

## QUOC-HUY NGUYEN

(714) 837-0205 | [quochuyn1999@gmail.com](mailto:quochuyn1999@gmail.com) | [linkedin.com/in/quochuyn](https://www.linkedin.com/in/quochuyn) | [quochuyn.github.io](https://github.com/quochuyn)

### EDUCATION

**University of Michigan, Ann Arbor**  
Master of Applied Data Science

May 2022 - Aug 2023  
3.9 GPA

**University of California, Irvine**  
Bachelor of Science in Mathematics - Data Science Specialization, Statistics Minor

Sep 2017 - Jun 2021  
Summa Cum Laude, 3.9 GPA

### EXPERIENCE

**Upwork** | California  
Data Scientist (Freelancer)

May 2022 - Present

- Evaluated and fine-tuned an overfitting ResNet-based TensorFlow convolutional neural network (CNN) on brain tumor imaging and EEG brain signal data, aligning it with the requirements of a client's medical research within the healthcare domain.
- Developed Python scripts using GeoPandas for analytical insights and geographic visualizations, offering a comprehensive understanding of patterns and trends in malicious hacking attempts.
- Implemented SQL queries to extract valuable insights from customer billing transactions that were presented to a stakeholder.
- Delivered expert consulting on Python programming: data engineering, code reviews, project planning, and documentation.

**Curacao** | Los Angeles, California  
Data Analyst

Sep 2021 - Apr 2022

- Engineered a data pipeline for our mobile application product, enabling the extraction, transformation, and loading (ETL) of 16+ GB of Open Banking data from FinTech APIs to underwrite personal loans.
- Designed and developed a scikit-learn random forest classification model that surpassed a 70% prediction accuracy, discerning key thresholds of diverse credit risk levels.
- Leveraged customer data analytics to guide data-driven decision making, identifying high-volume user segments and optimizing physical space utilization.
- Optimized reports in SAS/Excel by removing unnecessary joins, functions, and references, leading to ~10% time reduction.
- Acquired proficiency in SQL and SAS within 2 months, leveraging these tools to execute data analytic tasks.

### ACADEMIC DATA SCIENCE PROJECTS

**Youtube NLP Recommender**

June 2023 - Present

- [Developed](#) an end-to-end solution that integrated with the Youtube API and a website for real-time content viewing and filtering with Natural Language Processing (NLP) techniques.
- Configured an AWS RDS PostgreSQL database for user preferences that integrated with the Streamlit web app.

**NHL Draft Predictions**

May 2023 - June 2023

- [Created](#) a machine learning model to predict NHL draft positions for prospective hockey players and cluster players together into groups based on the nearest neighbor algorithm.
- Harnessed the cutting-edge power of Language Models (LLMs) including Google's BERT and OpenAI's Chat-GPT to encode word embeddings as model features.

### RESEARCH

**Atwood Lab** | Irvine, California  
Undergraduate Researcher

Aug 2019 - Mar 2021

- Implemented a Pytorch graph-based convolutional neural network (CNN) for semi-supervised classification model from a highly cited research article for identifying cancerous cells.
- Researched skin cancer single-cell data using methods for principal graphs and graph neural networks.

**MathBioU & Math EXPLR** | Irvine, California  
Undergraduate Researcher

Jul 2019 - Aug 2019

- Performed computational biology research at a 6-week summer research program.
- Collaborated as a co-author in a research report to present findings to a diverse >15 person audience.

### SKILLS

**Programming:** Python (4 years), SQL (2 years), PostgreSQL, R, SAS, C++, Matlab, Bash, Shell, Unix/Linux Terminal

**Libraries:** Numpy, Pandas, Scikit-learn, PyTorch, TensorFlow, Hugging Face, PySpark, Statsmodels, NLTK, Matplotlib, Seaborn, Altair, Geopandas, Pyscopg2, Streamlit

**Datasets:** Customer, Financial, B2B, Healthcare/Clinical, Geospatial, Causal Inference, JSON, Experiments, Web Scraping

**Tools:** Version Control (Git, Github, DVC), Jupyter Notebook, Business Intelligence (Power BI, Domo), Microsoft Excel, Cloud Infrastructure (AWS S3/EC2/RDS), DBEaver, RStudio, LaTeX