

Sushant Mahadik

Boston, MA | 857-544-8790 | sushantmahadik108@gmail.com | [Portfolio](#) | [LinkedIn](#) | [Tableau](#) | [GitHub](#)

SKILLS AND CERTIFICATIONS

Programming/Cloud:	SQL, Python, R, GCP(Healthcare API, Datastore, Dataflow), AWS (Redshift, EC2, S3), Microsoft Azure
DW/ETL:	Microsoft SQL Server, RDBMS, Google BigQuery, dbt, Snowflake, Apache Airflow, REST API, NoSQL, SSIS
BI/Analytics:	Tableau, Looker, Microsoft PowerBI, Microsoft Excel, Google Analytics, DAX, SAS, Power Query, VBA, JIRA
Certifications:	Apache Airflow Fundamentals, Azure AZ-900, Certified Scrum Master, Certified Scrum Product Owner

EDUCATION

Northeastern University, Boston, USA | Master of Science, Engineering Management (Data Analytics) September 2021 – May 2023
Courses: Database Management, Data Mining, Data Science Engineering, Probability and Statistics, Project Management, Visualization
University of Mumbai, Mumbai, India | Bachelor of Engineering, Electronics and Telecommunications August 2013 – June 2017
Courses: Data Compression & Encryption, Wireless Networks, OOPM, Image & Video processing, Advanced Statistics

WORK EXPERIENCE

Accenture Services Pvt. Ltd. **Mumbai, India**
Data Analytics Engineer January 2019 – December 2020

- Streamlined data migration and data warehousing process for GRP application by implementing ELT pipeline using **Apache Airflow** and **Snowflake**, increasing storage capacity by 3x, and reducing costs by 30%
- Reduced manual intervention by 80% by creating a data aggregation pipeline in Airflow to combine data from **Redshift**, **BigQuery**
- Collaborated with DE team to migrate terabyte-scale finance data from **Amazon Redshift** to Snowflake reducing infra cost by 30%
- Authored complex **SQL** queries to perform **data exploration** and **data validation** over 1 million transactional records using constructs such as joins, window functions, materialized views, and CTEs
- Performed data profiling and data modeling on financial data using **dbt** to compute infra costs and developed metrics using Looker which reduced the operational cost by approximately 150k dollars
- Assisted in the implementation of a customer loyalty program, analyzing **customer data** to identify high-value customers, and developing personalized offers to increase customer loyalty and retention
- Analyzed pain points, identified areas of improvement, and implemented **A/B testing**, increasing marketing effectiveness by 25%
- Curated dashboards using **Tableau** to visualize KPIs such as lead generation, conversion rates, revenue attribution, and user journeys for non-technical stakeholders, facilitating data-driven decision-making

Associate Data Analytics Engineer November 2017 – December 2018

- Built a data pipeline to redact PII data from 30k medical images concurrently using **GCP Healthcare API** and cloud **Datastore** making 10 million+ DICOM images GDPR compliant
- Conducted data mining and exploratory analysis on medical claims data using **Python**, resulting in the identification of billing errors and cost savings of \$500K annually, while improving revenue cycle efficiency by 20%
- Analyzed patient **demographic data** to identify target populations for preventive care programs, resulting in a 30% increase in patient participation and early detection of health risks

rDNA Corp. **Mumbai, India**
Data Analytics Intern June 2016 – May 2017

- Performed extraction, manipulation, and summarization of data from multiple sources such as **SPSS**, and Excel sheets using SQL
- Conducted **pricing analysis** to optimize pricing strategies, increasing profitability by 8% without compromising competitiveness
- Created interactive dashboards using **MS PowerBI**, to display days sales of inventory (DSI) KPI, improving sales efficiency by 40%

PROJECTS

Airbnb Listing Price Prediction (Python, Jupyter, Seaborn, Scikit-learn, Linear Regression, XGBoost, PCA) March 2023 – April 2023

- Developed a prediction model to estimate the prices of the Airbnb listings of major cities in the United States using linear regression, univariate and multivariate analysis, and XGBoost and achieved an r2 score of 89%

US Traffic Accidents Analysis (Python, Jupyter, Logistic Regression, Decision Tree, Random Forest) November 2022 – December 2022

- Built a prediction model to anticipate the occurrence of an accident using logistic regression, Decision Tree, and Random Forest by cleaning and feature engineering a large dataset and achieved an F-1 score of 0.92

Worldwide Coffee Market Analysis (R, Dplyr, Data Modeling, Data Profiling, Tidyr, Tableau, Excel Pivot) July 2022 – August 2022

- Performed analysis of the Coffee market around the world by collecting data using web scrapping, performing data transformation using R, and developing dashboards using Tableau. Bagged the best project award