Brief Overview of Solution

Introduction

The WiseAnalytics Data Engineering project focuses on processing the Online Retail Dataset to enable customer analytics. The solution consists of three main components:

1. **ETL Pipeline**: Extracts, transforms, and loads (ETL) retail data.
2. **Database & Data Modeling**: Structures and stores cleaned data in MySQL.
3. **API Development**: Provides an interface for accessing analytics using FastAPI.

Technologies Used

- **Jupyter Notebook**: Used for developing and executing the ETL pipeline.
- **MySQL**: Stores processed data in a structured format.
- **FastAPI**: Provides API endpoints for querying customer and product insights.

Solution Architecture

1. ETL Pipeline

The ETL pipeline processes the Online Retail dataset as follows:

- **Extract**: Reads raw data from CSV.
- **Transform**:
    - Handles missing values and duplicates.
    - Ensures correct data types.
    - Implements business rules (e.g., positive quantities).
- **Load**: Stores cleaned data in MySQL tables.

2. Database & Data Modeling

- The cleaned data is stored in MySQL using structured tables:
    - transactions: Stores sales transactions.
    - customers: Aggregates customer-level purchases.
    - products: Summarizes product-level sales.
- Proper indexing ensures optimized query performance.

3. API Development

- Developed using FastAPI to enable data access via REST endpoints.
- Provides:
    - Customer purchase summary.
    - Product sales insights.

- Implements basic error handling and input validation.

Key Features

- **Data Quality Checks**: Ensures completeness, correctness, and consistency.
- **Scalability**: Structured for future enhancements.
- **Ease of Use**: Simple setup and execution.

Conclusion

This solution successfully processes, structures, and exposes retail data, enabling insightful customer analytics through a well-integrated ETL pipeline, MySQL database, and FastAPI interface.