Johnny Yu

Johnnyyu14@gmail.com | 415-312-2772 | johnnyyu14 | johnnyyu14.github.io

Education

University of California Santa Barbara Santa Barbara, CA

Bachelor of Science in Statistics and Data Science *September 2018 – June 2022*

Awards: Dean's Honors L&S, Awarded $5000 to participate in UCSB's NSF-funded Central Coast Data Science Fellowship for the 2021-2022 academic year

Skills

Programming Languages: Python, SQL, R, SAS, HTML, CSS, Excel

Visualizations/Libraries/Analysis: Tableau, Power BI

Work Experience

Live Proxies Remote

Business Analyst *March 2020 – Present*

- Generated $1.5M in revenue for the fiscal year 2021, boosted company revenue by 30%, and traffic flow by 50% Year-Over-Year by identifying key market opportunities
- Decreased operating costs by 15% Year-Over-Year by implementing a restructured set of procedures for payment collection
- Pioneered a business plan encompassing overarching sale, growth, and marketing strategies in order to capture the target audience
- Utilized SQL and Power BI to analyze business processes and identifying improvement opportunities for business growth
- Liaised business relationships with large distributors and stakeholders such as Cox, CenturyLink, and CharterSpectrum

ENVENT Labs Santa Barbara, CA

Data Analyst *September 2021 – June 2022*

- Analyzed over 2500 fragmented and legacy data sets to explore how demographic attributes are predictors of climate opinion in different countries and time periods
- Doubled the speed of SQL queries used for data extraction by implementing efficiencies and developed complex supervised and unsupervised ML models to make predictions of an individual's climate score.
- Utilized Excel, SQL, and Power BI for data entry, processing, cleaning, visualizations, and analysis to help create data backed reports
- Presented the results of analyses to numerous technical and non-technical stakeholders to encourage data-driven decision making.

Projects

Forecasting Short-Term Future COVID-19 Cases *September 2021 - December 2021*

- Created two time-series forecasting models that predicts the number of new daily COVID cases trained on the estimated percentage of COVID related outpatient doctor visits
- Identified Decision Tree Regressor performed 52% better than SVR model with the accuracy score being able to capture the overall trend in the rise and fall of new cases
- Evaluated that neither model accurately predicted the number of new daily COVID cases using historical data due to wide range of values of the dataset and the features not being good predictors of new COVID cases

2016 Presidential Election Prediction Model *January 2021 - June 2021*

- Explored various models to answer whether or not Donald Trump would win a county based on demographic variables
- Examined that Random Forest had a 0.8% higher accuracy score compared to Logistic Regression. Using the optimal threshold for Logistic Regression, the true negative rate increased but total misclassification rate increased to 19%
- Detected that Trump won around 87% of the counties in the test data. Which implies if the misclassification rate in a model was above 13%, there is no improvement using a model to predict election results