

Koustubh Singh

Pune, Maharashtra | +91 9309384317 | eharshit.info@gmail.com

[LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

D Y PATIL INTERNATIONAL UNIVERSITY

Bachelor of Technology – Computer Science and Engineering (Data Science)

Sep 2022 – May 2026

Pune, India

SKILLS

Languages: Python, SQL

Libraries and Frameworks: NumPy, Pandas, Matplotlib, dbt

Data Visualization: Power BI

Cloud and Developer Tools: Git, GitHub, Docker, Postman, VS Code, AWS, Azure

Databases: MySQL, SQLite, DuckDB, MongoDB, Snowflake

Coursework: Big Data Analytics, DSA, Operating Systems, DBMS, Computer Networks

EXPERIENCE

LawLevelUp

Data Analyst Intern

May 2025 – Aug 2025

Ghaziabad, Uttar Pradesh

- **Developed** a KPI dashboard in **Power BI**, integrating real-time and predictive analytics to **forecast trends**, **saving more than 5 hours per week** in reporting.
- **Migrated** Excel-based reports to **Power BI**, enhancing **accessibility** and enabling interactive data exploration.
- **Analyzed** learning data from **2K+ enrollments** to identify the most engaging courses and recommend improvements that boosted learner participation.

PROJECTS

End-to-End Vendor Insights

[GitHub](#) | [Live Demo](#)

- Built a **Python and SQL pipeline** that merged multi-source CSV files into a single, query-ready database.
- Designed a **Power BI dashboard** to track sales, margins, pricing issues, and stuck inventory.
- Identified **low-performing vendors** and **pricing gaps**, **improving profit margins**.

Real-Time Insurance Claims Pipeline

[GitHub](#) | [Live Demo](#)

- Simulated insurance data with **Python**, ingesting from **MongoDB** to **Snowflake** via **Airbyte**.
- Transformed data using **dbt** into clean, aggregated, **OLAP-ready datasets**.
- Visualized claims trends and fraud patterns via **Power BI reporting**.

Heart Disease Prediction & Failure Analysis

[GitHub](#) | [Live Demo](#)

- Built an **end-to-end ML system** for heart disease risk prediction using **Python**, **scikit-learn**, and **Streamlit**.
- Trained **Random Forest** and **XGBoost** models achieving **88.6% accuracy** and **0.93 ROC-AUC**, with **SHAP-based explainability**.
- Developed an interactive **Power BI dashboard** to visualize **survival trends**, demographic patterns, and clinical risk factors for decision support.

CERTIFICATIONS

IBM Data Analyst Professional Certificate – Coursera

Apr 2025 – Aug 2025

ACHIEVEMENTS

Best Innovation Award – Smart India Hackathon (Internal round)

Sep 2024

Reached Final Round – Sheryians Reimagine Hackathon

Jun 2024