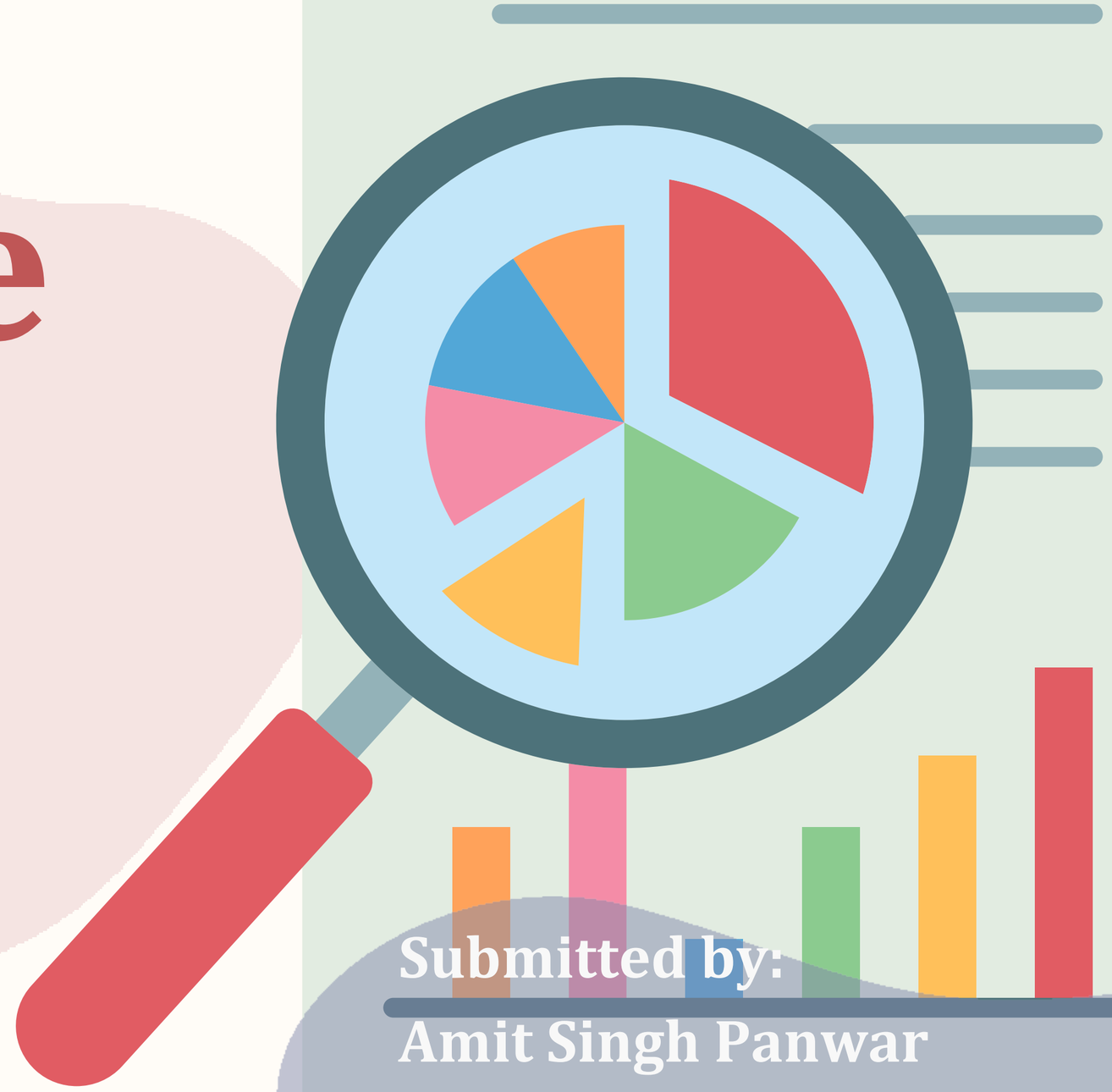
Capstone Project

Business Analysis using SQL & Power BI



Submitted by:
Amit Singh Panwar

Project Overview

This capstone project focuses on analyzing sales and profitability trends using **Power BI** and **SQL** to derive meaningful insights. The **goal** is to identify **key patterns** that can drive business decisions.

Dataset Summary:

- **Total Records: 700**
- **Total Columns: 16**

o **Categorical Columns (5):** Segment, Country, Product, Discount Band, Year
o **Numerical Columns (10):** Sales, Profit, Discounts, Units Sold, etc.
o **Date-Time Column (1):** Order Date (16 unique values)

- **Geographical Scope:** Western countries (Canada, Germany, France, Mexico, USA)
- **Key Metrics:** Sales, Profit, Discounts, Cost of Goods Sold (COGS), and Segment-based performance
- **Timeframe:** 2013 (last 4 months) & 2014 (full year)

Actionable Steps

A. Data Cleaning Process

- Checked for missing values: Minimal missing values found, handled appropriately.
- Removed duplicates: No duplicate records detected.
- Filtered outliers: Identified extreme outliers in Discount and Profit that could distort insights.

B. Data Type Conversion

- Converted date fields to datetime format for accurate time-based analysis.
- Changed numerical columns (Sales, Profit, Discount) to proper float types.
- Categorical variables (Segment, Country) converted to category dtype for optimized performance.

C. Statistical Summary

Statistical Summary of Western Countries Financial Data									
Descriptive Statistics	Measures	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit
Central Tendency	Mean	1608.29429	96.47714286	118.4285714	182759.4264	13150.35463	169609.0718	145475.2114	24133.8604
	Median	1542.5	10	20	37980	2585.25	35540.2	22506.25	9242.2
	Mode	727	10	20	37050	0	32670	17430	0
Dispersion	Minimum	200	3	7	1799	0	1655.08	918	-40617.5
	Maximum	4492.5	260	350	1207500	149677.5	1159200	950625	262200
	Range	4292.5	257	343	1205701	149677.5	1157544.92	949707	302817.5
	Standard Deviation	867.427859	108.6026122	136.7755146	254262.2844	22962.92877	236726.3469	203865.5061	42760.6266
	Sample Variance	752431.091	11794.52737	18707.54139	64649309257	527296097.9	56039363321	41561144585	1828471184
	Kurtosis	-0.315318	-1.42896268	-1.176789008	2.054300583	7.905712444	2.188633088	1.608462973	8.67861622
	Skewness	0.43615356	0.592583952	0.771281871	1.673921656	2.685038938	1.696295217	1.549047562	2.71215126
Sum		1125806	67534	82900	127931598.5	9205248.24	118726350.3	101832648	16893702.3
Standard Error		32.7856914	4.104792908	5.169628528	9610.211032	867.9171273	8947.414896	7705.391858	1616.19977
	Profit and Sales Correlation			Discount & Profit Correlation					

Dashboard Creation in Excel

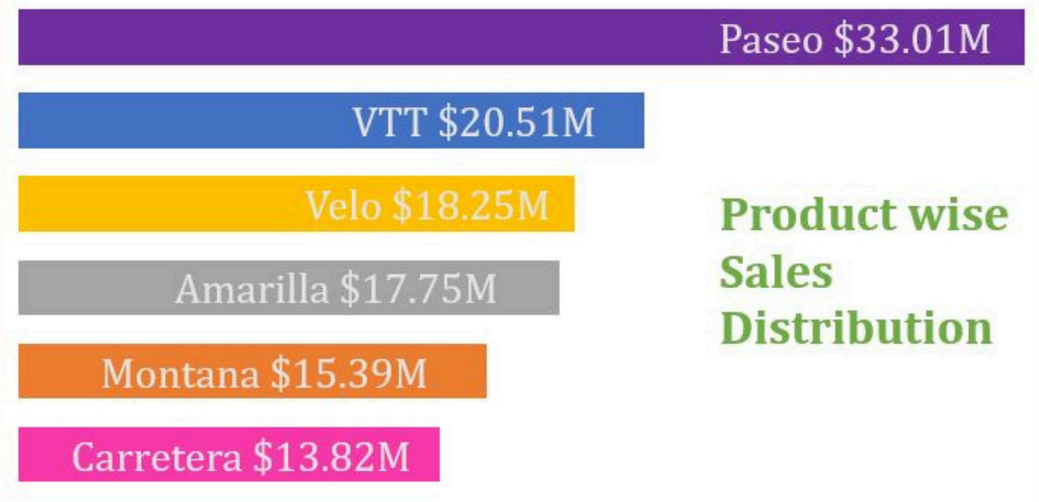
Dashboards Designed:

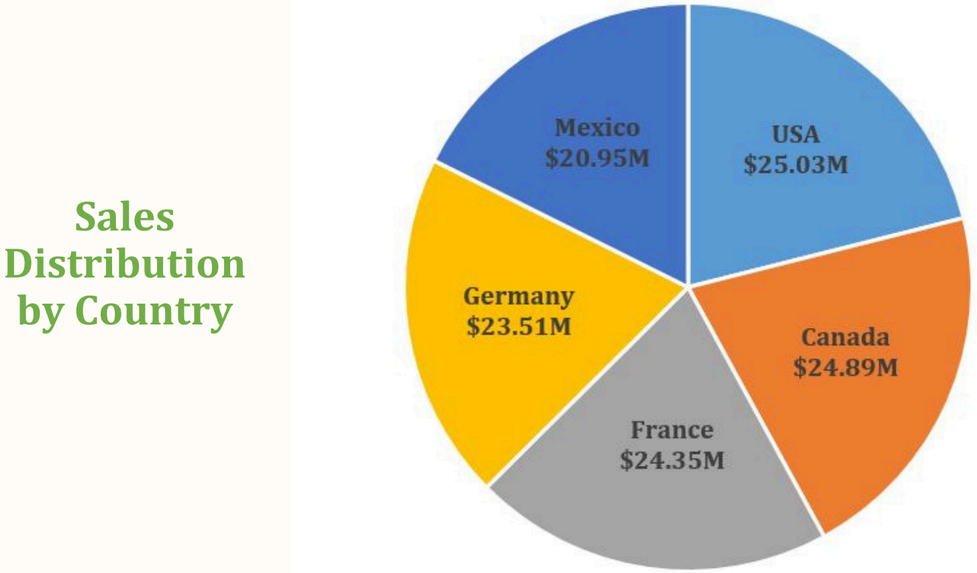
1. Sales Performance Overview

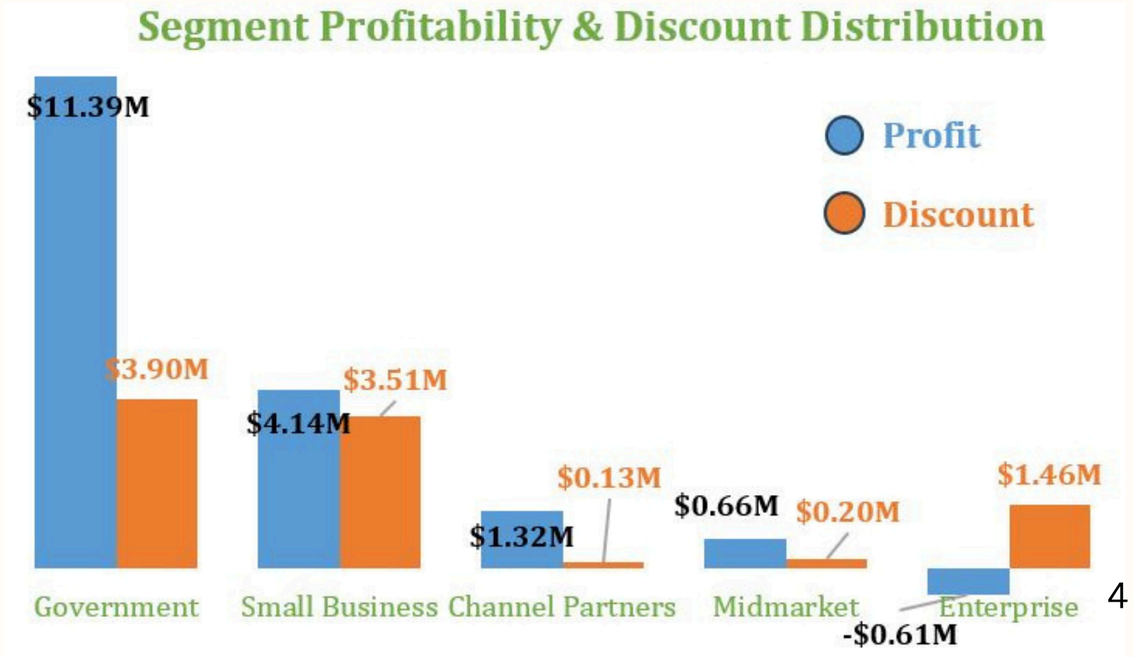
- **KPIs:** Total Sales, Total Profit, Total Units Sold
- **Visualizations:**
  - Monthly Sales & Profit
  - Quarterly Sales & Profit
  - Country-wise Sales Distribution
  - Country-wise Profit & Total Discount Provided

2. Discount vs Profitability Dashboard

- **Visualizations:**
  - Segment-wise Discount & Profit Distribution
  - Profit & Discount Relationship
  - Product-wise Total Sales

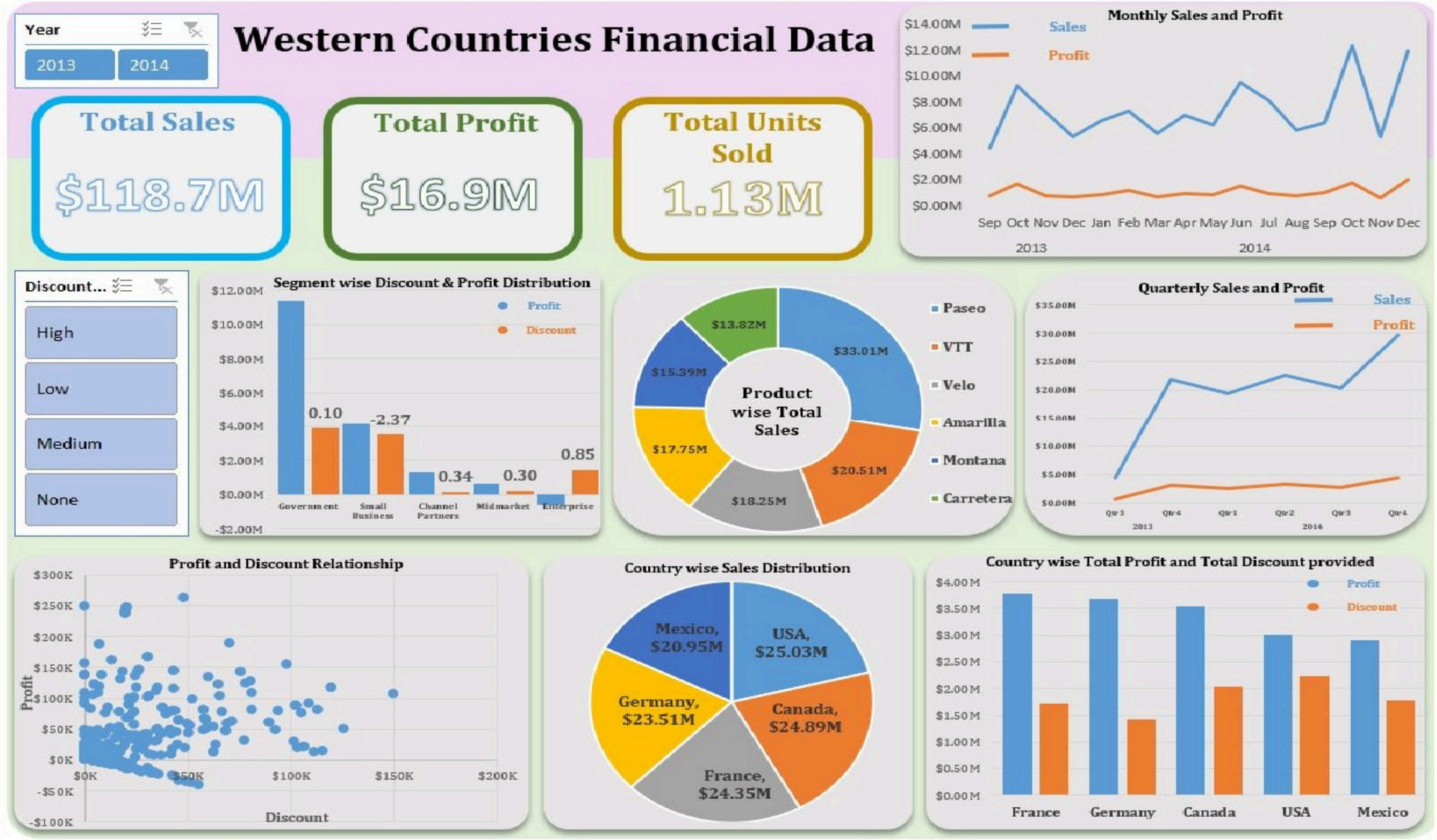






3. Quarterly Sales Breakdown & Slicers

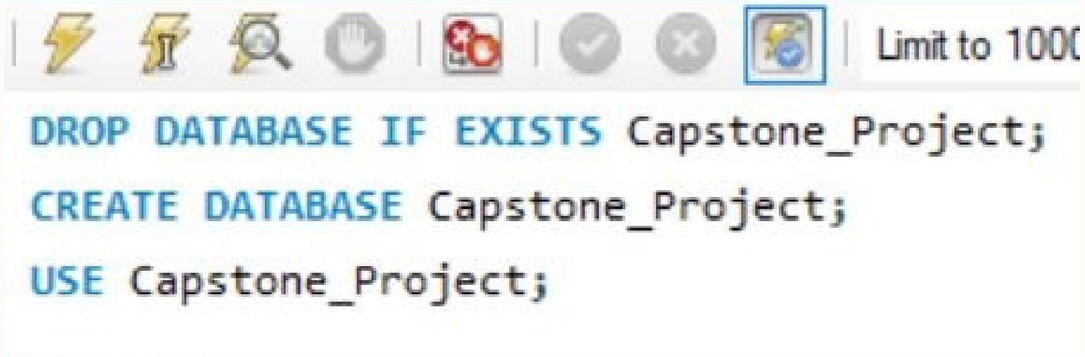
- **Visualizations:**
  - Best & Worst Performing Quarters
  - YoY (Year-over-Year) Growth
- **Slicers for Dynamic Analysis:**
  - Year (2013 & 2014)
  - Discount Bands (High, Medium, Low, None)

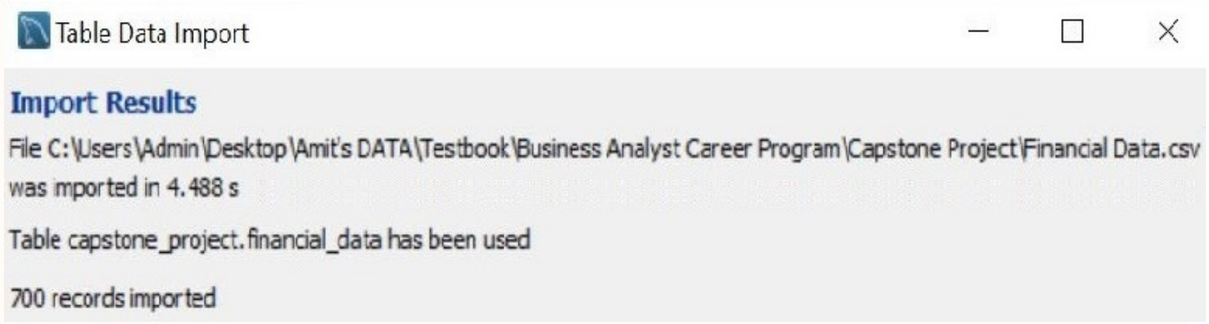


SQL & Power BI Implementation

Database Creation in MySQL

Created a structured database in MySQL Server.



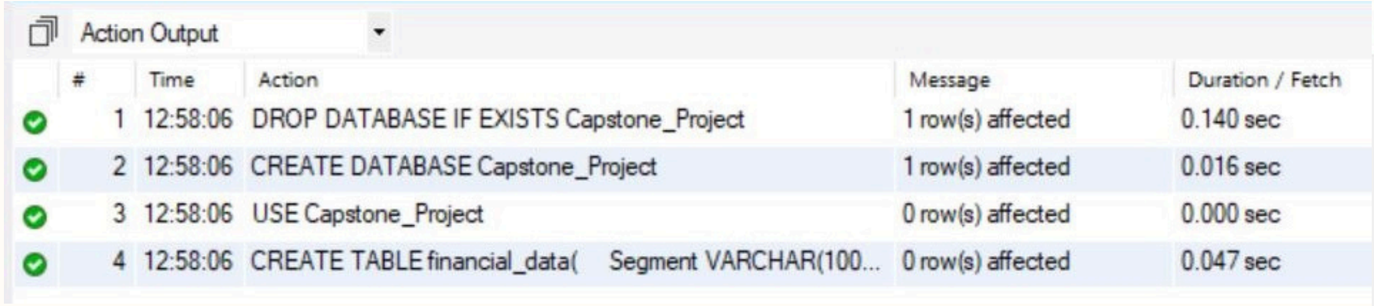


Defined appropriate data types and imported the dataset.

```
CREATE TABLE financial_data(
    Segment VARCHAR(100),
    Country VARCHAR(100),
    Product VARCHAR(100),
    Discount_Band VARCHAR(100),
    Units_Sold INT,
    Manufacturing_Price DECIMAL(10, 2),
    Sale_Price DECIMAL(10, 2),
    Gross_Sales DECIMAL(10, 2),
    Discounts DECIMAL(10, 2),
    Sales DECIMAL(10, 2),
    COGS DECIMAL(10, 2),
    Profit DECIMAL(10, 2),
    Date DATE,
    Month_Number INT,
    Month_Name VARCHAR(100),
    Year VARCHAR(100)
);
```

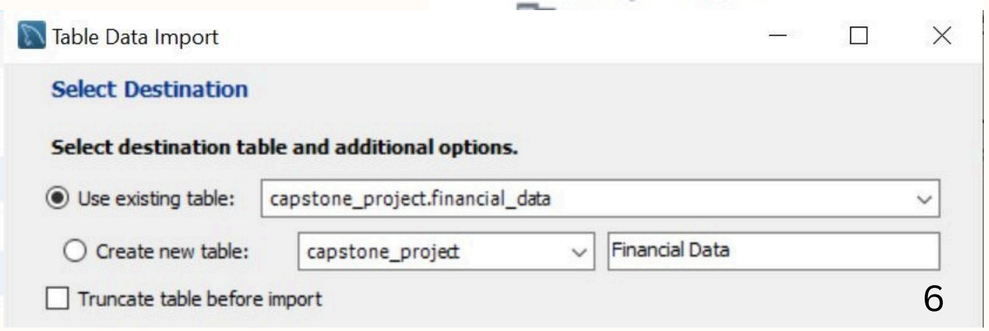
Schema Created

Query Output for Database and Table creation



Action Output

#	Time	Action	Message	Duration / Fetch
1	12:58:06	DROP DATABASE IF EXISTS Capstone_Project	1 row(s) affected	0.140 sec
2	12:58:06	CREATE DATABASE Capstone_Project	1 row(s) affected	0.016 sec
3	12:58:06	USE Capstone_Project	0 row(s) affected	0.000 sec
4	12:58:06	CREATE TABLE financial_data(Segment VARCHAR(100...	0 row(s) affected	0.047 sec



Verified the data import using SQL queries and Excel pivot tables.

Result Grid | Filter Rows: | Export:

country	sum_sales	profit
United States of America	25029830.18	2995540.68
Canada	24887654.89	3529228.89
France	24354172.29	3781020.79
Germany	23505340.82	3680388.82
Mexico	20949352.11	2907523.11

```
SELECT country, SUM(sales) AS sum_sales, round(SUM(profit),2) as profit
FROM financial_data GROUP BY country ORDER BY sum_sales DESC;
```

Row Labels	Sum of Sales	Sum of Profit
United States of America	$25,029,830.17	$2,995,540.67
Canada	$24,887,654.89	$3,529,228.89
France	$24,354,172.28	$3,781,020.78
Germany	$23,505,340.82	$3,680,388.82
Mexico	$20,949,352.11	$2,907,523.11

Result Grid | Filter Rows:

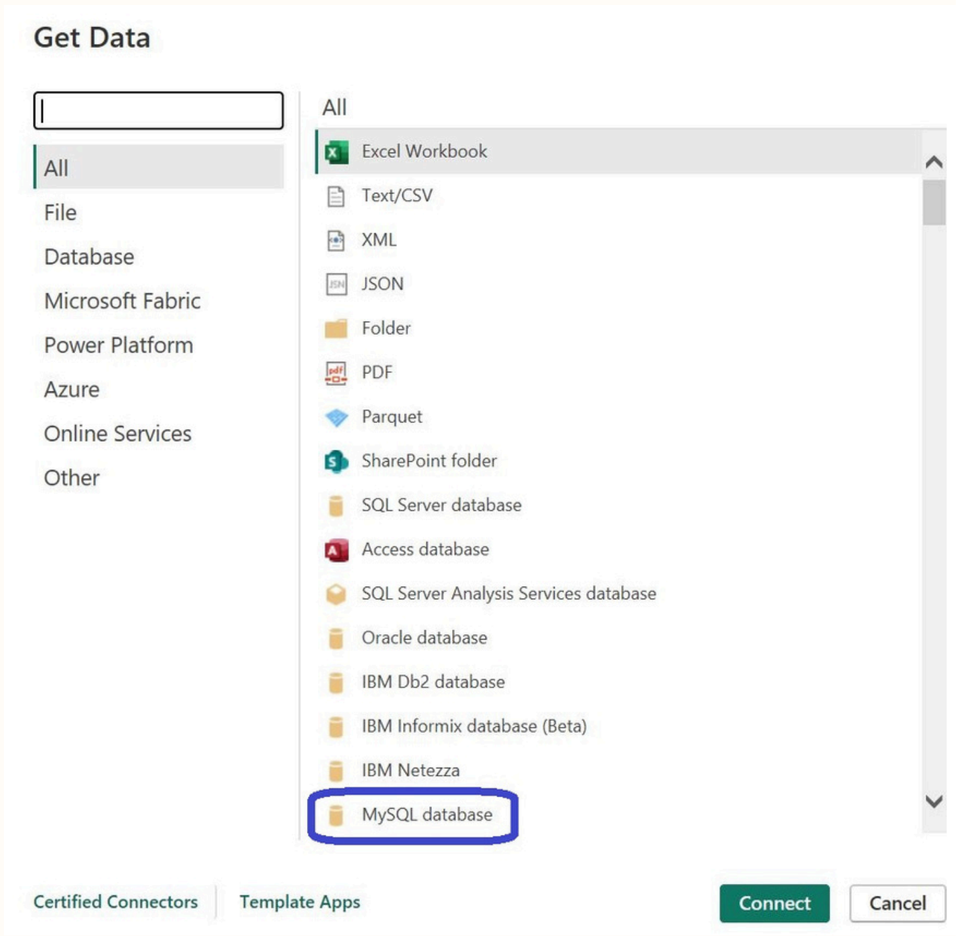
segment	sum_sales	profit
Government	52504260.68	11388173.18
Small Business	42427918.50	4143168.50
Enterprise	19611694.38	-614545.62
Midmarket	2381883.09	660103.09
Channel Partners	1800593.64	1316803.14

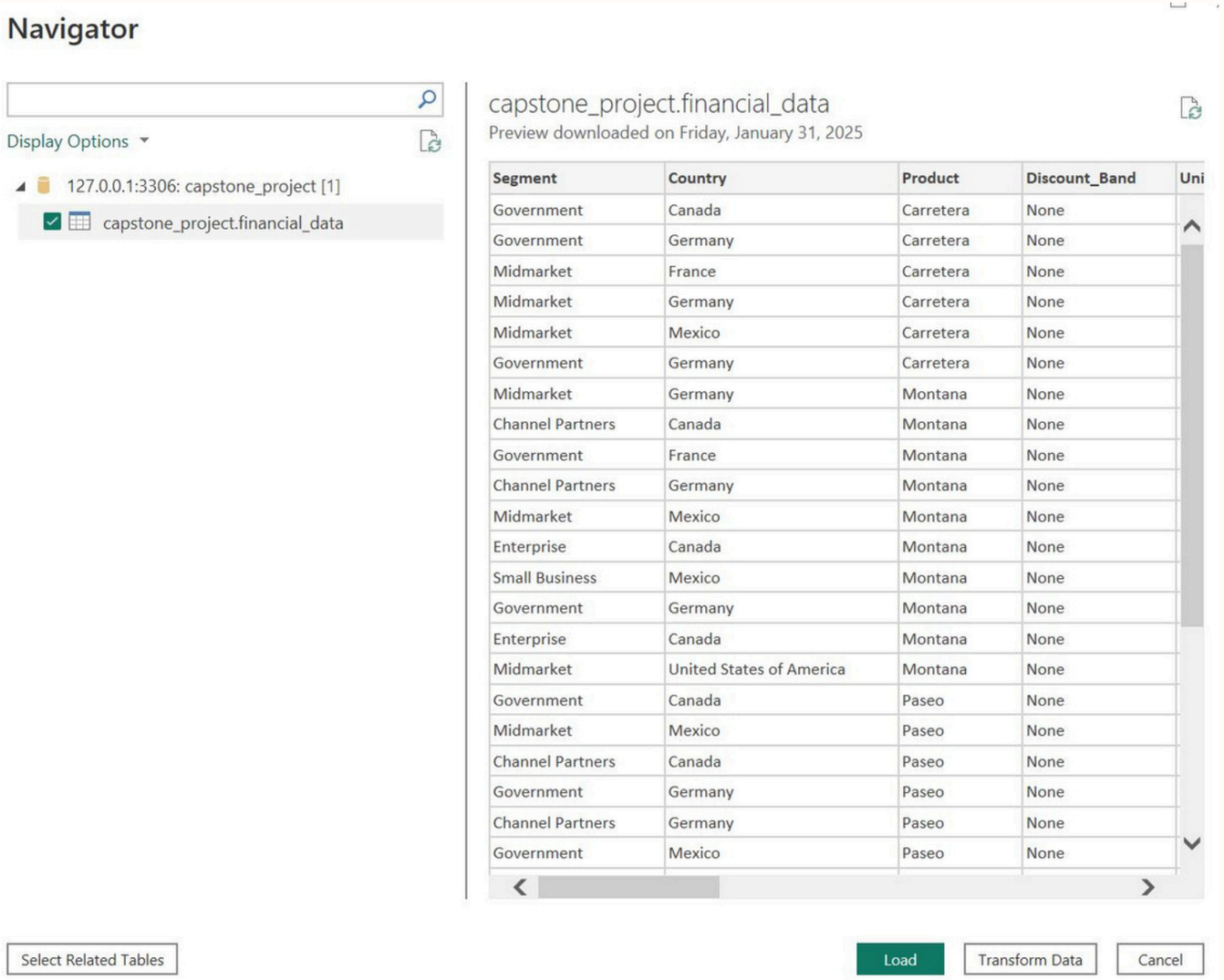
```
SELECT segment, SUM(sales) AS sum_sales, round(SUM(profit),2) as profit
FROM financial_data GROUP BY segment ORDER BY sum_sales DESC;
```

Row Labels	Sum of Sales	Sum of Profit
Government	$52,504,260.67	$11,388,173.17
Small Business	$42,427,918.50	$ 4,143,168.50
Channel Partners	$ 1,800,593.64	$ 1,316,803.14
Midmarket	$ 2,381,883.08	$ 660,103.08
Enterprise	$19,611,694.38	$ (614,545.63)

Power BI Data Integration & Transformation

- Connected Power BI to the MySQL database for real-time analysis.

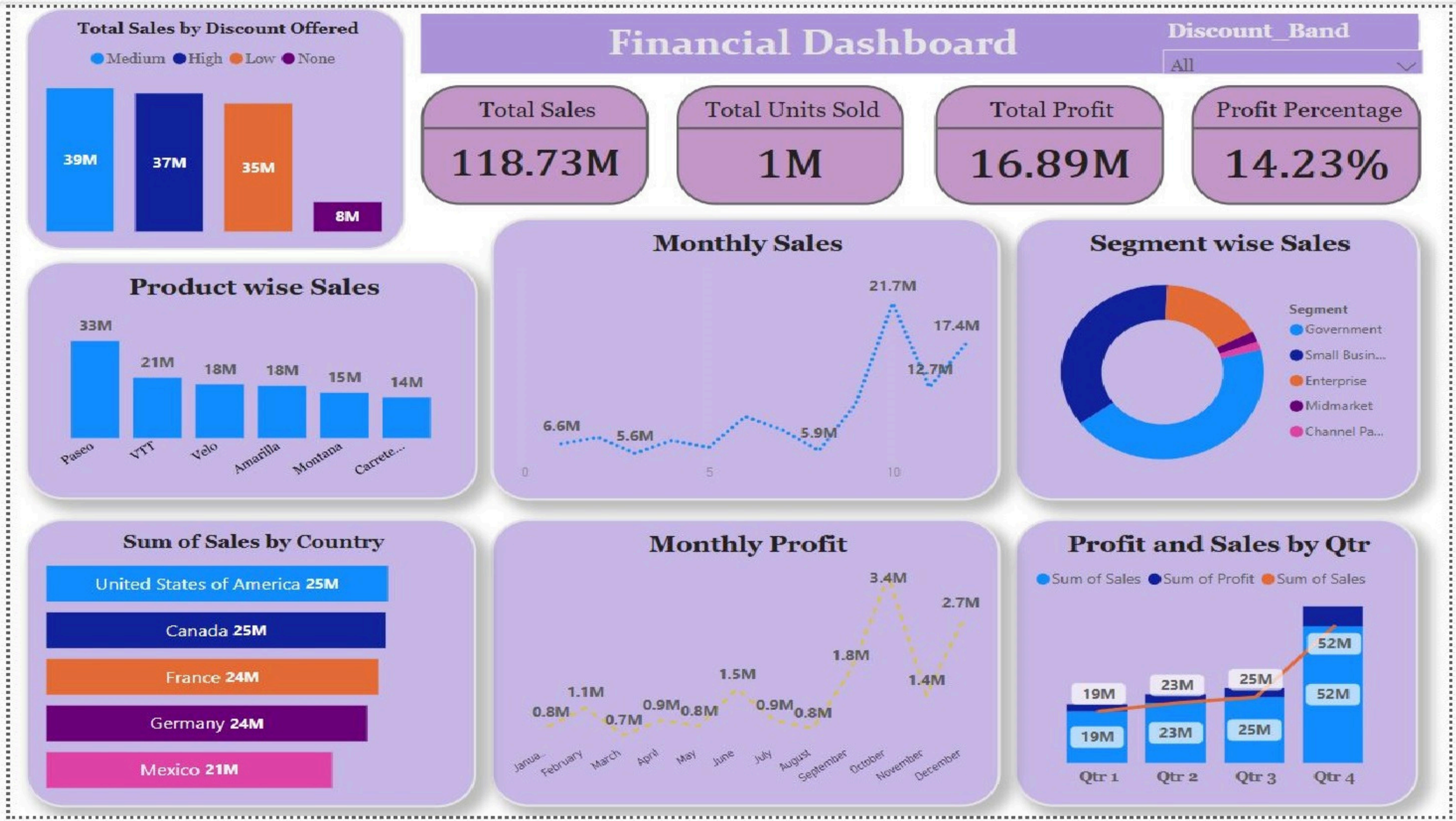




Created calculated fields:

Profit Margin (%) = (Profit / Sales) * 100

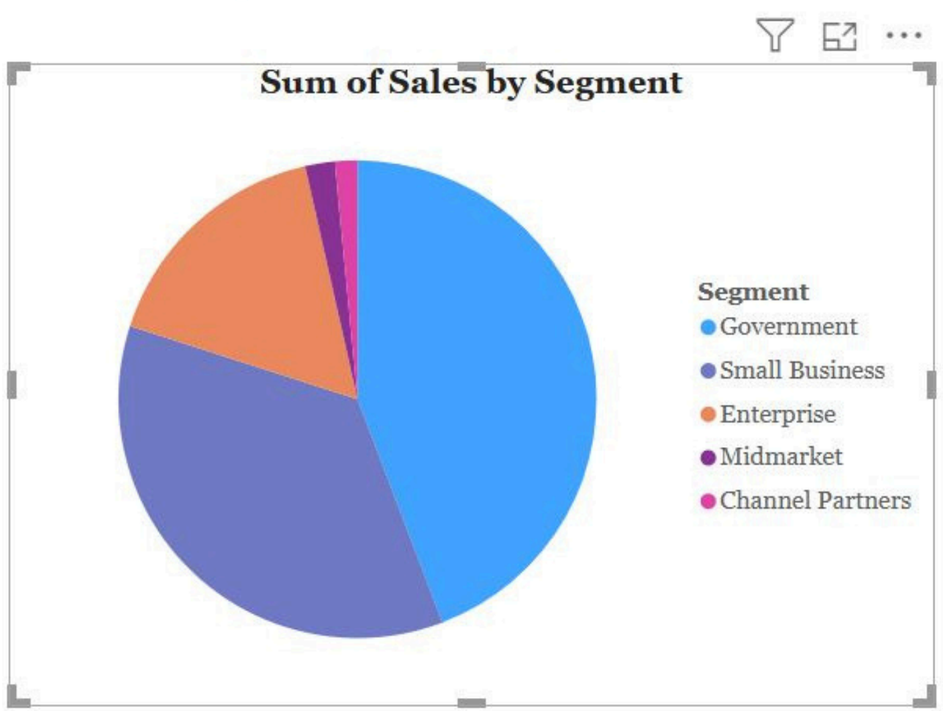
Power BI Dashboard



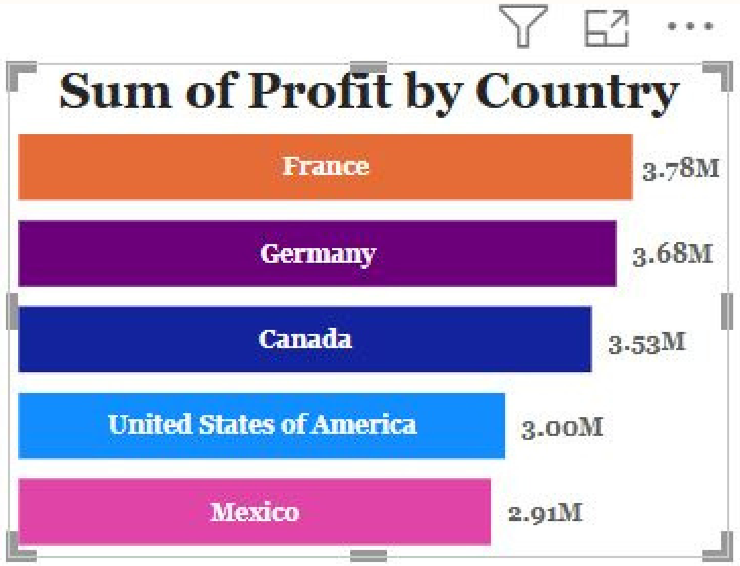
Insights & Key Takeaways

A. Sales & Seasonal Trends

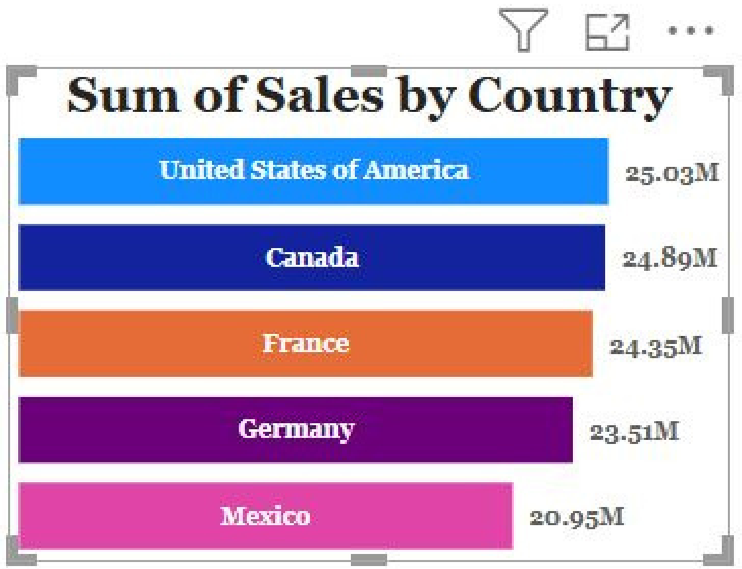
Q4 was consistently the highest revenue-generating quarter, indicating a strong year-end shopping trend.
Sales dropped in October-November but recovered in December 2014, unlike in 2013.
The Government sector drove nearly 50% of total sales, making it the most valuable segment.



B. Segment & Profitability Analysis

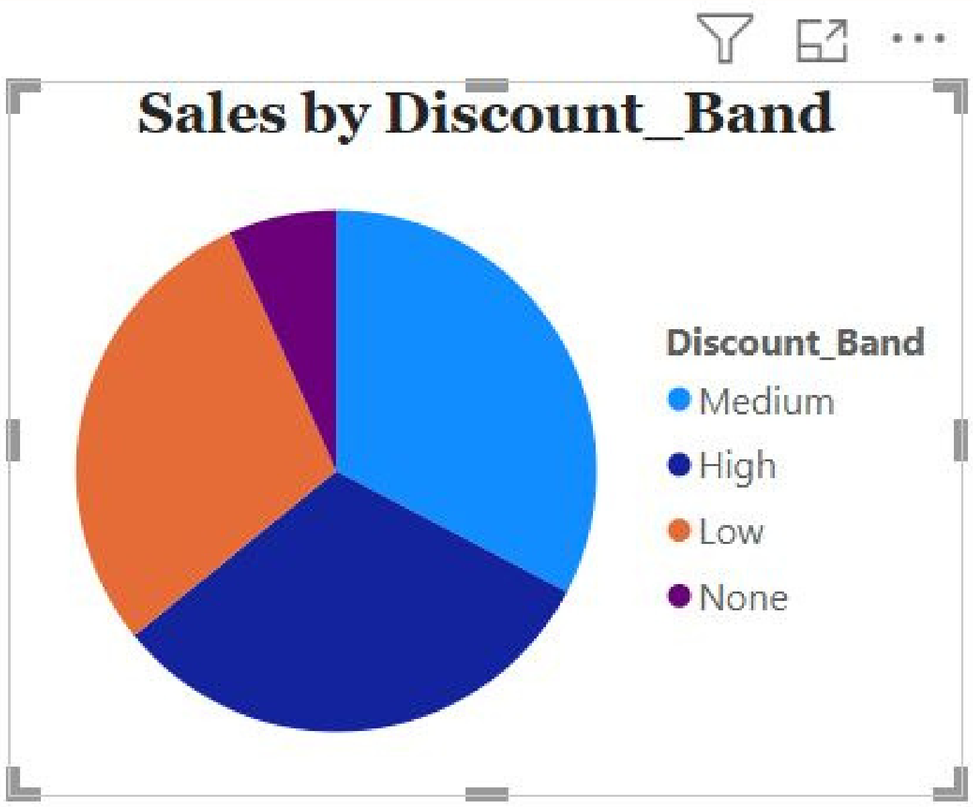


Enterprise discounts led to negative profits, signaling a flawed pricing strategy.
France had the highest profit margins, while the USA had the highest revenue.
Paseo emerged as the best-selling product.

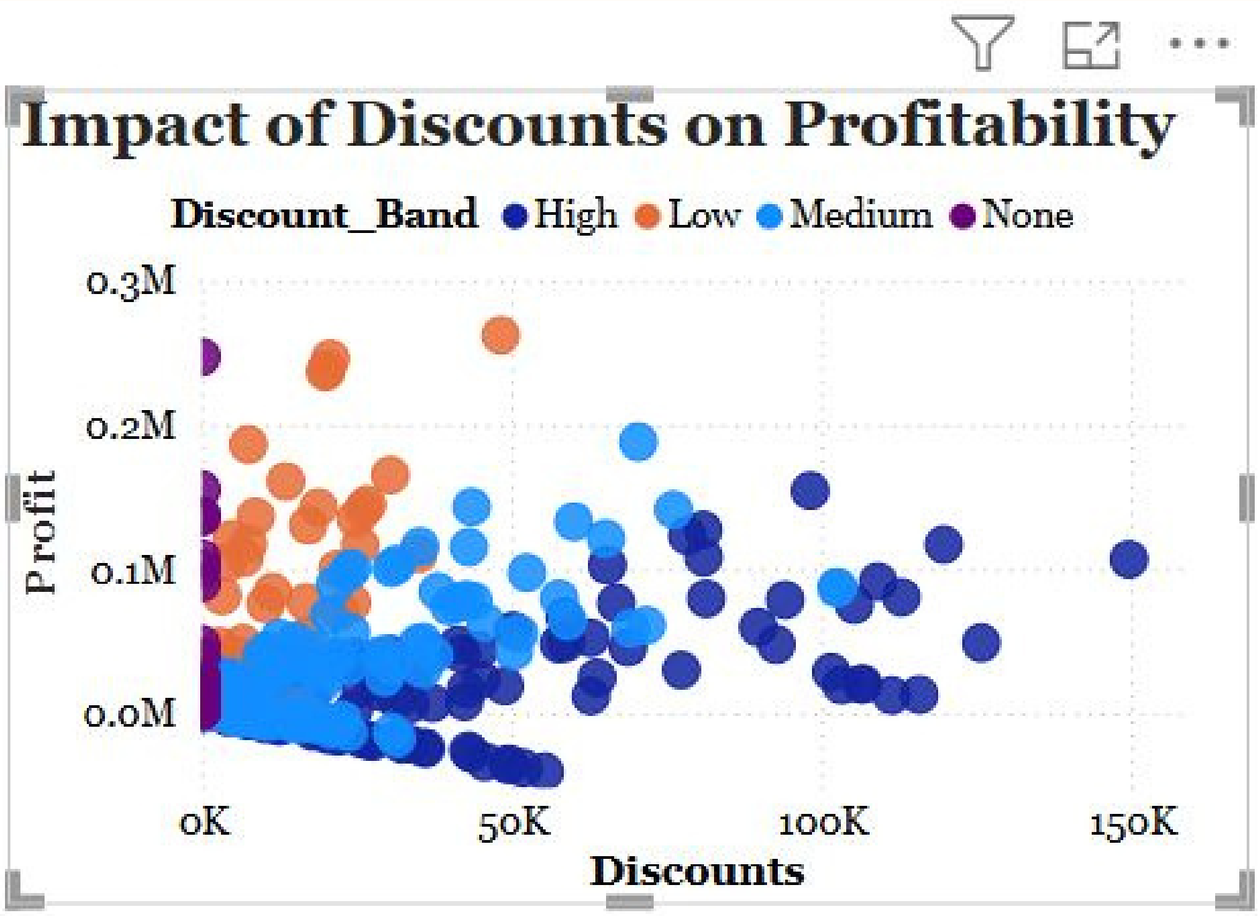


C. Discount & Revenue Relationship

647 out of 700 records had some level of discount, proving discounts were a key sales driver.



The correlation between Discounts and Profits was moderate (0.38), meaning excessive discounts didn't always result in better profitability.



D. Actionable Recommendations

Refine discount strategies for Enterprise customers to avoid losses.

Focus more on high-margin products in France instead of just volume sales in the USA.

Prepare inventory and marketing campaigns for Q4 in advance, as it's the most profitable period.