Kavish Rathod

Mumbai, India | kavishrathod2004@gmail.com | +91 7208684694 | Portfolio | LinkedIn | GitHub

Summary

Data Engineering-focused fresher with hands-on experience building **ETL and data pipelines in Python and PySpark**, including star schema design and workflow orchestration with Airflow. Ranked **Top 10 out of 120+ teams** in an AI/ML Hackathon. Proficient in Pandas, NumPy, SQL(MySQL), Power BI, and Excel for data preprocessing, transformation, EDA, and delivering actionable insights. Exposure to AWS-based data workflows, including storage (S3) and compute (EC2) concepts. Seeking **Data Engineer / Data Analyst** roles. Available within 15 days.

Education

Bachelor of Engineering – Electronics and Telecommunication | CGPA: **8.20**   *Aug 2022 – Present*
Vidyavardhini's College of Engineering and Technology (VCET), Mumbai

Technical Skills

Languages: Python, SQL, R (Basics), MATLAB (Basics)
Libraries / Frameworks: Pandas, NumPy, PySpark (Apache Spark), Scikit-learn, TensorFlow, PyTorch
Tools & Platforms: Power BI, Excel, MySQL, Git/GitHub, Apache Airflow, VS Code, Jupyter Notebook
Concepts: Data Engineering, ETL Pipelines, Star Schema Design, EDA, Machine Learning, Data Visualization, AWS (S3, EC2)

Experience

Intern – National Test House (WR), Mumbai   *Dec 2024 – Jan 2025*

- Processed and structured test data for 12+ materials using **Python and Excel**, improving data consistency and reducing manual errors.
- Performed analysis on test datasets and generated structured reports for engineering insights.

Projects

Telco Customer Churn — ETL Pipeline (Pandas + PySpark)   `[GitHub-ETL]` `[GitHub-PySpark]`

- Built a modular **ETL pipeline (Pandas + MySQL)** on 7,043 telecom records with data cleaning, feature engineering, row-count validation, and a JSON audit log; auto-falls back to CSV if MySQL is unavailable.
- Rebuilt using **PySpark** for scalability — Spark DataFrames, `approxQuantile` median imputation, Parquet and CSV output; generated 4 churn aggregations revealing **42.7% churn** for month-to-month vs. 2.8% for two-year contracts.

Sales Performance Analysis — SQL, EDA, Power BI & Data Engineering   `[GitHub]`

- Analysed Superstore Sales dataset with **18 SQL queries** covering revenue trends, top products, customer segments, and regional performance; designed a **Star Schema** (1 fact + 5 dim tables) and orchestrated the full pipeline via an **Apache Airflow DAG** (DQ checks → MySQL load → reports).
- Implemented **14 automated data quality checks** and built a **Power BI dashboard** (5+ visuals: KPIs, regional map, monthly trends) with full Python EDA across 11 sections in Jupyter.

Exploratory Data Analysis on Zomato Restaurant Data using Python   `[GitHub]`

- Performed end-to-end exploratory data analysis on Zomato Bangalore dataset (7K+ records) using Pandas, Matplotlib, and Seaborn, including data cleaning, outlier detection, and feature extraction.
- Analyzed 1,943+ **North Indian cuisine** listings and identified trends in ratings (avg 3.48), pricing (avg Rs. 536), and customer engagement; observed moderate correlation (0.32) between votes and ratings.

Human Perception-Based Deep Learning Model for Multi-Exposure Image Fusion

- Developed a custom MATLAB-based image fusion algorithm to generate high-quality fused outputs under varying exposure conditions; Currently developing a supervised deep learning pipeline using algorithm-generated outputs to train an LSTM model, addressing variable input challenges in multi-exposure image fusion.

Certifications

- **SQL (Basic & Intermediate) Certificate** – HackerRank | **SQL and Relational Databases 101** – IBM
- **Databricks for Data Engineering** – Databricks

Achievements

- Ranked **Top 10 out of 120+ teams** in an AI/ML Hackathon, building an ML solution under time constraints.
- **2nd place** at **VNPS 2024** (technical project showcase) | **2nd prize** at **Oscillations 2025** (research paper presentation).

Extracurricular Activities

Treasurer – IETE-SF VCET   *Jul 2024 – Present*

- Maintained financial records and managed INR 60,000+ in total event fee collections.
- Coordinated 7+ technical workshops and events on AI, IoT, and emerging technologies.