Vinay Kiran Raju

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

PROFILE SUMMARY

Data Analyst & Engineer with a Master's in Data Science and expertise in capital markets, delivering 15% cost savings through trade data insights and automated ETL pipelines. Specialize in transforming complex datasets into actionable strategies using SQL, Python, Tableau, and AWS. Proven track record of accelerating decision-making with real-time BI dashboards and risk assessment models, enhancing business efficiency and collaborating with cross-functional teams.

WORK EXPERIENCE

Cognizant   **New Jersey, USA**

Data Analyst   *Apr 2024 - Present*

- Analyzed over 5 million rows of trade data using MySQL and Excel, identifying key trends that informed strategic decisions and led to a 15% reduction in operational costs across departments.
- Automated the processing of FX transactions, achieving a 40% decrease in reporting time through integrating SQL and Tableau, while delivering real-time BI reports to over 50 key stakeholders.
- Developed three Python-driven machine learning models for risk assessment on trade data; implemented custom visualizations with Plotly and Matplotlib, enhancing decision-making efficiency by 25% within the trading team.

Natsoft Corporation   **New Jersey, USA**

Data Engineer   *Feb 2023 - Apr 2024*

- Designed and automated ETL workflows using AWS Lambda, S3, Glue, and Redshift, optimizing 250TB+ data processing and improving query performance by 40%.
- Optimized data transformations on user-generated data using PySpark on EMR, reducing null values and anomalies by 30%, while improving workflow efficiency through Airflow orchestration.
- Developed interactive Tableau dashboards, optimizing data pipelines to deliver real-time user insights, improving stakeholder decision-making and reducing reporting time by 30%.

                                                                                                                 Hyderabad, India

Data Engineer   *Sep 2021 - Jan 2022*

- 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.
- Enhanced data ingestion workflows by integrating 10+ diverse sources into Amazon S3 and Redshift, optimizing data processing for internal systems and cutting data latency by 20% during peak loads.
- 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%.

PORTFOLIO PROJECTS

Zillow End-to-End Automated ETL Pipeline   *Jan 2024 - Feb 2024*

- Developed an end-to-end ETL pipeline for Zillow, extracting real estate data from their API, transforming it on an EMR cluster using PySpark, and loading it into S3.
- Extracted Automated the workflow with **Apache Airflow** and integrated **Tableau** for data visualization, showcasing expertise in data transformation and analysis.

Loan Default Prediction   *Jan 2023 - Mar 2023*

- Developed a Python-driven Decision Tree ML model for loan default prediction with 95.21% accuracy and 92.17% AUC on 35,747 entries.
- Optimized Improved data quality by handling missing values, encoding categorical features, and eliminating duplicates; visualized key insights using Tableau, enhancing loan risk assessment.

SKILLS & INTERESTS

Programming: Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Plotly), SQL, PySpark, Scala.

Data Visualization & BI: Tableau, Matplotlib, Seaborn

Big Data: Spark (RDDs, DataFrames), AWS EMR, Hadoop (HDFS, Hive), Sqoop

Cloud: Amazon Web Services, Microsoft Azure

ETL & Automation: Airflow, AWS Glue, Kafka

Databases: MySQL, Snowflake, Redshift, Hive

Tools: Git, Excel (VBA, Macros)

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems	**New York, USA**
*MS in Data Science	Concentration: Data Analysis, Visualization & Machine Learning*	*Date: May 2023*
Gandhi Institute of Technology and Management	**Andhra Pradesh, India**
Bachelor of Technology in Electronics and Communication Engineering	*Date: Apr 2021*

CERTIFICATES

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