

HEMANTH SAI KUMAR PENTAKOTA

+91 83400 65854 | p.hemanth.sk@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Entry-level Analytics Engineer bridging data modeling and analytics. Builds Medallion-architecture pipelines (Bronze-Silver-Gold) on Spark and Databricks, then exposes them through SQL-driven dashboards - comfortable owning data from raw ingestion through to the metric a stakeholder actually reads.

EDUCATION

B.Tech in Computer Science and Engineering, Lovely Professional University
CGPA: 7.46 / 10

2022 - 2026
Phagwara, Punjab

SKILLS

Data Engineering	Apache Spark, Kafka/Redpanda, Databricks, Delta Lake, ETL
Languages	Python (Pandas, NumPy, PySpark), SQL(MySQL, PostgreSQL)
Visualization & BI	Power BI, Tableau, Streamlit, Excel (Advanced)
Cloud & DevOps	AWS S3, Azure Fundamentals, Docker, Git
AI & Other	Gemini API, GenAI, Machine Learning, DBMS

PROJECT EXPERIENCE

Real-Time Crypto Market Analytics Pipeline with LLM-Generated Insights | *Python · PySpark (Structured Streaming) · Redpanda · Docker · PostgreSQL · Streamlit · Gemini API* | [GitHub](#) Jul 2026

- Engineered a real-time ingestion pipeline streaming live Binance trade data across 5 pairs into **Redpanda**, processed via **PySpark Structured Streaming** with rolling **Z-score anomaly detection**.
- Persisted window aggregates and anomaly flags into **PostgreSQL** via idempotent, at-least-once upserts.
- Built a **Streamlit** dashboard integrating the **Gemini API** for batched market commentary, cutting API calls 5x within free-tier limits.

E-Commerce Sales Analytics Data Pipeline | *Databricks, Apache Spark, Spark SQL* | [GitHub](#) Dec 2025

- Built an end-to-end **Medallion Architecture** (Bronze-Silver-Gold) pipeline processing 100K+ transactions.
- Implemented Silver-layer **data quality checks** and validation to improve reliability for downstream analytics.
- Designed Gold-layer fact & dimension models enabling **₹3.4B revenue analysis** across customers and channels.
- Developed **Databricks SQL dashboards** to evaluate revenue trends and coupon impact.

Customer Shopping Trends Analytics | *Python, Pandas, PostgreSQL, Power BI* | [GitHub](#) Nov 2025

- Performed **EDA on ~4,000 customer records**, uncovering key spending patterns and product category trends.
- Cleaned transactional data with **Pandas** and loaded it into **PostgreSQL** via SQLAlchemy for querying.
- Built a **Power BI dashboard** delivering KPI insights on revenue trends and customer behaviour.

CERTIFICATIONS

SAP Certified - Data Analyst - SAP Analytics Cloud | [SAP](#)
IBM AI Engineering Professional Certificate | [Coursera](#)

Mar 2026
Aug 2025

ACTIVITIES & VOLUNTEERING

- Solved **300+** DSA problems including SQL challenges on LeetCode, HackerRank and DataLemur.
- Active member of **Indian Data Club**; participated in the **Databricks AI Challenge** and earned a certificate of participation.