SAGARIKA SARDESAI

LinkedIn: /sagarikasardesai | Tableau: /sagarikasardesai | GitHub: /sagarikasardesai | Website: sagarikasardesai.github.io
sagarikasardesai13@gmail.com | 6199537007 | San Francisco Bay Area | Open to Relocation

EDUCATION

MS, Data Science — **San Diego, CA**
UC San Diego, CGPA 3.90 — *Sep 2022 - Jun 2024*
Btech, Computer Science and Engineering — **Vellore, India**
Vellore Institute of Technology, CGPA 8.61 — *Jul 2016 - Jun 2020*

TECHNICAL SKILLS

Languages: SQL, Python, R, Cypher, PySpark, SparkSQL
Databases: MySQL, PostgreSQL, Neo4j, MongoDB.
Data Viz and BI Tools: Tableau, Qlik Sense, Advanced Excel (Formulae, XLOOKUP, VLOOKUP, Pivot Tables, Stat tools, Charts)
Cloud Technologies: AWS (Redshift, Glue, S3)
Other: Jupyter Notebook, Google Colab, Anaconda, Git (Version Control), Jira, Confluence

EXPERIENCE

Senior Operations Analyst — **Mountain View, CA**
Athelas — *Jul 2024 - Aug 2024*
- Utilized SQL, JavaScript, and Retool to maintain and build dynamic internal dashboards that integrated real-time data formatted in JSON.
- Monitored KPIs that tracked live medical claim submissions, denials, rejections, and reconciliations for 10+ medical facilities across 50 US states.
- Conducted gap analysis and root cause analysis to streamline medical claims submissions for 10+ medical facilities.

Data Science Research Assistant — **San Diego, CA**
UC San Diego — *May 2023 - Nov 2023*
- Ensured data accuracy by 5% through cleaning, exploration (pandas, numpy), and visualization (matplotlib, seaborn) in the Data Quality Report, revealing significant patterns and trends.
- Utilized exploratory data analysis, statistical analysis, and NLP techniques (nltk, transformers) such as VADER and RoBERTa (LLM) to measure sentiments and derive meaningful conclusions about consumer behavior.
- Derived valuable insights into consumer attitudes and decision-making regarding End-of-life choices.

Business Analyst — **Pune, India**
Credit Suisse — *May 2021 - Jun 2022*
- Analyzed production and test data sourced from company databases and software using SQL and Advanced Excel for company tooling defect mitigation.
- Performed ad-hoc analysis to support business strategy, reported & delivered actionable insights to cross-functional teams.
- Identified root causes, patterns, and trends from data analysis results, resulting in a 70% decrease in defects and unveiling of new product development areas.
- Led defect management processes to streamline Investment Trade Settlements lifecycle and business processes.

PERSONAL PROJECTS

Investment Scope of SBIR Awarded Companies
- Conducted in-depth analysis of the SBIR Awarded Companies containing 165,000+ records using Python and PostgreSQL, revealing patterns in industry sectors and year-wise award amounts across US agencies and branches.
- Improved data accuracy by 20% by cleaning methods including data type adjustments, keyword extraction, and imputation.
- Established relationships between US departments and agencies using Cypher (Neo4j) graphs, to enhance analysis context.
- Web scraped (Beautify Soup, Selenium Webdriver) recent news (Python data pipeline for ETL) about relevant companies, stored in MongoDB for further reference.

Covid19 Data exploration and Visualization
- Utilized PostgreSQL to analyze the OWID Covid-19 Dataset identifying patterns in global infection and mortality counts, country-specific statistics, and regional peak counts.
- Forecasted infection numbers for highly affected countries and presented insights visually with Tableau to aid decision-making.

Predictive Analysis for Detecting Credit Card Transaction Fraud
- Utilized supervised machine learning to detect credit card fraud, resulting in estimated savings of 21M USD.
- Tuned, trained, tested, and compared performance of Logistic Regression, Decision Trees, Random Forest, LGBM, LGBM with SMOTE, MLP classifier, Gradient Boosting Classifier, CatBoost, XGBoost, SVM.
- Analyzed credit card transactions and conducted data exploration, cleaning, and visualization (Matplotlib, Seaborn) to uncover insights.
- Feature engineered (pandas, numpy, scikit-learn) variables and feature selected (mlxtend, lightgbm) from existing dataset fields for comprehensive analysis.
- Employed the Kolmogorov-Smirnov statistical test to feature select relevant variables reducing training time by 30%.