Project Proposal

Title:

Superstore Sales Dataset

Data Cleaning, Analysis, and Visualization Using Python, SQL, and Power BI

Introduction:

In today's data-driven world, raw data collected from different sources is usually incomplete, inconsistent, or unstructured. This makes it difficult for organizations to make informed decisions. Our project aims to solve this problem by taking a real dataset, cleaning it, analyzing it, and finally visualizing it in a way that helps extract meaningful insights.

Objectives:

- To clean and preprocess raw data using Python (Pandas, NumPy).
- To store the cleaned data in SQL database for easy management and accessibility.
- To perform queries in SQL to verify the correctness and usability of data.
- To create interactive and easy-to-understand dashboards using Power BI.
- To provide meaningful insights for decision-making.

Tools & Technologies:

- Python → For data cleaning & preprocessing.
- SQL → For storing and managing the cleaned data.
- Power BI → For visualization and dashboard creation.
- Dataset → A real-world dataset (from Kaggle or other open-source platform).

Scope of the Project:

- Import dataset (CSV/Excel).
- Clean the dataset by handling missing values, duplicates, and formatting issues.
- Store the cleaned dataset into an SQL database.
- Run SQL queries to analyze the data.
- Visualize important KPIs and insights in Power BI.

Expected Outcome:

- A fully cleaned dataset stored in SQL.
- Queries that demonstrate data correctness and usability.
- Interactive Power BI dashboards that showcase insights and trends.
- A clear understanding of how raw data is transformed into valuable information.

Conclusion:

This project will demonstrate the complete journey of data: from raw form → to cleaned data → to stored database → to visual insights. It will help in understanding how real industries work with data for better decision-making.