

Long Nguyen

ngkhlng189@gmail.com | 309-631-4516
[linkedin.com/in/khlongg/](https://www.linkedin.com/in/khlongg/) | [khlong189.github.io/longknguyen.github.io/](https://github.com/khlong189/longknguyen.github.io/)

SUMMARY

Data Scientist with a strong foundation in machine learning, statistical analysis, **SQL**, and **Python/R** programming. Adept in conducting **A/B testing**, leveraging **Bayesian** statistics, and machine learning methods, including **Linear regression**, **Logistics regression**, **XG Boost**, **Decision Trees**, **Random Forest**. Anticipated graduation in December 2023, eagerly available to contribute and add value from January 1st onwards.

EDUCATION

University of Rochester

Rochester, Rochester, NY

Aug 22 - Expected Dec 2023

- 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.

WORK EXPERIENCE

Teaching Assistant (R, regression analysis, hypothesis testing)

Aug 2023 - Present

University of Rochester - Rochester, NY

- 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 (Python, OpenAI, chatGPT, NLP)

July 2023 - Aug 2023

Rel8ted.to Analytics - Rochester, NY

- Collaborated 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.
- Engineered a **Python** data scraper using multithreading to validate 1.4 million URLs simultaneously using XPath queries, improving data collection efficiency by 80%.

Data Analyst Intern (Python)

Jan 2022 - May 2022

Leafprint Enterprises - Cedar Hills, UT

- Aided in gathering 100,000 geographical data points via **BeautifulSoup** for an internal Google Maps-style company dictionary, streamlining operations.
- Employed Pandas to preprocess data, reducing data entry errors by 80% for enhanced analysis.

Data Analyst Intern (Python, SQL)

Mar 2021 - Jul 2021

BENIT PTY LTD - Hanoi, Vietnam

- Designed SQL schemas, seamlessly integrating scraped TwelveAPI data into a structured database, boosting operational efficiency.
- Proficiently integrated diverse datasets, enhancing data organization and accessibility, resulting in a 70% reduction in data retrieval time and informed decision-making.
- Employed **Pandas** for data cleaning, formatting, and quality enhancement processes, reducing data preparation time by 60%.

PROJECT

Twitter Sentimental Analysis on Asian Hate During COVID-19 (Python)

- Collected over 16,000 Asian and COVID-related tweets using **snsrape**.
- Conducted **EDA** to quantify sentiment, achieving goals including hashtag frequency, targeted language filtering, date conversion, and refined sentiment analysis.
- Employed **WordClouds** to visualize essential discussions and themes within the realm of Asian COVID-related tweets.