

RAJ SARODE

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

EDUCATION

Master of Science in Information Systems (GPA: 3.82) | Northeastern University

Sept 2022 - May 2024

Coursework: Designing Data Architecture for Business Intelligence, Data Management and Database Design

Boston, MA

Bachelor of Engineering in Computer Engineering | University of Mumbai

Aug 2018 - May 2022

Coursework: Data Structures, Analysis of Algorithms, Machine Learning, Natural Language Processing

Mumbai, India

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Java, HTML, JavaScript

ETL and Business Intelligence Tools: Spark SQL, Power BI, Tableau, Alteryx, Talend, ER Studio(ETL), Navicat

Databases: RDBMS(MySQL, NoSQL, PostgreSQL, Microsoft SQL Server), NoSQL(MongoDB)

ML Libraries: Scikit-learn, Pandas, NumPy, Keras, Matplotlib, Plotly, Seaborn, PyTorch, Tensorflow

Cloud Platforms: AWS (S3, Redshift, EC2), Microsoft Azure, Google Cloud Platform (GCP)

Other Skills: Streamlit, API Integrations, Rest APIs, Docker, Microsoft Excel, Statistical Analysis, Data Modeling, Git, GitHub

WORK EXPERIENCE

Northeastern University | Graduate Teaching Assistant | Boston, MA

Jan 2024 - Present

- Instructed and supported **200+** students in "Prompt Engineering for Generative AI" and "Data Science Engineering Methods and Tools" courses, emphasizing actionable methods to improve prompt quality and project performance
- Assisted the instructor in translating academic concepts into hands-on exercises, significantly boosting students proficiency in leveraging generative AI for real-world applications

Wizcom Tech | Data Engineer Intern | Mumbai, India

Feb 2022 - Jul 2022

- Streamlined data cleaning using **SQL**, NumPy, Pandas, and **Amazon S3** for dataset storage, resulting in a **20%** error reduction and improved forecasting accuracy
- Applied advanced dimensional modeling techniques creating fact and dimension tables, to represent complex financial data using **PostgreSQL** in a **Snowflake** environment, enabling easy navigation and reporting
- Enhanced data management and reporting capability by automating routine ETL tasks by designing **Alteryx** workflows and **AWS Glue** for advanced ETL processes, reducing manual operations by up to **35%**
- Leveraged **Tableau** to create financial data-driven visualizations, optimizing ad spend based on financial metrics and resulting in a significant increase in marketing campaign ROI

Wizcom Tech | Data Operations Trainee | Mumbai, India

May 2021 - Aug 2021

- Optimized **Python** scripts and models for a client in the insurance sector by leveraging **Spark SQL** on distributed dataframes to meet their specific data needs
- Facilitated the integration of data sources by architecting advanced **SQL** data pipelines, enabling seamless data ingestion
- Generated comparative **Power BI** charts to illustrate variations in annual premiums and claims across diverse insurance products

PROJECTS

Web Wise - RAG Chatbot (Python, Streamlit, Chroma DB, LangChain, Flask, Beautiful Soup) |

- Orchestrated the integration of a Retrieval-Augmented Generation (RAG) chatbot, leveraging LangChain for website interaction, ensuring the system's capability to parse and utilize web content effectively
- Developed a data processing pipeline, incorporating techniques for web scraping, text preprocessing, vectorization, and Chroma DB for embedding storage, which significantly improved the chatbot's performance and response relevance
- Designed a user-centric interface through Streamlit, making the chatbot accessible and navigable for users

Collision Scope: NYC Traffic Dynamics (Azure Data Studio, Talend, Navicat, Power BI, Tableau) |

- Analyzed the robust NYC-MVC dataset with Azure Data Studio, executing SQL queries to extract valuable insights through data validation and cleaning techniques
- Led the optimization of data reliability and integration by employing Talend ETL processes, culminating in a **35%** boost in dataset accuracy, reflecting collision causes and effects
- Architected a cutting-edge dashboard through Tableau and Power BI, seamlessly integrating data from Azure Data Studio to display collision trends in high-risk areas

IOWA Liquor Sales Analysis (ER Studio, Alteryx, Talend, Tableau, Power BI) |

- Conducted a comprehensive analysis of a dataset comprising **25 million** liquor sales records from Iowa to identify trends and patterns
- Utilized Alteryx as an ETL tool for performing staging, modeling, and integration of the data, further improving data quality and significantly reducing inconsistencies
- Created impactful visualizations with Power BI and Tableau, resulting in efficient drill down for specific data points to gain a deeper understanding of trends