**Project: E-Commerce Sales & Customer Analytics Project**

****Date:**** October 16, 2025

## 1. Project Overview & Goal

This project simulates a real-world data engineering and analytics scenario. It involves processing a raw, messy e-commerce dataset through a complete data pipeline. The primary goal is to clean, transform, and structure the data to make it suitable for accurate business analysis, reporting, and insights generation.

The project will demonstrate a full-stack data workflow, starting from raw data ingestion, moving through cleaning and validation with SQL and Python, and culminating in an interactive business intelligence dashboard in Power BI.

## 2. Key Objectives

***Data Cleaning & Transformation:**** Identify and rectify common data quality issues in a raw e-commerce dataset.
***Establish Data Pipeline:**** Implement a structured workflow using MySQL, Python, and Excel for data processing.
***Reporting & Visualization:**** Develop an interactive dashboard in Power BI to report on key business metrics.
***Documentation:**** Document the entire process, from initial data problems to final insights.

## 3. Scope & Methodology

### 3.1. Dataset
The project will use a synthetic raw e-commerce dataset containing 150+ orders. This dataset is designed with intentional errors to simulate real-world challenges, including:
- Null / Missing Values
- Duplicate Order IDs
- Invalid Date Formats
- Textual Inconsistencies (e.g., "UPI", "upi")
- Incorrect Data Types (e.g., "five" instead of 5)
- Invalid Email Addresses

### 3.2. Tools & Technologies
***Database:**** MySQL
***Data Cleaning & Validation:**** SQL, Python (Pandas)
***Reporting & Visualization:**** Microsoft Excel, Power BI

### 3.3. Workflow

1. **Load Raw Data:** Import the raw `.csv` or `.xlsx` data into a staging table in MySQL.
2. **Data Profiling:** Run initial queries to identify the scope and nature of data quality issues.
3. **Data Cleaning (SQL):** Execute SQL scripts to handle duplicates, standardize categorical data (e.g., payment methods), fill missing values, and format dates and numbers correctly.
4. **Data Transformation:** Create a final, clean table in MySQL with appropriate data types, primary keys, and constraints (e.g., `NOT NULL`).
5. **Data Validation (Python):** Use Python and the Pandas library to perform CRUD operations (e.g., updating prices, inserting new records) and to cross-validate the integrity of the cleaned data.
6. **Export & Report:** Export the final, clean dataset to Excel and build an interactive Power BI dashboard to visualize Key Performance Indicators (KPIs) such as:
   * Sales by Region
   * Top Customers
   * Monthly Revenue Trends

## 4. Expected Outcomes & Deliverables

1. **SQL Scripts:** A complete set of SQL scripts for data cleaning, transformation, and loading.
2. **Python Scripts (Optional):** Scripts for any data validation or manipulation tasks.
3. **Cleaned Dataset:** The final, structured dataset exported to an Excel file.
4. **Power BI Dashboard:** A fully interactive dashboard for business analysis.
5. **Final Documentation:** A summary report documenting the workflow, challenges, solutions, and key insights discovered.

## 5. Business Relevance & Impact

This project is designed to mirror the demands of modern Data Analytics, Business Intelligence, and Data Engineering roles. It demonstrates the critical ability to handle imperfect, real-world data and transform it into a reliable source for decision-making. By building an end-to-end pipeline, this project showcases a comprehensive understanding of the entire data lifecycle, from raw ingestion to actionable visualization, a skill set highly valued across all industries.