

Yunjun Xia

(858) 666-5306 | yx2569@columbia.edu | linkedin.com/in/yunjunxia | github.com/xiaiyunj
485 Marin Blvd, Apt 1103, Jersey City, NJ 07302

EDUCATION

Columbia University, Columbia Engineering, New York, NY

Expected Dec 2020

Master of Science in Data Science (Current GPA: 4.0/4.0)

Coursework: Probability & Statistics for Data Science; Exploratory Data Analysis & Visualization; Algorithms for Data Science; Elements of Data Science; Machine Learning; Applied Machine Learning; Statistical Inference & Modeling; Databases

University of California San Diego, Revelle College, La Jolla, CA

Sep 2015 – Mar 2019

Bachelor of Science in Probability & Statistics with Minor in Economics (GPA: 3.83/4.0)

Coursework: Mathematical Statistics; Computational Statistics; Stochastic Processes; Combinatorics; Time Series; Bootstrap Analysis; Real Analysis; Actuarial Mathematics; Mathematics of Finance; Microeconomics; Macroeconomics; Econometrics

Honors: Provost Honor in Freshman year, Fall 2016, Winter 2017, Winter 2018, Spring 2018

SKILLS & TECHNOLOGIES

Programming Languages R, Python, Java, SQL, NoSQL, Swift, Matlab, Stata

Programming Packages Python: Pandas, Numpy, Scikit-Learn, Matplotlib, Seaborn, Keras, TensorFlow; R: tidyverse, dplyr, ggplot2, Shiny

Cloud Services Google Colab, Google Cloud Platform, AWS

Databases MySQL, MySQLWorkbench, Neo4j

Machine Learning Regression, Classification, SVM, Random Forests, Gradient Boosting, Clustering, NLP, Neural Networks

Web Technologies JavaScript, D3

PROFESSIONAL EXPERIENCE

Silver Cell, New York, NY

Jul 2020 - Present

Internship, Summer Startup Track, Columbia Business School

- The project is focused on building a mobile application which aims to make technology more accessible to senior citizens;
- Built a mobile application prototype using Invision Studio and developed the application using Xcode - Swift.

China Everbright Bank, Chengdu, China

Aug 2017 - Sep 2017

Internship, Electronic Banking Department, Chengdu Branch

- Collected and published daily articles about the branch's main products in season on CEB Chengdu WeChat Official Account;
- Improved the contents of articles by analyzing the popular previous ones, which increased the page reading from 200 to 800 on average;
- Signed QR Code payment agreements with 30+ business owners and analyzed the information of business owners.

PROJECT EXPERIENCE

Summer Invitational Datathon: Regional Economic Impact on Host Countries from Olympic Games (2010, 2012, 2016)

Jul 2020

Group Project, The Data Open, Citadel

- Analyzed regional GDP, income, unemployment rate and tourism data for the UK, Canada and Brazil before and after hosting the Olympics;
- Concluded that there was no significant difference in regional economic impacts of hosting the Olympics by ANOVA tests and visualization;
- Published the report with R Bookdown and created a Datafolio for data storytelling and visualization.

Python project: Machine Learning - Regression Model Evaluation

Spring 2020

Individual Project, Columbia University

- Evaluated Ridge, Lasso and Elastic Net on price prediction using cross-validation for the Sydney housing dataset on Kaggle;
- Preprocessed categorical and continuous variables with One Hot Encoder and Standard Scaler respectively, which formed a pipeline;
- Tuned parameters for the resulting models by grid search which improved the mean score for Elastic Net from 0.58 to 0.76.

SQL project: Databases of Game of Thrones

Spring 2020

Individual Project, Columbia University

- Created a new schema on MySQLWorkbench with 10 data tables of Game of Thrones and built up an entity-relationship model;
- Cleaned up the tables by altering column Data Types, setting Not Null constraints, adding Primary/Foreign keys and defining Check constraints;
- Developed a view of interactions between characters in the same scene, where Daenerys and Drogon interacted the most with 172 scene counts.

Data Visualization R project: 2019 Airbnb Analysis - New York City

Fall 2019

Group Project, Columbia University

- Visualized distributions of Airbnb room types in NYC by boroughs using stacked bar chart and Google Maps Platform;
- Analyzed Airbnb prices by NYC Community Districts and transportation convenience using choropleth map, histogram, and violin plot;
- Extracted emotional words expressed in customer comments over years to compare with customer review scores;
- Created interactive Shiny heat maps of price and rating scores, for which Midtown Manhattan has the highest price but the lowest ratings.