

Wilson Cheung

Email: info@wilsoncheung.me

Personal Website: wilsoncheung.me

PROFESSIONAL EXPERIENCE

Shutterstock

Dublin, IE

Data Engineer II - Business Analytics Engineering Team

September 2023 - Present

- Developed and implemented caching mechanism for internal web application, resulting in decreased page load times by 90% and enabling new brands to utilize predefined behavioral metrics in their experimentation use cases
- Enhanced Streamlit app template with best coding practices in dependency management and CI/CD deployment; successfully integrated in high-profile applications for database cost and revenue projection analysis
- Supported dbt model migration from MySQL database to Snowflake by collaborating with business intelligence teams to identify upstream dependencies and conducting regression testing to ensure successful data migration

Accenture Federal Services

Arlington, VA, USA

Senior Data Scientist - Applied Intelligence Division

August 2022 - August 2023

- Optimized the CI/CD deployment process for fraud detection models by refactoring code to comply with best practices, leading to increased efficiency and reduced lead times for operations and maintenance
- Facilitated migration of risk modeling production jobs to upgraded database by refactoring SQL queries with nested CTEs, reducing high virtual memory usage by 70% and improving execution times by 55%
- Developed interactive dashboard visualizations featuring counterfactual analysis and algorithmic fairness assessments based on cutting-edge research to offer model recommendations for various client initiatives

Booz Allen Hamilton

Washington, DC, USA

Data Scientist - Senior Consultant in Strategic Innovations Group

April 2019 - December 2021

- Led technical initiatives to enhance scalability in an existing strategic investment project by integrating geospatial data transcoding and implementing advanced spatial preprocessing capabilities
- Contributed to the development of MLOps pipeline by containerizing a custom-built API with GPU support to train reinforcement learning models for simulating agent strategies from PDF documents
- Implemented text analytics vendor solution that reduced Total Cost of Ownership by \$2M over three years for a government client, fulfilling business and technical requirements identified through end-user surveys
- Developed ETL pipelines with Python and shell scripts to process and store 5.5 TB of position-delimited Medicare data in a Postgres database, reducing data ingestion lead time by 90%

Morgan Stanley

New York, NY, USA

Data Scientist - Contingent Hire under Fixed Income Division

October 2017 - March 2019

- Maintained internal application housing 43 GB of proprietary warehouse data by integrating improved data storage API, refactoring legacy codebase, and upgrading software dependencies
- Created an internal data visualization application for 50 TB of proprietary financial data, utilizing Highmaps API to generate statistical aggregations of reported U.S. household incomes
- Developed scripts to validate proprietary pricing models, conduct post-migration data quality checks, and provide analysts with daily real-time results through automatically generated email reports

SKILLS

Technical: Python (numpy, scipy, pandas, polars, scikit-learn, xgboost, PyTorch), dbt, SQL (Snowflake, Postgres, MySQL, MongoDB), API Development (FastAPI, Flask), Docker, UNIX, Git Version Control, GitHub Actions

Quantitative: Machine Learning (classification, regression, clustering), Experimental Design (A/B Testing, Multi-Armed Bandits), Probability and Statistics (confidence intervals, hypothesis testing), Data Mining

EDUCATION

Georgia Institute of Technology — Master of Science in Analytics

December 2021

University of South Carolina — Master of Science in Mathematics

December 2015

State University of New York at Geneseo — Bachelor of Arts in Mathematics

May 2013