Rahul Kengeri

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

EDUCATION

ROCHESTER INSTITUTE OF TECHNOLOGY   Rochester, United States
Master of Data Science   August 2022 – December 2024
Relevant Coursework: Foundations of Data Science & Analytics, Software Construction,
Data Management and Analytics, Database Design and Implementation, Non-Relational Data Management

GLOBAL ACADEMY OF TECHNOLOGY   Bengaluru, India
Bachelor of Computer Science and Engineering   August 2018 – July 2022

SKILLS

Programming Languages / Frameworks: Python, SQL, R, Java, PySpark, Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn
Databases: Snowflake, MySQL, MongoDB, PostgreSQL
Tools: AWS (EC2, S3, EMR, Lambda), Azure (Synapse, Data Factory, Events Hub), Apache Airflow, Tableau, Git, Microsoft Excel

WORK EXPERIENCE

Radiant Info Systems Ltd   Bengaluru, India
Analytics Engineer Intern   August 2021 – February 2022
- Tested a scalable **ELT data pipeline** ingesting real-time passenger data from 900+ city buses running a computer vision-based APC (Automated Passenger Counter) into **Snowflake data warehouse**, using **Apache Kafka** for messaging and **Spark** for transformations, enhancing the experience for 4M+ daily commuters.
- Identified an opportunity to optimize the Snowflake data ingestion process by implementing **micro-batching** in the Kafka Connector for Snowflake configuration, reducing the number of individual database connections and enhancing throughput by 15% during peak hours, enabling handling 8M+ events daily.
- Tweaked 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 such as occupancy rates and route performance metrics to facilitate informed decision-making and drive profitability.

PROJECTS

Global Hotels and Resorts (GHR) Data Analysis | (Python, Pandas, Matplotlib, Excel, Tableau) | Link
- Led a challenging project in a university-wide competition, using data analysis tools and problem-solving skills, to analyze **25 KPI metrics** such as occupancy rate, average daily rate (ADR), booking channels, and others to understand customer demands, drive revenue, and significantly lower operating costs.
- Conceptualized and created customized data visualizations using Python, Excel, and Tableau that led to **data-driven recommendations** to increase efficiency and productivity within the hospitality sector.

DataWorld Insights: Transforming, Loading, and Visualizing Jobs Data | (AWS, PySpark, Airflow, Snowflake, Tableau) | Link
- Orchestrated the **ETL data pipeline** using Airflow for a 10k-row dataset containing jobs and salaries in data science domain, leveraging Apache PySpark and AWS services (EC2, EMR, S3) for efficient data extraction, transformation, and loading.
- Implemented **batch processing** and conducted **Star Schema** data modeling using SQL to optimize data storage and retrieval in Snowflake enhancing efficiency in managing large-scale datasets.
- Built tailored visualizations by loading data from Snowflake into Tableau containing 4-5 key insights for job seekers.

Supermarket Sales Analysis Using Azure Services | (PySpark, Azure, Power BI) | Link
- Engineered an **end-to-end data processing pipeline** to analyze 10k data points from a superstore leveraging Azure services like Data Factory for data ingestion into Data Lake Gen2 storage and transforming data in Azure Databricks using PySpark.
- Analyzed Data in Azure Synapse Analytics using Python to calculate sales and profit, analyzed key factors influencing sales such as discount ranges, shipping methods, and others to identify lucrative product categories across different states.
- Designed an effective dashboard in Power BI to **communicate key findings** to stakeholders to drive more sales and profitability.

Insurance Price Prediction Using Machine Learning | (Python, Pandas, Numpy, Scikit-Learn, AWS) | Link
- Developed an **end-to-end machine learning** application predicting health insurance charges, utilizing key metrics and incorporating 1300+ data points.
- Applied and trained a diverse set of **regression models**, including Random Forest, Decision Tree, and Gradient Boosting, achieving an average accuracy of 87% with the best model on the test data.
- Leveraged AWS services, such as Elastic Beanstalk and CodePipeline, for efficient deployment and **continuous integration (CI)** from GitHub to production.

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