

Sparsh Marwah

Boston, MA 02130 | marwah.sp@northeastern.edu | +1 (857) 225-9142 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Education

Northeastern University, Boston, MA

May 2025

Master of Science in Data Analytics Engineering, GPA: 3.75/4.0

Relevant Coursework: Data Management in Analytics, Data Mining in Engineering, Machine Learning Operations, Financial Management for Engineers

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

Data Analysis & Visualization: Tableau, Power BI, Salesforce

Programming & Automation: Python (Pandas, PySpark, NumPy, TensorFlow), SQL

Database Management: Oracle, MS SQL Server, PostgreSQL, Redshift, Hive, Relational Databases

Data Systems & Cloud tools: ETL processes, AWS (Glue, S3, Athena), BigQuery, API Integration, Airflow, GCP

Statistical Analysis: Machine Learning, Forecasting, Modeling, Time Series Analysis

Collaboration Tools: Git, Jira, Confluence, SharePoint, Microsoft Office Suite

Certifications: Python ([Programming](#), [Data Structures](#)), [Data Science & AI](#), [Intro to Cloud Data Analytics](#), [ETL in Python and SQL](#)

Work Experience

Teaching Assistant, Northeastern University

Sep 2024 – Dec 2024

- Instructed students in **Python**, database management and data analysis, offering tailored guidance towards data visualization
- Directed labs and workshops on **Tableau**/storytelling with data; resolved problems for students to deliver 20+ projects

Data Analyst, Tredence Analytics Solutions Pvt. Ltd., Bengaluru, India

Jun 2021 - Jul 2023

- Developed **SQL**-based analytics solutions by querying massive datasets containing millions of rows to uncover trends, patterns, and actionable insights for business optimization. Ensured data accuracy and integrity while supporting strategic decision-making
- Managed **A/B testing** frameworks by designing and executing controlled experiments to evaluate various marketing strategies. Analyzed experiment results to ensure statistical significance, generating data-driven recommendations to improve campaign effectiveness
- Analyzed consumer shopping behavior across websites, apps, and other digital platforms to identify key trends in customer engagement, retention, and purchasing habits. Used findings to optimize advertising strategies and enhance marketing ROI
- Designed and automated **ETL** pipelines using **SQL**, **Python**, and **AWS Glue** to streamline data integration processes. Improved data accessibility and operational efficiency by automating repetitive tasks and ensuring timely data availability for analytics
- Optimized **SQL** queries, reducing execution times by 30% through query restructuring, indexing strategies, and performance tuning. Enhanced reporting processes for faster and more efficient data retrieval and analysis
- Collaborated with commercial teams to bridge the gap between technical data analysis and business needs. Communicated complex analytical findings in a clear, concise manner to both technical and non-technical stakeholders, facilitating informed decision-making

Data Analyst 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 ([View Project](#))

Sep 2024 – Dec 2024

- Developed **machine learning** models by using a time series data to predict PM2.5 and PM10 levels using OpenAQ API
- Applied advanced **data preprocessing**, **feature engineering**, & **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 ([View Project](#))

Jan 2024 – May 2024

- Executed comprehensive **data preprocessing**, **exploratory data analysis (EDA)**, & time series analysis on IMDb movie data, followed by predictive modeling using ML algorithms **Linear Regression** and **XGBoost**, achieving an accuracy of 91.2%
- Developed an interactive **Tableau** dashboard with dynamic visuals, slicers, and filters, enabling users to explore and analyze movie data, providing stakeholders with personalized insights and actionable results