In [4]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Data Import"

```
"We begin by loading the financial dataset for analysis
```

In [5]:
```
pd.read_csv(r"C:\Users\HP\Documents\Financials.csv")
```

Out[5]:

	Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	
0	Government	Canada	Carretera	None	$1,618.50	$3.00	$20.00	$32,370.00	$-	$32,370.00	$16,185.00	$1
1	Government	Germany	Carretera	None	$1,321.00	$3.00	$20.00	$26,420.00	$-	$26,420.00	$13,210.00	$1
2	Midmarket	France	Carretera	None	$2,178.00	$3.00	$15.00	$32,670.00	$-	$32,670.00	$21,780.00	$1
3	Midmarket	Germany	Carretera	None	$888.00	$3.00	$15.00	$13,320.00	$-	$13,320.00	$8,880.00	$
4	Midmarket	Mexico	Carretera	None	$2,470.00	$3.00	$15.00	$37,050.00	$-	$37,050.00	$24,700.00	$1
...	...	...	...	...	...	...	...	...	...	...	...	
695	Small Business	France	Amarilla	High	$2,475.00	$260.00	$300.00	$7,42,500.00	$1,11,375.00	$6,31,125.00	$6,18,750.00	$1
696	Small Business	Mexico	Amarilla	High	$546.00	$260.00	$300.00	$1,63,800.00	$24,570.00	$1,39,230.00	$1,36,500.00	$
697	Government	Mexico	Montana	High	$1,368.00	$5.00	$7.00	$9,576.00	$1,436.40	$8,139.60	$6,840.00	$
698	Government	Canada	Paseo	High	$723.00	$10.00	$7.00	$5,061.00	$759.15	$4,301.85	$3,615.00	
699	Channel Partners	United States of America	VTT	High	$1,806.00	$250.00	$12.00	$21,672.00	$3,250.80	$18,421.20	$5,418.00	$1

700 rows × 16 columns

In [ ]:
```
### Assigning data to a dataframe (finance_df)
```

In [6]:
```
finance_df=pd.read_csv(r"C:\Users\HP\Documents\Financials.csv")
```

In [ ]:
```
"## Initial Glimpse"
    "Previewing the information(info), shape and missing values(isnull().sum() of of the dataset to understand the available c
```

In [7]:
```
finance_df.info
```

Out[7]:
```
<bound method DataFrame.info of              Segment                  Country    Product  Discount Band  \
0          Government                   Canada   Carretera          None   
1          Government                  Germany   Carretera          None   
2           Midmarket                   France   Carretera          None   
3           Midmarket                  Germany   Carretera          None   
4           Midmarket                   Mexico   Carretera          None   
..                ...                      ...         ...           ...   
695    Small Business                   France    Amarilla          High   
696    Small Business                   Mexico    Amarilla          High   
697        Government                   Mexico     Montana          High   
698        Government                   Canada       Paseo          High   
699   Channel Partners  United States of America         VTT          High   

     Units Sold   Manufacturing Price   Sale Price     Gross Sales   \
0     $1,618.50                 $3.00       $20.00      $32,370.00   
1     $1,321.00                 $3.00       $20.00      $26,420.00   
2     $2,178.00                 $3.00       $15.00      $32,670.00   
3       $888.00                 $3.00       $15.00      $13,320.00   
4     $2,470.00                 $3.00       $15.00      $37,050.00   
..          ...                   ...          ...             ...   
695   $2,475.00               $260.00      $300.00    $7,42,500.00   
696     $546.00               $260.00      $300.00    $1,63,800.00   
697   $1,368.00                 $5.00        $7.00       $9,576.00   
698     $723.00                $10.00        $7.00       $5,061.00   
699   $1,806.00               $250.00       $12.00      $21,672.00   

         Discounts          Sales            COGS         Profit         Date  \
0              $-       $32,370.00      $16,185.00    $16,185.00   01/01/2014   
1              $-       $26,420.00      $13,210.00    $13,210.00   01/01/2014   
2              $-       $32,670.00      $21,780.00    $10,890.00   01/06/2014   
3              $-       $13,320.00       $8,880.00     $4,440.00   01/06/2014   
4              $-       $37,050.00      $24,700.00    $12,350.00   01/06/2014   
..             ...             ...             ...           ...          ...   
695   $1,11,375.00    $6,31,125.00    $6,18,750.00    $12,375.00   01/03/2014   
696     $24,570.00    $1,39,230.00    $1,36,500.00     $2,730.00   01/10/2014   
697      $1,436.40       $8,139.60       $6,840.00     $1,299.60   01/02/2014   
698        $759.15       $4,301.85       $3,615.00       $686.85   01/04/2014   
699      $3,250.80      $18,421.20       $5,418.00    $13,003.20   01/05/2014   

     Month Number  Month Name   Year  
```

```
0               1     January   2014
1               1     January   2014
2               6        June   2014
3               6        June   2014
4               6        June   2014
..            ...         ...    ...
695             3       March   2014
696            10     October   2014
697             2    February   2014
698             4       April   2014
699             5         May   2014

[700 rows x 16 columns]>
```

In [8]:
```
finance_df.shape
```

Out[8]:
```
(700, 16)
```

In [9]:
```
finance_df.isnull().sum()
```

Out[9]:
```
Segment                  0
Country                  0
 Product                 0
 Discount Band           0
 Units Sold              0
 Manufacturing Price     0
 Sale Price              0
 Gross Sales             0
 Discounts               0
  Sales                  0
 COGS                    0
 Profit                  0
Date                     0
Month Number             0
 Month Name              0
Year                     0
dtype: int64
```

In [ ]:
```
### Calliing out te names of each column in the dataset
```

In [10]:
```
finance_df.columns = finance_df.columns.str.strip()
```

In [11]:
```
print(finance_df.columns)
```

```
Index(['Segment', 'Country', 'Product', 'Discount Band', 'Units Sold',
       'Manufacturing Price', 'Sale Price', 'Gross Sales', 'Discounts',
       'Sales', 'COGS', 'Profit', 'Date', 'Month Number', 'Month Name',
       'Year'],
      dtype='object')
```

In [ ]:
```
### Converting date and month into datetime(feature engineering)
```

In [12]:
```
finance_df["Date"] = pd.to_datetime(finance_df["Date"], errors='coerce')
```

In [13]:
```
finance_df["Month"] = finance_df["Date"].dt.month_name()
```

In [14]:
```
# calculate the sum of product and sales for each product
profit_margin_by_product = finance_df.groupby("Product")[["Profit", "Sales"]].sum()
```

In [15]:
```
print(profit_margin_by_product )
```

```
                                                          Profit  \
Product
 Amarilla     $2,47,500.00  $17,577.00  $21,097.50  $18,990...
 Carretera    $16,185.00  $13,210.00  $10,890.00  $4,440.00...
 Montana      $4,605.00  $22,662.00  $18,990.00  $13,905.00...
 Paseo        $2,920.00  $4,870.00  $22,662.00  $90,540.00 ...
 VTT          $1,00,050.00  $25,542.00  $10,890.00  $4,440....
 Velo         $2,986.00  $9,020.00  $19,449.00  $90,540.00 ...

                                                           Sales
Product
 Amarilla     $9,62,500.00  $23,436.00  $5,27,437.50  $37,9...
 Carretera    $32,370.00  $26,420.00  $32,670.00  $13,320.0...
 Montana      $13,815.00  $30,216.00  $37,980.00  $18,540.0...
 Paseo        $5,840.00  $14,610.00  $30,216.00  $3,52,100....
 VTT          $6,00,300.00  $34,056.00  $32,670.00  $13,320...
 Velo         $10,451.00  $2,25,500.00  $25,932.00  $3,52,1...
```

EXPLANATION OF sum of product and sales for each product:

-Amarilla:

i) Profit figures such as $2,47,500.00,$17,577.00, $21,097.50, and$18,990.00 indicate variability in performance.

ii) This suggests that while one category of transactions might be highly profitable, other segments perform with lower margins.

-Carretera:

i)Values like $16,185.00,$13,210.00, $10,890.00, and$4,440.00 show a relatively steady performance, possibly reflecting consistent pricing strategies or market conditions.

-Montana:

i) Profit numbers range from $4,605.00 to$22,662.00 and $18,990.00 to$13,905.00.

ii) This variation might indicate mixed performance across different segments or time periods.

-Paseo:

i)With profits such as $2,920.00,$4,870.00, $22,662.00, andanexceptional$90,540.00, there appears to be a significant peak in profitability, which could be due to a particularly successful campaign or market strategy in one instance.

-VTT:

i) One entry with $1,00,050.00 standsout, alongwithlower figureslike$25,542.00, $10,890.00, and$4,440.00.

ii)This wide dispersion signals that while VTT can generate very high profits under certain conditions, it also has periods or segments with lower profitability.

-Velo:

i) Profit values of $2,986.00,$9,020.00, $19,449.00, and$90,540.00 illustrate that performance can vary substantially, pointing to potential differences in market effectiveness or cost structures across regions or periods.

-------------------------Sales Overview -Amarilla:

i) The sales figures, including $9, 62, 500.00,$23,436.00, $5,27,437.50, etc., mirror the variability seen in profit, suggesting that high sales do contribute to large revenue streams—although they do not always correlate directly with profit margins.

- Carretera:

ii) With sales numbers like $32, 370.00,$26,420.00, $32, 670.00, and$13,320.00, this product seems to maintain consistent performance, which may indicate stable demand and pricing.

- Montana, Paseo, VTT, and Velo:

iii) Similar to their profit behavior, these products show a spread of sales figures. For instance, despite Paseo having very high sales in one instance ($3,52,100.00), its profits are not uniformly high, hinting at possible cost issues or discount strategies that impact the margin.

--------------------------Key Insights

```
Variable Product Performance:
```

i) The range of profit and sales figures across products like Amarilla, Paseo, and VTT suggests that different factors (such as regional market dynamics, seasonal trends, or discount policies) may significantly influence performance.

```
Profit vs. Sales Relationship:
```

-While some products achieve high sales volumes, this does not necessarily translate into uniformly high profits. For example, Paseo’s high sales are paired with both modest and very high profits. This discrepancy can point to varying cost structures or inconsistent discounting practices that need further investigation.

Opportunities for Targeted Strategies: -Examining these trends in detail might reveal opportunities to optimize pricing, improve cost management, or reallocate resources to enhance the most profitable segments. For instance, a focused review on why certain entries (like the standout profit figures in VTT and Paseo) perform so well can provide clues to scale similar practices across other segments.

Actionable Recommendations:

i) Investigate Outliers: Focus on the high-profit outliers for Paseo and VTT to understand the drivers behind exceptional performance.

ii) Cost Management: Analyze cases where high sales do not convert into high profits to streamline operations or adjust discount strategies.

ii) Market Segmentation: Consider more granular segmentation that could help explain the variability and support targeted marketing or sales initiatives.

In [16]:
```
# Clean monetary columns
monetary_columns = [
    "Units Sold",
    "Manufacturing Price",
    "Sale Price",
    "Gross Sales",
    "Discounts",
    "Sales",
    "COGS",
    "Profit"
]

for col in monetary_columns:
    finance_df[col] = (
        finance_df[col]
        .astype(str)
        .str.replace('[\$,]', '', regex=True)
        .str.replace('[\(\)]', '', regex=True)  # Remove parentheses
        .str.replace('-', '0')  # Replace '-' with '0'
        .str.strip()
        .replace('', '0')  # Replace empty strings with '0'
        .astype(float)
    )
```

Product Profit Margin Analysis

In [17]:
```
# Calculate profit margin
profit_margin_by_product = finance_df.groupby("Product").apply(lambda x: (x["Profit"].sum() / x["Sales"].sum())).reset_index(n
```

In [18]:
```
print(profit_margin_by_product)
```

```
     Product  Profit Margin
0   Amarilla       0.172309
1  Carretera       0.165688
2    Montana       0.144749
3      Paseo       0.152631
4        VTT       0.159814
5       Velo       0.140104
```

Explanation of

Product Profit Margin Analysis:

------Key Insights Highest Margin:

i) Amarilla has the highest profit margin (approximately 17.23%). This indicates that, relative to its sales, Amarilla is more efficient in converting revenue into profit. This product may benefit from continued investment or further strategic focus.

------ Competitive Performance:

-Carretera follows closely, with a profit margin of around 16.57%. Its performance is competitive with Amarilla, suggesting that it maintains strong cost controls and pricing strategies.

```
-------Lower Margins:
```

-Montana and Velo show lower margins, at about 14.47% and 14.01%, respectively. These figures could warrant further investigation into cost management or pricing strategies to identify if improvements can be made.

```
----Mid-Range Performers:
```

-Paseo and VTT lie in between the highest and lowest ranges, with margins around 15.26% and 15.98%. This indicates moderate efficiency and room for potential optimization.

```
-----Business Implications
```

Strategic Focus: Products with higher profit margins like Amarilla and Carretera may be prioritized for increased marketing or distribution efforts since they are inherently more profitable.

-----------Cost Optimization: For products such as Montana and Velo that have lower margins, a detailed analysis of cost structures and pricing strategies might uncover opportunities to boost profitability.

-------------Uniformity vs. Variation: Although the profit margins are relatively close (ranging from ~14% to ~17%), even small differences can significantly impact overall profitability, especially when dealing with large revenue volumes. Understanding these nuances can drive targeted actions that improve overall financial performance.

In [ ]:

```
```

segment-level profit data

In [19]:

```
#calculate profit by segment
profit_by_segment = finance_df.groupby("Segment")["Profit"].sum().reset_index()
```

In [20]:

```
print(profit_by_segment)
```

```
            Segment       Profit
0  Channel Partners   1316803.14
1        Enterprise    940096.88
2        Government  11388173.18
3         Midmarket    660103.09
4    Small Business   4143168.50
```

Explanation of segment-level profit data

---------Key Insights:

1. Government Segment

a)Profit: $11,388,173.18

b)This segment leads significantly in profit generation. It suggests that government contracts or operations within the government market are highly profitable.

c)Implication: Consider dedicating more resources or creating specialized strategies to further enhance performance in this area.

2. Small Business Segment

a) Profit: $4,143,168.50

b) As the second-highest profit contributor, small businesses represent an important revenue stream.

c) Implication: Tailored offerings, pricing strategies, and personalized support may further boost performance in this segment.

3. Channel Partners Segment

a) Profit: $1,316,803.14

b) While not as high as government or small business, channel partners still contribute a solid profit figure.

c) Implication: Leveraging partnerships and exploring ways to optimize collaboration with channel partners can potentially drive additional growth.

4. Enterprise Segment

a) Profit: $940,096.88

b) Enterprise-level deals, despite typically larger transaction sizes, show lower aggregated profit compared to other segments in this dataset.

c) Implication: This may indicate higher costs or competitive pricing pressures in the enterprise market, warranting a review of enterprise pricing strategies and cost controls.

5. Midmarket Segment

a) Profit: $660,103.09

b) With the lowest profit among the segments listed, the midmarket segment could be a target for strategic improvements.

c) Implication: Investigating factors such as price sensitivity, discount strategies, or cost structures here might reveal areas for optimization to boost profitability.

--------------------Business Implications:

1. Resource Allocation:

Given the dominance of the Government segment in profit, it may be beneficial to explore further investment or strategic partnerships in this area. Equally, understanding the lower profit margins in the Midmarket and Enterprise segments can help in reallocating marketing or operational resources where they have the most impact.

2. Segment-Specific Strategies:

Each segment shows unique performance characteristics. Tailored strategies—including differentiated pricing models, targeted promotions, or improved cost management—can help improve overall profitability across segments.

3. Growth Opportunities:

For segments like Small Business and Channel Partners, there might be untapped growth potential. Identifying the drivers behind their current performance, then scaling up best practices, can lead to significant incremental profit increases.

In [21]:
```
#calculate the discount and profit
discount_impact = finance_df.groupby("Discount Band")[["Discounts", "Profit"]].mean().reset_index()
```

In [22]:
```
print(discount_impact)
```

```
  Discount Band     Discounts        Profit
0          High  21702.148122  18912.670367
1           Low   5535.473813  38680.360687
2        Medium  12407.215579  24336.478678
3          None      0.000000  32763.301887
```

Analysis of Discount Bands, Discounts, and Profit

------------------Key Insights

1. High Discount Band:

a) When large discounts are applied (totaling about $21,702), profitisrelativelylow($(18,912).

b) Implication: Offering very high discounts significantly reduces profit margins; thus, aggressive discounting might be counterproductive unless it drives substantially higher sales volumes.

2)Low Discount Band:

a)Although a small amount is discounted (around $5, 535), thisbandgeneratesthehighestprofit($38,680).

b)Implication: A modest discount strategy appears to yield the best balance between incentivizing sales and maintaining strong profitability. It suggests that slight price concessions can boost volume without significantly eroding margins.

3)Medium Discount Band:

a) With moderate discount totals (about $12, 407), theprofitislowercomparedtothelowdiscountband, at$24,336.

b) Implication: Intermediate discount levels may still dilute profit, indicating there may be an optimal threshold below which discounts are beneficial.

4)No Discount Applied:

a) In the absence of any discount, profit is \$32,763, which is strong, but slightly lower than the low discount band.

b) Implication: Not offering any discount at all maintains the price level but might miss opportunities to stimulate additional sales that can boost profit further.

-------------Business Implications

1. Optimizing Discount Strategies:

The evidence suggests that small discounts (the low discount band) can enhance profitability by striking a balance between driving sales and preserving margins. It may be worthwhile to focus on maintaining minimal discounts rather than aggressive price cuts.

2. Cost vs. Volume Trade-Off:

While high discounts reduce the profit margin per unit, the low discount band shows that limited discounting can effectively stimulate demand without overly sacrificing profit.

3. Review of Pricing Tactics:

Businesses may consider analyzing pricing tactics to ensure discounts are set at levels that optimize overall profitability. Regularly reviewing the impact of discount strategies can reveal if adjustments are needed to improve financial performance.

In [23]:
```
#calculate Profit by Country
profit_by_country = finance_df.groupby("Country")["Profit"].sum().sort_values(ascending=False).head(10).reset_index()
```

In [24]:
```
print(profit_by_country)
```

```
                    Country      Profit
0                    France  4032488.29
1                   Germany  3961381.32
2                    Canada  3858206.39
3  United States of America  3397345.68
4                    Mexico  3198923.11
```

explanation of Country level Profit Analysis

-------Key Insights

1. Top Performer – France:

------France leads with a profit of approximately $4,032,488. This strong performance could be attributed to favorable market conditions, effective pricing strategies, or a robust demand for products.

2. Competitive Markets:

------Germany and Canada follow closely, with profits nearing or exceeding $3.8 million. This indicates that these markets are also performing well and contribute significantly to overall profitability.

-------United States and Mexico:

-Although the United States and Mexico have slightly lower profit figures compared to the top three countries, their contributions remain substantial.

-The lower profit in the United States might point to higher competition or increased cost structures compared to European markets, while Mexico's position could highlight opportunities for growth or require further review of market strategies.

-------Business Implications --Resource Allocation: Given the strong performance in France, it may be advantageous to further invest in marketing and distribution channels in this region.

---Tailored Strategies: For markets with slightly lower performance (United States and Mexico), consider reviewing pricing tactics, cost management, or promotional activities to boost profitability further.

----Balanced Expansion: With multiple regions showing robust financial results, it may also be worthwhile to explore cross-market strategies to leverage best practices from the top-performing regions

In [25]:
```
#calculate Cost of goods sold vs Profit
cogs_vs_profit = finance_df[["COGS", "Profit"]]
```

In [26]:
```
print(cogs_vs_profit)
```

```
         COGS    Profit
0     16185.0  16185.00
1     13210.0  13210.00
2     21780.0  10890.00
3      8880.0   4440.00
4     24700.0  12350.00
..        ...       ...
695  618750.0  12375.00
696  136500.0   2730.00
697    6840.0   1299.60
698    3615.0    686.85
699    5418.0  13003.20

[700 rows x 2 columns]
```

Analysis of COGS and Profit

--------------Key Insights --Variability in Profit Margin Ratios:

a) In some cases, such as the first two rows, profit is equal to the COGS. This suggests that for these transactions, the pricing or revenue might be structured such that the generated profit mirrors the cost, indicating an extremely favorable or possibly a one-off scenario.

b) In other cases (e.g., rows 2, 3, and 4), profit is approximately half of COGS. This variability indicates that the profit margin differs significantly from transaction to transaction, which might be due to different discount rates, pricing strategies, or cost structures.

-----High COGS vs. Low Profit:

a) Toward the later rows, such as where COGS is extremely high (e.g., 618,750.0), the corresponding profit is relatively low (e.g., 12,375.00).

b) This drop in the profit-to-COGS ratio signals potential inefficiencies in cost management or aggressive discounting strategies that lower profit margins when the cost base is large.

------Outlier Transactions:

a) Some rows exhibit unusual behavior; for instance, in the last row the COGS is 5,418.0 while profit is 13,003.20. Such outliers prompt a deeper look to understand if there are special cases, pricing errors, or one-time adjustments that could be skewing the data.

---------Business Implications

1. Margin Management:

The varying profit-to-COGS ratios indicate that a standardized approach might not apply across all products or transactions. Identifying why some transactions generate a higher profit margin can guide targeted strategies for improving overall profitability.

2. Cost Control and Pricing Strategy:

For transactions with a low profit relative to high COGS, reviewing cost control measures or re-evaluating pricing strategies may be essential. A detailed analysis might reveal opportunities to adjust pricing, reduce costs, or modify discounting practices to enhance margins.

3. Further Segmentation:

To better understand these variances, it may be useful to segment the data further by factors such as product type, region, or discount band. Such segmentation can uncover patterns that are not immediately obvious in the aggregate data, leading to more tailored recommendations.

In [ ]:

In [ ]:
```
### Monthly profit Comparison from JAanuary 1, 2013 to January 1, 22014
```

In [27]:
```
monthly_profit = finance_df.groupby(finance_df["Date"].dt.to_period("M"))["Profit"].sum().reset_index()
```

In [28]:
```
print(monthly_profit)
```

```
      Date       Profit
0  2013-01   4323139.51
1  2014-01  14125205.28
```

Explanation Monthly profit Comparison from JAanuary 1, 2013 to January 1, 22014

-------Key Insights 1)Significant Year-over-Year Growth: There was a substantial increase in profit from January 2013 to January 2014, with profits more than tripling. This growth indicates a strong upward trend in the company's financial performance during this period.

2. Potential Contributing Factors:

Several factors could have contributed to this growth, including:

a) Increased Revenue: A rise in sales or service income could have boosted profits.

b) Cost Management: Improved efficiency and cost-cutting measures may have enhanced profit margins.

c) Market Expansion: Entering new markets or expanding product lines might have contributed to higher profits.

-----Business Implications

1. Strategic Planning:

Understanding the drivers behind this profit growth can inform future business strategies. Identifying successful initiatives from this period can help replicate positive outcomes.

2. Investment Opportunities:

The significant profit increase may attract investors and provide opportunities for reinvestment into the business for further growth.

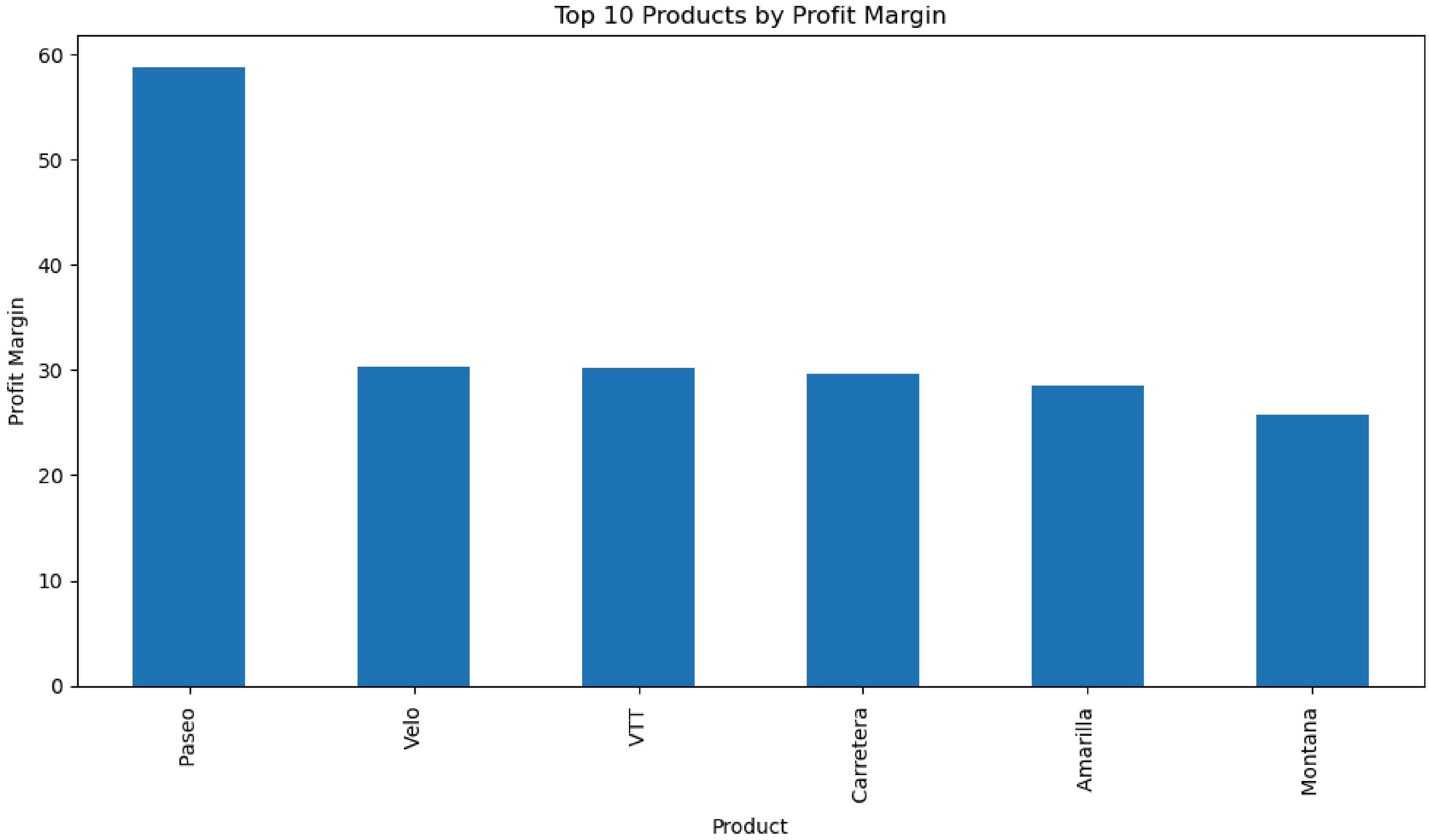
3. Performance Benchmarking:

January 2014's profit figures can serve as a benchmark for evaluating future performance and setting financial targets.

In [29]:
```
# Calculate the profit margin for each product
finance_df['Profit Margin'] = finance_df['Profit'] / finance_df['Sales']
```

In [30]:
```
#the top 10 products by profit margin
top_10_products = finance_df.groupby("Product")["Profit Margin"].sum().nlargest(10)
```

In [31]:
```
# visualise top 10 products by profit margin
plt.figure(figsize=(10,6))
top_10_products.plot(kind="bar")
plt.title("Top 10 Products by Profit Margin")
plt.xlabel("Product")
plt.ylabel("Profit Margin")
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



Explanation of Top 10 products by Profit Margin

----key Observations:

1. Paseo leads significantly

Paseo stands out with a profit margin close to 60%, making it by far the most profitable product in this group. This suggests it's either priced very well, has low costs, or both.

2. Velo and VTT perform similarly

Both Velo and VTT show a profit margin around 30%, indicating consistent profitability. These could be considered reliable mid-range performers in terms of profitability.

3. Carretera, Amarilla, and Montana

These products show moderately lower profit margins (between ~26–30%). While still profitable, they lag behind Paseo and could benefit from better cost management or pricing optimization.

----Business Implications: a) Double Down on Paseo Given its high profitability, Paseo may be a strategic focus area. Consider boosting production, marketing, and availability.

b) Benchmarking Opportunities Investigate what makes Paseo so profitable and see if those strategies can be applied to lower-margin products like Montana.

c) Portfolio Optimization Products with consistently lower margins should be reviewed. Are they essential to the product line? Can they be improved or repositioned?

C) Data-Driven Decision-Making This visualization enables a quick, intuitive grasp of product performance, making it easier to inform pricing, promotion, and inventory strategies.

In [32]:
```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
import numpy as np
```

In [33]:
```
# Select features and target
features = [
```

```
    "Segment", "Country", "Product", "Discount Band", "Units Sold",
    "Sale Price", "COGS"
]
```

In [34]:
```
target = "Profit"
```

In [35]:
```
X = finance_df[features]
y = finance_df[target]
```

In [36]:
```
# Categorical and numeric columns
categorical_cols = X.select_dtypes(include="object").columns.tolist()
numeric_cols = X.select_dtypes(include="number").columns.tolist()
```

In [37]:
```
categorical_transformer = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="most_frequent")),
    ("onehot", OneHotEncoder(handle_unknown="ignore"))
])
```

In [38]:
```
numeric_transformer = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="mean"))
])
```

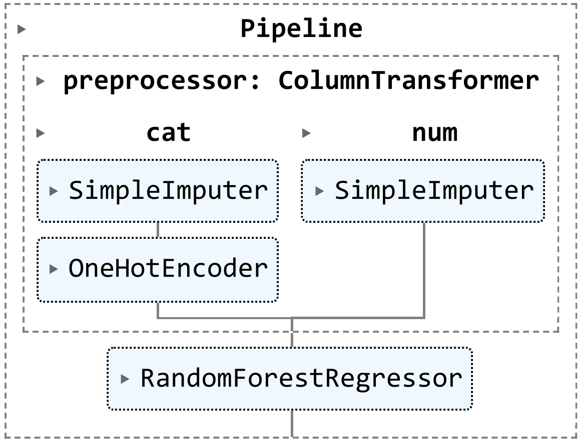
In [39]:
```
preprocessor = ColumnTransformer(transformers=[
    ("cat", categorical_transformer, categorical_cols),
    ("num", numeric_transformer, numeric_cols)
])
```

In [40]:
```
# Build the pipeline with a model
model = Pipeline(steps=[
    ("preprocessor", preprocessor),
    ("regressor", RandomForestRegressor(n_estimators=100, random_state=42))
])
```

In [41]:
```
# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

In [42]:
```
# Train model
model.fit(X_train, y_train)
```

Out[42]:



In [43]:
```python
# Predict and evaluate
y_pred = model.predict(X_test)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
```

In [44]:
```python
print("Model Performance:")
print(f"R² Score: {r2:.2f}")
print(f"MAE: ${mae:,.2f}")
print(f"RMSE: ${rmse:,.2f}")
```

```
Model Performance:
R² Score: 0.92
MAE: $5,357.37
RMSE: $13,337.45
```

------What These Metrics Mean:

1. R^2 Score (Coefficient of Determination)

--0.92 indicates that the model explains 92% of the variance in the target variable (likely profit or sales).

-This is a strong performance, suggesting the model is well-fitted and captures the underlying patterns in the data effectively.

2)MAE (Mean Absolute Error)

-On average, the model's predictions are off by $5,357.37 from the actual values.

-A relatively low MAE means predictions are fairly accurate in day-to-day business decision-making.

3)RMSE (Root Mean Squared Error)

-RMSE penalizes larger errors more heavily. At $13,337.45, this value is higher than MAE, which is expected.

-It provides a more conservative view of model error, highlighting the impact of any large deviations.

------Key Insights: -The high R^2 score confirms that the model captures most of the variation in the data, making it reliable for forecasting or analysis.

-The low MAE indicates the model is consistently accurate, which is crucial for budgeting, planning, or pricing strategies.

NOTE:The gap between RMSE and MAE suggests occasional large errors, which should be reviewed. Consider:

-Investigating outliers or anomalous transactions.

-Applying regularization or robust regression techniques if needed.

------------------Business Implications: -You can confidently use this model to make financial or operational decisions.

-Forecasting future performance, identifying top products, or evaluating promotional strategies can be backed by this level of model accuracy.

-Continual monitoring and retraining with new data will help maintain performance over time.

In [ ]:

```

```

MAKE PREDICTIONS WITH NEW DATA SET

In [45]:

```
import pandas as pd
```

In [46]:

```
# Sample new transactions
new_data = pd.DataFrame([
    {
```

```
        "Segment": "Midmarket",
        "Country": "Germany",
        "Product": "Carretera",
        "Discount Band": "None",
        "Units Sold": 1800,
        "Sale Price": 15.0,
        "COGS": 10500.0
    },
    {
        "Segment": "Government",
        "Country": "Canada",
        "Product": "Montana",
        "Discount Band": "Low",
        "Units Sold": 2500,
        "Sale Price": 20.0,
        "COGS": 17000.0
    },
    {
        "Segment": "Enterprise",
        "Country": "Mexico",
        "Product": "VTT",
        "Discount Band": "Medium",
        "Units Sold": 2200,
        "Sale Price": 18.0,
        "COGS": 14500.0
    }
])
```

In [47]:
```
# Predict profit using the trained model pipeline
predicted_profits = model.predict(new_data)
```

In [48]:
```
# Attach predictions to the new data
new_data["Predicted Profit ($)"] = predicted_profits.round(2)
```

In [49]:
```
# View results
print(new_data)
```

```
       Segment  Country    Product Discount Band  Units Sold  Sale Price  \
0    Midmarket  Germany  Carretera          None        1800        15.0
1   Government   Canada    Montana           Low        2500        20.0
2   Enterprise   Mexico        VTT        Medium        2200        18.0

      COGS  Predicted Profit ($)
0  10500.0               7007.17
1  17000.0              21129.99
2  14500.0              17418.72
```

In [ ]: