Junseok 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
Bachelors in Data Science (Concentration: Economics)

Relevant Coursework: Principles of Data Science, Intro to Machine Learning, Data Engineering, Data Mining and Analytics, Intro to Probability, Data Structures and Algorithms, Econometrics, Economic Models

Experience

Atlassian   May 2024 - Aug 2024
Data Science Intern   *San Francisco, California*
- Aligned FY25 OKR for Atlas Teams, utilizing Databricks SQL and PySpark
- Developed a linear regression forecast model of Atlas Teams Usage with Python scikit-learn
- Boosted usage predictions by 25% by analyzing Atlas usage across 10+ segments and identifying key target customers
- Created live interactive Tableau dashboards that highlight major trends of Atlas Teams usage, reducing analysis time by 20%

Quizlet   Aug 2023 - March 2024
Data Science Intern   *San Francisco, California*
- Designed and implemented unified data pipeline and dashboard for all marketing A/B test analytics utilizing Google BigQuery, dbt, and Periscope, reducing the cost of the marketing A/B test process by 75%
- Developed data pipeline model using BigQuery for year-in-review recap email and in-app notification updating 20+ millions active users' product usage metrics
- Built methods for multivariate A/B tests for Quizlet's stats library, speeding up A/B test process by 15%

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 analysis on social media datasets using Python pandas, numpy, and plotly
- Performed a t-test on the campus ambassador program to determine effectiveness

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

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 technique to predict housing price after feature-engineering 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, PySpark, Scipy, Plotly, Ggplot2, Dplyr
Tools: Google BigQuery, Tableau, dbt, Airflow, Microsoft PowerBI, Git, Databricks, Microsoft Excel, Jupyter