

SABRINA YANG

DATA SCIENCE + MACHINE LEARNING

EXPERIENCE

Metis New York, NY
Data Scientist May 2021 to Aug. 2021
Completed Metis's 14-week accredited data science bootcamp focused on Python programming, machine learning, deep learning, data engineering, NLP, data visualization, project design, and communication. **Selected Projects include:**

Bag Hunter: Brands Identify Classification App (Deep Learning)

- Built a bag brand image classification model, based on neural networks, by web scraping Farfetch.com and storing data on MongoDB and constructed a recursive pipeline to build a bi-weekly updated database
- Created a Streamlit app that allowed users to dynamically input bag images to identify bag brands and their probability distribution chart

Loan Repayment Prediction (Classification)

- Predicted whether borrowers would default on their loans by wrangling 3 million data points with 140 features into only 50 features; applied data visualization analyses such as correlation matrixes, which are aimed at helping lending institutions maximize profit
- Utilized various classification models (logistic regression, KNN, random forest, XGBoost) to optimize the F beta score for model selection

Charging Away From Climate Change: Where To Build EV Charging Stations in NY (Business)

- Created a location optimizing plan for Volta Charging to build its electric vehicle (EV) charging stations; analyzed EV data across geographic regions using Tableau and Excel
- Forecasted EV sales and profitability by NY county locations to rank the best locations for EV charging station development; created a Tableau map visual to display optimal location choices

Tesla Stock Tweets (NLP)

- Built an NLP model to analyze Twitter sentiment for Tesla's (\$TSLA) stock based on tweets; used NLTK and Regex to pre-process tweets, TFIDF to vectorize text, and NMF for topic modeling
- Performed Vader sentiment analysis and NLP topic modeling; compared its output with stock price movements pulled from Yahoo Finance

Box Office Gross Prediction (Regression)

- Predicted box office gross revenue by building a multi-linear regression model; web scrapped Box Office Mojo movie feature data using BeautifulSoup
- Applied feature engineering and used sklearn for model evaluation (MAE) and pyplot diagnostic plots for model selection; identified the best model to predict gross revenue for pre-release movies by utilizing LassoCV

Newman Myers Law Firm New York, NY
Data Analyst - Corporate Finance Aug. 2018 to May 2019
Optimized company expenses and cash flows by analyzing and reconciling cost trends across numerous legal cases and settlements; generated cash flow budgets and targets which reduced expenditures across the firm

Rock Properties Newark, NJ
Database Administrator - Property Management Apr. 2017 to July 2017
Maintained and developed multiple databases using SQL and Salesforce by tracking property occupancy totals, vacancy reports, leasing call statistics, and resident survey scores; increased database efficiency and insight generation

Compass New York, NY
Data Analyst - Real Estate Nov. 2016 to Feb. 2017
Analyzed regional sales data, current market trends, and seasonal price sensitivity to develop valuation models on multiple real estate properties in New York; presented findings to real estate agents and brokers

Bloomington New York, NY
Sales Data Associate - International Business May 2016 to Feb. 2017
Developed database analytics to monitor and extract consumer spending data for the International Business segment; analyses were used to develop and refine the international marketing and promotions strategy on social media

TC Bank Taipei, Taiwan
Trader - Foreign Exchange Oct. 2010 to Dec. 2013
Executed and performed large foreign exchange transactions, totaling over \$3 million per month, for major corporate clients; ensured optimal trade timing using market research and technical analysis

Procter & Gamble Taipei, Taiwan
Data Analyst - Sales Oct. 2007 to May 2010
Analyzed revenue data, industry trends, brand competition, and customer needs to generate presentations for the SKII skincare team; presentations were used by account managers to develop sales strategies

EDUCATION

Harrisburg University of Science and Technology
M.S. Analytics 2019
Coursework: Sentiment Analysis, Data Visualization, Quantitative Decision Making, Machine Learning
Grad Essay: US Presidents Speech NLP Analysis in R

Baruch College CUNY
Certificate Finance 2015

National Taiwan University
B.S. Public Health Biostatistics 2007

CONTACT

✉ sabrinayang.ds@gmail.com
☎ 929-232-5220
📍 New York, NY
🏠 SYNYC

SKILLS

PROGRAMMING

Python
SQL
R
HTML

MACHINE LEARNING

Regression
Classification
Natural Language Processing
Neural Network
Random Forest
XGBoost
K-Means
Dimensionality Reduction
Naive Bayes
A/B Testing
Clustering
LSTM

DATA VISUALIZATION

Tableau
Matplotlib
Seaborn

DATA MANAGEMENT

AWS
Google Cloud Platform
MongoDB
Spark

LIBRARIES

Numpy
Pandas
Scikit-Learn
Keras / Tensorflow
NLTK
Beautiful Soup
Selenium

OTHERS

Excel
Streamlit