

Xu Han

 xu.han@nyu.edu |  xh852 |  Xu Han

EDUCATION

New York University, Center for Data Science, New York, NY
Master of Science in Data Science

Expected 2023

New York University, College of Art and Science, New York, NY
Bachelor of Arts in Mathematics and Economics

May 2020

SKILLS

Programming Languages/Tools: Python (NumPy, Pandas, Scikit-Learn, Jupyter notebook, Geopandas, Plotly), SQL (MySQL, PostgreSQL), Dash, R, Tableau, MATLAB, C++, Java, Excel, Google sheets, Git.

Relevant Courses: Machine Learning for Healthcare, **Natural Language Processing**, Natural Language Understanding, **Machine Learning**, Big Data, Probability and Statistics for Data Science, Optimization and Linear Algebra for Data Science, Introduction to Data Science, Introduction to Computer Science.

EXPERIENCE

Data Science Research Intern, **Marron Institute of Urban Management**, Brooklyn, NY

May. 2022 – present

- Collaborated with 2 other interns and the Illinois Department of Correction to build a job recommendation and a job application system for **20239 parolees** across **104 prisons** in Illinois.
- Generated an interactive and geographical data visualization dashboard to present distribution of parolees and prisoners by counties in Illinois and neighborhoods in Chicago with libraries such as Geopandas and Plotly.
- Analyzed parole and prison population datasets to allow policy makers at Department of Correction to implement better policy for helping parolees and to prevent crimes in the future.

Top 50 teams, **Humana-Mays Healthcare Analytics Case Competition**, New York, NY

Sep. 2021 – Oct. 2021

- Performed feature engineering on **367 features** of **10 million** Humana customers with resampling, one-hot encoding and PCA.
- Implemented ML models including XGBoost, SVM and Random Forest to identify customers' hesitant towards COVID vaccine.
- Categorized vaccine resisters and explored reasons (pre-existing conditions, access, misinformation, lack of trust) for the hesitance.
- Recommended policies for helping Humana to encourage customers to receive vaccines.

Universal Intern, **Aflac**, New York, NY

June 2020 - Aug. 2020


- Rotated through managerial, sales, marketing and financial operations departments and collaborated with 10 interns.
- Developed actionable recommendations, such as podcasts, to increase company's marketing effectiveness by **10%**.
- Analyzed and visualized Aflac's financial data by using Excel dashboard and presented to program supervisor to showcase insights.

PROJECTS

User/ item representation algorithm – Capstone @ Zillow | *neural network, big data, Python*


Sep. 2022 – present

- Working with Zillow AI team and constructed user/item representations for short- and long-term prediction with transformers4rec library
- Modifying the transformers4rec library to create and evaluate model on Zillow's dataset
- Improving the model by testing additional hypotheses, varying variables such as loss function, variables, sampling method, dataset

Data Analysis on Among Us Dataset  | *Hypothesis Testing, Logistic Regression, Multiple Linear Regression, Python*

Dec. 2021

- Conducted hypothesis testing on approx. **2000 Among Us players** with **14 features**; suggested winning strategy for crewmates.
- Created a new dataset regarding crewmates' game data to predict player's imposter win rate by multiple linear regression.

Breast Cancer Prediction  | *Scikit-Learn, Logistic Regression, Random Forest Classifier, KNN, Python*

May 2021

- Cleaned and turned **146.8+ KB** of numerical and text data into training and test sets.
- Constructed **3 machine learning models** to predict beingness of breast lumps based on **30 clinical parameters** of **569 patients**.
- Conducted exploratory data analysis and visualized information from dataset by using histograms, scatter plots, and heatmaps.
- Tuned 4 hyperparameters to increase accuracy of Random Forest Classifier model from 96% to 97%.