EXPLORATORY DATA ANALYSIS AND USE OF VISUALIZATIONS TO EXPLAIN A DATA SET USING TOOLS LIKE POWER BI (DASHBOARD), PYTHON AND EXCEL.

BUSINESS ANALYTICS

**REPORT**

**SECTION: A**

TO BE SUBMITTED TO:

**DR. AMMAR AFTAB RAJA**

A Report by;

**Mehmood Ali Khan | S2021054004**

(Student of BBIS [7th Semester])

**TABLE OF CONTENTS**

[**INTRODUCTION** 2](#_Toc166868682)

[**METHODOLOGY** 2](#_Toc166868683)

[**EXPLORATORY DATA ANALYSIS** 3](#_Toc166868684)

[**DATA VISUALIZATION AND DASHBOARD using POWER BI** 6](#_Toc166868685)

[**PIE CHART** 6](#_Toc166868686)

[**LINE AND CLUSTURED COLUMN CHART** 7](#_Toc166868687)

[**STACKED BAR CHART** 7](#_Toc166868688)

[**STACKED AREA CHART** 9](#_Toc166868689)

[**100% STACKED AREA CHART** 9](#_Toc166868690)

[**STACKED COLUMN CHART** 10](#_Toc166868691)

[**MATRIX** 10](#_Toc166868692)

[**SUGGESTIONS/RECOMMENDATIONS** 12](#_Toc166868693)

# **INTRODUCTION**

In this this report, using various tools data exploration and visualization is conducted in a way which can enable the reader to have a very easy outlook of what the dataset is trying to tell. The primary objective is to perform the systematic process of transforming raw data into actionable insights, utilizing the tools like Microsoft Power BI, Microsoft Excel, and Python 3.0 to architect an intuitive dashboard.

Recognizing how important data is for making decisions nowadays, we understand the need to go through big sets of data carefully. That's where exploratory data analysis (EDA) comes in. It helps us find hidden patterns and connections in the data.

In this report, we'll explain each step of the process. We start by getting and cleaning the data, then move on to designing and refining the dashboard. We pay close attention to detail to make sure our dashboard is easy to use and gives important insights.

But it's not just about the technical stuff. We also want to make data exploration and visualization easy for everyone, no matter how much experience they have. Our goal is to share insights that everyone can understand, promoting a culture where decisions are based on data.

# **METHODOLOGY**

**Tools Used:**

* MICROSOFT EXCEL (CSV FILE)
* PYTHON 3.0 (FOR EXPLORATORY DATA ANALYSIS)
* MICROSOFT POWER BI (FOR VISUALIZATIONS AND DASHBOARD)

**STEPS:**

* **Getting the Data:** Start by getting the data from the CSV file. This means bringing the information into the software you're using.
* **Cleaning Up:** Make sure the data is neat and tidy. Remove any mistakes or things that don't belong, and fix any missing bits.
* **Checking the Data in Python:**

1. Bring the cleaned-up data into a Python notebook.
2. Use tools like Pandas, Matplotlib, and Seaborn to look at the data closely.
3. Do some basic math to see what the data is like, and make pictures to help understand it better.
4. Finding Interesting Stuff: Look for cool things in the data. This could be stuff like trends, or things that stand out as unusual.

* **Making the Dashboard in Power BI:**

1. Move the cleaned and checked data into Power BI.
2. Design how you want the dashboard to look, picking the right kinds of pictures and words to show what you found.
3. Make sure people can play with the dashboard to learn more about the data.

* **Fixing and Checking Again:** Look over the dashboard to make sure it's easy to understand and doesn't have any mistakes. Fix anything that needs fixing.

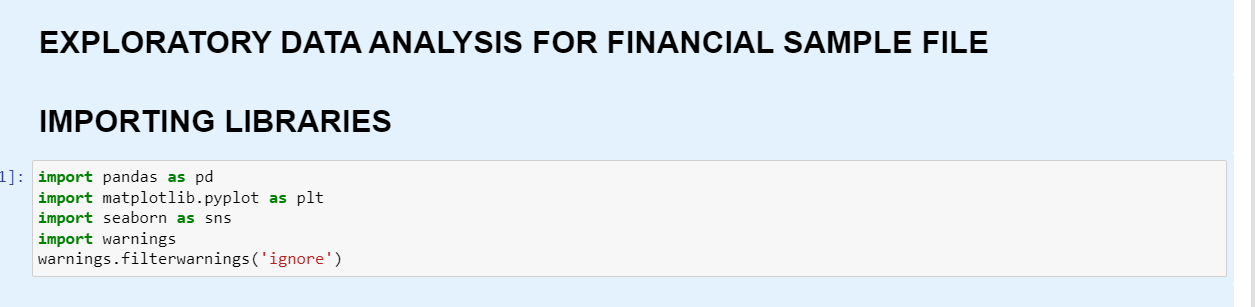
This methodology outlines the steps involved in using a CSV file to create a Power BI dashboard and perform exploratory data analysis in a Python notebook, using simple language and easy-to-understand steps.

# **EXPLORATORY DATA ANALYSIS**

Exploratory Data Analysis (EDA) is a crucial step in understanding and preparing data for further analysis. In this report, we'll walk through the steps of EDA performed on a financial dataset.

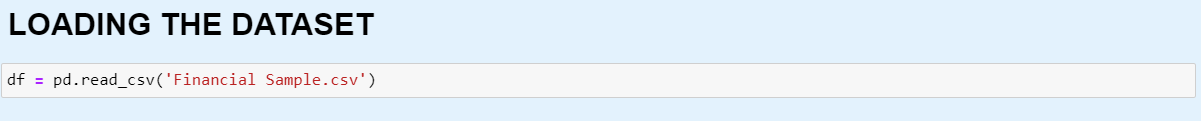
**Step 1: Importing Libraries**

We begin by importing necessary libraries such as Pandas for data handling, Matplotlib and Seaborn for data visualization, and warnings to manage any potential warnings.



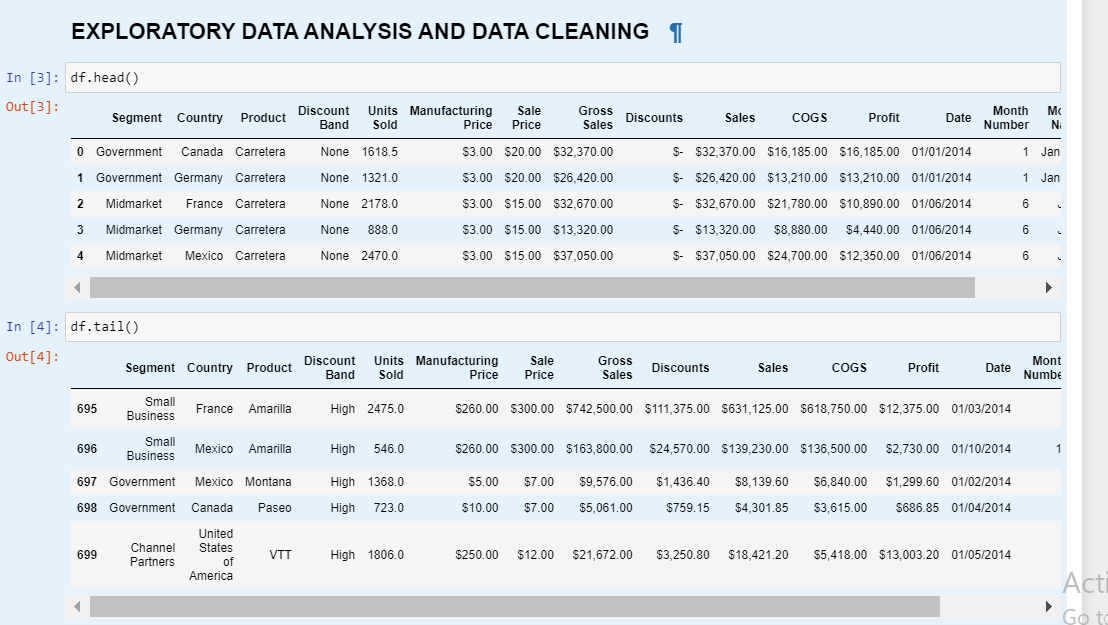
**Step 2: Loading the Dataset**

We load the financial dataset from a CSV file into a Pandas DataFrame called 'df'.

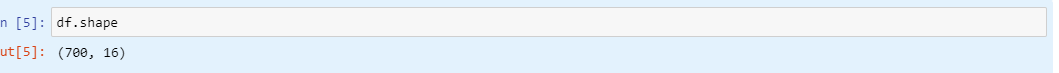


**Step 3: Exploratory Data Analysis and Data Cleaning**

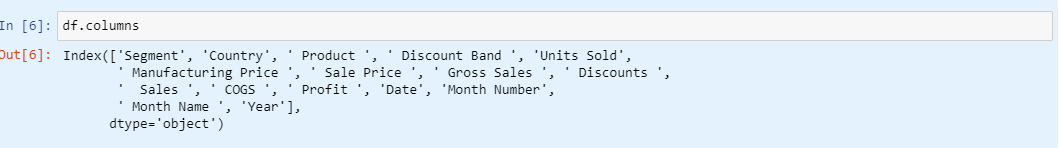
1. **Initial Data Overview:** We start by examining the first and last few rows of the dataset to get a glimpse of its structure.



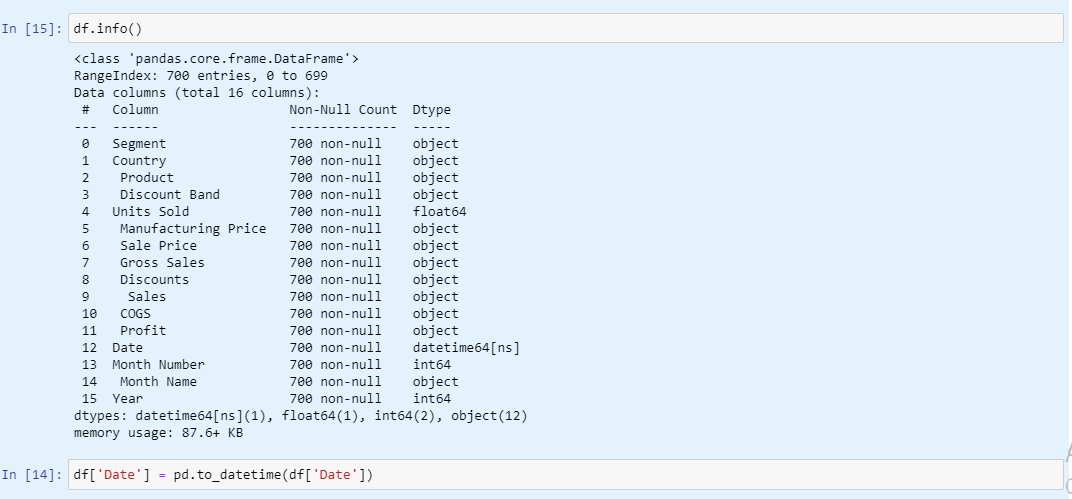
1. **Dataset Dimensions:** We check the number of rows and columns in the dataset using 'df.shape'.



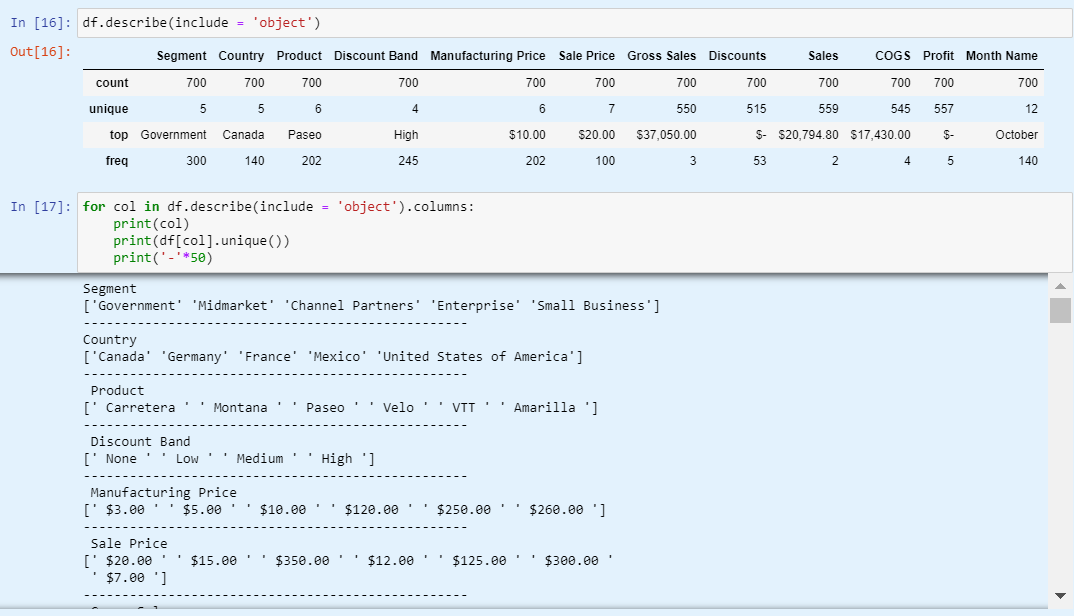
1. **Column Names:** We list all the column names in the dataset using 'df.columns'.



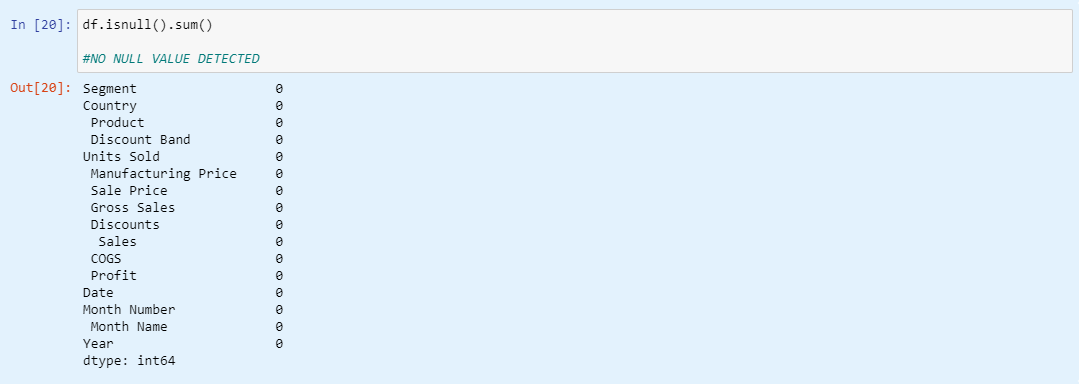
1. **Data Types and Missing Values:** We inspect the data types of each column and check for any missing values using 'df.info()'. The 'Date' column is converted to a datetime format for better handling.



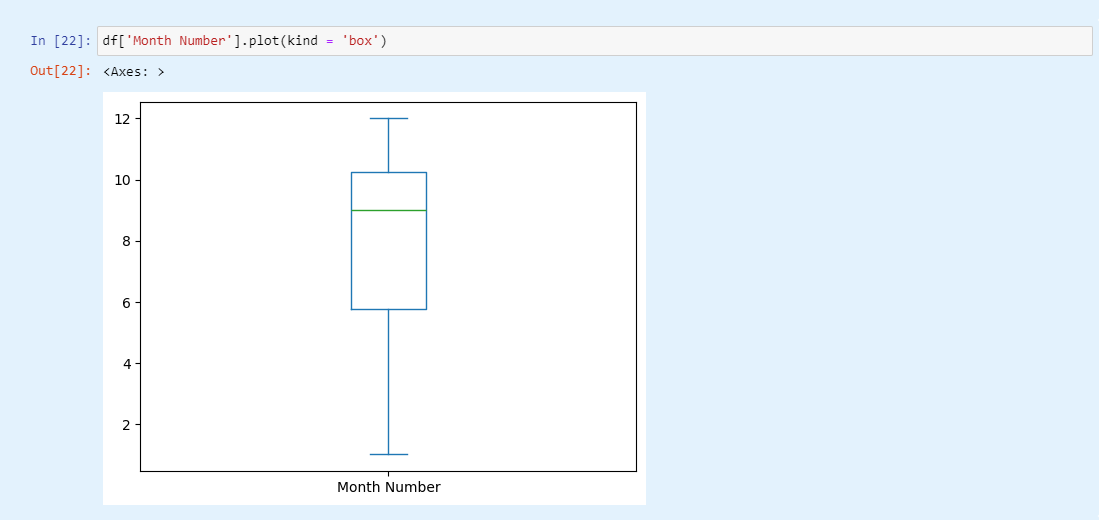
1. **Descriptive Statistics:** We compute descriptive statistics for numerical and categorical columns using 'df.describe()' and 'df.describe(include='object')', respectively.
2. **Unique Values:** For each categorical column, we list unique values to understand the variety of data in those columns.



1. **Missing Values Check:** We verify if there are any missing values in the dataset. Fortunately, no missing values are detected.



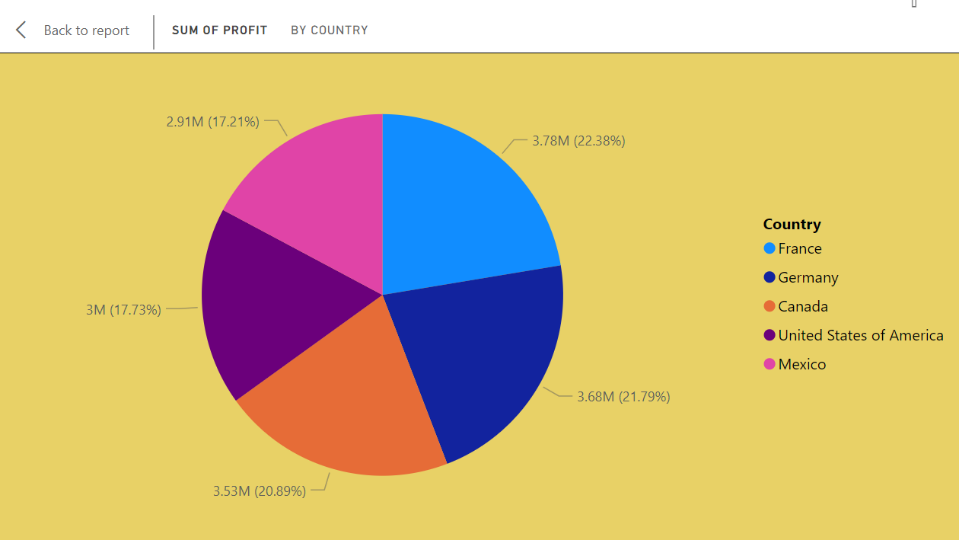
1. **Box Plot:** We visualize the distribution of values in the 'Month Number' column using a box plot to identify any outliers.



In conclusion, this exploratory data analysis provided valuable insights into the financial dataset's structure and characteristics. It involved examining the data, cleaning it for further analysis, and conducting basic statistical analysis to better understand the data's features.

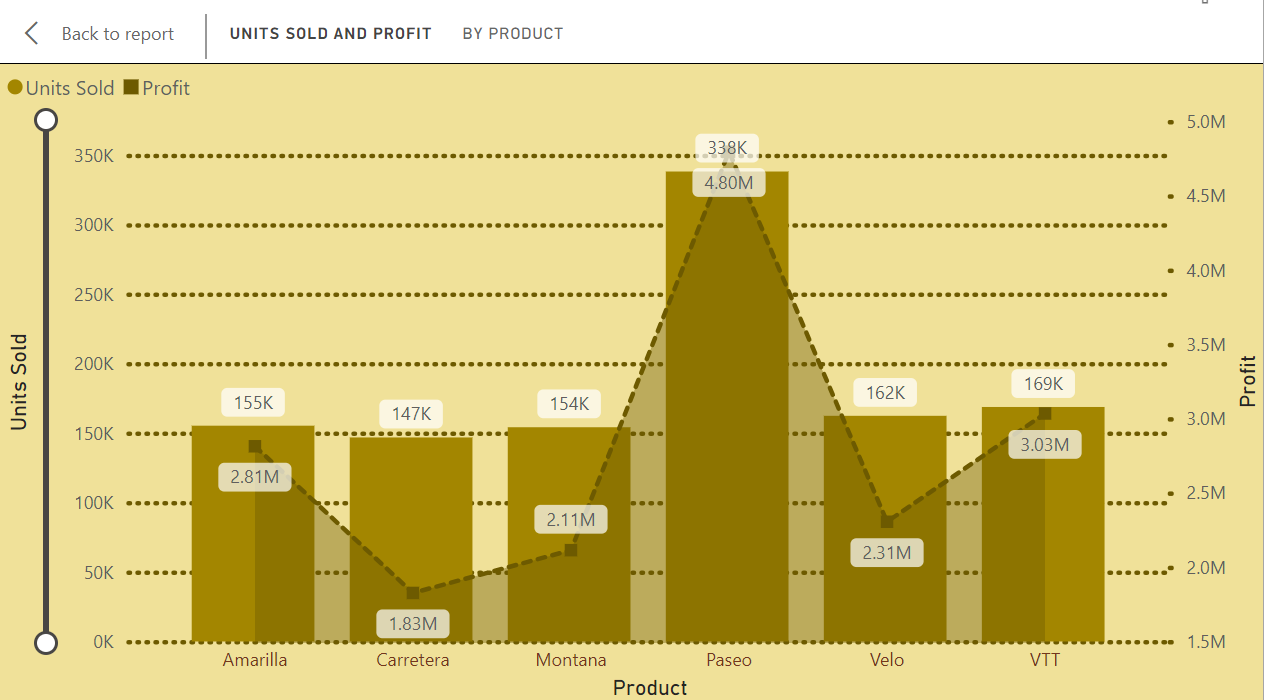
# **DATA VISUALIZATION AND DASHBOARD using POWER BI**

## **PIE CHART**

  
﻿ ﻿

At 338,239.50, Paseo had the highest Units Sold and was 130.34% higher than Carretera, which had the lowest Units Sold at 146846.﻿﻿ Units Sold and total Profit are positively correlated with each other.﻿﻿ Paseo accounted for 30.04% of Units Sold.﻿﻿ Across all 12 Month, Gross Profit ranged from 669,866.87 to 3,439,781.02.﻿﻿

## **LINE AND CLUSTURED COLUMN CHART**



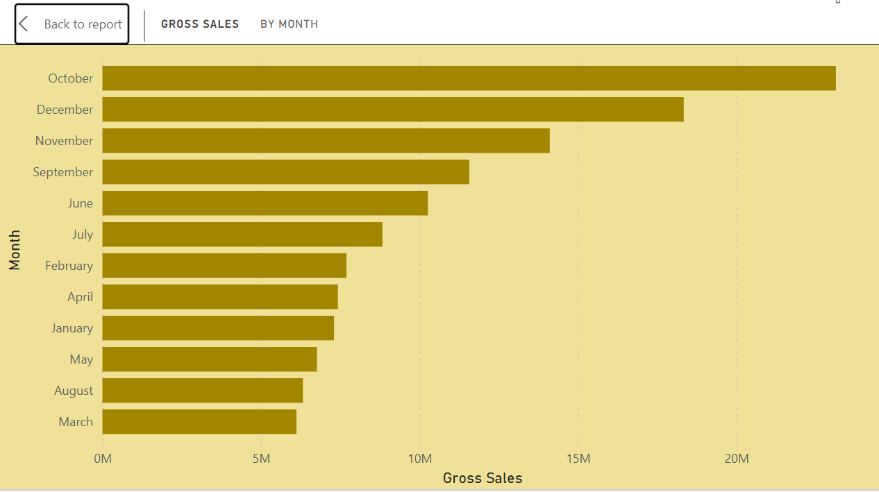
At a staggering 338,239.50 units sold, Paseo dominated the sales figures, surpassing Carretera by a significant margin of 130.34%, with the latter recording the lowest units sold at 146,846.

There exists a positive correlation between units sold and total profit.

Paseo constituted 30.04% of the total units sold.

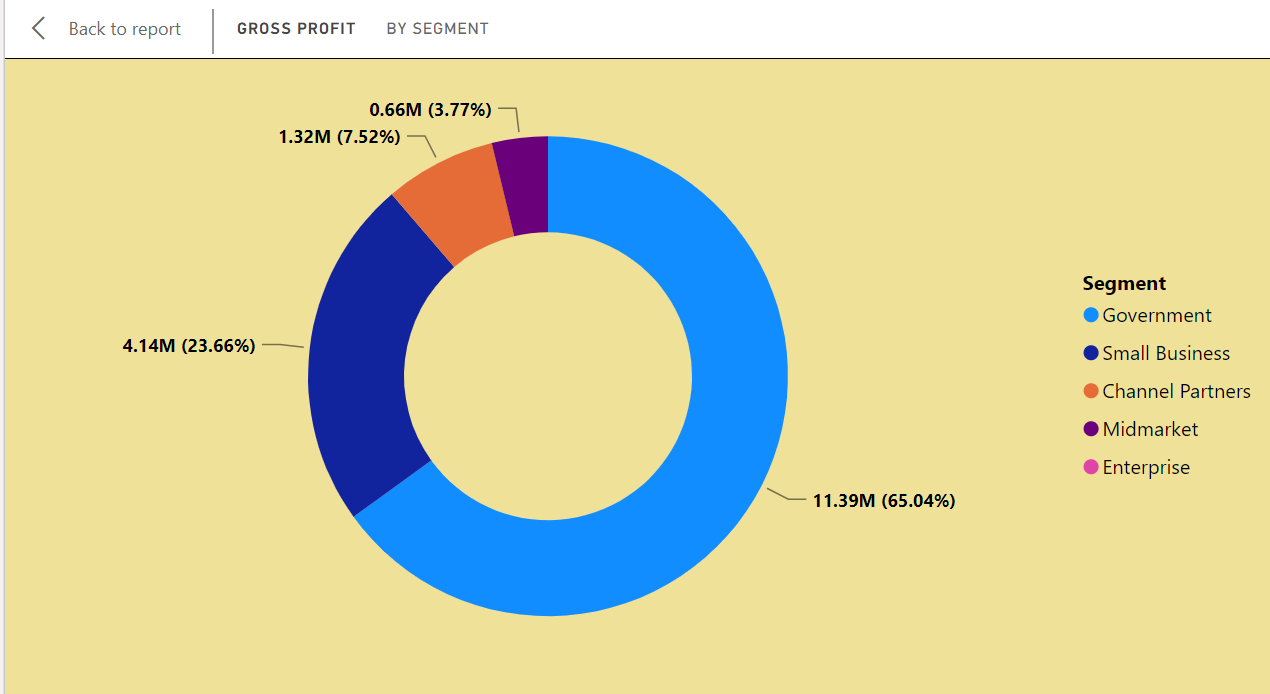
Across all 12 months, the gross profit varied substantially, ranging from 669,866.87 to 3,439,781.02.

## **STACKED BAR CHART**



1. The graph represents gross sales data by month.
2. The months are listed on the Y-axis, and the gross sales amounts are on the X-axis.
3. Sales data is shown for each month from March to October.
4. December has the highest gross sales, followed by October and November.
5. March has the lowest gross sales.

﻿﻿**DONUT CHART**



﻿ The chart represents gross profit data by segment.

There are five segments, each represented by a different color:

Government (purple)

Small Business (red)

Channel Partners (orange)

Midmarket (light blue)

Enterprise (dark blue)

The gross profit amounts and percentages for each segment are as follows:

Government: No visible numerical data or percentage provided.

Small Business: 1.32M (7.52%)

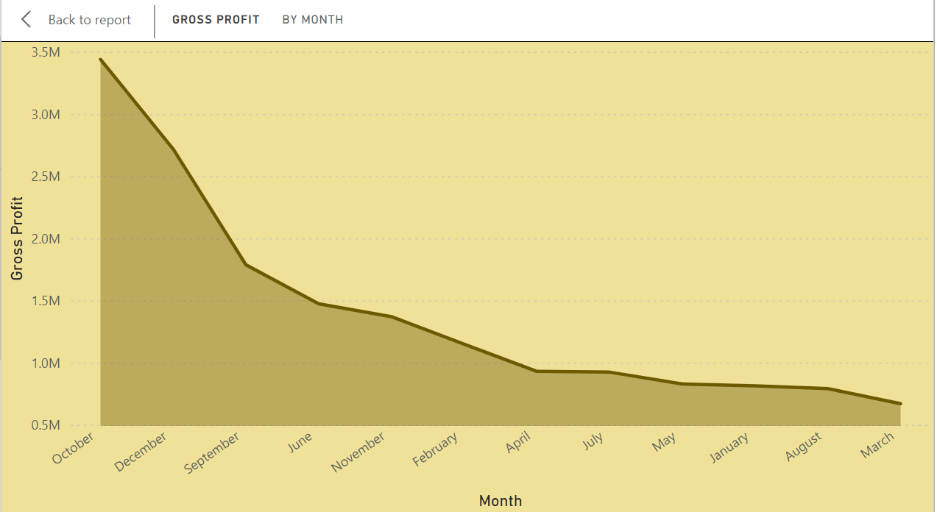
Channel Partners: 0.66M (3.77%)

Midmarket: 4.14M (23.66%)

Enterprise: 11.39M (65.04%)

The Enterprise segment occupies the largest portion of the chart.

## **STACKED AREA CHART**



The Y-axis represents “Gross Profit” and ranges from 0 to 3.5 million.

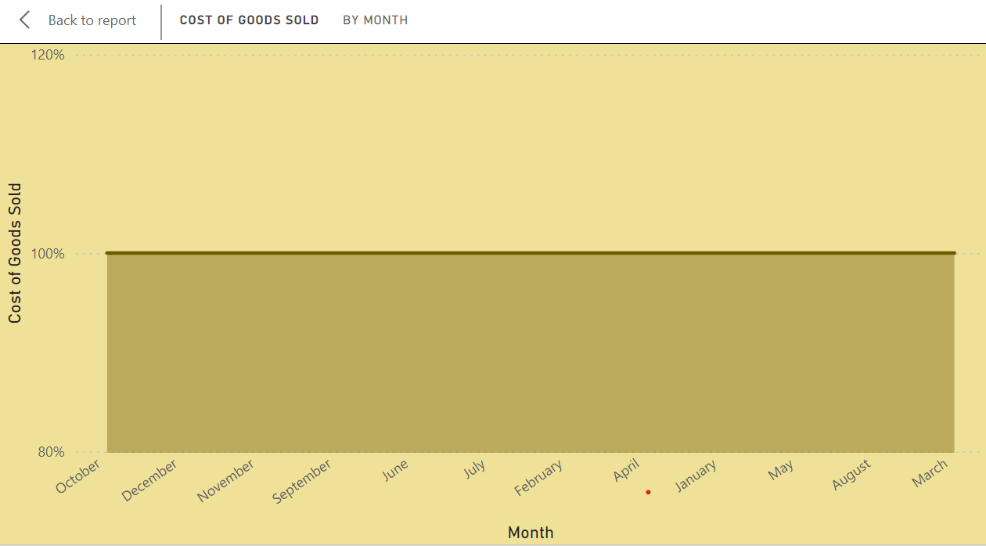
The X-axis corresponds to months, starting from October and ending in March.

Notably, there’s a significant peak in gross profit in October, reaching approximately 3 million.

After October, there’s a sharp decline in November, and the gross profit remains relatively flat for the subsequent months.

The shaded area under the line graph visually represents the total gross profit each month.

## **100% STACKED AREA CHART**



The graph shows the cost of goods sold by month.

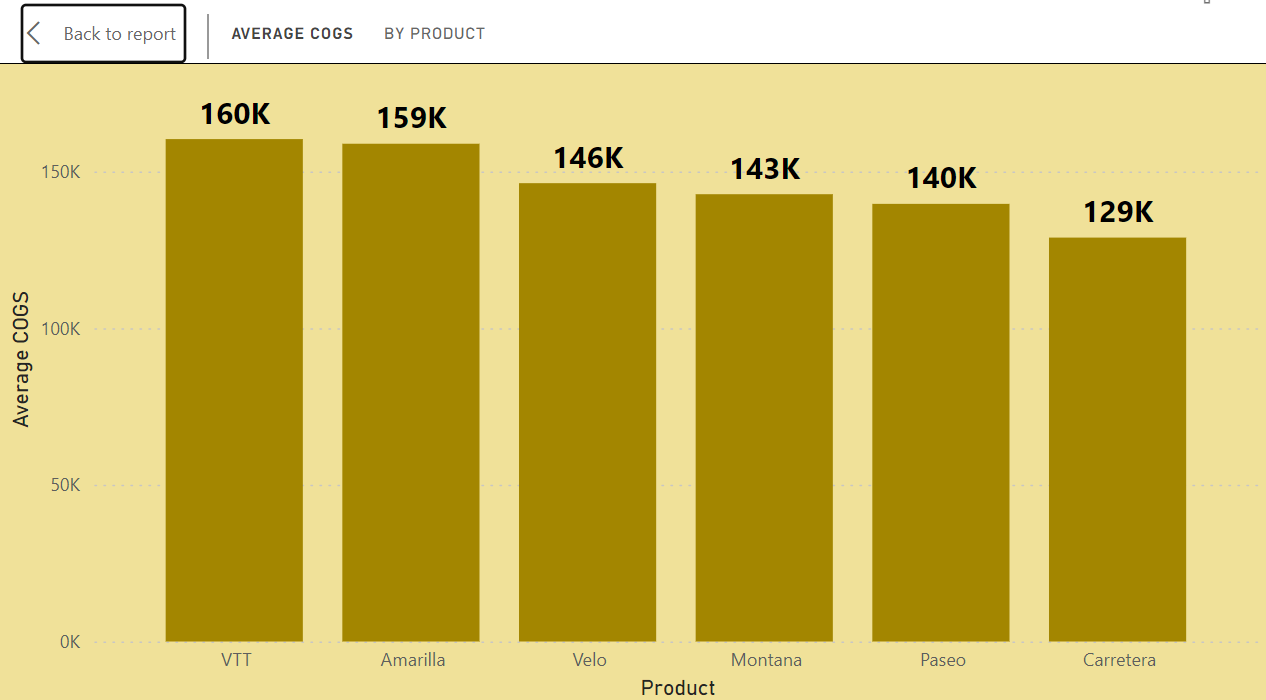
All months have the same value, which is unusual and might indicate an error or placeholder data.

The Y-axis represents the “Cost of Goods Sold,” ranging from 80% to 120%.

The X-axis lists the months of the year, but they are not in chronological order.

Each month is represented by a green bar reaching up to 100% on the Y-axis.

## **STACKED COLUMN CHART**



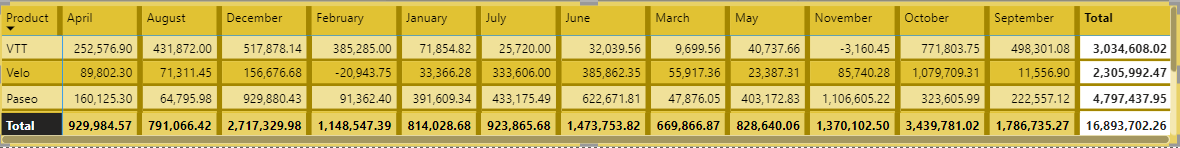
The graph displays six vertical bars, each corresponding to a different product.

The products are labeled at the bottom of each bar: “VTT,” “Amarilla,” “Velo,” “Montana,” “Paseo,” and “Carretera.”

The average COGS values (in ‘K’ units) for these products are as follows:

VTT: 160K, Amarilla: 159K, Velo: 146K, Montana: 143K, Paseo: 140K, Carretera: 129K

## **MATRIX**



* **Products and Months:**

The spreadsheet includes three products: “VTT,” “Velo,” and “Paseo.”

The months are listed from April to September, with additional months like December, February, January, July, June, March, May, November, and October.

* **Sales Figures:**

Each product has its own row, and the corresponding sales figures are listed under each month.

Notably, there are negative numbers in red font, indicating losses or returns for certain months (specifically for the “Velo” product).

The background color of the cells alternates between yellow and green to distinguish rows.

* **Total Sales:**

At the bottom of the spreadsheet, there’s a total row that sums up all sales figures for each month.

The overall total is provided at the far right end.

* **Product Sales:**

1. **VTT (Vehicle Tracking Technology):**

In April, VTT sales were approximately $252,576.90.

The highest sales occurred in August, reaching around $43,187,200.

December saw significant sales at $51,787,814.

Overall, VTT contributed significantly to the total revenue.

1. **Velo (Bicycle Brand):**

Velo had varying performance:

April: Sales of $8,980,230.

August: Negative sales (returns) of $7,131,145.

December: A substantial loss of $156,767,668.

February: Another negative month with sales at -$2,094,375.

Overall, Velo faced challenges, especially in December.

1. **Paseo (Product/Service):**

Paseo experienced significant losses:

April: -$16,012,530.

August: -$6,479,958.

December: -$92,988,043.

February: -$9,136,240.

Overall, Paseo struggled throughout the year.

1. **Montana:**

In April, Montana sales were approximately $337,205.50.

August saw sales of $162,685.90.

December: $78,177.82.

Overall, Montana contributed significantly to the total revenue.

1. **Carretera:**

Carretera faced challenges:

April: Negative sales (returns) of -$24,018.

August: More negative sales at -$46,531.

December: A substantial loss of -$636,525.

Overall, Carretera struggled throughout the year.

1. **Amarilla:**

Amarilla’s data is incomplete in the image, but we can see some values:

June: -$13440.

July: -$61640.

March: -$19187.

Overall, Amarilla also faced difficulties.

* **Monthly Totals:**

The total sales for each month are summarized at the bottom of the spreadsheet.

For example, in September, the total sales across all products amounted to $3,034,608,082.

* **Overall Performance:**

The total revenue for all products combined is shown in the last column.

Unfortunately, the overall total is negative: -$16,893,702,226.

This suggests that the company faced significant challenges during this period.

# **SUGGESTIONS/RECOMMENDATIONS**

1. **Sales Strategy Adjustment:**

Identify and focus on products with consistent positive sales performance, such as VTT and Montana.

Explore ways to improve sales for struggling products like Velo, Paseo, and Carretera, possibly through marketing campaigns or product enhancements.

1. **Cost Management:**

Investigate and address the reasons behind high costs and losses, especially for products like Velo, Paseo, and Carretera.

Implement cost-saving measures where possible to improve overall profitability.

1. **Inventory Management:**

Optimize inventory levels to avoid overstocking of products that are facing challenges, which can lead to further losses.

Ensure adequate stock availability for consistently performing products to meet customer demand.

1. **Market Analysis:**

Conduct market research to understand changing consumer preferences and market trends, which can inform product development and marketing strategies.

1. **Customer Feedback:**

Gather and analyze customer feedback to identify areas for improvement in product quality, customer service, and overall customer experience.

1. **Financial Planning:**

Develop a robust financial plan that accounts for fluctuations in sales and expenses, ensuring the company remains financially resilient in challenging times.

1. **Continuous Monitoring and Adaptation:**

Regularly monitor sales and financial performance metrics to track progress and identify emerging trends or issues.

Be flexible and ready to adapt strategies based on market dynamics and evolving customer needs.

**Implementing these suggestions can help the company navigate challenges, optimize performance, and drive sustainable growth in the long run.**