GC-47

IT7993 Capstone Spring 2021
Key Professional Dataset- Data Spider

Abstract

The purpose of this project is to build a Web Crawler to extract personal information from a public website like Reddit and LinkedIn. We completed the Instagram crawling as a bonus for the project. The team will be using MySQL or any other open source relational database to organize the data and conduct a quantitative data analysis on it.

https://sites.google.com/view/it7993-spring-2021/home

Introduction

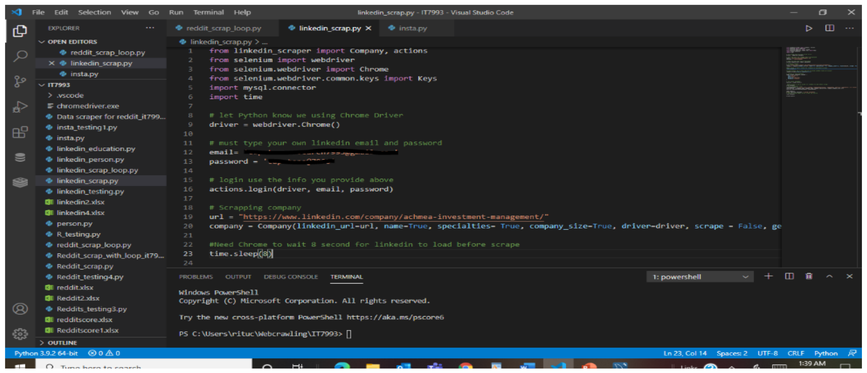
Our goal is to establish a one stop shop for institutional asset management distribution intelligence; the one place to go for mandates, documents, and profiles of consultants, investors, and managers with key professional contact information.

Materials and Methods

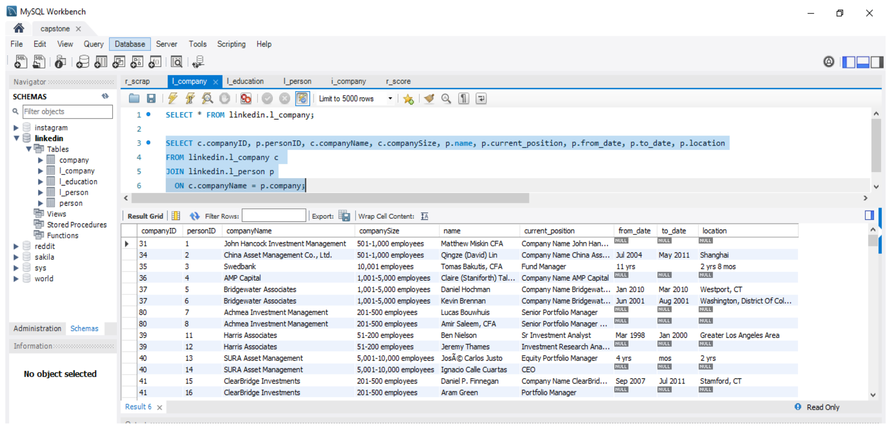
The Key Professionals dataset we call dataspider is aimed at delivering global coverage of investors and consultant key professional that are involved in making investment decisions beginning with US based firms. Jing Wang from eVestment has provided a list of 400 asset managers to crawl data, organize in a database and create appropriate visualization tools. By the end of this project, the goal is to provide coverage for the US Investors and Consultants found within eVestment platform.

- Asset manager list provided by Jing Wang
- Github list provided by Dr. Han
- Visual Studio Code
- Python
- MySQL Workbench/My SQL Installer for Window
- Juypter
- Db4free.net
- PowerBl
- Tableau 2020.4

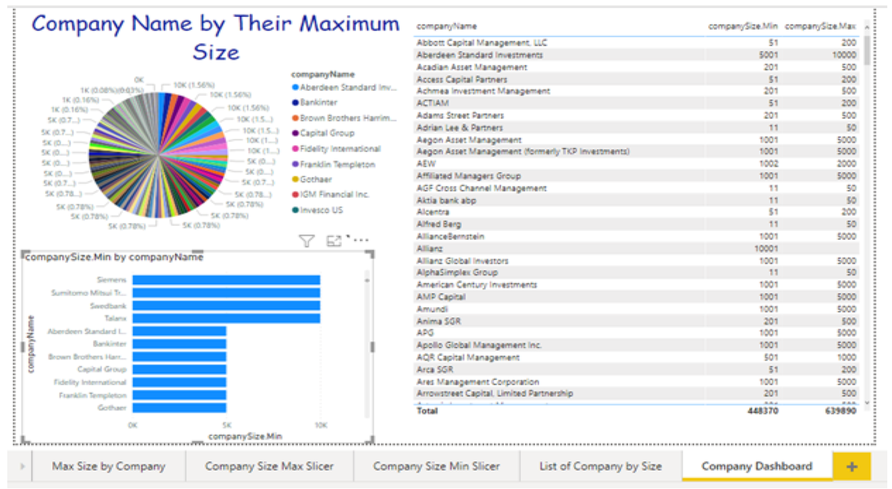
Code Snippet



Database



Visualization



Conclusions

The project is to create a data repository and visualization for 400 asset managers that are provided by our sponsor. The objective was achieved through the Data Spider system that we created. The system crawls different social media such as Linkedin, Reddit, and Instagram. The crawled data then persists into a centralized MySQL database. A visualization is then built on top utilizing Tableau dashboard. Since we do have a final product that is working as intended, the project is successful.

Acknowledgments

Dr. Meng Han - Professor
Jing Wang - Sponsor

Contact Information

Janell Westmoreland
jwestmo6@students.kennesaw.edu
Kajal Vaghani
kvaghani@students.kennesaw.edu
Ritu Choudhary
rchoudh2@students.kennesaw.edu
Vy Duong
vduong.it7@gmail.com
Nyong Nkereuwem
johnsonjuniornn@gmail.com

References

The Source for Institutional Intelligence. eVestment. (2021, April 9). https://www.evestment.com/.

Boe, B. (n.d.). *PRAW: The Python Reddit API Wrapper*. Retrieved from https://praw.readthedocs.io/en/latest/

Greening, C. (2021). *instascrape*. Retrieved from https://chris-greening.github.io/instascrape/

Sham, J. (2021, April 10). *Scrapes user data from LinkedIn*. Retrieved from https://pypi.org/project/linkedin-scraper/

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING