Justin Bates   Data Engineer

✉ jujbates@gmail.com   ☎ (818) 447-8632   ⚲ Santa Cruz, CA (Open to Remote)   **in** justin-j-bates   ⚯ jujbates

SUMMARY

Experienced data engineer with a strong track record of three years at a prominent data consulting firm. Skilled in Python, SQL, and proficient in data analysis. Specialized in designing, maintaining, and optimizing data infrastructure for comprehensive data lifecycle management. Excited to apply my expertise in a dynamic and challenging setting, utilizing my problem-solving prowess to deliver impactful solutions.

SKILLS

Languages: Python (pyspark, pandas, numpy, sqlalchemy, matplotlib, scikit-learn, pytorch, keras, tensorflow, seaborn), SQL

Platforms/Tools: AWS, GCP, Docker, Quicksight, Looker, Tableau, Git, Jupyter Notebook, Excel

Hard Skills: A/B Testing, Big Data (e.g., Hadoop, Spark), Business Analysis, Business Intelligence, Cluster Analysis, Data Collection, Data Mining, Data Modeling, Data Scraping, Data Validation, Data Visualization, Data Wrangling, Databases (MySQL, Postgres, Redshift, MS SQL), ETL, Data Pipeline, Exploratory Data Analysis

PROJECTS

Beer Recommendation Engine   *https://github.com/jujbates/Beer-Recommendation-System*
- Developed a beer recommendation engine with collaborative filtering techniques, namely K Nearest Neighbors and Singular Value Decomposition.
- Achieved low error rates (RMSE: 0.1317, MAE: 0.0547) with KNN model through 5-fold cross-validation, demonstrating effectiveness in predicting beer preferences.
- Leveraged the combined power of KNN and SVD to provide more relevant recommendations, indicating potential for improved user satisfaction and exploration of diverse beer options.

S&P Portfolio Optimization   *https://github.com/jujbates/Portfolio-Optimization*
- Designed an optimized portfolio aimed to maximize sharpe ratio, outpacing S&P 500 by over 10% in cumulative return with 2-year data analysis.
- Utilized web scraping and data preprocessing techniques to gather adjusted close prices of S&P 500 companies for portfolio optimization.
- Implemented bootstrap and Gibbs sampling methods to evaluate portfolio performance, achieving a significant increase in return.

Big Mountain Resort Analysis   *https://github.com/jujbates/Big-Mountain-Resort*
- Using data analysis and predictive modeling, the project recommends a weekend ticket price increase to $86 and extending the season to 130 days to achieve the desired profit margin.
- Insights from the models highlight White Mountain Resort's competitive advantages in skiable terrain and snowmaking, suggesting potential for further revenue growth through increased accessibility and pricing adjustments.

EXPERIENCE

ProductOps, *Data Scientist,* Santa Cruz, CA   *Jan. 2021 - Sept. 2023*
- Orchestrated end-to-end data management, including ingestion into data lake, ETL processes, and data warehousing using AWS Glue and Athena.
- Leveraged AWS CDK and implemented a robust CI/CD pipeline to ensure seamless operations throughout the data management lifecycle.
- Developed dynamic AWS Quicksight dashboards in collaboration with various teams to monitor progress against performance benchmarks.
- Implemented predictive models using AWS technologies (Kinesis Data Streams, Glue, Redshift, Athena, Step Functions, Sage Maker) to revolutionize machine repair downtime forecasting and real-time outlier detection in sensor data.
- Improved processing efficiency of claim data, optimizing the workflow from sales at the manufacturer level to installation by on-site contractors, and submission to utility companies nationwide, leading to streamlined operations and enhanced turnaround times.

Aceolution @ Google (Google Maps), *Field Special Data Analyst,* Mountain View, CA   *Jul. 2019 - Nov.. 2019*
- Collected and labeled location ground truth data for various projects to allow Google Engineers to improve models for Google Maps tools.
- Post-processing on GIS and GNSS datasets involves outlier detection, smoothing noisy data, interpolation to fill missing data, and data validation for ensuring reliable analysis results.

Warrior Media Inc. (Health & Fitness E-commerce Startup), *Full Stack Engineer*, Santa Cruz, CA   *Mar. 2017 - Aug. 2018*
- Over the first 12 months, efforts contributed to 3x subscriber growth and 200%+ revenue increase.
- Designed, developed, launched, and managed numerous apps for 3 corresponding company brands.
- Managed app development, test, and production deployment pipeline with Heroku and AWS.
- Developed dashboards and ETL workflow to process our KPIs such as CPA, CPC along with A/B testing results with various data sources.

Seagate Technology (eSecurity), *Security Intern*, Scotts Valley, CA   *Jun. 2016 - Sept. 2016*
- Developed an internal tool that analyzed network traffic in order to generate security reports.
- Trained international colleagues to operate the network analyzer and how to read and gain insights from the generated reports.

EDUCATION

University of California, Santa Cruz   *Sept. 2012 - June 2016*
Bachelor's of Science in *Computer Engineering* minor in *Computer Science*

Udacity, *Deep Learning Nanodegree* 🔗   *Jan. 2019*
Springboard, *Data Science Career Track*   *May 2020 - Dec. 2020*