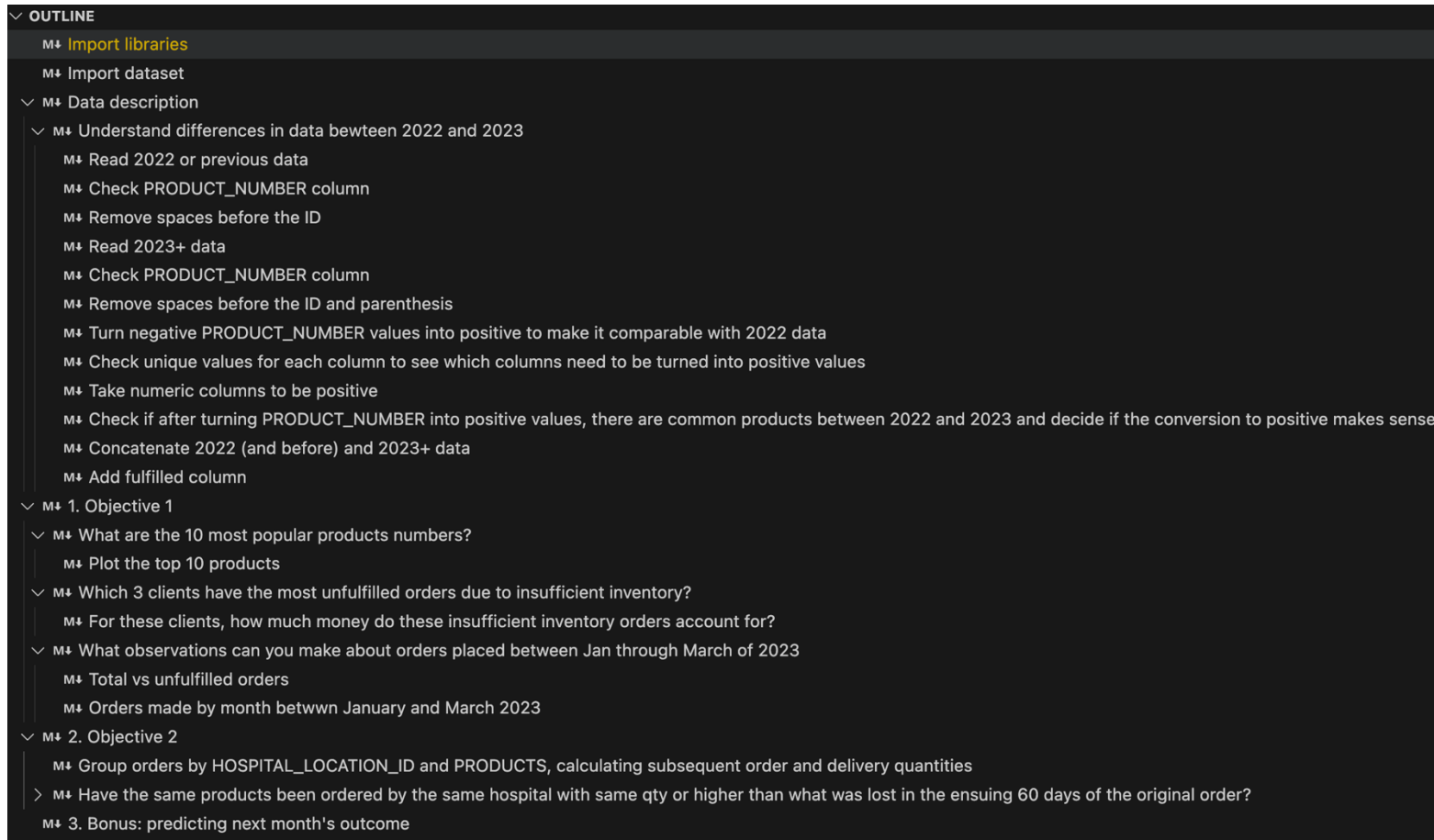


Order Data Analysis 2021 - 2023

Presenter: Laura Saldarriaga Higueta

Methodology

The proposed solution for this challenge was developed with the following outline:



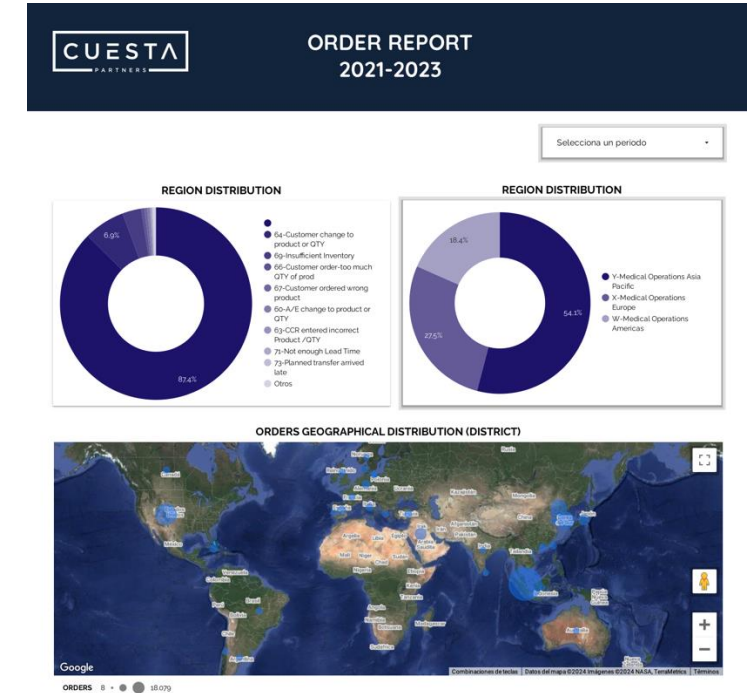
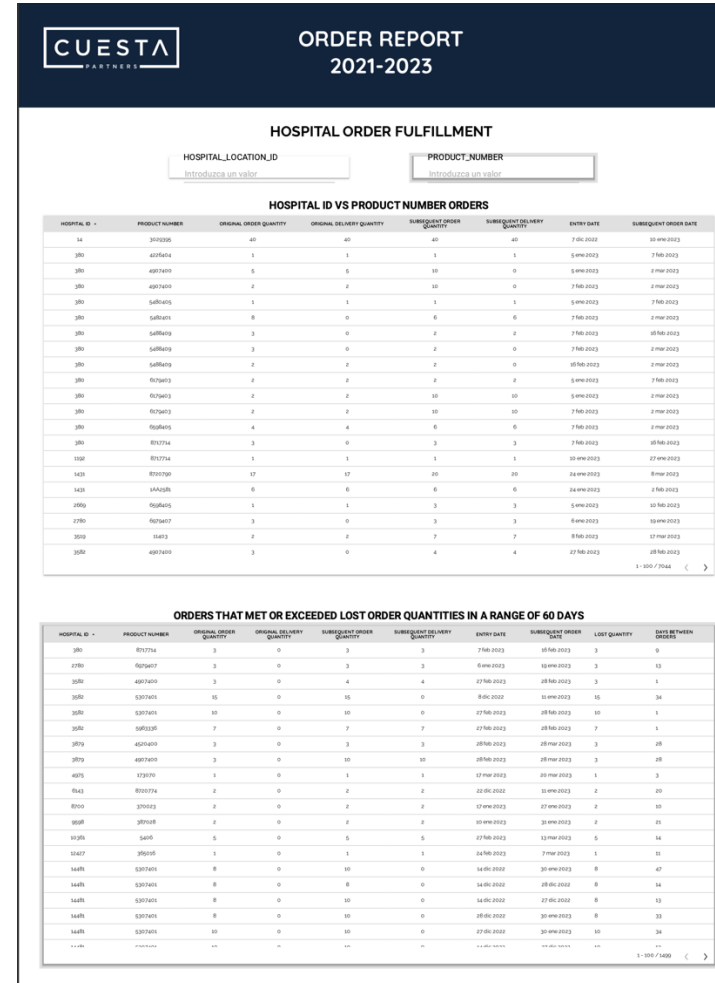
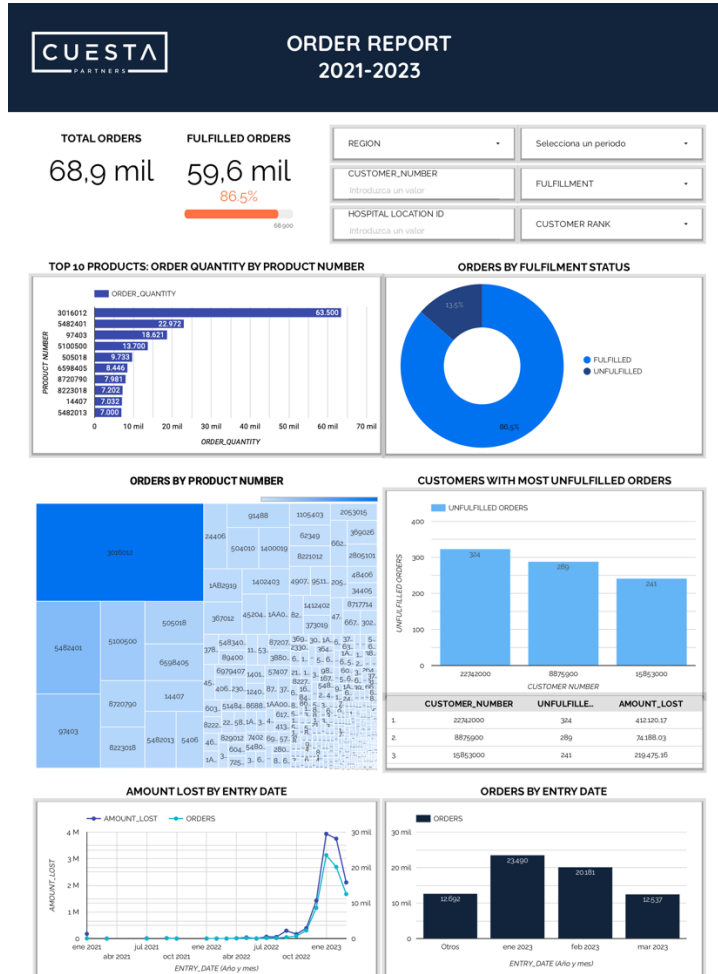
The screenshot shows a Jupyter Notebook outline with the following structure:

- OUTLINE
 - M+ Import libraries
 - M+ Import dataset
 - ✓ M+ Data description
 - ✓ M+ Understand differences in data between 2022 and 2023
 - M+ Read 2022 or previous data
 - M+ Check PRODUCT_NUMBER column
 - M+ Remove spaces before the ID
 - M+ Read 2023+ data
 - M+ Check PRODUCT_NUMBER column
 - M+ Remove spaces before the ID and parenthesis
 - M+ Turn negative PRODUCT_NUMBER values into positive to make it comparable with 2022 data
 - M+ Check unique values for each column to see which columns need to be turned into positive values
 - M+ Take numeric columns to be positive
 - M+ Check if after turning PRODUCT_NUMBER into positive values, there are common products between 2022 and 2023 and decide if the conversion to positive makes sense
 - M+ Concatenate 2022 (and before) and 2023+ data
 - M+ Add fulfilled column
 - ✓ M+ 1. Objective 1
 - ✓ M+ What are the 10 most popular products numbers?
 - M+ Plot the top 10 products
 - ✓ M+ Which 3 clients have the most unfulfilled orders due to insufficient inventory?
 - M+ For these clients, how much money do these insufficient inventory orders account for?
 - ✓ M+ What observations can you make about orders placed between Jan through March of 2023
 - M+ Total vs unfulfilled orders
 - M+ Orders made by month between January and March 2023
 - ✓ M+ 2. Objective 2
 - M+ Group orders by HOSPITAL_LOCATION_ID and PRODUCTS, calculating subsequent order and delivery quantities
 - > M+ Have the same products been ordered by the same hospital with same qty or higher than what was lost in the ensuing 60 days of the original order?
 - M+ 3. Bonus: predicting next month's outcome

- **Jupyter Notebook:** *orders_data_analysis.ipynb*
- **Scripting language:** Python
- **Dashboard tool:** Google Looker Studio
- The "Data description" section contains all data procedures used to have cleaner and more organized data.
- Also, data from 2022 and 2023 needed to be standardized to have accurate data in time.
- Then, Objective 1 is approached by answering 3 questions.
- Later, Objective 2 is developed by creating new *hospital_location_id-based* tables.
- Finally, bonus challenge is addressed
- 3 new files are generated: *order_data_processed.csv*, *orders_met_or_exceeded_by_hospital_id.csv*, and *subsequent_orders_by_hospital_id.csv*

Methodology

Looker Studio Dashboard



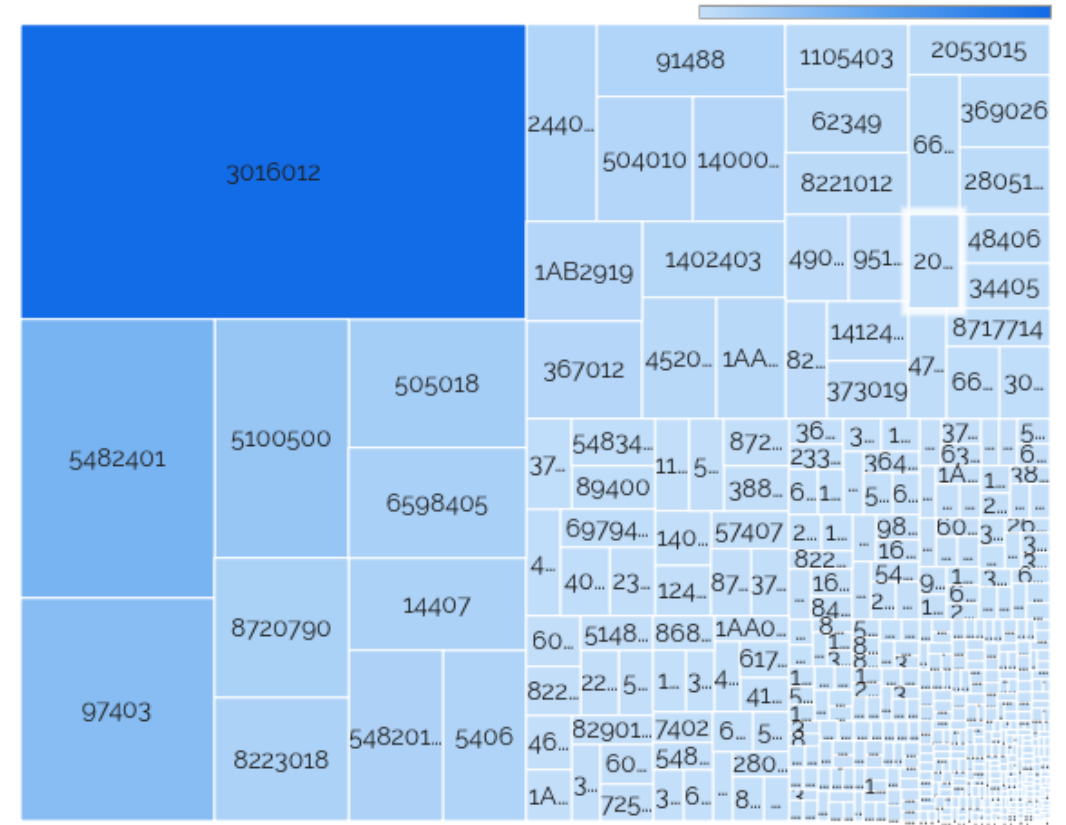
Objective 1

Q1: What are the 10 most popular products numbers?

TOP 10 PRODUCTS: ORDER QUANTITY BY PRODUCT NUMBER



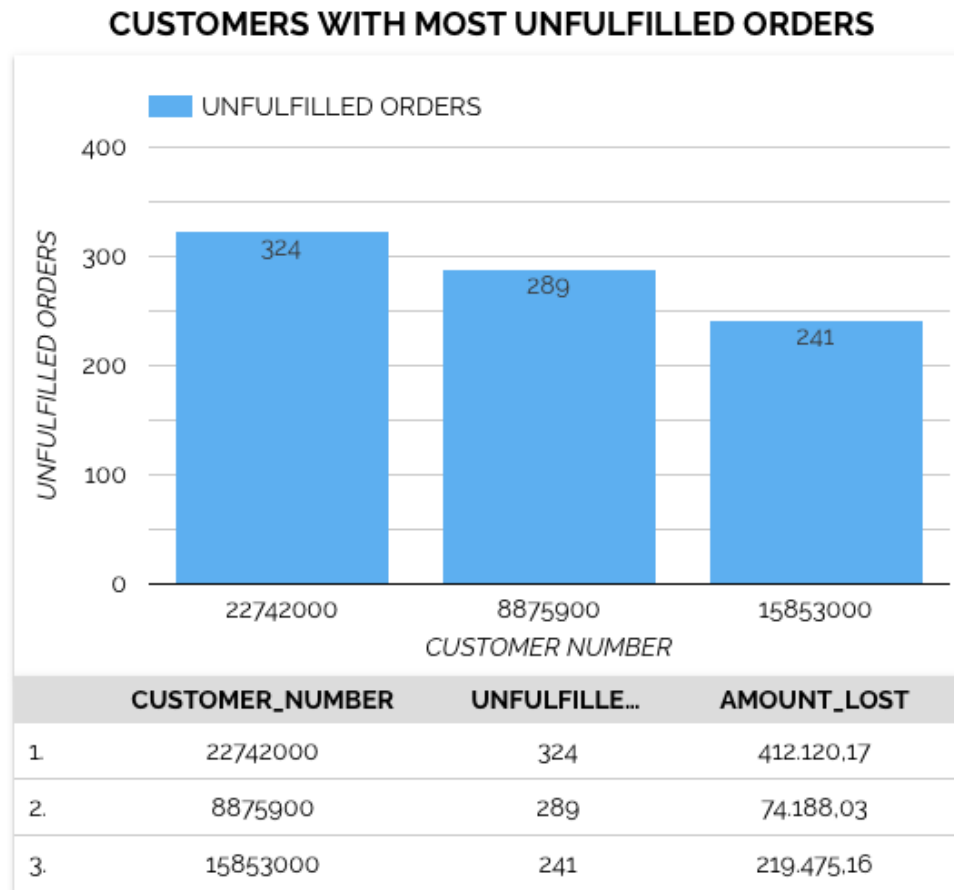
ORDERS BY PRODUCT NUMBER



Objective 1

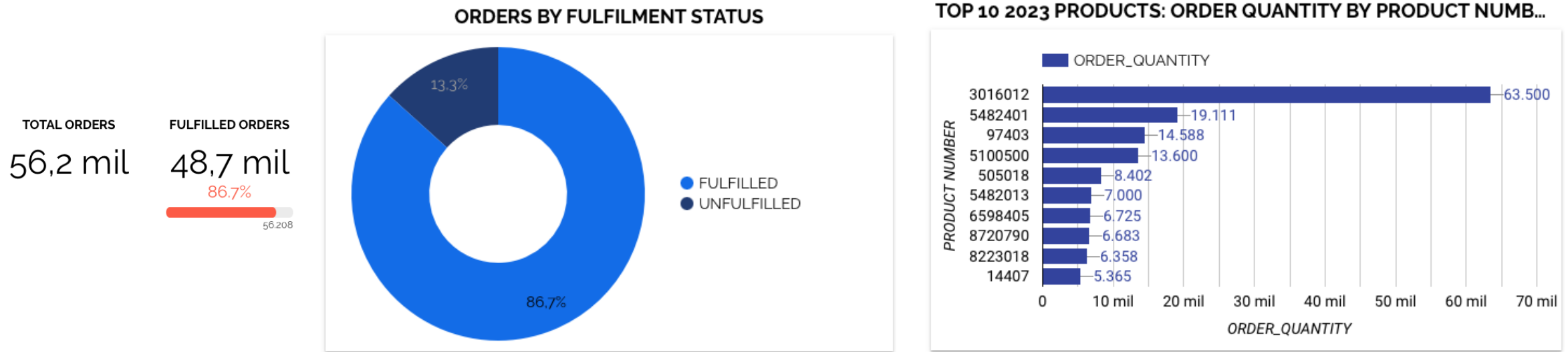
Q2.0: Which 3 clients have the most unfulfilled orders due to insufficient inventory?

Q2.1: For these clients, how much money do these insufficient inventory orders account for?



Objective 1

Q3: What observations can you make about orders placed between Jan through March of 2023

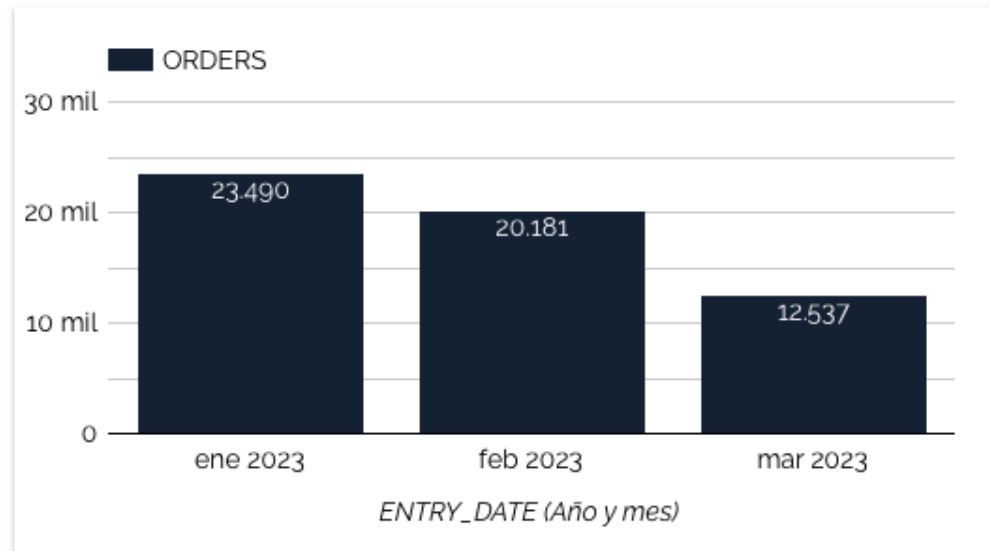


- 56.2k orders placed between January and March 2023 and 86.7% of them were fulfilled successfully in terms of available and delivered inventory. This suggests a relatively strong fulfillment rate overall
- Orders of most popular products were mostly placed during this period

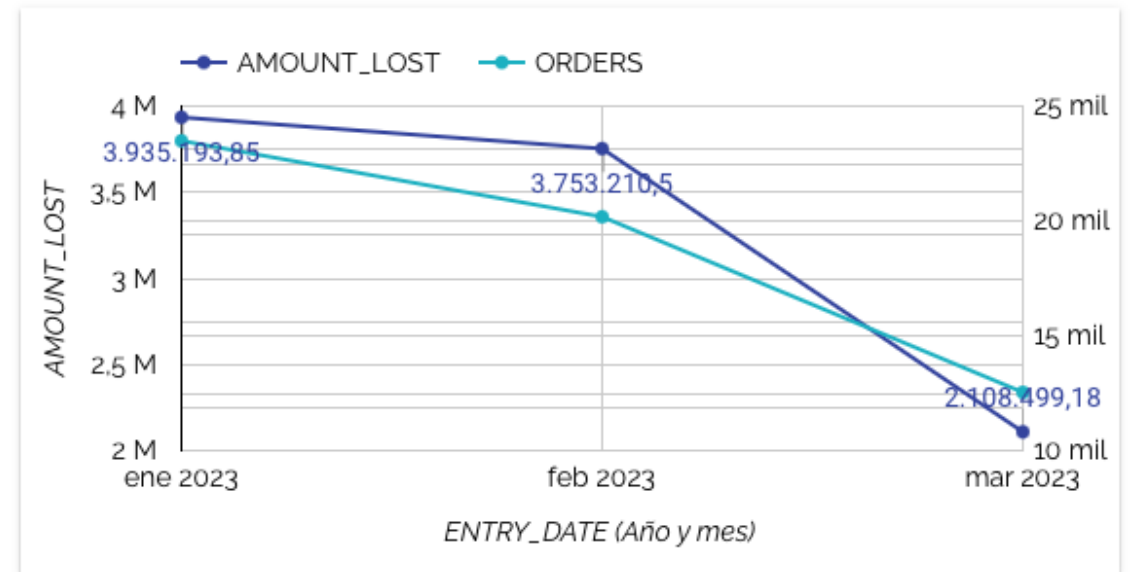
Objective 1

Q3: What observations can you make about orders placed between Jan through March of 2023

ORDERS BY ENTRY DATE: JAN TO MARCH 2023



AMOUNT LOST BY ENTRY DATE: JAN TO MARCH 2023



- First quarter of 2023 presented the highest number of orders placed, when compared to years 2021 and 2022. The demand of products only in January, exceeded the orders placed from 2021 to 2022
- It's noted that orders and amount lost both started with a decreasing tendency in 2023, where January presented the highest value for both variables

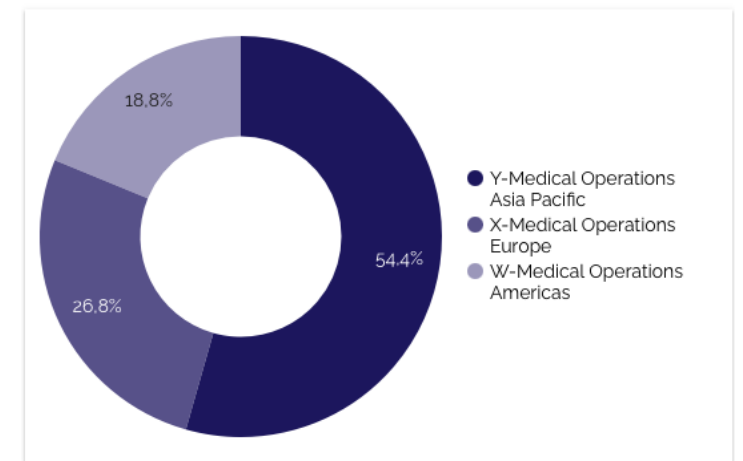
Objective 1

Q3: What observations can you make about orders placed between Jan through March of 2023

ORDERS GEOGRAPHICAL DISTRIBUTION (DISTRICT)



REGION DISTRIBUTION



- In this period, districts of Asia Pacific represented the highest order demand with 54.4%
- Europe surpassed the order placement of the United States with a 26.8%, which indicates that markets in this area should be explored more

Objective 2

CEO statement:

“Sure, we may leave a bit of money on the table when we are unable to fulfill orders based on insufficient inventory; however, I am not too worried about it because our products are essentials, and our hospitals will re-order the same product with the same quantity or higher within the ensuing 60 days the majority of the time”

Key consideration: have the same products been ordered by the same hospital with same qty or higher than what was lost in the ensuing 60 days of the original order?

HOSPITAL ID ▲	PRODUCT NUMBER	ORIGINAL ORDER QUANTITY	ORIGINAL DELIVERY QUANTITY	SUBSEQUENT ORDER QUANTITY	SUBSEQUENT DELIVERY QUANTITY	ENTRY DATE	SUBSEQUENT ORDER DATE
380	4907400	5	5	10	0	5 ene 2023	2 mar 2023
380	4907400	2	2	10	0	7 feb 2023	2 mar 2023
380	5480405	1	1	1	1	5 ene 2023	7 feb 2023
380	5482401	8	0	6	6	7 feb 2023	2 mar 2023
380	5488409	3	0	2	2	7 feb 2023	16 feb 2023
380	5488409	3	0	2	0	7 feb 2023	2 mar 2023
380	5488409	2	2	2	0	16 feb 2023	2 mar 2023
380	6179403	2	2	2	2	5 ene 2023	7 feb 2023
380	6179403	2	2	10	10	5 ene 2023	2 mar 2023

Case 1: the subsequent order is delivered but is smaller than the original order

Case 2: the order is placed again but it's not delivered as the first time

Note: Reason code variable was not significant to understand the underlying causes of the unfulfilled orders

Objective 2

CEO statement:

“Sure, we may leave a bit of money on the table when we are unable to fulfill orders based on insufficient inventory; however, I am not too worried about it because our products are essentials, and our hospitals will re-order the same product with the same quantity or higher within the ensuing 60 days the majority of the time”

Key consideration: have the same products been ordered by the same hospital with same qty or higher than what was lost in the ensuing 60 days of the original order?

HOSPITAL ID ▲	PRODUCT NUMBER	ORIGINAL ORDER QUANTITY	ORIGINAL DELIVERY QUANTITY	SUBSEQUENT ORDER QUANTITY	SUBSEQUENT DELIVERY QUANTITY	ENTRY DATE	SUBSEQUENT ORDER DATE	LOST QUANTITY	DAYS BETWEEN ORDERS
27807	97403	15	0	15	15	3 ene 2023	24 feb 2023	15	52
27807	97403	15	0	15	15	3 ene 2023	6 feb 2023	15	34
27807	97403	15	0	15	0	3 ene 2023	13 ene 2023	15	10
27807	97403	15	0	15	15	13 ene 2023	24 feb 2023	15	42

Case 3: the subsequent order is delivered and the quantity ordered is the same or higher

Based on the analysis, **77.40%** of reorders placed by hospitals met or exceeded the quantities initially lost due to unfulfilled deliveries. This suggests that while most of the unfulfilled orders were successfully compensated through subsequent orders, there remains a notable **22.6%** where reorders fell short.

Conclusion and recommendations

- Addressing the gap between successfully fulfilled orders could be crucial for improving inventory reliability and ensuring that products reach their intended destinations in a consistent way. A focus on identifying and addressing the causes behind this lack of inventory may help enhance overall fulfillment rates, better meet demand, and decrease losses.
- Although most orders are recovered, an approach to reducing unfulfilled orders could boost revenue by maintaining clients satisfied.
- It's important to identify the causes of the decrease in orders for Q1 in 2023, since the downward tendency could extend to other months. This is relevant because losses are being proportional to the order placement, which could be reflecting inventory problems to fulfill customer orders.
- Despite having a high fulfillment rate, it's worth addressing the 13.3% of unfulfilled orders for the period between January – March 2023. This is a small percentage, but there are still important losses that are not being compensated by reordering from hospitals.
- European markets should be another focus of sales campaigns, although this could carry a more difficult set up since orders have been distributed in small different countries, instead of few bigger districts. Optimizing order transport should be considered.
- Since 22.6% of reorders did not meet initial quantities, a priority system for high-demand products could be implemented, especially for the most ordered products by the biggest clients. This suggests there is room for targeted inventory management adjustments. This could help avoid repeat shortages and further improve customer satisfaction for high-priority clients.
- After a time-based evaluation, a high-priority products list could be considered to have availability of inventory to supply the orders, reducing losses.

Bonus Question

- Set up and explain your methodology to predict the next month's Volume and/or sales amount using the current data set
- Predict the next month's Volume and/or Sales amount

First steps of the setup:

- Started by converting the Year_Month column to datetime format
- Took the columns for **order_quantity**, **delivery_quantity**, and **delivery_line_item_total** and put the cumulative values per month

Then, **with the aid of Claude.AI**, I used two approaches:

1. Trend-based forecasting:

Uses linear regression model to capture the trend in the data. The *create_trend_forecast()* function fits a linear model on the Days_From_Start feature and the target metric (Total_Order_Quantity, Total_Delivery_Quantity, Total_Sales_Revenue). It then predicts the value for the next month.

In the *analyze_and_forecast()* function, the Mean Absolute Percentage Error (MAPE) is calculated for the trend-based forecasts. This provides a measure of the forecast accuracy, which can be useful for assessing the reliability of the predictions.

2. Moving Average-based forecasting

This calculates the average of the **last 6 months** of data for each metric using the *calculate_moving_average()* function. This provides a forecast based on the recent historical patterns.

Bonus Question

- Set up and explain your methodology to predict the next month's Volume and/or sales amount using the current data set
- Predict the next month's Volume and/or Sales amount

Forecast Results:

Total_Order_Quantity:

Current value: 47,377.00
Trend-based forecast: 51,441.20
Moving average forecast: 57,562.00
Trend percent change: 8.58%
MA percent change: 21.50%
Model MAPE: 42.03%

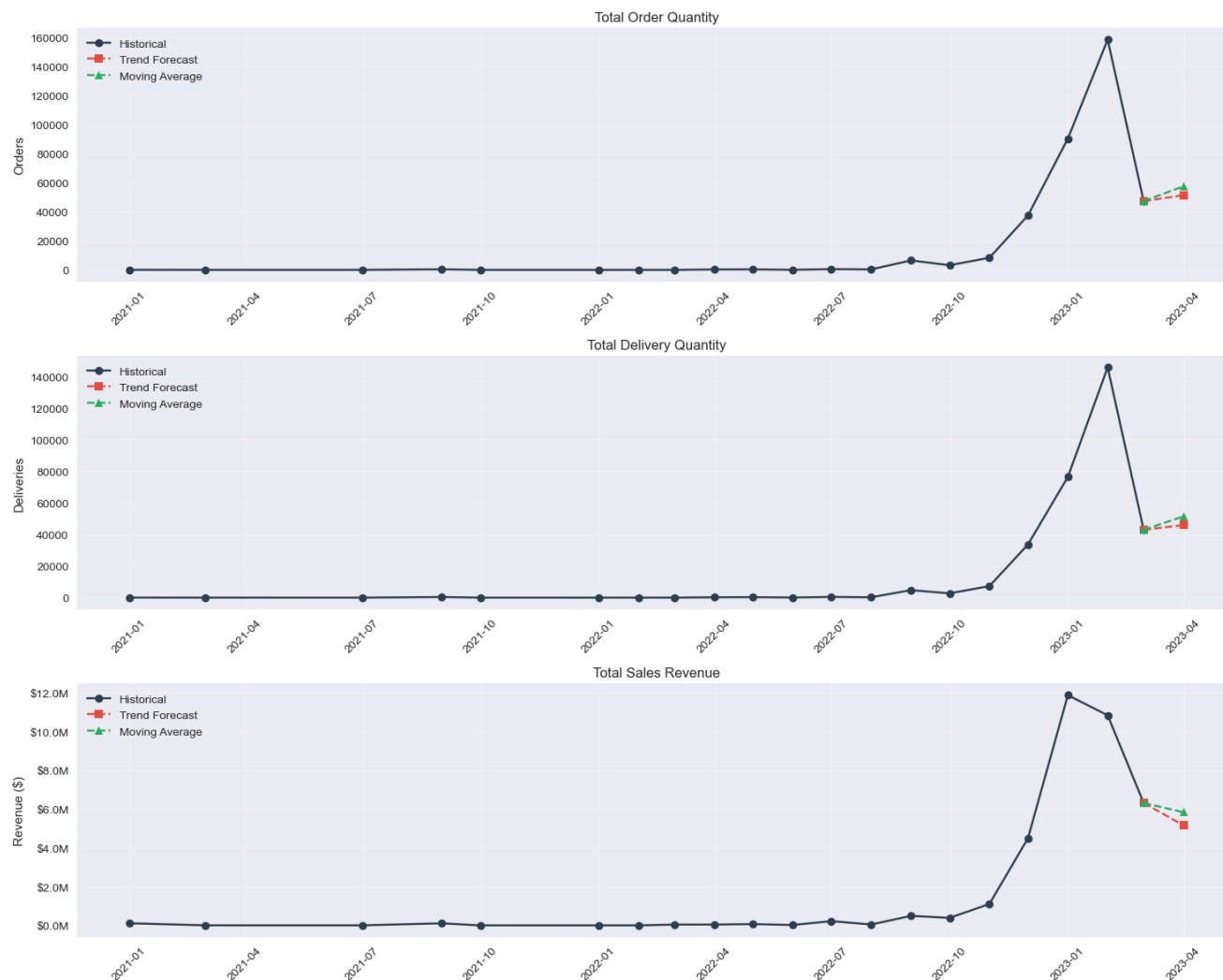
Total_Delivery_Quantity:

Current value: 42,936.00
Trend-based forecast: 45,932.76
Moving average forecast: 51,458.67
Trend percent change: 6.98%
MA percent change: 19.85%
Model MAPE: 40.92%

Total_Sales_Revenue:

Current value: 6,314,836.95
Trend-based forecast: 5,161,668.00
Moving average forecast: 5,836,116.54
Trend percent change: -18.26%
MA percent change: -7.58%
Model MAPE: 48.13%

The percent change between the current value and the forecasted value for both the trend-based and moving average methods is also calculated. In this case, it gives an indication of growth for quantities and decline for sales.



Bonus conclusion

- Other forecasting methods should be evaluated to try to capture more complex patterns or seasonality, since the methods that were shown, present indications of not being very accurate ($< 50\%$).
- Data from the available months may not be sufficient to address the problem, since most orders were placed during the last 3 months.
- ARIMA and SARIMA methods were tested but data was not enough to capture the pattern, since the model didn't have enough data points to work with. An evaluation with data of more months is suggested.
- A more detailed implementation should be evaluated. The one that was shown comes mostly from an AI tool and a proper revision needs to be made.