**Marketing Analytics Data Infrastructure Project Report: Olist Dataset**

**Overview**

This project involves the design and implementation of a comprehensive marketing analytics database using the Olist e-commerce dataset. The goal of this initiative is to create a scalable and relational data structure that supports multiple marketing-related analytical projects such as funnel analysis, customer segmentation, return on ad spend (ROAS), customer lifetime value (CLV), campaign performance, and regional sales trends.

This report documents the structure of the database, the rationale behind its design choices, and the tools used to manage the infrastructure, making it suitable for use in real-world marketing analytics workflows.

**Database Technology: PostgreSQL**

We selected **PostgreSQL** as the core database management system for this project for the following reasons:

* **Reliability**: PostgreSQL is a proven, robust, and open-source relational database known for its data integrity and ACID compliance, which are essential in analytical scenarios.
* **Rich SQL Support**: Its full support of SQL standards, window functions, and CTEs makes it ideal for complex queries used in marketing reports.
* **Extensibility**: PostgreSQL allows easy integration with advanced analytics libraries through extensions and Python connectors, enabling deeper analytical use cases.
* **Community and Ecosystem**: A large community and plenty of documentation make PostgreSQL a safe, long-term choice.

**Infrastructure and Execution: Docker**

To ensure consistency and portability, we use **Docker** to run PostgreSQL. Docker allows us to create isolated containers that include all dependencies and configurations necessary to run our database. This approach ensures:

* **Reproducibility**: The entire database environment can be easily recreated across machines or team members.
* **Isolation**: PostgreSQL runs in its own container, avoiding conflicts with other applications or local installations.
* **Deployment Speed**: With Docker, launching the database is quick and does not require manual setup.

**Database Schema Diagram Summary**

The diagram illustrates the structured design of the Olist PostgreSQL database, capturing the key relationships within an e-commerce environment. Orders are linked to customers, payments, products, reviews, and sellers. Each customer and seller is associated with a geographic location via zip code. Products are categorized and connected to their translations for ease of analysis.

****

**Schema Summary**

We analyzed the structure of the Olist datasets and translated them into a normalized relational schema that captures relationships such as:

* Orders related to Customers, Payments, and Reviews
* Products linked to Sellers and Categories
* Location information for customers and sellers for geographic marketing analysis

Key design features:

* Each table has a clearly defined **primary key** for integrity.
* **Foreign keys** enforce relational integrity across dimensions.
* Some tables use **synthetic keys** (e.g., combining order\_id + item\_id) to uniquely identify records where no natural key exists.
* Key fields like customer\_unique\_id and review\_id are marked as **UNIQUE** where applicable.
* Nullable fields such as order\_delivered\_date allow flexibility for orders still in progress.
* Indexes are planned for high-traffic fields such as order\_id, customer\_id, and product\_id to optimize performance in analytical queries.

**Intended Use**

This database is built to support multiple marketing-focused analytical projects. Each project will connect to the core database and use queries to extract insights. Example projects include:

* Analyzing customer behavior across order funnels
* Computing average acquisition costs and return on marketing investments
* Forecasting demand in specific cities or regions
* Tracking product category performance over time

This foundational work ensures that all future marketing analytics work is built on a clean, logical, and performance-optimized data structure.

**Conclusion**

This infrastructure project represents the foundation for a full suite of marketing analytics use cases. By choosing PostgreSQL and Docker, and carefully modeling the Olist dataset, we ensure the database is both accessible and scalable. This system is now ready to support a portfolio of data-driven marketing analyses that simulate real-world professional demands.