Rahul Kengeri

(516) 923-1045 | rk1087@rit.edu | rahulkengeri.com | GitHub | LinkedIn

SKILLS

Programming Languages: Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn) | SQL | R | Java | PySpark
Databases: MySQL | MongoDB | PostgreSQL | Snowflake
Cloud Platforms: AWS (EC2, S3, EMR, Lambda) | Azure (Synapse, Data Factory, Stream Analytics)
Data Engineering and Visualization: Apache (Airflow, Kafka, Spark) | Tableau | PowerBI | Excel | ETL Pipeline | Data Modeling

WORK EXPERIENCE

Analytics Engineer Intern   *Bengaluru, KA*
Radiant Info Systems Ltd   *August 2021 – February 2022*

- Collaborated on an automated **ELT data pipeline** ingesting real-time passenger data from 900+ city buses into **Snowflake data warehouse**, using **Apache Kafka** for streaming data and **Spark** for transformations.
- Streamlined data ingestion by implementing **micro-batching** in the Kafka Connector configuration which reduced the number of individual database connections and **enhanced throughput** by 15% during peak hours which enabled handling **8M+ events** daily.
- Overhauled the existing data warehouse in Snowflake by implementing a **Snowflake schema** which resulted in a **12% decrease in storage needs**, reducing costs.
- Created 5+ **data-driven** dashboards featuring comprehensive **analytics and visualizations**, providing **management teams** with **actionable insights** which in turn enhanced experience for 4M+ daily commuters.

PROJECTS

Supermarket Data Pipeline and Analysis Using Azure Services | (Azure, PySpark, Power BI) | Link

- Orchestrated an end-to-end **data processing pipeline** leveraging Azure services to analyze 10,000+ sales data from a nationwide superstore, establishing an automated pipeline for deriving sales insights.
- Utilized Data Factory to ingest raw data, performed data cleaning, and transformation using **PySpark** on Azure Synapse, and deployed **Dedicated SQL Pool** as a data warehouse, enabling ad-hoc and advanced querying.
- Designed an intuitive **Power BI dashboard** containing key findings (optimal discount range, lucrative product categories, sales figures) for stakeholders to drive more sales and profitability.

Global Hotels and Resorts (GHR) Data Analysis | (Python, Pandas, Matplotlib, Excel, Tableau) | Link

- Analyzed large dataset of 94K+ hotel bookings, **engineered 10+ new features** which revealed a $272K loss due to free upgrades and $1.2M cancellation loss over the study period.
- Developed 2 versions of the **CatBoost** model predicting hotel upgrades with 0.40 F1-score on an **imbalanced dataset** (85K:8.5K) - one optimized for recall to estimate maximum revenue leakage, the other for precision to identify customers likely to upgrade.
- Presented strategies to **grow revenue** from direct bookings by **15% -30 %**, improve room occupancy, resource planning, cancellation policy amendments to **Corporate Directors, VPs, and CIOs** from the hotel industry.

Data Science Job Market Analysis: ETL Pipeline and Visualization| (AWS, PySpark, Airflow, Snowflake, Tableau) | Link

- Engineered an **ETL data pipeline** using **Airflow** for a 10k-row dataset containing jobs and salaries in data science domain, leveraging Apache PySpark, AWS services (EC2, EMR, S3), and Snowflake, implementing **batch processing**.
- Performed **Star Schema** data modeling using SQL to build **data warehouse for analytics**, visualized processed data in Tableau.
- Built tailored **Tableau dashboards** with 5+ key **insights** for job seekers in data science domain, enhancing career decision-making.

EDUCATION

Master Of Data Science   *Rochester, NY*
Rochester Institute of Technology   *August 2022 – December 2024*

Bachelor of Computer Science and Engineering   *Bengaluru, KA*
Global Academy of Technology   *August 2018 – July 2022*

ACHIEVEMENTS

- Winner of both Round 1 and Round 2 of Saunders Business Analytics Competition 2023-2024
- 2nd Place in BrickHack X hackathon in the 'Most Commercially Viable Product' category