# SAGARIKA SARDESAI

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

## EDUCATION

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

## EXPERIENCE

**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 80,000+ production and test data records** sourced from company databases and software **using SQL and Advanced Excel** for defect mitigation.
- **Performed ad-hoc analysis** to support business strategy, **reported & delivered actionable insights to cross-functional teams**.
- **Identified root causes, patterns, and trends** by conducting 300+ SWIFT CSDR Test Trades (Hold & Release), resulting in a 70% decrease in defects and unveiling new product development areas.
- Created 300+ **functional & user acceptance test (UAT) Quality Assurance (QA) B2B scenarios in HP-ALM**.
- Led defect management processes to streamline the SWIFT trade lifecycle 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 of 85,000+ rows, **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 with a 3% FDR.
- 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 90,000+ credit card transactions, conducting **data exploration, cleaning, and visualization (Matplotlib, Seaborn) to uncover insights**.
- **Feature engineered ((pandas, numpy, scikit-learn) over 1000 variables and selected (mlxtend, lightgbm)** from existing dataset fields for comprehensive analysis.
- Employed the Kolmogorov-Smirnov statistical test to **feature select 20-25 relevant variables reducing training time by 30%**, improving model accuracy and fraud detection capabilities.

### NY Property Fraud Detection (Unsupervised Anomaly Detection)

- Detected potential property tax fraud in 10k+ NY property records using **statistical analysis (z-scoring) & autoencoder(MLPRegressor)**.
- **Uncovered insights through data exploration, cleaning (pandas, numpy, scipy), and visualization (matplotlib, seaborn).**
- **Feature engineered 59 z-scaled variables** using statistical analysis to enhance fraud detection accuracy.
- Employed heatmaps to **visualize variables influencing high fraud scores**, gaining deeper insights into fraud patterns.

## TECHNICAL SKILLS

Languages: **SQL, Python**, R, Cypher, PySpark, SparkSQL
Data Vis and BI Tools: **Tableau, Qlik Sense, Advanced Excel (Advanced Formulas, XLOOKUP, VLOOKUP, Pivot Tables, Stat tools, Macros, Charts)**
Cloud Technologies: AWS (Redshift, Glue, S3)
Other: Jupyter Notebook, Google Colab, Anaconda, **Git (Version Control), Jira**, Confluence