Vinay Kiran Raju

raj.vinay2408@gmail.com | +1 201-492-8306 | LinkedIn | Portfolio | Blog

SUMMARY

Data Analyst & Engineer with 4 years of experience and a Master's in Data Science, delivering 15% cost savings through actionable insights and automated **ETL** pipelines. Skilled in **Python, SQL, AWS,** and BI tools (**Tableau, Power BI**), I transform complex datasets into real-time dashboards, driving operational efficiency and cross-team success.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems	**Jan 2022 - May 2023**
*MS in Data Science	Data Analysis, Visualization & Machine Learning*	***GPA**: 3.71/4*	New York, USA
Gandhi Institute of Technology and Management	**Jun 2017 - Apr 2021**
Bachelor of Technology in Electronics and Communication Engineering	***GPA**: 8.38/10*	Vizag, India

SKILLS

- **Programming Languages**: Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Plotly), SQL, PySpark.
- **Visualization Tools:** Tableau, Power BI, Microsoft Excel
- **Big Data Technologies:** Spark, AWS EMR, Hadoop (HDFS, Hive), HBase, Sqoop, Airflow, Kafka.
- **Databases:** MySQL, MS SQL Server, PostgresSQL, Snowflake, Redshift, Hive.
- **Other Tools& Technologies:** Git, Linux, AWS, AZURE, Data Warehousing, Data Modelling. Agile Methodologies.

EXPERIENCE

Cognizant Technology Solutions   **Apr 2024 - Present**
Data Analyst   New Jersey, USA

- Analyzed 5M+ rows of trade data using SQL and Excel, resolving anomalies to deliver insights that reduced costs by 15% while ensuring data quality for data pipelines.
- Streamlined FX transaction reporting by designing Tableau dashboards with KPIs, reducing processing time by 40% and delivering real-time updates to improve team decision-making.
- Developed three Python-based machine learning models to predict cash requirements for trading, and created Plotly visualizations to reduce overdraft risks by 20% for trading teams.

Natsoft Corporation   **Feb 2023 - Apr 2024**
Data Engineer   New Jersey, USA

- Designed end-to-end ETL workflows with AWS Lambda, S3, Glue, and Redshift, processing 250TB+ datasets and boosting query performance by 40%.
- Optimized PySpark transformations on EMR, reducing null values by 20% and and streamlining Airflow-orchestrated pipelines to improve data reliability for downstream systems.
- Built interactive BI dashboards, leveraging AWS and SQL to deliver real-time insights, reducing reporting latency by 15% for business stakeholders.

                                                                                 Sep 2020 - Jan 2022
Data Engineer   Hyderabad, India

- Established and optimized clusters on Amazon EC2, managing over 120 nodes to deploy EMR for big data environments, resulting in a 35% reduction in processing time across multiple ETL pipelines.
- Integrated 10+ diverse data sources into Amazon S3 and Redshift, refining ingestion workflows to reduce peak-load latency by 20% for internal applications.
- Accelerated large-scale data processing using Spark RDDs, DataFrames, and SQL, optimizing semi-structured data and improving query performance by 10%.
- Orchestrated complex data pipelines using Apache Airflow, implementing dynamic task scheduling to improve workflow efficiency by 40%.

PROJECTS

Bank Loan Default prediction | *Python, Tableau, SQL.*

- Analyzed loan default trends for 35,747 borrowers using Python and visualized insights in Tableau.
- Achieved 96% accuracy in predicting loan defaulters by comparing Random Forest and Decision Tree models

End-to-End Automated ETL Pipeline for Zillow | *Python, PySpark, AWS, Tableau*

- Processed real estate data from Zillow's API using Python and visualized trends in Tableau.
- Built an automated ETL pipeline with PySpark on AWS EMR, stored data in S3, and scheduled workflows using Apache Airflow.

CERTIFICATIONS

Snowflake -BUILD 2023 LLM Bootcamp - Snowflake	**Dec 2023**
Azure Data Fundamentals - Microsoft	**Apr 2023**
Tableau Desktop Specialist- Tableau	**Mar 2023**
Google Data Analytics -Coursera	**Nov 2022**