

SANATH KUMAR BAJJURI

sanathkumar.bajjuri@stonybrook.edu | +1 3479016516 | [LinkedIn](#) | [Github](#)

EDUCATION

Stony Brook University

Master of Science, Data Science

Stony Brook, NY

Aug 2023 - May 2025

WORK EXPERIENCE

Data Engineering Analyst, ZeMoSo Technologies Pvt Ltd

Jul 2022 – Jul 2023

- Orchestrated and developed **ETL pipelines** using **AWS Glue** for automated data cataloging, and **Python** in the backend for data loads in downstream, reducing **10%** the DOJ call-in rate for field technicians in the last quarter.
- Applied transformations like harmonization, deduplication, Change data capture using Python and lambda functions on usage data extracted from Athena ensuring accurate dataloads to **AWS S3 buckets** and **AWS redshift**.
- Developed **9+ Tableau reports** for digital platform insights analyzing call rates, work-order volumes, and other Tableau dashboards such as Techmobile score card for analyzing metrics on quantified issues and activation attempts.
- Delivered recurring production reports and **Tableau dashboards** on viewer metrics, engagement metrics, region-wise content performance, and financial analytics for Connectivity Insights product and collaborated with the team for feature enhancements deployed on a **30 million** internet customer base.
- Leveraged technical expertise to support data-related aspects of the ad-insertion process in **Business Intelligence** applications by integrating the data from different sources before the **ETL** process.

Data Engineering Associate, ZeMoSo Technologies Pvt Ltd

Jun 2021 – Jun 2022

- Plunged dashboard refresh time by **12%** through customer data unification by revamping older **Power BI dashboards** and optimizing **SQL** queries using CTE, indexing, ingestion scripts, and user-defined functions to track KPIs, ensuring uptime.
 - Streamlined repetitive **data collection, processing, & analysis** through data pipelines to intermediate **S3 buckets** using **AWS Lambda & Python** scripts, resulting in a **30%** reduction in processing time.
 - Implemented self-service **EMR** clusters for trigger ETL jobs in a test environment using git hooks, saving **5 hours** of manual efforts for testing data products.
 - Maintained **Snowflake databases**, schemas, tables, views, and stored procedures, ensuring data consistency and accuracy throughout the **data pipeline**.
-

PROJECTS

Netflix Movie Recommendation System | Python- NumPy, Pandas, Scikit-Learn, Matplotlib

- Designed a movie recommendation system within Netflix through correlations in Python by dealing with over 40M+ rows accounting for over 17K movies and 500K customers data to recommend best shows to users based on self & others behavior, produced insights using Power BI visualizations.

Data science and Time Series Forecasting | SAS JMP, NLP, Tableau

- Utilized ARIMA & ARIMAX models in SAS Studio to forecast weekly sales, interpret trends, and seasonality for 45 Walmart stores in the US, and offered department-specific product availability recommendations through some Power BI dashboards based on 10 years of data.
-

TECHNICAL SKILLS

Languages : Python, R Programming, SAS, Java, C, C++, Data Structures and Algorithms, JavaScript & TypeScript.

Libraries: Pandas, NumPy, PySpark, Seaborn, Matplotlib, Scikit-Learn, openCv, Keras, Tensor Flow, SciPy, Boto3, Shiny.

Databases : Oracle, MS SQL Server, Cassandra, PostgreSQL, MongoDB, Snowflake, Database design.

Cloud Technologies : AWS(EC2, S3, Red Shift, RDS, Glue, Cloud Watch, EMR), Docker, linux, Google Cloud Platform.

Developer Tools : Jupyter, Spyder, Tableau, Power BI, Kafka, Splunk, Data Bricks, DBeaver, R Studio, GitHub, VS Code, GCP.

Statistical Skills : Exploratory Data Analysis, Regression Analysis, Classification Analysis, Clustering, Data Mining, PCA, Decision Trees, Neural Networks, Discriminant Analysis, K Nearest Neighbors, Naive Bayes, Text Analytics, Probability Distributions, Time Series Forecasting, Search Engine Optimization, ETL, Data Warehousing, Machine Learning Algorithms, NLP, Big Data Analysis, Artificial Intelligence, Data Analytics and statistics, Deep Learning, supervised learning, statistical methods.

CODING PROFILES

Leetcode: 260+ problems

HackerRank: 30,000+ Hackos (500+ problems)

Interviewbit: 32,400+ (200+ problems)