

SAHIL KERKAR

sahilkerkar1994@gmail.com ♦ New York, NY ♦ (973) 945-0339 ♦ www.sahilkerkar.com

Education

Duke University ♦ *Durham, NC*

August 2012-May 2016

♦ Dual B.S. in Mathematics and Economics ♦ **3.95 GPA** ♦ Summa Cum Laude ♦ Phi Beta Kappa ♦ Multi/Linear Algebra TA

Data Courses and Certifications: Data Scientist in Python Path/Data Analyst in R Path (Dataquest.io), MySQL for Data Analytics and BI/Tableau 2020 A-Z (Udemy.com), Stanford Machine Learning Course/Deep Learning Specialization (Coursera.org)

Data Science Portfolio

Visit sahilkerkar.com for in-depth analysis/findings and github.com/sahilkerkar for source code/relevant datasets

School Safety: Analyzed the US Department of Education's School Survey on Crime and Safety from the 2015-2016 school year

- ♦ Part 1: Removed multicollinearity, trained/optimized 8 regression models, and concluded that drug testing, requiring faculty IDs, training teachers on substance abuse, and community involvement in safety were associated with a reduction in violent incidents
- ♦ Part 2: Employed the SMOTE algorithm, trained/optimized 7 multiclass classification models, and determined that a random forest model was most effective at capturing nonlinearity and interaction between features used to predict bullying prevalence
- ♦ Part 3: Implemented dimensionality reduction/clustering to suggest disciplinary action at schools was influenced by student race

Evolution of Rap: Integrated data from Billboard, Spotify, and Genius APIs and applied NLP techniques to identify trends since 1989

- ♦ Part 1: Visually displayed changes in lyrics to demonstrate that present-day tracks are more profane, shorter, and more repetitive
 - ♦ Part 2: Concluded that audio features were useful but not sufficient in predicting if a track would break into the Billboard Top 5
-

Experience

Laurion Capital Management ♦ *New York, NY* ♦ *Cross-Asset Volatility Trader*

June 2019-May 2020

- ♦ Utilized SQL, Jupyter, pandas, and time series data to backtest a mean-reversion trading strategy for corporate credit indices, factoring in transaction costs and delays in execution time and concluding that the strategy was profitable in the long run
- ♦ Applied multiple linear regression on cross-sectional data to analyze the option implied volatilities of currencies, ETFs, and government bond futures and provided profitable recommendations for which assets statistically screened expensive versus cheap
- ♦ Built visualizations with matplotlib to convey the relationship between short-term price movements and future realized volatility

Barclays ♦ *New York, NY* ♦ *Trading Assistant Vice President/Summer Analyst*

Summer 2015/August 2016-May 2019

- ♦ Managed and hedged multi-dimensional portfolio risk by executing broker trades, producing daily risk metrics, and proposing trade ideas based on qualitative macroeconomic views and quantitative statistical analysis
- ♦ Combined Python, Bloomberg Terminal, and Excel to perform time series analysis on the correlation and volatility of various currencies and commodities and presented results to the sales team to drive client engagement and increase revenue

Cigna ♦ *Bloomfield, CT* ♦ *Actuarial Intern*

Summer 2014

- ♦ Reduced model runtime from 600 hours to 1 hour by redesigning and efficiently programming a SAS-based implementation of Monte Carlo simulations to forecast insurance product performance in future fiscal quarters

Fuqua School of Business ♦ *Durham, NC* ♦ *Research Assistant for Professor Christine Moorman*

August 2013-May 2016

- ♦ Built a Python tool to clean and apply NLP techniques to over 5,000 transcripts of pharmaceutical company quarterly presentations and determined that the length, tone, and factual density of these presentations have a significant impact on stock price reaction
-

Technical Skills

Programming: Python (NumPy, Pandas, Matplotlib, Seaborn, Sklearn, NLTK), R (Tidyverse), SQL/NoSQL, VBA, SAS

Applications: Jupyter, Command Line, API Data Retrieval, Web Scraping, Git, Spark, Tableau, Excel, Bloomberg Terminal

Modeling: Linear/Logistic Regression, SVM, KNN, Random Forests, Deep Learning (CNNs/RNNs), PCA, Clustering, NLP