

# Long Nguyen

47 Gold St, Rochester, NY 14620

(309)-631-4516 | [ngkhlng189@gmail.com](mailto:ngkhlng189@gmail.com) |

<https://www.linkedin.com/in/khlngg/> | <https://khlng189.github.io/longknguyen.github.io/>

## EDUCATION

---

### University of Rochester

Aug 22 - Expected Dec 2023

*Rochester, Rochester, NY*

- Master's Degree, **Data Science** - 40% tuition scholarship - 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 robust **Python** scripts for data scraping using hidden **APIs**, with a focus on leveraging network monitoring techniques and extracting JSON data efficiently and accurately to meet clients' specific requests.
- Utilizing **XPath** and **BeautifulSoup** (bs4) to extract structured data from HTML.
- Collaborating with the team to leverage **ChatGPT Turbo 5** in implementing and evaluating cutting-edge NLP models. Applying advanced techniques like tokenization and word embedding to drive the development of high-performing and efficient solutions.

### Data Analyst Intern

Jan 2022 - May 2022

*Leafprint Enterprises - Cedar Hills, UT*

- Collected and cleaned jurisdiction data from countries worldwide, utilizing **Excel** and **Python**, resulting in the creation of a comprehensive and accurate geographic system for World Archives.

### Data Analyst Intern

Mar 2021 - July 2021

*BENIT PTY LTD – Hanoi, Vietnam*

- **Forex Trading Web App** | [Code](#) | Contributor
- Created a forex trading engine for company use with **60%** accuracy.
- Implemented web scraping techniques using **BS4** and **TwelveData API** to collect structured data from websites.
- Employed **SQL** queries to create schemas and perform **JOIN** operations for efficient data organization and analysis.
- Utilized **pandas** to clean, format, and enhance data quality for further analysis.

## 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.