

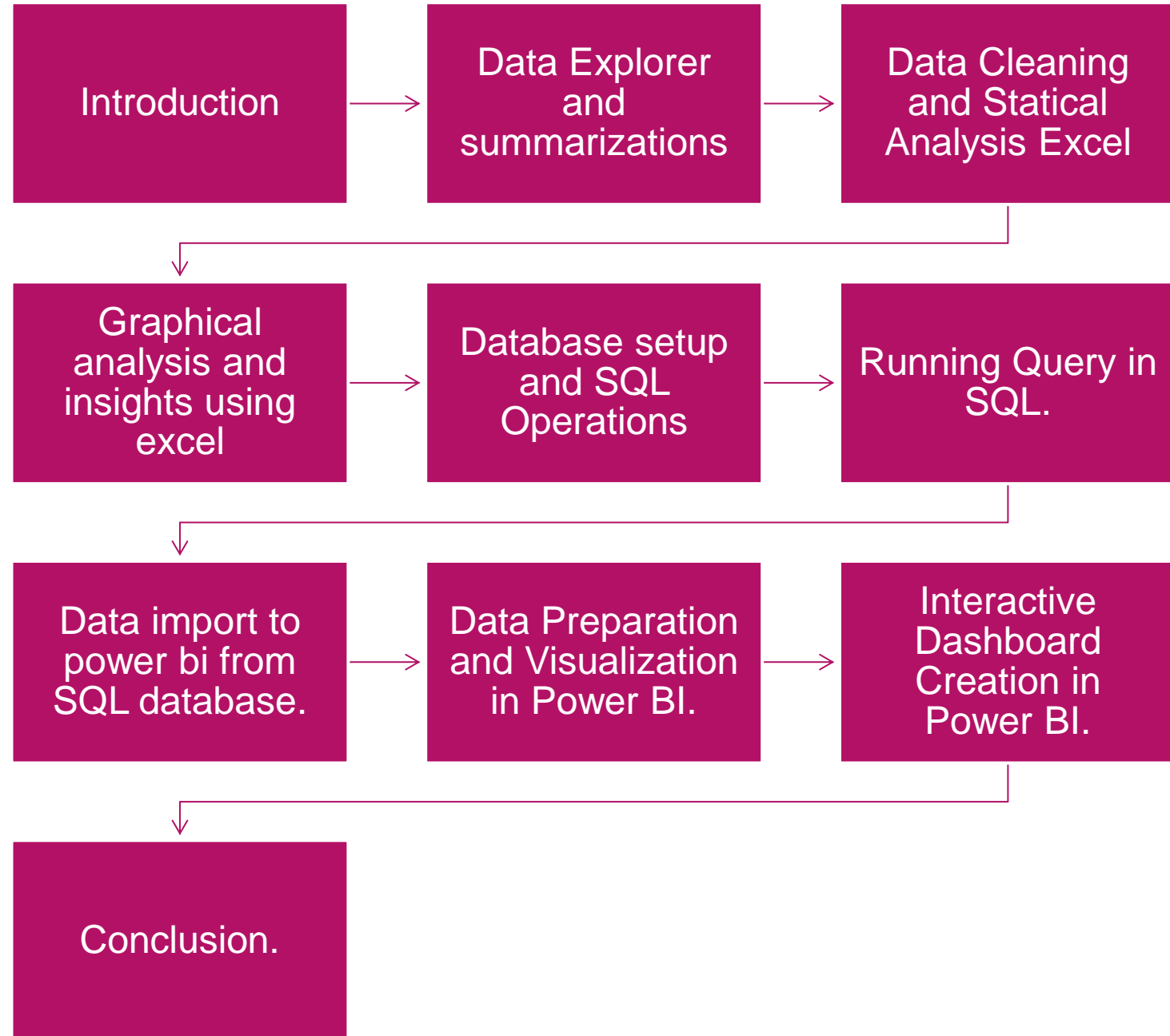


# Capstone Project

WESTERN COUNTRIES FINANCIAL DATA

Presented By : Sunil Kumar Prusty

# CONTENTS



# Introduction

## Brief Overview of the Data Lifecycle

- Begins with data collection from various sources.
- Followed by cleaning and preprocessing to ensure accuracy.
- Analysis and visualization help uncover patterns and insights.
- Final step involves reporting and decision-making based on findings.

## Importance of Data Cleaning, Analysis, and Visualization

- Ensures data integrity and reliability.
- Helps identify trends, outliers, and performance metrics.
- Supports strategic decisions with clear, visual evidence.

## Tools Used

- **Excel:** For initial data exploration, cleaning, and basic analysis.
- **SQL:** For structured data storage and advanced querying.
- **Power BI:** For dynamic visualizations and interactive dashboards.

# DATA EXPLORATION AND SUMMARIZATION

## Data Summary

- Total Variables: 16
- Total Observations: 700
- Missing Values:
  - Discounts: 53 Missing Values
  - Profit: 5 Missing Values
- Duplicate Rows: 0 (No Duplicates)
- Column Types:
  - Categorical Columns:  
Segment, Country, Product, Discount Band, Month Name
  - Numerical Columns: 10
  - Date-Time Column: 1

## Variable Overview

- **Categorical Variables:**
  - Segment (5 unique values), Country (5), Product (6), Discount Band (4).
- **Numerical Variables:**
  - Units Sold(510), Manufacturing Price(6), Sale Price (7), Gross Sales(550).
  - Discounts (515) → 53 missing values.
  - Sales (559), COGS(545), Profit (557) → 5 missing values in Profit.
- **Date-Time Variables:**
  - Date (16 unique values), Month Number (12), Month Name (12), Year (2).

# Data Cleaning Steps

- 1.Removed Duplicates:** All duplicate rows were eliminated.
- 2.Date Conversion:** Columns with date-like values were converted to proper datetime format.
- 3.Missing Values:**
  - 1. Numerical columns:** Filled with the median value.
  - 2. Categorical columns:** Filled with the most frequent (mode) value.
- 4.Data Type Correction:** Ensured all numerical columns are correctly typed.

# STATISTICAL ANALYSIS IN EXCEL

Stataical Value	Units_Sold	Discounts	Sales	Cogs	Profit
Total	↓ 1125806	↓ 9205248.24	↓ 118726350.3	↓ 101832648	↓ 16893702.26
MEAN	↓ 1608.29	↓ 13150.35	↓ 169609.07	↓ 145475.21	↓ 24133.86
MODE	↓ 727.00	↓ 0.00	↓ 32670.00	↓ 17430.00	↓ 0.00
MEDIAN	↓ 1542.50	↓ 2585.25	↓ 35540.20	↓ 22506.25	↓ 9242.20
Standard Deviation(Simple)	↓ 867.43	↓ 22962.93	↓ 236726.35	↓ 203865.51	↓ 42760.63
Variance(simple)	↓ 752431.09	↓ 527296097.92	↑ 56039363321.24	↑ 41561144584.94	↓ 1828471184.09
Standard Deviation(Population)	↓ 866.81	↓ 22946.52	↓ 236557.20	↓ 203719.84	↓ 42730.07
Variance(Population)	↓ 751356.19	↓ 526542817.78	↑ 55959307087.92	↑ 41501771521.25	↓ 1825859082.40
Maximum Value	↓ 4492.50	↓ 149677.50	↓ 1159200.00	↓ 950625.00	↓ 262200.00
Minimum Value	↓ 200.00	↓ 0.00	↓ 1655.08	↓ 918.00	↓ -40617.50
Second Largest Number	↓ 4251.00	↓ 125820.00	↓ 1038082.50	↓ 948375.00	↓ 247500.00
Range	↓ 4292.50	↓ 149677.50	↓ 1157544.92	↓ 949707.00	↓ 302817.50
First Quartile	↓ 905.00	↓ 798.96	↓ 15928.00	↓ 7490.00	↓ 2803.48
Second Quartile	↓ 1542.50	↓ 2585.25	↓ 35540.20	↓ 22506.25	↓ 9242.20

# GRAPHICAL ANALYSIS AND INSIGHTS USING EXCEL



## Western Countries Financial

Year

2013

2014

Segment

Channel Partners

Enterprise

Government



**Total Sales**  
118.73M



**Total Unit Solds**  
1125806



**Total COGS**  
101.83M



**Total Profit**  
16.89M



**Total Discount**  
9.21M

Mo...

January

Febru...

March

April

May

June

July

August

Septe...

Octo...

Nove...

Co...

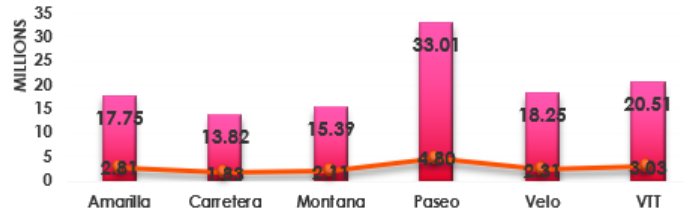
Can...

France

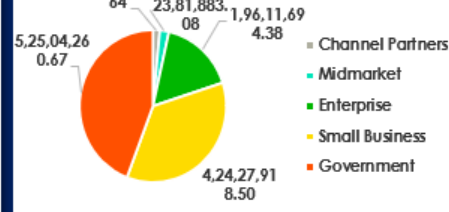
Germ...

Mexico

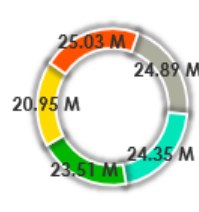
**Total Sales and Total Profit by Product**



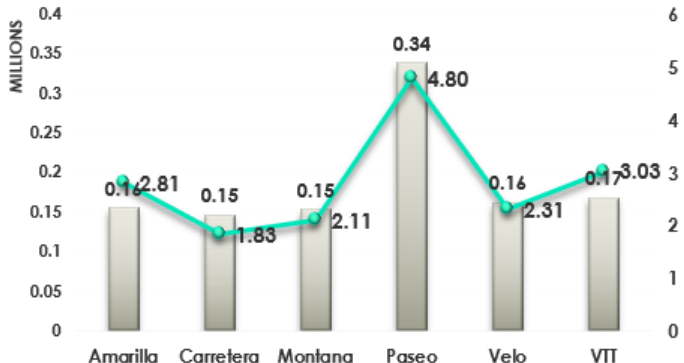
**Total Sales By Segemnt**



**Total Sales By Country**



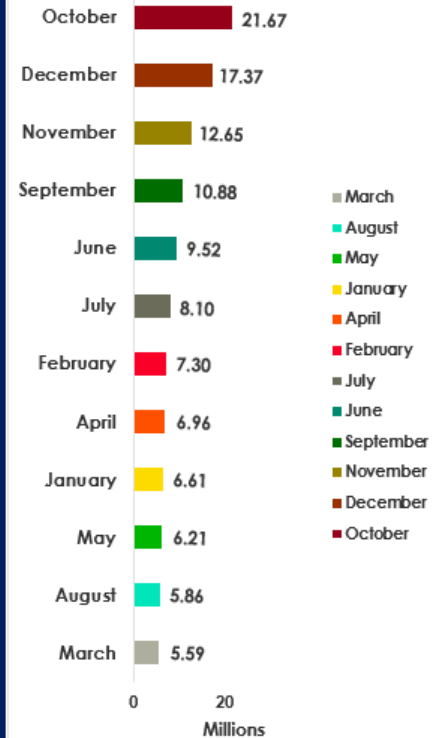
**Total Unit Sold vs Total Profit by Product**



**Total COGS BY Total Profit by Month**



**Total Sales By Month**

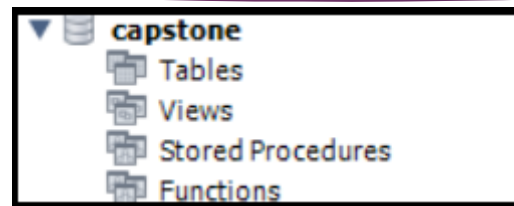




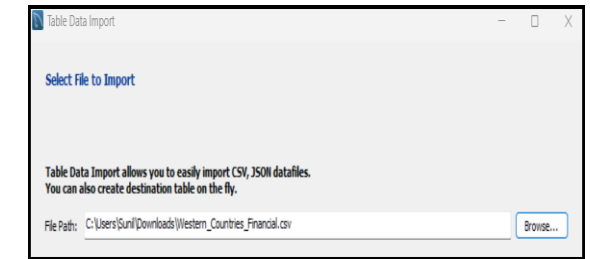
# DATABASE SETUP AND SQL OPERATIONS

- 1 • `CREATE DATABASE CAPSTONE;`
- 2 • `USE CAPSTONE;`

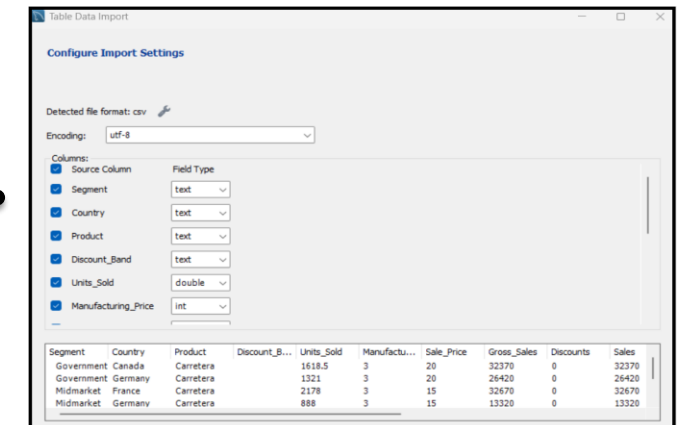
Database Creation



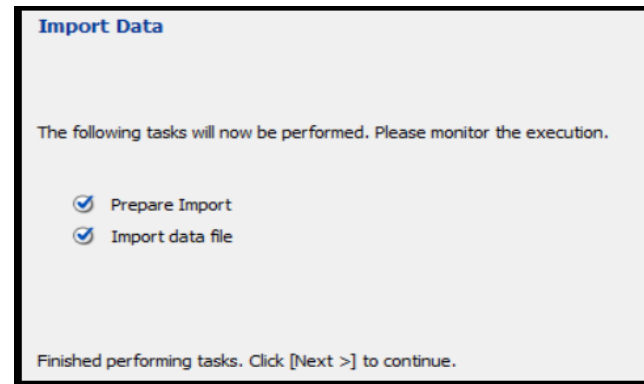
Database Created



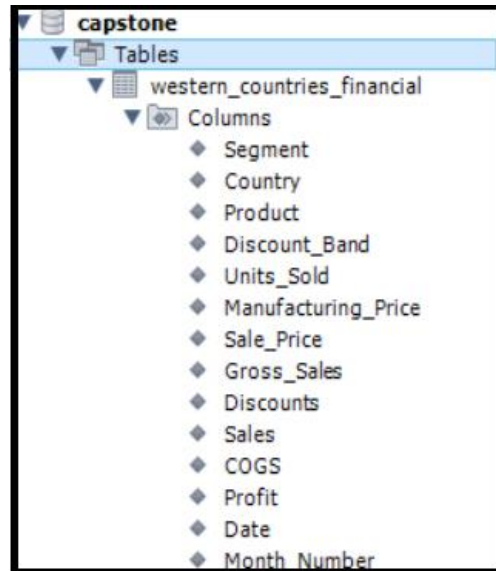
Importing Dataset



Datatype Edition



Dataset Loaded



Showing Database



# RUNNING QUERIES IN SQL

## Input Query

```
SELECT * FROM WESTERN_COUNTRIES_FINANCIAL;
```

```
# Max Profit by Product
SELECT PRODUCT, MAX(PROFIT) AS Max_Profit
FROM WESTERN_COUNTRIES_FINANCIAL
GROUP BY PRODUCT;
```

```
#TOP 5 SALES
SELECT PRODUCT, SALES
FROM WESTERN_COUNTRIES_FINANCIAL
ORDER BY SALES DESC LIMIT 5;
```

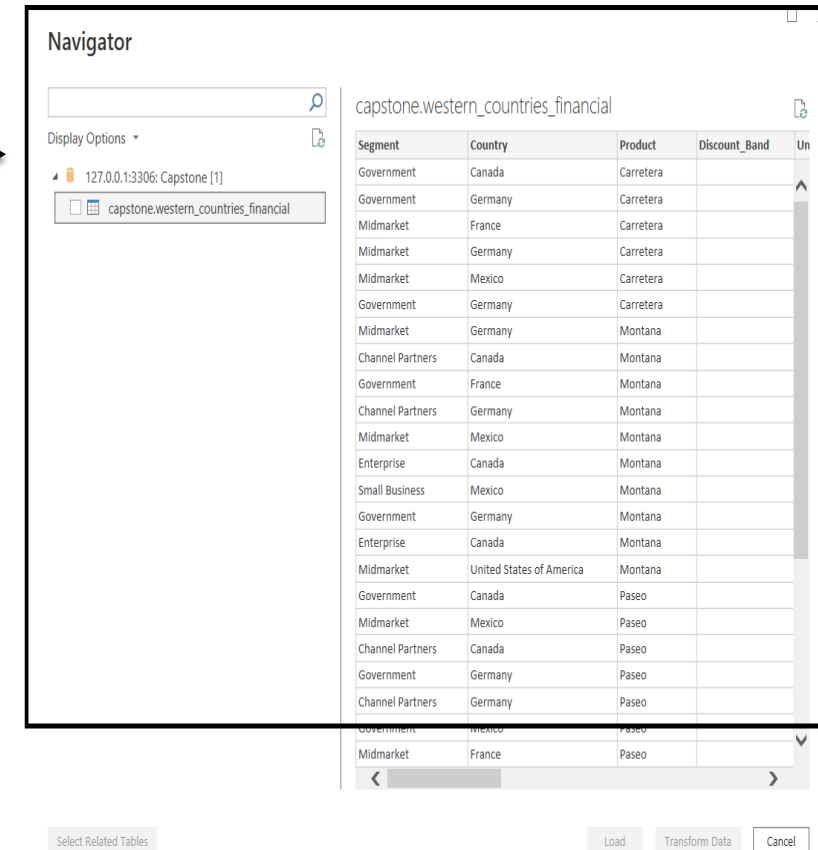
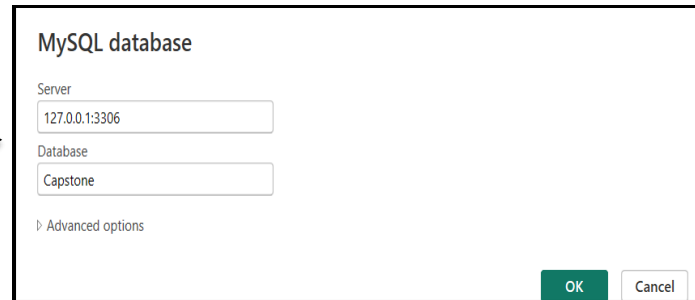
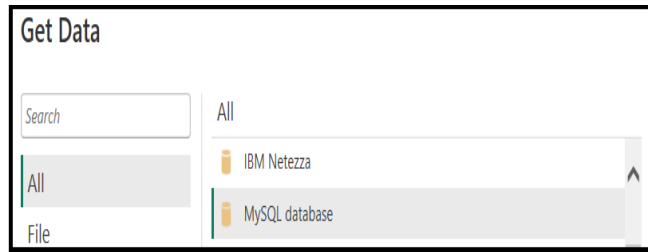
## Output Query

Segment	Country	Product	Discount_Band	Units_Sold	Manufacturing_Price	Sale_Price	Gross_Sales	Discounts	Sales	COGS	Profit	Date
Government	Canada	Carretera		1618.5	3	20	32370	0	32370	16185	16185	01-01-2011
Government	Germany	Carretera		1321	3	20	26420	0	26420	13210	13210	01-01-2011
Midmarket	France	Carretera		2178	3	15	32670	0	32670	21780	10890	01-06-2011
Midmarket	Germany	Carretera		888	3	15	13320	0	13320	8880	4440	01-06-2011
Midmarket	Mexico	Carretera		2470	3	15	37050	0	37050	24700	12350	01-06-2011
Government	Germany	Carretera		1513	3	350	529550	0	529550	393380	136170	01-12-2011
Midmarket	Germany	Montana		921	5	15	13815	0	13815	9210	4605	01-03-2011
Channel Partners	Canada	Montana		2518	5	12	30216	0	30216	7554	22662	01-06-2011
Government	France	Montana		1899	5	20	37980	0	37980	18990	18990	01-06-2011
Channel Partners	Germany	Montana		1545	5	12	18540	0	18540	4635	13905	01-06-2011
Midmarket	Mexico	Montana		2470	5	15	37050	0	37050	24700	12350	01-06-2011
Enterprise	Canada	Montana		2665.5	5	125	333188	0	333188	319860	13328	01-07-2011
Small Business	Mexico	Montana		958	5	300	287400	0	287400	239500	47900	01-08-2011
Government	Germany	Montana		2146	5	7	15022	0	15022	10730	4292	01-09-2011
Enterprise	Canada	Montana		345	5	125	43125	0	43125	41400	1725	01-10-2011
Midmarket	United S...	Montana		615	5	15	9225	0	9225	6150	3075	01-12-2011
Government	Canada	Paseo		292	10	20	5840	0	5840	2920	2920	01-02-2011

	PRODUCT	Max_Profit
▶	Carretera	236716
	Montana	142862
	Paseo	262200
	Velo	246178
	VTT	238791
	Amarilla	247500

	PRODUCT	SALES
▶	Paseo	1159200
	Montana	1038083
	Velo	1035626
	Amarilla	1017338
	Velo	1017338

# DATA IMPORT TO POWER BI FROM SQL DATABASE



1. Open Power BI Desktop
2. Click on **Get Data** → choose **More**
2. Under **Database**, select **MySQL database**
3. Enter:
  - **Server name** (127.0.0.1:3306)
  - **Database name** (Capstone)
4. Choose **Authentication type**:  
Typically, **Database authentication** with username(root) and password
5. Click **OK** and select the tables or views you want to use
6. Load or transform the data using Power BI's interface

# Dashboard

## Western Countries Financial Dashboard

Total Sales

119M

Total Unit Solds

1.13M

Total Profit

17M

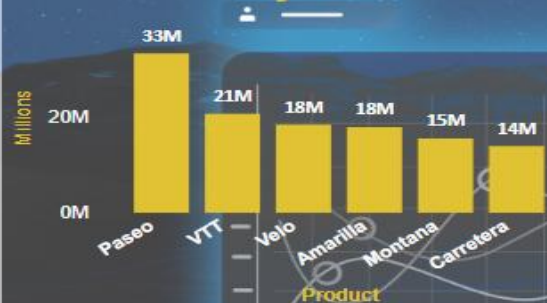
Total Discount

9M

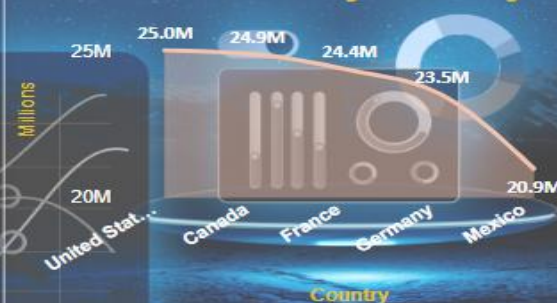
Total Profit by Year



Sales by Product



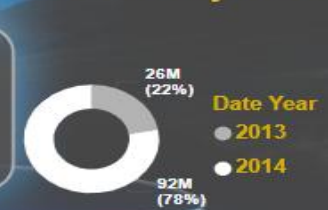
Sum of Sales by Country



Product Wise Discount



Total Sales by Year



Profit by Segment



Profit and Sales by Quarter



Segment Wise Product Sales



Units Sold by Year



# INSIGHTS AND CONCLUSIONS

- **Top-Selling Product:** Paseo recorded the highest sales of \$33,011,143.96.
- **Top Sales Segment:** The Government Segment counts for the highest number of sales.
- **Top Sales Country:** USA has the highest number of sales.
- **Profitability:** Paseo yields more profit due to its high sales volume.
- **Segment Performance:** Countries with high sales in the Government Segment also generate higher profits.
- **Top Profit-Contributing Countries:** Francena Germany account for the highest profits.
- **Best Performing Quarter:** Q4 recorded the highest sales and profit.
- **Business Implication:** Focus on high-performing segments and countries while addressing losses in the Enterprise sector.
- **Loss Area:** The Enterprise Sector is experiencing losses.



# Thank You