Jake Byford

https://jakebyford.github.io/portfolio/

Email : jake_byford@outlook.com
Mobile : +1-732-861-7994

Analytical problem solver who seeks to interpret data to make decisions and answer coding questions while communicating findings effectively.

Summary of Qualifications

- 4 years' experience: Python, R, SQL, Tableau to create big data models
- 4 years' experience: Coding math, data analysis, data science
- M.S. Data Science; 4+ years' business, analytical, and technical writing

Education

- **Master of Science, Data Science** | NJIT, Newark, NJ, GPA: 3.75 *Dec. 2023*
- **Bachelor of Science, Finance** | Kean University, Elizabeth, NJ, GPA: 3.67 *May 2020*

Certifications

- **AWS Cloud Practitioner Certification** *Apr. 2024*
- **[PCEP-30-02] Certified Entry-Level Python Programmer** *Feb. 2024*
- **CompTIA Data+ Certification** *Jan. 2024*
- **Rutgers University Data Science Certification** *Nov. 2020*

Experience

- **Data Scientist** | Incedo, Florham Park, NJ *June 2024 - Present*
- **Data Teaching Assistant** | edX (University of Pennsylvania), Remote *May 2021 - Present*
  - **Classroom Leadership** : Led a class of 30 students in daily analytical coding exercises, leading to weekly KPI's consistently exceeding 90% in terms of overall qualifications, in contrast to 80% for my peers
  - **Tutoring** : Tutor students in the following analytical coding areas - **Python, R, SQL databases, and Excel**.
  - **Company Growth** : Improved company code using **Git** and presented to management findings yielding a 5% increase from employee appraisal.

Projects

- **NLP Custom ATS**: Coding, mining, and modeling a custom applicant tracking system with LinkedIn job description data and NLP techniques - scoring resumes from a range of 50 to 100 based on important words for upwards 7% increases in job matches between the resume and job description.
- **NFL Betting Model**: Scraped, analyzed, and modeled NFL data testing the ELO algorithm for prediction - tuning thresholds for a 9.56% increase in accurate bets - and using cross validation data for classification seeing our cost functions decrease 7.62% and accuracy increase 4.18%
- **Nintendo Sales Analysis**: Rolled out a comprehensive 2023 sales report using **SQL database, Pandas, and Python** to identify data trends such as Nintendo published 37% of the top 50 best-selling video games ever sold worldwide
- **TD Ameritrade Investment Data**: Built an automated **Python** program with **Pandas, BeautifulSoup, and Selenium** that scrapes companies' analytical data and stores it inside a **SQL** database to perform data analysis against competitors
- **Starbucks Coffee Recommendation**: Deployed a **Python Flask data user experience** with **Pandas and BeautifulSoup** to collect data in **MongoDB No SQL** database and provide customer recommendations by way of the KNN algorithm, Cosine Similarity, and Collaborative/Content filtering techniques
- **Big Data Flight Analysis**: Used 4 on demand **AWS EC2 cloud instances** and integrated **MySQL database, Hadoop, Oozie workflows** with **Java** to perform parallel MapReduce jobs across 11GB of flight data to reduce execution time from 850 seconds with 2 instances to 478 seconds with 4 instances

Technical Qualifications

- **SQL**: ETL, PostgreSQL, MySQL, API connection, Python integration, Star/Snowflake schema
- **Machine Learning**: Regression, Regularization, Cross Validation, Bayes, SVM, KNN, DTC, testing
- **Coding**: Python ML & app development testing, R statistical packages/visualization coding
- **Excel Power User**: Formulas, pivot tables, dashboard coding, VBA automation coding
- **Data Visualization**: D3.js, Leaflet, ggplot, Matplotlib, Plotly, Seaborn, Tableau