Harshita Pasupulety

Boston, MA | +1(857-991-8754) | pasupulety.h@northeastern.edu | Linkedin | github

Education

Northeastern University — Boston, United States
*Masters of Science in Analytics — **GPA 3.7*** — *September 2023 – December 2025*

Andhra University — Vizag, India
Bachelors of Computer Science — *June 2019 – May 2022*

Technical Skills

Programming Languages: R, Python, SQL, Java, C++, SAS.
Technologies: Cloudera, Hadoop, Apache Spark, AWS, Microsoft Azure, Google Cloud Platform (GCP), Linux, HTML, CSS.
Cloud Technologies: AWS, Microsoft Azure, Google Cloud Platform (GCP).
Tools: Tableau, Power BI, ETL, Jupyter Notebook, R-Studio, MS Excel, Google Data Studio, Docker, Kubernetes.
Version Control: Git, GitHub.
Databases: MySQL, SQLite, NoSQL.
Machine Learning: Supervised Learning, Unsupervised Learning, Regression, Classification, Clustering, Decision Trees, Random Forest, SVM, Neural Networks, TensorFlow, Scikit-learn, Model Deployment.

Experience

Data Analyst Apprenticeship — January 2025 – Present
*Allagash Brewing Company — **Optimization of Wholesaler Shipping*** — *Boston, United States*

- Collaborating with **Supply Chain and Sales leadership** to optimize outbound shipments across **23 states** with **49 distribution points**, targeting a **10 percent reduction in freight costs**.
- Analyzing **lane rates vs. market rates**, historical average rates per mile, and carrier performance using **SQL, Python, R,** and **Excel** to identify cost-saving opportunities and recommend optimal carriers.
- Developing **interactive dashboards** in **Power BI** for inventory and shipment optimization, integrating data on **Days on Hand (DOH)**, upcoming purchase orders, and market trends to maximize truck capacity and improve supply chain efficiency.

Research Assistant — June 2020
Andhra University — *India*

- **Optimized** a data pipeline using **Python** and **SQL** to analyze smart city IoT data, boosting processing **efficiency by 30 percent** through improved data handling and query optimization.
- Applied **supervised machine learning models** (Random Forest, Logistic Regression) to predict peak traffic hours with **85 percent accuracy**, enhancing urban mobility strategies.
- Designed and implemented **interactive Tableau dashboards** for real-time data visualization, enabling **faster data-driven decisions** for traffic management and environmental monitoring.

Projects

Predictive Analysis of Spotify Song Popularity | *Python, scikit-learn, NumPy, Pandas, Matplotlib, Seaborn* — April 2024

- Designed a data pipeline using **Python** to process and clean data from the Spotify API, optimizing feature engineering for improved model accuracy with **scikit-learn**.
- Conducted extensive **exploratory data analysis (EDA)** using **Pandas** and **NumPy** to identify patterns in song attributes like danceability, energy, and loudness.
- Created insightful visualizations with **Matplotlib** and **Seaborn**, including heatmaps and scatter plots, to showcase trends and popularity metrics for business stakeholders —resulting in a **15 percent improvement in prediction accuracy**.

FIFA World Cup Analytics and Interactive Visualization | *Tableau, SQL, Excel, Data Studio, R* — March 2024

- Developed an interactive dashboard in **Tableau** to analyze FIFA World Cup data, providing insights into player performance, team strategies, and match outcomes
- Utilized **SQL** for data extraction, **Excel** for preprocessing, and created over **10 custom visualizations** like heat maps, bar charts, and line graphs for trend analysis
- Applied advanced Tableau techniques with **R** and **Google Data Studio**, incorporating calculated fields, parameter controls, and custom filters for dynamic data exploration

Sales Forecasting and Demand Analysis for Retail Chains | *Python, SQL, Power BI, Scikit-learn* — December 2023

- Developed a predictive model using time-series analysis techniques like **ARIMA** and **Prophet** to forecast sales trends and demand fluctuations for retail chains
- Analyzed historical sales data with **SQL** and **Pandas**, identifying seasonal patterns, product performance, and growth opportunities.
- Created an interactive **Power BI** dashboard to visualize key metrics, enabling data-driven decisions—achieving a **12 percent increase in forecasting accuracy** for demand predictions.