

Hua Shi

New York, NY | U.S. Permanent Resident | melaniesoek0120@gmail.com | [LinkedIn](#) | [GitHub](#) | [Blog](#) | [Video](#)

INTRODUCTION

Experience in data analysis and machine learning with a Data Science background. Having earned a master's degree in Economics has provided me a solid foundation and a good intuition in Machine Learning and Statistical Analysis. My experience is ideal for helping data-driven companies make informed and optimal decisions.

TECHNICAL SKILLS

Programming & Data Visualization: Python, Tableau, Google Data Studio

Statistical Techniques: Hypothesis Testing, A/B Testing, Linear Regression, Logistic Regression, Time Series

Machine Learning: K-nearest Neighbors, Random Forest, XGBoost, Support Vector Machines, Naive Bayes, Principal Component Analysis, Clustering, Recommendation Systems, Deep Learning, Natural Language Processing

Other Technical Skills: SQL, Google Analytics, Google Sheets, web scraping, Git

EMPLOYMENT HISTORY

Junior Data Analyst - H1, New York, NY

Jan 2021 – Current

- Responsible for researching and collecting data on leading healthcare professionals to the proprietary database
- Used analytics methodologies and Excel/Google Sheet/ Python to manipulate and clean the data
- Built functions & procedures, writing complex queries, using SQL
- Developed strategies to capture additional data and client use-cases to effectively partner with data engineering teams for better insights into operational processes

Career Break - Full-Time Mother

Feb 2018 – Jan 2020

Dior Beauty Advisor - Macy's Inc., Flushing, NY

Aug 2017 – Jan 2018

- Collaborated with the counter manager to organize products and keep correct inventory
- Ranked #1 in sales out of eight based on employee satisfaction and sales record

Paralegal Assistant - Geng & Associates, P.C. Flushing, NY

May 2017 – Jul 2017

- Reviewed and corrected clients Co-op contract to make sure all the details were accurate
- Scheduled and conducted interviews with clients and professionals to discuss the details

Assistant Store Manager - The Yeon Beauty, Flushing, NY

Jun 2016 – Apr 2017

- Hired and trained employees with professional sales skills and maintained daily inventory count to keep its accuracy
- Maintained statistical and financial records and put forward new strategies for holiday events with the store manager

Data Analysis Intern - ChengXing Real Estate, Jilin, China

Nov 2012 - Mar 2013

- Cleaned the data and converted data into actionable insights by predicting and modeling future outcomes
- Applied OD/PA to analyze the direction and development trend of the population commuting in cities or regions
- Analyzed and synthesized large data streams across multiple inputs with Excel
- Created dashboards with Arcmap and compiled reports for the manager

TECHNICAL PROJECTS

AI Chatbot | [GitHub](#) | [YouTube](#)

Created an interactive chat system to help answer frequently asked questions

- Extracted data from Reddit and web scraped Stack Overflow
- Sanitized and re-arranged text data by leveraging Natural Language Processing (NLP)
- Trained and fine-tuned the hyperparameters to optimize the sequence-to-sequence model
- **Leveraged:** Python, Pandas, Tkinter, Reddit API, NLP

Time Series Panel Data Analysis | [GitHub](#)

Predicted air pollution (PM2.5) in Beijing in order to provide the government with environmental recommendations

- Sanitized data and applied Exploratory Data Analysis to obtain actionable insights
- Used the stationary process to check stochasticity
- Applied Econometrics and Machine Learning algorithms
- **Leveraged:** Python, Pandas, Numpy, Matplotlib, Hypothesis Test, Fixed Estimator, Random Estimator, Pooled OLS, LSTM

Ternary Classification Analysis | [GitHub](#)

Predicted whether water pumps are broken, not broken, or need repair in order to decrease expenditure

- Identified specific regions in Tanzania that need more repair than others
- Increased the F1 score from 75% to 80% by tuning the hyperparameters
- Interpreted complex data sets patterns and provided a detailed data visualization of findings
- **Leveraged:** Python, Pandas, Seaborn, Matplotlib, Oversampling Method, Random Forest, Decision Trees, Logistic Regression

EDUCATION

Flatiron School, New York, New York

Jan 2020 – Apr 2020

Immersive Data Science Bootcamp program

Stony Brook University, Stony Brook, New York

Aug 2015 – Dec 2016

Master of Science, Economics

Yanbian University, Jilin Province, China

Aug 2009 – Dec 2013

Bachelor of Science, Economics