

# RAUNAK SINGH MATHARU

Boston, MA | +1(857)-891- 4811 | [matharu.r@northeastern.edu](mailto:matharu.r@northeastern.edu) | [LinkedIn](#) | [GitHub](#) | [Tableau](#)

## SUMMARY

Have over three years of experience in data management, leveraging Agile methodologies with expertise includes requirements gathering, data analysis, and reporting, with advanced proficiency in ETL tools and cloud architectures such as AWS and Azure. I am skilled in data visualization using Tableau, Power BI and programming languages such as Python and SQL

## EDUCATION

### NORTHEASTERN UNIVERSITY

Master of Science (MS), Information System

September 2022 – April 2024

Boston, MA

### MUMBAI UNIVERSITY

Bachelor of Engineering (B.Tech.), Electronics and Telecommunications

June 2018 – June 2022

Mumbai, India

## WORK EXPERIENCE

### EPIC

May 2023 – April 2024

#### Data Analyst

Lowell, MA

- Processed real time healthcare and patient data with 1 billion+ rows via **Apache Spark ETL**, ingested into **Databricks** from different sources like **Snowflake** database, and **Flat files**
- Crafted 20+ insightful data visualizations in **Tableau** to analyze patient health metrics and assist healthcare leaders in understanding long-term trends in patient outcomes
- Influenced decision making by root cause analysis and time-series analysis to answer business questions related to GCM sensor failures, projecting a 20% reduction in downtime
- Utilized **advanced SQL** for comprehensive ad-hoc analysis of patient treatment cycles across 'Health Plans' and 'Revenue Cycle', improving operational analytics by 25%
- Developed a Random Forest-based machine learning model, achieving 92% accuracy in predicting claims for Containers, while optimizing accuracy by 15% through insightful data analysis and feature engineering
- Collaborated with cross-functional teams to define data requirements and develop solutions in Patient Flow and Digital Patient Experience modules, enhancing productivity and outcomes by 30%

### MAGNA INFOTECH

January 2021 – August 2022

#### Data Engineer

Mumbai, India

- Designed and Engineered ETL architecture and Data Pipelines from scratch using **Azure Data Factory**, **Azure Databricks** and **Logic Apps** enabling efficient loading of operational data into the **Data Warehouse** achieving an accuracy rate of 99.9% for loaded data
- Led successful data migration from on-premise to Cloud using **Databricks** and **Azure Synapse**, leading to a 25% increase in performance
- Pioneered comprehensive data solution that increased data log accuracy by 24% reducing errors and improving overall efficiency
- Integrated comprehensive data cleansing framework in **Python** and **SQL**, resulting in a 60% reduction in manual effort for cleaning data
- Identified changes to assortment, distribution, and marketing through financial and behavioral analysis enabling client to better target their customers, resulting in 25% increase in customer base
- Streamlined data processing by automating data cleaning transformation like **JSON parsing** with **Python Scripts**
- Eliminated manual efforts by **Power Query Scripts** and saved team from receiving ad-hoc requests by Interactive Dashboard in **PowerBI** using filters, parameters, and **DAX** functions and customization according to **business KPIs**

## PROJECTS

### Query Cuisine – Restaurant Data Scraping | SQL, Python (Data Cleaning, Web Scraping)

- Designed and established a real-time web scraping system using **BeautifulSoup** and **RestAPIs**, designing schema with 8 tables and integrating **CI/CD pipelines** for daily data extraction and management workflows using **Airflow**
- Developed user interface with Streamlit for robust interaction with **SQLite database**, and enhancements were subsequently deployed using FastAPI and created interactive visualizations with Looker

## SKILLS

**Core Competencies :** Data Lake Architecture, RFM Analysis, EDA Analysis, Normalization, Data Warehousing, CI/CD Pipeline

**Programming Languages :** PySpark, Python, SQL, Spark, Scala, C, Data Structures and Algorithm

**Cloud Services and Technologies :** Azure Databricks, Azure Data Factory, Azure Synapse, Apache Kafka, Azure Synapse, Data AWS Glue, Snowflake, AWS Redshift

**Machine Learning :** Feature Engineering, Exploratory Data Analysis, Classification, Decision Tree, SVM, Random Forest

**Database and visualization tools :** Tableau, Power BI, SQL (PostgreSQL, MySQL, MS SQL Server), MS Excel, MS Visio, Oracle

**Other Tools:** Microsoft Excel (Pivot table, VLOOKUP), PowerPoint, REST APIs, Jira

## PROFESSIONAL CERTIFICATIONS

- Microsoft Azure Data Engineer Associate (DP-203)
- Google Data Analytics
- Tableau Desktop Specialist
- Microsoft Azure Cloud Fundamentals (AZ-900)