

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	<i>Sept 2022 - May 2024</i>
Coursework: Data Science, Advanced Data Science, Designing Data Architecture for Business Intelligence	Boston, MA
Bachelor of Engineering in Computer Engineering University of Mumbai	<i>Aug 2018 - May 2022</i>
Coursework: Machine Learning, Digital Signal and Image Processing, Natural Language Processing	Mumbai, India

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Java

ETL and Business Intelligence Tools: 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, Predictive Modeling, 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*

- Managed the extraction and integration of data pipelines from a travel agency client, prioritizing thorough data validation and cleansing to validate data accuracy and reliability, resulting in a **25%** increase in booking success rate
- Crafted **SQL** queries using **SQL Server 2008** to streamline the data extraction process, which included consolidating and migrating data sets from blob storage to **Snowflake** for more efficient data analysis
- Engineered an **Alteryx** based automation framework to optimize data workflows, enabling a seamless integration of data and simultaneously improved the speed and accuracy of expense forecasting by **20%**
- Leveraged **Power BI** and **Tableau** to create data-driven visualizations, optimizing ad spend allocation and resulting in a significant increase in marketing campaign ROI


Wizcom Tech | Data Operations Trainee | Mumbai, India *May 2021 - Aug 2021*

- Collaborated with clients under senior guidance to tailor **Python** scripts and models for data needs, achieving a **12%** increase in data quality and efficiency
- Facilitated the refinement of data extraction queries using **SQL**, ensuring the data is structured for further comprehensive analysis
- Generated comparative charts within **Power BI** to highlight annual revenue variations and across different brands in apparel industry


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 effectively parse and utilize web content
- Spearheaded the development of a sophisticated data processing pipeline, incorporating techniques for web scraping, text preprocessing, and vectorization, 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, thus facilitating immediate and relevant information retrieval based on website data

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

- Analyzed the robust NYC-MVC dataset to extract valuable insights and apply data validation and cleansing 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, which displayed collision trends in high-risk areas and provided actionable insights for traffic safety

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

- Conducted a comprehensive analysis of a massive dataset comprising **25 million** liquor sales records from Iowa to identify trends, anomalies, 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 that KPIs by **70%**, resulting in efficient drill down for specific data points to gain a deeper understanding of trends