

K Sai Datta

+91 9182261303 | ksaidatta.116@gmail.com | linkedin.com/in/saidatta | github.com/datta116

Portfolio: datta116.github.io

Professional Summary

Results-driven Data Analyst with hands-on experience in end-to-end data analytics projects, including web-scraped datasets, ETL pipelines, predictive modeling, and stakeholder-facing dashboards using SQL Server, Power BI, Tableau, and Python. Expertise in data quality, governance, feature engineering, and building interactive visualizations. Passionate about real-world evidence (RWE), patient-level analytics, and commercial forecasting in pharmaceutical and healthcare domains.

Experience

Telugu Film Industry (TFI) Box Office Performance Analysis: Independently built a production-grade analytics solution: web-scraped and aggregated box office data (worldwide gross, Day-1 collections, verdicts) for films of top 6 heroes of TFI from sources like Sacnilk, Wikipedia, and Andhra Box Office; cleaned, modeled, and validated data using Excel and MS SQL; developed an interactive Power BI dashboard with dynamic filters, trend analysis. *Tech:* Python (Web Scraping - Requests/BeautifulSoup), Excel, MS SQL, Power BI (DAX, Visualizations)
GitHub

Road Accident Analysis: Led end-to-end Power BI solution on large-scale accident data: cleansed and engineered features in MSSQL Excel, designed interactive dashboards with drill-throughs, dynamic KPIs, and slicers — reduced reporting time by 50% and drove data-informed decisions.

Tech: Power BI (DAX, Power Query), MSSQL, Excel Advanced
GitHub

Projects

E-Commerce Sales Analysis (Tableau): Developed fully interactive executive dashboard tracking sales KPIs, YoY growth, regional performance, product-wise profitability, and top/bottom performers using LOD expressions, parameters, action filters, and drill-downs for business stakeholders.
GitHub

Data Warehouse ETL Developer: Designed and deployed a compliant-ready data warehouse using SQL Server T-SQL with Bronze-Silver-Gold medallion architecture. Built automated ETL pipelines, enforced rigorous data quality validation rules, created star schema, and accelerated query performance by 60%+. Enabled self-service KPI reporting for stakeholders.
Tech: T-SQL, SQL Server, ETL, Data Modeling, Data Governance
GitHub

Customer Behaviour Churn Prediction (Python): Conducted customer segmentation using RFM analysis and built predictive churn model (85%+ accuracy) with Pandas and Scikit-learn; uncovered key retention drivers and delivered actionable business recommendations.

Tech: Python, Statistical Modeling, Segmentation
GitHub

[More on GitHub →](#)

Skills

Programming: Python (Pandas, NumPy, Scikit-learn), SQL / T-SQL, DAX

BI Visualization: Tableau (LODs, Parameters, Dashboards), Power BI (DAX, Power Query), Excel (Advanced)

Machine Learning: Random Forest, Regression, Feature Engineering

Data Engineering: ETL, Data Warehousing, Medallion Architecture, Star Schema

Data Management: Data Quality, Validation, Governance

Domain: Real-World Evidence (RWE), Patient-Level Analytics, Commercial Forecasting

Tools: SQL Server (SSMS), Git, Jupyter, VS Code

Education

B.Tech – Computer Science Engineering: 2021 – 2025

GPA: 7.32

Certifications

Data Analysis with Python – NPTEL (2024):

Cambridge English B2 – First Certificate (2022)