

## Professional Experience

### Data Consultant, Systems Limited

Aug'23 – Present

- Developed a **Flask**-based data ingestion web application leveraging **Snowpark** for streamlined data integration
- Implemented robust data integrity checks in the Python backend, utilizing **Pandera** for schema validation against sample files stored in **Azure Blob Storage**
- Orchestrated access control via an admin interface, allowing precise management of user permissions and file upload privileges, ensuring **data security and compliance**
- Implemented **CI/CD** pipelines on **Azure DevOps**, automatically refreshing the web application on **Azure App Service** upon changes being pushed
- Analyzed various HR data tables hosted on Snowflake to extract features for **multivariate forecasting** models
- Employed **AzureML datasets** and Databricks managed **MLflow** to version datasets and models, ensuring reproducibility and traceability
- Utilized the **Snowflake** table within a **Plotly Dash** web application hosted on Azure App Service to present a dashboard, enabling HR business users to derive insights from the predictions
- Developed an end-to-end credit scoring pipeline utilizing **PySpark** and **DataRobot**, enabling predictive analysis of high-risk customers and empowering the client to make informed decisions regarding loan disbursements

### Data Scientist, Afiniti

Jan'21 – Aug'23

- Developed CRM **SCD** and processed call outcomes data for a telecom client. Automated **ETL jobs** using **Talend** and **PostgreSQL** stored procedures on **Greenplum**, with triggers for immediate data processing upon receipt
- Engineered **hierarchical join logic** for multiple tables and created corresponding stored procedures. Developed materialized views to streamline data access from joined tables
- Implemented and configured **data integrity alerts**, identifying and resolving quality issues in billions of rows and terabytes of data
- Built feature engineering pipelines using **SQL**, **Talend**, and **Python**. Engineered advanced features in Python and CRM features in SQL. Consolidated features into a MySQL view for AI model consumption.
- Created **Bayesian regression** models using **R** and **Stan**. Utilized decision trees for **customer segmentation** and **Item Response Theory** (IRT) models to evaluate customer purchasing potential and agent selling capabilities.
- Developed a **notification bot** for Webex to provide real-time updates on KPIs and metrics for newly received data, facilitating immediate data validation
- Generated data integrity and **BI reports** using **RMarkdown**. Distributed reports incorporating predictive model outcomes to stakeholders for monitoring AI model progress.
- Led key accounts in Europe and North America, contributing to \$4.0 million in monthly revenue
- Designed and maintained **CI/CD** pipelines, ensuring runtime and historical data mismatches remained below 1% across billions of records
- Utilized **Apache Superset** and **SQL** to create dashboards and sensors, **visualizing data integrity, data drift, join rates, and mismatches**, ensuring a user-friendly experience for stakeholders

## Education

### B.S. Mechanical Engineering, Ghulam Ishaq Khan Institute

Aug'16 – July'20

Relevant Courses: Engineering statistics, Operations Management, Numerical Analysis, Linear Algebra, Robotics, Control Systems, Microcontrollers, Digital Logic Design,

## Skillset

**Programming Languages:** Python | R | SQL | PostgreSQL | noSQL | Stan | Bash

**Data Science:** Tabular ML | Deep learning | Large Language Models | RAG | Vector DB | Regression and Classification | Tree based algorithms | Bayesian Causal Modeling

**Databases:** MySQL | GreenPlum | Snowflake | MongoDB

**Cloud Technologies:** Azure (App Service | Databricks | ML | Blob Storage | DevOps | Pipelines), AWS (Rekognition | Comprehend | Personalize | Lambda | S3 | EC2 | CloudWatch | Glue | Kinesis | Boto3)

**Tools:** RStudio | Anaconda | Jupyter | VSCode | Selenium | CI/CD | Airflow | MLOps | Git | Docker | AirFlow | Talend

**Soft Skills:** Agile | Scrum | Leadership | Communication | Problem Solving | Strategic Planning

## Projects

### Real-Time Data Pipeline on AWS

Developed a comprehensive real-time data pipeline to ingest, process, and visualize data using AWS services. Automated data ingestion from a local generator to Kinesis, transformed and stored data in S3 as Parquet files, cataloged data with Glue Crawler, enabled querying via Athena, and created dashboards with QuickSight.

**Technologies Used:** AWS S3, Kinesis Data Streams, Glue ETL, Glue Crawler, Athena, QuickSight, Python, Boto3

### Sales Maturity Forecasting

Pioneered a sales maturity forecasting system to predict customer holdings 30 days after a call, ensuring compliance with the client's requirement of a 30-day maturity period. Predictions were stored in a Greenplum (Postgres) database, and an interactive dashboard was created using RMarkdown, providing real-time insights on current gain numbers and 30-day predictions to stakeholders.

**Technologies Used:** Python, R (data.table, tidyverse, RMarkdown, knitr), Greenplum (PostgreSQL)