End-to-End Data Analytics Project: Customer Sales Analysis

Project Context

This project is a personal, industry-style simulation aimed at understanding the end-to-end workflow of a real-world data analytics project. The goal was to experience how raw datasets are processed, analyzed, and visualized to generate actionable business insights, simulating an actual organizational scenario.

Dataset Overview

- Dataset Source: Kaggle
- Type: Customer Sales Data
- Format: CSV file
- Purpose: To analyze customer behavior, sales patterns, and business performance trends.

Tools & Technologies Used

- Excel: Initial inspection of CSV data
- Python (Jupyter Notebook, Pandas): Data cleaning, preprocessing, feature engineering
- MySQL: Database creation, storage, and SQL-based analysis
- Power BI: Interactive dashboard creation and visualization

Data Preprocessing in Python

- Loaded CSV data into Jupyter Notebook using Pandas
- Explored dataset using df.info() and df.describe()
- Checked for missing values using df.isnull().sum()
- Filled null values in review_rating using median per category
- Standardized column names (lowercase, underscores)
- Created age_group and frequency category columns
- Dropped duplicate column promo_code_used
- Final cleaned dataset ready for database

Database Creation & SQL Analysis

- Installed MySQL and connected to Jupyter Notebook
- Created database 'data_analysis' and table 'sales_data'
- Loaded cleaned data to MySQL
- Conducted analysis using GROUP BY, ORDER BY, WHERE, LIMIT, LIKE
- Extracted totals, averages, trends, category-based comparisons

Data Visualization in Power BI

- Connected MySQL to Power BI
- Built interactive dashboard with multi-card visuals, slicers, and charts
- Designed to support data-driven decisions

Challenges Faced

- Handling null values in significant attributes
- Removing duplicate columns
- Mapping numerical data to categories
- Integrating data across tools
- Designing interactive, insightful dashboard

Key Learnings

- Importance of preprocessing
- Using multiple tools effectively in a pipeline
- SQL aggregation and filtering
- Clear communication through visualization
- Planning and iterative refinement in end-to-end projects

Conclusion

Project provided hands-on experience in the complete data analytics workflow, strengthening skills in data cleaning, feature engineering, database management, SQL querying, and visualization, simulating real-world analytics projects.