Long Nguyen

https://www.linkedin.com/in/khlongg/ | https://khlong189.github.io/longknguyen.github.io/
ngkhlong189@gmail.com

WORK EXPERIENCE

Data Scientist Intern   July 2023 - Present
Rel8ted.to Analytics - Rochester, NY
- Developing a **Python** data scraper with multithreading support to verify 1.4 million URLs using XPath queries in parallel.
- Utilized web scraping techniques, including cookie requests and **CURL**, to collect essential data from a diverse array of 10,000 Canadian companies.
- Collaborating with the team to utilize **GPT**-based embedding models for predicting NAICS codes of 30,000 Canadian companies based on their homepage's similarity to NAICS code descriptions.

Data Analyst Intern   Jan 2022 - May 2022
Leafprint Enterprises - Cedar Hills, UT
- Collected 100,000 geographical data points from diverse administrative levels using **BeautifulSoup**.
- Utilized Python's **pandas** library to clean and standardize the data for future EDA.

Data Analyst Intern   Mar 2021 - July 2021
BENIT PTY LTD – Hanoi, Vietnam
- Utilized **SQL** queries to create schemas for the scraped data and perform table joins, enhancing the organization and integration of the collected information.
- Employed **pandas** to streamline data cleaning, formatting, and quality enhancement processes.

PERSONAL PROJECTS

Soccer Scouting Database
- Employed Python's random library to create synthetic data for youth players' details.
- Created a comprehensive **MySQL** database handling 3,300 player records across 10 distinct tables.

Mental Health Data Mining
- Utilized data mining to analyze 1,259 data points, unveiling factors impacting employee mental health treatment via **association rules**.
- Demonstrated a 74.6% accuracy rate with the top-performing SVM model, bolstering predictive capabilities and informing strategies for mental health intervention.

Twitter Sentimental Analysis on Asian Hate During COVID-19
- Collected over 16,000 Asian and COVID-related tweets using **snscrape**.
- Conducted thorough **EDA** to quantify positive and negative sentiment.
- Employed **WordClouds** to visualize essential discussions and themes within the realm of Asian COVID-related tweets.

EDUCATION

University of Rochester   Aug 22 - Expected Dec 2023
Rochester, Rochester, NY
- Master's Degree, **Data Science** - 3.67/4.0 GPA

Augustana College - Bachelor of Arts, **Applied Mathematics**   Aug 2018 - May 2022
Certificates: *Google Data Analytics Professional Certificate, Udacity's Data Analyst Nanodegree.*

TECHNICAL SKILLS

Programming: Python, R, SQL.
Technologies: pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, dplyr, ggplot2.
Tools: Github, Jupyter Notebook, Google Colab, Rstudio, MySQL Workbench, Excel, Tableau.