

# Raj Sarode

Seattle, WA | [rajsarode05@gmail.com](mailto:rajsarode05@gmail.com) | 617-653-0918 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

## EDUCATION

<b>Northeastern University</b> , Boston, MA <i>Master of Science in Information Systems - Concentration in Data Engineering</i> Relevant Coursework: Designing Data Architecture for Business Intelligence, Data Management and Database Design	<b>Sep 2022 – May 2024</b> GPA: 3.84/4.0
<b>University of Mumbai</b> , Mumbai, India <i>Bachelor of Engineering in Computer Engineering</i> Relevant Coursework: Data Warehouse and Mining, Big Data Analytics, Database Management Systems	<b>Aug 2018 – May 2022</b> GPA: 3.5/4.0

## PROFESSIONAL EXPERIENCE

<b>CREWASIS</b>   Brooklyn, New York <b>Data Science Intern</b>	<b>Jul 2024 – Present</b>
<ul style="list-style-type: none"><li>Integrated <b>ETL</b> with BeautifulSoup for social media &amp; news data, generating pickle files and <b>NLP</b> insights on <b>2M+</b> data points</li><li>Generated clusters to analyze relationships in textual data from various brands, improving accuracy by <b>60%</b> using <b>LSI</b> with <b>Python</b></li><li>Collaborated with the Marketing team to create <b>Power BI</b> reports, boosting campaign analysis and cutting reporting time by <b>15%</b></li></ul>	
<b>Utsavi Trading Co.</b>   Mumbai, India <b>Data Engineer Intern</b>	<b>Feb 2022 – Jul 2022</b>
<ul style="list-style-type: none"><li>Employed <b>SQL</b> and <b>Python</b> for advanced querying and data cleansing, achieving a significant reduction in data discrepancies</li><li>Optimized data processing with <b>SQL Server</b> to boost accuracy in financial reporting, revealing <b>20%</b> more cost-saving opportunities</li><li>Streamlined <b>ETL</b> workflows and improved risk analysis efficiency for pharmaceutical clients using <b>Alteryx</b>, resulting in a <b>15%</b> revenue growth and a <b>60%</b> reduction in manual operations</li><li>Automated monthly reporting with advanced <b>Excel (VLOOKUP, Pivot tables)</b>, reducing data compilation and analysis time by <b>35%</b></li><li>Implemented <b>Power BI</b> dashboards for financial analysis of <b>10,000+</b> metrics, enhancing data quality and <b>saving 18 hours/week</b></li></ul>	
<b>Data Operations Intern</b>	<b>May 2021 – Aug 2021</b>
<ul style="list-style-type: none"><li>Conducted data analysis for pharmaceutical clients with <b>1M+</b> monthly active users, employing <b>Python</b> and <b>SQL</b> to optimize sales and distribution strategies, leading to a <b>25%</b> increase in company revenue</li><li>Executed <b>ETL</b> using <b>SQL</b> and <b>Alteryx</b> to integrate terabytes of data, reducing data processing time by <b>40%</b> and increasing efficiency for the analytics team</li><li>Transformed data by integrating <b>10,000</b> records with <b>Excel VLOOKUP</b> and creating <b>Pivot tables</b> for real-time sales trend analysis</li><li>Engineered a dynamic inventory dashboard with <b>Tableau</b>, reducing overstock by <b>50%</b> and optimizing warehouse space by <b>10%</b></li><li>Analyzed <b>5M+</b> rows of shipping data via <b>SQL</b> and visualized shipping delays in <b>Tableau</b>, resulting in a <b>30%</b> drop in late deliveries</li></ul>	
<b>The Sparks Foundation</b>   Mumbai, India <b>Data Science and Business Analytics Intern</b>	<b>Feb 2021 – Mar 2021</b>
<ul style="list-style-type: none"><li>Developed a classification model for predicting heart disease using <b>ML</b> and analyzed the results on the basis of age, sex, etc</li><li>Performed exploratory data analysis and worked on predicting the results using <b>KNN</b> model and achieved an accuracy of <b>86%</b></li><li>Created visualizations like <b>scatter plots, heatmaps, and box plots</b> to show correlations between symptoms and heart disease</li></ul>	

## PROJECTS

<b>IMDB Project (Hadoop, MS SQL, Talend, Alteryx, Tableau)</b>	<b>Jan 2024 – Mar 2024</b>
<ul style="list-style-type: none"><li>Designed IMDB data warehouse model using <b>ER studio</b>; created a schema of DW model in <b>MS SQL</b> for loading data</li><li>Integrated, cleaned, mapped, and loaded around <b>30 million</b> records in the IMDB data warehouse using <b>Alteryx</b></li><li>Analyzed data by writing complex queries, performed data profiling, developed visualizations, and dashboards on <b>Tableau</b></li></ul>	
<b>IOWA Liquor Sales (Alteryx, SQL, Power BI)</b>	<b>Sep 2023 – Dec 2023</b>
<ul style="list-style-type: none"><li>Conducted in-depth analysis of <b>25M+</b> unclean liquor sales records from Iowa, identifying key trends and insights</li><li>Developed a dimensional model in <b>Alteryx</b>, reducing data processing time by <b>50%</b> and achieving <b>99.5%</b> accuracy</li><li>Created data visualizations in <b>Power BI</b>, improving <b>KPI</b> understanding by <b>70%</b> and decision-making efficiency by <b>30%</b></li></ul>	
<b>Collision Scope: NYC Traffic Dynamics (SQL, BigQuery, Azure Data Studio, Talend, Power BI, Tableau)</b>	<b>Feb 2023 – May 2023</b>
<ul style="list-style-type: none"><li>Extracted NYC-MVC data from <b>BigQuery</b> and ingested in <b>Azure Data Studio</b> for data validation and cleaning utilizing <b>SQL</b></li><li>Performed data integration through <b>Alteryx</b> ETL processes, resulting in successful ingestion into the data model</li><li>Generated advanced dashboards in <b>Tableau</b> and <b>Power BI</b> for NYC traffic, focusing on collision trends in high-risk areas</li></ul>	

## SKILLS

<b>Programming</b>	Python, SQL, R, JavaScript
<b>Database Systems</b>	MySQL, PostgreSQL, Microsoft SQL Server, BigQuery
<b>Visualization Tools</b>	Tableau, Microsoft Power BI, SSRS, MS Excel, Matplotlib, Seaborn
<b>Big Data</b>	Alteryx, Talend Open Studio, DBT, ER Studio
<b>Cloud Services</b>	Amazon Web Services (AWS), Microsoft Azure, Snowflake, Google Cloud Platform (GCP)
<b>Certifications</b>	AWS Certified Cloud Practitioner, Google Analytics, Alteryx Designer Core