

A2: Web Performance - Team Assignment

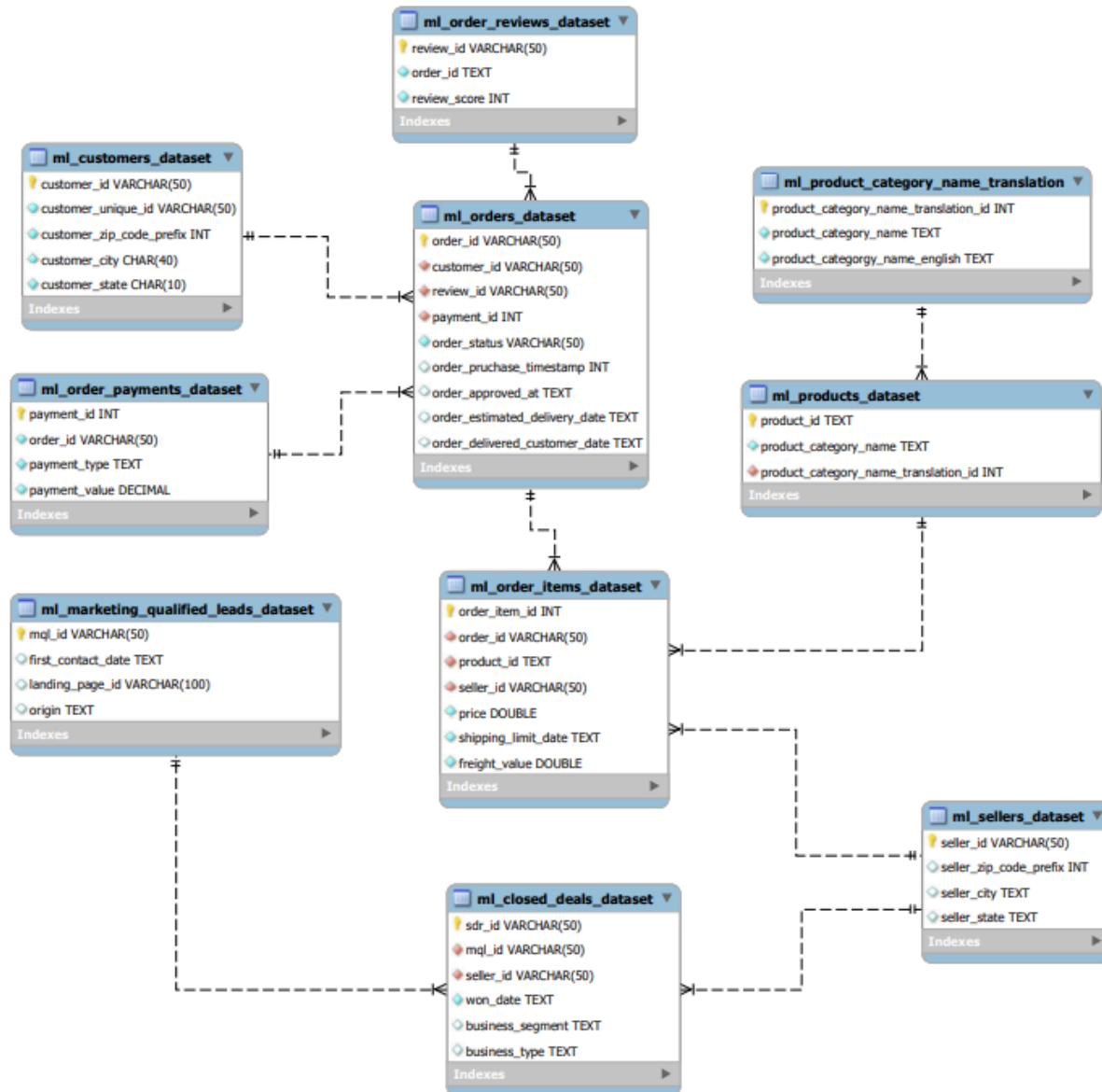
Group 09

- Rudrani Sakhare
- Ania de Chavigny
- Ahmed Raza
- Hakeem Garcia
- Janati

Video Link: <https://youtu.be/xvIZh2UYQ2w>

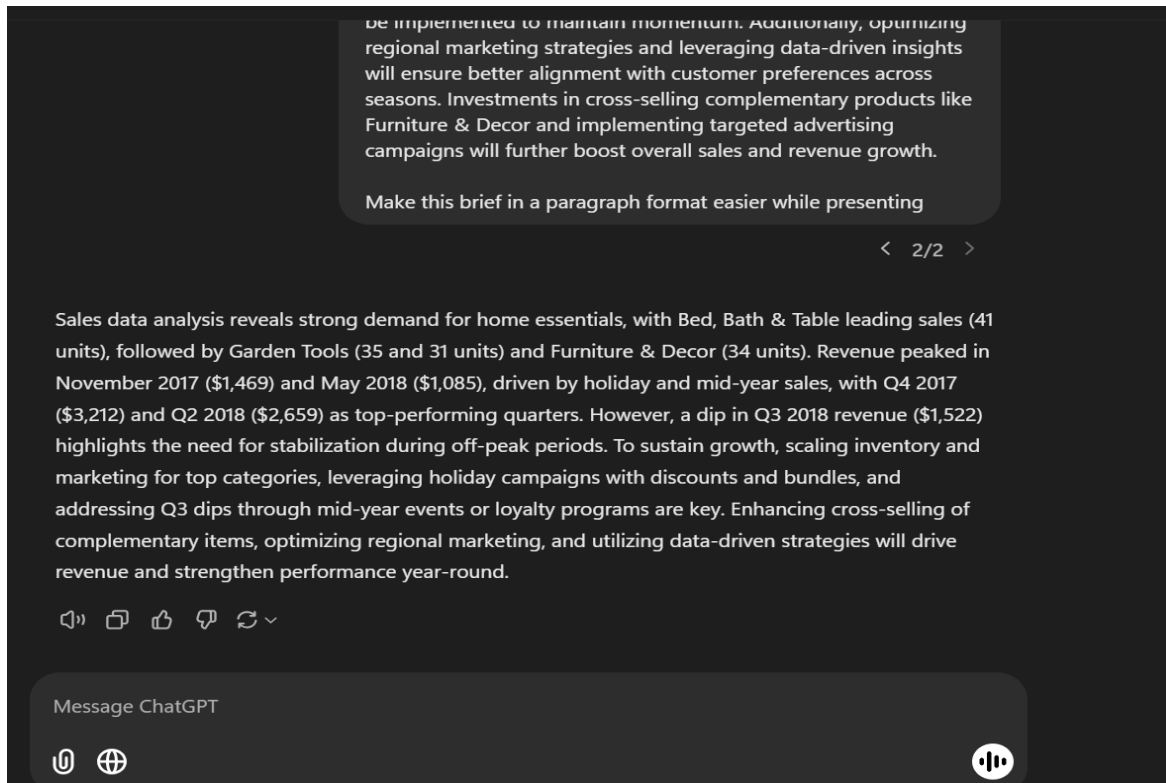
Appendix:

Entity Relationship Diagram



References:

- OpenAI. (2024). *ChatGPT*. <https://chatgpt.com>
We utilized ChatGPT to assist in creating SQL queries for data analysis, ensuring accurate syntax and optimization.



create a summary and conclusion script for this information: The analysis of revenue distribution across states highlights significant disparities in performance: High-Performing States: SP (São Paulo) is the top-performing state, contributing the highest revenue at \$700.56 USD, accounting for a significant share of the total revenue. RJ (Rio de Janeiro) follows with a notable revenue contribution of \$277.21 USD, indicating a strong customer base. Moderate-Performing States: States such as DF (Distrito Federal), PE (Pernambuco), and MS (Mato Grosso do Sul) show moderate revenue generation, with contributions ranging from \$80 to \$96 USD.

ChatGPT 4o

Share

Introduction

"Good day, everyone. Today, I will provide an analysis of the revenue distribution across states, highlighting significant disparities in performance, key insights, and actionable recommendations to drive sustainable growth."

Summary

1. High-Performing States:

- "The state of **São Paulo (SP)** leads as the top-performing region, contributing the highest revenue of **\$700.56 USD**, accounting for a significant share of the company's total revenue."
- "Following closely is **Rio de Janeiro (RJ)**, with a notable revenue

Message ChatGPT



ChatGPT 4o

Based on section 1 which is the sales performance I need you to create queries in regards to question 3 and 4. Please analyse the dataset and create the queries. also please verify the erd diagram



To address your questions regarding Section 1 (Sales Performance) and specifically for questions 3 and 4, I will:

1. Analyze the ERD to ensure the structure aligns with your dataset and queries.
2. Create SQL queries for questions 3 and 4.
3. Verify the ERD diagram for correctness based on the provided structure.

sql

Copy code

```
SELECT
    c.customer_state,
    COUNT(DISTINCT o.order_id) AS total_orders, -- Total distinct orders per state
    SUM(oi.price * 0.25) AS total_revenue_usd -- Convert BRL to USD for revenue
FROM
    ml_customers_dataset c
JOIN
    ml_orders_dataset o
ON
    c.customer_id = o.customer_id
JOIN
    ml_order_items_dataset oi
ON
    o.order_id = oi.order_id
GROUP BY
```

ChatGPT 4o 

```
ORDER BY
total_revenue_usd DESC; -- Highlight states with the highest revenue
```

 Copy code

What This Query Does

1. Joins the Tables:

- `ml_customers_dataset` provides the `customer_state`.
- `ml_orders_dataset` links orders to customers via `customer_id`.
- `ml_order_items_dataset` provides the revenue (via `price`) for each order.



Hi, You're a SQL Data Expert and I want you to answer these queries for my job. Please share the write syntax for MySQL and don't make mistakes.

I want you to answer the following questions.

What are the most frequently used payment methods?

You can get the below information from the following table (sharing the column name as well for each table)

Table: ml_order_payments_dataset
order_id
payment_sequential
payment_type
payment_installments
payment_value



To determine the most frequently used payment methods in MySQL, you can use the following query. This query counts the occurrences of each `payment_type` and orders the results in descending order of frequency.

SQL Query:

```
sql                                                                    Copy code

SELECT
    payment_type,
    COUNT(payment_type) AS frequency
FROM
    ml_order_payments_dataset
GROUP BY
    payment_type
ORDER BY
    frequency DESC;
```

Explanation:

Unable to share conversation

Sharing conversations with user uploaded images is not yet supported



Selects the `payment_type` column and counts its occurrences.

Message ChatGPT



ChatGPT can make mistakes. Check important info.

```
correct code
SELECT DISTINCT
  (order_status)
FROM
  mban_a2_fall2024.ml_orders_dataset
WHERE
  order_status = 'unavailable'
  AND order_status = 'canceled';
```

Your SQL query has a few issues that need correction:

1. The parentheses around `(order_status)` are unnecessary.
2. There is a typo in `oder_status` which should be `order_status`.
3. The `WHERE` clause checks two mutually exclusive conditions (`order_status = 'unavailable'` and `order_status = 'canceled'`), making the query logically invalid. No single record can meet both conditions simultaneously.

Here's the corrected code:

```
sql Copy code

SELECT DISTINCT
  order_status
FROM
  mban_a2_fall2024.ml_orders_dataset
WHERE
  order_status IN ('unavailable', 'canceled');
```

Explanation:

- `DISTINCT`: Ensures that duplicate `order_status` values are removed from the result.
- `IN ('unavailable', 'canceled')`: Checks if `order_status` is either 'unavailable' or 'canceled', resolving the logical error.

This query retrieves all distinct order statuses that are either 'unavailable' or 'canceled'.

I want code to eliminate order_status that are unavailable and canceled

To eliminate `order_status` values that are either 'unavailable' or 'canceled', you can use the `NOT IN` operator in the `WHERE` clause. Here's the corrected query:

```
sql Copy code

SELECT DISTINCT
  order_status
FROM
  mban_a2_fall2024.ml_orders_dataset
WHERE
  order_status NOT IN ('unavailable', 'canceled');
```

Explanation:

- `NOT IN ('unavailable', 'canceled')`: Excludes rows where the `order_status` is 'unavailable' or 'canceled'.
- `DISTINCT`: Ensures that only unique `order_status` values are included in the result.

This query returns all distinct order statuses that are neither 'unavailable' nor 'canceled'.

👍 🗲️ 🗲️ 🗲️ 🗲️