SHERRY DUONG

408-442-8444 | sherryduong@gmail | in/sherryduong | sherryduong93

TECHNICAL QUALIFICATIONS

Data Science Applications:

- Analytics: Data Visualization, Statistical Inference, Hypothesis Testing, Sampling Methods, A/B Testing
- Machine Learning: Linear/Logistic Regression/Classification, Regularization: Ridge & Lasso, Decision Trees, Random Forest, Neural Networks (RNN/CNN), Natural Language Processing, Time Series, Recommender Systems, Graphs

Programming & Analytics:

- Python Libraries: NumPy, Pandas, Scikit-Learn, Tensor-Flow, NLTK, Matplotlib, Seaborn, Beautiful Soup, Flask
- Database Management: Postgres & Psycopg2, Microsoft SQL Server, Spark & PySpark, MongoDB & Pymongo
- Additional Technologies: Microsoft Excel, Tableau, Amazon AWS (Sagemaker, EC2, S3), Docker, Heroku, Netsuite

DATA SCIENCE PROJECTS

Analyzing the Chase Center Impact on Crime in Dogpatch/Mission Bay -

- Conducted hypothesis testing with python using Welch's T-test and Mann Whitney U-test to confirm whether or not the opening of the Chase Stadium had a statistically significant impact on crime rates in the Dogpatch/Mission Bay areas.
- Additional Tools/Libraries Used: Numpy, Pandas, Scipy, Matplotlib

Predicting AirBnB Listing Prices in San Francisco - |

- Analyzed current Airbnb daily listing prices in San Francisco to predict prices of future listings using Linear Regression, Random Forest, and Gradient Boosting. Performed feature engineering and hyper-parameter tuning, eventually achieving 75% improvement on RMSE for final RMSE of $54 on cross-validated data.
- Additional Tools/Libraries Used: Numpy, Pandas, Matplotlib, Sklearn, NLTK

Your Anime Match Maker: An Anime Recommender - | |

- Created a recommender system to recommend anime based on popularity, content similarity (measured using cosine similarity), and matrix factorization collaborative filtering. Used Spark's ALS collaborative filter model and achieved 1.13 RMSE on predictions of 1.3M user ratings (from 1-10). Web app created with Flask and is live on Heroku.
- Additional Tools/Libraries Used: Numpy, Pandas, Matplotlib, Sklearn, Spark ML, AWS Sagemaker

PROFESSIONAL EXPERIENCE

Galvanize Inc.

Data Scientist in Residence — 2020 - Present

- Instructed 18 data science students on topics including python programming, machine learning, and statistical inference.
- Innovated on curriculum material, providing targeted content in order to reinforce retention of program materials.

Personalized Beauty Discovery Inc. (IPSY)

Business Operations (eCommerce | Supply Chain | Logistics | Inventory Management) — 2017 - 2020

- Analyzed inbound shipments annually to identify opportunities with quality control, reducing cost by 40% in one year by implementing quality control procedures and standards for squeeze tube inspection.
- Created a monthly report template using Microstrategy and Excel to streamline inventory movements for the subscription program. This resulted in a significant reduction of errors during the monthly process.
- Optimized supply chain operations through scrutiny of data: identifying risks and bottlenecks in the supply chain process among 200+ SKUs, 20+ global suppliers, and 20+ internal team members.

EDUCATION

University of California: Berkeley: Bachelor's of Science: Business Administration — 2011 - 2015

Galvanize Data Science Immersive: 600+ hours of Data Science applications/techniques/tools — 2020

- Maximized expected profit of $200K for a ticket sales and distribution company by identifying fraudulent users using machine learning and cost-benefit matrix. Achieving an AUC of 99.4% with Random Forest Classifier.