Sameer Enjapuri

Tampa, FL-33613 | (813)389-8792 | enjapurisameer@gmail.com | [LinkedIn](#) | [GitHub](#) | [Tableau](#) | [Portfolio](#)

EDUCATION

University of South Florida
Master of Science in Business Analytics and Information Systems Aug 2023 – May 2025 | *GPA – 3.87/4*

Shah & Anchor Kutchhi Engineering College
Bachelor of Engineering Jul 2014 - Jun 2018 | *CGPA – 8.37/10*

TECHNICAL SKILLS

Languages:	Python (Pandas, NumPy, Scikit Learn etc.), R, C, SQL, HTML, CSS
Databases:	Oracle, MySQL, PostgreSQL, MongoDB, MS SQL
Data Warehousing:	ETL (Extract, Transform, Load), Apache Kafka, Confluent Kafka, Informatica PowerCenter, Databricks
Tools:	GitHub, Jenkins (CI/CD), Jupyter Notebook, RStudio, Apache Airflow, AWS (Amazon S3, AWS Glue, AWS Lambda, AWS Athena), Docker
Data Analytics:	Exploratory Data Analysis, Data Cleaning, Data Visualization, Data Modeling, Statistical Analysis
Data Visualization:	Tableau, MS Power BI, MS Excel, Qlik Sense, Looker Studio, Amazon QuickSight
Project Management:	Agile, PowerPoint, Scrum, JIRA, Kanban, Confluence, SDLC

WORK EXPERIENCE

Graduate Research Assistant (University of South Florida) Jan 2024 - Present
- Spearheaded the development of comprehensive reports using Tableau to **analyze Tampa Bay's economic performance** in areas such as economic outcomes, affordability, and talent pipeline relative to 19 Metropolitan Statistical Areas (MSAs).
- Collaborated with faculty and peers demonstrating strong communication skills to **gather, analyze, and visualize data**, providing insights that informed policy recommendations for enhancing the region's economic standing.
- Demonstrated proficiency in Tableau for **data presentation**, creating visually compelling dashboards and reports, effectively conveying complex economic data.

Data Analyst, Capgemini (Client – Länsförsäkringar Bank) Oct 2022 - Jul 2023
- Streamlined data ingestion and processing using **Python**-based APIs and Confluent Kafka connectors, achieving a **40% reduction** in processing time and significantly enhancing data preparation efficiency for downstream analytics.
- Developed and maintained **Spark**-based ETL processes, **enhancing data quality and reliability by 25%**. These processes improved data **throughput by 30%** and supported real-time analytics, driving business decisions.
- Applied agile methodologies to data analysis projects, emphasizing iterative development and continuous improvement. This approach resulted in a **30% reduction** in project delivery time, ensuring timely delivery of analytical solutions to meet business objectives.

Data Analyst, Cognizant (Client – Adidas) Feb 2022 - Sep 2022
- Engineered and implemented a real-time data integration solution using **Confluent Kafka**, resulting in a **25% increase** in data streaming efficiency and enabling continuous data streaming and analytics. This **Change Data Capture** (CDC) system facilitated timely insights into market trends and customer behaviors.
- Optimized data transformation workflows and **reduced processing time by 30%** by analyzing and processing large-scale data sets using **PySpark**.
- Analyzed rental car market trends using statistical methods and data visualization tools, resulting in strategic recommendations that drove a **15% increase** in revenue. This involved extracting, transforming, and analyzing large datasets to identify key trends and opportunities for business growth.

Data Analyst, Tata Consultancy Services (Client – Avis Budget Group) Sep 2018 - Jan 2022
- Orchestrated the entire Software Development Life Cycle (SDLC) for **ETL data pipelines** in Informatica PowerCenter, from requirements gathering to production deployment. Managed a cross-functional team of four members, achieving a **95% success rate** in transitioning ETL workflows to Kafka, ensuring seamless data integration and processing.
- Utilized advanced **data modeling** techniques and SQL Trace to **optimize database performance by 15%**, enhancing the efficiency of data retrieval and analysis processes. Applied business intelligence principles to ensure effective data modeling and mapping, facilitating accurate and insightful reporting.
- Engineered and fine-tuned ETL processes and Kafka topics to ingest real-time data from IoT sensors on rental vehicles, **reducing data inaccuracies by 30%** and improving data quality for subsequent analytics and reporting tasks.
- Cultivated and nurtured client relationships, leveraging strong networking skills to facilitate project success and satisfaction. Engaged with stakeholders to understand their data needs and requirements, ensuring alignment with project objectives, and delivering actionable insights that drive business value, resulting in a **20% increase** in client satisfaction and a **15% improvement** in project deliverables' alignment with strategic goals.

PROJECTS

Realtime Data Streaming Pipeline (Apache Kafka, Spark, Airflow, Docker, Cassandra, Python)
- Developed an end-to-end Data Engineering project using Spark, Kafka, Airflow, Docker, Cassandra, and Python. The project involved fetching data from an API via scheduled **scripts**, sending it to Kafka using **Airflow**, processing it with Spark Structured Streaming, and storing it in Cassandra.
- All components were containerized with Docker for seamless integration and scalability, resulting in a **30% reduction** in data processing time.

Data Engineering YouTube Analysis (AWS Services - Athena, Glue, S3, QuickSight)
- Designed a scalable data pipeline using AWS Glue ETL to transform over **100k records**, securely stored in an S3 bucket. Employed **Lambda functions** to clean and preprocess JSON data, converting it into optimized Parquet format, and automated the process with a file load trigger, **reducing manual work by 10%**.
- Used QuickSight to analyze engagement metrics, identifying top-performing videos and enhancing marketing strategies.

Movie Shows Clustering and Recommendation System (EDA, Clustering, Recommender System)
- Conducted **exploratory data analysis** on Netflix content, revealing key trends such as the dominance of movies over TV shows and the significant increase in content produced in the United States over time.
- Developed and optimized **clustering algorithms** (K-Means, Agglomerative) and a content-based recommender system using cosine similarity, providing 10 personalized show recommendations based on user preferences.

CERTIFICATIONS

- Microsoft Certified: Power BI Data Analyst Associate