Eric Lim
Data Scientist / Software Engineer

(202) 834-8536 📱
Rosslyn, VA
ml1859@georgetown.edu

Website
GitHub
LinkedIn

SKILLS

Data Science	Python (pandas, numpy, scikit-learn, scipy, tensorflow, pytorch, nltk), SQL (PostgreSQL, MySQL, MongoDB), R (ggplot2, tidyverse)
Data Engineering	AWS, Docker, Kubernetes, Kubeflow, Jenkins, Gradle
Analytics	Tableau, D3, Plotly, matplotlilb, seaborn
Back-end	Java, C/C++, Solidity
Front-end	HTML, JS, CSS
Tools	Version Control (Git, SVN), Jira

PROFESSIONAL EXPERIENCE

Data Engineer   **Jul 2023 — Present**
Freddie Mac   *McLean, VA (Hybrid)*

Software Development Engineer   **May 2022 — Aug 2022**
Amazon   *Nashville, TN (Remote)*
- Deployed data flow of existing prediction models to a new AWS database by creating SQL queries to generate inputs for prediction models into a new database
- Reduced AWS S3 bucket storage needs by 55% by refactoring prediction models to require a shorter time frame for generating forecasts
- Refactored the Java back-end to allow API calls and DJS jobs to handle different data types for better scalability
- Initiated documentation process for technologies set-up to accelerate the onboarding process for future hires

PROJECTS

Data Visualization: Transportation Accidents in the U.S.   **May 2023**
- Created data visualizations of railroad and airplane accidents in the U.S. to analyze the state of the transportation sector
- ✤ **Techniques:** Data Mining, Data Wrangling, EDA

Statistical Analysis: Hans Niemann Cheating Scandal   **Dec 2022**
- Led a statistical analysis on a chess cheating scandal to provide insight on the validity of Magnus Carlsen's allegations against Hans Niemann
- Discovered relevant KPIs in determining a player's legitimacy via EDA and linear regression analysis that shaped the rest of the analysis
- ✤ **Techniques:** Data Mining, Data Wrangling, EDA, Shapiro-Wilk, Kruskal-Wallis, Linear Regression, Multicollinearity tests, Bootstrapping, T-tests

Machine Learning: Gender Equality in the U.S.   **Dec 2022**
- Analyzed gender equality in the U.S. by observing educational, occupational, and societal metrics with supervised and unsupervised machine learning models
- Gathered data from websites, such as Wikipedia and Twitter, and government sources, such as the Bureau of Labor Statistics and the Census Bureau, by creating Python scripts, making API calls, and data mining
- ✤ **Techniques:** Data Mining, Data Wrangling, EDA, Naive Bayes, Decision Trees, Random Forests, SVM, Clustering, ARM, Networking

Media Playlist Simulator   **Dec 2021**
- Implemented a Spotify-like app that stores users' log-in information and respective playlists of songs, movies, TV shows, and podcasts with Python and SQL

ATM Simulator   **Dec 2021**
- Developed the functions of an ATM, such as handling users' banking details, depositing/withdrawing money, and sending spending reports with Java and Python

EDUCATION

Master of Science in Data Science and Analytics, *Georgetown University*, GPA: 4.0   Aug 2022 - May 2024
Bachelor of Science in Computer Science and Statistics, *Georgetown University*, GPA: 3.6   Aug 2019 - May 2023