Raj Sarode

Boston, MA | sarode.r@northeastern.edu | 617-653-0918 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

Northeastern University, Boston, MA   **Sep 2022 – May 2024**
Master of Science in Information Systems - Concentration in Data Engineering   **GPA:** 3.82/4.0
Relevant Coursework: Designing Data Architecture for Business Intelligence, Data Management and Database Design

University of Mumbai, Mumbai, India   **Aug 2018 - May 2022**
Bachelor of Engineering in Computer Engineering   **GPA:** 3.5/4.0
Relevant Coursework: Data Warehouse and Mining, Big Data Analytics, Database Management Systems

PROFESSIONAL EXPERIENCE

Northeastern University | Boston, Massachusetts
Graduate Teaching Assistant   **Jan 2024 – Present**
- Mentored **more than 200** students in Prompt Engineering and Data Science, resulting in improved performance and grades
- Assisted professor in teaching graduate students, with a focus on statistical and predictive modeling techniques
- Facilitated class discussions, provided feedback on assignments, and led study sessions to help grasp complex concepts
- Graded assignments, exams, and developed questionnaires; collaborated with the professor to assess class performance

Wizcom | Mumbai, India
Data Engineer Intern   **Feb 2022 – Jul 2022**
- Employed **SQL** and **Python** for advanced querying and data cleansing, achieving a significant reduction in data discrepancies
- Optimized data processing with **SnowSQL** to boost accuracy in financial reporting, revealing 15% more cost-saving opportunities
- Streamlined ETL workflows using **Alteryx** and **AWS Glue** to enhance risk analysis, achieving a 60% reduction in manual operations
- Designed and executed a data security protocol for sensitive financial information using **Amazon S3's** encryption and access management features minimizing data breaches
- Implemented **Tableau** dashboards for key financial metrics, revealing a 10% increase in operational efficiencies

Data Operations Intern   **May 2021 – Aug 2021**
- Performed **Spark SQL** query optimization by partition indexing resulting in a 70% faster response time from **Azure Delta Lake**
- Collaborated with cross-functional teams to implement real-time monitoring and data pipelines in **Azure Data Factory**, yielding a 20% increase in fraud prevention efficiency
- Developed T-SQL objects in **Azure Synapse Analytics**, enhancing **ETL** and data flow in **OLTP** environments for better performance
- Created comparative **Power BI** charts, illustrating variations in annual premiums and claims across diverse insurance products

PROJECTS

Web Wise - RAG Chatbot (Python, Streamlit, Chroma DB, LangChain, Flask, Beautiful Soup)   **Jan 2024 – Mar 2024**
- Orchestrated the integration of a RAG chatbot using **LangChain** to enable dynamic interaction with and parsing of website content
- Engineered a high-performance data processing pipeline utilizing **Python**, integrating web scraping, text preprocessing, vectorization, and **Chroma DB** for embedding storage, resulting in a substantial boost to the chatbot's functionality
- Designed an accessible chatbot interface with **Streamlit**, improving the overall user experience by making interactions engaging

Open Weather Cloud Analytics (AWS, Snowflake, Python, Sreamlit)   **Sep 2023 – Dec 2023**
- Built a Python/AWS ETL pipeline hosted on **EC2** to ingest data from 100+ US cities scraped from weather API into **Snowflake**
- Illustrated **data visualizations** to identify relationships between AQI, temperature, and humidity using **matplotlib, seaborn, plotly**
- Formulated an interactive **Streamlit** application to analyze relationships between climatic factors impacting major cities

Collision Scope: NYC Traffic Dynamics (Azure Data Studio, Talend, Navicat, Power BI, Tableau)   **Feb 2023 – May 2023**
- Analyzed NYC-MVC dataset using **Azure Data Studio** for insights, enhancing data accuracy with **SQL** data validation and cleaning
- Performed data integration through **Talend Open Studio** ETL processes, resulting in successful ingestion into the data model
- Generated advanced dashboards in **Tableau** and **Power BI** for NYC traffic, focusing on collision trends in high-risk areas

SKILLS

Programming	Python, SQL, T-SQL, R, JavaScript
Database Systems	Microsoft SQL Server, PostgreSQL, NoSQL, MongoDB, Spark SQL, Azure Data Studio, Azure Synapse Analytics
Visualization Tools	Tableau, Microsoft Power BI, MS Excel
Big Data	Alteryx, Talend Open Studio, Apache Spark, PySpark
Cloud Services	Amazon Web Services (AWS), Microsoft Azure, Snowflake, Azure Data Factory, Azure Databricks
Certifications	AWS Certified Cloud Practitioner

LEADERSHIP EXPERIENCE AND RECOGNITIONS

- DATA Club - Active Member   **Sep 2022 – Present**
- Computer Society of India - Treasurer   **Jul 2021 – June 2022**