Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Department of Mathematics and Physics

Master Program in Data Science

Forecasting and automating accurate levels of reserve for Contra Revenue, using artificial intelligence

---

THESIS to obtain the **GRADE** of

MASTER IN DATA SCIENCE

Presented by: **DAVID ARMANDO CISNEROS CHAVIRA**

Directed by: **DR. IVÁN ESTEBAN VILLALÓN TURRUBIATES**

Tlaquepaque, Jalisco. May, 2023.

Instituto Tecnológico
y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física

Maestría en Ciencia de Datos

El uso de la inteligencia artificial para mejorar la predicción de niveles de reserva de Contra revenue

---

TRABAJO RECEPCIONAL que para obtener el **GRADO** de **MAESTRO EN CIENCIA DE DATOS**

Presenta: **DAVID ARMANDO CISNEROS CHAVIRA**

Director: **DR. IVÁN ESTEBAN VILLALÓN TURRUBIATES**

Tlaquepaque, Jalisco. Mayo de 2023.

ACKNOWLEDGEMENTS

The author wishes to thank HP for all the support provided and the scholarship given, ITESO for taking me back once again after my engineering degree, Dr. Juan Carlos Martínez and Dr. Rocio Carrasco who were the most supportive along these 3 years studying, Dr. Ivan Esteban Villalón Turrubiates for stepping up as my thesis assessor and guide me through finalizing TOG and automatization program, and my wife Mariel Zavala for all the love and support provided during my busy homework nights, heavy classes schedules and impatient behavior when something wasn’t coming out as planned.

AGRADECIMIENTOS

El autor quiere agradecer a HP por todo el apoyo provisto y la beca otorgada, al ITESO por tomarme una vez más como alumno dentro de su casa de estudios, al Dr. Juan Carlos Martínez y la Dr. Rocío Carrasco por todo el apoyo a lo largo de estos 3 años de maestría, al Dr. Iván Esteban Villalón Turrubiates por apoyarme y tomarme como asesorado en este proceso de crear mi TOG y automatizar el programa final, y a mi esposa Mariel Zavala por todo el amor y el apoyo provisto, principalmente durante las largas noches de tarea y los horarios de clases, así como por el mal comportamiento cuando las cosas no estaban saliendo conforme a lo planeado.

DEDICATION

The author dedicates this thesis to his mother, who is no longer with us, and would have been the proudest.

DEDICATORIA

El autor dedica esta tesis a su madre, que ya no está con nosotros.

SUMMARY

The process of balance sheet forecasting from the Contra revenue team, within HP Inc, is currently a manual process that takes a lot of manual work to make, currently only two forecasts are provided, one per business, and the scope needs to be expanded to the market level organizations in the company.

The main objectives are to improve the forecast, to expand the scope and to automatize the process. The solution is a friendly executable that uses Python and an ARIMA algorithm to provide the next value of reserve entering an Excel file with organized data, that goes through the algorithm, and exiting an Excel file for the analyst to keep working on their own deliverable.

RESUMEN

El proceso de predicción de balance de Contra Revenue en HP Inc. es un proceso manual que requiere de muchas horas de trabajo y actualmente solo se proveen dos predicciones, una por negocio, y el enfoque se necesita expandirse al nivel de mercado de la compañía.

Los objetivos principales del proyecto son mejorar el estimado de la predicción, expandir el enfoque al nivel mercado, y automatizar el proceso. La solución fue crear un ejecutable amigable que utiliza Python y un algoritmo ARIMA para proveer el siguiente valor de la reserva, el ejecutable tiene como entrada un archivo de Excel, que pasa por el algoritmo y tiene como salida también un archivo de Excel con los resultados para que el analista pueda seguir trabajando en su entregable.

TABLE OF CONTENTS

MAESTRÍA EN CIENCIA DE DATOS **2**
1. INTRODUCTION **13**
1.1 BACKGROUND 14
1.2 JUSTIFICATION 17
1.3 PROBLEM 18
1.4 HYPOTHESIS 23
1.5 OBJECTIVES 23
1.5.1 General objective: 23
1.5.2 Specific objectives 23
1.6 SCIENTIFIC OR TECHNOLOGICAL NOVELTY OR CONTRIBUTION 25
2. STATE OF THE ART **26**
2.1 CURRENT APPLICATIONS OF AI IN FINANCE [3] [4] [5] 27
3. THEORETICAL FRAMEWORK **29**
3.1 ARIMA [6] 30
3.2 CASH FLOW [7] 30
3.3 BALANCE SHEET 32
3.4 PROFIT AND LOSS 32
3.5 ARMA [11] 32
3.6 GARCH [11] 32
3.7 DIFFERENCES BETWEEN ARIMA, ARMA AND GARCH 33
4. METHODOLOGY DEVELOPMENT **34**
4.1 REQUIREMENTS 35
4.2 RACI MATRIX 35
5. RESULTS AND DISCUSSION **36**
5.1 RESULTS 37
5.2 DISCUSSION 41
6. CONCLUSIONS **42**
6.1 CONCLUSIONS 43
6.2 FUTURE WORK 43
BIBLIOGRAPHY **45**

LIST OF FIGURES

Figure 1. HP Go to market route from HP to the final customer 15
Figure 2. HP P&L statement sample (Financial Annual Report: https://investor.hp.com/financials/annual-reports-and-proxies/default.aspx) 16
Figure 3. HP Balance sheet statement sample. Financial Annual Report (https://investor.hp.com/financials/annual-reports-and-proxies/default.aspx) 17
Figure 4. Manually stored data example 19
Figure 5. Tabs in data storage.. 19
Figure 6. Linearity data history showing the columns for the data forecasted and the rows the month when it was forecasted 20
Figure 7. Rate estimation data- the table on the bottom right shows the MIN, MAX, AVE example... 20
Figure 8. BS rates data showing the example same as figure 7 for MIN, MAX, AVE........ 21
Figure 9. Payment profile data showing the profile in columns and the historical months on the rows 21
Figure 10. Linear regression for payments showing the algorithm to calculate the payment value 22
Figure 11. Flash data example showing the P&L description on the rows and the time in the columns.. 23
Figure 12. Clustering in Python result.. 37
Figure 13. ARIMA decomposition showing a clear seasonality of 3 months in the third graph 38
Figure 14. ARIMA forecasting result... 38
Figure 15. Revenue decomposition showing a clear season every 3 months in the third graph 39
Figure 16. New model input showing what needs to be entered... 39
Figure 17. Organized date for input showing the table example of the data that needs to be inputed 40
Figure 18. Filter selection on executable.. 40
Figure 19. Output of executable model 41

LIST OF TABLES

Table 1. Accuracy table showing the actual accuracy data 24

Acronyms and abbreviations

AR		Auto regressive
ARIMA		Auto Regressive Integrated Moving Average
BS		Balance sheet
CI		Channel Inventory
HP		Hewlett Packard
MA		Moving average
ARMA		Auto Regressive Moving Average
GARCH		Generalized Autoregressive Conditional Heteroskedasticity
PL or P&L		Profit and loss
WW		Worldwide
AI		Artificial Intelligence
HPQ		Hewlett Packard Company (Stock market symbol)
HPE		Hewlett Packard Enterprise
HPI		Hewlett Packard Inc
PS		Personal Systems
GRU		Gated Recurrent Unit
TCN		Temporal Convolutional Network

1. INTRODUCTION

The whole Finance Corporate world is managed trough Excel and Excel only, the usual analyst doesn't use the powerful tool that technology and AI (Artificial Intelligence) is providing. It is today the perfect time to take advantage of those tools and begin to change the corporate financial world.

Nothing can replace the judgment of a good analyst, yet. But the tools can help that same analyst save time, improve the outcome of their work, and align faster to the always-changing environment that corporations live perpetually in.

Literature is always telling that finance is hard to predict, yet there is a lot of artificial intelligence implemented in the stock market. ARIMA (Autoregressive Integrated Moving Average) has proven to be a good predictor in a stable environment with the right judge of the experienced analyst.

Combining Python and ARIMA we could automatize and improve the balance sheet forecasting deliverable for Contra Revenue in within HP (Hewlett Packard) Inc.

1.1 Background

Time series have been used to study several cases finance cases with ARIMA models, one thesis from New Jersey Institute of Technology called Time series forecasting with applications to Finance by Viswapriya Misra published in 2021 were they test several forecasting models to predict stock values.
Contra Revenue is the second biggest line of the Profit and loss statement of the company, it includes all the spend occurred during a certain period related in its vast majority (by definition) to sell incentives and product returns, which would naturally deduct money from the gross sells of the company.

Since its origins, HPQ (Hewlett Packard Company) had managed the Contra Revenue spent through the business division of the company, without a major control or scrutiny revisions to it, letting the business managers and directors decide what was better to offer for a certain deal or a certain customer, amount or percentages allowed to return by partner, etc.

During 2010 HPQ company's CEO Mark Hurd was involved in a set of sex and money fraud scandals. According to Michael Holston, HP's general counsel: "The investigation revealed numerous instances where the contractor received compensation and/or expense reimbursement where there was not a legitimate business purpose. And the investigation found numerous instances where inaccurate expense reports were submitted by Mark or on his behalf that intended to or had the effect of concealing Mark's personal relationship with the contractor." [1]

After this inconvenient company's event, and because the deviation of expenses occurred through a Contra Revenue account, that whole operation of discounting, sells incentives, returns and any other related account moved to create a new team under Controllership supervision, whit an improved scrutiny, audits, and several controls, and we can now say 12 years later, this strategy had worked better so far.
After the split of HPQ back in 2015 [2] into two new companies, HPE and HPI (Hewlett-Packard Enterprise and Inc respectively), the Contra revenue operations remain under the Controllership supervision at least for the HPI portion of the new company, that its main core business are the manufacturing and reselling of computers and printers at a world-wide level.
As it was mentioned the Contra revenue spend is high, second biggest line of the P&L statement, and the reason of this is because HPI does not have an own distribution channel, so most of the incentives are given as a backend incentive to the biggest partners that redistribute and resell HP's product into the market or the final customer.

Diagram of HP's Go to market:

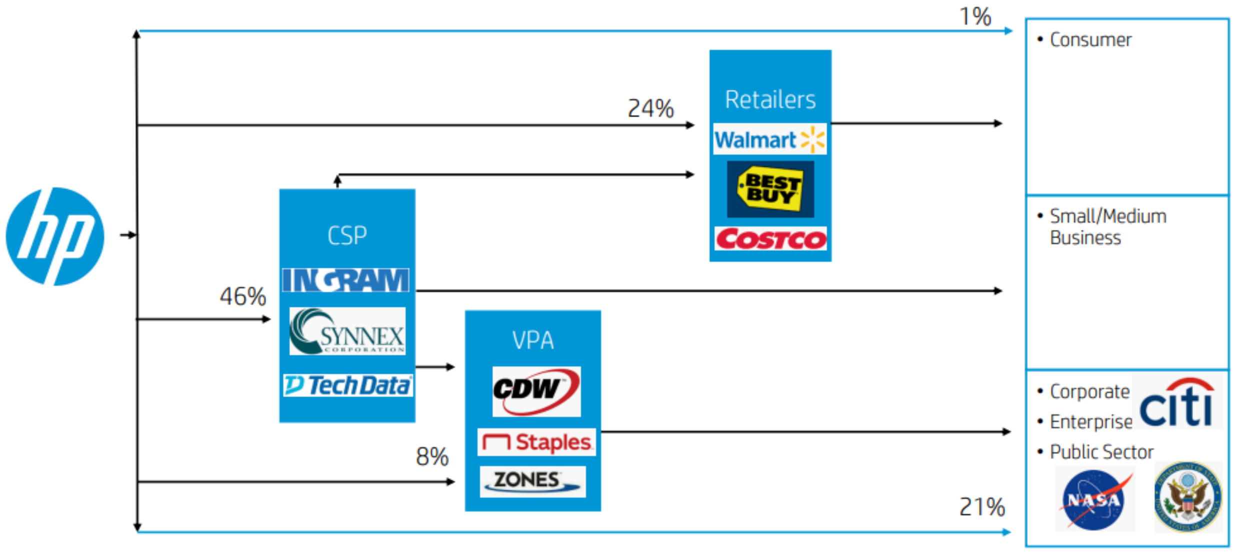


Figure 1. HP Go to market route from HP to the final customer

At the moment of the sell in, or from HP to the redistributor partners (as seen un figure above CSP and VPA partners) , and because of the US GAAP requirement, every spend associated to each sell needs to be recognized at the moment of the sell, so at the moment of the sell in we need to accrue for a certain percentage to recognize the spend in the P&L statement as well as to recognize the future obligation (debt) in the Balance sheet, this is the portion that needs to be forecasted from a Contra revenue team stand point.

On HP's accounting manual, contra revenue is:

Channel Sales incentives typically refers to programs used to stimulate sales, particularly sales of our channel partners which can include distributors, resellers, system integrators, retail partners and others.
Sales incentives are offers from HP that can be used by a customer to receive a reduction in the price of a product or service.

The only contra we accrue for is the backend portion, but every sell has a percentage of upfront discounting associated to it as well:

Upfront Contra – sales incentive that is offered and granted to the customer at the time of invoicing as part of the pricing of the products or services.

Backend Contra – that is offered and granted either the sale or after the initial sale of products and which need to be claimed subsequently by the customer by submitting the claim.
The main contra types are:

3103-Sales Goal Attainment Incentives
3104-Special Negotiated Discounts
3105-Promotions
3106-Price Protection
3107-Cooperative Partner Marketing
3109-End of Life
3113-Other discounts

The biggest ones are 3104, 3103 and 3105.
Location of Contra in the P&L statement:

It comes right before Net revenue (as seen in figure 2 below); it is not published because it is part of the company strategy.

HP Inc. and Subsidiaries

Consolidated Statements of Earnings

	2019	2018	2017
	FOR THE FISCAL YEARS ENDED OCTOBER 31		
	IN MILLIONS, EXCEPT PER SHARE AMOUNTS		
Net revenue	$58,756	$58,472	$52,056
Costs and expenses:			
Cost of revenue	47,586	47,803	42,478
Research and development	1,499	1,404	1,190
Selling, general and administrative	5,368	5,099	4,532
Restructuring and other charges	275	132	362
Acquisition-related charges	35	123	125
Amortization of intangible assets	116	80	1
Total costs and expenses	54,879	54,641	48,688
Earnings from operations	3,877	3,831	3,368
Interest and other, net	(1,354)	(818)	(92)
Earnings before taxes	2,523	3,013	3,276
Benefit from (provision for) taxes	629	2,314	(750)
Net earnings	$3,152	$5,327	$2,526
Net earnings per share:			
Basic	$2.08	$3.30	$1.50
Diluted	$2.07	$3.26	$1.48
Weighted-average shares used to compute net earnings per share:			
Basic	1,515	1,615	1,688
Diluted	1,524	1,634	1,702

Figure 2. HP P&L statement sample (Financial Annual Report: https://investor.hp.com/financials/annual-reports-and-proxies/default.aspx)

Location of Contra in the Balance sheet statements:

It is embedded in Other current liabilities (as seen in figure 3 below), it is not published as an individual line, same because it is part of the company's business strategy.

HP Inc. and Subsidiaries

Consolidated Balance Sheets

	AS OF OCTOBER 31 2019	AS OF OCTOBER 31 2018
	IN MILLIONS, EXCEPT PAR VALUE	
ASSETS		
Current assets:		
Cash and cash equivalents	$4,537	$5,166
Accounts receivable, net	6,031	5,113
Inventory	5,734	6,062
Other current assets	3,875	5,046
Total current assets	20,177	21,387
Property, plant and equipment, net	2,794	2,198
Goodwill	6,372	5,968
Other non-current assets	4,124	5,069
Total assets	$33,467	$34,622
LIABILITIES AND STOCKHOLDERS' DEFICIT		
Current liabilities:		
Notes payable and short-term borrowings	$357	$1,463
Accounts payable	14,793	14,816
Other current liabilities	10,143	8,852
Total current liabilities	25,293	25,131
Long-term debt	4,780	4,524
Other non-current liabilities	4,587	5,606
Commitments and contingencies		
Stockholders' deficit:		
Preferred stock, $0.01 par value (300 shares authorized; none issued)	—	—
Common stock, $0.01 par value (9,600 shares authorized; 1,458 and 1,560 shares issued and outstanding at October 31, 2019, and 2018 respectively)	15	16
Additional paid-in capital	835	663
Accumulated deficit	(818)	(473)
Accumulated other comprehensive loss	(1,225)	(845)
Total stockholders' deficit	(1,193)	(639)
Total liabilities and stockholders' deficit	$33,467	$34,622

Figure 3. HP Balance sheet statement sample. Financial Annual Report (https://investor.hp.com/financials/annual-reports-and-proxies/default.aspx)

1.2 Justification

The need to save time is always urgent in the corporate finance world, the need to have a malleable, adaptable one-fits-all model is a must.

Then the corporation just recently moved to a market level finance, which is a new organized geographical grouping system that consists into having 10 markets with 10 individual controllers and finance owned organizations.

Currently the balance sheet forecast was made only for 2 forecast series, PS (Personal systems or computers) and Print which are the main businesses of the company, at a WW (Worldwide) level. In within the 2 big groups there are also sub-divisions or categories, these needed to be also shown when forecasting, so the 2 original series turned into 10 and then times the market, turned into a hundred series. If the average analyst takes 2 working days to complete the 2 series, it would have taken 100 days to complete the full 100 needed series as per the new organization figures.

ARIMA has proven to work excellent with time series, and at the end the balance sheet is also a time series, so the exercise was to prove that right, once more.

1.3 Problem

The Balance sheet portion of the Contra reserves need to be forecasted at least 2 times every quarter, the requirement comes from the Financial Planning and Analysis (FP&A) team of the company and the final purpose of this forecast is for them to be able to forecast the cash flow of the company.

Currently we are facing three main issues with this process:

-Granularity: All the setup of contra rates forecast and actual revies of the company are being done at a market level, and this forecast's granularity today is only made at a global level, by business (computing and printing) meaning we have 2 main forecasts by account.

-Accuracy: The aimed required accuracy for out short-term forecast is for it to be below 5% error, currently we are sitting at a historical of 5.2%, with this change to market and artificial intelligence implementation, the idea is to improve accuracy for it to be below 3%.

-Automatization: The actual process is a massive excel sheet, semi-automated with a lot of manual inputs to it, the goal would be to automate it, reduce manual inputs and increase its autonomy.

Descriptive data analysis (actual)

In many large multinational companies such as HPI, there is often data issues, due to incomplete prior projects, multiple unaligned systems, and pretty much because there are lots on lots of data everywhere; even a major director described us (HP employees) as data junkies once because of the vast data and maybe not enough analysis.

Providing this context, we can now say that some of the process to complete the data needed for the Contra Balance sheet forecast is treated manually before it gets processed.

There are two main sources of information, the actual RP of the company that reports the Balance sheet, and there is the P&L source that tracks the monthly impacts and revenue.

Internally Contra Revenue team is divided into three sub-teams being:

-Delivery: Pretty much in charge of Mont End Close activities, hard core processes, interaction with all stakeholders, the face of contra to the outside.

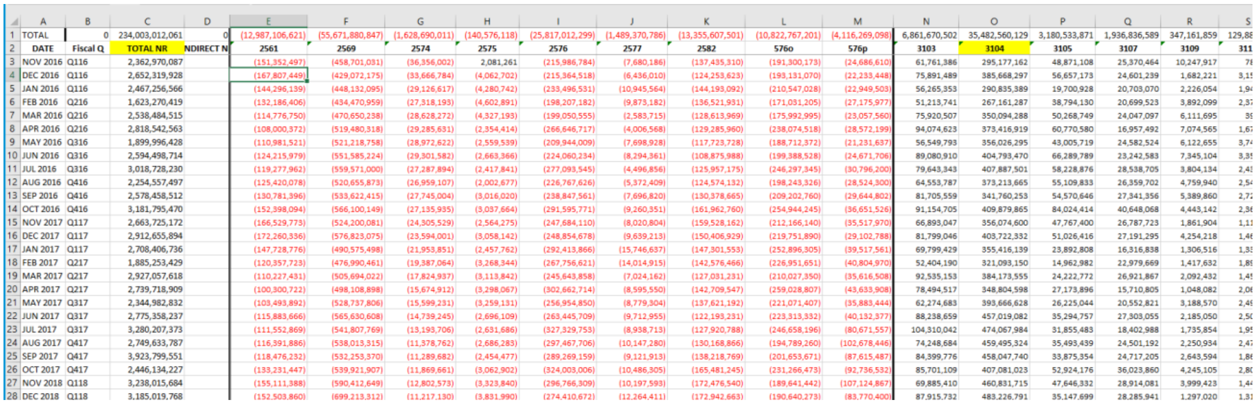
-Design: Internal management of processes, looking for improvements, expert owners of reporting, master data, definitions, documentation.

-Analytics: We call it that, but it is pretty much the dashboard's team, the maintain and constantly create new views adapting to business requirements, a lot of power BI and ad-hoc reporting.

Balance sheet and P&L data

Delivery oversees the Balance sheet forecast, and the Analytics team is the one that provides the raw reporting data to begin with the process.

As for now, as explained in the introduction, the forecast is made at a WW level, by business by account, and that is how data is manually stored today (showing example in figure 4):



	A	B	C	D	E	F	G	H	I	J
1	TOTAL	0	234,003,012,061	0	(12,987,106,621)	(55,671,880,847)	(1,628,690,011)	(140,576,118)	(25,817,012,299)	(1,489,370,786)
2	DATE	Fiscal Q	TOTAL NR	NDIRECT N	2561	2569	2574	2575	2576	2577
3	NOV 2016	Q116	2,362,970,087		(151,352,497)	(458,701,031)	(36,356,002)	2,081,261	(215,986,784)	(7,680,186)
4	DEC 2016	Q116	2,652,319,928		(167,807,449)	(429,072,175)	(33,666,784)	(4,062,702)	(215,364,518)	(6,436,010)
5	JAN 2016	Q116	2,467,256,566		(144,296,139)	(448,132,095)	(29,126,617)	(4,280,742)	(233,496,531)	(10,945,564)
6	FEB 2016	Q216	1,623,270,419		(132,186,406)	(434,470,959)	(27,318,193)	(4,602,891)	(198,207,182)	(9,873,182)
7	MAR 2016	Q216	2,538,484,515		(114,776,750)	(470,650,238)	(28,628,272)	(4,327,193)	(199,050,555)	(2,583,715)
8	APR 2016	Q216	2,818,542,563		(108,000,372)	(519,480,318)	(29,285,631)	(2,354,414)	(266,646,717)	(4,006,568)
9	MAY 2016	Q316	1,899,996,428		(110,981,521)	(521,218,758)	(28,972,622)	(2,559,539)	(209,944,009)	(7,698,928)
10	JUN 2016	Q316	2,594,498,714		(124,215,979)	(551,585,224)	(29,301,582)	(2,663,366)	(224,060,234)	(8,294,361)
11	JUL 2016	Q316	3,018,728,230		(119,277,962)	(559,571,000)	(27,287,894)	(2,417,841)	(277,093,545)	(4,496,856)
12	AUG 2016	Q416	2,254,557,497		(125,420,078)	(520,655,873)	(26,959,107)	(2,002,677)	(226,767,626)	(5,372,409)
13	SEP 2016	Q416	2,578,458,512		(130,781,396)	(533,622,415)	(27,745,004)	(3,016,020)	(238,847,561)	(7,696,820)
14	OCT 2016	Q416	3,181,795,470		(152,398,094)	(566,100,149)	(27,135,935)	(3,037,664)	(291,595,771)	(9,260,351)
15	NOV 2017	Q117	2,663,725,172		(166,529,773)	(524,200,081)	(24,305,529)	(2,564,275)	(247,684,110)	(8,020,804)
16	DEC 2017	Q117	2,912,655,894		(172,260,336)	(576,823,075)	(23,594,001)	(3,058,142)	(248,854,678)	(9,639,213)
17	JAN 2017	Q117	2,708,406,736		(147,728,776)	(490,575,498)	(21,953,851)	(2,457,762)	(292,413,866)	(15,746,637)
18	FEB 2017	Q217	1,885,253,429		(120,357,723)	(476,990,461)	(19,387,064)	(3,268,344)	(267,756,621)	(14,014,915)
19	MAR 2017	Q217	2,927,057,618		(110,227,431)	(505,694,022)	(17,824,937)	(3,113,842)	(245,643,858)	(7,024,162)
20	APR 2017	Q217	2,739,718,909		(100,300,722)	(498,108,898)	(15,674,912)	(3,298,067)	(302,662,714)	(8,595,550)
21	MAY 2017	Q317	2,344,982,832		(103,493,892)	(528,737,806)	(15,599,231)	(3,259,131)	(256,954,850)	(8,779,304)
22	JUN 2017	Q317	2,775,358,237		(115,883,666)	(565,630,608)	(14,739,245)	(2,696,109)	(263,445,709)	(9,712,955)
23	JUL 2017	Q317	3,280,207,373		(111,552,869)	(541,807,769)	(13,193,706)	(2,631,686)	(327,329,753)	(8,938,713)
24	AUG 2017	Q417	2,749,633,787		(116,391,886)	(538,013,315)	(11,378,762)	(2,686,283)	(297,467,706)	(10,147,280)
25	SEP 2017	Q417	3,923,799,551		(118,476,232)	(532,253,370)	(11,289,682)	(2,454,477)	(289,269,159)	(9,121,913)
26	OCT 2017	Q417	2,446,134,227		(133,231,447)	(539,921,907)	(11,869,661)	(3,062,902)	(324,003,006)	(10,486,305)
27	NOV 2018	Q118	3,238,015,684		(155,111,388)	(590,412,649)	(12,802,573)	(3,323,840)	(296,766,309)	(10,197,593)
28	DEC 2018	Q118	3,185,019,768		(152,503,860)	(699,213,312)	(11,217,130)	(3,831,990)	(274,410,672)	(12,264,411)

	K	L	M	N	O	P	Q	R	S
1	(13,355,607,501)	(10,822,767,201)	(4,116,269,098)	6,861,670,502	35,482,560,129	3,180,533,871	1,936,836,589	347,161,859	129,88
2	2582	576o	576p	3103	3104	3105	3107	3109	311
3	(137,435,310)	(191,300,173)	(24,686,610)	61,761,386	295,177,162	48,871,108	25,370,464	10,247,917	78
4	(124,253,623)	(193,131,070)	(22,233,448)	75,891,489	385,668,297	56,657,173	24,601,239	1,682,221	3,15
5	(144,193,092)	(210,547,028)	(22,949,503)	56,265,353	290,835,389	19,700,928	20,703,070	2,226,054	1,94
6	(136,521,931)	(171,031,205)	(27,175,977)	51,213,741	267,161,287	38,794,130	20,699,523	3,892,099	2,37
7	(128,613,969)	(175,992,995)	(23,057,560)	75,920,507	350,094,288	50,268,749	24,047,097	6,111,695	39
8	(129,285,960)	(238,074,518)	(28,572,199)	94,074,623	373,416,919	60,770,580	16,957,492	7,074,565	1,67
9	(117,723,728)	(188,712,372)	(21,231,637)	56,549,793	356,026,295	43,005,719	24,582,524	6,122,655	3,74
10	(108,875,988)	(199,388,528)	(24,671,706)	89,080,910	404,793,470	66,289,789	23,242,583	7,345,104	3,35
11	(125,957,175)	(246,297,345)	(30,796,200)	79,643,343	407,887,501	58,228,876	28,538,705	3,804,134	2,43
12	(124,574,132)	(198,243,326)	(28,524,300)	64,553,787	373,213,665	55,109,833	26,359,702	4,759,940	2,54
13	(130,378,665)	(209,202,760)	(29,644,802)	81,705,559	341,760,253	54,570,646	27,341,356	5,389,860	2,72
14	(161,962,760)	(254,944,245)	(36,651,526)	91,154,705	409,879,865	84,024,414	40,648,068	4,443,142	2,36
15	(159,528,162)	(212,166,140)	(35,517,970)	66,893,047	356,074,600	47,767,400	26,787,723	1,861,904	1,11
16	(150,406,929)	(219,751,890)	(29,102,788)	81,799,046	403,722,332	51,026,416	27,191,295	4,254,218	1,46
17	(147,301,553)	(252,896,305)	(39,517,561)	69,799,429	355,416,139	23,892,808	16,316,838	1,306,516	1,33
18	(142,576,466)	(226,951,651)	(40,804,970)	52,404,190	321,093,150	14,962,982	22,979,669	1,417,632	1,89
19	(127,031,231)	(210,027,350)	(35,616,508)	92,535,153	384,173,555	24,222,772	26,921,867	2,092,432	1,45
20	(142,709,547)	(259,028,807)	(43,633,908)	78,494,517	348,804,598	27,173,896	15,710,805	1,048,082	2,06
21	(137,621,192)	(221,071,407)	(35,883,444)	62,274,683	393,666,628	26,225,044	20,552,821	3,188,570	2,49
22	(122,193,231)	(223,313,332)	(40,132,377)	88,238,659	457,019,082	35,294,757	27,303,055	2,185,050	2,50
23	(127,920,788)	(246,658,196)	(80,671,557)	104,310,042	474,067,984	31,855,483	18,402,988	1,735,854	1,95
24	(130,168,866)	(194,789,260)	(102,678,446)	74,248,684	459,495,324	35,493,439	24,501,192	2,250,934	2,47
25	(138,218,769)	(201,653,671)	(87,615,487)	84,399,776	458,047,740	33,875,354	24,717,205	2,643,594	1,86
26	(165,481,245)	(231,266,473)	(92,736,532)	85,701,109	407,081,023	52,924,176	36,023,860	4,245,105	2,80
27	(172,476,540)	(189,641,442)	(107,124,867)	69,885,410	460,831,715	47,646,332	28,914,081	3,999,423	1,44
28	(172,942,663)	(190,640,273)	(83,770,400)	87,915,732	483,226,791	35,147,699	28,285,941	1,297,020	1,31

Figure 4. Manually stored data example

Figure 5. Tabs in data storage

*Two tabs containing the same data in the same order, one for PS and one for Print business (figure 5 above for reference).

Historical since Nov 2016 (Nov is the first month of HPI's fiscal year) by business, containing revenue data, and data by accounts, respectively BS and P&L.

We ca say the data is:

-Complete: we have no NULL values, there has always been either a reserve or an impact for every account by business by month.

-Valid: The data is pulled, populated, and validated by the Analytics team.

-Stored: Safely stored in a Teams group folder, and in One drive.

-Handy: Easy access to the storage folders.

-Reliable: Even when is manually processed to get to the view shown above, after adding it, there is a validation process that makes data reliable.

-Enough: With this data there is enough history and enough columns to process the forecast.

Linearity data

There is another set of really important data that we call linearity (figure 6 below), also stored by PS and Print, where we keep track of previous revenue estimations; on the lines we keep the month of the estimation, on the columns the month estimated, and how it varied along time, it also has the inter-quarter distribution of expected sales, and then when the data turns from flash to actuals, it gets updated too.

A	MM	MN	MO	MP	MQ	MR	MS	MT	MU	MV	MW	MX	MY	F
	MAY 2021	JUN 2021	JUL 2021	Q321	AUG 2021	SEP 2021	OCT 2021	Q421	FY21	NOV 2022	DEC 2022	JAN 2022	Q122	F
DATE	Q321M1	Q321M2	Q321M3	Q321	Q421M1	Q421M2	Q421M3	Q421	FY21	Q122M1	Q122M2	Q122M3	Q122	C
MAY 2021	3,132	3,559	4,254	10,945				11,400	43,503				10,560	
JUN 2021	2,733	3,728	4,356	10,817				11,391	43,366				10,560	
JUL 2021	2,733	3,417	4,675	10,825				11,270	43,253				10,560	
AUG 2021	2,733	3,418	4,255	10,406	3,110	3,209	4,163	10,481	42,045				10,560	
SEP 2021	2,733	3,418	4,255	10,406	3,255	3,358	4,357	10,970	42,534				10,969	
OCT 2021	2,733	3,418	4,255	10,406	3,255	3,908	4,229	11,392	42,956				11,701	
NOV 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359	3,672	4,054	3,974	11,701	
DEC 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359	3,885	4,204	4,161	12,249	
JAN 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359	3,885	4,290	4,046	12,221	
FEB 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359	3,885	4,290	4,021	12,196	
MAR 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359					
APR 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359	32%	35%	33%		
MAY 2022	2,733	3,418	4,255	10,406	3,255	3,908	4,631	11,795	43,359					

Figure 6. Linearity data history showing the columns for the data forecasted and the rows the month when it was forecasted

A useful set of data that has been also used to analyze variations between periods, example: if the revenue is higher in % at the end of one Quarter in comparison to another that could mean more BS accrued for in that specific period.

This data is provided directly from FP&A team, monthly and gets collected in the tabs as shown in the image.

Rates data

With all the data inputs from before we manually create our own set of data for rates, especially for the Model M, Model Q and Budget.

Historical spend rates by business by account, with an interaction table where you can set either MAX, MIN, or AVE (as seen in figure 7 below) depending on the expectations and set the time frame of this calculations, for a minimum of 1 year and no maximum.

	A	B	C	D	E	F	G	H	I	J	K	V
1	PS	RATES										
2	TOTAL				3103	3104	3105	3107	3109	3112	3113	
3	DATE	Mont	Year	Quart	3103	3104	3105	3107	3109	3112	3113	
4	NOV 2016	NOV	2016	Q1								
5	DEC 2016	DEC	2016	Q1								
6	JAN 2016	JAN	2016	Q1	2.59%	12.99%	1.67%	0.94%	0.19%	0.08%	0.30%	alignment
7	FEB 2016	FEB	2016	Q2								alignment
8	MAR 2016	MAR	2016	Q2								alignment
9	APR 2016	APR	2016	Q2	3.17%	14.19%	2.15%	0.88%	0.24%	0.06%	0.20%	alignment
10	MAY 2016	MAY	2016	Q3								
11	JUN 2016	JUN	2016	Q3								
12	JUL 2016	JUL	2016	Q3	3.00%	15.56%	2.23%	1.02%	0.23%	0.13%	0.14%	
13	AUG 2016	AUG	2016	Q4								
14	SEP 2016	SEP	2016	Q4								
15	OCT 2016	OCT	2016	Q4	2.96%	14.03%	2.42%	1.18%	0.18%	0.10%	0.20%	
16	NOV 2017	NOV	2017	Q1								
17	DEC 2017	DEC	2017	Q1								
18	JAN 2017	JAN	2017	Q1	2.64%	13.46%	1.48%	0.85%	0.09%	0.05%	0.20%	
19	FEB 2017	FEB	2017	Q2								
20	MAR 2017	MAR	2017	Q2								
21	APR 2017	APR	2017	Q2	2.96%	13.96%	0.88%	0.87%	0.06%	0.07%	0.12%	
22	MAY 2017	MAY	2017	Q3								
23	JUN 2017	JUN	2017	Q3								
24	JUL 2017	JUL	2017	Q3	3.03%	15.77%	1.11%	0.79%	0.08%	0.08%	0.14%	
25	AUG 2017	AUG	2017	Q4								
26	SEP 2017	SEP	2017	Q4								
27	OCT 2017	OCT	2017	Q4	2.68%	14.53%	1.34%	0.93%	0.10%	0.08%	0.15%	
28	NOV 2018	NOV	2018	Q1								
29	DEC 2018	DEC	2018	Q1								
30	JAN 2018	JAN	2018	Q1	2.75%	14.58%	1.22%	0.75%	0.07%	0.04%	0.23%	
31	FEB 2018	FEB	2018	Q2								

W	X	Y	Z	AA	AB	AC	AD	AE	
	4	5	6	7	8	10	11	13	
M/M	3105	3104	3103	3109	3112	3113	3107	TOTAL	
NOV	1.2%	13.8%	2.5%	0.4%	0.0%	0.0%	0.4%	18.3%	Q1
DEC	2.1%	14.9%	2.2%	0.5%	0.0%	0.0%	0.3%	20.0%	Q1
JAN	1.5%	13.3%	2.2%	0.4%	0.0%	0.0%	0.3%	17.8%	Q1
FEB	1.9%	16.4%	2.7%	0.5%	0.0%	0.0%	0.4%	21.89%	Q2
MAR	1.7%	13.8%	2.2%	0.5%	0.0%	0.0%	0.3%	18.48%	Q2
APR	1.6%	13.8%	2.2%	0.4%	0.0%	0.0%	0.3%	18.4%	Q2
MAY	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q3
JUN	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q3
JUL	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q3
AUG	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q4
SEP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q4
OCT	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Q4
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		

	3105	3104	3103	3109	3112	3113	3107
NOV	MAX	MAX	MAX	MAX	MAX	MAX	MAX
DEC	MAX	MAX	MAX	MAX	MAX	MAX	MAX
JAN	AVG	AVG	AVG	AVG	AVG	AVG	AVG
FEB	MAX	MAX	MAX	MAX	MAX	MAX	MAX
MAR	MAX	MAX	MAX	MAX	MAX	MAX	MAX
APR	MAX	MAX	MAX	MAX	MAX	MAX	MAX
MAY	MAX	MAX	MAX	MAX	MAX	MAX	MAX
JUN	MAX	MAX	MAX	MAX	MAX	MAX	MAX
JUL	MAX	MAX	MAX	MAX	MAX	MAX	MAX
AUG	MAX	MAX	MAX	MAX	MAX	MAX	MAX
SEP	MAX	MAX	MAX	MAX	MAX	MAX	MAX
OCT	MAX	MAX	MAX	MAX	MAX	MAX	MAX

	3105	3104	3103	3109	3112	3113	3107

Figure 7. Rate estimation data- the table on the bottom right shows the MIN, MAX, AVE example

Also, a set of rates is calculated for the historical BS position on respect to the Revenue of the period (figure 8):

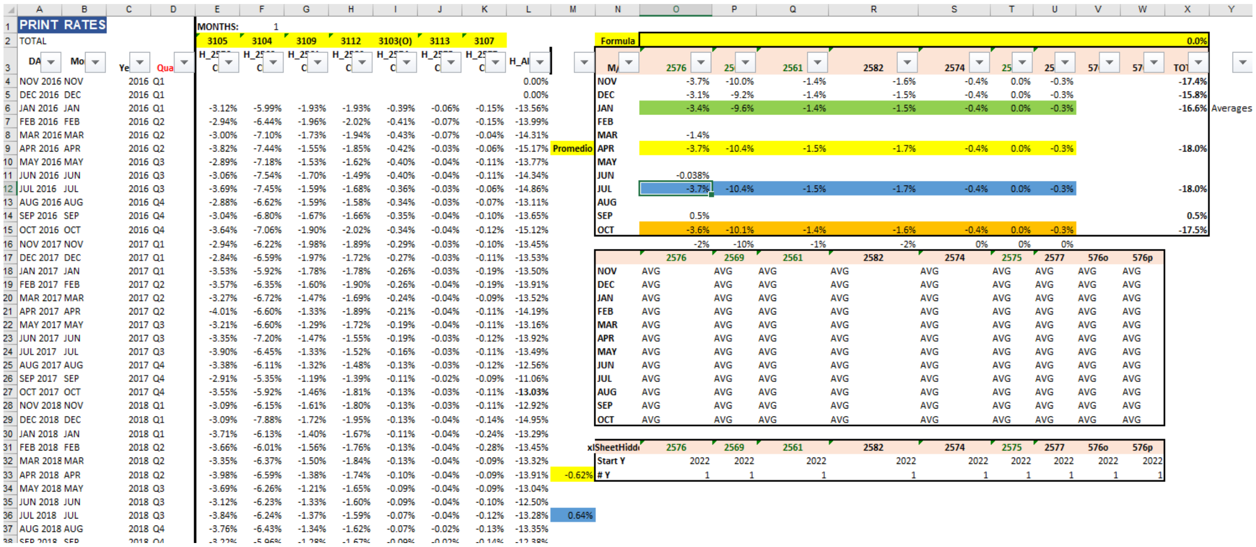


Figure 8. BS rates data showing the example same as figure 7 for MIN, MAX, AVE

Similar selection table to adapt to the forecast on every period.

Payment data

The last set of data is on the payments, there is no actual source to get this data in the company, because it would require a lot of consolidation, so we just drove it from the formula of the reserve:

New reserve = Previous reserve + (sell in x contra rate by month) – payments

Therefore:

- New reserve + Previous reserve + (sell in x contra rate by month) = payments

This is how this data set is calculated and created, and there are two methodologies to forecast payments:

-Payment profile: There is a report that only contains two accounts in it (3104 and 3105) where you can identify the period to which the payment pertains to, and therefore you can calculate the future payments by projecting this payment profile distribution, good news is that the two accounts with this data are the largest, but unfortunately there is not available for all of them.

Showing an example of a payment profile payment forecast (figure 9):

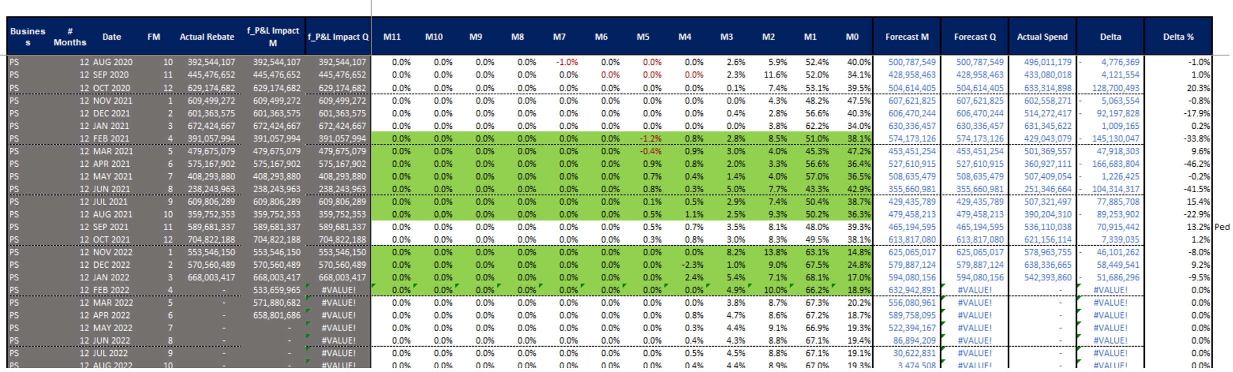


Figure 9. Payment profile data showing the profile in columns and the historical months on the rows

-Linear regression: For the rest of the accounts where we don't have any basis to predict payments, we base our forecast on a linear regression and historical created values.

Showing an example (figure 10):

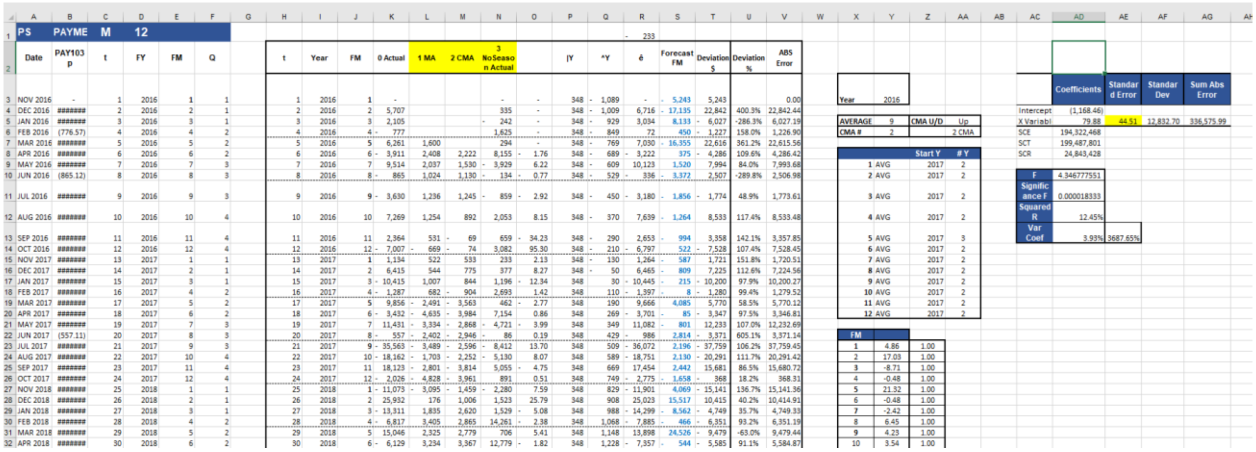


Figure 10. Linear regression for payments showing the algorithm to calculate the payment value

It is a basic linear regression, with moving averages to normalize them, tested to be significative with the period took as historical, with a seasonality factor added to it by month by year, it has shown good accuracy when forecasting in the short term (1-3 months).

We have described the source of the actual data needed to run the forecast process. And now we will concentrate on the alternatives of sources for the new data to supply the objectives of this project, which would be two options:

Data for the project

Source data by market

The data is available by market with the same splits for the project in the same system where the Analytics team goes and pull the actual data used for the forecast process. But there are some inconveniences with the completeness of the data as there are some months where in small markets there aren't any spend, not precisely an inconvenience as the data is what the data is.

The date would work the same from the source, manual (now automated) process to consolidate it and have enough history to start projecting with the new tools.

Flash data

There is the availability of this flash information by market where we could get the spend rate, and the forecasted revenue at the same time:

Input Cells

Greater China

Note: FY'22 Outlook financials pre-seeded with data fror

HPS Printers	Q120		Q220		Q320		Q420		FY20	
	M$	%	M$	%	M$	%	M$	%	M$	%
$M										
Gross Trade Revenue	295	100.0%	286	100.0%	326	100.0%	345	100.0%	1,253	100.0%
Contractual Discount	87	29.6%	85	29.6%	97	29.7%	103	29.8%	372	29.7%
NDP	208	70.4%	202	70.4%	229	70.3%	242	70.2%	881	70.3%
Special Negotiated Discounts	19	6.5%	17	5.9%	17	5.3%	19	5.4%	72	5.8%
Promotions	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Product Returns	0	0.0%	(0)	0.0%	(0)	-0.1%	0	0.0%	(0)	0.0%
Price Protection	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Partner Programs (Compensation/MD	11	3.7%	11	4.0%	11	3.3%	9	2.6%	42	3.4%
Other Discounts	3	0.9%	3	1.0%	3	0.9%	3	0.9%	12	0.9%
Discretionary Discounts	33	11.1%	31	10.9%	31	9.4%	31	9.0%	126	10.0%
Total Discounts	120	40.7%	116	40.6%	127	39.1%	134	38.8%	498	39.7%
Net Currency	4	1.4%	1	0.2%	0	0.1%	(4)	-1.0%	1	0.1%
Net Revenue	179	60.6%	171	59.6%	199	61.1%	208	60.2%	756	60.4%
Owned Gross Margin	23	13.0%	23	13.2%	30	15.2%	37	17.8%	113	15.0%
Prior period adjustments	0		0		0		0		0	
Non recurrent Investments									0	
Net adjustments	0		0		0		0		0	
Adjustments re-allocated to pertainir	0		(1)		(1)		1		(2)	
Normalized Discretionary Discounts	33	11.1%	30	10.5%	30	9.2%	32	9.2%	124	9.9%
Normalized Gross Margin	23	13.0%	24	13.9%	31	0.155837	36	17.6%	115	15.1%
Weeks of stock [in weeks]	2.1		1.7		1.4		1.3			
Channel Inventory $	26		21		19		19			
Sell-thru $	159		167		177		185			

HPS Printers	Q121		Q221		Q321		Q421		FY21		Q122		Q222	
	M$	%	M$	%	M$	%	M$	%	M$	%	M$	%	M$	%
$M														
Gross Trade Revenue	387	100.0%	301	100.0%	222	100.0%	322	100.0%	1,232	100.0%	298	100.0%	357	100.0%
Contractual Discount	115	29.7%	89	29.7%	65	29.4%	96	29.7%	365	29.7%	87	29.3%	106	29.7%
NDP	272	70.3%	211	70.3%	157	70.6%	226	70.3%	866	70.3%	210	70.7%	251	70.3%
Special Negotiated Discounts	18	4.8%	13	4.4%	10	4.4%	13	4.0%	54	4.4%	14	4.7%		0.0%
Promotions	0	0.0%	0	0.0%	0	0.0%	(0)	0.0%	0	0.0%	0	0.0%		0.0%
Product Returns	(0)	-0.1%	1	0.3%	0	0.0%	0	0.1%	1	0.1%	0	0.1%		0.0%
Price Protection	0	0.0%	0	0.0%	1	0.3%	0	0.0%	1	0.1%	0	0.0%		0.0%
Partner Programs (Compensation/MD	10	2.6%	7	2.4%	5	2.4%	10	3.0%	32	2.6%	8	2.7%		0.0%
Other Discounts	4	1.0%	3	1.0%	2	0.9%	3	1.0%	12	1.0%	4	1.3%	33	9.4%
Discretionary Discounts	31.814	8.2%	24	8.1%	18	8.0%	25.984	8.1%	100	8.1%	26	8.7%	33	9.4%
Total Discounts	147	37.9%	114	37.9%	83	37.5%	122	37.8%	465	37.8%	113	38.0%	140	39.0%
Net Currency	(10)	-2.5%	(9)	-3.0%	(7)	-3.1%	(3)	-1.0%	(29)	-2.3%	(3)	-1.0%	(4)	-1.1%
Net Revenue	230	59.5%	178	59.2%	132	59.4%	197	61.2%	737	59.9%	181	60.9%	214	59.8%
Owned Gross Margin	57	25.0%	27	14.9%	13	10.0%	44	22.2%	141	19.1%	43	23.7%	54	25.1%
Prior period adjustments	(1)		(0)		(0)		0		(1)		0		0	
Non recurrent Investments									0					
Net adjustments	(1)		(0)		(0)		0		(1)		0		0	
Adjustments re-allocated to pertainir	(1)		(0)		1		0		0		0		0	
Normalized Discretionary Discounts	32	8.4%	24	8.1%	19	8.7%	26	8.0%	102	8.3%	26	8.7%	33	9.4%
Normalized Gross Margin	57	24.8%	27	15.0%	12	9.0%	44	22.3%	139	18.9%	43	23.7%	54	25.1%
Weeks of stock [in weeks]	1.1		2.0		3.4		2.0				2.2		0.0	
Channel Inventory $	17		24		32		29				30		0	
Sell-thru $	214		160		122		187				175		0	

Figure 11. Flash data example showing the P&L description on the rows and the time in the columns

The idea with these files would be to consolidate them in a SharePoint, and then just access the necessary data by the granularity desired, the source is directly from market and business representatives, and it is only part of the flash process for margin (figure 11 shows the raw flash data example).

1.4 Hypothesis

The usage of ARIMA and Python will help to automate, save time, and improve the forecast estimations.

1.5 Objectives

1.5.1 General objective:

Be practical, introduce the Python strength into the financial analysis, save money/time in the meantime, and provide a better estimate to the CFO office to have a TRUE cash flow every Quarter.

1.5.2 Specific objectives

1.Granularity

The current granularity, as explained before, is WW (1), by business (2), meaning we have 2 forecasts running at the same time in all the modeling.

Due to a recent change in how finance is structured, it is a must to go down to a market forecasting level, currently there are 10 markets, not all of them as big nor as important, an maybe we will end up having only 7-8 markets forecasted, but that is potentially going from 2 forecast to 20 individual forecast by 6 accounts that is a total of 120 time series, and if we add in the categories, we are looking at a 500+ time series.

There are two main options to approach this change:

-Tops down approach. Either keep considering the actual models, refine them and allocate reserves to the market behaviors according to history or most recent business/markets expectations based on sell in or any other factor that we find out relevant.

-Bottoms up approach. Build the reserve individually from scratch based on each individual market expectation of spend, sell in and/or payment, at the end consolidate these reserves to have a grand total WW reserve.

Both approaches would have both views, consolidated, and allocated by market, this will make much easier the analysis of variations, deltas, new accuracy by market, and it would be better to highlight anomalies in forecasting, or focus on a specific attention point, besides and among all the forecast process will be in the same language as all the new reviews by market level.

2.Accuracy

Currently the accuracy of the forecast is at a total level of 5.2% with good accuracy for the short term and a little bit worse on the mid-term deliveries as the table below shows:

Table 1. Accuracy table showing the actual accuracy data

Row Labels	Average of % accuracy
⊟ Budget	**5.7%**
Budget	5.8%
Model	3.7%
⊟ Flash	**5.0%**
Budget	6.1%
Model	2.8%
Grand Total	**5.2%**

Budget model is used for mid-term, Model is used for short term.

Under the grand scheme of accuracy parameters, we are under the green flag, but the idea of granularity, more than making it worse, with the AI planned to be involved in the future calculations we aim to be in the 3% range.

The error calculation was made at a total absolute level.

3.Automatization

The process is currently semi-automated in Excel, and as stated before it has a lot of manual inputs and supervision, so we always have to come back and give the number a sense of realness, from it was born the idea to introduce a certain AI checks that allow the model to run automatically and produce analytics itself, even learn from history and previous mistakes to not repeat them again and know what to look for, what to do, etc.

With this whole new market structure, there are new spreadsheets updated monthly feed with market expectations, from which the model could make several checks or even leverage from them.

The clustering analysis is another possibility, some of the markets or the accounts behave similar, and this will potentially reduce the number of forecasts needed to be provided.

The end state of the process is to have a flexible model or set of models that reflect the actual business/markets expectations, with low/or no supervision.

1.6 Scientific or technological novelty or contribution

The combination of ARIMA time series in finance is not new, but the ability for any person with a computer to run this program without even the necessity to have Python installed in your computer is a novelty. Besides the project is saving thousands of manual work hours per year.

2. STATE OF THE ART

The use of artificial intelligence (AI) in finance has been growing rapidly in recent years, driven by the need for faster and more accurate decision-making, improved risk management, and increased efficiency. AI technologies such as machine learning, natural language processing, and computer vision are being used to analyze large volumes of data, automate routine tasks, and provide personalized services to customers. This paper provides an overview of the state of the art of AI in finance, including current applications, challenges, and future directions. It also includes a review of relevant literature to support the discussion.

2.1 Current applications of AI in Finance [3] [4] [5]

- **Fraud Detection**: AI can be used to detect and prevent fraud in financial transactions by identifying patterns and anomalies that may indicate fraudulent activity. For example, machine learning algorithms can analyze large volumes of transaction data to identify unusual patterns in behavior, such as frequent transactions in odd amounts or at unusual times. Banks and financial institutions are increasingly using AI-powered fraud detection systems to prevent fraudulent activities, reduce losses, and protect customers' assets.

- **Trading and Investments**: AI is being used to analyze market trends, stock prices, and other financial data to make more accurate predictions about investments. Machine learning algorithms can be trained on historical data to identify patterns and relationships that can be used to make predictions about future market trends. These algorithms can also identify opportunities for high returns and provide recommendations on investments to clients. For instance, hedge funds and other investment firms are using AI-based trading systems to analyze market trends, automate trading, and optimize returns.

- **Risk Management**: AI can help assess risks associated with financial transactions and predict potential outcomes. Machine learning algorithms can analyze various types of data, such as credit ratings, economic indicators, and financial market trends, to assess the likelihood of potential risks and identify opportunities for risk reduction. For example, AI-powered credit risk assessment systems are being used to evaluate borrowers' creditworthiness and determine whether they are eligible for loans or credit.

- **Customer Service**: AI-powered chatbots are being used to provide personalized customer service 24/7. These chatbots can answer customer queries and provide advice based on customers' individual needs and preferences. Natural language processing (NLP) technologies enable chatbots to understand and respond to customers' queries in a human-like manner. Financial institutions are using AI chatbots to enhance customer engagement, improve customer satisfaction, and reduce operational costs.

- **Credit Scoring:** AI is being used to develop more accurate credit scoring models that consider a range of factors, such as credit history, income, employment status, and other personal data. Machine learning algorithms can analyze large volumes of data to identify patterns and relationships that can be used to predict creditworthiness. AI-powered credit scoring systems are being used to assess credit risk and make lending decisions.

- **Compliance**: AI can help ensure regulatory compliance by monitoring transactions and identifying potential violations. Machine learning algorithms can analyze financial data to detect unusual transactions or patterns that may indicate fraudulent activity or compliance breaches. AI-powered compliance systems can help financial institutions meet regulatory requirements and reduce the risk of penalties or legal action.

- **Robo-Advisors**: AI-powered robo-advisors are being used to provide personalized investment advice to clients based on their risk appetite, investment goals, and financial circumstances. These systems use machine learning algorithms to analyze client data and make investment recommendations. Robo-advisors are becoming increasingly popular among retail investors who are looking for low-cost investment advice.

- **Natural Language Processing**: AI is being used to analyze unstructured data such as news articles, social media feeds, and other sources of information to identify trends and sentiment related to financial markets.

The current state of art that this thesis is leveraging from is mostly on the trading and investments as said in Misra's thesis: "In our analysis we found that ARIMA, GRU (Gated Recurrent Unit) and TCN (Temporal Convolutional Network) are one of the best performing models for univariate single step price prediction" [6]

As the values that are being predicted for this forecasting thesis are only the BS as a % of NDP the univariable entry for prediction proved once more to work well for ARIMA model.

3. THEORETICAL FRAMEWORK

3.1 ARIMA [6]

ARIMA stands for Auto Regressive Integrated Moving Average. The ARIMA model is a general use of the ARMA (Autoregressive Moving Average) remodel that is suited to explain non-stationary time-series. The dominance of using the ARIMA model is model that coverts a non-stationary series into a series without trend or seasonality by applying a defined number of differentiation of data points. A time series that is stationary has its statistical properties constant over time. If the trend is not defined, its variance is around its very similar to its mean and it has an amplitude that could consider constant. Also, its autocorrelations are constant. Based on the below assumptions, a stationary time series could be defined as a combination of signal and noise. The ARIMA model once you separate the signal from the noise, outputs a single step-ahead to produce forecasts (or the immediate after value). An ARIMA model has three components, the Autoregression (AR) component, the Moving Average (MA) component, and the Integrated (I) differencing component. Autoregression is a type of regression in which past lagged values determine the current value. The order of the lag can be determined by looking into the plots for autocorrelation or partial autocorrelation. The differencing component makes the time series stationary, and the differenced values replace the actual values. The Moving average (MA) part incorporates the dependency between an observation and a residual error that is obtained from a moving average model applied to lagged observations. The order of each of the three components is denoted by the terms p, d, and q. Here p is the number of lag observations required in the AR model, d is the number of times the raw data needs to be differenced to make it stationary, and q is the size of the moving average window.

The full model can be written as:

Equation 1. Equation of ARIMA modeling

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

where y’ is the differenced series (it may have been differenced more than once). The right side of the equation include the lagged values of y’ and the lagged errors. In the study, the ARIMA model is implemented in Python with the help of the stat model library and the auto ARIMA library.

3.2 Cash flow [7]

Cash Flow (CF) is the increase or decrease in the amount of money a business, institution, or individual has. In finance, the term is used to describe the amount of cash (currency) that is generated or consumed in each period. There are many types of CF, with various important uses for running a business and performing financial analysis. This guide will explore all of them in detail.

Types of Cash Flow

There are several types of Cash Flow, so it's important to have a solid understanding of what each of them is. When someone refers to CF, they could mean any of the types listed below, so be sure to clarify which cash flow term is being used.

Types of cash flow include:

1. Cash from Operating Activities – Cash that is generated by a company's core business activities – does not include CF from investing. This is found on the company's Statement of Cash Flows (the first section).

2. Free Cash Flow to Equity (FCFE) – FCFE represents the cash that is available after reinvestment back into the business (capital expenditures). Read more about FCFE.

3. Free Cash Flow to the Firm (FCFF) – This is a measure that assumes a company has no leverage (debt). It is used in financial modeling and valuation. Read more about FCFF.

4. Net Change in Cash – The change in the amount of cash flow from one accounting period to the next. This is found at the bottom of the Cash Flow Statement.

Uses of Cash Flow

Cash Flow has many uses in both operating a business and in performing financial analysis. In fact, it's one of the most important metrics in all of finance and accounting.

The most common cash metrics and uses of CF are the following:

- Net Present Value – calculating the value of a business by building a DCF Model and calculating the net present value (NPV)
- Internal Rate of Return – determining the IRR an investor achieves for making an investment
- Liquidity – assessing how well a company can meet its short-term financial obligations.
- Cash Flow Yield – measuring how much cash a business generates per share, relative to its share price, expressed as a percentage.
- Cash Flow Per Share (CFPS) – cash from operating activities divided by the number of shares outstanding.
- P/CF Ratio – the price of a stock divided by the CFPS (see above), sometimes used as an alternative to the Price-Earnings, or P/E, ratio
- Cash Conversion Ratio – the amount of time between when a business pays for its inventory (cost of goods sold) and receives payment from its customers is the cash conversion ratio
- Funding Gap – a measure of the shortfall a company has to overcome (how much more cash it needs)
- Dividend Payments – CF can be used to fund dividend payments to investors.
- Capital Expenditures – CF can also be used to fund reinvestment and growth in the business

Cash Flow vs Income

Investors and business operators care deeply about CF because it's the lifeblood of a company. You may be wondering, "How is CF different from what's reported on a company's income statement?" Income and profit are based on accrual accounting principles, which smooths-out expenditures and matches revenues to the timing of when products/services are delivered. Due to revenue recognition policies and the matching principle, a company's net income, or net earnings, can be materially different from its Cash Flow.

Companies pay close attention to their CF and seek to manage it as carefully as possible. Professionals working in finance, accounting, and financial planning & analysis (FP&A) functions at a company spend significant time evaluating the flow of funds in the business and identifying potential problems.

Learn more from Harvard about the difference between Cash Flow and Net Income here.

3.3 Balance sheet

Balance Sheet is the financial statement of a company which includes assets, liabilities, equity capital, total debt, etc. at a point in time. Balance sheet includes assets on one side, and liabilities on the other. For the balance sheet to reflect the true picture, both heads (liabilities & assets) should tally (Assets = Liabilities + Equity). [8]

3.4 Profit and loss

A profit and loss statement (P&L), or income statement or statement of operations, is a financial report that provides a summary of a company's revenues, expenses, and profits/losses over a given period. The P&L statement shows a company's ability to generate sales, manage expenses, and create profits. It is prepared based on accounting principles that include revenue recognition, matching, and accruals, which makes it different from the cash flow statement. [9]

3.5 ARMA [11]

The ARMA model is a popular tool for analyzing and forecasting time series data. It combines two fundamental components: autoregressive (AR) and moving average (MA) processes. The AR component captures the dependence of the current value on its own past values, while the MA component captures the dependence on past error terms or random fluctuations.

In an ARMA model, the current value of a variable is expressed as a linear combination of its own past values and past error terms. By including both autoregressive and moving average components, the model can effectively capture various patterns and dynamics in the data.

ARMA models are widely used in fields such as finance, economics, engineering, and environmental sciences for time series analysis and forecasting purposes. They offer a flexible framework that accommodates both short-term and long-term dependencies in the data.

3.6 GARCH [11]

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model is a widely used time series model that allows for the analysis of conditional heteroskedasticity, which refers to the changing volatility or variability of a time series over time. Unlike the ARMA model, which focuses on the mean of the series, the GARCH model focuses on the volatility or variance of the series.

The GARCH model captures the idea that the variance of a series at a given time depends not only on past values of the series but also on past variances or squared error terms. It incorporates autoregressive components for the variance, similar to the AR part of the ARMA model, along with moving average components for the squared error terms.

By capturing the time-varying volatility, the GARCH model is particularly useful in financial econometrics, where asset returns often exhibit clustering of high or low volatility periods. It helps in modeling and forecasting the volatility of financial assets, which is crucial for risk management and option pricing.

3.7 Differences between ARIMA, ARMA and GARCH

The main differences between ARIMA, ARMA, and models lie in their focus and components.

ARIMA: The ARIMA model is designed to capture the dependencies in a time series while handling non-stationarity. It includes autoregressive (AR), moving average (MA), and differencing components. The AR component models the linear relationship between the current value and lagged values, the MA component captures the relationship between the current value and past error terms, and the differencing component helps achieve stationarity by removing trends and seasonality.

ARMA: The ARMA model is focused on capturing the dependencies in a stationary time series. It combines autoregressive (AR) and moving average (MA) components. The AR component models the linear relationship between the current value and lagged values, and the MA component captures the relationship between the current value and past error terms. Unlike ARIMA, ARMA assumes the data is already stationary, without the need for differencing.

GARCH: The GARCH model is used to model and forecast the conditional volatility or variability of a time series, particularly in financial data. It incorporates autoregressive components for the variance, similar to the AR part of ARMA, along with moving average components for the squared error terms. GARCH models address the presence of time-varying volatility, which is crucial for risk management and option pricing.

In summary, ARIMA models handle non-stationary data through differencing, while ARMA models assume stationarity. ARMA models focus on capturing dependencies in the mean of the series, while GARCH models focus on modeling the volatility or variability of the series. The choice between these models depends on the characteristics of the data and the specific aspect of the time series being analyzed or forecasted.

4. METHODOLOGY DEVELOPMENT

The process of creating a systematic approach or method for conducting research, problem-solving, or achieving a specific goal. The methodology provides a framework that outlines the steps to be taken, the tools and techniques to be used, and the data to be collected and analyzed.

4.1 Requirements

The system to be developed will primarily cater to the needs of the Contra Revenue analyst, who is responsible for ensuring that the company's revenues are accurately reflected in the financial statements. One of the key issues that this system is expected to solve is the need to adapt the forecasting model to the latest business requirements, specifically by moving towards market and category level forecasting. By doing so, the system aims to minimize the time required for forecasting and allow the analyst to focus more on analysis, which is the primary responsibility of any financial analyst.

The system is designed to run seamlessly on any Windows system, with a minimum requirement of Windows 11. From a software standpoint, the only requirement is to have Excel installed, preferably from Excel 2013, as the output will appear automatically in an Excel format. This feature will enable the Contra Revenue analyst to access and analyze the data easily without the need for any specialized software or tools.

The system is developed to be highly adaptive and user-friendly. Even if changes are made to the hierarchy, the system will not require any recommendations or changes. Instead, there will be a single step-by-step training process to update the script only in the sections that require it. This approach will ensure that the system can evolve and adapt to the changing needs of the business, without causing any disruption or downtime.

The input of the Excel that will run through the Python algorithm needs to be in a specific program that the analyst will be trained for. This requirement is crucial to ensure that the data is clean and consistent, which is essential for accurate forecasting and analysis.

In summary, the system aims to streamline the forecasting process and enable the Contra Revenue analyst to focus more on analysis. The system is developed to be highly adaptive, user-friendly, and compatible with any Windows system. By using Excel as the output format, the system ensures that the data is easily accessible and can be analyzed by anyone with basic Excel skills. With a user-friendly interface and a minimal learning curve, the system is expected to significantly improve the efficiency and effectiveness of the Contra Revenue analyst, enabling them to deliver more value to the organization.

4.2 RACI Matrix

Responsible- Contra analyst developer and forecaster
Accountable- Contra analyst developer and forecaster
Consulted- Contra manager and director
Informed- FP&A

*The Contra analyst developer and forecaster are the 2 people in charge of the process, the developer is the one who developed and designed the model, the forecaster is the one who presents the information to FP&A, the rest of the RACI people are self explanatory.

5. RESULTS AND DISCUSSION

5.1 Results

The clustering power of Python showed us that we could cluster accounts (figure 12):

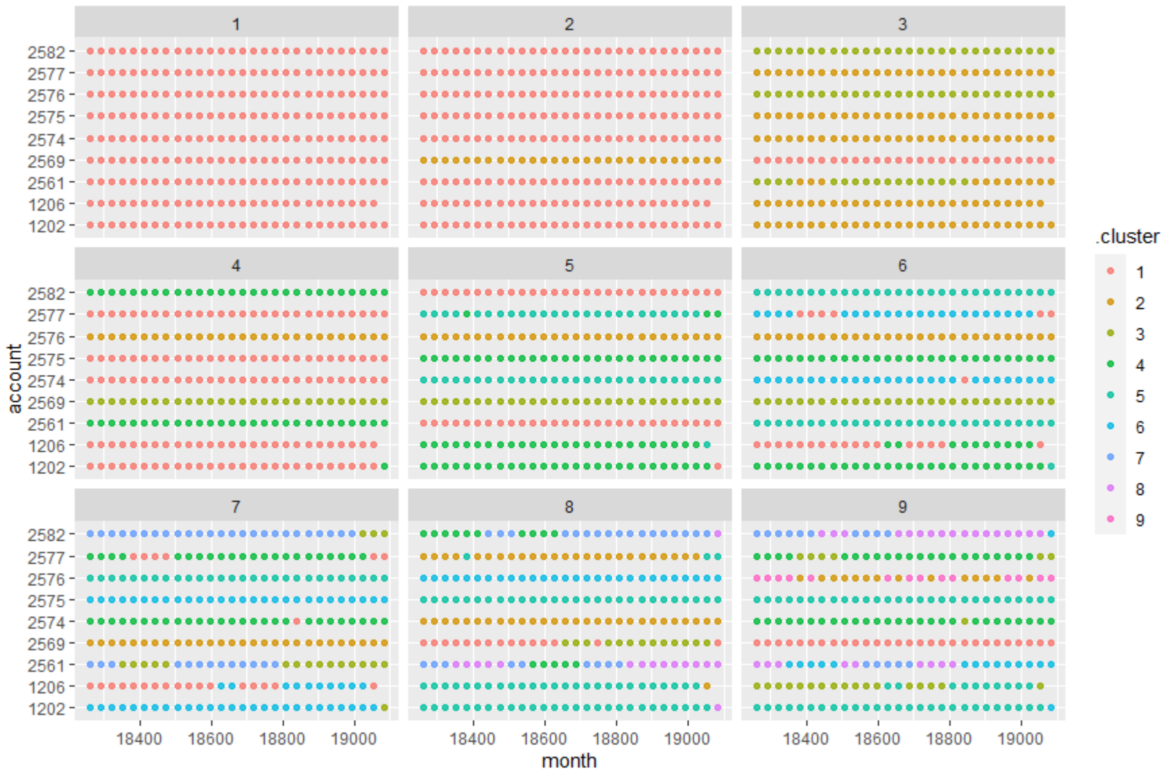


Figure 12. Clustering in Python result

And go from 9 accounts to only 4 clusters of accounts:

- C 2569
- C 2561
- C 2576 2582
- C Other

The forecast decomposition showed an interested 3-month season to all the clusters, aligning to the Quarterly statements that the company presents to stakeholders (figure 13).

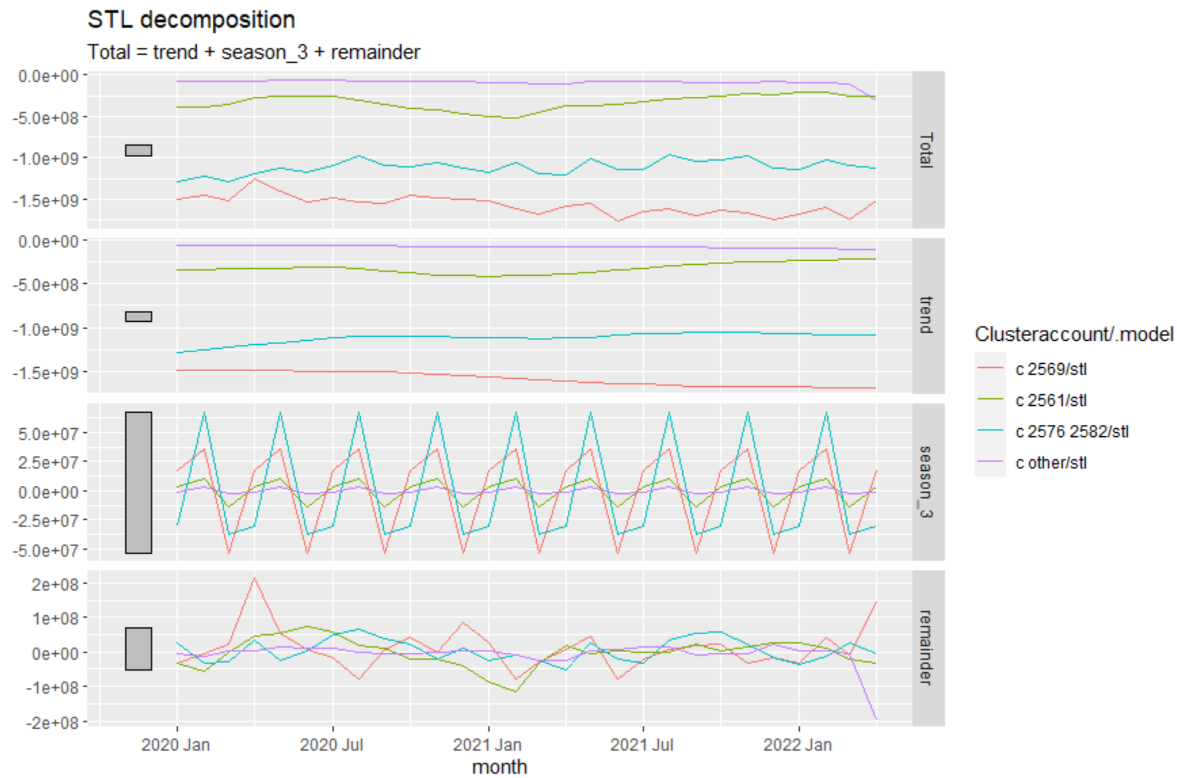


Figure 13. ARIMA decomposition showing a clear seasonality of 3 months in the third graph

Then with the ARIMA clearly signal of seasonality, a forecast was made at an account level (figure 14):

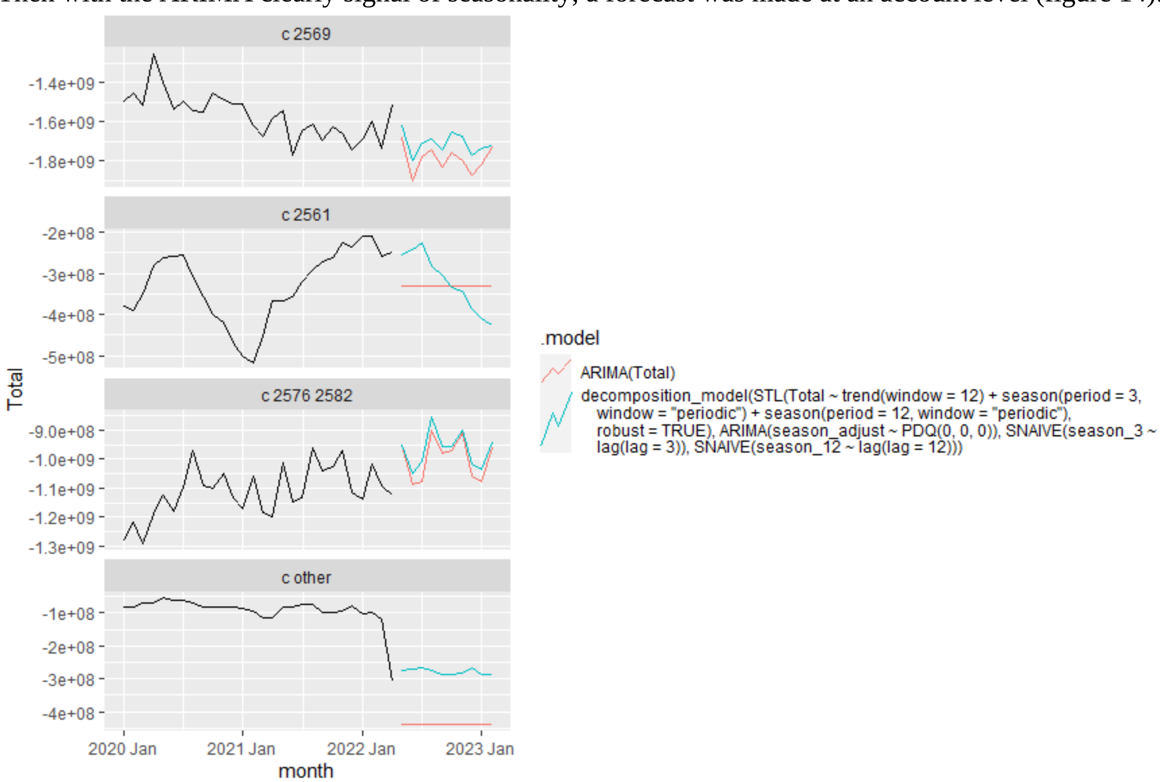


Figure 14. ARIMA forecasting result

From here the decision was made to forecast at a total category/market level, and the allocate by clusters of accounts, because forecasting at the account level would create unnecessary noise.

Also, the decision was made to forecast only at the end of the quarter level (month 3), because that is the data that we need to provide and forecasting for the months in the middle would just create noise.

As explained before, the main input for the forecast is the revenue that FP&A team provides, and just to prove that it has same seasonality as out reserve levels, the series was decomposed (figure 15):

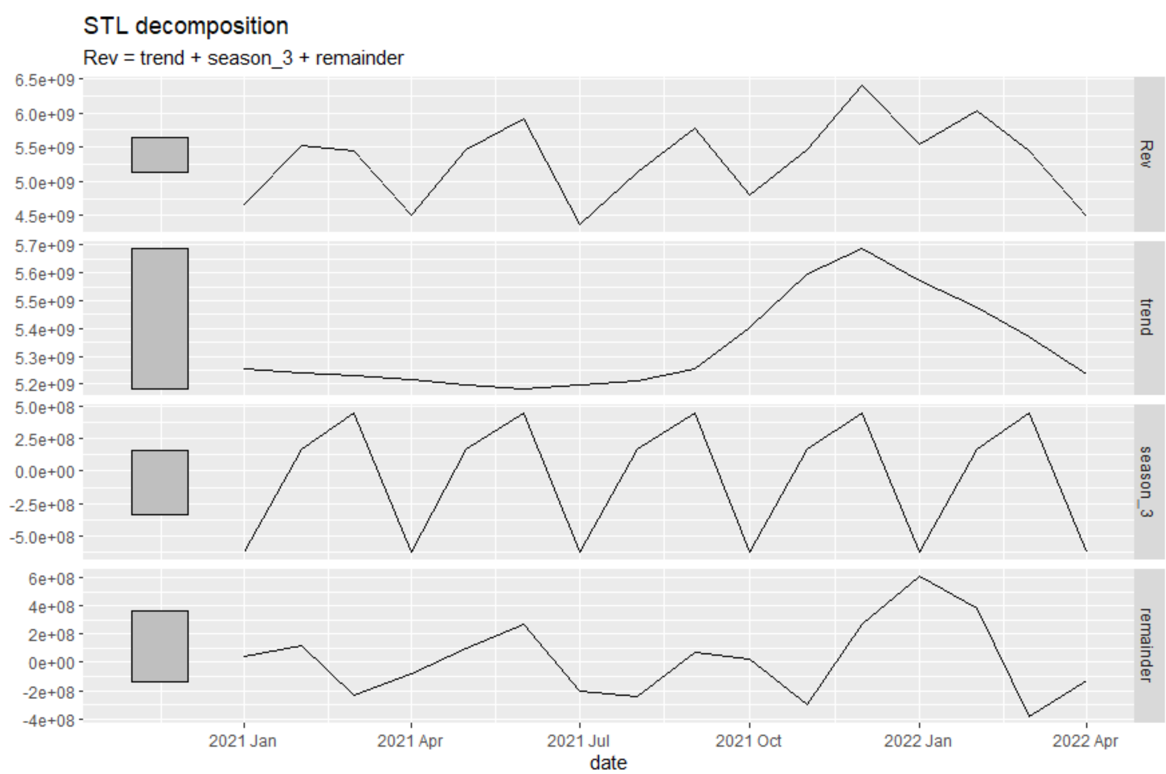


Figure 15. Revenue decomposition showing a clear season every 3 months in the third graph

And proven to be right, the reserve is highly correlated at a total level to the revenue trend.

As not everyone knows how to use Python, the idea to have an automatic executable came to mind where the inputs are placed and then retrieved with the results in an Excel sheet.

Input screen (figure 16):

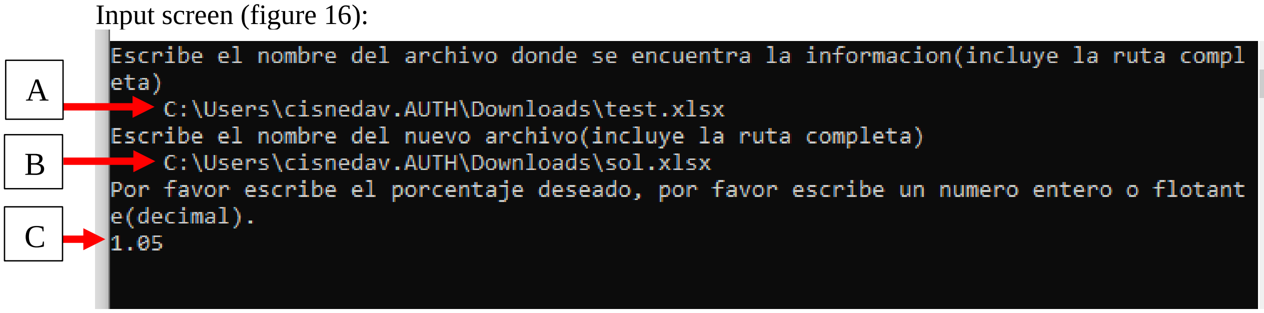


Figure 16. New model input showing what needs to be entered

A: it is the input file, where the data needs to be organized in the following format (figure 17):

	A	B	C	D	E	F	G	H
1	**GBU 2**	**Market**	**Quarter**	**NDP**	**Backend**	**BS**	**priorBS**	**Payments**
2	**Business PC Solutions**	**Greater Asia**	Q321	$1,436 M	$639 M	$142 M	$142 M	$639 M
3	**Business PC Solutions**	**Greater Asia**	Q421	$1,505 M	$684 M	$154 M	$142 M	$697 M
4	**Business PC Solutions**	**Greater Asia**	Q122	$1,780 M	$789 M	$156 M	$154 M	$790 M
5	**Business PC Solutions**	**Greater Asia**	Q222	$1,600 M	$723 M	$161 M	$156 M	$729 M
6	**Business PC Solutions**	**Greater Asia**	Q322	$1,439 M	$640 M	$169 M	$161 M	$648 M
7	**Business PC Solutions**	**Greater Asia**	Q422	$1,389 M	$632 M	$167 M	$169 M	$630 M
8	**Business PC Solutions**	**Greater Asia**	Q123	$1,545 M	$784 M	$186 M	$167 M	$803 M

Figure 17. Organized date for input showing the table example of the data that needs to be inputed

The user needs to respect the formats and do not over right formulas.

B: it is the output file; it is a standard format that needs to be provided and respected as well.

C: it is the coefficient needed to give, as we are always forecasting the following reserve quarter, what this number indicates is how much (in %) is the new revenue forecast in comparison to the last, for the example using 1.05 it means the revenue is expected to grow 5% from the prior period.

Then the executable asks for what combination of data are you looking for (figure 18):

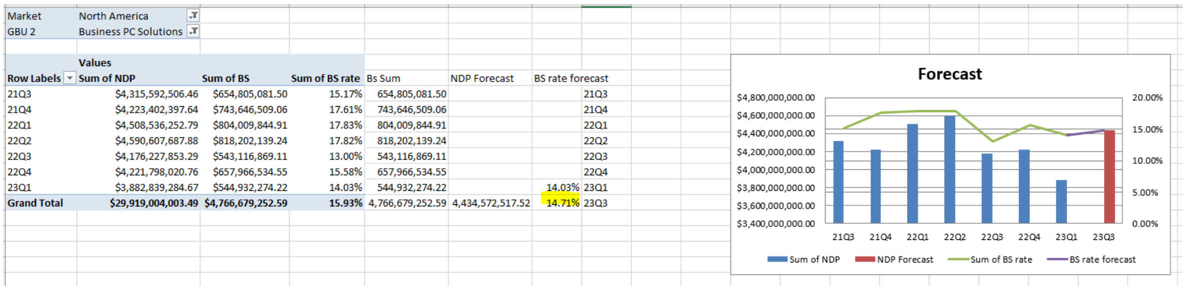
```
Escribe el nombre del archivo donde se encuentra la informacion(incluye la ruta
completa)
    C:\Users\cisnedav.AUTH\Downloads\test.xlsx
Escribe el nombre del nuevo archivo(incluye la ruta completa)
    C:\Users\cisnedav.AUTH\Downloads\sol.xlsx
Por favor escribe el porcentaje deseado, por favor escribe un numero entero o fl
otante(decimal).
1.05
Por favor escribe el numero que deseas de los filtros, si vas a ocupar varios fi
ltros, separalos con una coma,
      Market:
      1.Central and Eastern Europe
      2.Greater Asia
      3.Greater China
      4.India, B&SL
      5.Latin America
      6.North America
      7.Northwest Europe
      8.Southern Europe Middle East and Africa

      GBU 2:
      9.Business PC Solutions
      10.Consumer PC
      11.Non-Supplies
      12.Other Print
      13.Supplies
```

Figure 18. Filter selection on executable

You can select as many markets and as many GBU (categories) as needed, here there are a lot of possible combinations to make and forecast.

Once you select the combinations needed, hit enter and the program is going to automatically create the new structure in the file you selected as the solution file in the following form (figure 19):



Market	North America						
GBU 2	Business PC Solutions						
	Values						
Row Labels	Sum of NDP	Sum of BS	Sum of BS rate	Bs Sum	NDP Forecast	BS rate forecast	
21Q3	$4,315,592,506.46	$654,805,081.50	15.17%	654,805,081.50			21Q3
21Q4	$4,223,402,397.64	$743,646,509.06	17.61%	743,646,509.06			21Q4
22Q1	$4,508,536,252.79	$804,009,844.91	17.83%	804,009,844.91			22Q1
22Q2	$4,590,607,687.88	$818,202,139.24	17.82%	818,202,139.24			22Q2
22Q3	$4,176,227,853.29	$543,116,869.11	13.00%	543,116,869.11			22Q3
22Q4	$4,221,798,020.76	$657,966,534.55	15.58%	657,966,534.55			22Q4
23Q1	$3,882,839,284.67	$544,932,274.22	14.03%	544,932,274.22		14.03%	23Q1
Grand Total	**$29,919,004,003.49**	**$4,766,679,252.59**	**15.93%**	4,766,679,252.59	4,434,572,517.52	14.71%	23Q3

Figure 19. Output of executable model

For the example, the selections were: 6,9 (North America and Business PC solutions)
We can see that the ARIMA is estimating a total reserve at the 14.71% of the revenue forecasted, historical data sits at around 15%, so the data does make sense.

The fixed ARIMA after trying with different assumptions and auto ARIMA was (3,1,4).
As the data per Market goes as back as 2021 the data is still missing longer history, but as the ARIMA is considering the longer period to be at 4, the data seems to be just the necessary.

5.2 Discussion

The idea of the executable is something completely useful that I will recommend, because not every financial person is necessarily an AI person, nor knows how to use Python, se the beauty of this implementation is that is easy to manager.

The ARIMA was decided to be kept as fixed after several back testing and the usage of the function auto ARIMA, the result for the fixed ARIMA is (4,1,3).

The one-fits-all model may have variances if some of the time series have a different behavior, but in general it will provide a great estimate of the level of reserves.

There are plenty of models such as GARCH, ARMA, AR, MA, etc., the use of ARIMA was proven to work well with the set of data available.

6. CONCLUSIONS

6.1 Conclusions

The specific objectives were achieved by:

The Balance sheet forecast for Contra Revenue was experiencing a lot of manual labor invested, and needed to be updated due to some internal market organization in HP.

Three main objectives were pretended to be achieved, which were:
-Automate the process
-Adapt to a market level forecast
-Improve accuracy

Taking advantage of the estate of art of AI applied to finance the objectives were completed.

The result is an automated executable that any analyst could use.

The use of Python for clustering and ARIMA for forecasting proved to be a valuable tool for financial planning and analysis. The implementation of an executable program that allows users to input data and receive forecasts in a user-friendly format is a great way to bridge the gap between those who have expertise in AI and those who do not. The results of the forecasting showed a clear seasonality in the data, which allowed for accurate predictions of future reserves. While there are other models that could be used, the ARIMA model was found to be the best fit for the available data. Overall, the use of these techniques can greatly improve financial forecasting and decision making

In conclusion, the objectives of this study were successfully achieved. By implementing the new system, the amount of manual work invested in the calculations was massively reduced. This allowed the Contra Revenue analyst to focus more on analysis, which is the primary responsibility of any financial analyst. Additionally, the accuracy of the forecasting model was improved, as evidenced by the back testing results, which showed a reduction in forecasting errors from 5.2% to 4.1%.

Another significant accomplishment was achieving the goal of having individual time series or even combinations of categories and markets. This feature allows for more accurate forecasting and analysis, especially when dealing with complex and dynamic markets.

6.2 Future work

However, there is still a lot of work to do. One of the primary areas of future work is to automate the allocation down to the account level. This would further reduce manual work and enable the system to handle larger and more complex datasets.

Another area of future work is the implementation of neural networks and auto ARIMA functions. These advanced techniques can improve the accuracy of the forecasting model even further, especially when dealing with complex and dynamic markets.

Furthermore, adding the rate estimation from original Flash files, as described in the introduction, would provide additional insights, and enhance the overall accuracy of the forecasting model.

Finally, a crucial area of future work is to create an interactive dashboard in Power BI as a stage 2 of the executable. This would allow the Contra Revenue analyst to visualize the data and insights more effectively, facilitating better decision-making and analysis.

In summary, while this study achieved its specific objectives, there is still a lot of room for improvement. The future work outlined in this paper provides a roadmap for further research and development in this area. The implementation of these improvements would significantly enhance the accuracy, efficiency, and effectiveness of the system, providing significant benefits to the Contra Revenue analyst and the organization.

BIBLIOGRAPHY

[1] Patrick Thibodeau, "5 reasons HP's Mark Hurd resigned," *Computerworld*, 2010.

[2] J. Cornell, "Hewlett-Packard Completes Spin-Off Of Enterprise Business," *Forbes*, 2015.

[3] M. &. S. H. Ghiassi, Artificial Intelligence in Finance: A Review, Journal of Financial Management of Property and Construction, 2019.

[4] W. L. S. &. Y. S. Jang, A Review of Artificial Intelligence Applications in Finance, 2020.

[5] Y. Z. J. &. C. T. Li, Artificial Intelligence in Finance: A Bibliometric Analysis and Research Agenda, Journal of Intelligent & Fuzzy Systems, 2021.

[6] V. Misra, Time series forecasting with applications to finance, New Jersey: New Jersey Institute of Technology, 2021.

[7] Athanasopoulus-Hyndman, Forecasting: Principles and Practice, Melbourne, Australia, 2021.

[8] CFITeam, "Cash Flow," *Corporate Financial Institute*, 2023.

[9] CFITeam, "What is Balance sheet," *The Economic Times*, 2023.

[10] CFITeam, "What is the Profit and Loss Statement (P&L)?," 2023.

[11] G. E. P. J. G. M. R. G. C. &. L. G. M. Box, Time Series Analysis: Forecasting and Control, 2015.