

SABRINA YANG

DATA SCIENCE + MACHINE LEARNING

EXPERIENCE

Metis Data Scientist

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 project section for more details.**

New York, NY
May 2021 to Aug. 2021

Newman Myers Law Firm Data Analyst - Corporate Finance

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

New York
May 2019 to Oct. 2019

Rock Properties Database Administrator - Property Management

Maintained and developed multiple databases by tracking property occupancy totals, vacancy reports, leasing call statistics, and resident survey scores; increased database efficiency and insight generation

Newark, NJ
Apr. 2017 to July 2017

Compass Data Analyst Intern - Real Estate

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

New York, NY
Nov. 2016 to Feb. 2017

Bloomington Support Associate, International Business

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
Implemented online social media campaigns for seasonal and special events by researching consumer behavior, search channels/trends, and online traffic; improved brand awareness and sales across target markets

New York, NY
May 2016 to Feb. 2017

TC Bank Trader, Foreign Exchange

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

Taipei, Taiwan
Oct. 2010 to Dec. 2013

Procter & Gamble Data Analyst, Sales

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

Taipei, Taiwan
Oct. 2007 to May 2010

DATA SCIENCE PROJECTS

Bag Hunter - Brands Identifier Image Classification App

- Web scrapped Farfetch.com using a self-created API and stored it in MongoDB; pulled data/images on Python and cloud computing to build an image classification model, based on neural networks, for predicting bag brands from images
- Constructed a recursive pipeline to build a bi-weekly updated database, connecting Farfetch.com and MongoDB
- Created a Streamlit app that allowed users to dynamically input bag images to identify bag brands and its probability distribution chart

Loan Repayment Prediction

- Taken historical loan data from Lending Club.com to build a classification model to help lending institutions to profit from being able to predict the risk of the borrower would default on a loan or not
- Wrangled 3 million data points with 140 features into 50 features by checking its definitions on Loan Dictionary and created data visualizations for observing the relations between target and variables
- Ran various classification models (logistic regression, KNN, random forest, XGBoost) with GridSearchCV tuning parameters and managed class imbalance issue to determine the best prediction model on F beta score

Charging Away From Climate Change - Where To Build EV Charging Stations in NY

- To help Volta Charging.com expand by choosing the optimal locations to place EV charging stations in NY to benefit from President Biden EV network plan; while taking into account its unique business model objectives
- Analyzed electric vehicle (EV) data across geographic regions (i.e., EV ownership, EV sales, population density, sales tax per capita, and daily traffic) using data manipulation and visualization tools
- 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

- To monitor if there's any relationship between Tesla stock price movement and Tweets sentiments
- Extracted Tesla stock (\$TSLA) tweets from Twitter using API Tweepy; collected various features (i.e., tweet text, dates, usernames, locations, following, followers, total tweets, hashtags, retweet counts, etc.)
- Used NLTK and Regex to clean and pre-process text data through string extraction, tokenization, lemmatizing, and stop words; created a document term matrix with a TFIDF vectorizer and used NMF for topic modeling to find relevant stock topics from reoccurring terms
- Performed Vader sentiment analysis and NLP topic modeling contents compared it with stock price movements pulled from Yahoo Finance with web data reader

Box Office Gross Prediction

- Web scrapped Box Office Mojo data for Top Lifetime Domestic Grossing movies using BeautifulSoup and extracted its profile features (such as budget, genres, running time, MPAA rating, and Cast & Crew etc.) to predict box office gross revenue through performing multilinear regression models
- Used sklearn for model evaluation (MAE) and pyplot diagnostic plots for visualization to compare model performances; found the best model to predict gross revenue for pre-release movies

CONTACT

✉ nyyang.ntu@gmail.com

📞 929-232-5220

📍 New York/Toronto

🌐 SYNYC

SKILLS

PROGRAMMING LANGUAGES

Python

SQL

R

DATA VISUALIZATION

Matplotlib

Seaborn

Tableau

LIBRARIES

Numpy

Pandas

Scikit Learn

Keras

Tensorflow

NLTK

Beautiful Soup

Selenium

MODELS/ALGORITHMS

Regression

Random Forest

XGBoost

Convolutional Neural Network

NLP

BIG DATA

Google Cloud

MongoDB

AWS

EDUCATION

Harrisburg University of Science and Technology
Masters of Science Analytics 2019

Baruch College CUNY
Certificate of Finance 2015

National Taiwan University
Bachelor of Science Public Health
Biostatistics 2007