

Rakesh Kumar

DATA ENGINEER

LinkedIn: <https://www.linkedin.com/in/rakesh-kumar-275870148/> | rakesh.10april.1998@gmail.com | 7387526086

Summary: Data Engineer with **2+ years of experience** specializing in building scalable data pipelines, developing automation frameworks, and delivering big data solutions to support robust data processing and analysis

SKILL SET

- **Languages.** : Python, SQL
- **Big Data** : Spark (PySpark), Apache Kafka, Apache Airflow, Datawarehouse, Docker, Kubernetes
- **Databases.** : MySQL, MongoDB
- **Cloud** : AWS (Redshift, S3), GCP, Big Query
- **Tools.** : GitHub, CI/CD, Jenkins, Jira
- **Visualization** : Power BI, Tableau, Microsoft Excel, Google Sheet

WORK EXPERIENCE

Infosys Limited

Feb 2022 – May 2023, Chandigarh

Role: Data Engineer

- Developed Python Panda-based validation workflows for 5TB+ of daily claims data, ensuring schema enforcement, duplicate detection, and format validation, reducing data anomalies by 25% and improving data accuracy to 98%.
- Built Python Panda-based jobs to process hundreds of thousands of claim records monthly, applying deduplication, enrichment, and aggregations, optimizing processing efficiency and reducing computation time.
- Designed optimized Big Query data models for 500TB+ of claims data, leveraging partitioning and clustering to enhance query performance by 20% for reporting and analytics.
- Automated ETL pipelines using Apache Airflow, reducing manual intervention by 40%, improving pipeline reliability by 20%, and integrating monitoring and alerting for seamless execution.

Bharat Dynamics Limited (Ministry of Defence)

Jan 2021 – Jan 2022, Hyderabad

Role: Data Trainee (Graduate Apprentice)

- Created 10+ structured reports by querying large datasets using SQL and visualizing insights in Excel dashboards.
- Performed data cleaning and transformation to support analytics on defense systems.
- Collaborated with engineering teams to create data-driven insights, improving manufacturing accuracy and reducing defects.
- Reduced reporting time by 20% through automation and improved data access.

Personal Project

Sales Performance Dashboard (Power BI & SQL)

Feb 2025 – Mar 2025, Lucknow

- Tools Used: SQL, Power BI, Microsoft Excel
- Analyzed 100,000+ sales records and created interactive dashboards.
- Used DAX functions & SQL queries for advanced data calculations.
- Optimized sales prediction models, resulting in an 18% lift in accuracy.

Stakeholder-Insights-and-Performance-Tableau-Dashboard(Tableau & Python)

Mar 2025 – Apr 2025, Lucknow

- Tools Used: Tableau, Python, MS Excel
- Enabled key stakeholders to track KPIs, identify performance trends.
- Conducted exploratory data analysis (EDA) to identify critical business insights, improving strategic planning.
- Improved data consistency and validation processes, achieving a 90% increase in data accuracy.

EDUCATION

Bharati Vidyapeeth college of engineering, Pune, India

2016-2020

Course: B.Tech CGPA: 8.70