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: Data Science, Advanced Data Science, Designing Data Architecture for Business Intelligence Boston, MA

Bachelor of Engineering in Computer Engineering | University of Mumbai *Aug 2018 - May 2022*
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*

- Assisted in teaching the principles of prompt engineering, including crafting effective prompts for **Large Language Models (LLMs)**. Managed and evaluated assignments, projects, and exams, providing constructive feedback to foster students growth
- Guided students through comprehensive **Python programming** assignments, concentrating on **data preparation** techniques for **machine learning** applications

Wizcom Tech | Data Engineer Intern | Mumbai, India *Feb 2022 - Jul 2022*

- Extracted and connected data pipelines from travel agency client with a focus on performing thorough data validation and cleansing to ensure the accuracy and reliability of the data
- Crafted **SQL** queries using **SQL Server 2008** to streamline the data extraction process, which included consolidating and migrating various 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%**
- Developed insightful visualizations using **Power BI** and **Tableau** to simplify complex data for informed decision-making

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

- Provided support in gathering client's business requirements and utilized **Python** for the initial data extraction and transformation process, focusing on adapting scripts and models to consumer data needs under senior supervision
- Contributed to refining data extraction queries using **SQL**, ensuring the data was structured for further comprehensive analysis
- Generated comparative charts within **Power BI** to highlight annual revenue variations and across different brands in apparel industry

PROJECTS

Collision Scope: Unveiling 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
- Employed ETL processes via Talend to refine and integrate data, thereby enhancing the reliability of the dataset to represent collision causes and effects accurately
- Generated a comprehensive dashboard using 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
- Leveraged 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

Twitter Sentiment Analysis (Python, NLP, SVM, Random Forest, XGBoost, Twitter API) |

- Builded an advanced XGBoost model that boasts a remarkable **96%** accuracy in discerning hate speech from non-hate speech, tested on a dataset exceeding **100,000** tweets
- Implemented NLP strategies by orchestrating an ensemble of algorithms, including SVM, Random Forest, and XGBoost, achieving analytical precision with an F1 score of **0.66**
- Efficiently managed data acquisition through the Twitter API, analyzing real-time streams and historical data, encompassing more than **2 years** of tweet data, to build a comprehensive sentiment analysis model