

Peixuan Song

ps3193@columbia.edu | (646)407-5291 | 70w 95thst, New York, NY, 10025

EDUCATION

Columbia University	New York, NY
M.A. in Statistic (Expected). GPA: 3.75/4.00	Sep 2021 - Dec 2022
Shandong University	Shandong, China
M.S. in Artificial Intelligence (Expected). GPA: 87.1/100	Sep 2020 - Jun 2023
Shandong University	Shandong, China
B.S. in Statistics (Finance). GPA: 85.75/100. (Top 10%, Class Monitor)	Sep 2016 - Jun 2020

SKILLS

Programming language: Python, R, SQL, MATLAB, C, C++

Tools: Tableau, Jupyter Notebook, R Studio, PyCharm, LaTeX, Microsoft Excel, Linux, GCP, Spark, Azure

WORK EXPERIENCE

Columbia University	New York, NY
Data Analyst (Part-time)	Mar 2022 - present
<ul style="list-style-type: none">◆ Compiled over 1,000 respondents' feedbacks from the Columbia Engineering Graduate Quality of Life Survey, processed data using Python and chose valid data with response time greater than 200s.◆ Visualized charts that identified statistical trends of students' satisfaction level and significance level of school services and sources. Gave recommendations and advocated for meaningful and lasting change for Columbia Engineering student to Student Council.	
CCB Life Insurance	Shandong, China
Sales Analyst Intern	Apr 2020 - Jun 2020
<ul style="list-style-type: none">◆ Extracted over 1,000 daily insurance sales data of different city-branches in Shandong Province from the company database using SQL, reported and summarized the sales data for each city-branch.◆ Created pivot tables and bar plots weekly to analyze the performance of tens of products in Excel.◆ Built advanced formula to realize the financial regulation terms for pricing insurance and salary calculation of product managers based on the profits of each city-branch.	
Inspur Corporation	Shandong, China
Analyst Intern	Jun 2018 - Aug 2018
<ul style="list-style-type: none">◆ Collected the sales data of cloud services from company OA using Excel, assisted the department in calculating the revenue of different state-owned enterprises.◆ Helped created dashboards and presentations to display the weekly KPIs, implemented meetings and improved the department work efficiency.	

PROJECT EXPERIENCE

Yelp Restaurants Analysis and Star Rating Prediction	Mar 2022
<i>Columbia University, Department of Statistics</i>	
<ul style="list-style-type: none">◆ Visualized different features of the 35,000 currently opened restaurants from Yelp. Conducted basic text mining and plot review word cloud of different ratings.◆ Performed a 5-level stars rating prediction of restaurants using review counts, opening hours and categorical features for random forest regression with 62.56% accuracy.◆ Refined model by using gradient boosting classifier and improved the accuracy to 70.56%. Conclude the most important features, and gave suggestions on the restaurant operation.	
R Shiny App for the Open Status of NYC Parks and Facilities under COVID-19	Feb 2022
<i>Columbia University, Department of Statistics</i>	
<ul style="list-style-type: none">◆ Deployed an R Shiny app including both interactive map and static map using google map API for visualization of the open status of park facilities in NYC.◆ Investigated on the park crime numbers in 5 different boroughs and used R for regression and time series analysis of last 5 years.◆ Gave guidance for New Yorkers to choose when and which park facilities to go after COVID-19.	
Predictive Modeling for Image Classification for Semi-supervised Image	Feb 2022
<i>Columbia University, Department of Statistics</i>	
<ul style="list-style-type: none">◆ Carried out a 3-layer CNN image classifier model for noise data using Python on an imbalanced image data with noise labels generated from CIFAR-10 with 65% accuracy.◆ Applied "label-correction" approach for weakly-supervised data before feeding the training set to the predictive model. Optimized the model accuracy to 70% by refitting the model with corrected label and accelerated the training speed by 10%.	