Sean **Ammirati**

✉ sammirati@statsworks.info | ▫ (347) 628 9155 | ♀ New York, NY | ⚲ www.statsworks.info

Work Experience

Major League Baseball
New York, NY

DATA SCIENTIST
January 2020 – Present

- Developed a COVID-19 pandemic scenario simulator by creating a regression model and scenario simulation. The regression model was trained using LASSO for league-wide, game level baseball attendance and revenue considering game features as well as external factors. Then, simulated potential outcomes in terms of opening day date, economic condition and other sports' responses were used to produce ranges for both attendance and revenue in each scenario. These results were used at both a league and club level to give a sense of how to best prepare resources.
- In addition to the scenario simulator, developed a COVID Correction Factor (CCF) to account for the diminished attendance due to fear, government regulation, or extreme economic strain. This factor was estimated using Markov Chain Monte Carlo simulations with a Bayesian framework, modeling the CCF following an exponential decay over time and integrating survey results to construct posterior distributions for the initial CCF and half-life of its resonance.

CKM Analytix
New York, NY

DATA SCIENTIST
June 2018 – December 2019

- Leveraged various machine learning and statistical techniques in natural language processing, such as topic modeling and keyword analysis using NMF and DBSCAN, to determine primary areas of automation based on description text of IT tickets.
- Directed the development and research behind a series of time-series forecasting models (Seasonal Means, Holt-Winters Exponential Smoothing, SARIMA and Prophet) for IT Service Level Agreement (SLA) violations, forecasting whether a particular priority of ticket would breach at the end of a period using both proportional and average measures.
- Coordinated and developed for the analytics team on a Pro Bono project for a not-for-profit gun safety advocacy group to produce accurate depictions of public sentiment concerning gun control, gun safety, and background checks over time using Local Polynomial Regression and Recursive Bayesian Filtering. This allowed them to track the public sentiment reliably over time from both public and private polls.
- Administered and maintained a relational database to store historical and current transaction and holdings information for a major law firm and architected a generalized historical ETL software package to map historical records of varying formats into the database consistently and accurately. Overall effort reduced client work from weeks of searching through Excel files to seconds querying the database.
- Spearheaded the training of new joiners in data manipulation with python using pandas and numpy, visualization using plotly, and object-oriented programming. Orchestrated lectures on statistics to ensure the use of the scientific method across the firm.

AXIS Capital Insurance Company
New York, NY

ACTUARIAL INTERN
January 2017 – December 2017

- Developed a Generalized Hierarchical Linear Model (HLM) with log link function using R to score potential business opportunities, estimating the probability to book and to quote a policy.

CVS Health
New York, NY

CASHIER
July 2015 – December 2016

- Created a Logistic Regression model for probability of running out of on-floor inventory of particular selected items (like toilet paper, paper towels, soda) to create a schedule of restocking for each particular item (every day, every other day, every week, etc.) Resulted in a 81% boost in customer satisfaction surveys in the "were you able to find what you were looking for?" sections.

Education

Hunter College
New York, NY

M.A./B.A IN STATISTICS
September 2012 – December 2017

GPA: 3.97

Skills

Languages: Python, R, SQL, JavaScript, Golang, Bash, VBA
Tools: Pandas, NumPy, Sci-Kit Learn, Plotly, Matplotlib, SciPy, Spark, Keras/Tensorflow, Git, pymc3
Development: ETL Pipelines, Object Oriented Programming, Test Driven Development
Machine Learning: Tree Based Methods, Gradient Boosting Machines, Artificial Neural Networks, Clustering
Statistics: GLMs, Multivariate Analysis, Bayesian Inference, Stochastic Processes

Projects (statsworks.info/category/projects)