

LONG KHANH NGUYEN

ngkhlong189@gmail.com | 309-631-4516

[Linkedin](#) | [Tableau Viz](#) | [Portfolio](#)

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

Data Science Intern

Rochester, NY

Feb 2023 - Aug 2023

- Led an end-to-end LLM project to classify North American companies into their respective industries, employing **Python**, OpenAI, and utilizing the ChatGPT chatbox function.
- Achieved 90% in data retrieval time by building 5 **Python** scrapers to extract homepage information for 10M companies.
- Collaborated with business teams to convert complex business logics, and successfully built 6 **Excel** dashboards incorporating pivot tables, XLOOKUP based on client requests, reducing delivery time by 70%.

LEAFPRINT ENTERPRISES

Data Analyst Intern

Cedar Hills, UT

Jan 2022 - May 2022

- Led global municipal research, collected 100k+ data points, and delivered actionable recommendations to enhance World Archives's municipal dictionary.
- Utilized **Python**'s pandas for data cleaning and transformation, reducing errors and retrieval time by 90%.
- Developed and maintained interactive dashboards and reports using **Tableau** to support data driven decision-making.

BENIT PTY LTD

Data Analyst Intern

Hanoi, Vietnam

Mar 2021 - Jul 2021

- Developed an optimized **SQL** data model for TwelveAPI's Forex data, improving access time by 80%.
- Utilized **AWS S3** and **MongoDB** for scalable database management, resulting in a 70% reduction in overall latency.
- Integrated **Python** with **SQL** databases to streamline **ETL** processes, ensuring accurate financial insights via dynamic dashboards and reports.

DATA ANALYTICS PROJECTS

- Behavioral Segmentation for a hotel and resort company (ongoing)** Building a K-means clustering model in **Python** on a hotel and resort dataset of more than 80k customers to segment them into different behavioral groups, and provided key strategies to enhance customer experience and maximize revenue for each segment.
- [NYC bike tracking forecast](#) Led a team of 4 fellow Data Science graduates to achieve 70% accuracy in building a forecasting model using **Prophet** for Citibike Bike Sharing app on **Databricks**. Collaborated on **GitHub**, developed end-to-end **ETL** pipelines, and created bronze, silver, and gold data tables using **PySpark** and **Python** to deploy a solution for monitoring and optimizing bike inventory at designated NYC stations.
- [Mental health mining and prediction](#) Preprocessed and analyzed 1,259 employee observations, using association rules and SVM with Grid-CV to predict mental health treatment necessity with 74.6% accuracy in **Python** (sklearn, pandas, numpy, matplotlib, seaborn), **R** published: [Mental Health prediction on Rpubs](#) (an implementation in R with statistical analysis).

TECHNICAL SKILLS

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