en Operació	Noticia - elozano - 12/22/2015 - 14:3	Martes, Diciembre 22, 2015 - 11:00
11	Más de dos toneladas de clorhidrato de cocaín	Noticia - elozano - 08/30/2016 - 10:0	Martes, Agosto 30, 2016 - 10:00

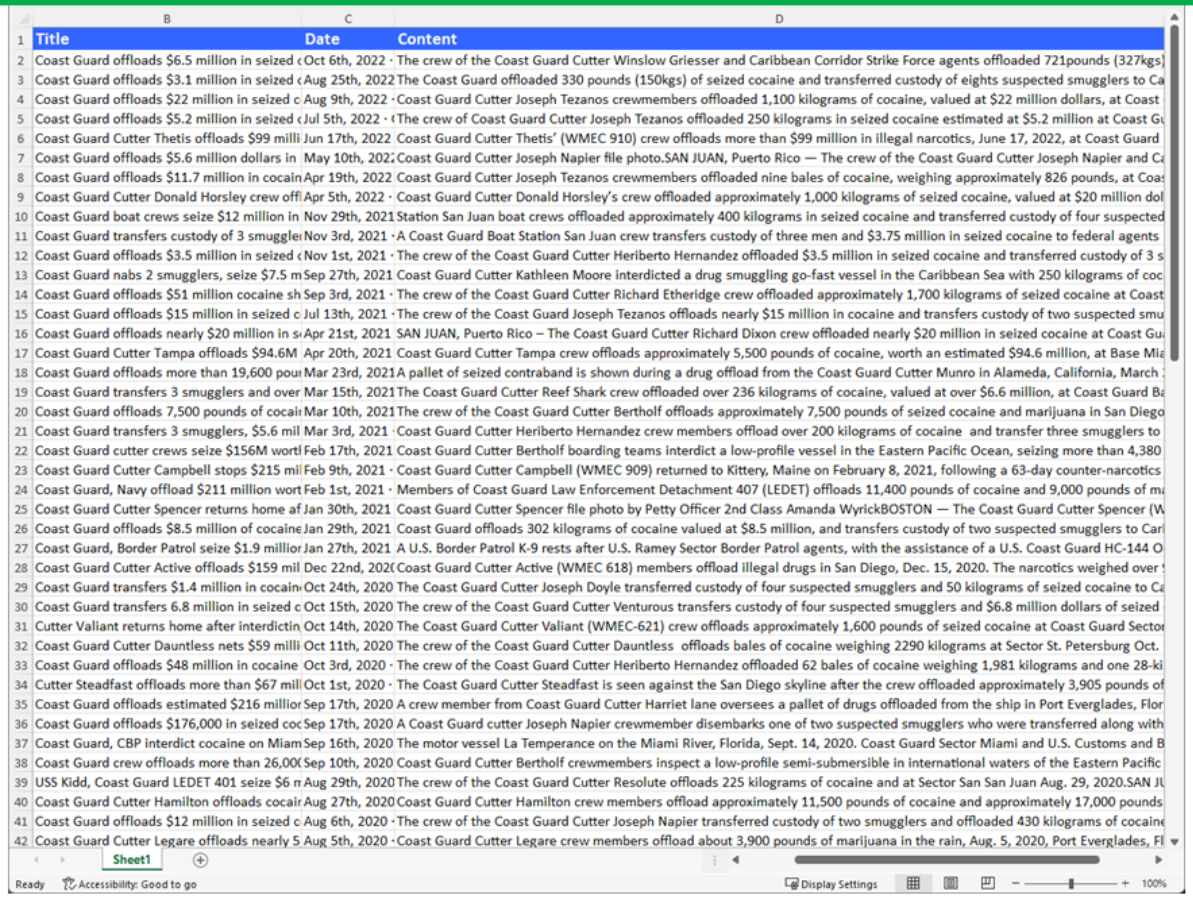
This program extracted data just in the first scraping level.

Most of the time is scraping dates in the content field.

Web Sundew -> USCG website



Web Sundew -> USCG website results



	Title	Date	Content
2	Coast Guard offloads $6.5 million in seized c	Oct 6th, 2022	The crew of the Coast Guard Cutter Winslow Griesser and Caribbean Corridor Strike Force agents offloaded 721pounds (327kgs)
3	Coast Guard offloads $3.1 million in seized c	Aug 25th, 2022	The Coast Guard offloaded 330 pounds (150kgs) of seized cocaine and transferred custody of eights suspected smugglers to Ca
4	Coast Guard offloads $22 million in seized c	Aug 9th, 2022	Coast Guard Cutter Joseph Tezanos crewmembers offloaded 1,100 kilograms of cocaine, valued at $22 million dollars, at Coast
5	Coast Guard offloads $5.2 million in seized c	Jul 5th, 2022	The crew of Coast Guard Cutter Joseph Tezanos offloaded 250 kilograms in seized cocaine estimated at $5.2 million at Coast Gu
6	Coast Guard Cutter Thetis offloads $99 milli	Jun 17th, 2022	Coast Guard Cutter Thetis' (WMEC 910) crew offloads more than $99 million in illegal narcotics, June 17, 2022, at Coast Guard
7	Coast Guard offloads $5.6 million dollars in	May 10th, 202	Coast Guard Cutter Joseph Napier file photo.SAN JUAN, Puerto Rico — The crew of the Coast Guard Cutter Joseph Napier and Ca
8	Coast Guard offloads $11.7 million in cocain	Apr 19th, 2022	Coast Guard Cutter Joseph Tezanos crewmembers offloaded nine bales of cocaine, weighing approximately 826 pounds, at Coas
9	Coast Guard Cutter Donald Horsley crew offl	Apr 5th, 2022	Coast Guard Cutter Donald Horsley's crew offloaded approximately 1,000 kilograms of seized cocaine, valued at $20 million dol
10	Coast Guard boat crews seize $12 million in	Nov 29th, 2021	Station San Juan boat crews offloaded approximately 400 kilograms in seized cocaine and transferred custody of four suspected
11	Coast Guard transfers custody of 3 smuggle	Nov 3rd, 2021	A Coast Guard Boat Station San Juan crew transfers custody of three men and $3.75 million in seized cocaine to federal agents
12	Coast Guard offloads $3.5 million in seized c	Nov 1st, 2021	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded $3.5 million in seized cocaine and transferred custody of 3 s
13	Coast Guard nabs 2 smugglers, seize $7.5 m	Sep 27th, 2021	Coast Guard Cutter Kathleen Moore interdicted a drug smuggling go-fast vessel in the Caribbean Sea with 250 kilograms of coc
14	Coast Guard offloads $51 million cocaine sh	Sep 3rd, 2021	The crew of the Coast Guard Cutter Richard Etheridge crew offloaded approximately 1,700 kilograms of seized cocaine at Coast
15	Coast Guard offloads $15 million in seized c	Jul 13th, 2021	The crew of the Coast Guard Joseph Tezanos offloads nearly $15 million in cocaine and transfers custody of two suspected smu
16	Coast Guard offloads nearly $20 million in s	Apr 21st, 2021	SAN JUAN, Puerto Rico – The Coast Guard Cutter Richard Dixon crew offloaded nearly $20 million in seized cocaine at Coast Gu
17	Coast Guard Cutter Tampa offloads $94.6M	Apr 20th, 2021	Coast Guard Cutter Tampa crew offloads approximately 5,500 pounds of cocaine, worth an estimated $94.6 million, at Base Mia
18	Coast Guard offloads more than 19,600 pou	Mar 23rd, 2021	A pallet of seized contraband is shown during a drug offload from the Coast Guard Cutter Munro in Alameda, California, March
19	Coast Guard transfers 3 smugglers and over	Mar 15th, 2021	The Coast Guard Cutter Reef Shark crew offloaded over 236 kilograms of cocaine, valued at over $6.6 million, at Coast Guard Ba
20	Coast Guard offloads 7,500 pounds of cocai	Mar 10th, 2021	The crew of the Coast Guard Cutter Bertholf offloads approximately 7,500 pounds of seized cocaine and marijuana in San Diego
21	Coast Guard transfers 3 smugglers, $5.6 mil	Mar 3rd, 2021	Coast Guard Cutter Heriberto Hernandez crew members offload over 200 kilograms of cocaine and transfer three smugglers to
22	Coast Guard cutter crews seize $156M worth	Feb 17th, 2021	Coast Guard Cutter Bertholf boarding teams interdict a low-profile vessel in the Eastern Pacific Ocean, seizing more than 4,380
23	Coast Guard Cutter Campbell stops $215 mil	Feb 9th, 2021	Coast Guard Cutter Campbell (WMEC 909) returned to Kittery, Maine on February 8, 2021, following a 63-day counter-narcotics
24	Coast Guard, Navy offload $211 million wort	Feb 1st, 2021	Members of Coast Guard Law Enforcement Detachment 407 (LEDET) offloads 11,400 pounds of cocaine and 9,000 pounds of ma
25	Coast Guard Cutter Spencer returns home af	Jan 30th, 2021	Coast Guard Cutter Spencer file photo by Petty Officer 2nd Class Amanda WyrickBOSTON — The Coast Guard Cutter Spencer (W
26	Coast Guard offloads $8.5 million of cocaine	Jan 29th, 2021	Coast Guard offloads 302 kilograms of cocaine valued at $8.5 million, and transfers custody of two suspected smugglers to Cari
27	Coast Guard, Border Patrol seize $1.9 millior	Jan 27th, 2021	A U.S. Border Patrol K-9 rests after U.S. Ramey Sector Border Patrol agents, with the assistance of a U.S. Coast Guard HC-144 O
28	Coast Guard Cutter Active offloads $159 mil	Dec 22nd, 202	Coast Guard Cutter Active (WMEC 618) members offload illegal drugs in San Diego, Dec. 15, 2020. The narcotics weighed over
29	Coast Guard transfers $1.4 million in cocain	Oct 24th, 2020	The Coast Guard Cutter Joseph Doyle transferred custody of four suspected smugglers and 50 kilograms of seized cocaine to Ca
30	Coast Guard transfers 6.8 million in seized c	Oct 15th, 2020	The crew of the Coast Guard Cutter Venturous transfers custody of four suspected smugglers and $6.8 million dollars of seized
31	Cutter Valiant returns home after interdictin	Oct 14th, 2020	The Coast Guard Cutter Valiant (WMEC-621) crew offloads approximately 1,600 pounds of seized cocaine at Coast Guard Sector
32	Coast Guard Cutter Dauntless nets $59 milli	Oct 11th, 2020	The crew of the Coast Guard Cutter Dauntless offloads bales of cocaine weighing 2290 kilograms at Sector St. Petersburg Oct.
33	Coast Guard offloads $48 million in cocaine	Oct 3rd, 2020	The crew of the Coast Guard Cutter Heriberto Hernandez offloaded 62 bales of cocaine weighing 1,981 kilograms and one 28-ki
34	Cutter Steadfast offloads more than $67 mil	Oct 1st, 2020	The Coast Guard Cutter Steadfast is seen against the San Diego skyline after the crew offloaded approximately 3,905 pounds of
35	Coast Guard offloads estimated $216 millior	Sep 17th, 2020	A crew member from Coast Guard Cutter Harriet lane oversees a pallet of drugs offloaded from the ship in Port Everglades, Flor
36	Coast Guard offloads $176,000 in seized coc	Sep 17th, 2020	A Coast Guard cutter Joseph Napier crewmember disembarks one of two suspected smugglers who were transferred along with
37	Coast Guard, CBP interdict cocaine on Miam	Sep 16th, 2020	The motor vessel La Temperance on the Miami River, Florida, Sept. 14, 2020. Coast Guard Sector Miami and U.S. Customs and B
38	Coast Guard crew offloads more than 26,00	Sep 10th, 2020	Coast Guard Cutter Bertholf crewmembers inspect a low-profile semi-submersible in international waters of the Eastern Pacific
39	USS Kidd, Coast Guard LEDET 401 seize $6 n	Aug 29th, 2020	The crew of the Coast Guard Cutter Resolute offloads 225 kilograms of cocaine and at Sector San San Juan Aug. 29, 2020.SAN JU
40	Coast Guard Cutter Hamilton offloads cocair	Aug 27th, 2020	Coast Guard Cutter Hamilton crew members offload approximately 11,500 pounds of cocaine and approximately 17,000 pounds
41	Coast Guard offloads $12 million in seized c	Aug 6th, 2020	The crew of the Coast Guard Cutter Joseph Napier transferred custody of two smugglers and offloaded 430 kilograms of cocaine
42	Coast Guard Cutter Legare offloads nearly 5	Aug 5th, 2020	Coast Guard Cutter Legare crew members offload about 3,900 pounds of marijuana in the rain, Aug. 5, 2020, Port Everglades, Fl

Sheet1

Ready Accessibility: Good to go Display Settings 100%

This program extracted data in the first and second scraping level.

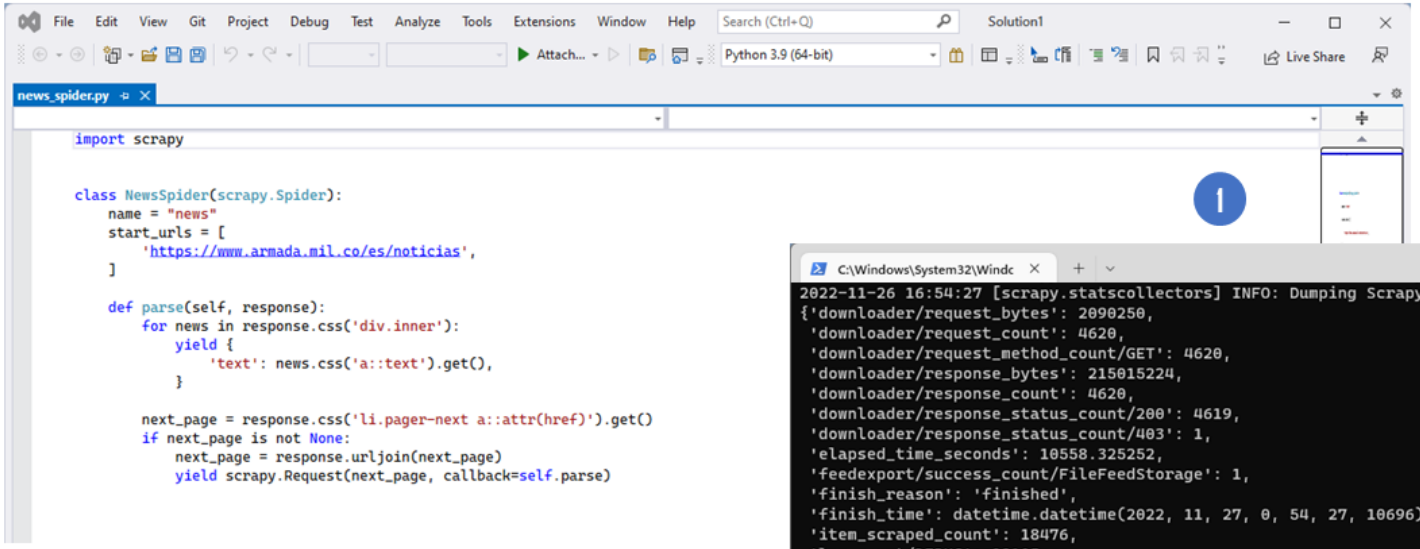
Web Sundew -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
PARTIAL	PARTIAL	PARTIAL	High	90.00	300.00	16.33	0.65	NO	NO	2000	High	High	NO
NO			1	450.00		16.98		0	0	1	3	3	1

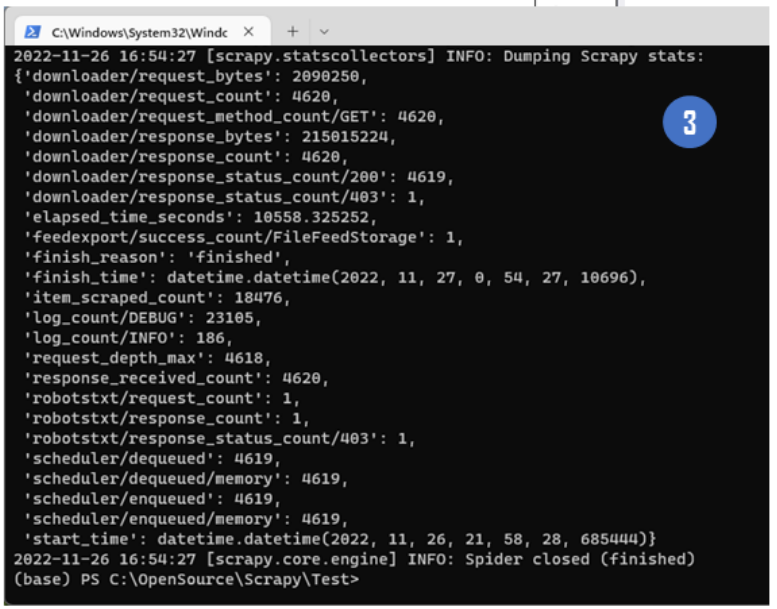
Remarks:

1. The state was not consistent with the selected field and erroneous content was extracted.
2. Pagination options were not useful for the ARC website.
3. Pagination options couldn't solve the news extension in the ARC website's second scraping level. Therefore, content extraction was incomplete.

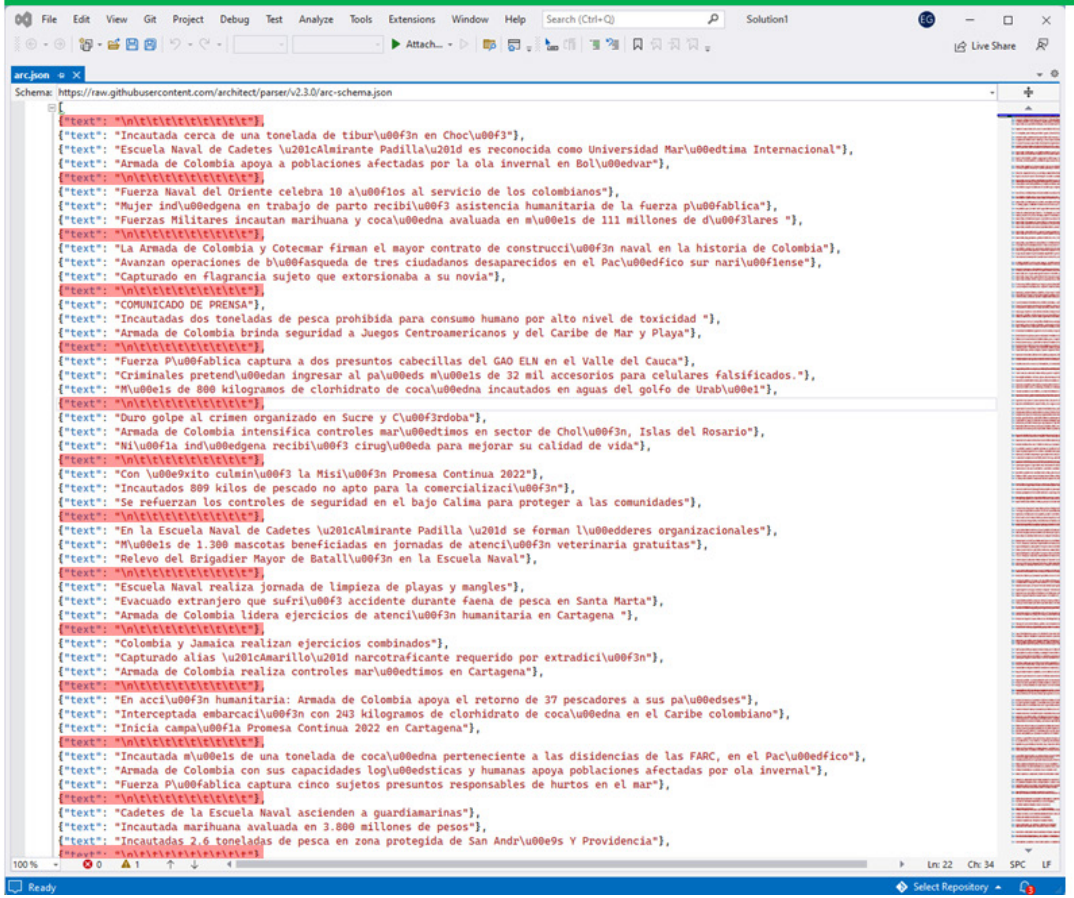
Scrapy -> ARC website



I was able to implement the extraction script of the titles for the first scraping level on the ARC website. More knowledge in Python is required to implement crawling and extraction in other levels.



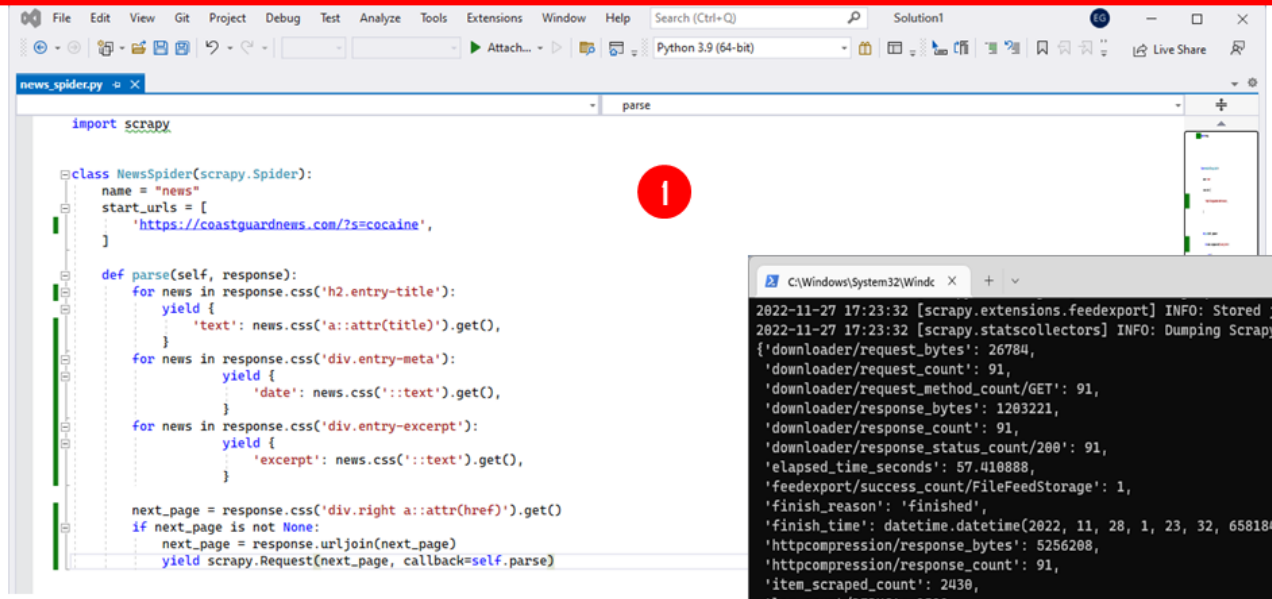
Scrapy -> ARC website results

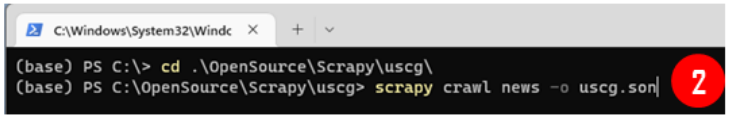


This program extracted data just in the first scraping level.

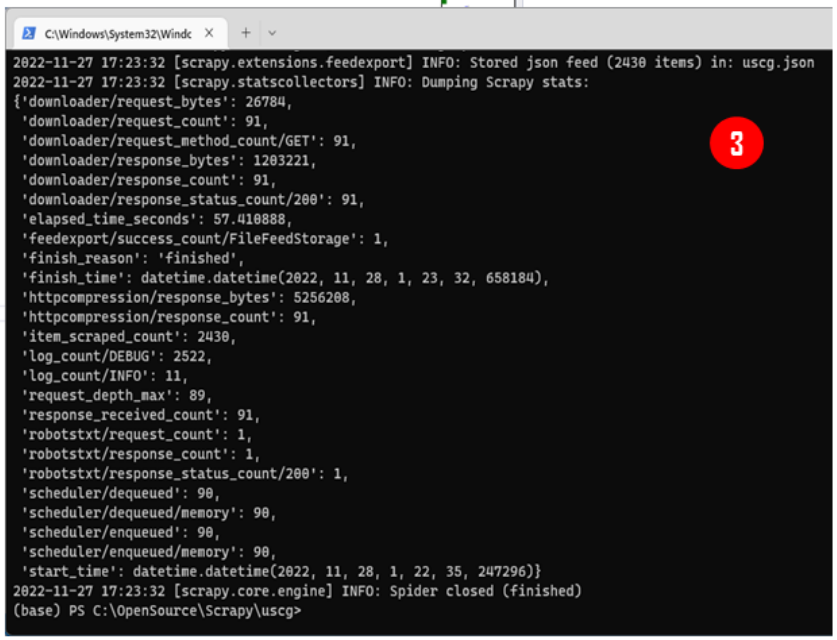
Every three titles the scraper is extracting a erroneous line. Also accent in Spanish are not presented properly.

Scrapy -> USCG website

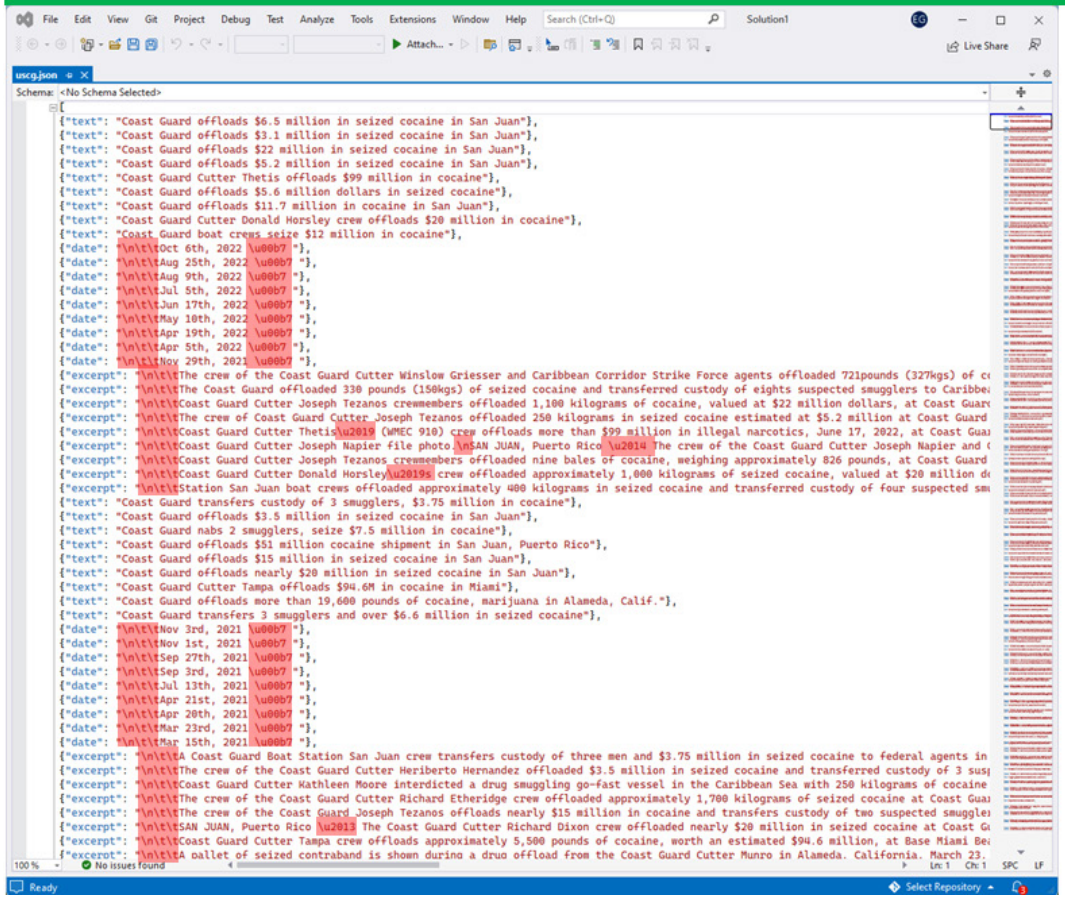




I was able to implement the extraction script of the titles, dates, and excerpts for the first scraping level on the USCG website. More knowledge in Python is required to implement crawling and extraction in other levels.



Scrapy -> USCG website results



This program extracted data just in the first scraping level.

All dates and excerpts contain HTML coding traces as well as parts of the excerpts' content.

Scrapy -> Results

Mandatory features			Complexity	Implementation time [min]		Scraping time [min]		Human behavior			Language fidelity		Blocked by NPS firewall
Automatic scraping	Automatic crawling	Scraping depth	GUI	Tutorial	Solution	ARC 100 web pages	USCG 100 web pages	IP rotation	CAPTCHA solution	Delay [ms]	Spanish	English	
PARTIAL	YES	NO	High	240.00	300.00	0.95	0.04	NO	NO	NO	Low	Medium	NO
NO			1	540.00		0.99		0	0	0	1	2	1

Remarks:

1. During scraping it creates erroneous entries in the ARC website case.
2. More programming knowledge is necessary to create the scraping depth capability.
3. Scrapy provides the fastest solution for scraping.
4. The language fidelity is a problem because there are HTML coding traces and misinterpretation of Spanish accents, and HTML coding traces when the web page is in English.

APPENDIX G. OCTOPARSE TASKS DETAILS

A. TASK 1: ARC NEWS WEBSITE

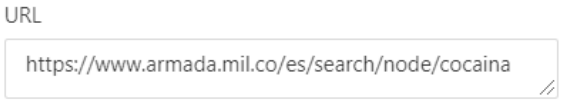


Figure 34. Exclusive URL for ARC website

Figure 35. ARC semi-automatic field detection

Figure 36. Workflow for the ARC website

B. TASK 2: PONAL NEWS WEBSITE

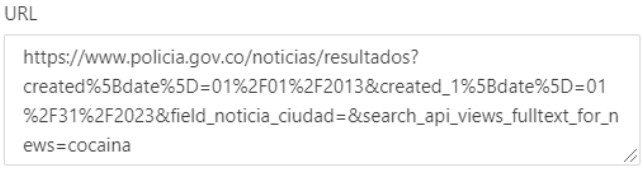


Figure 37. Exclusive URL for PONAL website

Figure 38. PONAL semi-automatic field detection

Figure 39. Workflow for the PONAL website

C. TASK 3: USCG NEWS WEBSITE

Figure 40. Exclusive URL for USCG website

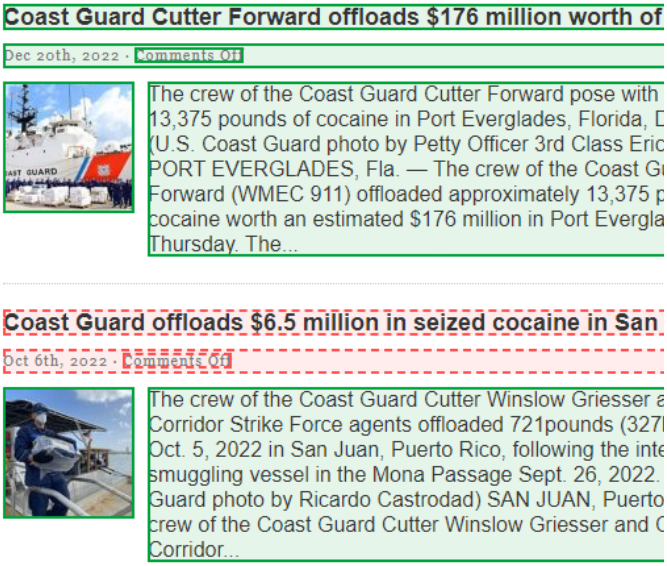


Figure 41. USCG semi-automatic field detection

Figure 42. Workflow for the USCG website

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H. EXCEL FILE WITH SCRAPED DATA FOR COCAINE SEIZURES DATASET

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I. HIVE SCRIPTS FOR COCAINE SEIZURES DATASET

A. HIVE SCRIPT FOR DATA SCRAPED FROM ARC NEWS WEBSITE

--IS4205-VM

--Local directory is /home/training/thesis/1.1-Colombia-Sea

--Ingestion of the scraped data 1.1-Colombia-Sea.csv (1484 records) from the local directory to the HDFS URI (Uniform Resource Identifier) hdfs://localhost/user/training/

$ hadoop fs -put 1.1-Colombia-Sea.csv /user/training/

--Execute Hive

$ hive

> set hive.cli.print.header=true;

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 11_colombia_sea;

--Create a new table structure in Hive's warehouse

> CREATE TABLE 11_colombia_sea (texto STRING);

--Load data from the 1.1-Colombia-Sea.csv file in HDFS to the table 11_colombia_sea Hive's warehouse

> LOAD DATA INPATH '1.1-Colombia-Sea.csv' INTO TABLE 11_colombia_sea;

> SELECT * FROM 11_colombia_sea LIMIT 2;

> SELECT COUNT(*) FROM 11_colombia_sea;

--Result 1484

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 11_colombia_sea_location;

--Create table with locations from 11_colombia_sea

> CREATE TABLE 11_colombia_sea_location AS SELECT REGEXP_REPLACE(texto, "\\\u{fffd}", "\\z") AS texto FROM 11_colombia_sea;

> CREATE TABLE 11_colombia_sea_location_final AS SELECT REGEXP_EXTRACT(LOWER(texto), '(mar caribe)', 0) AS caribe, REGEXP_EXTRACT(LOWER(texto), '(urabz)', 0) AS uraba, REGEXP_EXTRACT(LOWER(texto), '(paczfico)', 0) AS pacifico, REGEXP_EXTRACT(LOWER(texto), '(cauca)', 0) AS cauca, REGEXP_EXTRACT(LOWER(texto), '(valle)', 0) AS valle, REGEXP_EXTRACT(LOWER(texto), '(chocz)', 0) AS choco, REGEXP_EXTRACT(LOWER(texto), '(narizo)', 0) AS narino FROM 11_colombia_sea_location;

> SELECT * FROM 11_colombia_sea_location_final LIMIT 10;

> SELECT COUNT(*) FROM 11_colombia_sea_location_final;

--Result 1484

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 11_fecha_subtexto_tm;

--Create table in tm from 11_colombia_sea

> CREATE TABLE 11_fecha_subtexto_tm AS SELECT REGEXP_EXTRACT(texto,'(\\d{2}\\/\\d{2}\\/\\w+)',1) AS fecha, SPLIT(REVERSE(TRIM(SUBSTRING(texto,1,INSTR(texto,'tonelada')-1))),' ') AS subtexto FROM 11_colombia_sea;

> SELECT * FROM 11_fecha_subtexto_tm LIMIT 2;

> SELECT COUNT(*) FROM 11_fecha_subtexto_tm;

--Result 1484

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 11_fecha_cantidad_tm;

--Create table from 11_fecha_subtexto_tm

> CREATE TABLE 11_fecha_cantidad_tm AS SELECT SUBSTRING(fecha,7,4) AS year, SUBSTRING(fecha,1,2) AS month, SUBSTRING(fecha,4,2) AS day, REVERSE(subtexto[0]) AS cantidad FROM 11_fecha_subtexto_tm;

> SELECT * FROM 11_fecha_cantidad_tm LIMIT 20;

> SELECT COUNT(*) FROM 11_fecha_cantidad_tm;

--Result 1484

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 11_fecha_subtexto_kg;

--Create table in kg from 11_colombia_sea

> CREATE TABLE 11_fecha_subtexto_kg AS SELECT REGEXP_EXTRACT(texto,'(\\d{2}\\/\\d{2}\\/\\w+)',1) AS fecha, SPLIT(REVERSE(TRIM(SUBSTRING(texto,1,INSTR(texto,'kilo')-1)))," ") AS subtexto FROM 11_colombia_sea;

> SELECT * FROM 11_fecha_subtexto_kg LIMIT 2;

> SELECT COUNT(*) FROM 11_fecha_subtexto_kg;

--Result 1484

--Create table from 11_fecha_subtexto_kg

> CREATE TABLE 11_fecha_cantidad_kg AS SELECT SUBSTRING(fecha,7,4) AS year, SUBSTRING(fecha,1,2) AS month, SUBSTRING(fecha,4,2) AS day, REVERSE(subtexto[0]) AS cantidad FROM 11_fecha_subtexto_kg;

> SELECT * FROM 11_fecha_cantidad_kg LIMIT 20;

> SELECT COUNT(*) FROM 11_fecha_cantidad_kg;

--Result 1484

--Exporting files

$ hadoop fs -cat /user/hive/warehouse/11_colombia_sea_location_final/000000_0 > 11_colombia_sea_location.txt

$ hadoop fs -cat /user/hive/warehouse/11_fecha_cantidad_kg/000000_0 > 11_colombia_sea_fecha_cantidad_kg.txt$ hadoop fs -cat /user/hive/warehouse/11_fecha_cantidad_tm/000000_0 > 11_colombia_sea_fecha_cantidad_tm.txt

B. HIVE SCRIPT FOR DATA SCRAPED FROM PONAL NEWS WEBSITE

--IS4205-VM

--Local directory is /home/training/thesis/1.2-Colombia-Urban

--Ingestion of the scraped data 1.2-Colombia-Urban.csv (414 records) from the local directory to the HDFS URI (Uniform Resource Identifier) hdfs://localhost/user/training/

$ hadoop fs -put 1.2-Colombia-Urban.csv /user/training/

--Execute Hive

$ hive

> set hive.cli.print.header=true;

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 12_colombia_urban;

--Create a new table structure in Hive's warehouse

> CREATE TABLE 12_colombia_urban (title STRING, url STRING, place STRING, fecha STRING, texto STRING) row format delimited fields terminated BY “,”;

--Load data from the 1.2-Colombia-Urban.csv file in HDFS to the table 12_colombia_urban Hive's warehouse

> LOAD DATA INPATH '1.2-Colombia-Urban.csv' INTO TABLE 12_colombia_urban;

> SELECT * FROM 12_colombia_urban LIMIT 2;

> SELECT COUNT(*) FROM 12_colombia_urban;

--Result 414

--Discard the possibility of having a same name table in Hive's warehouse

```
> DROP TABLE IF EXISTS 12_colombia_urban_place_date;

--Create table with locations and dates from 12_colombia_urban

> CREATE TABLE 12_colombia_urban_place_date AS SELECT REGEXP_EXTRACT(place,'(.*\\;)',1) AS place, REGEXP_EXTRACT(fecha,'(\\d{2})',1) AS day, REGEXP_EXTRACT(fecha, 'de (.*\\ )de',1) AS month, REGEXP_EXTRACT(fecha,'(\\d{4})',1) AS year FROM 12_colombia_urban;

> SELECT * FROM 12_colombia_urban_place_date LIMIT 10;

> SELECT COUNT(*) FROM 12_colombia_urban_place_date;

--Result 414


--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 12_urban_subtexto_tm;

--Create table in tm from 12_colombia_urban

> CREATE TABLE 12_urban_subtexto_tm AS SELECT SPLIT(REVERSE(TRIM(SUBSTRING(title,1,INSTR(title,'ton')))),' ') AS subtexto FROM 12_colombia_urban;

> SELECT * FROM 12_urban_subtexto_tm LIMIT 50;

> SELECT COUNT(*) FROM 12_urban_subtexto_tm;

--Result 414


--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 12_urban_cantidad_tm;

--Create table from 12_urban_subtexto_tm

> CREATE TABLE 12_urban_cantidad_tm AS SELECT REVERSE(subtexto[0]) AS unit, REVERSE(subtexto[1]) AS qty FROM 12_urban_subtexto_tm;
```

```
> SELECT * FROM 12_urban_cantidad_tm LIMIT 50;

> SELECT COUNT(*) FROM 12_urban_cantidad_tm;

--Result 414


--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 12_urban_subtexto_kg;

--Create table in kg from 12_colombia_urban

> CREATE TABLE 12_urban_subtexto_kg AS SELECT SPLIT(REVERSE(TRIM(SUBSTRING(title,1,INSTR(title,'gramo'))))," ") AS subtexto FROM 12_colombia_urban;

> SELECT * FROM 12_urban_subtexto_kg LIMIT 20;

> SELECT COUNT(*) FROM 12_urban_subtexto_kg;

--Result 414


--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 12_urban_cantidad_kg;

--Create table from 12_urban_subtexto_kg

> CREATE TABLE 12_urban_cantidad_kg AS SELECT REVERSE(subtexto[0]) AS unit, REVERSE(subtexto[1]) AS qty FROM 12_urban_subtexto_kg;

> SELECT * FROM 12_urban_cantidad_kg LIMIT 20;

> SELECT COUNT(*) FROM 12_urban_cantidad_kg;

--Result 414


--Exporting files
```

$ hadoop fs -cat /user/hive/warehouse/12_colombia_urban_place_date/000000_0 > 12_colombia_urban_pd.txt

$ hadoop fs -cat /user/hive/warehouse/12_urban_cantidad_kg/000000_0 > 12_colombia_urban_kg.txt

$ hadoop fs -cat /user/hive/warehouse/12_urban_cantidad_tm/000000_0 > 12_colombia_urban_tm.txt

C. HIVE SCRIPT FOR DATA SCRAPED FROM USCG NEWS WEBSITE

--IS4205-VM

--Local directory is /home/training/thesis/1.4-US-Sea

--Ingestion of the scraped data 1.4-US-Sea.csv (1 header +175 records = 176) from the local directory to the HDFS URI (Uniform Resource Identifier) hdfs://localhost/user/training/

$ hadoop fs -put 1.4-US-Sea.csv /user/training/

--Execute Hive

$ hive

> set hive.cli.print.header=true;

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 14_us_sea;

--Create a new table structure in Hive's warehouse

> CREATE TABLE 14_us_sea (title STRING, url STRING, image STRING, fecha STRING, place STRING, texto STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ",";

--Load data from the 1.4-US-Sea.csv file in HDFS to the table 14_us_sea Hive's warehouse

> LOAD DATA INPATH '1.4-US-Sea.csv' INTO TABLE 14_us_sea;

> SELECT * FROM 14_us_sea LIMIT 2;

> SELECT COUNT(*) FROM 14_us_sea;

--Result 176

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 14_us_sea_place_date;

--Create table with locations and dates from 12_colombia_urban

> CREATE TABLE 14_us_sea_place_date AS SELECT REGEXP_EXTRACT(fecha,'(\\w{3,5})',1) AS month, REGEXP_EXTRACT(fecha,'(\\d{2})',1) AS day, REGEXP_EXTRACT(fecha,'(\\d{4})',1) AS year , REGEXP_EXTRACT(place,'(Pacific)',1) AS place FROM 14_us_sea;

> SELECT * FROM 14_us_sea_place_date LIMIT 10;

> SELECT COUNT(*) FROM 14_us_sea_place_date;

--Result 176

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 14_ussea_subtexto_tm;

--Create table in tm from 14_us_sea

> CREATE TABLE 14_ussea_subtexto_tm AS SELECT SPLIT(REVERSE(TRIM(REGEXP_EXTRACT(texto,'(^.*) metric',1)))," ") AS subtexto FROM 14_us_sea;

> SELECT * FROM 14_ussea_subtexto_tm;

> SELECT COUNT(*) FROM 14_ussea_subtexto_tm;

--Result 176

--Discard the possibility of having a same name table in Hive's warehouse

> DROP TABLE IF EXISTS 14_ussea_cantidad_tm;

--Create table from 14_ussea_subtexto_tm

> CREATE TABLE 14_ussea_cantidad_tm AS SELECT REVERSE(subtexto[0]) AS qty FROM 14_ussea_subtexto_tm;

> SELECT * FROM 14_ussea_cantidad_tm;

> SELECT COUNT(*) FROM 14_ussea_cantidad_tm;

--Result 176

--Exporting files

$ hadoop fs -cat /user/hive/warehouse/14_us_sea_place_date/000000_0 > 14_us_sea_pd.txt

$ hadoop fs -cat /user/hive/warehouse/14_ussea_cantidad_tm/000000_0 > 14_us_sea_tm.txt

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX J. EXCEL FILE WITH ETL DATA FROM HIVE FOR COCAINE SEIZURES DATASET

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX K. EXCEL FILES WITH FILTERS AND PROCESSES FOR COCAINE SEIZURES AND HOMICIDES DATASET

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

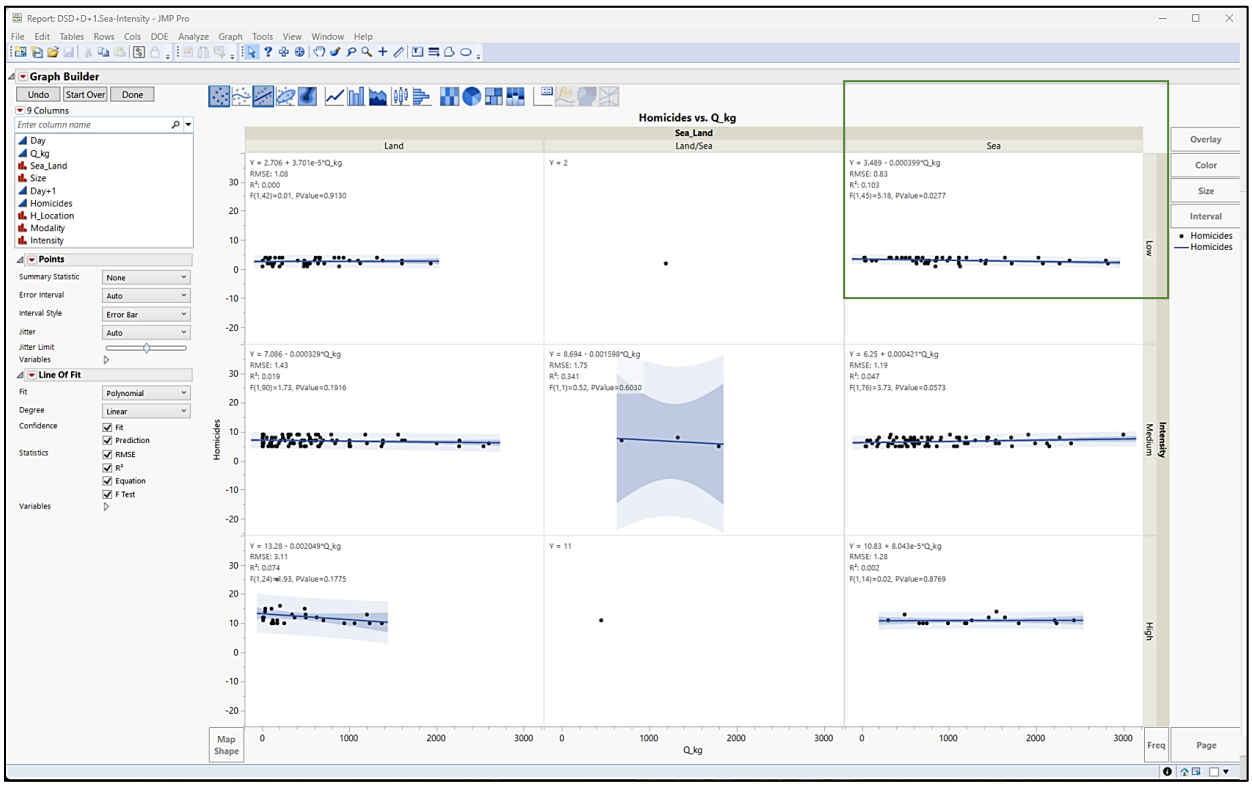
APPENDIX L. TIME SERIES DATASETS

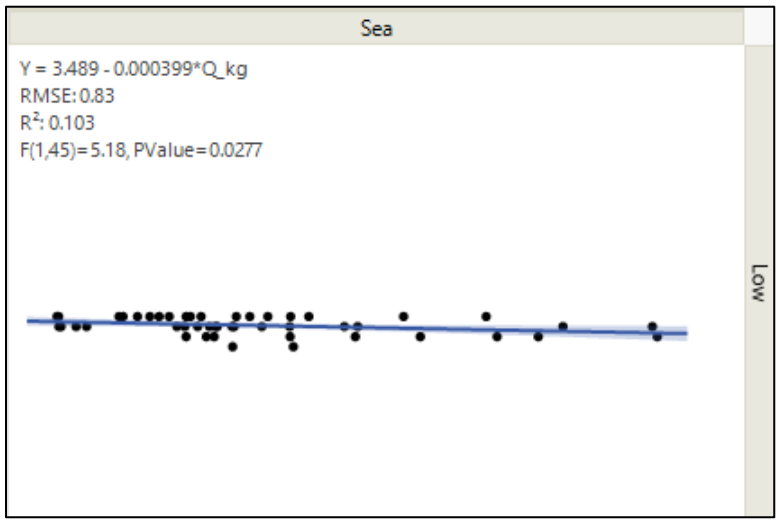
Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX M. FINDINGS ON POLYNOMIAL REGRESSIONS

A. FINDING 1



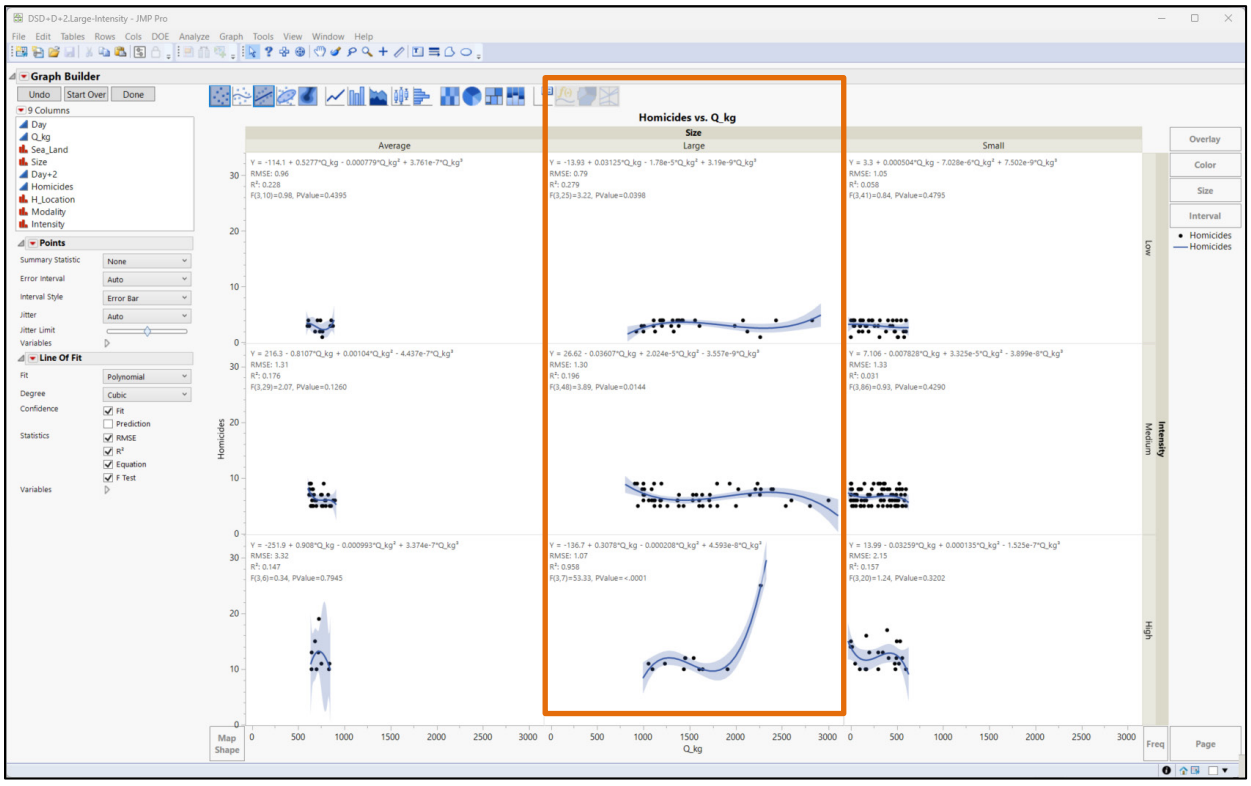


A significant correlation found between cocaine seizures on the Pacific Ocean and the rate of low-intensity homicides in the Colombian Pacific region (less than 5 homicides per day), supported by a low p-value of 0.02 and a sample size of 47 out of 308 records available (15.25%).

Nonetheless, this simple linear model only accounts for 10.3% of the variance (R^2) in the homicide rate within the Colombian Pacific region, attributed to the cocaine seizures on the Pacific Ocean from the preceding day. The RMSE indicates an average prediction error of roughly 0.83 homicides per day.

Also, considering most cocaine seizures were reported hours after the incident, as indicated by the gathered online data, an analogous association between land and sea/land categories was expected. Therefore, despite the correlation, it is not possible to generalize this relationship as indicative of causation.

B. FINDING 2



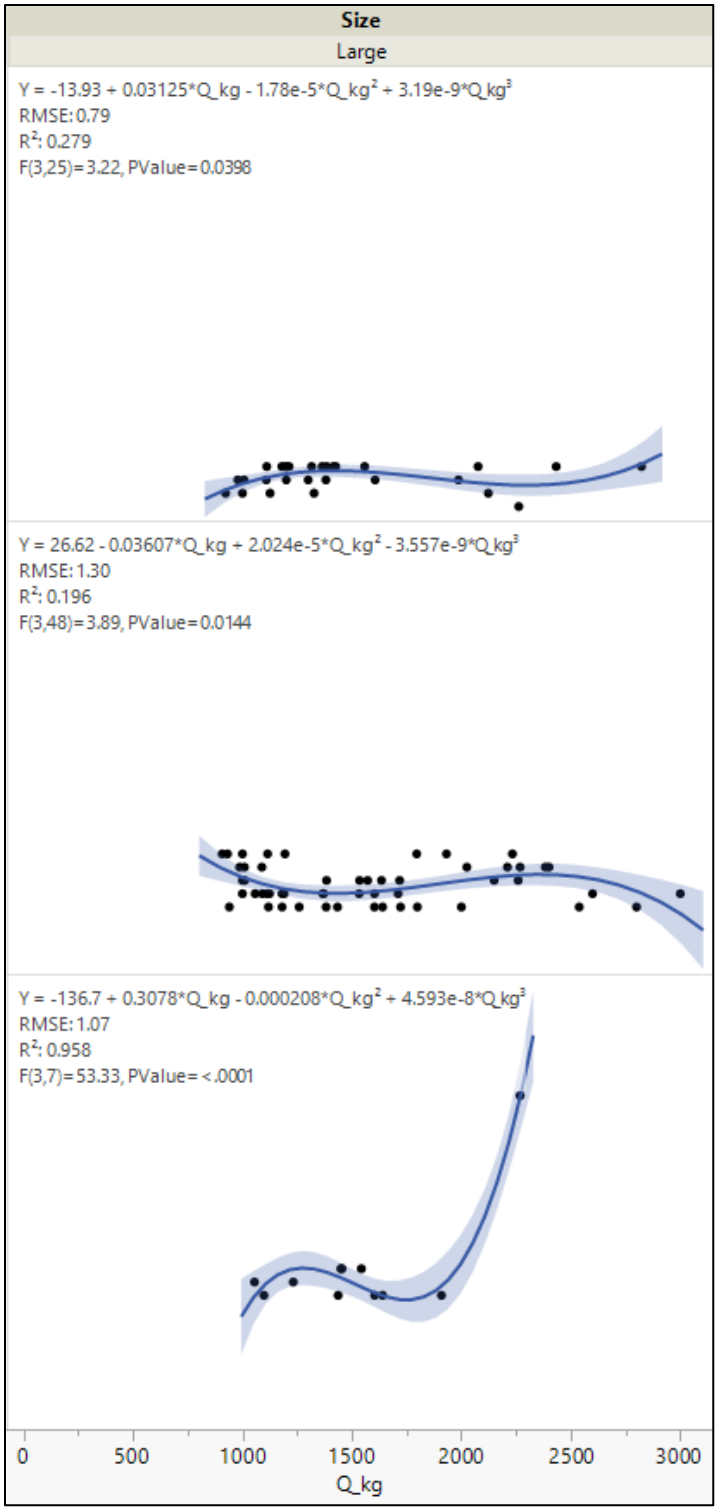
This second finding has a sample size of 92 out of 308 (29.87%). The absolute distribution of samples for each homicide intensity rate category is as follows: 29 out of 308 (9.41%) for low, 52 out of 308 (16.88%) for medium, and 11 out of 308 (3.5%) for high. The relative distribution of samples for each homicide intensity rate category is 29 out of 92 (31.52%) for low, 52 out of 92 (54.34%) for medium, and 11 out of 92 (11.95%) for high.

The p-values below 0.05 for both the low and medium categories of homicide intensity rates indicate noteworthy statistical support for linking large cocaine seizures with the homicide rate in the Colombian Pacific region. Moreover, the p-value below 0.001 for the high homicide rate category adds even greater support to this hypothesis.

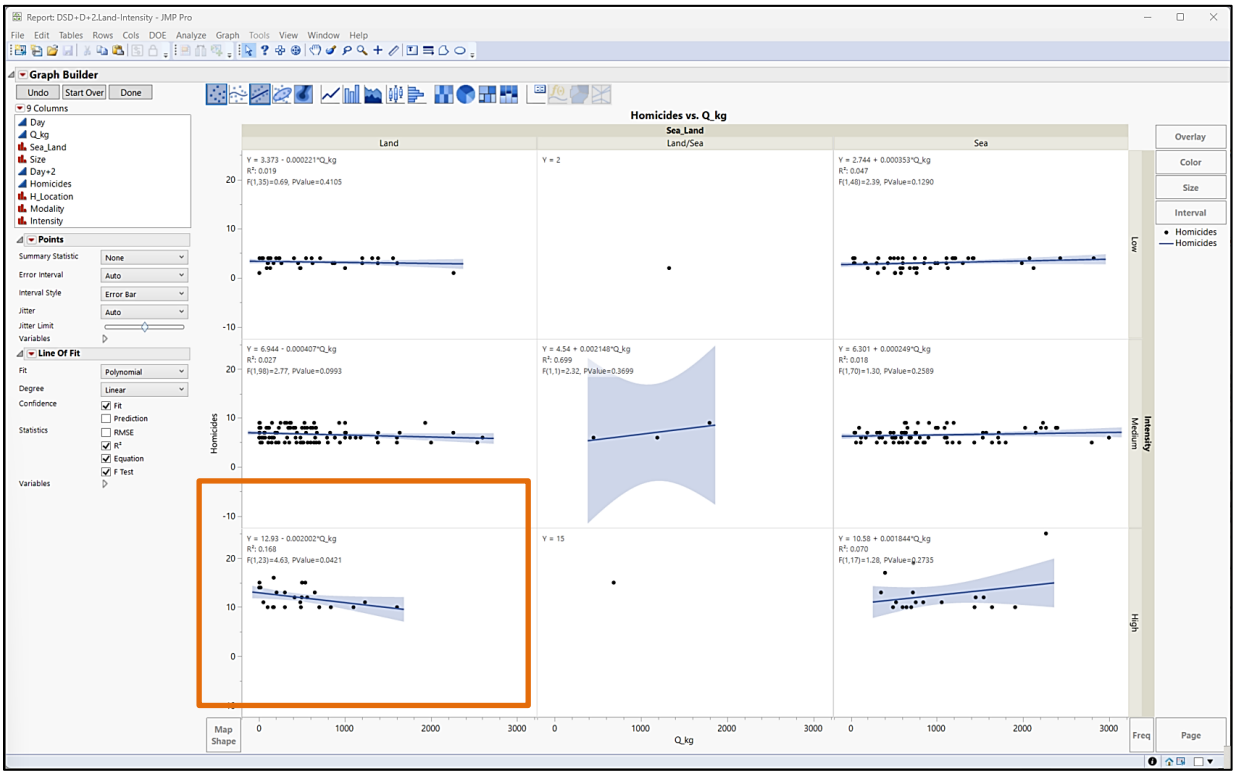
This cubic model accounts for 27.9% of the variability in the low-intensity category (< 5 homicides/day), 19.6% of the variability in the medium-intensity category (4 < homicides/day < 10), and a significant 95.8% of the variability in the high-intensity category (> 9 homicides/day) of homicides in the Colombian Pacific region attributed to large cocaine seizures (> 901.3 kg) occurring two days prior.

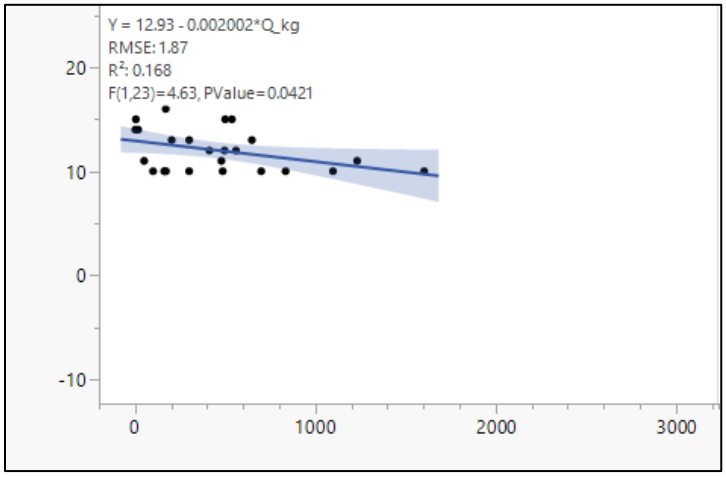
Additionally, the RMSE values for the model indicate that, on average, the predictions deviate by around 0.79 homicides/day for the low-intensity rate, 1.3 homicides/day for the medium-intensity rate, and 1.07 homicides/day for the high-intensity rate of homicides in the Colombian Pacific region.

The three cubic equations serve to illustrate the relationship between cocaine seizures and homicides in the Colombian Pacific region. As such, this discovery holds the potential to be a strong candidate for establishing a causal link between the variables under investigation.


Size
Large
Y = -13.93 + 0.03125*Q_kg - 1.78e-5*Q_kg² + 3.19e-9*Q_kg³
RMSE: 0.79
R²: 0.279
F(3,25)=3.22, PValue=0.0398
Y = 26.62 - 0.03607*Q_kg + 2.024e-5*Q_kg² - 3.557e-9*Q_kg³
RMSE: 1.30
R²: 0.196
F(3,48)=3.89, PValue=0.0144
Y = -136.7 + 0.3078*Q_kg - 0.000208*Q_kg² + 4.593e-8*Q_kg³
RMSE: 1.07
R²: 0.958
F(3,7)=53.33, PValue=<.0001
0
500
1000
1500
2000
2500
3000
Q_kg

C. FINDING 3





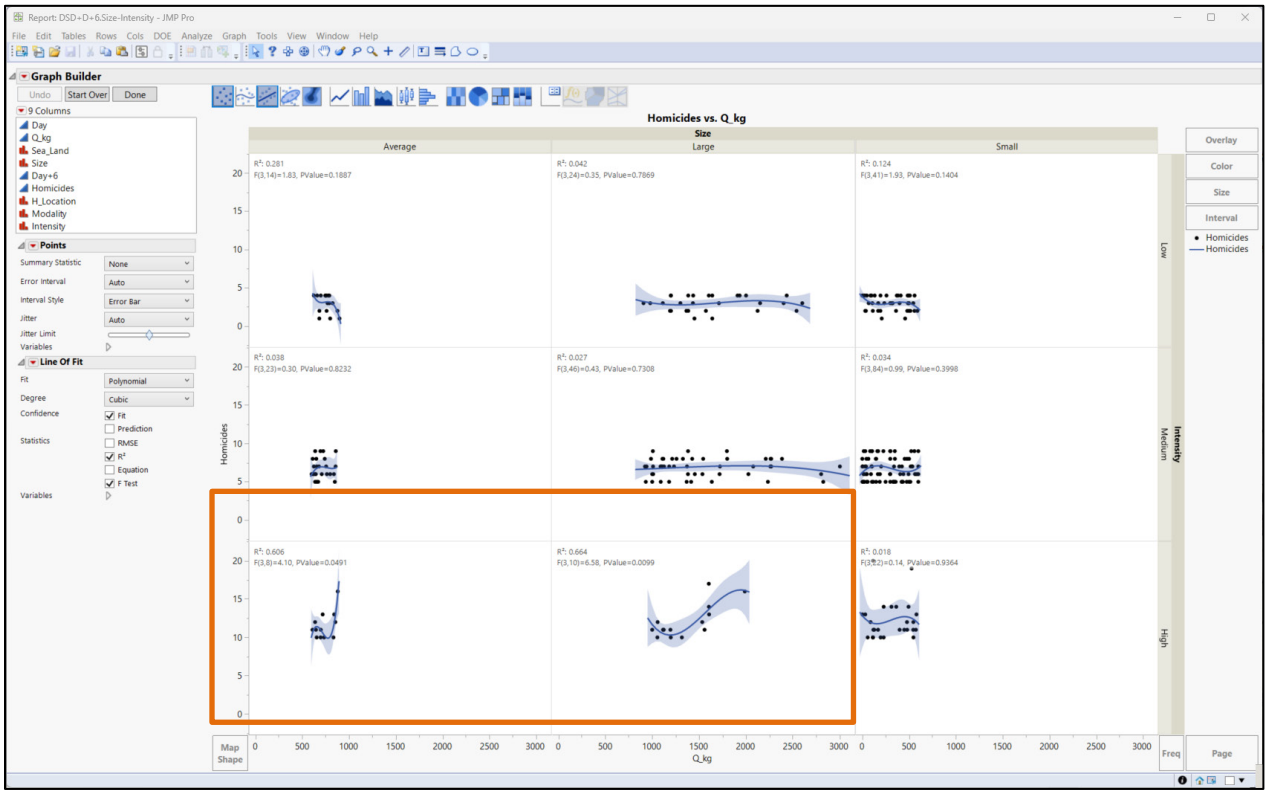
A significant correlation was found between cocaine seizures on the Colombian Pacific land and the high-intensity rate of homicides in the Colombian Pacific region (more than 9

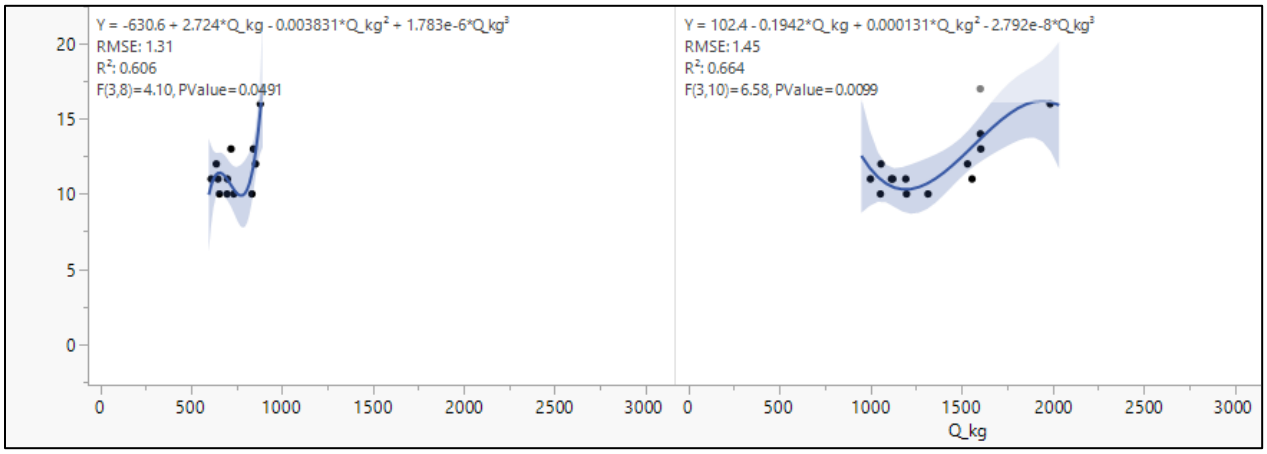
homicides/day). This association was supported by a low p-value (0.04) and was observed in a sample of 25 out of 308 instances (approximately 8.11%).

However, this linear model only explains 16.8% of the variation in the homicide rate within the Colombian Pacific region. This variance is attributed to the cocaine seizures from the preceding two days in the Pacific Ocean. Moreover, the RMSE of the model indicates an average prediction error of roughly 1.87 homicides per day.

Given that most cocaine seizures were typically reported hours after the actual incident, as indicated by the data collected from online sources, it was anticipated that a comparable connection between sea and sea/land categories would exist. Consequently, despite the observed correlation, it is important to note that this relationship cannot be generalized to imply causation.

D. FINDING 4





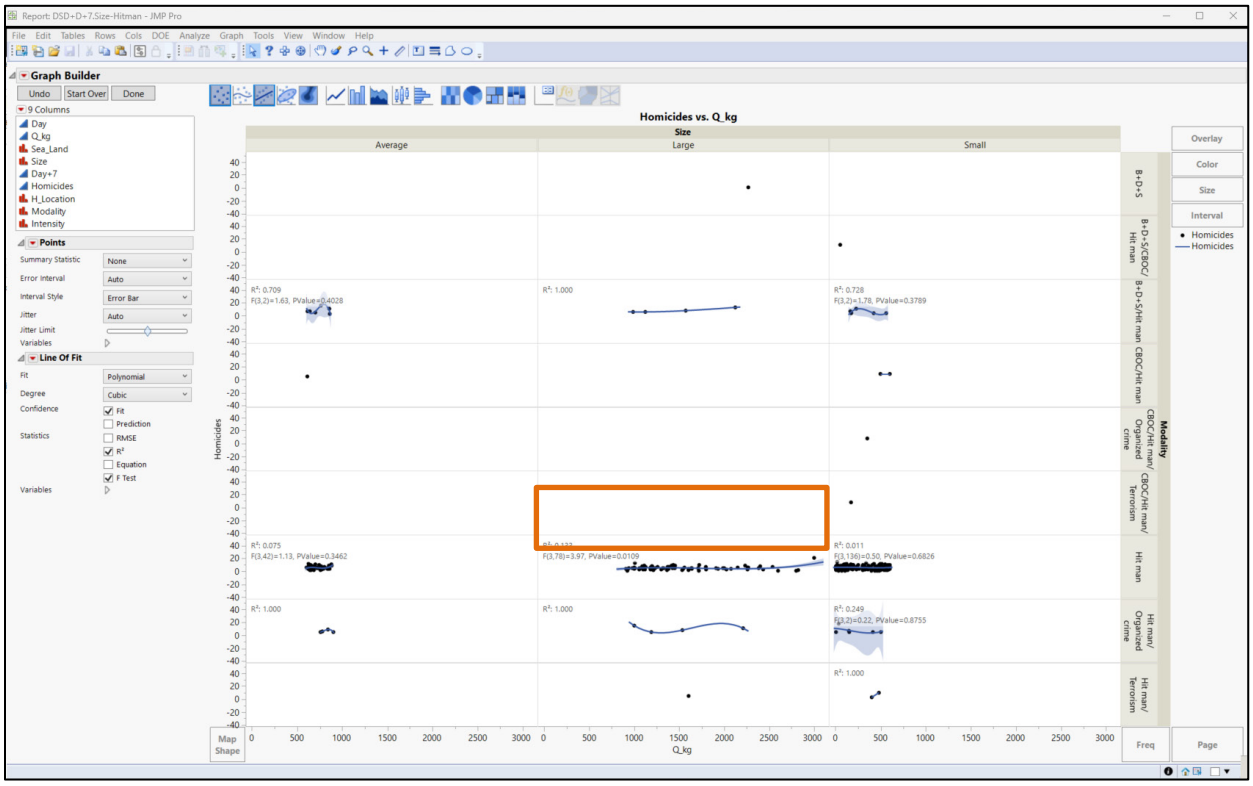
With a sample size of 26 out of 308 instances (8.44%), a p-value of less than 0.05 indicates a robust statistical significance between the occurrence of substantial cocaine seizures (> 901.3 kg) and moderate-sized seizures (601.3 kg to 901.3 kg) and the subsequent high-

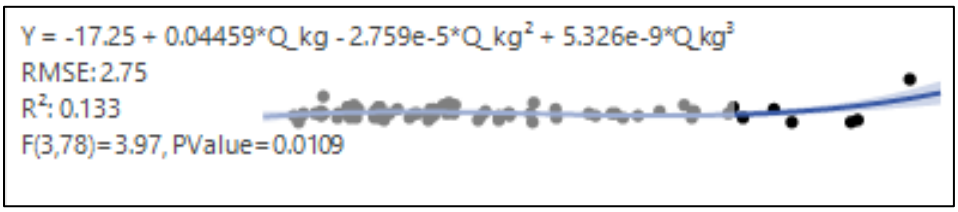
intensity homicide rate in the Colombian Pacific region six days later. The absolute distribution of the cocaine seizure sizes in the sample is as follows: 12 out of 308 instances (3.89%) for average-sized seizures and 14 out of 308 instances (4.54%) for large seizures. The relative distribution of the cocaine seizure sizes within the sample is 12 out of 26 instances (46.15%) for average-sized seizures and 14 out of 26 instances (53.84%) for large seizures.

The first cubic model explains 60.6% of the variance, whereas the second model explains 66.4% of the variance in the high-intensity (more than 9 homicides per day) homicide rate in the Colombian Pacific region. These variations are attributed to the occurrences of average and large cocaine seizures that took place in the Pacific six days prior. Also, the RMSE of the models indicates that, on average, the predictions deviate by approximately 1.31 and 1.45 homicides per day for the high-intensity homicide rate in the Colombian Pacific region.

The two cubic equations could potentially illustrate the linkage between cocaine seizures and homicides in the Colombian Pacific region. Consequently, this discovery holds promise as a compelling candidate to establish a causal relationship between the variables under scrutiny.

E. FINDING 5





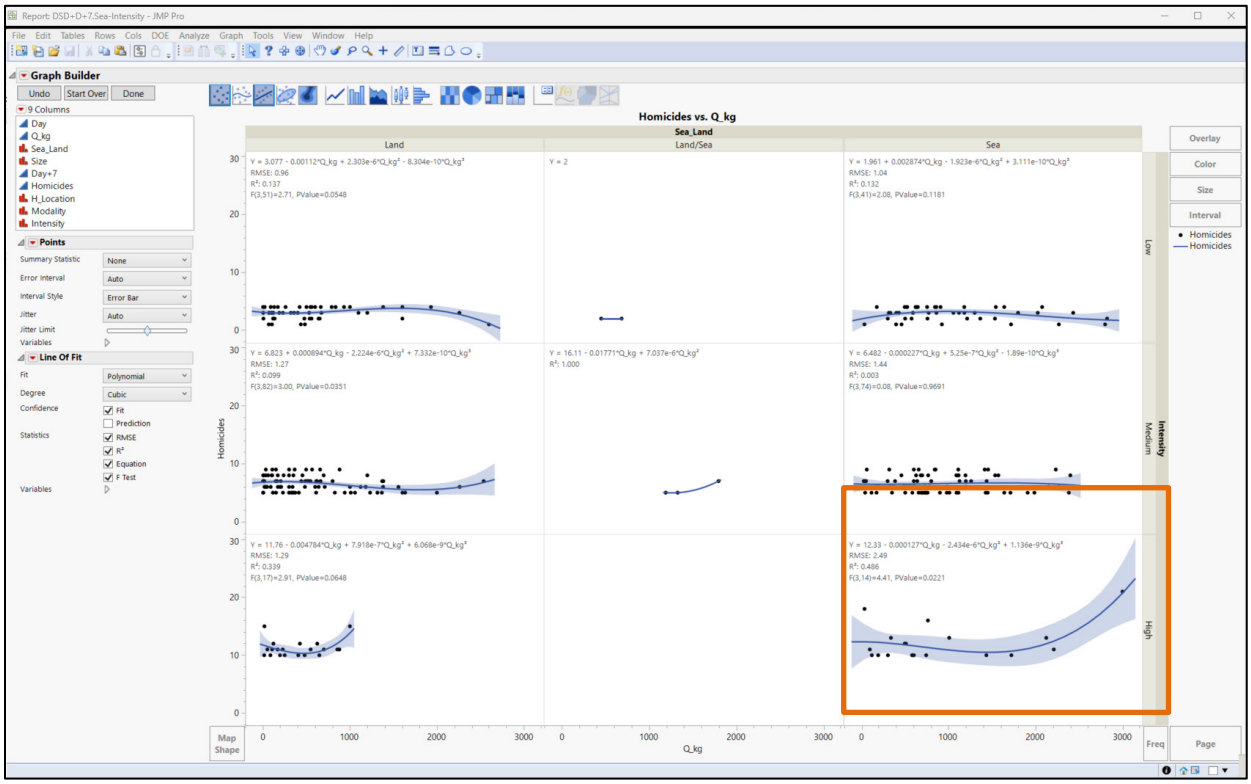
Based on a sample size of 82 out of 308 instances (26.62%), a p-value of 0.0109 indicates a strong statistical significance between the occurrence of large cocaine seizures (> 901.3 kg) and the occurrence of hitman-related homicides in the Colombian Pacific region seven days later.

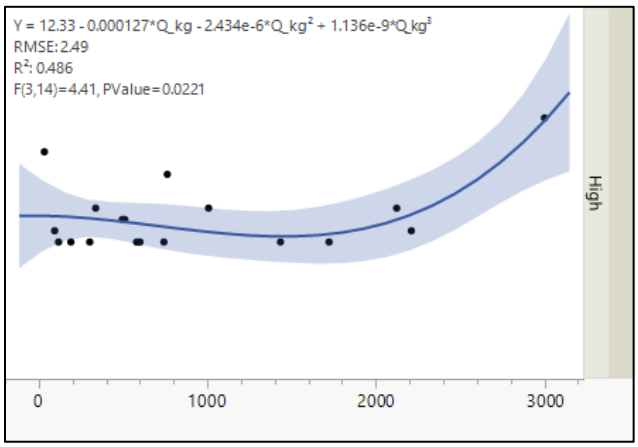
However, this cubic model only explains 13.3% of the variability in the hitman-related homicides modality within the Colombian Pacific region, attributed to the occurrence of large cocaine seizures that took place in the Pacific seven days before. Also, the RMSE of the model

indicates that, on average, the predictions deviate by approximately 2.75 homicides per day for the hitman modality of homicides in the Colombian Pacific region.

While the cubic equation may reveal a potential link between cocaine seizures and homicides in the Colombian Pacific region seven days later, it is important to note that this relationship should not be extended to imply causation.

F. FINDING 6



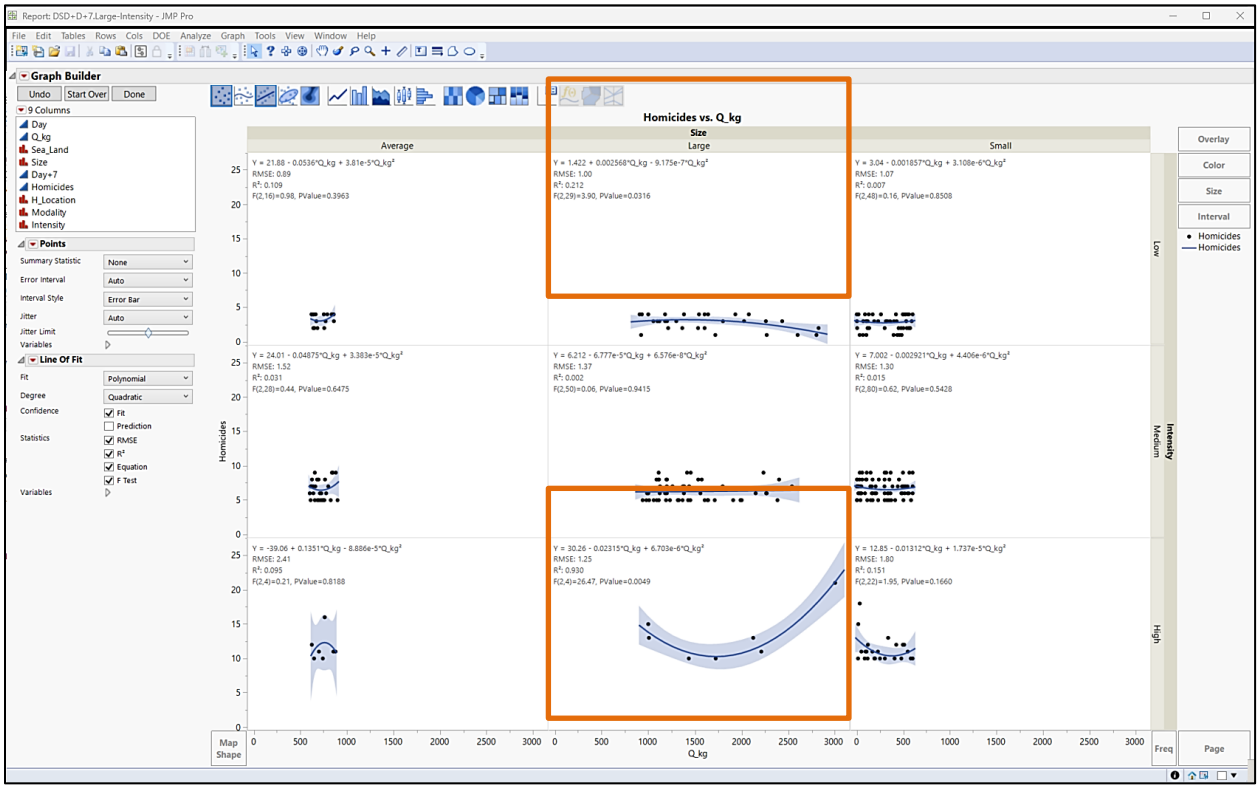


A significant correlation was found between cocaine seizures on the Pacific Ocean and the high-intensity rate of homicides in the Colombian Pacific region (more than 9 homicides per day), with a p-value (0.02) and a sample of 18 out of 308 (5.84%).

The cubic model explains 48.6% of the variance in the homicide rate within the Colombian Pacific region, attributable to the cocaine seizures that occurred in the Pacific Ocean over the preceding seven days. Moreover, the RMSE of the model indicates an average prediction error of roughly 2.49 homicides per day.

Although the cubic equation could show a connection between cocaine seizures and homicides in the Colombian Pacific region seven days after, this relationship cannot be generalized to show causality because the behavior is not consistent in the land and sea/land categories.

G. FINDING 7

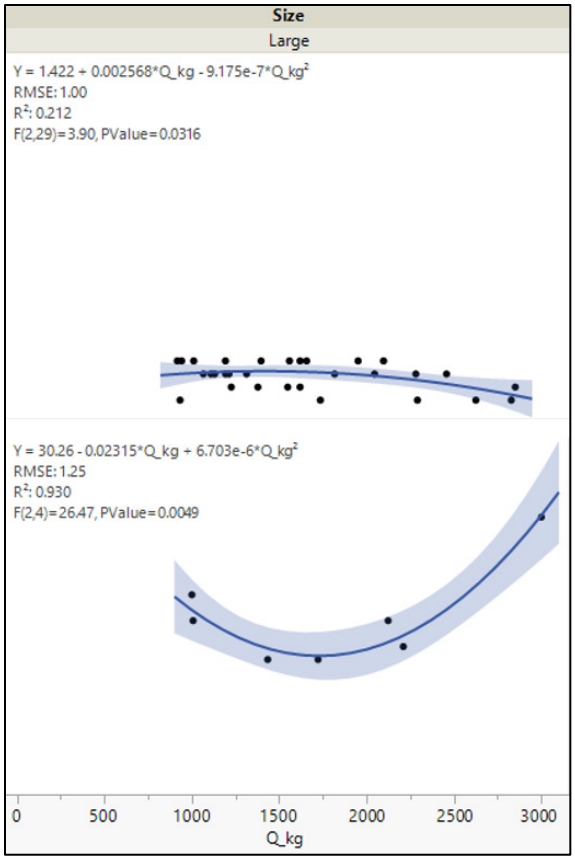


A significant correlation was found between cocaine seizures in the Pacific region and both the low-intensity (< 5 homicides/day) and high-intensity (> 9 homicides/day) rates of homicides in the Colombian Pacific region. This correlation is supported by a small p-value (< 0.05) and is based on a sample size of 25 out of 308 instances (approximately 8.11%).

The absolute distribution within the sample for the homicide intensity rates is as follows: 18 out of 308 instances (5.84%) for the low-intensity category and 7 out of 308 instances (2.27%) for the high-intensity category. The relative distribution within the sample for the homicide intensity rates is 18 out of 25 instances (72%) for the low-intensity category and 7 out of 25 instances (28%) for the high-intensity category.

The quadratic model explains 21.2% of the variation in the low-intensity homicide rate and a substantial 93% of the variance in the high-intensity homicide rate within the Colombian Pacific region. These variances are attributed to the occurrence of large cocaine seizures (> 901.3 kg) in the Pacific over the preceding seven days. Additionally, the RMSE of the model indicates an average prediction error of around 1 homicide per day for the low-intensity rate and 1.25 homicides per day for the high-intensity rate of homicides in the region.

While the quadratic equation might demonstrate an association between cocaine seizures and homicides in the Colombian Pacific Region occurring seven days later, it is important to emphasize that this connection should not be extended to imply a causal relationship. This is because there is not consistent evidence of the same effect in the medium-intensity homicide rate.


Size
Large
Y = 1.422 + 0.002568*Q_kg - 9.175e-7*Q_kg²
RMSE: 1.00
R²: 0.212
F(2,29)=3.90, PValue=0.0316
Y = 30.26 - 0.02315*Q_kg + 6.703e-6*Q_kg²
RMSE: 1.25
R²: 0.930
F(2,4)=26.47, PValue=0.0049
0
500
1000
1500
2000
2500
3000
Q_kg

APPENDIX N. GRAPH BUILDER FITTING SUBSETS

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX O. DSD+S+2-SUBSET MODELS

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX P. DSD+S+6-SUBSET MODELS

Refer to the “Supplementals” section.

THIS PAGE INTENTIONALLY LEFT BLANK

SUPPLEMENTALS

To access the supplemental material(s) listed here, contact the Dudley Knox Library or visit the thesis pages in the library's Calhoun database.

APPENDIX H SUPPLEMENTALS

1. Supplemental 1_Appendix_H-1.1-Colombia-Sea

   Data scraped from ARC news website using Octoparse.

2. Supplemental 2_Appendix_H-1.2-Colombia-Urban

   Data scraped from PONAL news website using Octoparse.

3. Supplemental 3_Appendix_H-1.4-US-Sea

   Data scraped from USCG news website using Octoparse.

APPENDIX J SUPPLEMENTALS

1. Supplemental 4_Appendix_J-1.1-Colombia-ARC-Hive

   ARC ETL data from Hive for cocaine seizures dataset.

2. Supplemental 5_Appendix_J-1.2-Colombia-PONAL-Hive

   PONAL ETL data from Hive for cocaine seizures dataset.

3. Supplemental 6_Appendix_J-1.4-US-USCG-Hive

   USCG ETL data from Hive for cocaine seizures dataset.

APPENDIX K SUPPLEMENTALS

1. Supplemental 7_Appendix_K-1.1-Colombia-ARC-Excel

   ARC filtered data for cocaine seizures dataset.

2. Supplemental 8_Appendix_K-1.2-Colombia-PONAL-Excel

   PONAL filtered data for cocaine seizures dataset.

3. Supplemental 9_Appendix_K-1.4-US-USCG-Excel

   USCG filtered data for cocaine seizures dataset.

4. Supplemental 10_Appendix_K-Pacific cocaine seizures dataset

   Pacific cocaine seizures dataset.

5. Supplemental 11_Appendix_K-Colombian Pacific homicides dataset

   Colombian Pacific region homicides dataset.

APPENDIX L SUPPLEMENTALS

1. Supplemental 12_Appendix_L-Cocaine seizures and homicides dataset

   Cocaine seizures and homicides time-series dataset.

2. Supplemental 13_Appendix_L-DSD+D+1

   Cocaine seizures and next day homicides time-series dataset.

3. Supplemental 14_Appendix_L-DSD+D+2

   Cocaine seizures and next two days homicides time-series dataset.

4. Supplemental 15_Appendix_L-DSD+D+3

   Cocaine seizures and next three days homicides time-series dataset.

5. Supplemental 16_Appendix_L-DSD+D+4

   Cocaine seizures and next four days homicides time-series dataset.

6. Supplemental 17_Appendix_L-DSD+D+5

   Cocaine seizures and next five days homicides time-series dataset.

7. Supplemental 18_Appendix_L-DSD+D+6

   Cocaine seizures and next six days homicides time-series dataset.

8. Supplemental 19_Appendix_L-DSD+D+7

   Cocaine seizures and next seven days homicides time-series dataset.

APPENDIX M SUPPLEMENTALS

1. Supplemental 20_Appendix_M-DSD+D+1.Sea-Intensity

   Cocaine seizures and next day homicides time-series dataset findings on sea and intensity categories.

2. Supplemental 21_Appendix_M-DSD+D+2.Land-Intensity

   Cocaine seizures and next two days homicides time-series dataset findings on land and intensity categories.

3. Supplemental 22_Appendix_M-DSD+D+2.Large-Intensity

   Cocaine seizures and next two days homicides time-series dataset findings on large and intensity categories.

4. Supplemental 23_Appendix_M-DSD+D+6.Size-Intensity

   Cocaine seizures and next six days homicides time-series dataset findings on size and intensity categories.

5. Supplemental 24_Appendix_M-DSD+D+7.Large-Intensity

   Cocaine seizures and next seven days homicides time-series dataset findings on large and intensity categories.

6. Supplemental 25_Appendix_M-DSD+D+7.Sea-Intensity

   Cocaine seizures and next seven days homicides time-series dataset findings on sea and intensity categories.

7. Supplemental 26_Appendix_M-DSD+D+7.Size-Hitman

   Cocaine seizures and next seven days homicides time-series dataset findings on size and hitman categories.

APPENDIX N SUPPLEMENTALS

1. Supplemental 27_Appendix_N-DSD+D+2 subset

   Cocaine seizures and next two days homicides time-series subset.

2. Supplemental 28_Appendix_N-DSD+D+2 subset

   Cocaine seizures and next two days homicides time-series subset graph builder regression.

3. Supplemental 29_Appendix_N-DSD+D+6 subset

   Cocaine seizures and next six days homicides time-series subset.

4. Supplemental 30_Appendix_N-DSD+D+6 subset

   Cocaine seizures and next six days homicides time-series subset graph builder regression.

APPENDIX O SUPPLEMENTALS

1. Supplemental 31_Appendix_O-DSD+D+2 subset-Fit Least Squares

   Cocaine seizures and next two days homicides time-series subset least squares fitting.

2. Supplemental 32_App._O-DSD+D+2 subset-Fit Least Squares no outliers

   Cocaine seizures and next two days homicides time-series subset least squares fitting excluding outliers.

3. Supplemental 33_Appendix_O-DSD+D+2 subset-Fit Least Squares no row 10

   Cocaine seizures and next two days homicides time-series subset least squares fitting excluding data point in row 10.

4. Supplemental 34_App._O-DSD+D+2 subset no row 10-Distribution of Q_kg

   Continuous distribution for cocaine seizures time-series subset excluding data point in row 10.

5. Supplemental 35_Appendix_O-DSD+D+2 subset (no row 10) simulation

   Simulation results table for cocaine seizures and next two days homicides time-series excluding data point in row 10.

APPENDIX P SUPPLEMENTALS

1. Supplemental 36_Appendix_P-DSD+D+6 subset-Fit Least Squares

   Cocaine seizures and next six days homicides time-series subset least squares fitting.

2. Supplemental 37_App._P-DSD+D+6 subset-Fit Least Squares no outliers

   Cocaine seizures and next six days homicides time-series subset least squares fitting excluding outliers.

3. Supplemental 38_Ap._P-DSD+D+6 subset no outliers-Distribution Q_kg

   Continuous distribution for cocaine seizures time-series subset excludng outliers.

4. Supplemental 39_Appendix_P-DSD+D+6 subset (no outliers) simulation

   Simulation results table for cocaine seizures and next six days homicides time-series excluding outliers.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Abi-Habib, M., Lopez, O., & Cegarra, A. (2022, August 31). 'Absolute Warfare': Cartels Terrorize Mexico as Security Forces Fall Short. *The New York Times*. https://www.nytimes.com/2022/08/31/world/americas/mexico-cartels-violence.html

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, *82*(2), 261–277. https://doi.org/10.1037/h0076477

Álvarez, C., & González, C. (2012). Análisis espacial de la violencia homicida en el pacífico colombiano [Spatial analysis of homicidal violence in the Colombian Pacific]. *Revista de Economía & Administración E-ISSN 2463–1035 ISSN 1794–7561*, *9*(1), Article 1. https://revistas.uao.edu.co/ojs/index.php/REYA/article/view/168

Anastasijevic, D. (2010). Getting Better? A Map of Organized Crime in the Western Balkans. In W. Benedek, C. Daase, V. Dimitrijević, & P. van Duyne (Eds.), *Transnational Terrorism, Organized Crime and Peace-Building: Human Security in the Western Balkans* (pp. 149–168). Palgrave Macmillan UK. https://doi.org/10.1057/9780230281479_9

Arias, E. D., & Grisaffi, T. (Eds.). (2021). *Cocaine: From Coca Fields to the Streets*. Duke University Press. https://directory.doabooks.org/handle/20.500.12854/78356

Armada de Colombia. (2022a). *Política de Seguridad de la Información | Armada Nacional* [Information security policy | Colombian Navy]. https://www.armada.mil.co/es/content/politica-de-seguridad-de-la-informacion

Armada de Colombia. (2022b, November 5). *Busqueda noticias Armada Nacional* [Search for news from the Colombian Navy]. https://www.armada.mil.co/es/search/node/cocaina

Asamoah, M. K. (2014). Re-examination of the limitations associated with correlational research. *Journal of Educational Research and Reviews*, *2*(4), 45–52.

Aschner, J. P., & Montero, J. C. (2021). Architectures, spaces, and territories of illicit drug trafficking in Colombia and Mexico. *Crime, Media, Culture*, *17*(3), 327–351. https://doi.org/10.1177/1741659020910212

Australian Institute of Criminology. (1999). *Drug Use*. Australian Institute of Criminology. https://www.aic.gov.au/statistics/drug-use

Aziani, A. (2020). Violent disequilibrium: The influence of instability in the economic value of cocaine markets on homicides. *Crime, Law and Social Change*, *74*(3), 245–272. https://doi.org/10.1007/s10611-020-09894-2

Barua, R., Maity, R., Minj, D., Barua, T., & Layek, A. K. (2019). F-NAD: An Application for Fake News Article Detection using Machine Learning Techniques. *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 1–6. https://doi.org/10.1109/IBSSC47189.2019.8973059

Bitkoska, N. (2022, September 6). *How Many websites Are There?* WebsiteBuilder.Org. https://websitebuilder.org/blog/how-many-websites-are-there/

Bright Mountain Media. (2022, November 5). *You searched for cocaine—Coast Guard News | Coast Guard News*. Coast Guard News. https://coastguardnews.com/?s=cocaine

Bright Mountain Media. (2023a). *Privacy Policy—Coast Guard News | Coast Guard News*. https://coastguardnews.com/privacy-policy/

Bright Mountain Media. (2023b). *Terms of Use—Coast Guard News | Coast Guard News*. https://coastguardnews.com/about/terms-of-use/

Caruana, A., & Ramaseshan, B. (2015). The effect of service quality and consumer trust on retail website Loyalty. In H. E. Spotts (Ed.), *Creating and Delivering Value in Marketing* (pp. 76–76). Springer International Publishing. https://doi.org/10.1007/978-3-319-11848-2_21

Castillo Gallardo, M., & Durand Guevara, A. (2008). Movimiento cocalero, política y representación: Los casos boliviano y peruano [Coca grower movement, politics, and representation: The Bolivian and Peruvian cases]. *Identidades, etnicidad y racismo en América Latina*.

Castro, J. (2017, December 9). *Tenemos más territorio que Estado* [We have more territory than the State]. El Tiempo. https://www.eltiempo.com/opinion/columnistas/jaime-castro/tenemos-mas-territorio-que-estado-160082

Charles, M. H. (2022). Occo análisis: Ecuador's battle against international drug trafficking. https://doi.org/10.48713/10336_36768

Congreso de la República de Colombia. (2008). *Ley 1266 de 2008: Ley hábeas data y regulación el manejo de la información contenida en bases de datos personales* [Law 1266 of 2008: Data protection law and regulation of the management of information contained in personal databases.]. http://www.secretariasenado.gov.co/senado/basedoc/ley_1266_2008.html

Congreso de la República de Colombia. (2012). *Ley 1581 de 2012: Ley para la protección de datos personales* [Law 1581 of 2012: Law for the protection of personal data]. http://www.secretariasenado.gov.co/senado/basedoc/ley_1581_2012.html

Congreso de la República de Colombia. (2014). *Ley 1712 de 2014: Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional* [Law 1712 of 2014: Transparency and the Right to Access National Public Information Law]. http://www.secretariasenado.gov.co/senado/basedoc/ley_1712_2014.html

Cotte Poveda, A. (2012). Violence and economic development in Colombian cities: A dynamic panel data analysis. *Journal of International Development*, *24*(7), 809–827. https://doi.org/10.1002/jid.2819

Cowls, J., & Schroeder, R. (2015). Causation, correlation, and Big Data in social science research. *Policy & Internet*, *7*(4), 447–472. https://doi.org/10.1002/poi3.100

Creswell, J. W. (2012). Educational research; planning, conducting, and evaluating quantitative and qualitative research. *Reference and Research Book News*, *26*(3). https://www.proquest.com/docview/869979977/citation/AEEE084E6C2F442EPQ/1

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Sage publications.

Crow, E. L., & Shimizu, K. (1987). *Lognormal distributions*. Marcel Dekker New York.

Dávalos, E., & Morales, L. F. (2022). *Diffusion of crime control benefits: Forced eradication and coca crops in Colombia – ProQuest*. https://www.proquest.com/docview/2746615551?pq-origsite=primo

Department of Homeland Security. (2023). *United States Coast Guard (USCG)*. https://www.news.uscg.mil/

Desimone, J. (2001). The effect of cocaine prices on crime. *Economic Inquiry*, *39*(4), 627–643. https://doi.org/10.1093/ei/39.4.627

Di Sotto, S., & Viviani, M. (2022). Health misinformation detection in the social web: An overview and a data science approach. *International Journal of Environmental Research and Public Health*, *19*(4), Article 4. https://doi.org/10.3390/ijerph19042173

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). Web scraping: State-of-the-art and areas of application. *2019 IEEE International Conference on Big Data (Big Data)*, 6040–6042. https://doi.org/10.1109/BigData47090.2019.9005594

Dreyer, A., & Stockton, J. (2013, July 15). *Internet "Data Scraping": A Primer for Counseling Clients | New York Law Journal*. https://www.law.com/newyorklawjournal/almID/1202610687621/

Duijn, P. A. C., Kashirin, V., & Sloot, P. M. A. (2014). The relative ineffectiveness of criminal network disruption. *Scientific Reports (Nature Publisher Group)*, *4*, 4238. https://doi.org/10.1038/srep04238

Durand-Ochoa, U. (2012). *Coca, contention and identity: The political empowerment of the Cocaleros of Bolivia and Peru* [Phd, London School of Economics and Political Science]. http://etheses.lse.ac.uk/560/

Durán-Martínez, A. (2015). Drugs around the corner: domestic drug markets and violence in Colombia and Mexico. *Latin American Politics and Society*, *57*(3), 122–146. https://doi.org/10.1111/j.1548-2456.2015.00274.x

Echandía, C. (2013). Narcotráfico: Génesis de los paramilitares y herencia de bandas criminals [Narcotrafficking: Genesis of the paramilitaries and inheritance of criminal gangs]. *Informes FIP*, *19*, 5–32.

Ertam, F. (2018). Deep learning based text classification with Web Scraping methods. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–4. https://doi.org/10.1109/IDAP.2018.8620790

European Monitoring Centre for Drugs and Drug Addiction. (1993). *EMCDDA home page | www.emcdda.europa.eu*. https://www.emcdda.europa.eu/emcdda-home-page_en

Evans, W. N., Garthwaite, C., & Moore, T. J. (2018). *Guns and Violence: The Enduring Impact of Crack Cocaine Markets on Young Black Males* (Working Paper 24819). National Bureau of Economic Research. https://doi.org/10.3386/w24819

Fajardo Cely, D. M. (2014). Grupo de Memoria Histórica, ¡Basta ya! Colombia: Memorias de guerra y dignidad [Historical memory group, "Enough already!" Colombia: Memories of war and dignity]. (Bogotá: Imprenta Nacional, 2013), 431 pp. *Historia y Sociedad*, *26*, 274–281.

Farber, D. (2019). Crack: Rock Cocaine, Street Capitalism, and the Decade of Greed. Cambridge University Press.

Ferdinand, R. F., Blüm, M., & Verhulst, F. C. (2001). Psychopathology in adolescence predicts substance use in young adulthood. *Addiction*, *96*(6), 861–870. https://doi.org/10.1046/j.1360-0443.2001.9668617.x

Ferro, T. M., & Castro, E. (2019). *¿Es eficaz la erradicación forzosa de cultivos de coca? La evidencia indica que no* [Is forced eradication of coca crops effective? The evidence indicates it is not]. https://cesed.uniandes.edu.co/wp-content/uploads/2019/11/Es-Eficaz-La-Erradicacion-Forzosa-De-Cultivos-De-Coca-Corregido-20nov2019-1.pdf

Ford, A. (2022, February 10). How Brazil's gangs took their war to Santa Cruz, Bolivia. *InSight Crime*. https://insightcrime.org/news/brazil-gangs-took-their-war-santa-cruz-bolivia/

Fox, J. A. (2000). Demographics and U.S. homicide. *The crime drop in America*, (Cambridge Studies in Criminology, pp. 288-318). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616167.010.

Gaglani, J., Gandhi, Y., Gogate, S., & Halbe, A. (2020). Unsupervised WhatsApp Fake News Detection using Semantic Search. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 285–289. https://doi.org/10.1109/ICICCS48265.2020.9120902

Galvis-Aponte, L. A., Moyano-Támara, L. M., & Alba-Fajardo, C. A. (2017). La persistencia de la pobreza y sus factores asociados [The persistence of poverty and its associated factors]. *Chapters*, 49–102.

Gillies, A., & Hume, M. (2022). *GDPO Working Paper Series*. https://www.swansea.ac.uk/media/Prime-Criminal-Real-Estate-Illicit-Economies-and-State-Power-in-Choc%C3%83%C2%B3-Colombia.pdf

Giommoni, L., & Gundur, R. V. (2018). An analysis of the United Kingdom's cannabis market using crowdsourced data. *Global Crime*, *19*(2), 85–106. https://doi.org/10.1080/17440572.2018.1460071

Gobierno de la Republica de Colombia. (2016, November 12). *Acuerdo Final para la Terminacion del Conflicto y la Construccion de una Paz Estable y Duradera* [Final agreement for the termination of the conflict and the construction of a stable and lasting peace]. https://www.jep.gov.co:443/Normativa/Paginas/Acuerdo-Final.aspx

Godoy, J. F. (2018). Violencia en la producción de cocaína: Laboratorios y grupos armados [Violence in cocaine production: Laboratories and armed groups]. *Documentos CEDE*, Article 016351. https://ideas.repec.org//p/col/000089/016351.html

Goldman, E. (2013). *The Computer Fraud and Abuse Act Is a Failed Experiment*. Forbes. https://www.forbes.com/sites/ericgoldman/2013/03/28/the-computer-fraud-and-abuse-act-is-a-failed-experiment/

González Peña, A., & Dorussen, H. (2021). The reintegration of ex-combatants and post-conflict violence. An analysis of municipal crime levels in Colombia. *Conflict Management and Peace Science*, *38*(3), 316–337. https://doi.org/10.1177/0738894219894701

Gonzalez, J., Sierra, J., & Fajardo-Toro, C. (2019). The prevention of the risk of terrorism in shopping centers: Towards public security in Colombia for the protection of critical infrastructures. *RISTI – Revista Iberica de Sistemas e Tecnologias de Informacao*, *2019*.

Google Developers. (2021, March 12). *What are extensions?* Chrome Developers. https://developer.chrome.com/docs/extensions/mv3/overview/

Grassi, P., & Sánchez-García, J. (2021). "Della coca, la piazza, gli spari." Rap e dinamiche di (ri)territorializzazione: Immaginari, pratiche ed economie transnazionali ["Of cocaine, the square, and gunshots." Rap and (re)territorialization dynamics: Transnational imaginaries, practices, and economies]. *Tracce Urbane. Rivista Italiana Transdisciplinare Di Studi Urbani*, *N. 10*, pratiche ed economie transnazionali. https://doi.org/10.13133/2532-6562/17738

Guerrero Castro, J. (2017). Estudios de ignorancia, inteligencia y la guerra contra las drogas en Colombia/ Ignorance studies, intelligence and the war on drugs in Colombia. *URVIO. Revista Latinoamericana de Estudios de Seguridad*, *21*, Article 21. https://doi.org/10.17141/urvio.21.2017.2943

Guzmán, J. P. S., & Sánchez, N. T. (2021). Desarrollo social, factor de mitigación de la criminalidad en el Pacífico colombiano [Social development as a factor in mitigating crime in the Colombian Pacific]. *Pensamiento Americano*, *14*(27), Article 27. https://doi.org/10.21803/penamer.14.27.351

Hinestroza, B. C., Sánchez, H. B., Aidar, T., & Palloni, A. (2021). Estimación de la tasa de mortalidad en contexto de altos homicidios: Caso de la región Pacífica colombiana 1993–2013 [Estimation of the mortality rate in the context of high homicides: The case of the Colombian Pacific region 1993–2013]. *Revista Latinoamericana de Población*, *15*(29), Article 29. https://doi.org/10.31406/relap2021.v15.i2.n29.5

Hirschey, J. K. (2014). Symbiotic relationships: Pragmatic acceptance of data scraping. *Berkeley Technology Law Journal*, *29*, 897–927.

Holmes, J. S., Gutiérrez de Piñeres, S. A., Curtin, K. M., & Gutiérrez de Piñeres, S. A. (2009). *Guns, Drugs, and Development in Colombia*. University of Texas Press. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=3443354

Isacson, A. (2018, April 29). *Coca in Colombia: What are the options?* Adam Isacson. https://adamisacson.com/coca-in-colombia-what-are-the-options/

Ives, B., & Krotov, V. (2006). Anything you search can be used against you in a court of law: Data mining in search archives. *Communications of the Association for Information Systems*, *18*, 29. https://doi.org/10.17705/1CAIS.01829

Jaiswal, M., & Patel, D. (2015). Data mining techniques and KDD. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR), E-ISSN* (2015): 2348-1269 .

Jeynes, W. H. (2022). A meta-analysis of the relationship between cannabis, opiates, cocaine, heroin, or other illegal drug use and student academic and behavioral outcomes. *Education and Urban Society*, *54*(6), 656–694. https://doi.org/10.1177/00131245211004571

JMP Statistical Discovery LLC. (2023a). *Compare Multiple Variables Using Graph Builder*. https://www.jmp.com/support/help/en/16.0/?os=win&source=application#page/jmp/compare-multiple-variables-using-graph-builder.shtml

JMP Statistical Discovery LLC. (2023b). *Element Types and Options*. https://www.jmp.com/support/help/en/16.0/?os=win&source=application#page/jmp/element-types-and-options.shtml#

JMP Statistical Discovery LLC. (2023c). *Predictive Analytics Software*. https://www.jmp.com/en_us/software/predictive-analytics-software.html

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test*, *13*(1), 1–43. https://doi.org/10.1007/BF02602999

Junjoewong, L., Sangnapachai, S., & Sunetnanta, T. (2018). ProCircle: A promotion platform using crowdsourcing and web data scraping technique. *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*, 1–5. https://doi.org/10.1109/ICT-ISPC.2018.8524003

Keller, Gerald. (2012). Statistics for management and economics. In *Statistics for management and economics* (9th ed.). South-Western Cengage Learning.

Krotov, V., & Silva, L. (2018). Legality and ethics of web scraping. Twenty-fourth Americas Conference on Information Systems. New Orleans.

Krotov, V., & Tennyson, M. (2018). Research note: Scraping financial data from the web using the R language. *Journal of Emerging Technologies in Accounting*, *15*(1), 169–181. https://doi.org/10.2308/jeta-52063

Landers, R. N., Cavanaugh, K. J., Brusso, R. C., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/met0000081

Li, J., Xu, Q., Shah, N., & Mackey, T. K. (2019). A machine learning approach for the detection and characterization of illicit drug dealers on Instagram: Model evaluation study. *Journal of Medical Internet Research*, *21*(6), e13803. https://doi.org/10.2196/13803

LLC, T. by C. E. (2022, March 10). Unprecedented" cocaine seizures in Europe from South America. *CE Noticias Financieras, English Ed.* https://www.proquest.com/docview/2638249998/citation/E7B836D9442A4AD8PQ/1

LLC, T. by C. E. (2023, January 29). Colombia records record cocaine seizures in 2022. *CE Noticias Financieras, English Ed.* https://www.proquest.com/docview/2770937366/citation/C91D1502E1AD4464PQ/1

LSE IDEAS, The London School of Economics and Political Science. (2014). *Ending the drug wars: Report of the LSE Expert Group on the economics of drug policy*. https://www.lse.ac.uk/ideas/Assets/Documents/reports/LSE-IDEAS-Ending-the-Drug-Wars.pdf

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Mackey, T. K., Li, J., Purushothaman, V., Nali, M., Shah, N., Bardier, C., Cai, M., & Liang, B. (2020). Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infoveillance Study on Twitter and Instagram. *JMIR Public Health and Surveillance*, *6*(3), e20794. https://doi.org/10.2196/20794

Man, N., Chrzanowska, A., Price, O., Bruno, R., Dietze, P. M., Sisson, S. A., Degenhardt, L., Salom, C., Morris, L., Farrell, M., & Peacock, A. (2021). Trends in cocaine use, markets and harms in Australia, 2003–2019. *Drug & Alcohol Review*, *40*(6), 946–956. https://doi.org/10.1111/dar.13252

Marín Llanes, L. (2022). The killing of social leaders: An unintended effect of Colombia's illicit crop substitution program. *International Journal of Drug Policy*, *101*, 103550. https://doi.org/10.1016/j.drugpo.2021.103550

Martens, B., Aguiar, L., GGmez, E., & Mueller-Langer, F. (2018). The digital transformation of news media and the rise of disinformation and fake news. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3164170

Matta Colorado, N. R. (2022a, June 28). *Dubai, el lujoso edén para quienes trafican la cocaína colombiana* [Dubai, the luxurious haven for those trafficking Colombian cocaine]. www.elcolombiano.com. https://www.elcolombiano.com/colombia/dubai-el-lujoso-eden-para-quienes-trafican-la-cocaina-colombiana-IO17912486

Matta Colorado, N. R. (2022b, September 26). *Cocaína colombiana coloniza el mercado negro de Oceanía* [Colombian cocaine colonizes the black market in Oceania]. www.elcolombiano.com. https://www.elcolombiano.com/colombia/cocaina-colombiana-coloniza-el-mercado-negro-de-oceania-LO18708535

Maududie, A., Retnani, W. E. Y., & Rohim, M. A. (2018). An approach of web scraping on news website based on regular expression. *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 203–207. https://doi.org/10.1109/EIConCIT.2018.8878550

Maybir, J., & Chapman, B. (2021). Web scraping of ecstasy user reports as a novel tool for detecting drug market trends. *Forensic Science International: Digital Investigation*, *37*, 301172. https://doi.org/10.1016/j.fsidi.2021.301172

McCarthy-Jones, A. (2018). Brokering transnational networks: Emerging connections between organised crime groups in the Pacific and Indian Oceans. *Journal of the Indian Ocean Region*, *14*(3), 343–353. https://doi.org/10.1080/19480881.2018.1519057

Meddeb, P., Ruseti, S., Dascalu, M., Terian, S.-M., & Travadel, S. (2022). Counteracting French fake news on climate change using language models. *Sustainability*, *14*(18), Article 18. https://doi.org/10.3390/su141811724

Medel, M., & Thoumi, F. (2014). Mexican Drug "Cartels." In L. Paoli (Ed.), *The Oxford Handbook of Organized Crime* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730445.013.013

Mejía, D., & Restrepo, P. (2008). The war on illegal drug production and trafficking: An economic evaluation of Plan Colombia. *Documentos CEDE*, Article 005123. https://ideas.repec.org//p/col/000089/005123.html

Mendels, G., Cooper, E., Soto, V., Hirschberg, J., Gales, M. J. F., Knill, K. M., Ragni, A., & Wang, H. (2015, September 6). *Improving speech recognition and keyword search for low resource languages using web data* [Proceedings Paper]. INTERSPEECH 2015 : 16th Annual Conference of the International Speech Communication Association; International Speech Communication Association (ISCA). https://www.isca-speech.org/archive/interspeech_2015/i15_0829.html

Milev, P. (2017). Conceptual Approach for Development of Web Scraping Application for Tracking Information. *Economic Alternatives*, 3, 475-485.

Millán-Quijano, J. (2020). Internal cocaine trafficking and armed violence in Colombia. *Economic Inquiry*, *58*(2), 624–641. https://doi.org/10.1111/ecin.12771

Ministerio de Defensa Nacional-Observatorio de Derechos Humanos y Defensa Nacional. (2023). *Homicidios en Colombia 2012–2022* [Homicides in Colombia 2012-2022] [dataset].

Ministerio de Tecnologías de la Información y las Comunicaciones. (2021). *Datos Abiertos Colombia* [Open Data Colombia]. https://www.datos.gov.co/

Ministerio del Interior de la Republica de Colombia. (2017). *Decreto 896 de 2017* [Decree 896 of 2017]. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=81878

Miron, J. A. (2001). Violence, guns, and drugs: A cross-country analysis. *The Journal of Law & Economics*, *44*(S2), 615–633. https://doi.org/10.1086/340507

Morales, W. Q. (2017). The political empowerment of the cocaleros of Bolivia and Peru—By Durand Ochoa, Ursula. *Bulletin of Latin American Research*, *36*(2), 273–275. https://doi.org/10.1111/blar.12629

Morselli, C., & Petit, K. (2007). Law-enforcement disruption of a drug importation network. *Global Crime*, *8*(2), 109–130. https://doi.org/10.1080/17440570701362208

Moyle, L., & Coomber, R. (2015). Earning a score: An exploration of the nature and roles of heroin and crack cocaine "user-dealers." *The British Journal of Criminology*, *55*(3), 534–555.

National Institute on Drug Abuse. (1986). *D.A.W.N.-Instruction Manual for Emergency Departments*. Bethesda.

Nestler, E. J. (2005). The neurobiology of cocaine addiction. *Science & Practice Perspectives*, *3*(1), 4–10. https://doi.org/10.1151/spp05314

Nutt, D., King, L. A., Saulsbury, W., & Blakemore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse. *The Lancet*, *369*(9566), 1047–1053. https://doi.org/10.1016/S0140-6736(07)60464-4

Octopus Data Inc. (2023). *Web Scraping Tool & Free Web Crawlers | Octoparse*. https://www.octoparse.com/

Ortega, M. (2022, August 3). Naval Campaign Orion, a Cornerstone in the Fight Against Narcotrafficking. *Diálogo Américas*. https://dialogo-americas.com/articles/naval-campaign-orion-a-cornerstone-in-the-fight-against-narcotrafficking/

Pande, S., Rathod, S., Joshi, R., Chvan, G., Jadhav, D., Phutane, P., Gonge, S., & Kada, K. (2022). Fake News Identification Using Regression Analysis and Web Scraping. *International Journal of Safety and Security Engineering*, *12*, 311–318. https://doi.org/10.18280/ijsse.120305

Paoli, L., & Vander Beken, T. (2014). Organized crime: A contested concept. In L. Paoli (Ed.), *The Oxford Handbook of Organized Crime* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730445.013.019

Paoli, L., Greenfield, V. A., & Zoutendijk, A. (2013). The harms of cocaine trafficking: Applying a new framework for assessment. *Journal of Drug Issues*, *43*(4), 407–436. https://doi.org/10.1177/0022042613475614

Pardo, R. (2000). Colombia's Two-Front War. *Foreign Affairs*, *79*(4), 64–73. https://doi.org/10.2307/20049809

Pfrimer, M., & Motta, A. L. (2021). Threats to national security from trafficking network: From urban environments to the transnational circulation of illicit drugs. *Revista de La Escuela Superior de Guerra Naval*, *27*, 149–168. https://doi.org/10.21544/1809-3791.v27n1.p161-180

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Rowman & Littlefield.

Phillips, E. M. (2015). How do kidnappers kill hostages? A comparison of terrorist and criminal groups. *Homicide Studies*, *19*(2), 123–148.

Pierce, M., Hayhurst, K., Bird, S. M., Hickman, M., Seddon, T., Dunn, G., & Millar, T. (2015). Quantifying crime associated with drug use among a large cohort of sanctioned offenders in England and Wales. *Drug and Alcohol Dependence*, *155*, 52–59. https://doi.org/10.1016/j.drugalcdep.2015.08.018

Policía Nacional de Colombia. (2015, December 24). *Política de privacidad y Uso, de la Policía Nacional* [Privacy and usage policy of the Colombian police]. Policía Nacional de Colombia. https://www.policia.gov.co/normatividad-politicas/privacidad-uso

Policía Nacional de Colombia. (2023a). *Policía Nacional: Noticias* [Colombian Police: News]. Policía Nacional de Colombia. https://www.policia.gov.co/noticias

Policía Nacional de Colombia. (2023b). *Transparencia y acceso a la información* [Transparency and access to information]. Policía Nacional de Colombia. https://www.policia.gov.co/transparencia-acceso-informacion

Polit Dueñas, G. (2013). *Narrating Narcos: Culiacán and Medellín*. University of Pittsburgh Press. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=2039425

Poulin, C., Fralick, P., Whynot, E. M., el-Guebaly, N., Kennedy, D., Bernstein, J., Boivin, D., & Rinehart, J. (1998). The Epidemiology of Cocaine and Opiate Abuse in Urban Canada. *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique*, *89*(4), 234–238.

Raubenheimer, J. E., & Barratt, M. J. (2018). Digital era drug surveillance: Quo vadis, Australia? *Drug and Alcohol Review*, *37*(6), 693–696. https://doi.org/10.1111/dar.12853

Research Scholar, School of Computing Science and Engineering, Galgotias University, Uttar Pradesh, India, Ponmaniraj*, S., Kumar, Prof. Dr. T., Doctorate, Professor, School of Computing Science and Engineering, Galgotias University, Uttar Pradesh, India, Goel, Prof. Dr. A. K., & Doctorate, Professor, School of Computing Science and Engineering, Galgotias University, Uttar Pradesh, India. (2020). Identification of website Reliability through Data Scrapping at Web Crawler's Navigation. *International Journal of Recent Technology and Engineering (IJRTE)*, *9*(1), 139–144. https://doi.org/10.35940/ijrte.A1463.059120

Reuter, P. (2014). Drug Markets and Organized Crime. In L. Paoli (Ed.), *The Oxford Handbook of Organized Crime* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730445.013.004

Richardson, L. (2022). *Beautiful Soup*. https://www.crummy.com/software/BeautifulSoup/

Rincón-Ruiz, A., & Kallis, G. (2013). Caught in the middle, Colombia's war on drugs and its effects on forest and people. *Geoforum*, *46*, 60–78. https://doi.org/10.1016/j.geoforum.2012.12.009

Salazar, L. G. S. (2010). Corredores y territorios estratégicos del conflicto armado colombiano: Una prioridad por territorializar en la geopolítica de los actores armados / Strategic corridors and territories in the colombian armed conflict: a priority for territorializing for. *Perspectiva Geográfica*, 9–36. https://doi.org/10.19053/01233769.1729

SARR, E. N., SALL, O., & DIALLO, A. (2018). FactExtract: Automatic Collection and Aggregation of Articles and Journalistic Factual Claims from Online Newspaper. *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 336–341. https://doi.org/10.1109/SNAMS.2018.8554421

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Personality*, *6*, Article 6. https://doi.org/10.1016/j.jrp.2013.05.009

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Semana. (2022, October 4). *Política antidrogas de Gustavo Petro: ¿Qué tan novedoso es el enfoque integral que pregona el mandatario?* [Anti-drug policy of Gustavo Petro: How novel is the comprehensive approach advocated by the president?]. https://www.semana.com/politica/articulo/politica-antidrogas-de-gustavo-petro-que-tan-novedoso-es-el-enfoque-integral-que-pregona-el-mandatario-analisis/202226/

Shabani, S., & Sokhn, M. (2018). Hybrid machine-crowd approach for fake news detection. *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, 299–306. https://doi.org/10.1109/CIC.2018.00048

Shah, N., Li, J., & Mackey, T. K. (2022). An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram. *Substance Abuse*, *43*(1), 273–277. https://doi.org/10.1080/08897077.2021.1941508

Shao, C., Hui, P.-M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PLOS ONE*, *13*(4), e0196087. https://doi.org/10.1371/journal.pone.0196087

Simanjuntak, M. L., Mansur, S., Saragih, N., Hayati, S., & Endri, E. (2022). The role of quality and trust on using website news. *International Journal of Data and Network Science*, *6*(3), 683–692. https://doi.org/10.5267/j.ijdns.2022.4.004

Singrodia, V., Mitra, A., & Paul, S. (2019). A review on web scrapping and its applications. *2019 International Conference on Computer Communication and Informatics (ICCCI)*, 1–6. https://doi.org/10.1109/ICCCI.2019.8821809

Sirisuriya, D. S. (2015). A Comparative Study on Web Scraping. *Proceedings of 8th International Research Conference, KDU* (pp. 135–140).

Slifker, J. F., & Shapiro, S. S. (1980). The Johnson System: Selection and Parameter Estimation. *Technometrics*, *22*(2), 239–246. https://doi.org/10.2307/1268463

Snell, J., & Menaldo, N. (2016). Web Scraping in an Era of Big Data 2.0. *Bloomberg Law News*.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections: Further analysis of the data by akaike' s. *Communications in Statistics – Theory and Methods*, *7*(1), 13–26. https://doi.org/10.1080/03610927808827599

Sundaramoorthy, K., Durga, R., & Nagadarshini, S. (2017). NewsOne—An Aggregation System for News Using Web Scraping Method. *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*, 136–140. https://doi.org/10.1109/ICTACC.2017.43

SyndiGate Media. (2022, June 28). Record cocaine seizures in West Africa. *Agence de Presse Africaine (APAnews), English Ed.* https://www.proquest.com/docview/2681503318/citation/B5874EBEE69F4FB7PQ/1

Tarapues, D. (2012). An Overview of the Terrorism in Colombia: Context, National Legislation and Anti-Terrorism Measures (SSRN Scholarly Paper 2435813). https://papers.ssrn.com/abstract=2435813

Terenghi, F. (2020). The financial management of cocaine trafficking in Italy. *European Journal of Criminology*, 1477370820980448. https://doi.org/10.1177/1477370820980448

Thoumi, F. (2014). Organized crime in Colombia: The actors running the illegal drug industry. In L. Paoli (Ed.), *The Oxford Handbook of Organized Crime* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730445.013.018

Toth, A. G., & Mitchell, O. (2018). A qualitative examination of the effects of international counter-drug interdictions. *International Journal of Drug Policy*, *55*, 70–76. https://doi.org/10.1016/j.drugpo.2018.02.012

U.S. Copyright Office. (1998). The Digital Millennium Copyright Act of 1998.

U.S. Department of State. (2021). *2.1 Drug production, cultivation, and eradication*. Tableau Public. https://public.tableau.com/app/profile/dcdd/viz/2_1DrugProductionCultivationandEradication/DrugProduction

United Nations Office on Drugs and Crime. (1997). *About UNODC*. United Nations: Office on Drugs and Crime. //www.unodc.org/unodc/en/about-unodc/index.html

United Nations Office on Drugs and Crime. (2008). *The threat of narco-trafficking in the Americas* (p. 44). https://www.unodc.org/documents/data-and-analysis/Studies/OAS_Study_2008.pdf

United Nations Office on Drugs and Crime. (2012). *Transnational Organized Crime in Central America and the Caribbean: A Threat Assessment*. United Nations Office on Drugs and Crime. https://doi.org/10.18356/493ae18b-en

United Nations Office on Drugs and Crime. (2021). *World Drug Report 2021*. United Nations : Office on Drugs and Crime. //www.unodc.org/unodc/en/data-and-analysis/wdr2021.html

United Nations Office on Drugs and Crime. (2022a). *Informes Programa Nacional de Sustitucion de Cultivos Ilicitos* [Reports from the National Program for the Substitution of Illicit Crops]. https://www.unodc.org/colombia/es/da2013/publicaciones.html

United Nations Office on Drugs and Crime. (2022b, October 20). *El cultivo de coca alcanzo niveles historicos en Colombia con 204.000 hectareas registradas en 2021* [The cultivation of coca reached historic levels in Colombia with 204,000 hectares registered in 2021]. https://www.unodc.org/colombia/es/el-cultivo-de-coca-alcanzo-niveles-historicos-en-colombia-con-204-000-hectareas-registradas-en-2021.html

United Nations Office on Drugs and Crime. (2023). *World Drug Report 2023*. United Nations : Office on Drugs and Crime. //www.unodc.org/unodc/en/data-and-analysis/world-drug-report-2023.html

UNODC-Integrated Illicit Crop Monitoring System [SIMCI]. (2022). *Colombia: Monitoring of territories affected by illicit crops 2021* (p. 169). https://www.biesimci.org/fileadmin/2022/documentos/survey_of_territories_affected_by_illegal_crops_2021_english.pdf

Ursin, M. (2014). "Crack ends it all?" A study of the interrelations between crack cocaine, social environments, social relations, crime, and homicide among poor, young men in urban Brazil. *Contemporary Drug Problems*, *41*(2), 171–199. https://doi.org/10.1177/009145091404100203

Vallerand, R. J., Deshaies, P., Cuerrier, J.-P., Pelletier, L. G., & Mongeau, C. (1992). Ajzen and Fishbein's theory of reasoned action as applied to moral behavior: A confirmatory analysis. *Journal of Personality and Social Psychology*, *62*(1), 98–109. https://doi.org/10.1037/0022-3514.62.1.98

Van Meter, H. J. (2020). Revising the DIKW pyramid and the real relationship between data, information, knowledge and wisdom. *Law, Technology and Humans*, *2*(2), 69–80. https://doi.org/10.5204/lthj.1470

Verisign. (2022). *Domain Name Industry Brief (DNIB)—Verisign*. https://www.verisign.com/en_US/domain-names/dnib/index.xhtml

VMware, Inc. (2023). *Windows VM | Workstation Pro*. VMware. https://www.vmware.com/products/workstation-pro.html

Wainwright, T. (2016). Narconomics: How to run a drug cartel. In *Narconomics: How to run a drug cartel* (First edition.). PublicAffairs.

Wehrey, F., & Boukhars, A. (Eds.). (2013). *Perilous Desert: Insecurity in the Sahara*. Brookings Institution Press. https://doi.org/10.2307/j.ctt6wpjcm

White, T. (Tom E.). (2015). Hadoop: The definitive guide. In *Hadoop: The definitive guide* (Fourth edition.). O'Reilly.

Wood, A. J., Mendelson, J. H., & Mello, N. K. (1996). Drug therapy: Management of cocaine abuse and dependence. *The New England Journal of Medicine*, *334*(15), 965–972.

Zapata, G. D. P. (2003). Terrorism in Colombia. *Prehospital and Disaster Medicine*, *18*(2), 80–87. https://doi.org/10.1017/S1049023X00000807

Zhao, B. (2017). Web Scraping. In *Encyclopedia of Big Data* (pp. 1–3). https://doi.org/10.1007/978-3-319-32001-4_483-1

Zollo, F., & Quattrociocchi, W. (2018). Misinformation spreading on Facebook. Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks, 177–196.

Zuleta, H. (2022). Coca, Cocaína y Violencia [Coca, Cocaine, and Violence]. *Libros En Línea*, Article 020415. https://ideas.repec.org//p/col/000490/020415.html

Zyte. (2022). Scrapy | A fast and powerful scraping and web crawling framework. https://scrapy.org/

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

DUDLEY KNOX LIBRARY

NAVAL POSTGRADUATE SCHOOL

WWW.NPS.EDU

WHERE SCIENCE MEETS THE ART OF WARFARE