**Pizza Data Analytics**

**Introduction:**

This project is dedicated to crafting a sophisticated end-to-end data pipeline, harnessing cutting-edge technologies for batch processing. The primary objective is the analysis of sales data from a pizza outlet, drawing insights from both local storage and Google Cloud Storage. Our chosen ETL tool, MAGE, renowned for its advanced capabilities, orchestrates seamless data processing. The robust foundation for our analytics lies in Google BigQuery, a powerful and scalable data warehouse. To present comprehensive insights, we utilize Tableau for visualization and dashboard creation. This amalgamation of technologies ensures a professional, efficient, and insightful data journey from inception to visualization.

**Dataset:**

Our pizza sales dataset comprises 12 relevant features, providing a granular view of each transaction:

* **order\_id:** Unique identifier for each order.
* **order\_details\_id:** Unique identifier for each pizza within an order.
* **pizza\_id:** Unique key tying the pizza to its details.
* **quantity:** Quantity ordered for each pizza type and size.
* **order\_date:** Date the order was placed.
* **order\_time:** Time the order was placed.
* **unit\_price:** Price of the pizza in USD.
* **total\_price:** Calculated as unit\_price \* quantity.
* **pizza\_size:** Size of the pizza (Small, Medium, Large).
* **pizza\_type:** Unique identifier tying the pizza to its details.
* **pizza\_ingredients:** Ingredients used in the pizza.
* **pizza\_name:** Name of the pizza as shown in the menu.

The dataset is publicly available on the Maven Analytics website, and you can access it [here](https://www.mavenanalytics.io/blog/maven-pizza-challenge).

**Technologies Used:**

* Programming Languages: Python, SQL
* Storage: Google Cloud Storage, Local storage
* Virtual Machine (VM): Compute Engine (GCP)
* ETL: MAGE
* Data Warehouse: Google BigQuery
* Visualization: Tableau

**Pipeline Architecture:**

A diagram of a website

Description automatically generated**Data Model:**

**A graph paper with a diagram

Description automatically generated with medium confidenceDashboard:**

**A graph of pizza sales

Description automatically generated with medium confidence**

**Conclusion:**

This comprehensive integration of technologies not only ensures a robust analysis of pizza sales data but also facilitates scalability and efficiency throughout the entire data processing pipeline.