Jared Berman Data Scientist

github.com/jaymberman

- Inquisitive and dedicated technical professional with three years of open-source software experience analyzing data data & developing models
- Experienced in supervised and unsupervised learning techniques, deep learning, data pipelines, model scoring & deployment
- Solid general programming skills, knowledge in a wide range of technologies, and a strong aptitude and desire to learn more

Experience

2020- — **Data Scientist, Advertisement Analytics & Insights, Spark Foundry**

- Collaborated to write clean and efficient code in Python to automate data quality assurance, ingestion, and reporting, pull from APIs and push to data lakes, and glue together various business processes.
- Used various data storage technologies from local sqlite to MySQL and S3 buckets.
- Created, maintained, and reported data dashboards and presentations to high-profile clients.
- Worked with media teams and vendors to pull, quality assure, and transform data from various sources and square away discrepancies, on the scale of tens of millions of advertisement impressions per week and dozens of attributes per row.
- Compiled & presented reports delivering insights and statistical analyses including hypothesis & A/B testing and regression to clients with annual advertising budgets exceeding $10MM.

Skills

Programming Languages: Python, SQL, C++, Java, Bash
Development Technologies: Git, GitHub, Docker, Heroku, Jupyter Notebook, Conda
Cloud Technologies: AWS
Machine Learning Frameworks: Scikit-Learn, XGBoost, Keras, MLLib, Statsmodels
Distributed Frameworks: Hadoop/Cloudera Ecosystem (Ambari) including Apache Spark, Hive, Hbase, and Pig
Data Analysis Libraries: Pandas, NumPy, Matplotlib, Seaborn
Basic Skills: Linux platform, package & dependency control, documentation reading & troubleshooting/debugging, data structures & algorithms

Education

Master's Degree, Applied Mathematics
Hunter College, New York, New York, 2017-2019
Bachelor's Degree, Mathematics
Hunter College, New York, New York, 2014-2017
E-learning Courses
Udemy - Deployment of Machine Learning Models; Git Complete; Data Science for Business; The Ultimate Hands-On Hadoop; Spark and Python for Big Data; GCP: Complete Google Data Engineer and Cloud Architect Bootcamp