

Wangrui Hou (Wendy)

Jersey City, NJ | (929)990-5985 | wh916@nyu.edu | [linkedin.com/in/wangruihou](https://www.linkedin.com/in/wangruihou)

EDUCATION

M.S. in Data Science, *New York University*

September 2019 – May 2021

- **GPA: 3.8/4.0**

- Relevant Courses: Optimization Linear Algebra; Probability & Statistics; Machine Learning; Big Data; Computational Cognitive Modeling, Database Systems, Recommender Systems

B.S. in Media, Culture, and Communications, *New York University* September 2015 – May 2019

- Minors in Mathematics, Psychology

- **Magna Cum Laude: 3.832/4.000**

- Relevant Courses: Algorithmic Culture; The Rise of Internet Media; Advertising and Society

SKILLS

Programming | Python (PyTorch, Pandas, NumPy, Scikit-learn, Matlab etc.), SQL, GraphQL, Java, Tableau, Git/GitHub, DMPs

Languages | Chinese (native), English (full professional proficiency)

Business | Data ETL Pipeline, Data Analytics, Data Visualization, Google Analytics, Verbal & Written Communication, User Research, Market Research, Client Relationship Management

WORK EXPERIENCE

Product Data Scientist Intern, *Johnson & Johnson Health Tech*

June 2020 – August 2020

Remote from Jersey City, NJ

- Built 3 data ETL pipelines based on Shopify APIs using Python, SQL, and GraphQL to help MyStore, Johnson & Johnson's internal e-commerce platform, obtain raw and live data on order history, 47k+ customers, products, and daily inventory status
- Implemented and evaluated 6 nightly or monthly AWS Lambda functions to calculate customer metrics such as recency, frequency, monetary value, customer lifetime value, and churn rate
- Generated 20+ detailed and interactive Tableau dashboards to visualize data insights including customer behaviors, revenue and order volume history, customer segmentations, frequently bought together products and etc.
- Advised J&J Health Tech team leadership of appropriate business strategies and performed ad hoc data analyses to prepare presentation materials for the strategy team

Client Strategy Intern, *Lotame Solutions Inc.*

September 2019 – December 2019

New York, NY

- Used Decision Tree algorithm and Logistic Regression algorithm for feature engineering on the topic of churn prediction and discovered the important relationship between "Client Touchpoint Score" and client churn
- Interpreted first-party and third-party data to access user behaviors/campaign performances and presented these business insights to clients through written reports and presentations
- Communicated with clients on their requests to ensure client satisfaction, engagement, and product usage

DATA SCIENCE PROJECTS (<https://github.com/hwendy12>)

Spoken Language Understanding, *Deep Learning Project*

September 2020 – December 2020

- The goal of this project is to create a generally-purposed SLU model for intent classification
- Assembling and augmenting various SLU corpora into a super corpus to train SLU models
- Improving Lugosch's SLU model for generally-purposed tasks through data augmentation, manipulating hidden units and model structure, and transfer learning.

COVID-19 Infection Rate Prediction, *Machine Learning Project*

April 2020 – May 2020

- Cleaned and analyzed Johns Hopkins' COVID data and U.S. Census data on U.S. counties' population, education, public and private healthcare, popular transportation methods etc.
- Trained Logistic Regression, Support Vector Machine, Perceptron, Gradient Boosting, and Neural Network to predict COVID infection rates for the 529 counties that were missing from Johns Hopkins' dataset based on their county features data from U.S. Census