Ecommerce Business Analysis

Presented by:

Sharil Ashar

Date Presented :

20th June 2023

Agenda

Business Scenario

You are working as a business/data analyst for **Shopee, Brazil** and the company has assign you to improve revenue through creating a customer base, enhance customer experience and also recommend relevant products to the customer. You are also given only 2 sets of data to work on by the stakeholders. You are to present your key findings/insights as well as the steps to achieve them to the stakeholders.

Problem Statement

01
Create a customer base

02
Enhance the customer experience

03
Recommend relevant products to customer

Proposed Solution

Conduct an Exploratory Data Analysis
To better understand the dataset and get a brief insight on the customers and their purchases.

Build a Sales Performance dashboard
For further analysis and insights.

Build a RFM Dashboard
To conduct a RFM analysis so as to improve customer experience.

Implement a Product Recommendation system
For the customer to choose relevant products

Data Preprocessing & Exploratory Data Analysis (EDA)

Objectives :
1. Confirm if the data is making sense in the context of the problem statement.
2. Uncover and resolve for data quality issues.
3. Drop unwanted columns and derive new variables.
4. Have a general understanding of the data.

Let's move to Anaconda for this section!

Data Preprocessing & EDA

Summary

Data Understanding

- The dataset contains information about the customer's information and purchases of **Shopee, Brazil** occurring between 15/08/2019 and 01/01/2023.

Data Cleaning

- The combined dataset only had 1 row with NULL values inside. It is dropped.
- No duplicates have been detected.
- Change the data type for 'order_date' and 'shipping_date' to datetime format.
- Generally, a clean dataset.

Data Transformation

- Extracted day, month, year , day_name, month_name from 'order_date'.
- Created new column for price after discount
- Removed data from 2023 as it is not needed.

Data Preprocessing & EDA

Findings

Univariate Analysis

1. Where are our customers from and its distribution?
   - They are from Brazil, Chile, Mexico as well as Chile and they are almost evenly distributed with Colombia being the highest and Mexico being the lowest.

2. What is the income distribution among customers?
   - It has a mean of approximately $84,692 with a standard deviation of $37,451. The range of values spans from $20,205 to $149,980, while the 25th, 50th, and 75th percentiles are $52,092, $83,944, and $116,821, respectively.

3. What is the distribution of customer genders?
   - It is relatively balanced at 50.6% and 49.4% for male and female respectively.

4. What is the distribution of customer ages?
   - The age distribution has a mean of approximately 41.34. The age ranges from 18 to 65, with the median age at 41. The interquartile range spans from 29 to 53.

Data Preprocessing & EDA

Findings

Multivariate Analysis

1.Who are our top 10 customers based on total revenue?
1. CS00609
2. CS00834
3. CS00348
4. CS00296
5. CS00336
6. CS00384
7. CS00322
8. CS00819
9. CS00178
10. CS00351

2. Who are our top 10 customers based on quantity purchased?
1. CS00176
2. CS00611
3. CS00703
4. CS00742
5. CS00997
6. CS00291
7. CS00563
8. CS00872
9. CS00293
10. CS00527

Data Preprocessing & EDA

Findings

Multivariate Analysis

3. What are the top 5 products based on revenue?
    1. Curtains
    2. Lighting
    3. Wall Art
    4. Rugs
    5. Kitchen Appliances

4. What are the top 5 products based on quantity?
    1. Sports Shoes
    2. Jeans
    3. Jacket
    4. Wireless Speaker
    5. Furniture

5. Who is the top performing country based on revenue?
    - Colombia with $13.5 million in total revenue.

Data Preprocessing & EDA

Findings

Time Series Analysis

1.Which time of year produce the most revenue?
- It is during the second half of the year starting from August, which then peaks at October before dipping slightly in November and December.

2. Which time of month produce the most sales?
- Customers tend to purchase more from the 18th to 23rd while the least is from 10th to 17th. This excludes the last day of the month as not all months have 31st, which explains the sudden dip.

3. Which time of the week produce the most sales?
- Most purchases are made on a Sunday and Tuesday in contrast to Friday, which has the least amount of purchases.

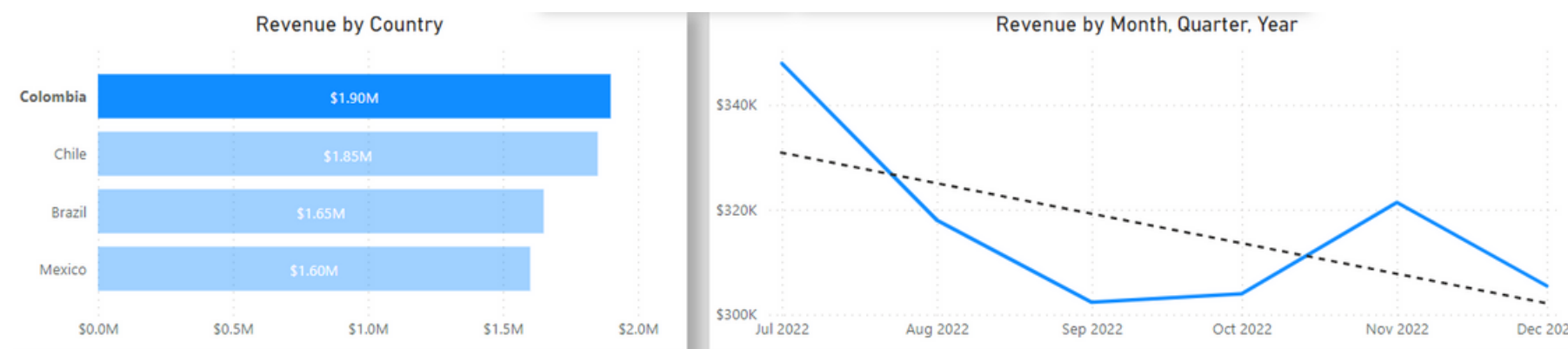# Data Visualization and Insights
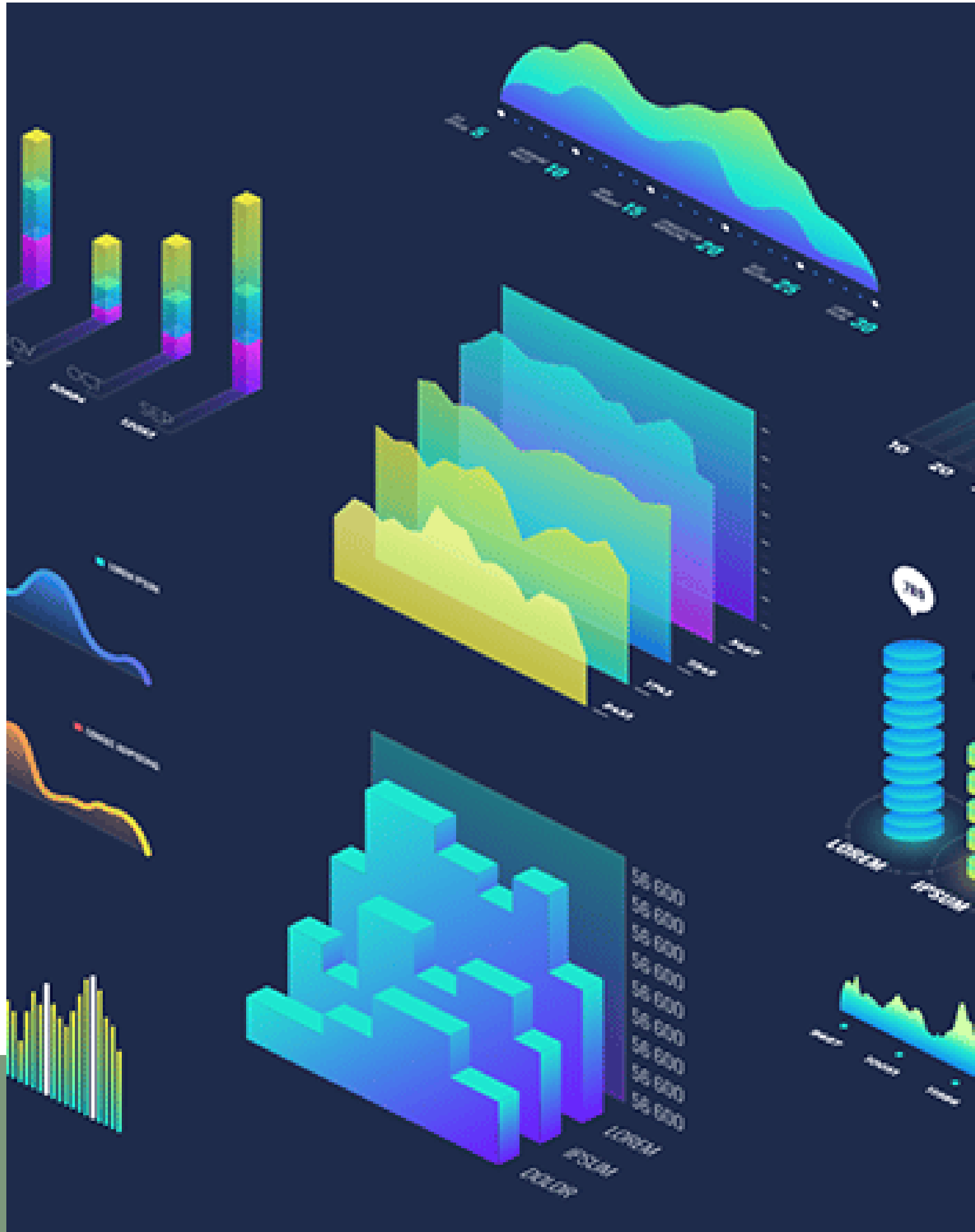
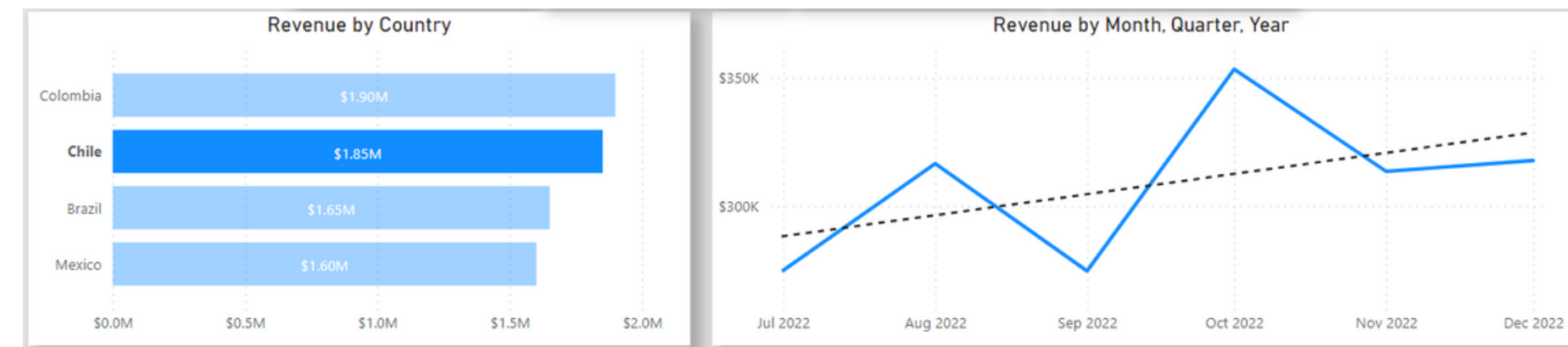Let's move to PowerBI for this section!

Insight #1 : Trend Analysis by Country

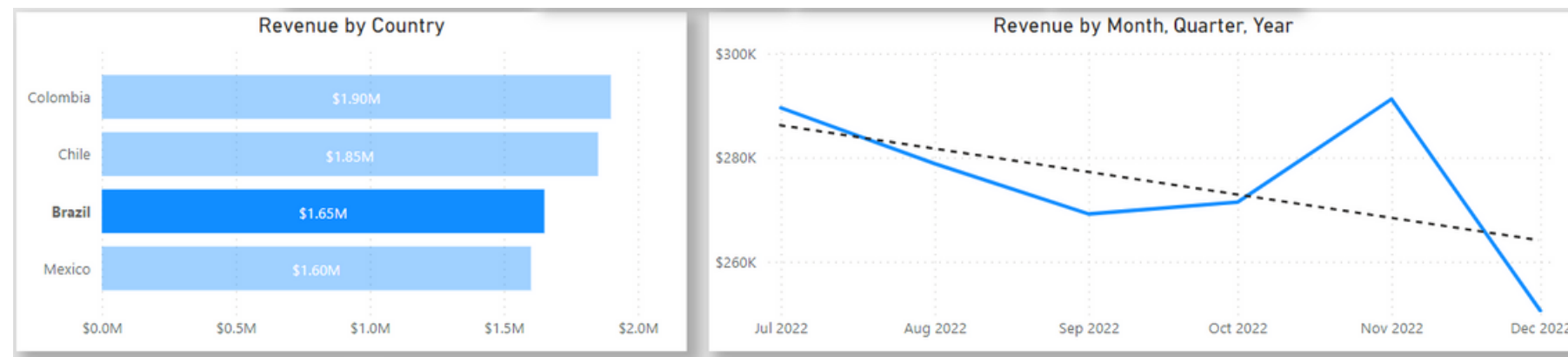Which country is currently experiencing an upward trend in revenue based on the past 6 months?
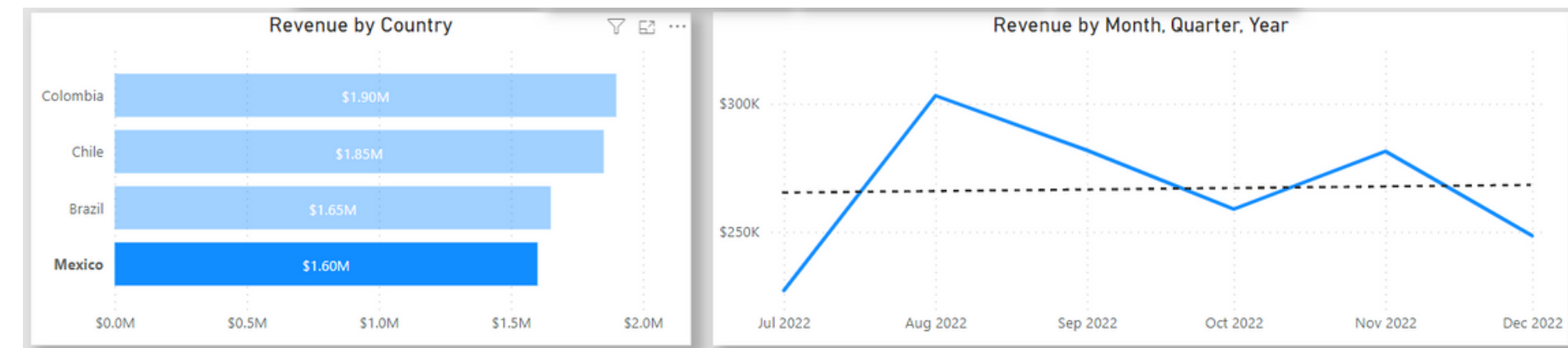
Colombia

Revenue by Country

Colombia	$1.90M
Chile	$1.85M
Brazil	$1.65M
Mexico	$1.60M

Revenue by Month, Quarter, Year

Chile

Revenue by Country

Colombia	$1.90M
Chile	$1.85M
Brazil	$1.65M
Mexico	$1.60M

Revenue by Month, Quarter, Year

Brazil

Revenue by Country

Colombia	$1.90M
Chile	$1.85M
Brazil	$1.65M
Mexico	$1.60M

Revenue by Month, Quarter, Year

Mexico

Revenue by Country

Colombia	$1.90M
Chile	$1.85M
Brazil	$1.65M
Mexico	$1.60M

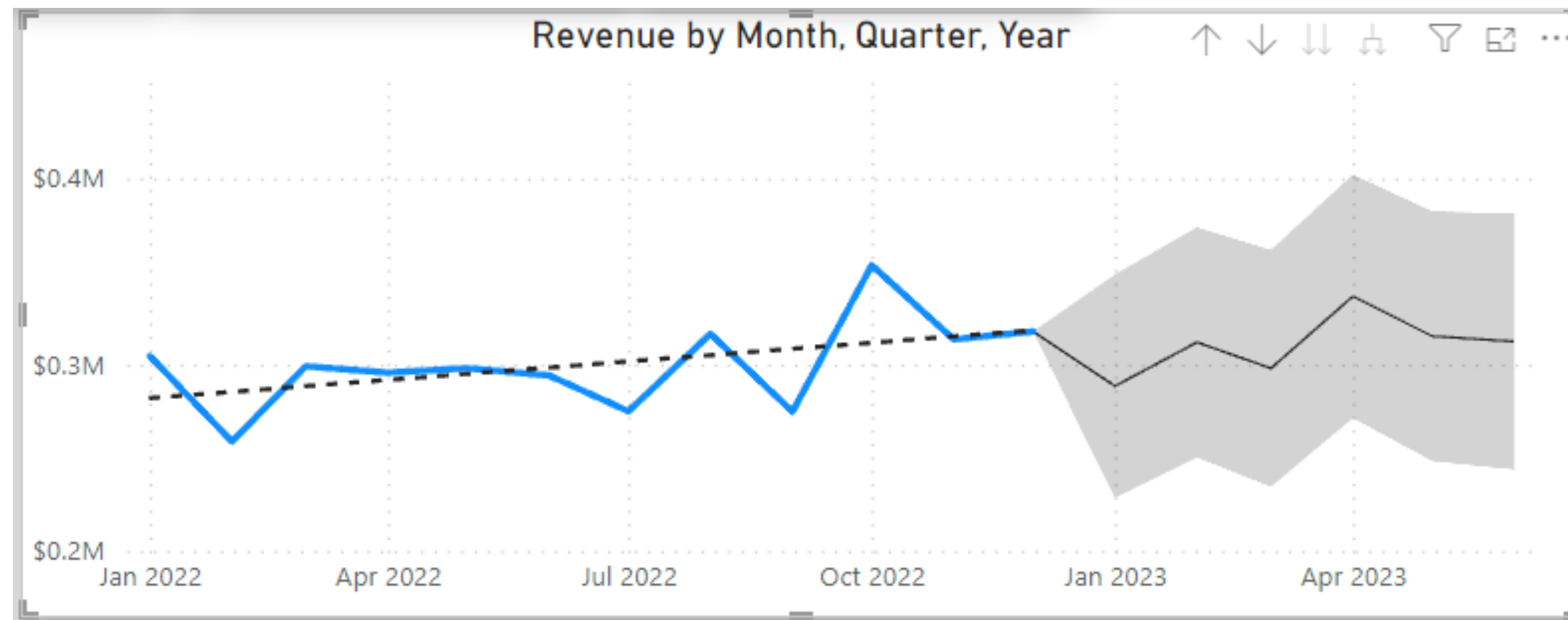**Revenue by Month, Quarter, Year**

- Despite generating the most revenue from Columbia, the trendline is going downwards. Indicating that they might not be doing well in the future.
- Chile is the only country with a positive trending line and worth looking into.

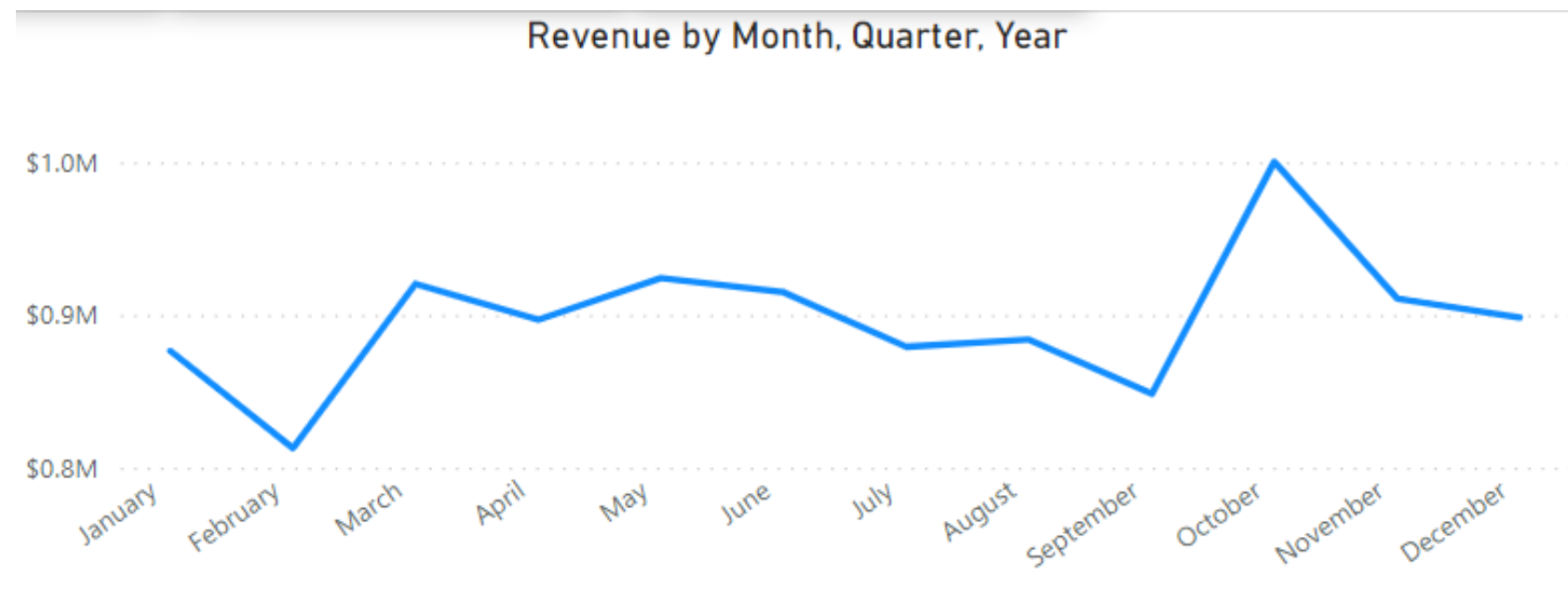# Insight #1 : Trend Analysis by Country

Chile

Based on the forecasting models, the trajectory for Chile in the next 6 months appears to be positive. This is a good indicator for us to decide on a customer base.

Insight #1 : Trend Analysis by Country

What are their weekly and monthly trend like since 2020?

Monthly Trend

Revenue by Month, Quarter, Year

- The month which generates the most revenue is October while the least is February.
- One of the possible reason for the peak is due to celebration of Chile's national holiday, known as "Fiestas Patrias" or Independence Day, which falls on September 18th and extends into early October.

Weekly Trend

Revenue by Day of Week
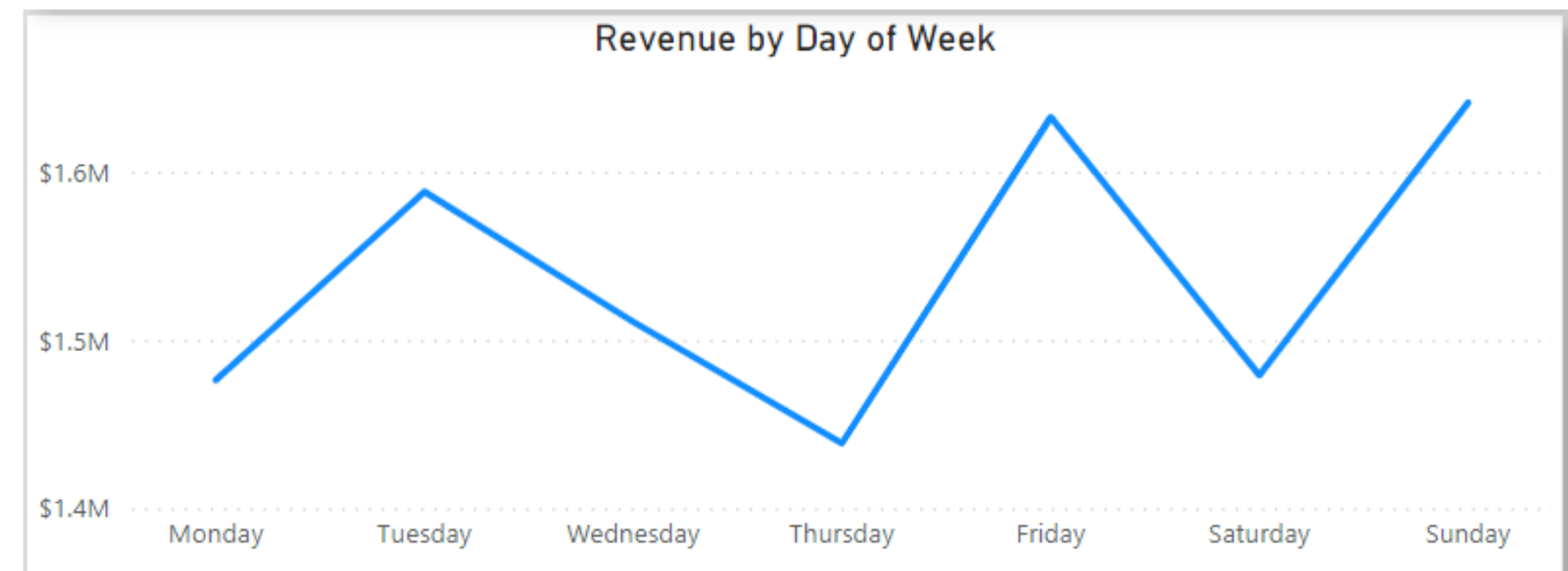
- Tuesday and Friday are the days that generates the most revenue.
- In contrast, Thursday and Saturday have very low revenue.

Insight #2 : Product Analysis

TOP PERFORMING PRODUCTS

Product	Revenue
Wall Art	$0.88M
Curtains	$0.88M
Lighting	$0.87M
Rugs	$0.83M
Dining Set	$0.83M
Furniture	$0.82M
Appliance Set	$0.78M
Home Decor	$0.76M
Bedding	$0.73M
Kitchen Appliances	$0.72M

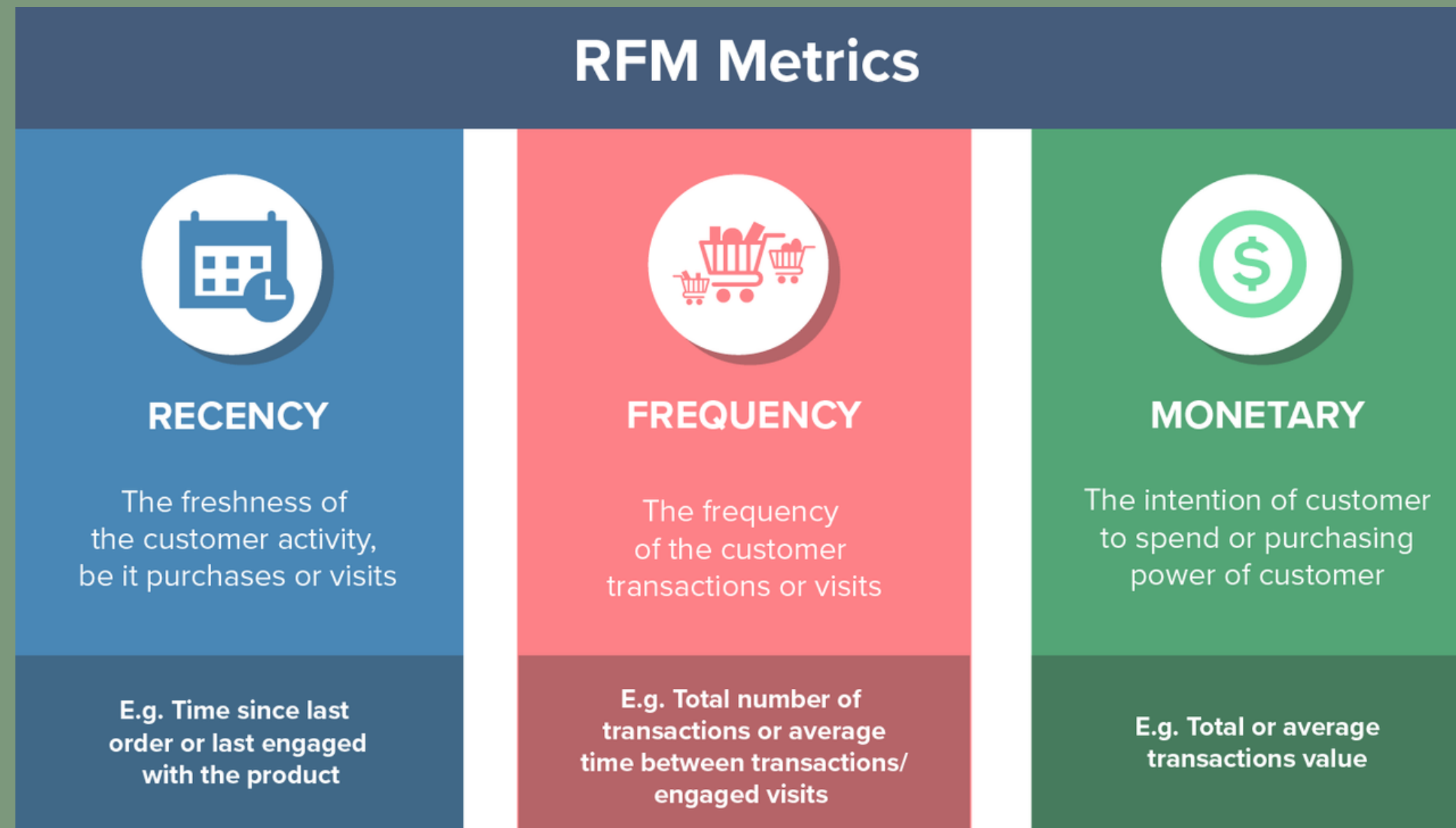- Wall Art is the most popular among Chileans, generating about $880k in revenue.

Insight #3 : Customer Analysis

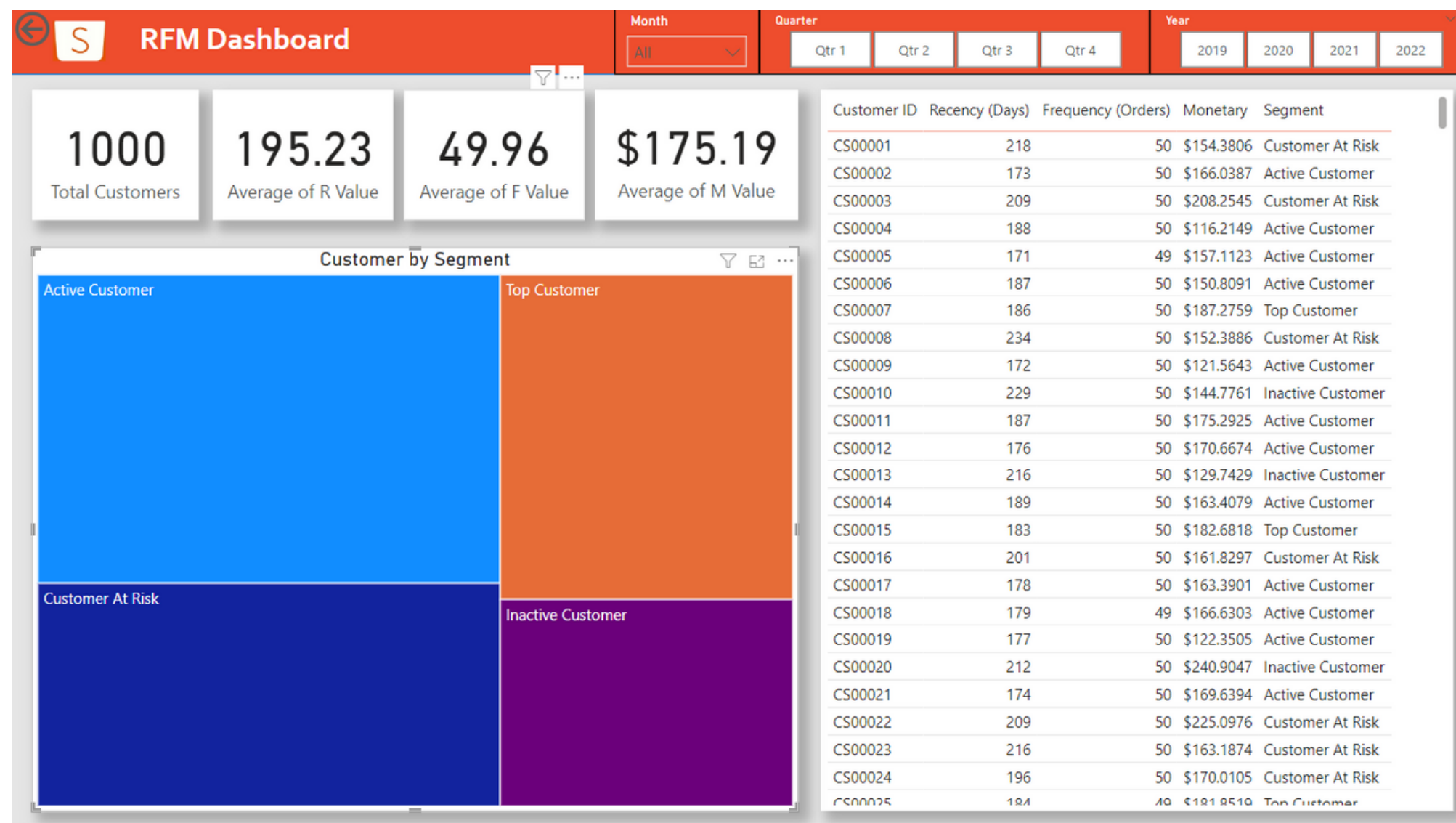- CS00384 is the most loyal generating about $66k in revenue.

RFM Analysis

RFM is a method used for analyzing customer behaviour. It is commonly used in database marketing and direct marketing and has received particular attention in retail and professional services industries.

RFM Metrics

RECENCY

The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product

FREQUENCY

The frequency of the customer transactions or visits

E.g. Total number of transactions or average time between transactions/ engaged visits

MONETARY

The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

Insight #4 : RFM Analysis

How is the customer segmentation like based on RFM?

- After creating the RFM scores and table, the customer will be segmentized into 5 different categories. They are : Top Customer, Active Customer, Average Customer, Customer At Risk, Inactive Customer.
- As there are some limitations to the dataset, the customers are segmentized into 4 categories instead.

Product Recommendation System

Let's move to Anaconda for this section!

Conclusion

- In conclusion, we can <u>create our customer base</u> from Chile, where we can double our marketing efforts between October and December as well as Tuesdays and Fridays.

- We can also <u>enhance customer experience</u> by :
  - Top Customers - Send birthday and anniversary (as being customer) cards with discount vouchers. Create a referral program where they can get a discount upon the first purchase of their referrals.
  - Active Customers - Create loyalty rewards program for this group where they can earn point for every purchase and convert these points into discount or voucher.
  - Customers at Risk - This group of customers need attention as they made some purchases but it's been a long time since their last purchase. Send them personalized emails as well containing promotional items or free samples of products to try to encourage them to be more active.
  - Inactive Customers - Send them personalized emails containing discounted items to encourage them to order and be active.

- Additionally, we can use the <u>product recommendation</u> system to encourage all of them to buy relevant products, thus increasing revenue for the company.

QnA