Yunjun Xia

(858) 666-5306 | yx2569@columbia.edu | linkedin.com/in/yunjunxia | github.com/xiayunj
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 **Mar. 2019**
Bachelor of Science in Probability & Statistics with Minor in Economics (GPA: 3.83/4.0)
Coursework: Mathematical Statistics; Computational Statistics; Stochastic Processes; Probability; Combinatorics; Time Series; Bootstrap Analysis; Linear Algebra; Real Analysis; Actuarial Mathematics; Mathematics of Finance; Differential Equations; Microeconomics; Macroeconomics; Econometrics; Decisions Under Uncertainty
Honors: Provost Honor in Freshman year, Fall 2016, Winter 2017, Winter 2018, Spring 2018

PROFESSIONAL EXPERIENCE

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;
- Classified and analyzed the information of business owners which signed the payment agreement with the branch.

PROJECT EXPERIENCE

Python project: Machine Learning - Regression Model Evaluation **Spring 2020**
Individual Project, **Columbia University**
- Visualized the dependency of the target on each continuous feature for Sydney housing dataset;
- Preprocessed both categorical and continuous variables by using a pipeline;
- Evaluated OLS, Ridge, Lasso and Elastic Net using cross-validation and tuned parameters using grid search;
- Analyzed the 20 most important coefficients of the resulting models.

SQL project: Databases of Game of Thrones **Spring 2020**
Individual Project, **Columbia University**
- Created a new schema on MySQLWorkbench and imported Game of Thrones data tables;
- Cleaned up the tables using SQL queries and built up an entity-relationship model;
- Developed views to analyze interactions between characters in the same scene in Game of Thrones.

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 Map 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;
- Used R packages Leaflet and Shiny to create interactive heat maps of price and rating score.

R project: Bootstrap methods on data analysis **Spring 2018**
Individual Project, **UCSD**
- Tested the mean of data samples with known variance through Bootstrap method;
- Computed the bootstrap pivotal confidence interval for the mean of data samples;
- Constructed the bootstrap studentized pivotal confidence interval for the mean of data samples.

Java application: Game 2048 **Winter 2018**
Individual Project, **UCSD**
- Built up a mathematical game 2048 based on Java;
- Coded and implemented 2048 backend using Vim editor;
- Constructed 2048 GUI for the fully functioning graphical 2048 game.

SKILLS & HOBBIES

- Programming language: R, Python, SQL, Java, Matlab, Microsoft Office, Stata.
- Hobbies: League of Legends; Baking; Cooking.