



**NYC Restaurants Health Inspection Analysis Project | Project Page***Class Final Project for DSCI 522, UBC***Python, R, Docker, Git**

2022 W I

- Designed the down-sampling approach for binary classifications to mitigate imbalanced data, aligning with our research objectives.
- Evaluated the performance of logistic regression and SVMs (balanced vs. imbalanced) with hyperparameter tuning and classification report analysis (including PR and ROC curves), leveraging data from NYC Open Data (achieving a **0.975 F1 test score**).
- Collaborated in developing automated pipelines in both Docker image and Python Makefile (with a full dependency diagram).

**ML Benchmarking Models Research Project – The Epic Games Ratings | Project Page***Research Assistant at Epic Games, Inc.***Python, R, MongoDB, L<sup>A</sup>T<sub>E</sub>X, Git**

11/2021 - 03/2022

- Categorized users in restricted game data as node classifications on graphs and trained Graph Convolutional Networks (GCNs) and Multi-layer Perceptron (MLP) models using Nvidia RAPIDS cuML to predict user age categories (Minors, Adults, Middle-aged).
- Trained a stacking ensemble of KNN and bootstrap random forest algorithms in Python scikit-learn with finely tuned hyperparameters, achieving better game rating predictions with **0.87 RMSE** and a **12% improvement** over the benchmark model.
- Implemented a gaming recommendation system that leverages users' game-playing histories, utilizing a weighted average of hybrid approaches combining unsupervised collaborative filtering and supervised content-based filtering.

**QGIS LA Local Market Analysis Project for Starbucks | Project Page***Class Final Project under Dr. Michael Shin, UCLA and the Undergraduate Business Society, UCLA***Python, R, SQL, QGIS, L<sup>A</sup>T<sub>E</sub>X, Git**

2017 Fall

- Analyzed the effects of population density, the presence of colleges and shopping centers within a 1-mile buffer, and income distribution on the QGIS Geo-distribution of Starbucks in LA, using the Python SQLite database from the US Census Bureau.
- Designed a linear mixed-effects model to analyze Starbucks revenues in ten US cities, estimating both nationwide and local effects.
- Developed a multi-layer Artificial Neural Network ANN model from scratch, **exclusively using basic NumPy and SciPy only**, for multi-class classification (achieving **0.97 test accuracy**).

**II. Finance Oriented Projects:****BlackRock Troubled Waters ESG Investment Research Project | Project Page***Research Assistant at BlackRock, Inc. under Dr. Yuxi Suo***Python, R, SQL, QGIS, L<sup>A</sup>T<sub>E</sub>X, Git**

03/2020 - 07/2020

- Utilized Python and QGIS to develop a dynamic, multivariate 3D mapping for visualizing 2014-2020 global water shortages.
- Conducted a three-level quantile analysis on US REITs with low, middle, and high returns in 2014, using water allocation and scarcity stress panel data from Bloomberg, WRI, the United Nations World Water Report, and the SNL database for REIT properties.
- Investigated the implications of multivariate time series predictions for 2030 on the US REITs market and water stress level proportions, based on sensitivity benchmark results from deep learning memoryless 2D CNN, short-memory RNN, and LSTM.

**US Finance Studies - The Effectiveness of the Big Four Auditors on Restatement Errors | Project Page***R Workshop Project under Dr. Ronghui (Lily) Xu, UCSD***Python, R, L<sup>A</sup>T<sub>E</sub>X**

2017 Spring

- Manipulated and classified the 2008-2016 internal audit data and US SEC multivariate raw panel data by firm industrial sector types and restatement error types, and applied STL decomposition and SARIMA to extract patterns of seasonality and autocorrelation.
- Confirmed the positive impact of the Big Four auditors in reducing the likelihood of committing restatement errors for their client firms, using binary logistic regression to compare firms with Big Four versus non-Big Four auditors.

**US Stock Portfolio Return Analysis on SIC Industry and Market Capitalization | Project Page***Class Final Project under Dr. Gordon B. Dahl, UCSD***R, Pine Script, L<sup>A</sup>T<sub>E</sub>X**

2016 Spring

- Visualized time series US stock trends and indicators from 2008 to 2015 by implementing GMMA, Bollinger Bands, moving average systems, and the reversal "9" indicator from DeMARK Analytics in Pine Script, coded from scratch on TradingView.
- Calculated and reported the five-year annualized compounded returns of three selected stock portfolios from 2011 to 2015 in R, based on different SIC first digits, volatility, and market capitalization rankings.

**III. Miscellaneous Projects:****US Labor Mobility Patterns Analysis | Project Page***Class Final Project under Dr. Abel Valenzuela and Dr. Kent Wong, UCLA***Python, R, L<sup>A</sup>T<sub>E</sub>X**

2018 Fall

- Conducted exploratory data analysis (EDA) using lagged variables in Python and R to compare movements across different regions and urban areas, utilizing panel data from the WBIS database and the US Bureau of Labor Statistics NLSY 79 survey.
- Analyzed and estimated the wage gap between movers and stayers across four selected regions and in urban/non-urban areas using plots and OLS regressions in R, controlling for factors such as age, employment, education, and gender.

**US GDP Stata Modeling Analysis***Class Final Project under Dr. Carroll B. Foster and Dr. James D. Hamilton, UCSD***Stata, R, L<sup>A</sup>T<sub>E</sub>X**

2016 Summer

- Tested both the Lagged Regressor Model and the Lagged Dependent Variable Model in Stata and R, conducting annual GDP regressions on the money supply, labor force, and government purchases using 1973-2013 data from the US Federal Reserve.

**UCSD Healthcare Labor Studies on Overtime Work Analysis for ER Physicians | Project Page***Class Final Project under Dr. Yuan Emily Tang, UCSD***Python, R, L<sup>A</sup>T<sub>E</sub>X**

2016 Spring

- Inspected data entry errors and preprocessed the data using POSIXct standard time format, focusing on patient flow and physician shift schedules in the emergency departments (ED) of UC San Diego Health Center.
- Analyzed ggplots depicting hourly patterns of patient arrivals and patient severity levels, and estimated the likelihood of ED physicians working overtime using binary logistic regression with four regressors.

## TEACHING EXPERIENCE

---

<b>Supplemental Student Self-Organized Workshop for DSCI 525 Web and Cloud Computing, UBC</b> <i>Google Cloud Platform Workshop &amp; Demo (80%+ of MDS cohort RSVP)</i>	2022 W II
<b>Student Services Workshop for Statistics, UBC</b> (medium post last updated on 10/15/2023) <i>Instrumental Variables Estimation Workshop (150+ students RSVP at UBC)</i>	2022 W I
<b>Math 115A Linear Algebra, UCLA</b> <i>Grading Reader (5.00/5.00 assessment review from the TAs and the professor)</i>	2018 Fall
<b>Coding in Finance Workshop Series hosted by the Undergraduate Business Society, UCLA</b> <i>Python Algorithms Workshop Series (120+ students enrolled at UCLA)</i>	2018 Spring
<b>Math 33B Differential Equations, UCLA</b> <i>Math Tutor (97.8% student recommendation)</i>	2018 Winter
<b>Math 181A Introduction to Mathematical Statistics I, UCSD</b> <i>R Code Workshops (90%+ of Math 181A students enrolled)</i>	2017 Spring
<b>Math 170A Introduction to Numerical Analysis, UCSD</b> <i>MATLAB Tutor (93.7% student recommendation)</i>	2017 Winter
<b>Econ 120B Econometrics B, UCSD</b> <i>Stata Tutor (96.1% student recommendation)</i>	2016 Fall
<b>Math 20D Introduction to Differential Equations, UCSD</b> <i>MATLAB Tutor (91.6% student recommendation)</i>	2016 Summer I

## LEADERSHIP EXPERIENCE & ACTIVITIES

---

<b>Volunteer at the PwC x Microsoft Transformational AI Conference</b>	06/22/2023
<b>Contributor of the Women in Data Science Conference, Vancouver 2023 hosted by Microsoft &amp; DataCan</b> <ul style="list-style-type: none"><li>• Interacted closely with guest speakers and sponsors from companies such as Microsoft, Nvidia, and Google.</li><li>• Collaborated in organizing Google's IamRemarkable movement event and contributed to its extension to SheIsRemarkable.</li></ul>	04/14/2023
<b>Volunteer at the Pacific Conference on Artificial Intelligence at UBC</b>	04/02/2023
<b>Member of the DataCan Network, Vancouver</b>	08/2022 - Present
<b>Member of the UBC Japan Association</b>	08/2022 - Present
<b>Contributor of the Wells Fargo Women in Leadership Event, San Francisco</b>	09/03/2021
<b>Committee of the Undergraduate Business Society, UCLA</b> <ul style="list-style-type: none"><li>• Interacted closely with US domestic sponsors and global clients from EY, Cornerstone Research, Accenture, etc.</li><li>• Led the runner-up team "Bruin Ace" in the Merrill Lynch Private Equity LBO &amp; Merger Model Case Competition.</li><li>• Managed a team to organize the UCLA 2018 Financial Services Night career fair.</li></ul>	09/2017 - 12/2018
<b>Semi-marathon Runner</b>	01/2019 - Present
<b>Member of the Japanese Student Association, UCLA</b>	09/2017 - 12/2018
<b>Member of Local Badminton Club</b>	09/2015 - Present