


Long Nguyen

✉ ngkhlng189@gmail.com ☎ 309-631-4516

🔗 [khlong189.github.io/longknguyen.github.io/](https://github.com/khlong189)  <https://www.linkedin.com/in/khlongg/>

EDUCATION

University of Rochester

MS in Data Science

Augustana College

Bachelor of Arts in Applied Mathematics

Aug 2022 - Expected Dec 2023

Current GPA: 3.67/4.0

Aug 2018 - May 2022

GPA: 3.77/4.0

SKILLS

Languages: Python, R, SQL.

Libraries: pandas, numpy, scikit-learn, matplotlib, seaborn, dplyr, ggplot2.

Tools: Github, Jupyter Notebook, Google Colab, Rstudio, MySQL Workbench, Excel, Tableau.

EXPERIENCE

Teaching Assistant | *University of Rochester* | *Rochester, NY, USA*

Aug 2023 - Present

- Assisting students in understanding a wide range of statistical topics, including: descriptive statistics, probability/combinatorics, confidence interval, hypothesis testing, regression analysis, and case studies.

Data Scientist Intern | *Rel8ed.to Analytics* | *Rochester, NY, USA*

Jul 2023 - Aug 2023

- Collaborated with the team to employ OpenAI embedding and GPT models for predicting NAICS codes by assessing homepage content similarity with corresponding code descriptions.
- Built a **Python** scraper with multithreading to concurrently validate 1.4 million URLs using XPath queries, boosting data collection efficiency by 80%.
- Employed web scraping techniques such as cookie requests and CURL to gather data from a diverse set of 10,000 Canadian companies.

Data Analyst Intern | *Leafprint Enterprises* | *Cedar Hills, UT, USA*

Jan 2022 - May 2022

- Collected 100,000 geographical data points using **BeautifulSoup** from various administrative levels.
- Applied **pandas** for data preprocessing, enhancing efficiency and accuracy in subsequent analyses.

Data Analyst Intern | *BENIT PTY LTD* | *Hanoi, VN*

Jan 2022 - May 2022

- Designed and implemented **SQL** schemas to integrate scraped data from TwelveAPI into a structured database.
- Conducted intricate table joins, linking diverse datasets to enhance data organization and accessibility.
- Applied **pandas** for streamlined data cleaning, formatting, and quality enhancement processes, reducing data preparation time by 60%.

PROJECTS

Soccer Scouting Database | *Python, SQL*

- 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 Mining and Prediction | *Python*

- Utilized data mining to analyze 1,259 data points, revealing factors affecting employee mental health treatment using **association rules**.
- Achieved a 74.6% accuracy rate using the top-performing **SVM** model, enhancing predictive capabilities and guiding strategies for mental health intervention.

Twitter Sentimental Analysis on Asian Hate During COVID-19 | *Python*

- Collected over 16,000 Asian and COVID-related tweets using **snsrape**.
- Conducted thorough EDA to quantify positive and negative sentiment.
- Utilized **WordCloud** to visualize key discussions and themes in Asian COVID-related tweets.