

## **Business Analysis with Structured Data**

### **A2 Assignment**

#### **Professor Federico Torreti**

**Group 5:** Juan Diego Hernández, Michael Holmes, Gabriel Yearwood, Abi Joshua, Geoff Patterson, Tariq Bin Jibreel

#### **Youtube link**

<https://youtu.be/WH1TGaVI0os?si=VChIWitqoRf-HUTm>

### **Script**

### **ETL**

After uploading the original SQL dataset into our AWS relational database, the first step was to understand the er diagram of our dataset.

Then the next step was to clean and prepare the data. Some tables, the ml\_marketing\_qualified\_leads\_dataset, ml\_orders\_dataset, and ml\_product\_category\_name\_translation, had generic column headers ( C1, C2).

These were replaced by setting the first row as the header to make the data easier to interpret and use.

Additionally we created a T\_table with the Orders table, added a Status column to categorize orders based on their delivery state.

Orders marked as "delivered" or "shipped" were classified as complete, while all others were marked as incomplete.

For the items table, Product IDs were linked to Portuguese category names and translated into English.

For the marketing dataset, we standardized the origin column by assigning 'unknown' to null values and entries labeled as "other."

Finally, we joined multiple tables to incorporate additional details such as payment data, geolocation, and customer demographics. This prepared the dataset for analysis and visualization.

For visualizations, we used Python to create geolocation maps and Amazon QuickSight for the other insights.

## **Assumptions**

We are going to assume that if an order got shipped or it got delivered to the customer that the order was completed.

Revenue Analysis is based off of the orders\_payments table, By doing this it is assumed that if an order has a payment value associated with it then a transaction has occurred thus increasing our revenue.

NULL Values plaguing this data set. for example there are some orders that got completed but there is no product associated with it or freight value but there is a price. Sometimes the opposite occurred and there was a product shipped out but no customer or order linked to it. This is to be expected since Mercado Libre acts as a middle man linking the customer to a seller. These Null values occurred in our data set for two possible reasons. First possible explanation is that the data got corrupted when it was sent to us or the seller forgot to fill in every detail before sending it to us. This would explain the above phenomenon. That being said we did provide an analysis on the information that was completely filled out just keep in mind that the vast majority of the data set is incomplete.

## **Marketing**

### **Say what the market channels are**

Our marketing channels are organic\_search(google searching for a product), paid\_search( sponsored ads on google's search engines), social media, direct\_traffic(direct access to our website via a link), email, referral, display ( visually stunning ads found online) , and unspecified (These are marketing channels that could be any of the above or things not listed here).

The top 3 marketing channels are

Our top three marketing channels are organic search, paid search, and unspecified . These account for ~66% of our marketing leads with organic search accounting for ~29% of our marketing leads.

## **Section 2: Operations**

**Payment Methods:** Starting with payment methods, credit cards are the most used, accounting for about 74% of transactions. This reflects their convenience and highlights the need to ensure smooth credit card payment processing.

**Delivery Speed Trends:** Next, delivery speed is a crucial factor. Faster delivery boosts orders, revenue, and customer retention. Customers value saved time, while slower delivery can drive them to seek alternatives, stressing the need for efficient logistics

**Review Trends by State:** In terms of reviews by state, 7,202 users rated us with an average score of 4.1 out of 5. Positive reviews dominate, with 4 and 5-star ratings making up 77%. This feedback strengthens customer trust and highlights overall satisfaction.

**Review Trends by Product Category:** This chart shows the average review scores across different product categories. Most categories maintain strong ratings, with averages between 4 and 5 once again. Notably, higher review scores are observed in categories like construction tools and electronics or musical instruments, suggesting strong customer satisfaction in these areas. Overall, the total average review score stands at 4.1, reflecting consistent positive feedback across the platform.

**Review Score Relationships:** From this, we can conclude that there is a relationship between the number of reviews and the average review score. Now let us take a look at Sales

## **Sales performance**

### REACH

We are able to reach customers all over Brazil with our biggest wins being in states like Sao Paulo, Minas Gerais, and Rio De Janeiro accounting for ~ 87% of all sales which is to be expected since they are the three richest states in Brazil.

### MOST POPULAR PRODUCTS

The top 3 most sold product categories are bed\_bath\_table, telephony and a tie in third place for sports\_leisure, furniture\_decore.

### correlation

We observed that there is no positive correlation (~8%) between expensive furniture and positive reviews. In fact we found that the more expensive an item is the worse review it got. This is to be expected because people will be more critical of the quality of an item if it is more expensive.

### REVENUE

#### **Quarterly:**

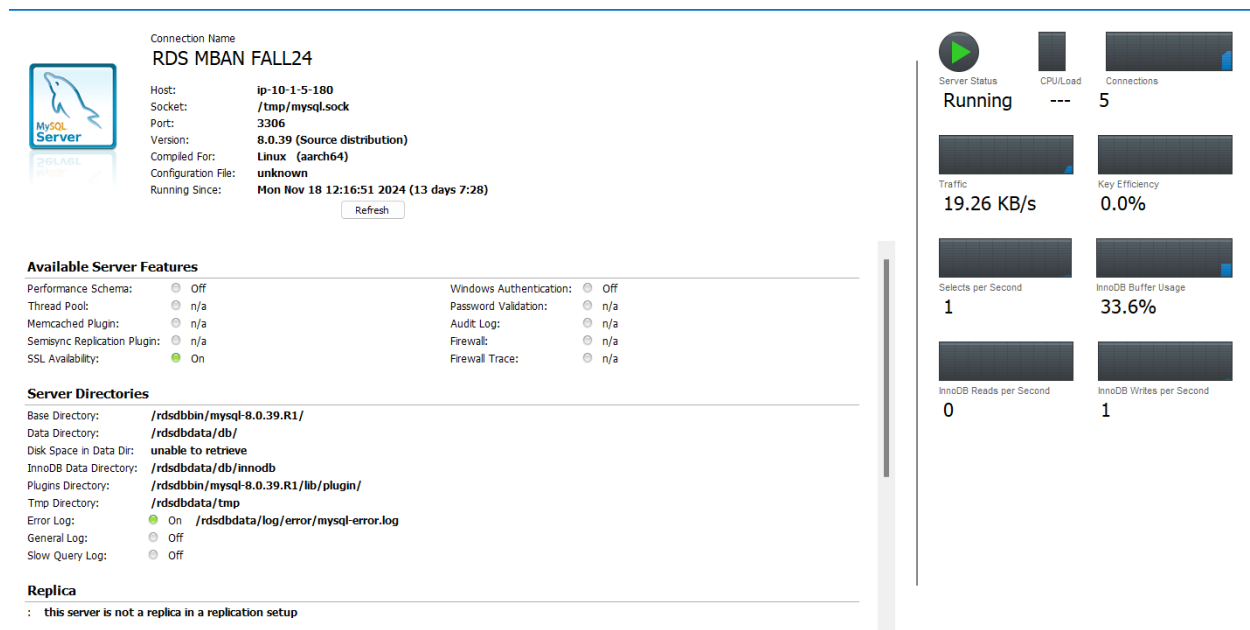
It is clear that this last fiscal year has been a success. We observed the company's revenue has grown throughout the fiscal year. Let's dive deeper into the monthly analysis to see why this growth has occurred

#### **MONTH:**

ML's best selling months occurred during the months of October and November seeing a spike in revenue during that time period. We hypothesize this is due to the Holiday season and everyone buying gifts for their loved ones. It is then observed that there is a drop off in december most likely due to christmas shopping being completed, but come January to April there is a steady growth in revenue compared to the previous year, we suspect that this means that people decided to try out our platform during the holidays, liked it and decided to use it more in the coming years.

## APPENDIX A

### RDS Database Connection



## ER-Diagram

