

Samprithi Ravisanker

Summary

saamprithiravisanker@gmail.com | Website | LinkedIn

Data Scientist with 4+ years of experience building scalable data pipelines, predictive and Bayesian models, and analytics solutions. Skilled in Python, SQL, PySpark, Databricks, Power BI, and QlikView, with expertise in forecasting, capacity planning, customer analytics, and A/B testing

Experience

Digital Data Scientist | CHEP, Orlando

Feb 2025 – Present

- Built scalable **Databricks** data pipelines using **PySpark** and **SQL** to process 10M+ logistics records, implementing validation checks and supporting investigative analytics that helped recover thousands of misused assets
- Developed **capacity planning** and **Bayesian** statistical models, to quantify operational drivers, estimate plant-level performance, and improve production efficiency by 7% through data-driven benchmarking and forecasting
- Applied **unsupervised learning** and statistical analysis to detect anomalous asset reuse and leakage patterns, while designing executive dashboards with 15+ risk and performance KPIs used by 20+ stakeholders, contributing to \$2M+ in contract renegotiation impact and proactive operational decision-making

Digital Operations Analyst | CHEP, Orlando

Sept 2023 – Feb 2025

- Developed and implemented automated **web scraping** pipelines using Python tools (**Beautiful Soup**, **Selenium**), reducing manual data collection efforts by **80%**, saving over **20 hours per 1000 locations**
- Executed **ETL** workflows by extracting and analyzing large datasets using complex **SQL** queries and performing data cleaning with Python libraries like **Pandas** and **NumPy** to enhance data usability
- Designed and developed data-driven **dashboards in Power BI** to visualize critical KPIs, such as leaks, reuse, maps and cycle time, enabling data-driven decision-making

Data Analyst | SS HVAC Engineers, India

Apr 2020 – Jul 2021

- Analyzed data from multiple sources, extracting insights on company performance metrics; built **ETL pipelines** to transform and integrate data, streamlining analysis and reporting resulting in a 40% reduction in data processing time
- Actively participated with cross-functional teams to design **Tactical Dashboards** for business managers using **Qlikview** to monitor departmental and team performance
- Collaborated with clients and team members to design week on week **Analytical Dashboards** to provide actionable insights that helped to increase the company's ROI

Data Analytics Intern | Arvind Internet, India

Jun 2019 – Mar 2020

- Collected and integrated data for 8M customers across 50+ beauty and wellness products and services from POS systems, validated APIs using Postman, and built data pipelines into **QlikView** to enable customer and product analytics
- Developed **Customer Lifetime Value (CLV)** and customer **churn models** using custom customer segmentation, and designed A/B tests for marketing and retention campaigns, increasing customer retention by 27%
- Performed customer behavior, segmentation, and product performance analysis to identify revenue opportunities, and partnered with cross-functional teams to build dashboards with **30+ KPIs**, enabling faster, data-driven business decisions

Education

Masters in Data Analytics | University of Central Florida, Orlando

Aug 2021 - Aug 2023

BE in Computer Science | Bannari Amman Institute of Technology, India

Aug 2016 - Mar 2020

Technical Skills

Programming Languages and Analytics - Python, R, MySQL, MSSQL, Tableau, Microsoft Power BI, DAX, QlikView
Others - Bit Bucket, Databricks, Jira, PyTorch, Tensorflow, Keras, TFLite, AWS, GCP

Projects

User Interest Evolution and Topic Drift Detection System

Python, Hugging Face, SBERT, OpenAI embeddings, Clustering

- Built a machine learning system using large-scale public text data to detect emerging, declining, and shifting topics over time. Applied embedding-based clustering and temporal drift analysis to proactively surface new behavioral and abuse trends.

LLM-Powered Email Categorization

Python, SQLite, LiteRT

- Built an offline-first email triage system using a locally deployed, quantized LLM to classify and label emails with low-latency inference. Added schema-constrained outputs and an SQLite-backed audit/evaluation workflow to minimize misclassification and enable safe automation.