**Challenge 4**:

* What other insights do you think would add value to the business that can be extracted using at least one of these tables?

Pick up to three, and explain why they might be useful and how we can get them. (If more tables are needed, list them.)

-- 1. Revenue Trends Over Time

SELECT created\_date, SUM(o.quantity \* p.price) AS total\_revenue

FROM orders o

JOIN products p ON o.product\_id = p.id

GROUP BY created\_date

ORDER BY created\_date;

--2. Price Sensitivity by Category

SELECT p.category,

ROUND(CORR(p.price, o.quantity), 2) AS price\_quantity\_correlation

FROM orders o

JOIN products p ON o.product\_id = p.id

GROUP BY p.category;

* What ETL/ELT tool would you use to extract this data and insert it into BigQuery? Explain the steps of creating this type of pipeline.

To build a robust data pipeline using DBT from PostgreSQL to BigQuery, the process begins

with extracting and loading the raw data. This can be accomplished using tools such as Airbyte,

Fivetran, or a custom Airflow job, which move data from the PostgreSQL source into

BigQuery's raw datasets.

Once the data is in BigQuery, DBT (Data Build Tool) takes over the transformation process.

Using SQL-based models, DBT allows you to clean, join, and aggregate this raw data into

structured, analytics-ready tables. These transformations are modular, version-controlled, and

easily testable.

Scheduling and orchestration of DBT jobs can be managed through platforms like Airflow, dbt

Cloud, or Google Cloud Composer, allowing the pipeline to run on a consistent schedule or

trigger.

Once the structured data is in place, an AI-based pipeline can be layered on top to unlock

predictive capabilities.

This begins with generating features—essentially curated datasets derived from the transformed

data, often built using DBT and stored in BigQuery. These feature sets are then used to train

machine learning models that could, for example, forecast future demand or provide product

recommendations.

The training process can be executed using platforms such as Vertex AI, SageMaker, or a

custom environment, depending on the team's preference and infrastructure. After training, the

models can be deployed for batch or real-time predictions, with results written back into

BigQuery for consumption by dashboards, reporting tools, or other business applications.

* What AI-based pipeline could you add to this pipeline? Describe it.

This integration of DBT and AI pipelines turns raw transactional data into both actionable analytics and forward-looking predictions, helping teams make better decisions on operations, marketing, inventory, and customer engagement.