

# RAJKUMAR KATTA

## DATA SCIENTIST

✉ kattarajkumar@gmail.com  
☎ 929-231-6479  
📍 New York, NY 10001  
in r-katta  
🌐 rkatta

## Summary

Fast-learning, motivated Data Scientist with a background in Finance. Spent the past year self learning and working with Metis to develop data science skills.

## Skills

### PROGRAMMING LANGUAGES

Python  
SQL  
HTML/CSS

### MACHINE LEARNING

Generalized Linear Models  
Random Forest  
Gradient Boosting  
Neural Networks  
Natural Language Processing  
Time Series

### VISUALIZATIONS

Tableau  
Matplotlib  
Seaborn

### LIBRARIES

Pandas  
Scikit-Learn  
Numpy  
NLTK  
BeautifulSoup  
Selenium  
TensorFlow  
Keras  
Flask

### DATABASES

Microsoft SQL Server  
Postgres SQL  
MongoDB

### OTHER

AWS  
GIT  
Google Cloud

## Education

CUNY Baruch  
BBA Finance 2014  
Minor History 2014

## Projects

### Predicting Computer Monitor Prices (Linear Regression)

- Scraped computer monitor specification data from Newegg using BeautifulSoup and Selenium python packages
- Performed feature transformations such as one-hot encoding and feature selection using LASSO regularization
- Fit both linear and polynomial regression models to the data
- Used the R squared and MAE metrics to determine model performance

### Renaissance and Modern Poetry Analysis (NLP)

- Scraped Renaissance and Modern Day poetry from PoetryFoundation using BeautifulSoup and Selenium Python packages
- Identified topics using CountVectorizer, TFIDF and Principal Component Analysis (PCA)
- Clustered similar poems using K-Means Algorithm

### Predicting Severity of Car Accidents (Classification)

- Used Postgres SQL to store, join, filter and transform 2015 US car accident data as downloaded from Kaggle
- Modeled the data using KNN with Cross Validation and Random Forest and was able to achieve 66% accuracy
- Determined that air bags and use of a seat belt are the most important features in determining severity of injury in car accidents

## Experience

### Metis

*Data Scientist*

New York, NY  
Apr. 2020 to June 2020

Metis is an ACCET accredited 12-week immersive Data Science Bootcamp focused on project oriented learning and covering topics such as Machine Learning, Statistics, Probability, Data Management, Programming, Communication, and Design. By the end of these twelve weeks, completed four individual end to end data science projects. Projects utilized a multitude of data science skills, including:

- Machine learning and statistical analysis in Python
- Data scraping and storage (PostgreSQL, MongoDB, Hadoop, Spark)
- Data transformation: reshaping, cleaning, type casting, filtering, broadcasting
- Scientific Python packages: Pandas, NumPy, SciPy, statsmodels, scikit-learn
- Supervised learning: regression, classification, ensemble methods, regularization
- Unsupervised learning: clustering, topic modeling, dimensionality reduction
- Natural Language Processing: vectorization, topic modeling, sentiment analysis
- Data visualization: Matplotlib, Plotly, Tableau

### Barclays Investment Bank

*Liquidity Risk Analyst*

New York, NY  
May 2018 to Apr. 2020

- Collaborated with the tech team as a subject matter expert to prioritize and refine business requirements, create user stories, perform UAT and support implementation of various tech projects
- Assessed and monitored liquidity risk metrics such a liquidity risk ratio (LCR) as impacted by the firm's existing and new business activities
- Coordinated with risk and technology partners to improve systems and data standards to enhance the accuracy, completeness and timeliness of reports
- Consolidated key findings and decompose data mappings used in the manual liquidity reporting processes of a major Barclays entity in the strategic transformation leading up to automation
- Spearheaded, implement and document the processes to identify high quality liquidity assets as required by Fed's marketability guidelines
- Effectively trained a newly formed team of 6 in Glasgow on the daily creation of critical reports submitted to the Federal Reserve

### Pride Technologies

*Data Integrity Specialist at PDT Partners*

New York, NY  
Mar. 2016 to Dec. 2017, Nov. 2013 to Oct. 2014

- Reconciled key financial model inputs such as security reference data, corporate actions and quarterly earnings reducing trading outages from once a week to once in four months on average
- Documented methods and procedures for data cleansing protocol that serves as a master reference document firmwide and has been crucial for training over 10 colleagues
- Identified underutilized resources such as obsolete data feeds or irrelevant products that saved the firm hundreds of thousands of dollars in subscription fees
- Gathered requirements from internal stakeholders regarding implementation of change management workflow in response to Brexit and then followed through with implementation and testing within the Atlassian framework
- Implemented and managed a complex JIRA project used firm-wide to track notifications from external parties and reduced major outages from an average of four a year to zero in my tenure
- Developed ad-hoc SQL queries to retrieve and update financial reference data from and to a database for analytical and business applications
- Created and maintained ad-hoc scripts in Python for projects that involve web scraping, email notifications and file processing