Harsh Kshirsagar

harshkshirsagar2403@gmail.com | +91-8823038682 | LinkedIn | GitHub

OBJECTIVE

Entry-level Data Analyst with hands-on experience in Python and SQL for ETL, data validation, and dataset preparation. Skilled in BI reporting, data quality checks, and issue investigation, with interest in financial data operations and governance.

SKILLS

Programming:	Python, SQL (joins, aggregations, filtering, validation queries)
Data Engineering / ETL:	Data ingestion, transformation, dataset validation, analytics-ready dataset creation
Databases:	MySQL, SQLite (tables, primary/foreign keys, relational concepts)
Business Intelligence:	Power BI (data modeling, dashboards, reports), Tableau
Data Processing:	pandas, NumPy, data cleaning, schema checks, feature engineering
Excel:	Pivot tables, lookup formulas, dataset checks, reporting support
Automation:	Python scripting, reusable preprocessing pipelines
Tools:	Git, GitHub, Jupyter Notebook, Google Colab, Excel
Domain Exposure:	Basic understanding of financial instruments, indices, and benchmark datasets (self-study)

EDUCATION

Samrat Ashok Technological Institute, Vidisha *Nov 2021 – June 2025*
Bachelor of Technology in Artificial Intelligence & Data Science

EXPERIENCE

Software Developer Trainee | Revature Consultancy Services Pvt. Ltd *Oct 2025 – Present*

- Exposure to full software development lifecycle (SDLC) and Agile-based execution in team environments.
- Collaborated with peers and mentors to understand requirement translation, iterative development, and code.
- Gained understanding of how engineering teams deliver scalable systems aligned with functional requirements.
- Developed strong foundations in clean code practices, system thinking, and cross-functional collaboration.
- Practiced debugging and analysing outputs to verify correctness of implemented features.

Artificial Intelligence Builder Intern | FITT, IIT Delhi *Jan 2024 – Mar 2024*

- Built Python ETL workflows for ingesting, cleaning, transforming, and validating structured datasets.
- Implemented data quality checks, schema validation rules, and consistency verification steps.
- Investigated missing and duplicate records and performed basic root-cause analysis.
- Generated analytics-ready tables to support dashboards, reporting, and downstream analysis workflows.

PROJECTS

Data-Driven Seamless Indoor Navigation Analysis — GitHub Repository

- Developed ETL pipelines to ingest and process 5k+ GeoJSON records into structured reporting datasets.
- Applied data cleaning, validation checks, and aggregation logic to ensure dataset reliability and consistency.
- Stored processed data in relational format for querying, analysis, and reporting use cases.
- Produced analytical tables and dashboards to support business-style reporting and insights.
- Built interactive Streamlit dashboards to visualize usage trends and traffic metrics.

Dementia Risk Detection App — GitHub Repository

- Processed and validated healthcare datasets containing 1k+ records for analysis and prediction workflows.
- Designed automated data preparation pipelines including cleaning, encoding, and validation of datasets.
- Implemented repeatable workflows to generate consistent analysis-ready datasets.
- Performed dataset integrity checks and structured outputs for reporting and application consumption.
- Developed an MVP application exposing processed data outputs and predictions to end users.

CERTIFICATIONS

- Prompt Design in Vertex AI Skill Badge – Google
- Learning Data Analytics: Foundations – LinkedIn
- Getting Started with Python – Coursera (University of Michigan)