K Sai Datta

Phone: +91 9182261303   Email: ksaidatta.116@gmail.com   LinkedIn: saidatta   GitHub: datta116

Portfolio: **datta.com**

Professional Summary

Entry-level Data Analyst skilled in Python, SQL, Power BI, and Excel with strong awareness of data governance, accuracy checks, and compliance-focused workflows. Experienced in EDA, KPI reporting, ETL pipelines, and documentation. Highly interested in real-world evidence (RWE), commercial analytics, patient-level insights, and data quality requirements within pharmaceutical and healthcare environments.

Education

- **B.Tech in Computer Science Engineering**: GPA: 7.32   (2021–2025)

Skills

- **Programming**: Python, SQL
- **Analytics Tools**: Power BI, Tableau, Excel (Advanced), SQL Server
- **Analytics Competencies**: EDA, Data Cleaning, Statistical Analysis, Data Quality Checks, ETL Pipelines, KPI Reporting
- **Libraries**: Pandas, NumPy, Matplotlib, Scikit-learn
- **Professional Skills**: Data Accuracy, Documentation, Reporting, Problem Solving

Projects

- **Commercial Insights Dashboard (Excel & Power BI)**: Developed an interactive dashboard analyzing 2023 workforce and compensation trends using pivot tables, slicers, and KPI indicators. Identified regional patterns, role-based variations, and outlier signals, enabling standardized reporting and faster insight extraction. GitHub
- **SQL Data Analytics on Retail Dataset**: Conducted detailed segmentation, revenue modelling, cohort performance, and retention analysis using CTEs and window functions. Built KPI reports for category-level trends and customer behaviour, improving clarity and enabling consistent data-driven decision making. GitHub
- **Data Warehouse (Medallion Architecture)**: Designed a SQL Server data warehouse integrating ERP and CRM datasets using Bronze–Silver–Gold layering. Built ETL workflows, standardized data cleaning rules, and developed a star schema to support BI dashboards, improving query performance and reporting consistency. GitHub
- **Diamond Price Prediction Model**: Created a Random Forest model trained on 53k+ records with engineered features (transformations, encodings) to enhance predictive accuracy. Converted USD pricing metrics to INR and built a clear workflow for data cleaning, model training, and performance evaluation. GitHub

View More Projects on GitHub

Certificates

- **Data Analysis with Python**: NPTEL (2023–2024)
- **B2 English**: Cambridge University (2021–2022)

Languages

- English, Hindi, Telugu

Interests

- Healthcare Analytics, Data Storytelling, Writing, Film Editing