

JASON KIM

✉ kimhjason@outlook.com
☎ (818) 935-0955
📍 New York, NY 10017
in jason-hkim
🌐 Jason-HKim

Recent NYU graduate with certification as a Spark Developer and AWS Cloud Practitioner, and passionate about the world of data. Skilled in coding with Python programming language and SQL. A professional with an interest in technology and the ability to explore tools to facilitate data analyses at scale and provide effective data processing for observation.

Skills

LANGUAGES

Python
SQL

PACKAGES

Pandas
NumPy
Scikit-learn
BeautifulSoup
NLTK

DATA VISUALIZATION

Matplotlib
Seaborn
Tableau

MACHINE LEARNING

Regression
Classification
Natural Language Processing

BIG DATA

Apache Spark
Databricks
Amazon Web Services

Education

New York University Sept. 2016 to May 2020
Bachelor's Economics 2020
BA in Economics with Policy Concentration; Minor in Business

Experience

Metis

Data Scientist

Completed intensive Data Science & Engineering bootcamp and gained experience utilizing Python, Pandas, web scraping tools (BeautifulSoup, Selenium), visualization tools (incl. Matplotlib, Seaborn), big data processing (Spark) and machine learning libraries (incl. SKLearn) to develop five projects that included exploratory data analysis, modeling, machine learning, building data pipelines, sentiment analysis, and natural language processing. See project section for in-depth description of projects.

New York, NY (Remote)
Mar. 2021 to June 2021

Localposh Inc.

Partner Success & Business Development Intern

• Pitched company services to potential partners to secure initial vendor contracts
• Researched and categorized local businesses in Excel to establish initial partner list to jump-start NYC entrance project
• Analyzed competitors' data with SWOT analyses to maximize capture of market share and presented insights to marketing director

New York, NY
Sept. 2019 to Jan. 2020

Comfy U.S.A. Apparel Inc.

Part-Time Assistant

• Assisted with procurement to ensure accurate client order fulfillment and timely product delivery
• Recorded, analyzed, and presented production expenses to management helping save approx. 8% in budget
• Organized booth setup at fashion trade shows in Jacob K. Javits Convention Center, NYC

Sylmar, CA
2015 to 2019

BBCN Bank Northridge

Intern

• Gained experience in account management by inputting and verifying checks
• Learned about banking security and significance of physical currency and reserves to businesses and financial institutions

Northridge, CA
June 2014 to Aug. 2014

Projects

Exploratory Data Analysis on MTA Data

- EDA project consisting of data cleaning and exploration to provide recommendation for best travel times for a delivery company using the MTA
- **Data:** MTA Turnstile dataset obtained from MTA database
- **Data Preprocessing & Manipulation:** Removing nan values and duplicate rows with pandas, and obtaining noncumulative values from cumulative entries/exits columns
- **Communication/Visualizations:** Matplotlib and Seaborn for visualization of ridership volume by time and date

Linear Regression on NBA Usage Rate

- Linear Regression project to predict NBA player usage rate based on per-game statistics
- **Data:** Web scraping per-game statistics obtained from basketball-reference.com using BeautifulSoup
- **Data Preprocessing & Manipulation:** Pandas to remove redundant data and create DataFrame from three seasons of data
- **Tools/Algorithms:** Predicting NBA player usage rate with K-Fold Cross Validation and Ridge Regression
- **Communication/Visualizations:** Matplotlib and Seaborn for visualization of usage rate prediction and important statistics with positive/negative impact on mean R-squared score

Data Engineering Pipeline with Spotify Data

- Utilize Spotify data to engineer a data pipeline and deliver an interface that allows data exploration and try to predict song popularity
- **Data:** Big Data Spotify dataset obtained from Kaggle (600,000 rows of track data and 1.1 million rows of artist data), stored in SQL database using DB Browser
- **Tools/Algorithms:** Linear Regression model with SKLearn to attempt to predict popularity of songs based on used features
- **Communication/Visualizations:** Matplotlib and Seaborn to display distribution of song popularity levels across all artists and genres of music among popular artists

Classification of Sepsis in Patients

- Build classification model to help predict whether or not a patient will develop Sepsis
- **Data:** Medical dataset obtained from Kaggle consisting of measurements of patients six days prior to Sepsis testing
- **Data Preprocessing/Manipulation:** Many null values were removed by dropping feature columns. High class imbalance remedied through Random Oversampling
- **Tools/Algorithms:** ROC-AUC and F-beta to score KNN, Logistic Regression, and Random Forest models. RandomSearchCV for hyperparameter tuning
- **Communication/Visualizations:** Matplotlib and Seaborn to display feature importance for final f-beta score

Natural Language Processing with Amazon Review Text

- Conduct NLP and Sentiment Analysis on Amazon reviews about digital video game products to explore the consistently utilized language around positively and negatively reviewed products
- **Data:** Digital Videogame reviews set obtained from Amazon Public Datasets contained review data
- **Data Preprocessing/Manipulation:** Removing all non-alphanumeric characters, lowercasing review text, using NLTK and custom words to remove stop words
- **Vectorization:** Created Document-Term matrix with TF-DF vectorizer and CountVectorizer. Fit to LSA and NMF for topic modeling seven topics
- **Visualizations:** Used Matplotlib and Seaborn for top ranked categories in positive and negative reviews. Scattertext to visualize positive and negative words across entire dataset