

# Xingyin Xu

201-978-7301 | [xingyinxu@gmail.com](mailto:xingyinxu@gmail.com) | [www.linkedin.com/in/xingyin-xu](https://www.linkedin.com/in/xingyin-xu) | <https://yinyin0916.github.io/>

## EDUCATION

### New York University

*Master of Science in Data Science*

GPA: 4.0/4.0 (top 2%)

Relevant Courses: Big Data, Visualization for Machine Learning, Machine Learning, Natural Language Processing.

New York, NY

Aug. 2023-May. 2025 (Expected)

### University of California, San Diego

*Bachelor of Science in Probability and Statistics, Minor in Data Science and Economics*

GPA: 3.98/4.0 (top 2%)

Relevant Courses: Data Analysis and Inference, Data Structures, Data Management, Machine Learning, Recommender Systems, Mathematical Statistics, Stochastic Processes, Mathematics of Finance, Econometrics, Linear Algebra.

San Diego, CA

Sept. 2019-Jun. 2023

## SKILLS

**Programming Languages:** SQL, Python, Excel, Java, R, VBA.

**Packages:** Pandas, Numpy, Scikit-learn, Seaborn, MongoDB, NLTK, matplotlib.

**Statistical Modeling:** Logistic Regression, SVM, Decision Tree, Random Forest, K-Means, Cross-Validation, PCA.

**Data Analysis Tools:** AWS, ETL, Tableau, PowerBI, Google Analytics, Microsoft products.

## WORK EXPERIENCE

### China Construction Bank

*Summer Intern, Data Analyst*

Suzhou, China

Jul. 2023-Aug. 2023

- Developed and implemented a **VBA-based automated system** for daily deposit reports, enhancing the detection of anomalies and inconsistencies. This tool streamlined the process of report generation, reducing manual workload by approximately **50%**.
- Created and deployed an interactive data dashboard using advanced visualization techniques (line and pie graphs, slider controls) in **Tableau**. This improved customer relationship management by providing clear, actionable insights, leading to a **40%** increase in operational efficiency.
- Evaluating **key performance indicators** for various company departments, aiding in strategic decision-making and performance assessment.

### Clear Creek Capital

*Remote Summer Intern, Data Analyst*

Los Angeles, CA

Jun. 2022-Sept. 2022

- Spearheaded a data scraping project to collect 50,000+ data points on crude oil, gold, and treasury bonds. Applied robust intermarket analysis during different macroeconomic cycles, refined profitability models and investment strategies.
- Updated existing datasets with **SQL queries** for market sentiment reports, optimizing join operations and utilizing window functions to lower runtimes and costs.

## PROJECTS

### Movie Ratings Analysis

*Applied Statistical Tests and Machine Learning to Analyze Ratings and Raters*

New York, NY

Oct. 2023-Dec. 2023

- Utilized **Mann-Whitney U tests** and found **significant** rating difference on engaged and not engaged audience.
- Built **Linear Regression Model** to predict movie ratings and utilized **grid search** to find best parameter for **Lasso regularization**.
- Examined quality consistency of franchise movies using **Kruskal-Wallis H-test**, and using consistent franchise movies ratings to build regression model. Improved R square by more than **150%** compared to the baseline model.

### Data Analyst/Lab Assistant

*CSEd at University of California, San Diego*

San Diego, CA

Sep. 2022-Jun. 2023

- Utilized **SQL** to extract usage data from the application's database, and collected usage data from **Google Analytics**.
- Analyzed feature usage data's correlation with users' course performance using **Pandas**, line graphs, and t-tests. Revealing a significant different for students using the quizzing feature.
- Constructed and composed experience report paper, manuscript submitted to SIGCSE TS 2023.

### Predicting Heart Disease from Health Information

*Applied Machine Learning to Predict Heart Disease*

San Diego, CA

Mar. 2022-Jun. 2022

- Built multiple **classification** machine learning models including SVM, Logistic Regression, Decision Tree.
- Analyzed the ROC-AUC graph and learning curve graph and adjusted the weight of positive and negative class, employed **Stratified KFold** to ensure proportion of positive and negative class.