Analytics Engineering Assignment: Task 1

Data Modeling & Transformation Strategy

1. Overview & Architecture

To address the need for a structured, flexible, and analytical dataset, I have implemented a data transformation pipeline using **dbt (Data Build Tool)**.

The project follows the **Medallion Architecture**, organizing data into four distinct layers to ensure data quality, traceability, and ease of analysis. This approach allows us to transform raw JSON logs into a clean, dimensional model suitable for business intelligence.

The Layers:

- **Bronze (Landing):** Raw data ingestion exactly as received from the source CSVs.
- **Silver (Standardized):** Cleaned data with parsed JSON payloads, standardized data types, and deduplication logic.
- **Gold (Enriched):** A Star Schema dimensional model. This includes **SCD Type 2** dimension tables (tracking item changes over time) and Fact tables for transactions.
- **Platinum (Marts):** Pre-joined, pre-aggregated "One Big Table" (OBT) models designed specifically to answer the business questions provided in the assignment.

2. Project Codebase & Documentation

I have emphasized transparency in my thought process and query logic. For a complete understanding of the data modeling decisions, DAG (Directed Acyclic Graph) lineage, and specific design choices, **please refer to the GitHub repository below:**

- **GitHub Repository:** https://github.com/BieNewBie/Wolt

⚠️ **Important:** Please read the **README.md** file located in the root of the repository. It contains:

- A detailed **Architecture Diagram**.
- **Setup instructions** for reproducing the environment.
- **Data Dictionary** and descriptions of the available Marts.
- **Assumptions** made regarding VAT calculations and delivery fees.

Note: In the event you cannot access GitHub, I have also uploaded the full dbt code to this drive in the folder named **"retail_data_model"**.

3. Data Deliverables

The final output datasets are available in the **"Datasets"** folder included in this submission. To make them easy to navigate, I have organized the folder structure to mirror the dbt project layers:

- **`marts/`**: Contains the final datasets required for the Analytics Team. These are the tables you should use to answer the business questions (e.g., `mart_sales_summary`, `mart_customer_summary`)
- **`dimensions/`** **&** **`facts/`**: Contains the intermediate Gold layer tables for dimensional modelling

4. Readiness for Analysis

The data model has been successfully built and tested. The `marts` layer specifically enables the Analytics Team to answer all 8 key business questions (such as store serving area, courier costs, and revenue trends) using simple `SELECT *` queries without requiring further complex joins

.