

# Jaya Ganesh Kumar

Data Engineer

[jayaganeshkumargudipati@gmail.com](mailto:jayaganeshkumargudipati@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Bengaluru, Karnataka

## PROFESSIONAL SUMMARY

---

Dynamic Data Engineer with a passion for transforming raw data into actionable insights. Proficient in SQL, ETL, and data warehousing techniques, implementing Medallion Architecture (Bronze, Silver, Gold). Expert in utilizing Apache Airflow for orchestrating data pipelines and leveraging PySpark for large-scale data processing. Experienced in developing ADLC-accelerated scripts (AI-Assisted Development Life Cycle) and creating reporting tables to support UI widgets. Skilled in leveraging cloud-based platforms like Redshift, Athena, and MongoDB to deliver scalable and efficient data solutions. Possess strong communication skills and a collaborative mindset, enabling effective teamwork and driving data-driven initiatives. Committed to delivering innovative, scalable, and impactful data solutions.

## PROFESSIONAL EXPERIENCE

---

**Visa** Bengaluru, Karnataka  
*Data Engineer* May 2025 – Present

- Working for Verifi Product on the data warehouse side, designing scalable data solutions.
- Developed 10+ data pipelines using PySpark following the Medallion architecture (Bronze, Silver, Gold) to ensure high data quality.
- Created 5+ reporting tables to support interactive widgets in the user interface.
- Leveraged ADLC (AI-Assisted Development Life Cycle) to accelerate script development, significantly reducing coding time.

**Vivriti Capital** Bengaluru, Karnataka  
*Data Analyst / Data Engineer 1* Jul 2023 – May 2025

- Designed & implemented web scraping solutions to extract data from 5+ websites, ensuring 98% data accuracy and reliability.
- Integrated scraped data into a database, reducing manual effort by 75%.
- Wrote and optimized 30+ SQL transformations to streamline the calculation of 10+ financial metrics, reducing manual intervention and error rates by 80%.
- Designed and implemented scalable data pipelines using Airflow and PySpark to extract, transform, and load data into Redshift for efficient analysis.
- Collaborated with four cross-functional teams to assess data needs and ensure that transformations provided accurate and relevant insights.
- Advanced data engineering role with increased responsibilities, leading complex data pipeline implementations and optimizations.
- Mentored junior team members on data engineering best practices.
- Implemented advanced ETL processes and data modeling techniques.

**PayPal** Remote  
*Data Analyst Intern* Aug 2022 – Jan 2023

- Developed an interactive dashboard using Tableau to monitor losses on day level and included filters for a deeper understanding of loss trends.
- Employed PySpark to optimize data processing, leading to a 40% increase in efficiency and timely data integration with Tableau.
- Worked on a User-experience Analytic project which predicts the number of good users blocked for a bad user.
- Utilized statistical models to forecast risks and guide strategic planning.

## SKILLS

---

- **Languages & Tools:** Python, SQL, Apache Spark, PySpark, Apollo Airflow, Hadoop, AWS, SQL Server, ETL

## EDUCATION

---

**SRM University AP** 2023  
Bachelor of Technology (Inferred)

## PROJECTS

---

### Driver Alertness Prediction

- A computer vision capability to assess driver fatigue in real-time. Leverages OpenCV and Python to process video streams and detect signs of drowsiness.

### Airline Price Prediction

- ML model for predicting flight prices based on historical data.

### **Crypto Tracker**

- Web scraper for realtime CoinMarketCap data tracking.

### **Phishing Detection**

- ML algorithm to classify malicious URLs and protect users.

### **CERTIFICATIONS**

---

- Build a Face Recognition Application using Python
- Python Crash Course
- Machine Learning with Python: Zero to GBMs
- Problem Solving (Basic)