Yi Heng Joshua Wu

+1-650-922-2084
joshuayhwu@gmail.com
/joshuayhwu
joshuayhwu.github.io

Education

University of California, Berkeley 2018 – Present
BA Data Science and Economics, CGPA: 3.86 *Berkeley, CA*

- Data Structures, Structures & Interpretation of Programs, Principles & Techniques of Data Science, Applied Econometrics, Research & Data Analysis, Behavioral Economics, Probability for Data Science

Projects

COVID-19 Mental Health Analysis Apr. 2021 – May. 2021

- Estimated mental health impacts of COVID-19 on US demographics based on CDC data
- Constructed *Classification Models* with *Random Forests* to enhance model accuracy up to 80%+
- Dimensionality reduction with *Principal Component Analysis* to find significant factors affecting wellness

World Happiness Analysis Dec. 2020 – Jan. 2021

- Examined the relationship between country's happiness index and macroeconomic variables
- Utilized *Spearman's rank correlation coefficient, Logistic Regression* for ordinal and binary variables
- Positive associations for life expectancy, social support, gdp; negative for urban population proportion

Airfare Analysis Jun. 2020 – Jul. 2020

- Modelled the relationship between number of passengers, airfare, seasonal fixed effects, and more
- Compared serial correlation treatments with *Cochrane–Orcutt, Prais–Winsten, and Newey-West estimators*
- Utilized *Chow Test* to evaluate effectiveness of the current models

Social Media Analysis Oct. 2019 – Dec. 2019

- Examined the relationship between social media usage, extroversion, and depressive symptoms
- Conducted multivariate regression analysis on R with the collection of *tidyverse* packages
- Effect of extroversion on depressive symptoms, indicating the former as a moderator on social media usage

Research

Overconfidence in Amazon Rekognition

- Utilizing *AWS Lambda* and *S3* with the *Rekognition API* to generate object labels for 1500+ images to test for overconfidence against MTurk reviewers

Overconfidence in Google Vision

- Employed *Google Cloud's Vision API* to illustrate overconfidence displayed by 9000+ objects

Prior vs. Desires

- Employed *K-Means Clustering* to understand how subjects behave against *Bayesian Belief Update*
- Replicated *regression with analytic weights* and *clustered standard errors* by subjects

Skills

Tools Python, Java, SQL, R, Stata, Tableau, Excel VBA Macros, LaTex, HTML5, CSS, JavaScript
Libraries Pandas, numpy, regex, tidyverse, scikit-learn, scipy, statsmodel, matplotlib, seaborn, plotly, d3
Languages English, Mandarin, Cantonese