

# Long Nguyen

47 Gold St, Rochester, NY 14620

(309)-631-4516 | [ngkhlong189@gmail.com](mailto:ngkhlong189@gmail.com) |

<https://www.linkedin.com/in/khlongg/> | <https://khlong189.github.io/longknguyen.github.io/>

## 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.*

## 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.
- Leveraging **hidden APIs**, gathered data from 10,000 local companies within the targeted search domain, specifically Alberta Health Services, and stored the results in **JSON** format.
- 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*

- Built a Python script using **BeautifulSoup** to scrape 100,000 geographical data points from diverse administrative levels.
- Utilized Python's **pandas** library to preprocess the data for further analysis and storage in **Excel**.

### Data Analyst Intern

Mar 2021 - July 2021

*BENIT PTY LTD – Hanoi, Vietnam*

- **Forex Trading Web App** | [Code](#) | Contributor
- Contributed to the development of a forex trading engine for company use with 60% accuracy.
- Utilized **SQL** queries to create schemas for the scraped data and perform table joins, enhancing the organization and integration of the collected information.
- Utilized **pandas** to clean, format, and enhance data quality.

## PERSONAL PROJECTS

---

### Soccer Scouting Database | [Code](#) | (Python, SQL)

- Generated dummy data using **Python**'s library - **random**, designed an efficient database schema, and used **SQL** queries to assess youth players based on key criteria for streamlined player scouting and decision-making.

### Mental Health Data Mining | [Code](#) | (Python)

- Investigated and predicted the factors contributing to employees seeking treatment for mental illness, including family history and workplace issues, using **association rules** and ML models. Achieved a **74.6%** accuracy rate with the best-performing model, **SVM**.

### Twitter Sentimental Analysis on Asian Hate During COVID-19 | [Code](#) | (Python)

- Scraped Asian and COVID-related tweets using **snsrape**, performed **EDA** on positive and negative sentiments, and visualized common words using **WordCloud**.

## 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.