Sparsh Marwah

Boston, MA 02130 | marwahsparsh24@gmail.com | +1 (857) 225-9142 | LinkedIn | GitHub | Portfolio

Education

Northeastern University, Boston, MA, GPA: 3.75/4.0 **Expected May 2025**
Master of Science in Data Analytics Engineering
Relevant Coursework: Data Management in Analytics, Data Mining in Engineering, Machine Learning Operations

SRM Institute of Science and Technology, Chennai, India **May 2021**
Bachelor of Technology in Computer Science Engineering
Relevant Coursework: Data Structures, Data Science and Big Data Analysis, Object Oriented Analysis and Design
Publication: AI Music Generator (Research paper)

Technical Skills

Programming & Databases: Python, Advanced SQL, NoSQL (MongoDB), MySQL, PostgreSQL
Data Engineering & Governance: ETL pipelines, data cataloging, data classification, auditing data access, and data lineage
Data Governance Tools: Master Data Management (MDM), Azure Data Factory, Snowflake, Databricks
Data Visualization & Reporting: Power BI, Tableau, interactive dashboards
Cloud Platforms: Azure, Google Cloud Platform, AWS
Process Integration: Engineering Procurement and Construction (EPC) processes (Procure to Pay, Cost Management)
Certifications: Python (Programming, Data Structures), Data Science & AI, Intro to Cloud Data Analytics, ETL in Python and SQL

Work Experience

Data Analyst, Tredence Analytics Solutions Pvt. Ltd., Bengaluru, India **Jun 2021 - Jul 2023**

- Designed and optimized scalable data pipelines using **SQL**, **PySpark**, and **Airflow** to process datasets efficiently, achieving a 30% improvement in execution times
- Developed automated tracking systems to ensure data consistency, quality, and security for e-commerce projects, enhancing reliability in decision-making processes
- Partnered with cross-functional teams, including data scientists and product managers, to design data-driven strategies that boosted customer retention by 20%
- Built interactive **Tableau** dashboards, simplifying complex datasets into actionable insights for stakeholders and improving decision-making capabilities

Data Integration Intern, SJVN Ltd., Shimla, India **Jun 2019 - Aug 2019**

- Gathered information about their different energy forms, analyzed their powerhouse tools inventory data by developing **SQL** queries to understand the stock levels and sales trends
- Developed data integration workflows documentation to decide the entire lifecycle of the project, ensuring seamless dataflow
- Performed data quality audits and troubleshooting to ensure data accuracy, integrity, consistency, contributing to improved decision making and operational efficiency

Project Experience

Air Quality Prediction (Link) **Sep 2024 – Dec 2024**

- Developed machine learning models to predict PM2.5 and PM10 levels using OpenAQ **API**
- Applied advanced data preprocessing, feature engineering, and model selection techniques to create a reliable prediction system
- Designed and implemented a comprehensive **MLOps** pipeline using **Airflow**, automating data ingestion, model retraining, and deployment of new data seamlessly through Google Cloud Platform
- Leveraged **MLflow** for model tracking, drift detection, and version control on **GitHub**, automating drift detection to flag accuracy deviations and enable timely retraining, maintaining performance standards across deployments

IMDb Movie Data Analysis & Visualization (Link) **Jan 2024 – May 2024**

- Conducted data preprocessing, EDA, and time series analysis on IMDb movie data, achieving 91.2% predictive accuracy with Linear Regression and XGBoost
- Created an interactive Power BI dashboard with dynamic visuals, empowering stakeholders with actionable insights

YELP Review System (Link**)** **Feb 2024 – Apr 2024**

- Built and implemented recommender systems using **K-Nearest Neighbors** (KNN), Singular Value Decomposition (SVD), and Neural Networks, achieving 51% accuracy in predicting user preferences and recommending items
- Forecasted customer sentiments using Logistic Regression, Random Forest, and XGBoost, reaching 91.4% accuracy in sentiment classification and included thorough performance evaluation

Cricket Auction Player Performance Tracking System (Link**)** **Sep 2023 – Dec 2023**

- Collected, cleaned, and optimized player stats, linking performance data with team outcomes for key relationship analysis
- Developed **SQL** queries and used **Python** with **Matplotlib** and **Seaborn** for visualizations for performance insights like batting average, top 10 batsmen, top 10 bowlers & bowling average
- Utilized **NoSQL** database for unstructured data in the dataset