Jun Hwang

jun.hwang@berkeley.edu | +1 949-299-6164 | Berkeley, CA | juns-hwang.github.io | Linkedin |Github

Education

University of California, Berkeley Graduation Date: Dec 2024
B.A. in Data Science (Concentration: Economics)

Relevant Coursework: Principles and Techniques of Data Science, Data Engineering, Data Mining and Analytics, Probability for Data Science, Data Structures and Algorithms, Econometrics, Economic Analysis - Micro, Economic Models

Experience

Atlassian May 2024 - Aug 2024
Incoming Data Science Intern *San Francisco, California*

Quizlet Aug 2023 - Present
Data Science Intern *San Francisco, California*

- Designed and implemented unified data pipeline and dashboard for all marketing A/B test analytics, reducing the monthly cost of the process by 75%
- Enhanced Quizlet's internal statistical testing library by developing customized Python functions for A/B testing methodology, speeding up A/B test process by 15%
- Developed data pipeline model for year-in-review recap email notification updating 20+ millions active users' product usage metrics using BigQuery and Airflow

Fremont Bank May 2023 - Aug 2023
Data Analyst Intern *Livermore, California*

- Analyzed historical IT service dataset consisting of 10k+ records by developing an interactive dashboard that outlines major trends using Microsoft Excel and PowerBI
- Proposed data-driven recommendations to reduce annual ticket volumes by 45%
- Enhanced IT support process by 20% by automating ticket categorization utilizing Microsoft Excel VBA

Callisto Feb 2023 - May 2023
Data Science Researcher *Berkeley, California*

- Conducted exploratory data analysis on social media datasets using Python pandas, numpy, and plotly for trend visualization
- Developed an interactive dashboard to visualize company's annual growth across various US regions using Microsoft Excel
- Performed a t-test on the campus ambassador program's efficacy to ensure the accuracy of the statistical conclusions

Projects

Validating Youtube Category (Python) Nov 2023 - Dec 2023

- Processed and transformed the dataset containing US Youtube videos' statistics by one-hot encoding, vectorizing and normalizing data using Python pandas and scikit-learn
- Applied K-means clustering and elbow method using scikit-learn to identify whether the optimal number of categories match current number of categories
- Utilized Google Gensim Word2vec model to find which current category pairs yield high similarity scores

Query Performance (PostgreSql) Sep 2023 - Oct 2023

- Examined the effects of views, predicate pushdowns, joins, and index managements on optimization and tuning for SQL Queries using PostgreSql

Housing Price Predictor (Python) Oct 2022 - Nov 2022

- Conducted EDA on a dataset of 500k+ housing records in Cook County, Illinois using Python pandas, numpy, seaborn
- Applied linear regression on feature-engineered data set with one-hot encoding using Python scikit-learn - achieved accuracy of 90%

Technical Skills

Languages: Python, SQL, NoSQL (MongoDB), R, VBA, Java, Html, CSS, LaTex
Libraries: NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, PyMongo, Scipy, Plotly, Ggplot2, Dplyr
Tools: Google BigQuery, dbt, Airflow, Tableau, Microsoft PowerBI, Git, Microsoft Excel, Jupyter