

# End-to-End Financial Analysis Using SQL, Python, and Power BI: Profitability, Forecasting & Optimization

This project presents a comprehensive financial analysis using a dataset of 700+ company transactions across products, countries, and segments. The objective was to uncover key drivers of profit and loss, identify underperforming regions and products, and forecast future performance using a combination of SQL, Python, and Power BI.

Key findings include:

- Small Business and Government segments are the most profitable overall, while Enterprise consistently incurs losses due to excessive discounting and high COGS.
- High discount rates erode profitability without boosting sales volume, especially in markets like Mexico and Canada.
- Low-margin products like Velo and Vtt still contribute high profits through sales volume but require pricing and cost control.
- A Prophet-based time series model forecasts steady profit growth to over \$6.3M by Jan 2015.

Through SQL-driven KPIs, Python-based machine learning, and interactive Power BI dashboards, this project demonstrates an integrated approach to actionable financial intelligence.

Keerthi Balaji

## Questions/Problem Statement

After initial inspection of the data, I focused on the following questions and areas for analysis.

1. Which product segments are most and least profitable by region and over time?
2. How do discounts affect total profit and sales volume?
3. Can we forecast monthly sales or profits for the next year?
4. Which countries consistently over- or underperform compared to global averages?
5. Which products yield high revenue but low profit margins?
6. How come the enterprise segment is incurring losses? What factors are causing it and how can it be improved?

## Dataset

I found a company financial dataset on Kaggle - [Company Financials Dataset](#) with 700 rows.

The dataset has the following features/columns:

1. Segment— The business or customer segment (e.g., Government, Midmarket, Channel Partners).
2. Country— Country where the transaction took place.
3. Product—The product involved in the transaction (e.g., Amarilla, Carretera).
4. Discount Band—The discount level applied to the sale (e.g., None, Low, Medium).
5. Units Sold—Estimated quantity of items sold. Used to calculate gross sales and COGS.
6. Manufacturing Price—Listed per-unit manufacturing cost. Note: this column may represent a static or placeholder value that does not align with actual cost calculations in the dataset.
7. Sale Price (\$/unit)The per-unit selling price charged to the customer.
8. Gross Sales (\$)Total sales amount before applying discounts: Units Sold × Sale Price.
9. Discounts (\$)Dollar value of discounts applied to the gross sales.
10. Net Sales (\$)—Actual revenue after discounts: Gross Sales – Discounts.
11. COGS (\$)Cost of Goods Sold — total cost to produce the items sold.
12. Profit (\$)Net profit from the transaction: Net Sales – COGS.
13. Date—Exact transaction date (format: dd/mm/yy or yyyy-mm-dd).
14. Month Number—Numerical representation of the transaction month (1 for January, etc.).
15. Month Name—Full name of the transaction month (e.g., January).
16. Year—Four-digit year of the transaction.

## Understanding the data and Preprocessing

Looking at the data, I was unsure of the columns and what they denoted. The data types of the columns were also not mentioned. For example, units sold should be whole number denoting the number of units sold but, in the dataset, it has a dollar sign which indicates currency. I also noticed that for certain products the manufacturing price is higher than the sale price and still shows positive profit for that product. This ambiguity makes it difficult to understand what the columns denote and what the data represents.

With this issue in mind, I proceeded with preprocessing the dataset. As suggested by the creator, I removed the '\$' and - first using python. I also replaced missing values with NaN values and replaced profit values in format (x) denoting negative numbers with float values in format -x. Then, I renamed column names for clarity.

For categorical features, I changed the datatype to 'category' for efficient processing. I, then, standardized capitalization and whitespace for these columns. Below is a snapshot of cleaned data.

	Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price (\$/unit)	Gross Sales (\$)	Discounts (\$)	Net Sales (\$)	...	Profit (\$)	Date	Month Number	Month Name	Year
0	Government	Canada	Carretera	None	1618.5	3.0	20.0	32370.0	NaN	32370.0	...	16185.0	2014-01-01	1	January	2014
1	Government	Germany	Carretera	None	1321.0	3.0	20.0	26420.0	NaN	26420.0	...	13210.0	2014-01-01	1	January	2014
2	Midmarket	France	Carretera	None	2178.0	3.0	15.0	32670.0	NaN	32670.0	...	10890.0	2014-01-06	6	June	2014
3	Midmarket	Germany	Carretera	None	888.0	3.0	15.0	13320.0	NaN	13320.0	...	4440.0	2014-01-06	6	June	2014
4	Midmarket	Mexico	Carretera	None	2470.0	3.0	15.0	37050.0	NaN	37050.0	...	12350.0	2014-01-06	6	June	2014

5 rows x 21 columns

I decided to conduct some feature engineering. I created some derived KPIs such as Profit Margin (%), Discount Rate (%), and COGS Ratio (%) to prepare the data for more useful and actionable insights. I added two new features – quarter and weekday to see some time-based trends. The snapshot below shows the added features.

Quarter	Weekday	Profit Margin (%)	Discount Rate (%)	COGS Ratio (%)	
1	Wednesday	50.00	NaN	50.00	
1	Wednesday	50.00	NaN	50.00	
2	Monday	33.33	NaN	66.67	
2	Monday	33.33	NaN	66.67	
2	Monday	33.33	NaN	66.67	

After close inspection of the manufacturing price column, I noticed it shows values significantly higher than the calculated cost per unit derived from COGS/Units Sold. Additionally, this value is not used in any computation of COGS, Sales, or Profit (in the data). Manufacturing Price column may represent a static or placeholder value that does not align with actual cost calculations in the dataset.

For accurate financial analysis, the COGS values should be used as the true representation of production costs. The Manufacturing Price column is retained only for reference and is not included in any profit or cost calculation in this analysis.

## Using SQL for data analysis (sqlite)

After creating and loading the financials data into SQL, I did some basic data analysis.

### Total KPIs by Segment:

```
--  
19  SELECT  
20      Segment,  
21      SUM("Net Sales ($)") AS Total_Sales,  
22      SUM("Profit ($)") AS Total_Profit,  
23      ROUND(AVG("Profit Margin (%)" ), 2) AS "Avg Profit Margin (%)"  
24  FROM financials  
25  GROUP BY Segment;  
26  ORDER BY Total_Profit DESC;  
27
```

	Segment	Total_Sales	Total_Profit	Avg Profit Margin (%)
1	Government	52504260.68	11388173.18	29.33
2	Small Business	42427918.5	4143168.5	9.67
3	Channel Partners	1800593.64	1316803.14	73.02
4	Midmarket	2381883.09	660103.09	27.67
5	Enterprise	19611694.38	-614545.62	-3.22

The above query checks the total net sales, total profit, and average profit margin across business segments. This allows us to analyze performance not only in terms of raw revenue or profit but also efficiency (profitability per dollar of sales).

Government – top performer

The Government has over \$52.5 million in total net sales and a total profit of \$11.39 million, it demonstrates both high volume and healthy profitability. Its average profit margin of 29.33% suggests that the products or services sold in this segment are not only in demand but also priced efficiently relative to their cost. This segment represents a highly successful business model and should be prioritized for continued investment, expansion, and optimization.

Small Business

Small Business ranks second in both total sales and profit, generating \$42.4 million in sales and \$4.14 million in profit. However, it operates with a relatively low average profit margin of 9.67%, indicating a volume-driven strategy. While it brings in significant revenue, the slim margins may pose a risk if operational costs increase or sales dip.

#### Channel Partners

Channel Partners displays a unique profile with lower total sales of \$18 million, but a remarkably high average profit margin of 73.02%. The total profit of \$1.32 million may not be the highest, but it is impressive given the sales base. This suggests that while the number of transactions may be limited, each deal is extremely lucrative. This segment's strategies can be used in other segments in order to increase the profitability of that segment, especially the Small Business segment.

#### Midmarket

The midmarket shows modest figures with \$2.38 million in net sales and \$660,103 in total profit but boasts a strong average profit margin of 27.67%. This indicates a potentially valuable segment that is currently underutilized. Given the high margin, there is an opportunity to expand this segment through targeted marketing or tailored product offerings.

#### Enterprise – has significant loss

The Enterprise segment is a major concern. Despite generating a substantial \$19.6 million in total sales, it results in a net loss of over \$614,000 and a negative average profit margin of -3.22%. This strongly indicates that the costs involved in serving this segment outweigh the revenues. This segment requires an immediate strategic review—potential actions include restructuring pricing, reassessing cost models, or even pulling back from this segment if profitability cannot be restored.

### COGS Efficiency by Segment

```
42     SELECT
43         Segment,
44         ROUND(AVG("COGS Ratio (%)" ), 2) AS Avg_COGS_Ratio
45     FROM financials
46     GROUP BY Segment
47     ORDER BY Avg_COGS_Ratio;
```

	Segment	Avg_COGS_Ratio
1	Channel Partners	26.98
2	Government	70.67
3	Midmarket	72.33
4	Small Business	90.33
5	Enterprise	103.06

This query and result show the average Cost of Goods Sold (COGS) ratio for each segment, which is a key metric for understanding how much it costs to produce or acquire the goods sold relative to their selling price.

The COGS ratio analysis highlights clear differences in cost efficiency across segments. Channel Partners is the most cost-effective, with a low average COGS ratio of 26.98%, meaning it retains the majority of its revenue as profit. In contrast, Government and Midmarket have high COGS ratios of 70.67% and 72.33%, respectively, indicating that a large portion of their revenue is consumed by costs, leaving tighter margins. Small Business is even more cost-heavy, with a very high ratio of 90.33%, suggesting that it operates on razor-thin margins and is highly sensitive to cost increases. Most critically, the Enterprise segment has a COGS ratio of 103.06%, meaning it spends more to produce goods than it earns in revenue, resulting in consistent losses and a fundamentally unsustainable cost structure.

## Top 5 Most Profitable Products

```
34     SELECT
35         Product,
36         SUM("Profit ($)") AS Total_Profit
37     FROM financials
38     GROUP BY Product
39     ORDER BY Total_Profit DESC
40     LIMIT 5;
41
42
```

	Product	Total_Profit
1	Paseo	4797437.96
2	Vtt	3034608.02
3	Amarilla	2814104.07
4	Velo	2305992.47
5	Montana	2114754.88

The top 5 profitable products are: Paseo, Vtt, Amarilla, Velo, and Montana.

## Low Margin Products (Profit Margin < 5%)

```

50  SELECT
51      Product,
52      AVG("Profit Margin (%)") AS Avg_Profit_Margin
53  FROM financials
54  GROUP BY Product
55  HAVING AVG("Profit Margin (%)") < 35
56  ORDER BY Avg_Profit_Margin ASC;
57
58

```

	Product	Avg_Profit_Margin
1	Vtt	26.3764220183486
2	Velo	26.4855963302752
3	Montana	27.5042391304348
4	Paseo	28.9808040201005
5	Amarilla	29.3033333333333
6	Carretera	29.521935483871

The products Vtt (26.38%), Velo (26.49%), and Montana (27.50%) have the lowest average profit margins, making them the least profitable among the group. These may need attention—either their costs are too high, or they are priced too low relative to their value. Paseo (28.98%), Amarilla (29.30%), and Carretera (29.52%) also fall below the 35% threshold, but are slightly better in terms of profitability. All six products are operating within a narrow margin range of ~26%–30%, which suggests that while not unprofitable, they are less efficient contributors to profit and may benefit from cost reductions, pricing adjustments, or promotional support to improve their margins.

Interestingly, I noticed the top 5 profitable products also appear to have the lowest profit margins. This reveals a valuable insight: while these products operate on relatively low profit margins (ranging from ~26% to ~30%), they still generate the highest total profits due to high sales volumes.

#### Implications:

- Paseo, with the highest total profit (\$4.79M), has an average profit margin of ~29%, indicating strong sales volume compensating for modest profitability per unit.
- Vtt and Velo, despite having the lowest margins (~26.4%), still bring in \$3.03M and \$2.31M in profit respectively, making them high-volume, low-margin success stories.
- Amarilla and Montana show a similar pattern, with moderate margins and solid overall profit.

These products are strategic drivers of overall profitability despite their low margins. They should be maintained and monitored, but any further cost optimization (e.g., supply chain efficiencies or reducing discounts) or pricing tweaks could significantly boost overall profit. However, aggressive price increases could backfire by reducing volume—so changes should be data-driven and carefully evaluated.

## Compare Discounts Across Countries

```
58  SELECT
59      Country,
60      ROUND(AVG("Discount Rate (%)) , 2) AS Avg_Discount_Rate
61      --COUNT(*) AS Num_Transactions
62  FROM financials
63  GROUP BY Country
64  ORDER BY Avg_Discount_Rate DESC;
65
66
```

	Country	Avg_Discount_Rate
1	Mexico	8.34
2	Canada	8.32
3	United States Of America	7.8
4	Germany	7.62
5	France	7.58

This analysis identifies the discount rates in countries, which offers insight into where the company is offering the most price reductions on sales. Mexico (8.34%) and Canada (8.32%) have the highest average discount rates, indicating these markets receive the most aggressive discounting overall. The United States (7.80%), Germany (7.62%), and France (7.58%) follow closely, suggesting discounts are consistently applied across all regions, though with slight variations.

Since discounting affects net sales and profit margins, the business should validate whether the higher discount rates in certain countries translate to increased sales volumes or profitability — or if they are eroding margins without sufficient return.

## Power BI – Visuals

I will use Power BI to answer some questions about the data to gain insights.

Which product segments are most and least profitable by region and over time?

Segment	Canada	France	Germany	Mexico	United States Of America	Total
Channel Partners	17,948.91	13,579.07	12,367.94	8,544.50	13,399.73	13,168.03
Enterprise	-6,075.44	-5,039.44	-5,340.72	-7,098.75	-8,756.75	-6,468.90
Government	37,641.19	45,165.25	44,619.60	33,985.99	28,390.85	37,960.58
Midmarket	6,624.45	8,227.10	4,267.74	7,527.32	6,358.54	6,601.03
Small Business	45,039.95	36,536.58	38,598.65	33,380.30	53,602.95	41,431.69
<b>Total</b>	<b>25,208.78</b>	<b>27,201.59</b>	<b>26,477.62</b>	<b>21,222.80</b>	<b>21,396.72</b>	<b>24,307.49</b>

Most Profitable Segments (Overall & By Region)

- Small Business is the most profitable overall, with a total of \$41,431.69, and dominates in USA (\$53,602.95), Canada, and Germany.
- Government is also consistently profitable across all regions, totaling \$37,960.58, showing particularly strong returns in Germany (\$44,619.60) and France.
- These segments are not only profitable overall but perform well across multiple countries, indicating broad market appeal and operational efficiency.

### Moderate/Variable Performance

- Channel Partners yields moderate profit overall (\$13,168.03), but the performance is regionally uneven—stronger in Canada and USA but much lower in Mexico and Germany.
- The midmarket also shows modest profits in all regions, totaling \$6,601.03.
- These segments indicates potential, but not a strong profit center — could benefit from strategic growth in better-performing regions.

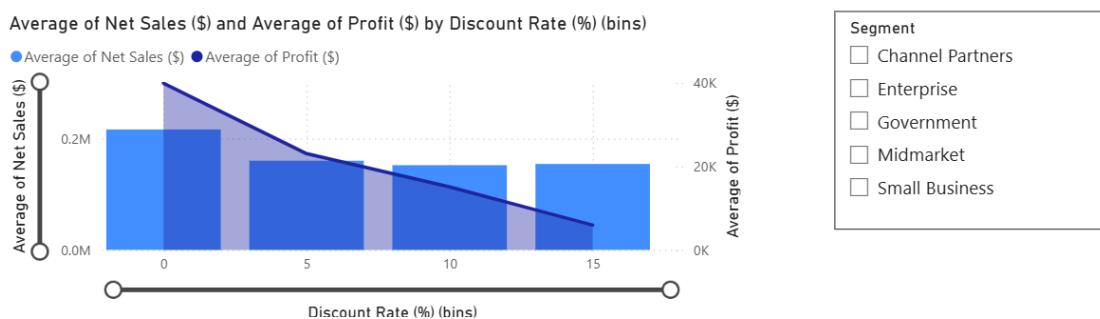
### Least Profitable Segment: Enterprise

- Enterprise stands out as the only segment consistently incurring losses across every single country, with a total loss of -\$6,468.90.
- Especially poor in USA (-\$8,756.75) and Mexico (-\$7,098.75).
- This signals a fundamental issue — possibly due to high servicing costs, discounts, or unfit offerings — and needs a deep strategic review.

### Regional Profitability Patterns

- Germany and France appear as strong markets overall, especially for Government and Small Business.
- USA delivers extremely high profit for Small Business, helping it dominate the overall totals.
- Mexico is a challenging market: profits are generally lower, and Enterprise losses are severe.
- Canada performs steadily across segments, with a good balance of profitability.

## How do discounts affect total profit and sales volume?



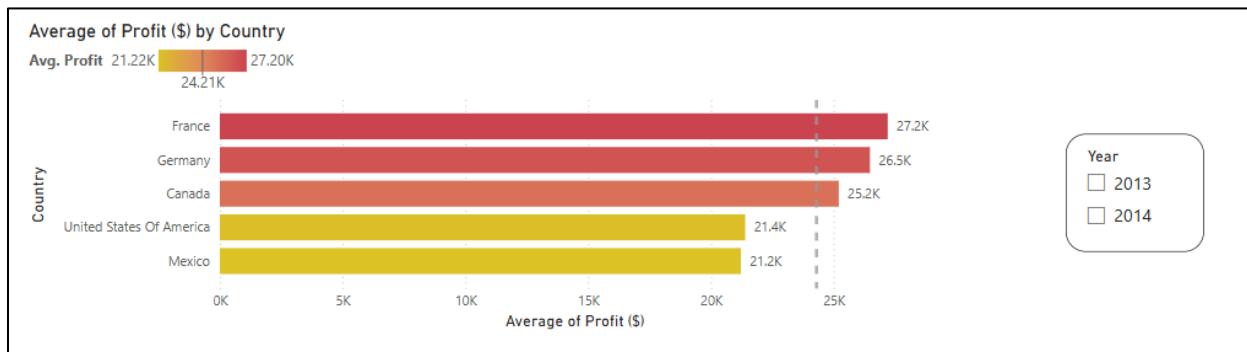
As discount rates increase, average profit (\$) consistently declines, forming a clear downward trend. Net sales initially stay stable but show a slight decline after the 5% mark, indicating that higher discounts do not significantly boost sales volume. The steep drop in profit without a proportional rise in net sales suggests diminishing returns from discounts.

While small discounts (0–5%) may be sustainable, larger discounts (10–15%) sharply reduce profit without meaningfully increasing sales. This indicates that over-discounting erodes margins and may not be an effective sales strategy.

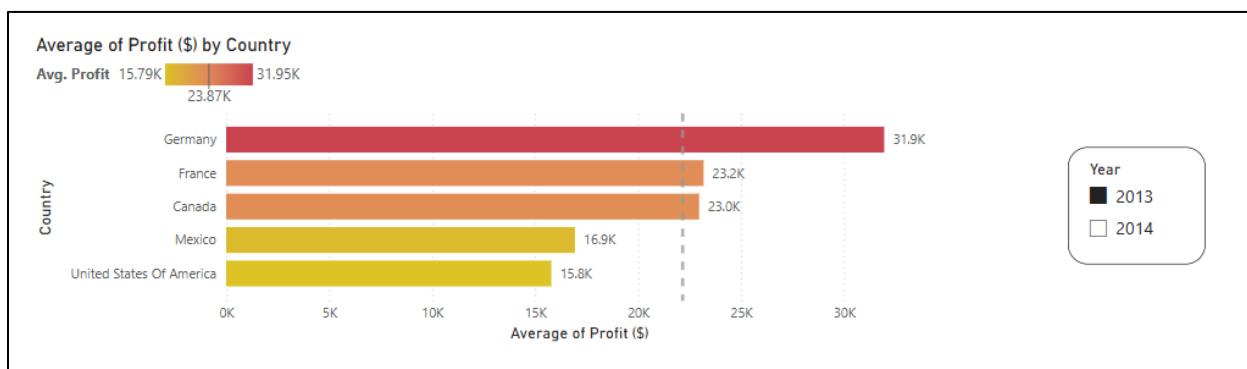
The business should consider:

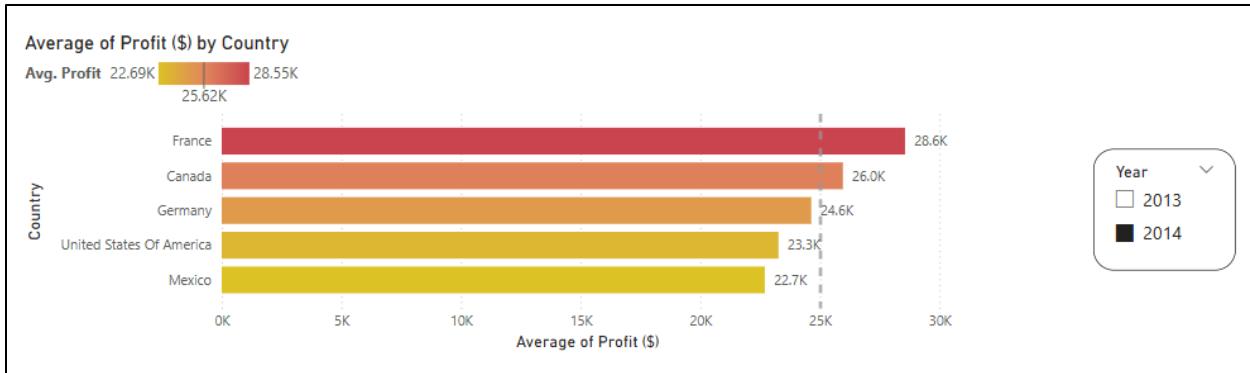
- Restricting high discounts on selective use (e.g., clearance or high-margin products).
- Testing lower discount tiers to find the sweet spot between conversion and profitability.
- Segmenting discount strategy by product or customer type to maximize ROI.

## Which countries consistently over- or underperform compared to global averages?



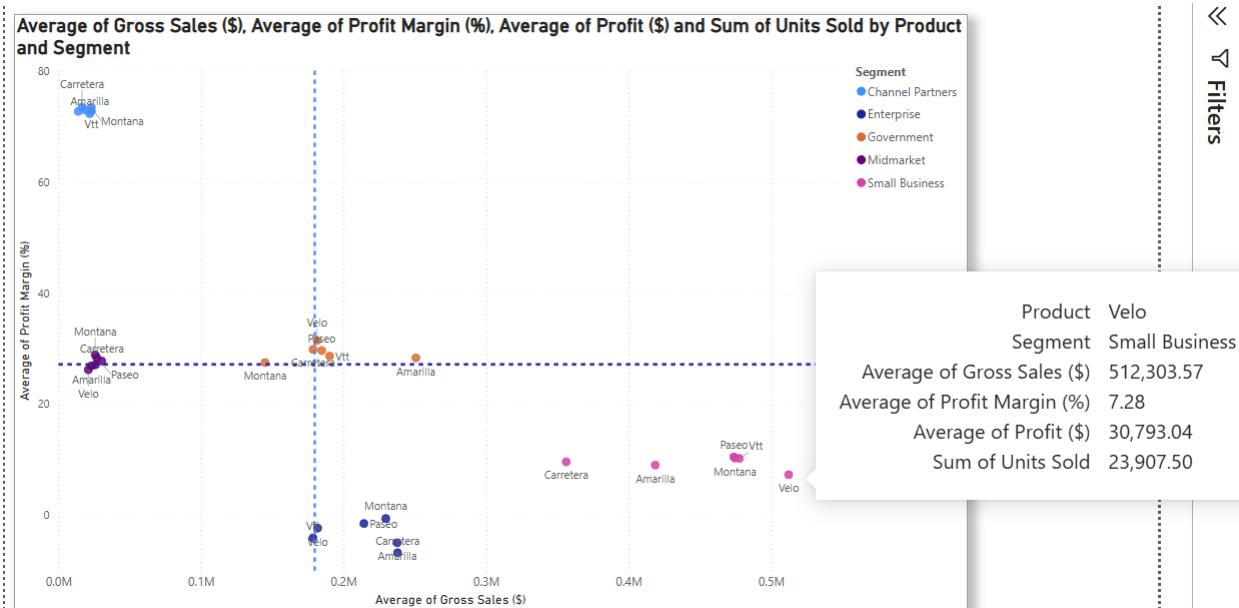
- Global average profit is marked with a dotted line at around \$25K.
- France (27.2K), Germany (26.5K), and Canada (25.2K) are all above the global average, indicating they are consistent overperformers in terms of profitability.
- United States (21.4K) and Mexico (21.2K) fall below the global average, making them underperformers.





- France is the most consistent overperformer, above global average in both 2013 and 2014.
- Germany was the top performer in 2013 but fell slightly below average in 2014.
- Canada improved from slightly below average in 2013 to above average in 2014 — a positive trend.
- USA and MX Mexico consistently underperformed in both years, with average profit below global benchmarks — confirming their position as underperforming markets.

## Which products yield high revenue but low profit margins?



This scatter plot visualizes individual product performance by plotting the Average Gross Sales (\$) on the x-axis and Average Profit Margin (%) on the y-axis, segmented by product and segment. The chart also provides tooltips for average profit and total units sold, enabling a nuanced view of profitability drivers across products.

## Velo

Velo stands out with exceptionally high gross sales (\$512K) and the highest number of units sold (23,907.50) among its peers. However, it operates on a low average profit margin of just 7.28%, indicating a volume-driven strategy. Despite thin margins, it still generates strong profits due to scale — a typical high-volume, low-margin success.

## Montana, Carretera, Adalta

These products fall into the high-margin, lower-sales quadrant, suggesting that they operate on premium pricing or have efficient cost structures. Their performance may not drive top-line revenue, but they are likely to contribute stable and efficient profits. These products could benefit from targeted marketing to increase sales while maintaining profitability.

## Paseo, Amarilla

Positioned near the middle, these products have balanced gross sales and mid-range margins, implying they are consistent contributors with room for either margin or volume improvement. Their moderate performance makes them dependable but not standout.

## Bottom-Left Cluster (e.g., low-margin, low-sales products)

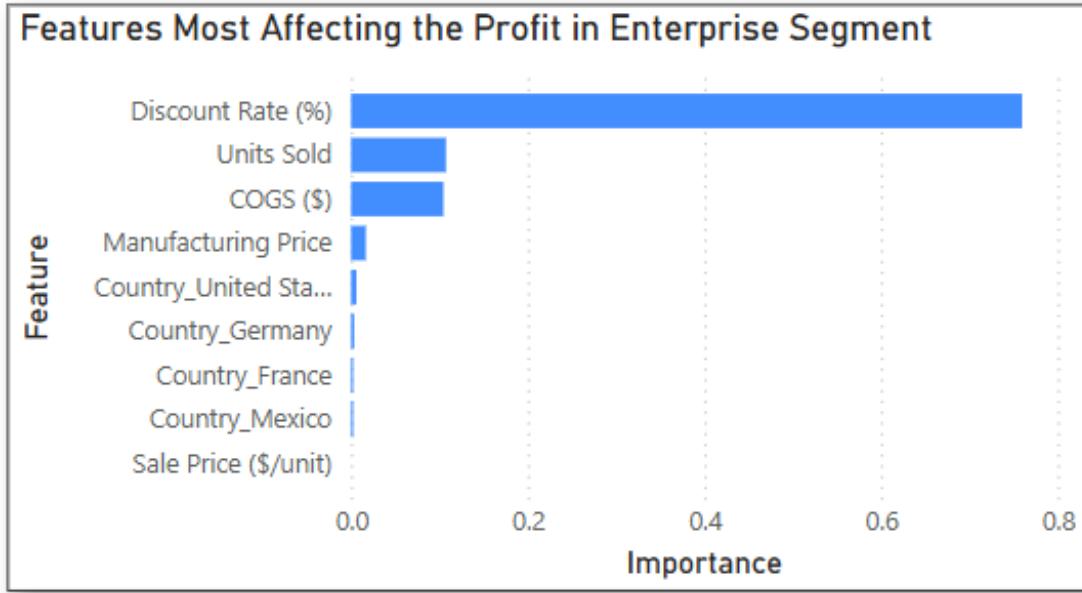
Products in this quadrant (low on both axes) are likely underperforming, offering neither volume nor profitability. These may require cost reduction, repositioning, or possible phase-out depending on their strategic relevance.

## Loss Analysis: Enterprise Segment



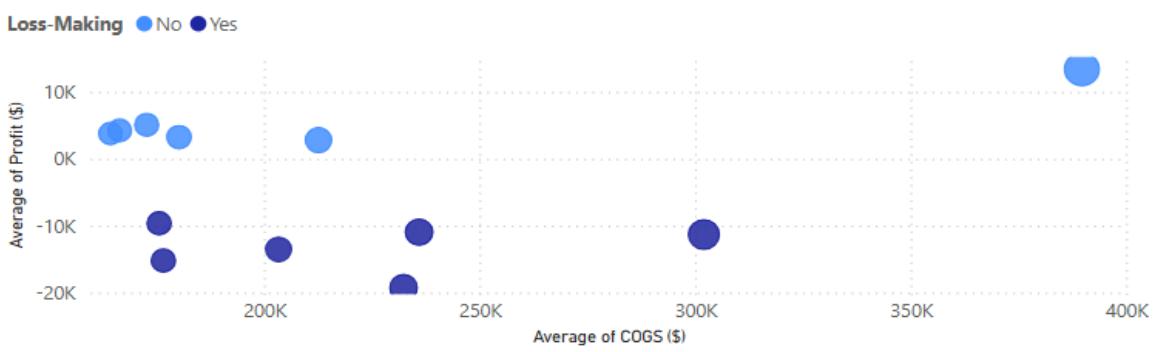
This dashboard uncovers key reasons for the -\$614K total loss in the Enterprise segment, with major regional losses seen in Mexico (-\$120K), Canada (-\$121K), and the USA (-\$175K). The Sankey diagram and discount band analysis reveal that high discounting and high COGS are heavily associated with these losses.

- Impact of Discounting: Products in the "High" discount band have an average loss of - \$18,859.66, compared to a positive average of \$10,166.67 in the "None" band. This suggests aggressive discounting erodes margins without delivering enough volume to compensate.
- Problematic Products: The bar chart for Profit Margin by Product shows deeply negative margins across all Enterprise products—Carretera (-101.85%), Vtt (-58%), Amarilla (-55%), and Montana (-54.33%) are the worst performers. These products are being sold below cost, or with such steep discounts that profits are reversed.
- COGS vs. Sale Price: The final chart shows that COGS (\$) is consistently higher than sale price across all products, reinforcing that—Sale prices are not covering costs and there is no profitability buffer, even with strong unit sales in some products.



The feature importance chart shows discount rate (%) as the most influential factor on profit — far outweighing units sold or COGS. This indicates that excessive discounting is the main reason for losses, making the pricing strategy a critical area for intervention.

Average of COGS (\$), Average of Profit (\$), Average of Units Sold, Average of Discount Rate (%) and Average of Sale Price (\$/unit) by Product and Loss-Making



The scatter plot shows that products with higher average COGS and lower profit fall into the loss-making category, confirming that COGS levels also play a secondary role. However, even products with moderate COGS (e.g., Vtt, Velo) still incur losses when discounts is high.

Product	Average of Profit (\$)	Average of Predicted Profit	Average of Discount Rate (%)	Average of COGS (\$)	Sale Price (\$/unit)	Loss-Making
Amarilla	13,393.75	13,375.89	2.00	389730.00	125	No
Carretera	2,799.38	4,128.04	2.00	212640.00	125	No
Montana	4,238.33	3,869.48	2.25	166530.00	125	No
Paseo	3,234.06	3,216.89	2.17	180280.00	125	No
Velo	5,086.72	5,086.46	1.80	172785.00	125	No
Vtt	3,776.25	3,785.22	1.67	164380.00	125	No
Amarilla	-13,548.89	-12,700.48	10.00	203380.00	125	Yes
Carretera	-19,259.17	-18,670.04	11.83	232375.00	125	Yes
Montana	-11,305.25	-11,162.25	8.00	302040.00	125	Yes
Paseo	-10,958.98	-11,062.24	8.64	235974.55	125	Yes
Velo	-9,650.48	-9,369.79	8.92	175670.77	125	Yes
Vtt	-15,217.50	-15,574.34	11.88	176655.00	125	Yes

- Products such as Amarilla, Carretera, Montana, Paseo, Velo, and Vtt appear in both profit and loss rows, indicating variation by transaction or region.
- All loss-making entries show very high average discount rates (8%–11.8%), compared to 2% or less in profitable cases — again reinforcing the impact of discounts on profit erosion.
- All products have the same sale price (\$125/unit), indicating a one-size-fits-all pricing model that does not reflect variations in COGS or demand. This is likely to contribute to losses, especially when high COGS and high discounts are applied together.

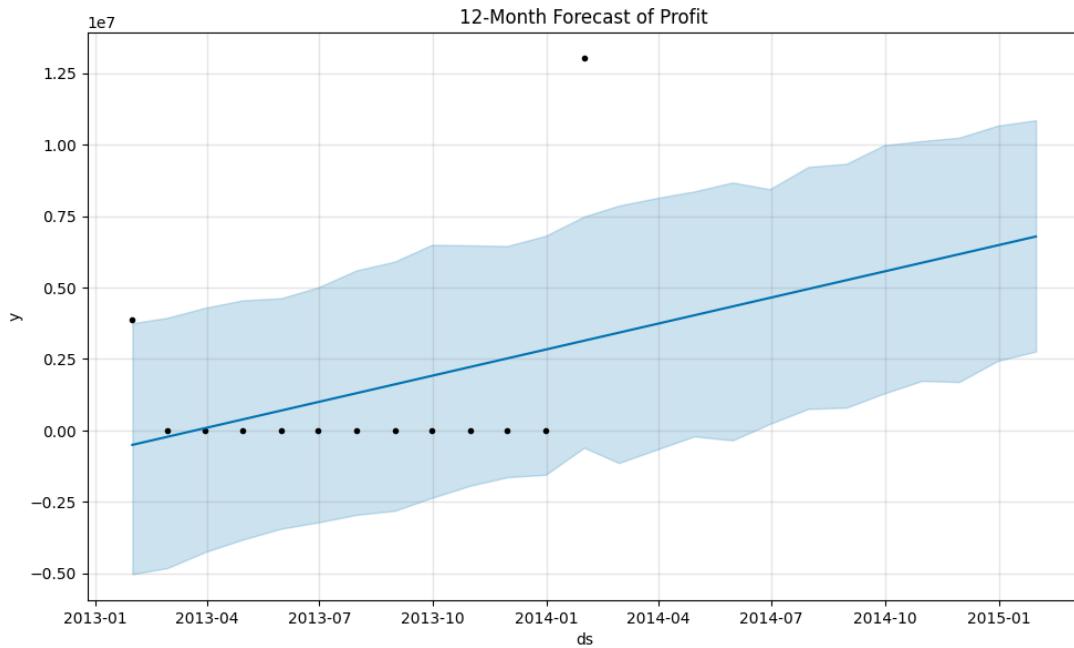
The Enterprise segment is unprofitable due to excessive discounts, high COGS, and uncompetitive pricing. To improve Enterprise segment profitability, start by **rationalizing discounts**, i.e., limit high discounts and apply them selectively to high-margin or overstocked products, while testing price sensitivity. **Revise pricing** by repricing loss-making items like Carretera and Vtt to cover costs and apply value-based pricing. **Optimize costs** by identifying inefficiencies in sourcing or logistics and exploring supplier consolidation. Finally, **review the product portfolio** to phase out consistently unprofitable products and focus on those with stronger ROI potential.

Can we forecast monthly sales or profits for the next year?

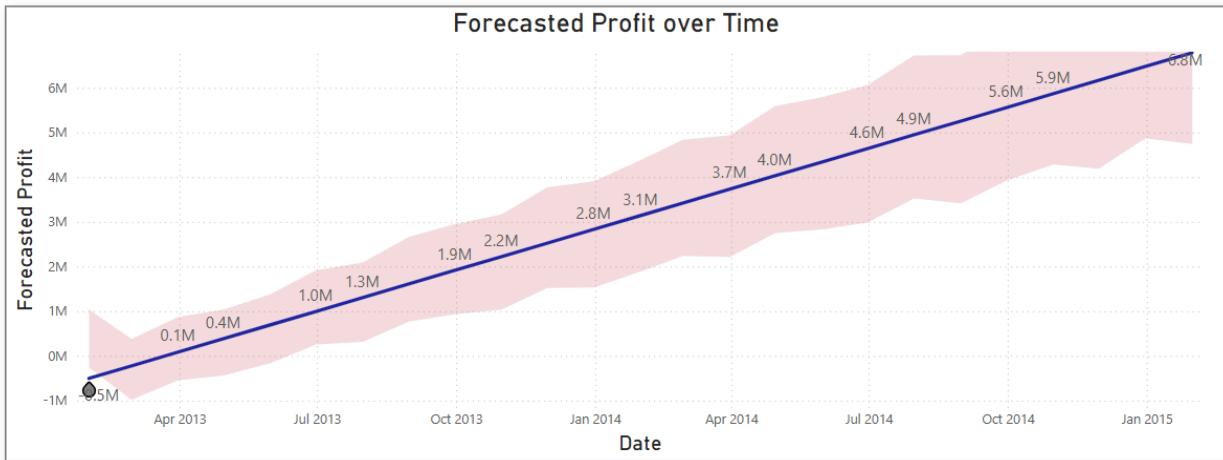
I wanted to use python and machine learning to forecast profits.

To project future profit trends, I used the Prophet library in Python, which is designed for time-series forecasting. The steps were:

1. Data Preparation: I loaded and grouped profit data (Profit (\$)) from financials\_processed.csv by month-end using Pandas, renaming the columns to fit Prophet's required format (ds for date, y for value).
2. Model Fitting: I created and trained a Prophet() model on the monthly aggregated profit data.
3. Forecast Generation: Prophet generated predictions (yhat) along with lower and upper bounds (yhat\_lower, yhat\_upper) to form the confidence interval.
4. Export & Visualization: The results were saved to forecast\_profit.csv, then imported into Power BI to build the “Forecasted Profit over Time” chart.



The Prophet model forecasts consistent profit growth, reaching over \$6.3M by Jan 2015. The narrowing confidence interval and upward slope indicate dependable and accelerating financial performance over time, assuming business trends remain stable.



This chart shows a **forecasted profit trend from Jan 2013 to Jan 2015**, created by importing a time-series dataset generated in Python. The visualization shows a clear upward slope, indicating strong expected profit growth — from **-\$0.5M** in early 2013 to **\$6.3M+** by the start of 2015. The shaded red area represents the **confidence interval**, accounting for variance in the forecast. As time progresses, the forecast becomes more optimistic and stable, implying a solid positive outlook if existing business conditions persist.