

SAMRUDDHI GAWANDE

San Jose, CA | +1 (408) 829-8173 | gawande.samruddhi2@gmail.com | [LinkedIn](#) | [GitHub](#) | [Medium](#)

EDUCATION

M.S., Data Analytics, Department of Applied Data Science, San Jose State University, San Jose, CA
Bachelor of Engineering, Computer Engineering, University of Mumbai, India

Aug 2023 - May 2025
Aug 2017 - Jun 2021

TECHNICAL SKILLS

Programming Languages: Python (Numpy, Pandas, scikit-learn, NLTK), SQL, C#, C, JavaScript
Data Engineering & Big Data Technologies: ETL Pipeline Development, Apache Airflow, Apache Spark, Kafka, PySpark, Pulsar, Redis, Cassandra
Databases: MySQL, Snowflake, PostgreSQL, Google BigQuery, Microsoft SQL Server, NoSQL, Amazon RedShift
Data Visualization & BI: Tableau, PowerBI, QuickSight, Streamlit, Seaborn, Matplotlib, Excel, Dashboard Development
Cloud Platforms: Google Cloud Platform (GCP), Amazon Web Services (AWS), Azure
Tools & Methodologies: Git, Docker, CI/CD, Jira, Asana, Agile, Flask
Statistical Analysis & Modeling: Machine Learning, Time-Series Analysis, Statistical Modeling
Additional: Project Management, Stakeholder Collaboration, Data Modeling, Business Intelligence, KPI Development, Data Warehousing

WORK EXPERIENCE

Business Intelligence Engineer Intern | Amazon

May 2024 - Aug 2024

- Spearheaded the development and optimization of an AWS QuickSight dashboard for the Profitability Evaluator (PE) program, focusing on package cost analysis by integrating real-time data from an Amazon Redshift cluster and refining queries to reduce report generation time from **8 hours to 2 hours**, providing faster insights for finance stakeholders.
- Implemented Amazon Q to enable natural language querying and self-service analytics within the PE dashboard, increasing dashboard adoption by **50%** and facilitating the creation of user stories that drove more informed, data-driven decisions.
- Developed a comprehensive cost-benefit analysis for USPS vs. AMZL shipping methods across 100+ postal codes, uncovering critical insights into cost efficiency and revenue impact, and orchestrated extract-and-load jobs to enhance data management processes.

Data Engineer | Capgemini

Jul 2021 - Apr 2023

- Architected high-volume ETL pipelines and advanced transformations, implementing near-real-time ingestion from GE Proficy into SQL Server and cloud data platform on AWS with Python and SQL, resulting in a **20%** reduction in data anomalies and faster operational insights.
- Optimized SQL Server databases using strategic indexing (clustered, non-clustered, composite) and date-based partitioning, leveraging SQL Profiler and Query Store for performance tuning, which cut query latency by **30%** and maintained a **98%** data accuracy through robust validation.
- Led cross-functional collaboration and knowledge transfer on CI/CD, Docker, Apache Airflow, and Spark, enabling predictive analytics and real-time OEE dashboards that significantly improved data-driven decision-making across multiple production lines.

Data Analyst Intern | Trivia Softwares

Nov 2019 - Feb 2020

- Engineered a robust Student Management System in Python with a Tkinter GUI, integrating SQLite for secure and optimized data handling, which improved reliability and efficiency in day-to-day student data operations.
- Accelerated data retrieval by **5x** through streamlined CRUD operations, significantly boosting overall system responsiveness and user satisfaction.
- Implemented automated data extraction via BeautifulSoup to deliver daily quotes at application startup, and utilized Pandas, NumPy, and Matplotlib for advanced data analysis and visualization. Ran A/B tests on different UI layouts and quote categories to measure engagement, improving data-driven insights and user interaction by **25%**.

PROJECTS

Bias Beacon

Jan 2024 - May 2024

- Fine-tuned large language models using Python and PyTorch for high-accuracy political sentiment classification on media coverage and news content, leveraging differential learning rates to ensure balanced generalization on unseen data.
- Implemented mixed-precision training and gradient accumulation, reducing computational load and training time by **40%**, and integrated SHAP values for interpretable and explainable predictions used in analyzing bias patterns across political media outlets.

Tradewise

Jan 2024 - May 2024

- Developed and optimized a machine learning platform for real-time stock trading using Python (Pandas, NumPy, TensorFlow, Spark) for predictive modeling with deep learning and time series forecasting, enhancing stock returns through high-performance computing.
- Engineered and maintained a real-time stock monitoring system and automated data ingestion pipeline with Docker, Kafka, PySpark, and MLlib, processing over **1 million** data points daily, improving trading strategy accuracy by **20%**, and deploying an advanced analytics dashboard for 45+ stocks, boosting trading efficiency and portfolio management with MACD and RSI indicators.

US Accidents Data Analysis

Aug 2023 - Dec 2023

- Orchestrated the end-to-end analysis of extensive US traffic accidents data (2016 to March 2023), employing python for EDA, Mage for ETL, Google Cloud Platform for analysis, and Power BI to extract meaningful insights.
- Streamlined processing and visualization of a significant dataset in BigQuery, showcasing trends and factors in accidents via a Power BI dashboard, improving trend communication.

Reputation Monitoring using Sentiment Analysis

Feb 2020 - Apr 2021

- Developed a Flask web application using NLTK to capture, categorize, and analyze sentiments from **1,000,000+** tweets about a brand or keyword in real-time via Twitter API, enabling comprehensive reputation and media perception monitoring.
- Enhanced user experience by presenting sentiment analysis results through dynamic graphs, charts, and word clouds using JavaScript libraries, offering a real-time pulse on social media sentiment trends. The interface was designed for user-friendliness, incorporating HTML, CSS, Bootstrap.

CERTIFICATES & ACHIEVEMENTS

- Microsoft Certified:** Azure Fundamentals, Certificates issued by DataCamp and Great Learning for **SQL** and **Tableau**. ([LinkedIn](#))
- Rockstar Award** recipient at Capgemini, India for outstanding performance and lasting contributions.
- Head of Public Relations** at CSI, VIT: Managed communication strategies and stakeholder relations.
- Adobe Student Ambassador:** Led campus initiatives and promoted products, enhancing Adobe's academic visibility.