

Harsh Mahajan

Boston • +1-(857)-390-5072 • harshmahajan08@gmail.com • linkedin.com/in/harsh-mahajan08 • harshmahajan.com

Education

Master of Science, Data Analytics Engineering April 2024
Northeastern University, Boston, Massachusetts GPA 3.72/4

- Relevant Coursework: Data Mining in Engineering, Statistical Learning in Engineering, Cloud Computing, Deterministic Operations Research, Computation and Visualization, Machine Learning Operations (MLOps), Data Management for Analytics

Bachelor of Technology, Data Science May 2022
NMIMS University, Mumbai, India GPA 3.53/4

- Relevant Coursework: Statistical Methods, Machine Learning (ML), Deep Learning (DL), Database Management Systems, Natural Language Processing, Computer vision (CV), Artificial Intelligence (AI), Marketing Management, Big Data, Data Gathering and Cleaning (ETL Process), Data Structures and Algorithms (DSA)

Technical Skills

Programming Languages: Python, SQL, R, NoSQL, LINGO, Git, Excel, C, C++, PL-SQL, Java

Tools: MySQL, Tableau, Power BI, Microsoft Office, Rshiny, Plotly, Kibana, IORTutorial, AWS, Azure, Google Cloud Platform (GCP)

Packages: Pandas, Numpy, TensorFlow, Keras, Scikit-Learn, sklearn, XGBoost, Matplotlib, Seaborn, Spark, NLTK, SpaCy, Cuda

Licenses & Certifications: AWS Academy Cloud Architecting, AWS Academy Cloud Foundations, Deep Learning and Computer Vision

Work Experience

Data Scientist April 2021 - July 2022
Aneka Labs, Pune, Maharashtra

- Extracted insights from **500GB+ datasets** to understand customer behavior and predict churn, leveraging AWS services for data storage, processing and model deployment tasks
- Utilized **AWS S3** for scalable data storage and **AWS Redshift** for high-performance data warehousing leveraging **AWS Glue** for data preparation and transformation tasks
- Collaborated cross-functionally with data engineering team to **optimize data pipelines** and ensure **data integrity**
- Developed machine learning models using **AWS SageMaker** and implemented **AWS Lambda** functions for automated data processing tasks and employed **AWS QuickSight for interactive data visualization**
- Identified key factors influencing customer engagement and the churn problem using machine learning models deployed on AWS to **achieve 85% accuracy** in churn prediction
- Implemented **data-driven recommendations** to provide analytical support, resulting in a **20% increase in customer retention**
- Presented insights to stakeholders using AWS QuickSight **dashboards**, ensuring alignment with business administration objectives

Projects

Inventory Management System (Excel) January 2024

- Collaborated with teammates for the development of an Automated Inventory Management System for optimization of inventory monitoring process and automation of stock updates, employing a PowerPoint presentation to showcase the findings
- Improved reporting efficiency by 20% through automated monthly reports featuring pivot tables and charts, streamlining data analysis and enhancing decision-making capabilities compared to manual method

Flight Cargo Planning (Python) December 2023

- Developed mathematical models using linear and mixed-integer programming techniques to optimize flight cargo distribution across 3 sections of an airplane
- Automated data processing and modelling for strategic planning, achieving 15% increase in efficiency in comparison to benchmark

Tourism Analysis in Europe (Tableau) March 2023

- Conducted in-depth analysis of various factors of tourism in Europe using data from 4 sources and visualization tools such as Tableau
- Created 3 dashboards consisting of 10+ graphs such as world heatmap and race bar chart to analyze tourist expenditure along with line graphs and pie charts to understand inland passenger transport in Europe and most visited tourist restaurants

Train Scheduling and Performance Analysis (Excel) April 2022

- Performed analysis using Excel, incorporating conditional formatting, lookup functions to calculate statistics of Indian railway ops using a dataset consisting of 100,00 rows, evaluating 10 factors
- Designed a robust KPI dashboard for track on-time performance and reliability, utilizing pivot tables and visualizations such as bar charts, heatmaps, line charts with rolling windows, discovering a 10% room for improvement in scheduling efficiency

Airlines Passenger Satisfaction (R) September 2021

- Developed a solution to identify customer satisfaction in airlines based on a dataset consisting of 130,000 records and 25 factors
- Implemented Logistic Regression to model and predict satisfaction level, using 70% of the data to train and 30% to validate and test
- Identified 6 key factors contributing highly to customer satisfaction in airlines, achieving 94% on key metrics like accuracy

Airlines Reservation System (Oracle APEX, MySQL) May 2019

- Created an airline reservation system on Oracle APEX, enabling seamless interaction for flight reservations
- Designed Entity-Relationship Diagrams and Schema Tables to create a relational model for the SQL database with 50+ fields
- Utilized SQL queries to retrieve data from the database and manage reservations and generate reports on flight status and passenger details