

LONG KHANH NGUYEN

ngkhlong189@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/) | [Tableau Viz](#)

EDUCATION

UNIVERSITY OF ROCHESTER *STEM Master of Science in Data Science*

- Relevant courses: Data Mining, Comp. Statistics, Statistical Machine Learning, Database Management, Deep Learning.

Rochester, NY

Aug 2022 - Dec 2023

AUGUSTANA COLLEGE *Bachelor of Arts in Applied Mathematics*

Rock Island, IL

Aug 2018 - May 2022

PROFESSIONAL EXPERIENCE

REL8ED.TO ANALYTICS

Rochester, NY

Data Science Intern

Jul 2023 - Aug 2023

- Partnered with Coface's Data Science team to enhance predictive capabilities, employing a **GPT**-based embedding model for accurately forecasting NAICS codes for 30,000 Canadian companies by leveraging homepage similarity analysis.
- Developed **Python** data scrapers with multithreading to validate over 1 million URLs concurrently, utilizing XPath queries and cURL, improving data collection efficiency by 80%.
- Utilized **Excel** (Pivot Table, XLOOKUP) for dynamic data analysis and implemented conditional formatting to visually highlight specific insights.

UNIVERSITY OF ROCHESTER

Rochester, NY

Teaching Assistant

Aug 2023 - Dec 2023

- Mentoring students in diverse statistical areas: descriptive statistics, probability and distribution, confidence intervals, hypothesis testing, inference, ANOVA, and regression analysis.
- Leading multiple review sessions for exams and projects, resulting in an 80% increase in students' overall grades.
- Empowering students in understanding **R** packages and creating impactful reports with **RMarkdown**.

LEAFPRINT ENTERPRISES

Cedar Hills, UT

Data Analyst Intern

Jan 2022 - May 22

- Assisted in collecting over 100,000 geographical data entries with **BeautifulSoup**, enriching the company's dataset for robust analytics.
- Applied **Pandas** for data wrangling, reducing data entry errors by 80% and expediting data-driven decision-making.
- Optimized the internal company dictionary, contributing to a 15% improvement in data accessibility.

BENIT PTY LTD

Hanoi, Vietnam

Data Analyst Intern

Mar 2021 - Jul 2021

- Designed data models, incorporating TwelveAPI data into a structured database to enhance operational efficiency.
- Managed database using **AWS** tools like PostgreSQL, Redshift Cluster, and S3, reducing data retrieval time by 70%.
- Applied **Pandas** for easy data cleanup, cutting data prep time by 60%.

PROJECTS

Mental Health Mining and Prediction | Python (sklearn, pandas, numpy, matplotlib, seaborn)

- Utilized **data mining** technique to analyze 1,259 observations, revealing factors affecting employees' mental health treatment status using **association rules**.
- Achieved 74.6% accuracy predicting treatment necessity with **SVM** model and Grid-CV hyperparameter tuning.

Twitter Sentimental Analysis on Asian Hate During COVID-19 | Python (pandas, plotly, re, nltk, textblob, WordClouds)

- Collected over 16,000 Asian and COVID-related tweets using sncrape.
- Conducted thorough exploratory data analysis to quantify positive and negative sentiment.
- Employed WordClouds to visualize essential discussions and themes within the realm of Asian COVID-related tweets.

TECHNICAL SKILLS

- Programming:** Python, SQL, R
- Data Visualization:** Tableau, Excel
- Database Management:** MySQL, PostgreSQL
- Familiar with:** Database design and development, Statistical modeling and Machine learning techniques