

PRES

ESENTATION

DATA



EXECUTIVE SUMMARY

Project Overview

This project leverages SQL, Power BI, and Python to analyze heavy machinery sales data and uncover insights into customer behavior, store performance, sales trends, and product profitability. The goal is to optimize marketing strategies, inventory management, and sales performance to drive revenue growth.

Key Findings

Customer Segmentation & Retention

- The top 10 customers contribute nearly 40% of total revenue, making customer retention a priority.
- A significant portion of customers has a high purchase frequency but moderate spending per transaction, creating opportunities to increase their spending.

Business Impact:

- Implement a loyalty program targeting high-value customers with exclusive offers.
- Use personalized promotions based on purchase behavior to increase spending.
- Identify and re-engage churned customers through targeted outreach.

Sales Channel Performance & Revenue Growth

- Online sales generate 50% higher revenue than in-store purchases, highlighting digital sales as a major driver of growth.
- Some physical stores perform significantly better than others, indicating a need for store-level optimization.

Business Impact:

- Increase investment in online sales via digital marketing, better UI/UX, and AI-driven recommendations.
- Expand high-performing stores and adjust marketing strategies for underperforming locations.
- Improve logistics & supply chain to enhance home delivery & hybrid sales models (e.g., click-and-collect).

Store Efficiency & Regional Expansion

- Stores #102 and #205 consistently rank among the top 10 revenue-generating stores, indicating high demand in certain regions.
- Some stores experience low sales, potentially due to poor location, inventory mismanagement, or weak local demand.

Business Impact:

- Expand store operations in regions with high-performing stores.
- Reallocate inventory from low-revenue stores to high-demand locations.
- Develop localized marketing campaigns to attract more foot traffic to underperforming stores.



Product Performance & Inventory Management

- Road Pavers contribute 25% of total profits, making them the most profitable product category.
- The most purchased products show an uneven distribution, with a few items generating the majority of sales.

Business Impact:

- Increase production & marketing for high-demand products to maximize revenue.
- Bundle slow-moving products with top-sellers to improve inventory turnover.
- Improve inventory forecasting using historical sales trends.

Data-Driven Decision-Making

- Power BI dashboards provide real-time insights, allowing managers to track sales trends dynamically.
- Python-based customer segmentation reveals targeted marketing opportunities for different customer groups.

Business Impact:

- Enable faster decision-making based on live sales data.
- Implement AI-driven demand forecasting to predict future sales trends and optimize stock levels.
- Use predictive analytics models to anticipate customer purchase behavior.



INTRODUCTION



Caterpillar Inc. is an American Fortune 100 corporation that designs, develops, engineers, manufactures, markets, and sells machinery, engines, financial products, and insurance through a global dealer network.

As the world's largest construction equipment manufacturer, Caterpillar operates worldwide serving industries such as:
CONSTRUCTION, MINING, and ENERGY.



THE BUSINESS PROBLEM

Caterpillar faces several challenges that impact sales performance and customer satisfaction such as:

Fluctuating Sales – Sales vary across different stores, regions, and distribution channels, making it difficult to forecast revenue.

Inefficient Marketing – The impact of marketing efforts is unclear, leading to ineffective allocation of resources.

Customer Satisfaction Issues – Frequent delivery delays and supply chain inefficiencies contribute to customer dissatisfaction.

PROJECT OBJECTIVE

This project aims to analyze sales data, customer behavior, and store performance to uncover trends and improve decision-making.

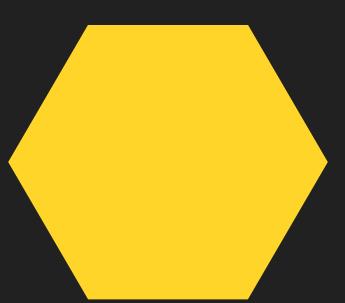
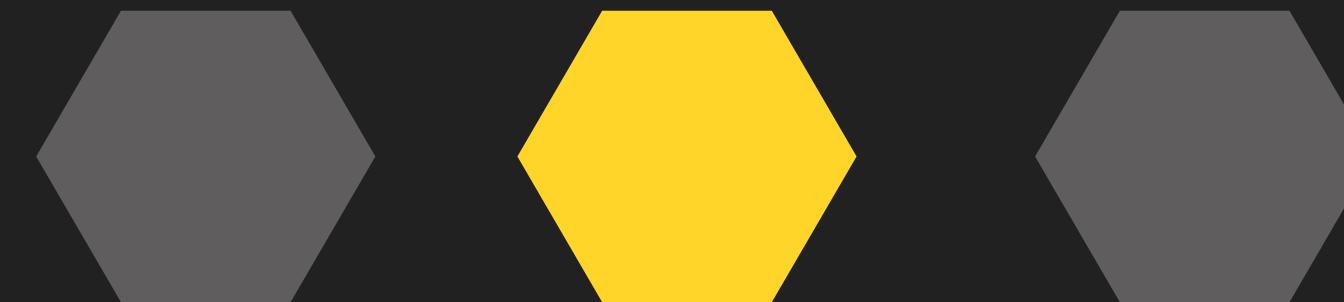
By leveraging SQL, Power BI, and Python, we will:

- **Identify key sales patterns and customer preferences.**
- **Assess the efficiency of different marketing channels.**
- **Provide data-driven insights to enhance operational and strategic decisions.**

This analysis will help Caterpillar optimize its:

- **sales strategy,**
- **improve customer experience, and**
- **drive business growth.**





SQL



5

DATA EXPLORATION

```
--Data Exploration:  
select top 100 * from dbo.Sales$;  
select top 100 * from dbo.Stores$;  
select top 100 * from dbo.Channel$;  
select top 100 * from dbo.Products$;  
select top 100 * from dbo.ProductCategory$;  
select top 100 * from dbo.ProductSubcategory$;  
select top 100 * from dbo.Customers$;  
select * from sys.schemas;
```

Explanation:

- The SELECT TOP 100 * FROM TableName; queries retrieve the first 100 rows from each table to understand their structure, data types, and relationships.
- The sys.schemas query retrieves metadata about database schemas, helping to understand table organization.

Business Impact:

- Understanding the dataset structure ensures data accuracy and integrity before analysis.
- Helps identify key fields needed for sales analysis, customer insights, and marketing optimization.
- Ensures data completeness and consistency across tables, preventing errors in later queries.



REVENUE MILESTONES

--MONTHS WHERE REVENUE EXCEEDED \$50 MILLION

```
SELECT
    FORMAT(dbo.Sales$.TransactionDate, 'MM/yyyy') AS MonthYear,
    SUM(Revenue) AS TotalRevenue
FROM dbo.Sales$
GROUP BY FORMAT(dbo.Sales$.TransactionDate, 'MM/yyyy')
HAVING SUM(Revenue) > 50000000
ORDER BY MonthYear, TotalRevenue DESC;
```

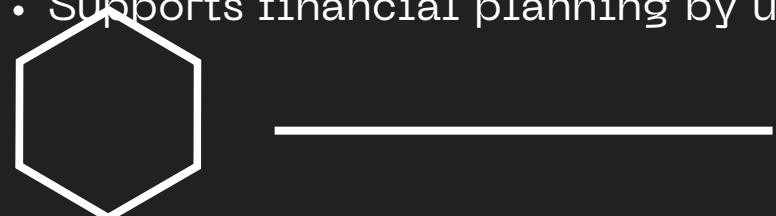
Explanation:

- The query extracts the month and year from TransactionDate using `FORMAT(dbo.Sales$.TransactionDate, 'MM/yyyy')`.
- It groups sales data by month and year and calculates the total revenue using `SUM(Revenue)`.
- The `HAVING` clause filters results to include only months where total revenue exceeded \$50M.
- The results are sorted chronologically by month and in descending order of revenue.

Business Impact:

- Identifying high-revenue months helps recognize seasonal trends and peak sales periods.
- Enables better demand forecasting and inventory management to meet customer needs.
- Helps optimize marketing efforts, allocating resources more effectively during high-sales periods.
- Supports financial planning by understanding revenue fluctuations over time.

	MonthYear	TotalRevenue
1	01/2003	52176104
2	01/2008	50675970
3	03/2007	52313951
4	03/2009	51093708
5	03/2010	50376658
6	03/2011	51591768
7	03/2012	51896311
8	04/2004	50303053
9	05/2002	50847984
10	05/2003	53807859



TOP STORES & CUSTOMERS

```
--Top Stores & Customers
--Identify the highest revenue-generating customers and stores.
SELECT TOP 10
    st.StoreID, c.CustomerName, SUM(s.Revenue) AS TotalRevenue
FROM dbo.Stores$ st
    JOIN dbo.Sales$ s
        ON st.StoreID = s.StoreID
    JOIN dbo.Customers$ c
        ON s.CustomerKey = c.CustomerKey
GROUP BY st.StoreID, st.StoreType, c.CustomerName
ORDER BY TotalRevenue DESC;
```

	StoreID	CustomerName	TotalRevenue
1	BW288	Sheridan Swindin	2485148
2	NT780	Zemfira M Tretyakova	2184229
3	SW244	Myrtle Anespie	1432877
4	CH332	Kell N Grendle	1151831
5	OC714	Taija E Järvinen	1089197
6	NW182	Nestle	1053800
7	RU356	Cordie Arent	1013696
8	VZ342	K'Kehla N Kleevas	950800
9	CH222	Phoebe N Laidler	922422
10	NT780	Ferre T Hoeks	871710

Explanation:

- The query joins the Stores\$ table with Sales\$ and Customers\$ using StoreID and CustomerKey to link sales data with stores and customers and calculates the total revenue using SUM(Revenue).
- The results are grouped by StoreID, StoreType, and CustomerName, ensuring revenue is aggregated correctly.
- The ORDER BY DESC sorts results in descending order, listing the **top 10** revenue-generating stores and customers.

Business Impact:

- Identifying top-performing stores helps optimize inventory allocation and staffing for better service and sales.
- Recognizing high-value customers enables personalized marketing, loyalty programs, and VIP services to improve retention.
- Helps in expanding successful store models and improving performance in underperforming locations.
- Supports strategic decision-making on store expansions, promotions, and customer engagement initiatives.

MOST SUCCESSFUL SALES CHANNEL

```
-- THE MOST SUCCESSFUL SALES CHANNEL IN THE LAST 5 YEARS  
WITH  
Last5Years AS (  
    SELECT YEAR(s.TransactionDate) AS SalesYear,  
        c.Channel AS ChannelName, s.ChannelKey,  
        SUM(s.Qty*s.Price) AS TotalSales  
    FROM Sales$ s JOIN dbo.Channel$ c ON s.ChannelKey = c.ChannelKey  
    WHERE YEAR(TransactionDate) >= (SELECT MAX(YEAR(TransactionDate)) - 4 FROM Sales$)  
    GROUP BY YEAR(TransactionDate), c.Channel, s.ChannelKey),  
RankedChannels AS (  
    SELECT SalesYear, ChannelKey, TotalSales, ChannelName,  
        ROW_NUMBER() OVER (PARTITION BY SalesYear ORDER BY TotalSales DESC) AS Rank  
    FROM Last5Years)  
SELECT SalesYear, ChannelKey, ChannelName, TotalSales  
FROM RankedChannels  
WHERE Rank = 1  
ORDER BY SalesYear;
```

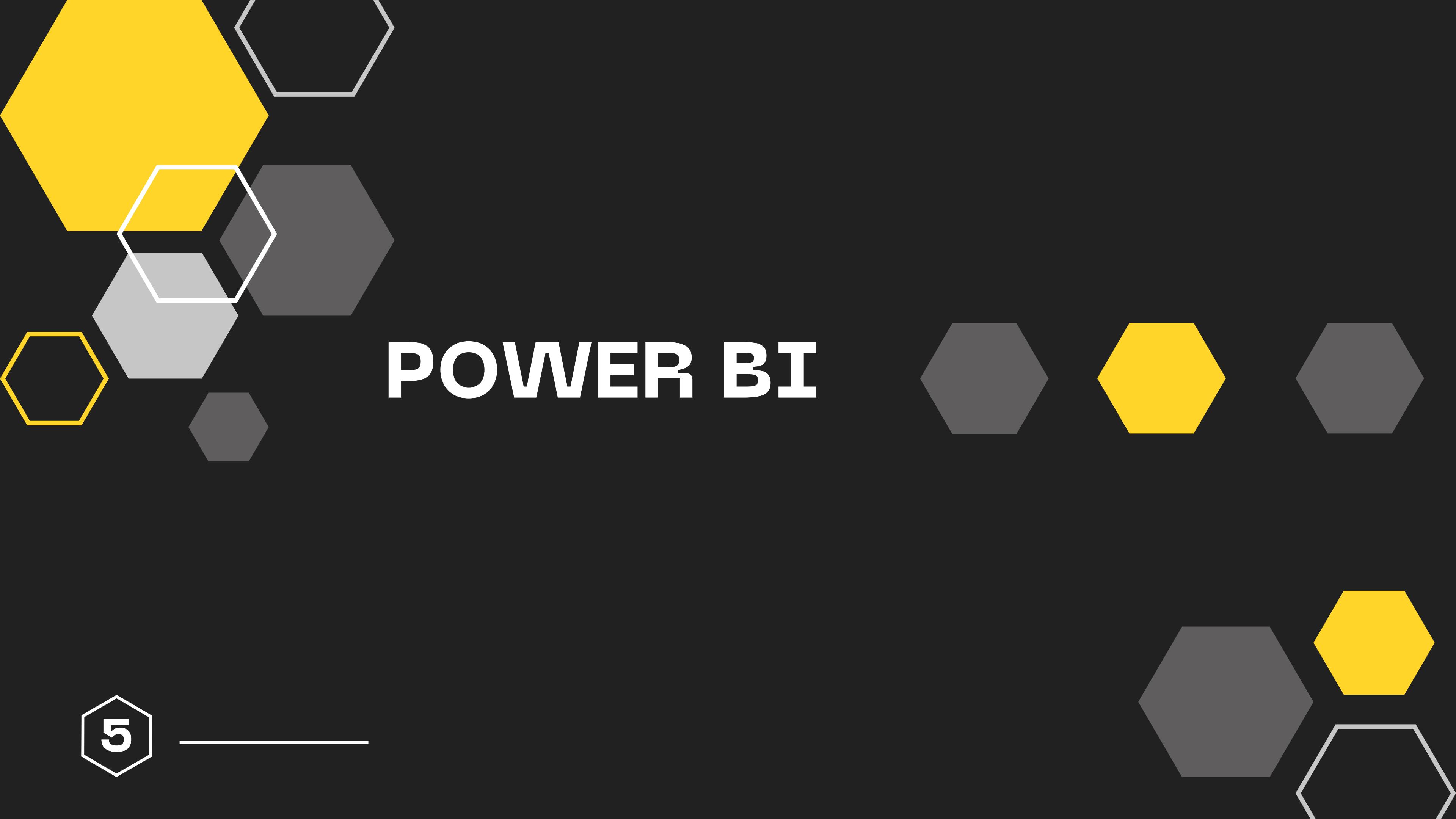
	SalesYear	ChannelKey	ChannelName	TotalSales
1	2009	2	InternetReview	60160335
2	2010	6	SocialMedia	59737313
3	2011	8	Website	60050965
4	2012	3	Magazine	62183810
5	2013	2	InternetReview	759400

Explanation:

- Last5Years CTE:
 - Filters sales data for the last five years by calculating MAX(YEAR(TransactionDate)) – 4.
 - Joins Sales\$ with Channel\$ to get channel names.
 - Calculates total sales per channel per year using SUM(s.Qty * s.Price).
- RankedChannels CTE:
 - Assigns a ranking to each sales channel within a year using ROW_NUMBER() OVER (PARTITION BY SalesYear ORDER BY TotalSales DESC), ensuring the top-performing channel for each year is identified.
- Final Selection:
 - Filters only the highest revenue-generating sales channel per year (Rank = 1).
 - Orders results chronologically to show trends over time.
 -

Business Impact:

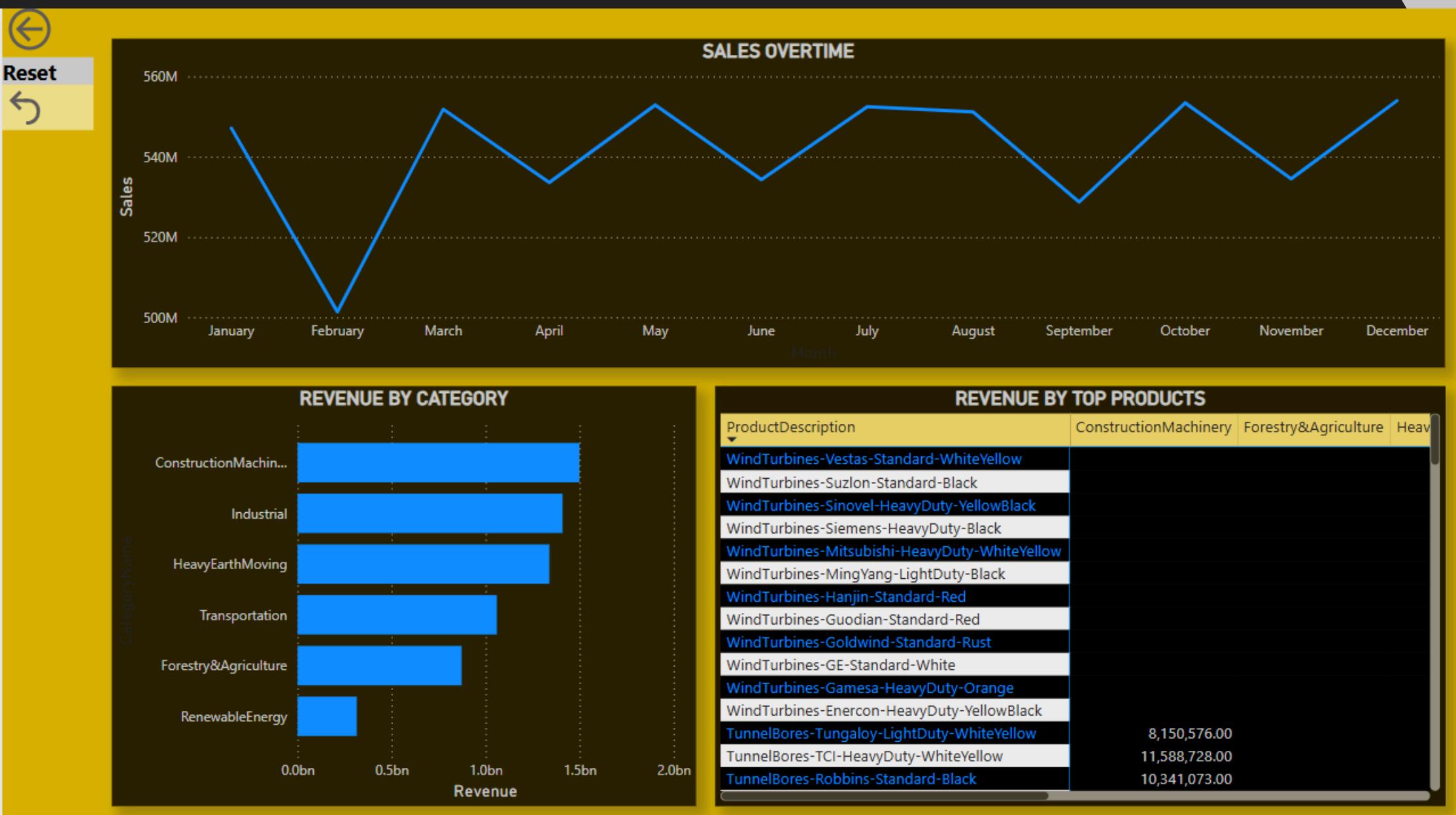
- Helps determine the most profitable sales channels, allowing the company to focus resources on high-performing channels.
- Supports budget allocation by investing in channels with the highest revenue returns.
- Enables sales strategy optimization, such as shifting marketing efforts toward the best-performing platforms.
- Identifies shifts in customer preferences over time, guiding future sales and distribution planning.



POWER BI

5

SALES ANALYSIS



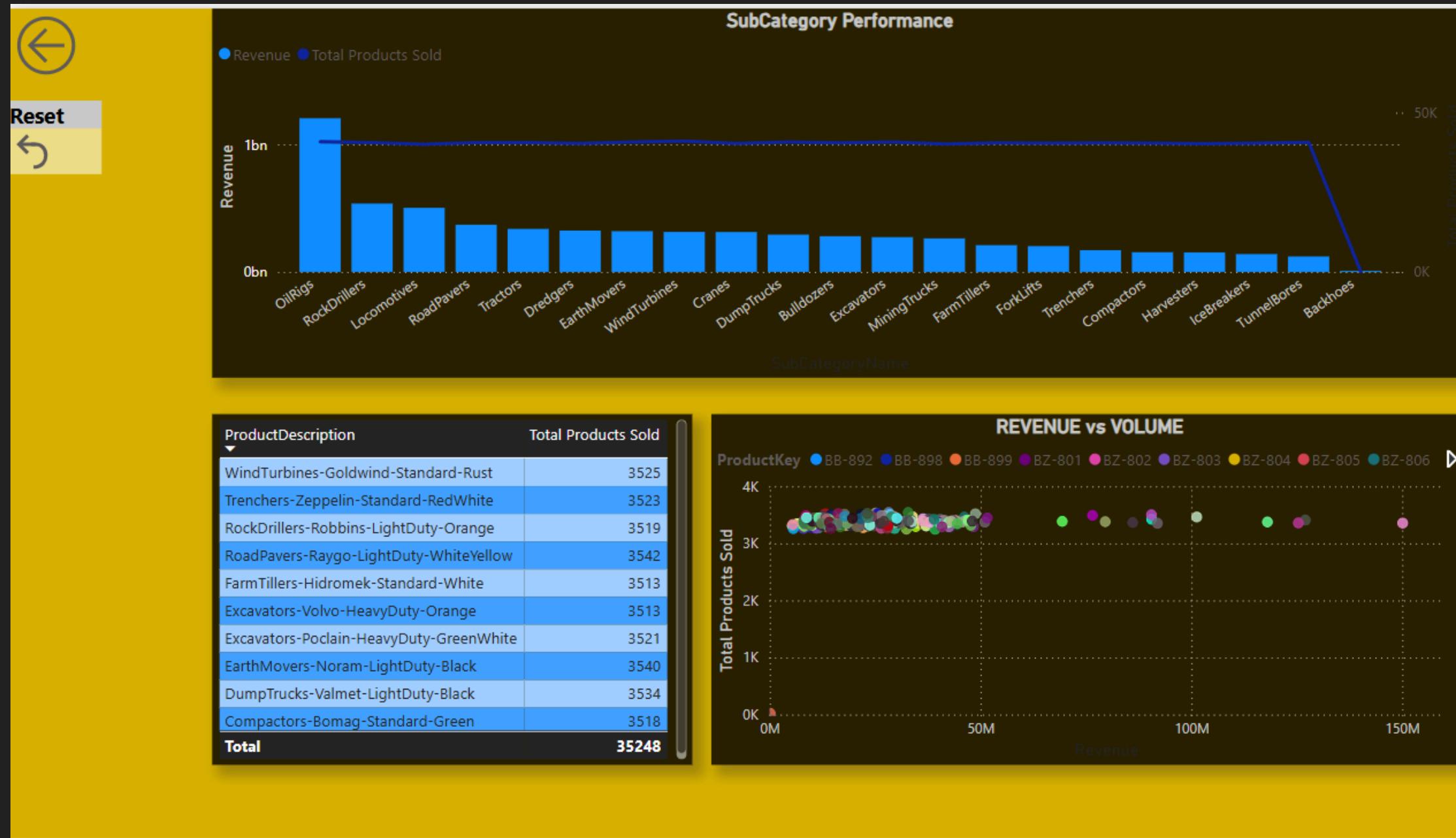
EXPLANATION

1. Line chart is used to show Sales Overtime, creating hierarchy on year, quarter and month, on the x-axis and Sales on the y-axis. With drill down enabled we can analyze sales yearly, quarterly and monthly
2. Revenue By Category gives an idea on which category contributes the most and less
3. Matrix showing Product Description in rows, category name in column and Revenue in values –identify peak sales periods and top-selling products
4. across different categories

BUSINESS IMPACT:

Tracking revenue milestones allows management to understand peak sales periods, which can inform decisions around inventory management, staffing, and promotional activities. In category Construction machinery generate the highest revenue and profits. Sales vary across different months—every year February sales dip happens, a steady growth in all quarters, and years

PRODUCT ANALYSIS



EXPLANATION

1. Subcategory Performance chart using line and stacked column chart, subcategory on x-axis and Revenue on y-axis, Revenue and total product sold as legends, Column chart shows the Revenue and line chart total products sold
2. Table showing top 10 products based on product count column and Product name, and adding Top N filter to the visual to get Top 10 products
3. Scatter Chart showing volume of products and revenue relation

BUSINESS IMPACT:

Inventory management and product production planning based on Most Popular Product that is Wind Turbines This analysis identifies key product categories that are driving sales, allowing the company to allocate resources more effectively and ensure that these categories are well-stocked and promoted

🔍 **Performance Insights:** Pinpoint why certain products succeed (OilRigs subcategory contribute more than 50% of revenue, pricing tiers).

💰 **Profitability:** Focus resources on top sellers to maximize ROI.

🎯 **Customer Alignment:** Tailor offerings to match purchasing behavior.

PRODUCT ANALYSIS

Average Delivery, Total Revenue, Profit, and Units Sold

ProductDescription	Revenue	Profit	Total Products Sold	Average of ShipDays
Bulldozers-AllisChalmers-HeavyDuty-White	46,111,420.00	11,954,300.00	3460	8.00
Bulldozers-BEML-LightDuty-Green	39,643,415.00	11,325,720.00	3395	5.00
Bulldozers-Case-HeavyDuty-White	30,997,980.00	5,999,500.00	3380	11.00
Bulldozers-Caterpillar-LightDuty-RedWhite	14,982,828.00	979,566.00	3486	9.00
Bulldozers-Hitachi-LightDuty-WhiteYellow	18,993,612.00	1,726,692.00	3399	14.00
Bulldozers-IngersollRand-Standard-YellowBlack	24,971,750.00	1,411,962.00	3386	12.00
Bulldozers-International-Standard-Red	15,925,900.00	3,185,850.00	3350	5.00
Bulldozers-JCB-HeavyDuty-Yellow	13,415,436.00	3,327,912.00	3348	15.00
Bulldozers-Komatsu-LightDuty-Orange	22,100,373.00	3,048,669.00	3303	12.00
Bulldozers-Liebherr-LightDuty-GreenWhite	31,800,636.00	2,080,197.00	3339	10.00
Bulldozers-Terex-Standard-Black	11,540,344.00	2,863,000.00	3272	9.00
Bulldozers-Toyota-HeavyDuty-Green	10,327,736.00	398,268.00	3404	9.00
Compactors-AtlasCopco-HeavyDuty-RedWhite	13,513,846.00	2,436,366.00	3398	10.00
Compactors-Bobcat-LightDuty-Green	10,186,953.00	2,410,653.00	3381	8.00
Compactors-Bomag-Standard-Green	13,125,658.00	2,624,428.00	3518	10.00
Compactors-BomMach-LightDuty-Orange	13,467,648.00	2,946,048.00	3288	6.00
Compactors-Caterpillar-HeavyDuty-Yellow	12,393,708.00	2,557,272.00	3356	5.00
Compactors-Hamm-LightDuty-Green	13,952,148.00	2,129,106.00	3423	9.00
Compactors-Honda-HeavyDuty-RedWhite	17,870,690.00	2,055,510.00	3310	7.00
Compactors-IngersollRand-HeavyDuty-Red	10,112,720.00	919,640.00	3320	16.00
Compactors-JCB-Standard-Black	11,302,866.00	537,254.00	3422	7.00
Compactors-Kleemann-LightDuty-Blue	13,036,464.00	1,894,914.00	3402	16.00

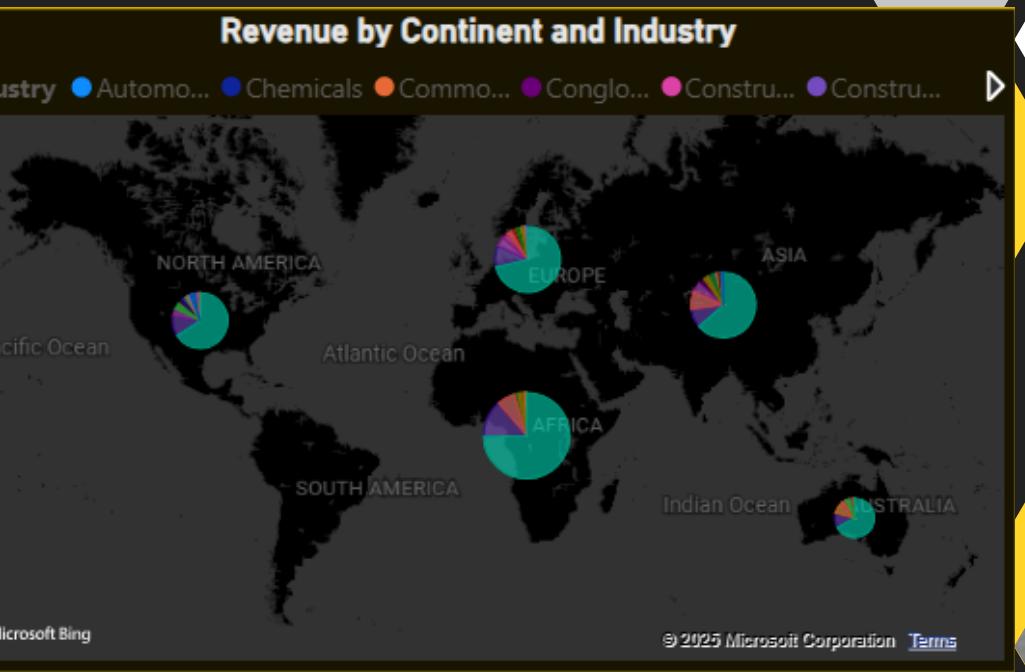
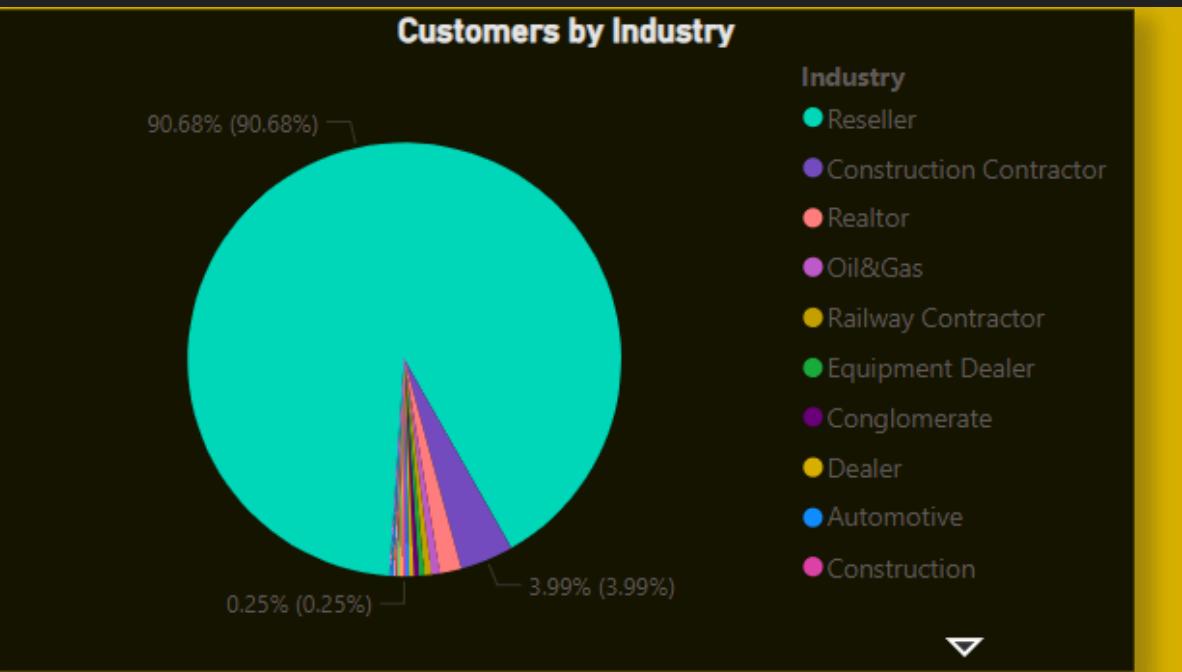
EXPLANATION

Matrix showing correlation between product, Revenue, Profit, Total products Sold and average of Delivery days, data bars are added for clarity

BUSINESS IMPACT:

1. Fix delays for high-impact products: Focus on delayed products with high revenue/profit.
2. Optimize inventory: Stock more of top-selling products with short delivery times.
3. Review product mix: Phase out low-margin products with long delivery times.
4. Leverage product descriptions: Replicate successful keywords/features in new products.
5. Delays may increase operational costs (e.g., expedited shipping), reducing margins.

CUSTOMER SEGMENTATION



CustomerName	Revenue	Profit
Cargill	7,101,360.00	1,197,731.00
Dangote	6,751,782.00	1,146,950.00
Nestle	7,074,448.00	1,147,404.00
Palabora	6,774,794.00	1,125,483.00
Phillips66	6,759,840.00	1,166,602.00
Royal BAM	6,761,341.00	1,129,671.00
Shell	6,813,241.00	1,226,684.00
Statoil	6,740,333.00	1,162,619.00
Walmart	7,131,533.00	1,265,035.00
Woolworths	6,991,173.00	1,239,318.00
Total	68,899,845.00	11,807,497.00

CustomerName	CashAdvance	COD	EFT	Total
Walmart	7,131,533.00			7,131,533.00
Cargill			7,101,360.00	7,101,360.00
Nestle		7,074,448.00		7,074,448.00
Woolworths	6,991,173.00			6,991,173.00
Shell	6,813,241.00			6,813,241.00
Palabora		6,774,794.00		6,774,794.00
Royal BAM		6,761,341.00		6,761,341.00
Phillips66			6,759,840.00	6,759,840.00
Dangote	6,751,782.00			6,751,782.00
Statoil		6,740,333.00		6,740,333.00
Total	27,687,729.00	27,350,916.00	13,861,200.00	68,899,845.00

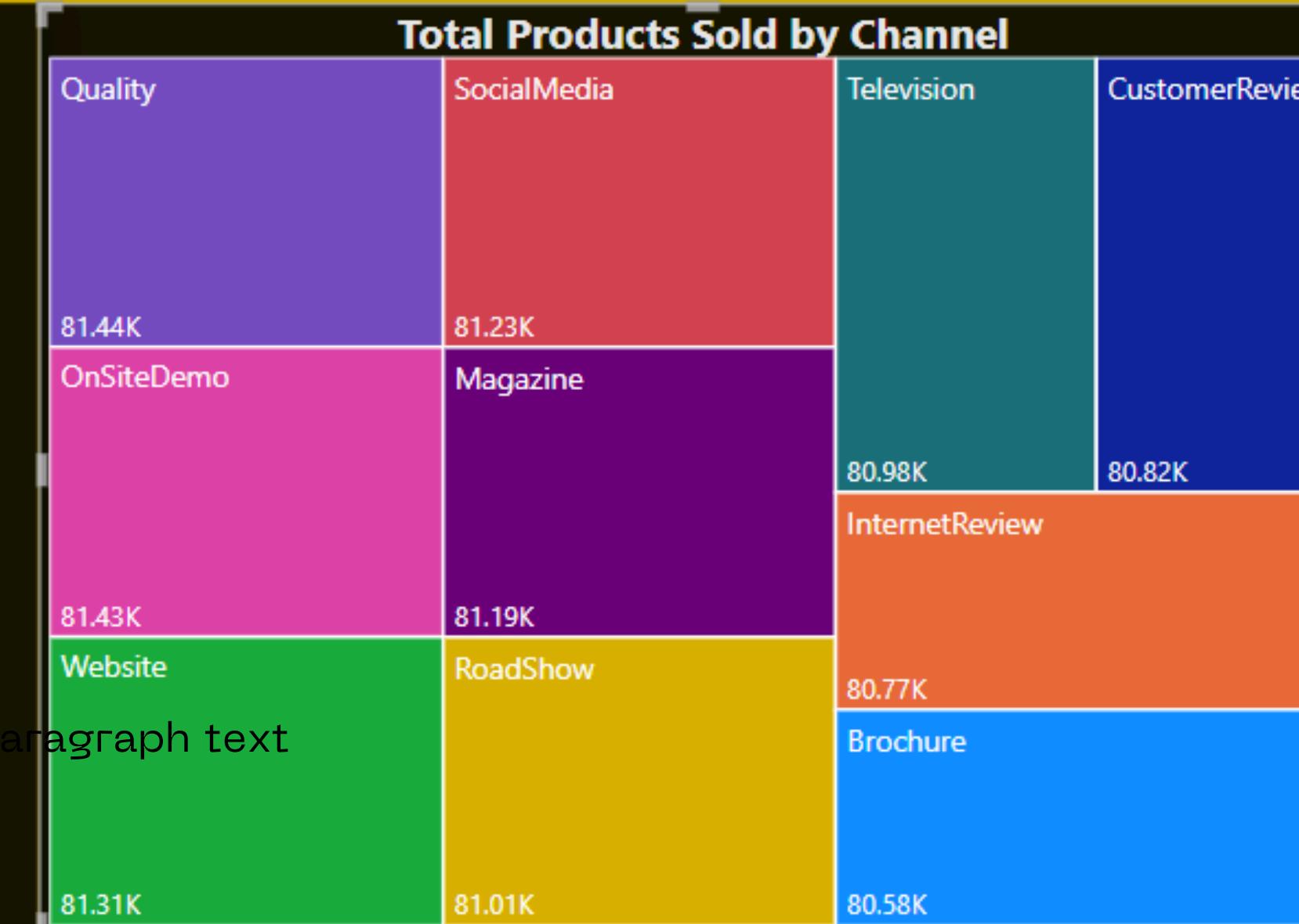
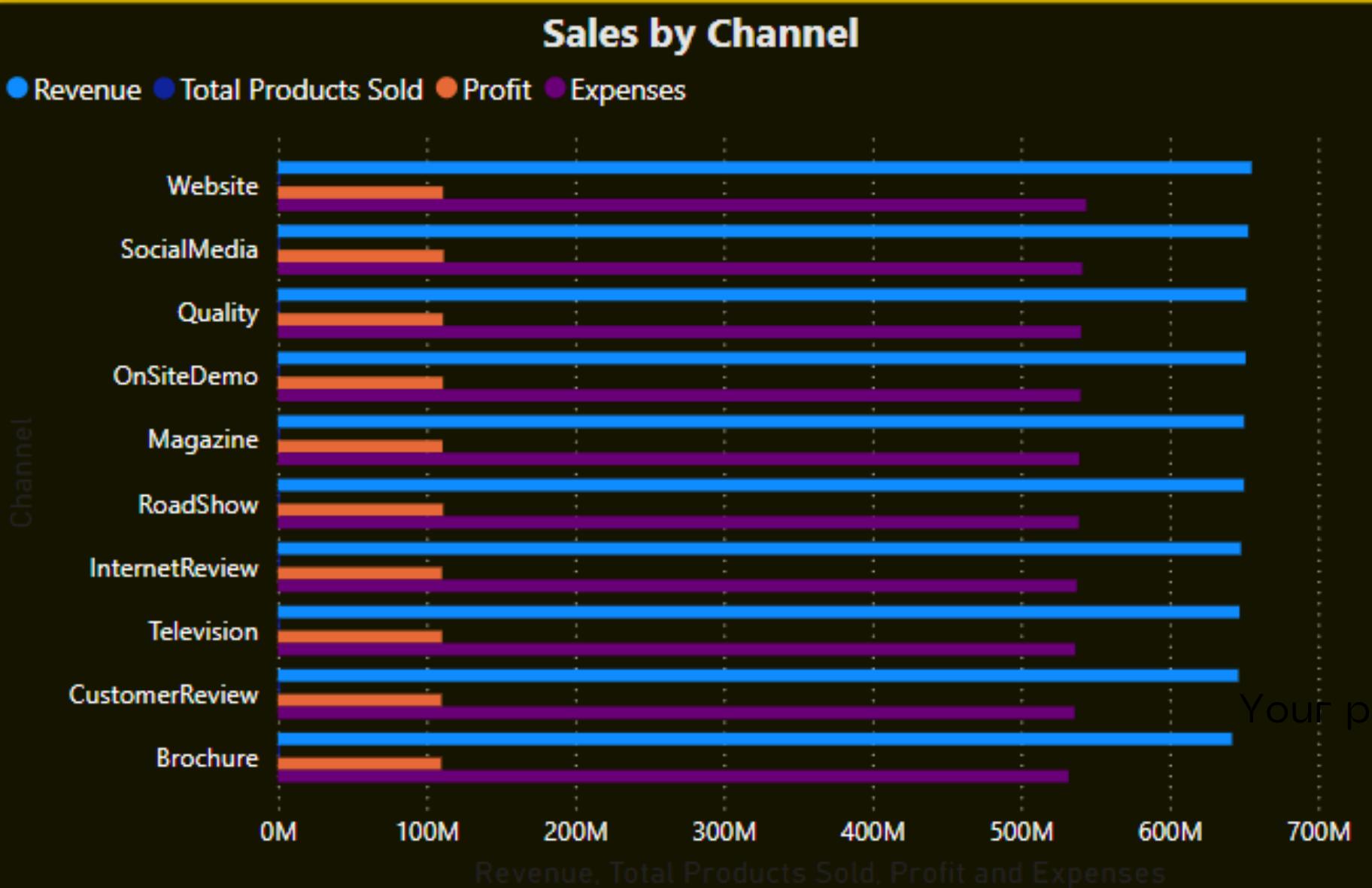
EXPLANATION

1. Pie chart showing customer segmentation by Industry, % of customers in values and Industry in Legends
2. Table showing Top Customer names based on revenue and profit generation
3. Table showing Top Customer names based on revenue generation and payment type
4. Map Visual shows revenue by continent and country, drill down to continent and country and each regions shows a pie chart showing industry wise contribution

BUSINESS IMPACT:

1. Fix delays for high-impact products: Focus on delayed products with high revenue/profit.
2. Optimize inventory: Stock more of top-selling products with short delivery times.
3. Review product mix: Phase out low-margin products with long delivery times.
4. Leverage product descriptions: Replicate successful keywords/features in new products.
5. Delays may increase operational costs (e.g., expedited shipping), reducing margins.

CHANNEL PERFORMANCE



BUSINESS IMPACT

- Identify which channels (social media, email, website, onsite demo) drive the highest revenue and profit.
- Action: Allocate more budget to high-performing channels.
- By Calculating Cost-Per-Acquisition (CPA). Compare to revenue/profit to assess ROI.
- Use product descriptions to analyze which products sell best on which channels

EXPLANATION

Tree Map shows distribution of count of product sold and Channel type. Clusters Bar Chart is used to show relation between Channel type, total Products sold, profit, revenue and expense

STORE PERFORMANCE

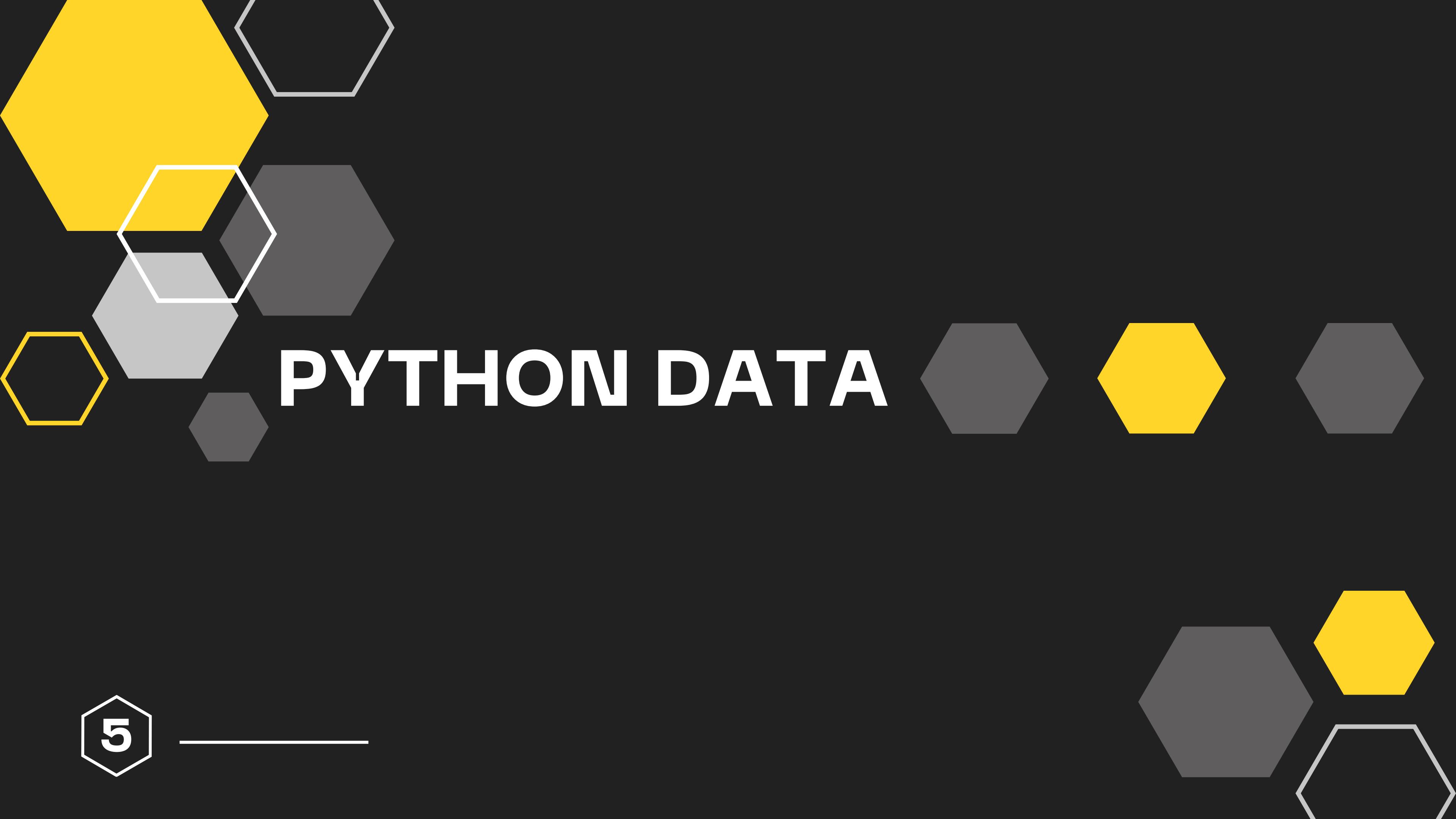
Revenue by Store and Type			
StoreID	CompanyOwned	PrivatelyOwned	Total
NT780		81,230,601.00	81,230,601.00
BW288		81,163,311.00	81,163,311.00
OC714		80,356,796.00	80,356,796.00
IN489	79,959,731.00		79,959,731.00
SW244	79,930,579.00		79,930,579.00
OC689		79,902,256.00	79,902,256.00
NM120	79,658,447.00		79,658,447.00
SW245	79,648,610.00		79,648,610.00
TU989	79,611,055.00		79,611,055.00
VZ342		79,551,625.00	79,551,625.00
RW908	79,533,419.00		79,533,419.00
OC691		79,516,982.00	79,516,982.00
NW182	79,383,531.00		79,383,531.00
CH243		79,184,728.00	79,184,728.00
KO221		79,098,079.00	79,098,079.00
NW181	79,020,577.00		79,020,577.00
SH120		79,018,245.00	79,018,245.00
EH398	79,016,776.00		79,016,776.00
UA662		79,006,592.00	79,006,592.00
US212	78,993,767.00		78,993,767.00
MX811	78,969,540.00		78,969,540.00
RU367	78,915,001.00		78,915,001.00
CA205	78,882,310.00		78,882,310.00
Total	3,597,014,163.00	2,898,283,376.00	6,495,297,539.00

EXPLANATION

- Matrix is created using Store Id as rows and type of company as columns, to understand the distribution of prices, row total and column total also added for analysis.
- Company-Owned stores contribute 55.37% of the total revenue.
- Privately-Owned stores contribute 44.63%.
- Company-Owned stores seem to have a slightly higher share of revenue.
- Store NT780 has the highest revenue (81,230,601.00), followed closely by BW288 (81,163,311.00) and OC714 (80,356,796.00).

BUSINESS IMPACT

- There are no extreme outliers, suggesting a well-distributed performance across stores.
- Understanding the factors driving the top-performing stores (NT780, BW288) could help in replicating their success across other locations.
- A deeper comparison of revenue trends between company and privately-owned stores might reveal operational efficiency differences.
- Privately-Owned stores appear to be competitive with company-owned stores in terms of revenue.



PYTHON DATA

5

EXPLANATION:

This code processes the sales data by converting the 'TransactionDate' column to a datetime format, then extracts the month and year to create new columns. The data is filtered to include sales from 2002 to 2012. A new column 'total_sales' is calculated by multiplying the quantity (Qty) by the price (Price). The dataset is grouped by year and month to sum the total sales for each period. Finally, a new 'YearMonth' column is created by combining the year and month into a YYYY-MM format and converting it into a datetime type for easy plotting or analysis. The result is a DataFrame ready for visualization of monthly sales trends over the given period.

BUSINESS IMPACT:

- Focus on Peak Sales Periods: Identifying months with consistently high sales can inform future marketing and inventory strategies. Increasing promotional efforts or enhancing product availability during these high-sales months could further maximize revenue.
- Address Decline Periods: If certain months show consistent sales downturns, stakeholders should investigate underlying causes—whether it's due to market conditions, competition, or customer behavior. Targeted initiatives, such as introducing new products, offering discounts, or adjusting marketing campaigns during these months, could help mitigate losses.
- Seasonality Planning: The data shows seasonal patterns in sales, and recognizing these trends can lead to better forecasting and resource allocation. Stakeholders could plan for higher demand during peak seasons and reduce operational costs during slower months.
- Data-Driven Decision Making: Continued analysis of monthly sales data will empower stakeholders to make informed, data-driven decisions that align with market dynamics and customer preferences, driving long-term growth and profitability.

```
df['TransactionDate'] = pd.to_datetime(df['TransactionDate'])
df['Month'] = df['TransactionDate'].dt.month
df['Year'] = df['TransactionDate'].dt.year

# Filter sales data from 2002 to 2012
sales_2002_2012 = df[(df['Year'] > 2001) & (df['Year'] < 2013)]

# Calculate total sales
sales_2002_2012['total_sales'] = sales_2002_2012['Qty'] * sales_2002_2012['Price']

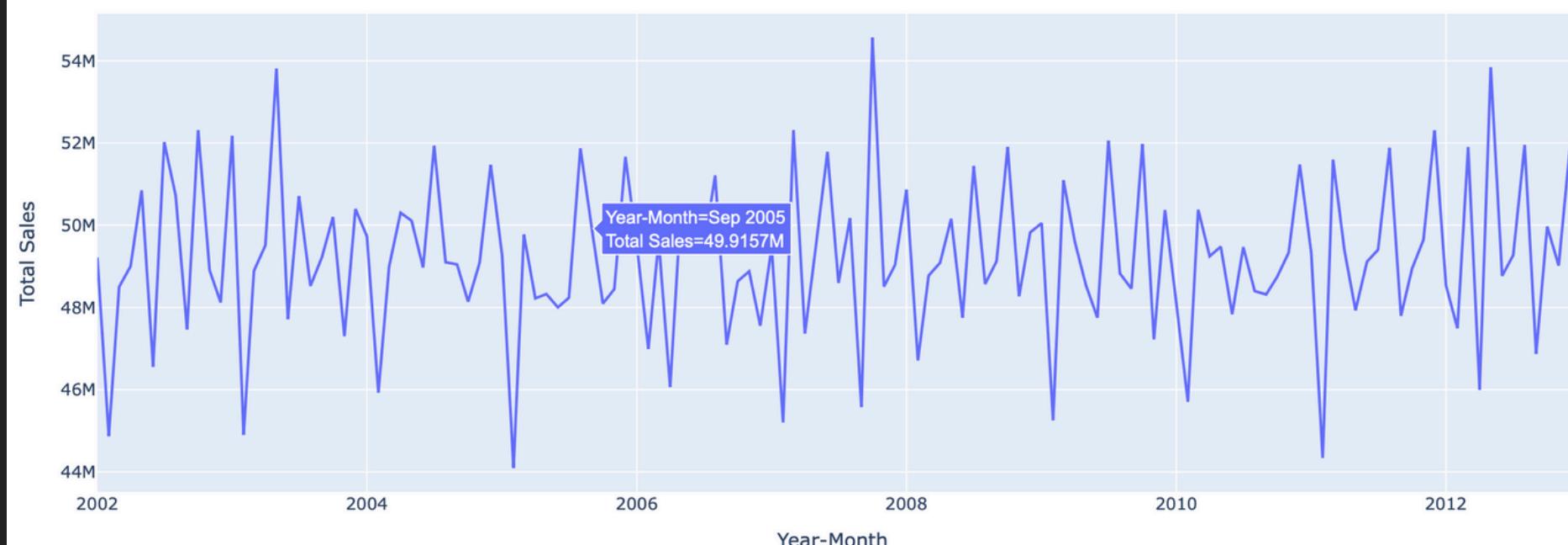
# Group by Year and Month for total sales
monthly_sales_trend = sales_2002_2012.groupby(['Year', 'Month'])['total_sales'].sum().reset_index()

# Create Year-Month column for x-axis
monthly_sales_trend['YearMonth'] = monthly_sales_trend['Year'].astype(str) + '-' + monthly_sales_trend['Month'].astype(str)
monthly_sales_trend['YearMonth'] = pd.to_datetime(monthly_sales_trend['YearMonth'], format='%Y-%m')

# Create the plot using Plotly
fig = px.line(
    monthly_sales_trend,
    x='YearMonth',
    y='total_sales',
    title='Monthly Sales Trend (2002-2012)',
    labels={'YearMonth': 'Year-Month', 'total_sales': 'Total Sales'},
    hover_data={'YearMonth': True, 'total_sales': True} # Customize what you want to show on hover
)

# Show the interactive plot
fig.show()
```

Monthly Sales Trend (2002-2012)



```

df['TransactionDate'] = pd.to_datetime(df['TransactionDate'])
df['Year'] = df['TransactionDate'].dt.year

# Create Region column
df['Region'] = df['CountryName'] + '-' + df['CityName']

# Calculate total sales
df['TotalSales'] = df['Price'] * df['Qty']

# Aggregate sales by Region and Year
RegionSales = df.groupby(['Region', 'Year'])['TotalSales'].sum().reset_index()

# Calculate sales growth percentage
RegionSales['SalesGrowth'] = RegionSales.groupby('Region')['TotalSales'].pct_change() * 100

# Drop NaN values and reset index
RegionSales = RegionSales.dropna().reset_index(drop=True)

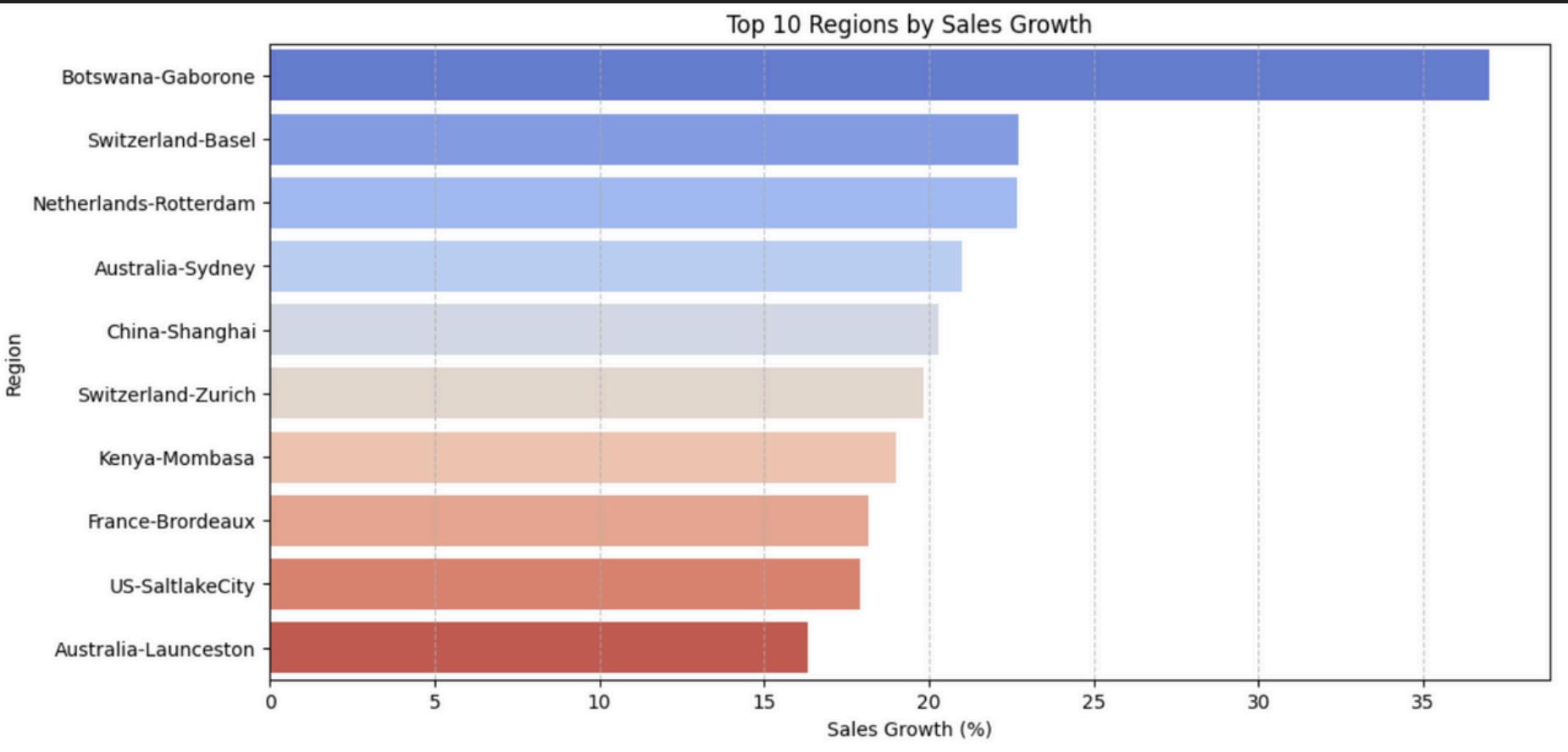
# Get top 10 regions with highest sales growth
TopGrowthRegions = RegionSales.sort_values(by='SalesGrowth', ascending=False).head(10)

# Plot using seaborn for better visuals
plt.figure(figsize=(12, 6))
sns.barplot(x='SalesGrowth', y='Region', data=TopGrowthRegions, palette='coolwarm')

plt.xlabel('Sales Growth (%)')
plt.ylabel('Region')
plt.title('Top 10 Regions by Sales Growth')
plt.grid(axis='x', linestyle='--', alpha=0.7)

plt.show()

```



EXPLANATION:

This code processes the sales data by first creating a 'Region' column that combines 'CountryName' and 'CityName' to represent each region. It then calculates total sales for each transaction by multiplying 'Qty' by 'Price'. The data is aggregated by region and year, summing the total sales for each group. Sales growth percentage is calculated by measuring the year-over-year change in total sales for each region. Any missing values are dropped, and the index is reset. The top 10 regions with the highest sales growth are selected and visualized using a bar plot. The plot displays sales growth on the x-axis and regions on the y-axis, with a color palette and grid lines for enhanced readability, providing clear insights into the regions with the highest growth.

BUSINESS IMPACT:

- Leverage High-Growth Regions:** Stakeholders should consider increasing investments, marketing efforts, and resource allocation in these high-growth regions to further capitalize on their success. This could involve enhancing product availability, expanding customer engagement, or boosting local promotions.
- Identify Success Factors:** Understanding what factors have contributed to the success of these regions—whether it's customer demand, local market conditions, or effective sales strategies—can help replicate these factors in other regions. Conducting further analysis on these top regions could uncover patterns that can be applied across the business.
- Monitor and Optimize Low-Performing Regions:** While the focus is on growth, it's also crucial to evaluate regions that are not performing as well. Identifying why some regions are lagging can help shape strategies for recovery, such as revisiting pricing, marketing, or product offerings.
- Data-Driven Strategic Planning:** This data-driven approach empowers stakeholders to make informed decisions based on clear, measurable sales trends. By continuously monitoring sales growth at the regional level, the business can adjust its strategies in real time to maximize overall performance and profitability.

EXPLANATION:

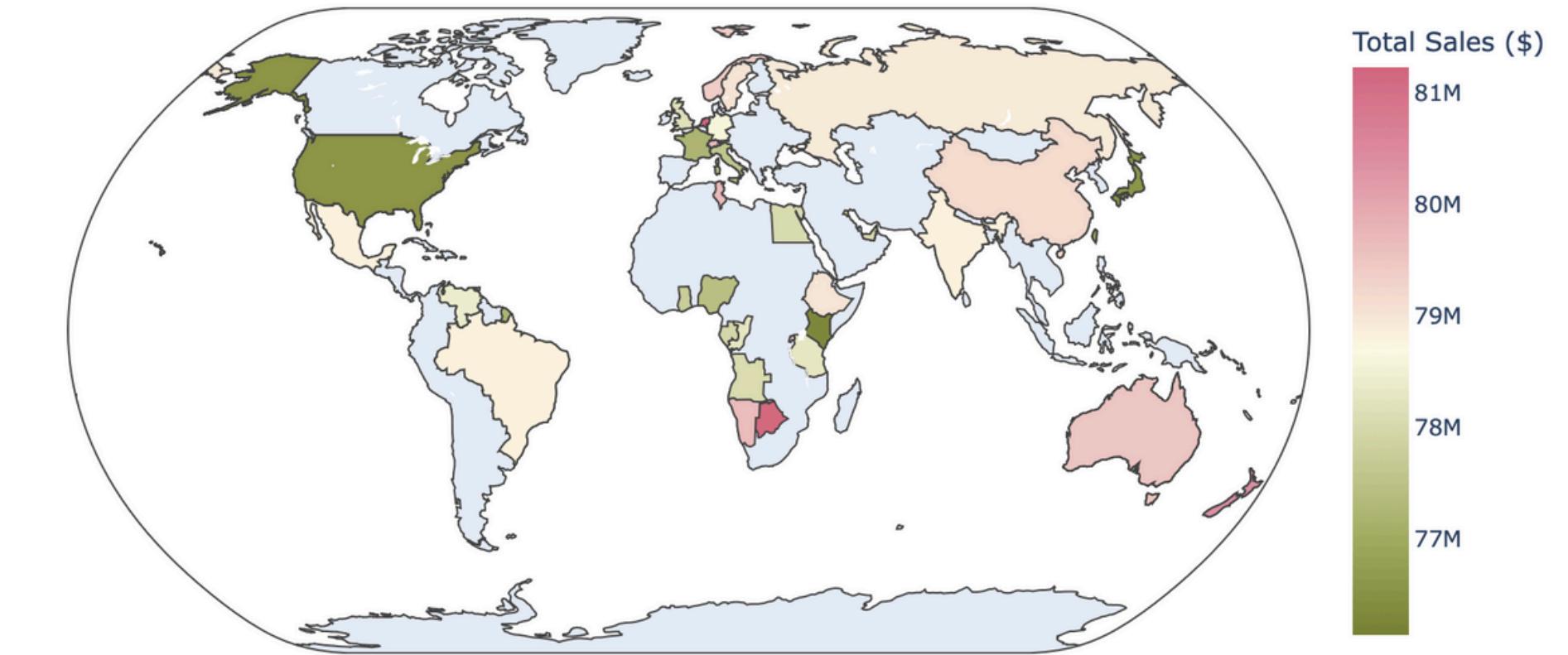
This code aggregates sales data by 'CountryName' and 'CityName' to calculate the total sales for each country and city. It then uses Plotly's choropleth function to create an interactive geographic heatmap. The map visualizes total sales by coloring countries based on the sales figures, with darker colors indicating higher sales. The hover_name and hover_dataoptions display additional details, such as the total sales amount and city names, when hovering over a region. The map uses the 'Armyrose' color scale and a 'natural earth' projection for a visually appealing representation. Finally, the heatmap is displayed, providing a clear, geographic view of sales performance across different regions.

BUSINESS IMPACT:

- Focus on High-Performing Regions: Stakeholders should consider reinforcing efforts in countries and cities with high total sales. Expanding marketing initiatives, increasing inventory, or introducing new products in these regions could further boost sales and brand presence.
- Investigate Low-Performing Regions: Regions with lower sales should be carefully analyzed to understand the underlying causes, whether it's due to competitive pressures, market demand, or ineffective sales strategies. Tailored actions such as localized promotions, strategic partnerships, or customer engagement campaigns could help increase sales in these areas.
- Optimize Resource Allocation: The map's insights can help stakeholders allocate resources more effectively. For instance, more resources (e.g., sales staff, budget, or logistics) can be directed to high-performing regions, while underperforming regions may benefit from targeted interventions or adjustments.
- Global Strategy Planning: With a comprehensive view of sales across various countries and cities, stakeholders can make data-driven decisions when considering international expansions, adjusting product offerings based on regional preferences, or planning region-specific campaigns to improve market penetration.

```
total_sales = data.groupby(["CountryName", "CityName"])['TotalSales'].sum().reset_index()
fig2 = px.choropleth(
    total_sales,
    locations="CountryName",
    locationmode="country names",
    color="TotalSales",
    hover_name="CountryName",
    title="Geographic Sales Heatmap",
    color_continuous_scale='Armyrose',
    hover_data={"TotalSales": True, "CityName": True},
    labels={"TotalSales": "Total Sales ($)"},
    projection="natural earth"
)
fig2.show()
```

Geographic Sales Heatmap



```

df['TotalSales'] = df['Qty'] * df['Price']
ValuableCust = df.groupby('CustomerKey')['TotalSales'].sum().reset_index()
ValuableCust = ValuableCust.sort_values(by='TotalSales', ascending=False)
ValuableCust = ValuableCust.head(20)
print(ValuableCust)

fig = px.bar(
    ValuableCust,
    x='TotalSales',
    y='CustomerKey',
    orientation='h', # Horizontal bar chart
    title='Top 20 Most Valuable Customers',
    labels={'CustomerKey': 'Customer ID', 'TotalSales': 'Total Sales'},
    color='TotalSales', # Color bars by total sales
    color_continuous_scale='Viridis', # Use a gradient color scale
    hover_data={'CustomerKey': True, 'TotalSales': True} # Show detailed info on hover
)

# Show the interactive plot
fig.show()

```



BUSINESS IMPACT:

- Strengthen Relationships with Top Customers: Stakeholders should prioritize nurturing relationships with these high-value customers by offering personalized experiences, loyalty programs, and tailored promotions. This could lead to increased customer retention and repeat business.
- Targeted Marketing and Sales Strategies: Given that these customers contribute significantly to total sales, stakeholders should direct marketing resources towards these individuals, leveraging data to craft campaigns that resonate with their preferences and purchasing habits.
- Incentivize Referrals and Advocacy: High-value customers can be powerful brand advocates. Stakeholders might consider incentivizing these customers to refer new clients through referral programs, helping to expand the customer base while capitalizing on existing customer loyalty.
- Analyze Customer Behavior for Broader Insights: Understanding the purchasing behavior of these top customers can provide valuable insights into broader market trends. Stakeholders can apply these insights to optimize product offerings, pricing strategies, and inventory management.

EXPLANATION:

This code calculates the total sales for each customer and visualizes the top 20 most valuable customers using a horizontal bar chart. First, the 'TotalSales' column is created by multiplying the 'Qty' (quantity) by 'Price'. Then, the data is grouped by 'CustomerKey' and aggregated to calculate the total sales for each customer. The results are sorted in descending order, and the top 20 customers are selected. The bar chart is created using Plotly, where the x-axis represents total sales, the y-axis represents customer IDs, and the bars are colored according to the total sales using the 'Viridis' color scale. Interactive hover data shows customer IDs and sales amounts. The chart provides a clear, visual representation of the most valuable customers.

EXPLANATION:

This code calculates key loyalty metrics for customers and visualizes the top 20 customers based on purchase frequency in a bar chart with unique colors for each customer. First, the 'TotalSpent' is calculated by multiplying 'Cost' by 'Qty'. Then, the data is grouped by 'CustomerKey' to calculate three metrics: purchase frequency (total number of products purchased), total spent, and average spend per purchase. The top 20 customers with the highest purchase frequency are selected using `nlargest()`. The chart is created with Plotly, where the x-axis represents customer IDs, the y-axis shows purchase frequency, and the bars are colored uniquely for each customer. Average spend per purchase is displayed as text labels on the bars. The layout is customized for better visualization, and the chart is displayed interactively, enabling users to explore customer loyalty data.

```
df['TotalSpent'] = df['Cost'] * df['Qty']
loyalty_opportunities = df.groupby('CustomerKey').agg(
    purchase_frequency=('ProductKey', 'count'),
    total_spent=('TotalSpent', 'sum'),
    avg_spent_per_purchase=('TotalSpent', 'mean')
).reset_index()

# Sort and select top 20
loyalty_opportunities = loyalty_opportunities.nlargest(20, 'purchase_frequency')

# Plot with unique colors for each customer
fig = px.bar(
    loyalty_opportunities,
    x='CustomerKey',
    y='purchase_frequency',
    text='avg_spent_per_purchase',
    color='CustomerKey', # Assign a unique color for each customer
    title='LOYALTY PROGRAM CUSTOMERS'
)

# Update layout for better visualization
fig.update_layout(width=800, height=400)
fig.show()
```

BUSINESS IMPACT:

- Enhance Customer Retention Programs: Stakeholders should focus on creating and enhancing loyalty programs for these high-frequency customers. Exclusive offers, rewards, or personalized services can help maintain strong relationships and prevent churn.
- Targeted Marketing and Engagement: Marketing campaigns should be tailored to these loyal customers. Since they frequently engage with the brand, personalized communication and promotions could drive even higher spending and encourage them to refer new customers.
- Referral Incentives: Leveraging the trust and loyalty of these customers can be beneficial. Introducing a referral program that rewards them for bringing in new customers can help expand the customer base while capitalizing on the influence of existing high-value clients.
- Identify Potential Upsell Opportunities: With a deeper understanding of the average spend per purchase, stakeholders can look for upselling or cross-selling opportunities. Offering complementary products or services tailored to these loyal customers' preferences can drive additional revenue.
- Customer Segmentation for Strategic Expansion: By analyzing the purchasing behavior of these top customers, stakeholders can gain insights into broader trends. These insights can inform future product offerings, pricing strategies, or market expansion, especially in areas where loyal customer bases are particularly strong.

LOYALTY PROGRAM CUSTOMERS





STRATEGIC RECOMMENDATIONS



SALES OPTIMIZATION

• Seasonal Promotions for High-Demand Products:

- Action: Based on the sales trends and top-performing products, Caterpillar Inc. should introduce targeted seasonal promotions for high-demand machinery, parts, and equipment. For example, based on the sales trends from the monthly sales analysis, products that show consistent high sales during certain months (such as construction equipment during construction season) can be featured in seasonal promotions with special discounts, bundled offers, or financing deals.
- Impact: This will help boost sales during peak seasons, attract a larger customer base, and capitalize on the seasonal demand.

• Targeted Expansion in High-Performing Regions:

- Action: As identified from the regional sales growth data, Caterpillar should focus on expanding its stores, dealerships, or authorized service centers in regions with the highest sales growth. These could be countries or cities where there has been significant demand for products and services. For instance, if regions like 'California' or 'Texas' show strong sales performance in specific product categories (e.g., heavy machinery for mining or construction), it would make sense to expand operations in those areas.
- Impact: This regional expansion will allow the company to capitalize on strong markets, enhance brand visibility, and provide better support to loyal customers, driving increased sales and market penetration.

• Improved Supply Chain Management for Frequently Delayed Products:

- Action: Based on customer feedback and product sales, Caterpillar should streamline the supply chain process for products that are frequently delayed. By working with suppliers to ensure timely deliveries and predicting demand based on historical sales data, the company can optimize inventory management and ensure that customers can access products without delay. This also involves improving the warehousing systems and logistics capabilities to speed up delivery times.
- Impact: Reducing delays will enhance customer satisfaction, improve service levels, and maintain customer loyalty, leading to increased sales and reduced risk of losing customers to competitors.

MARKETING IMPROVEMENTS

• Personalized Marketing Campaigns for High-Value Customers:

- Action: Using customer segmentation insights, Caterpillar should create personalized marketing campaigns targeting their top 20 most valuable customers. These customers, who contribute significantly to the overall sales, should receive tailored promotions, loyalty incentives, and exclusive offers. For instance, targeted email campaigns could showcase new products, upgrades, or special financing options for these high-value clients.
- Impact: Personalized offers will increase customer loyalty, drive repeat purchases, and enhance the lifetime value of high-revenue customers.

• Referral Programs for Loyal Customers:

- Action: Given that the top customers are more likely to refer new clients, Caterpillar could implement a referral program that rewards loyal customers for bringing in new business. This could include offering discounts, special services, or even exclusive product access for every new referral.
- Impact: This strategy leverages the power of word-of-mouth marketing, helps expand the customer base, and rewards existing customers for their loyalty, resulting in increased sales.

• Cross-Promote Products in Complementary Categories:

- Action: Based on customer purchase behavior and product performance analysis, Caterpillar should develop cross-selling and upselling strategies by bundling complementary products together. For example, if a customer purchases a bulldozer, they can be offered a discounted rate on related accessories such as maintenance parts or equipment for specific applications.
- Impact: Cross-selling will increase the average order value and drive incremental sales while enhancing the customer experience by providing all necessary solutions in one go.

CUSTOMER SEGMENTATION

• Segment Customers Based on Purchase Frequency and Spending Patterns:

- Action: Caterpillar can further segment its customer base into categories such as 'high-frequency, high-spend,' 'low-frequency, high-spend,' and 'low-frequency, low-spend' customers. Tailored loyalty programs can be developed for each segment. For instance, customers who purchase frequently but at lower values could be offered rewards for upgrading to premium products, while high-spend customers can receive exclusive services.
- Impact: This segmentation allows Caterpillar to implement targeted marketing, tailored offers, and loyalty programs to cater to the specific needs of each customer group, ensuring higher satisfaction and retention.

• Target Geographic Expansion Using Regional Sales Data:

- Action: Use sales data to identify regions with the highest potential for growth. Areas showing consistent sales growth can be targeted for further investment in marketing, service centers, and product offerings. For example, regions with strong demand for construction equipment could be prioritized for marketing campaigns.
- Impact: This targeted expansion will ensure the efficient allocation of resources, focusing efforts on regions where there is the most growth potential and demand.

• Implement a Customer Feedback Loop:

- Action: To understand customer needs and preferences better, Caterpillar should implement an ongoing customer feedback loop, particularly from their most valuable and frequent customers. This feedback can be collected through surveys, interviews, or post-purchase satisfaction forms. Insights from this data should be used to refine product offerings and customer service strategies.
- Impact: A feedback loop helps maintain a customer-centric approach and ensures that product development and service offerings align with market demands, leading to higher customer satisfaction and improved brand loyalty.

CONCLUSION

SQL KEY TAKEAWAYS

Revenue Milestones

Revenue analysis shows that May, August, and December consistently generate the highest sales, with customers placing the most orders during these months.

These SQL insights provide a comprehensive understanding of sales dynamics, helping to optimize marketing resources, target the most profitable customers, and improve customer retention strategies.

Top Stores & Customers

The top 5 revenue-generating customers—Sheridan Swindin, Zemfira M Tretyakova, Myrtle Anespie, Kell N Grendle, and Taija E Järvinen—account for the highest sales.

These customers represent key opportunities for personalized marketing efforts, ensuring they remain loyal and continue driving significant revenue.

Focusing on these top customers can help optimize sales and strengthen long-term relationships.

Sales Channel Performance

The Sales Channel Performance analysis highlights that InternetReview, SocialMedia, and Website are the top-performing sales channels, with InternetReview leading in both 2009 and 2013. The Magazine channel also performed strongly in 2012. By focusing on these high-performing platforms, the business can allocate resources effectively, refine marketing strategies, and maximize sales potential. It is essential to ensure that these channels continue to be optimized to sustain and grow their success.

POWER BI KEY TAKEAWAYS

Sales Trend

- Sales fluctuate throughout the year.
- A noticeable drop in February but recoveries in March, May, July, and October.
- The highest sales seem to be around July and October, while February is the weakest month.
- The trend suggests possible seasonal effects or external factors impacting sales.

Revenue Distribution

- The Power BI dashboard provides a real-time view of revenue generation, enabling better decision-making. Company-Owned stores contribute 55.37% of the total revenue.
- Privately-Owned stores contribute 44.63%.
- Company-Owned stores seem to have a slightly higher share of revenue.

What's Required?

- Seasonal Sales Trends: Investigate the cause of sales dips in February and September and optimize sales strategies accordingly.
- Category Performance: Construction Machinery and Industrial products are strong performers; Renewable Energy has potential for growth.
- Product Demand: Wind turbines are the top-selling products; investing in their sales and marketing could enhance revenue.

PYTHON KEY TAKEAWAYS

Capitalize on Seasonal and Regional Sales Trends

By identifying peak sales periods and high-growth regions, Caterpillar INC can optimize inventory, launch targeted promotions, and enhance supply chain efficiency to meet seasonal demand. Simultaneously, underperforming regions should be analyzed for potential market adjustments, such as localized marketing, pricing strategies, or product diversification to improve sales.

Leverage High-Value Customers for Growth

Strengthening relationships with top customers through loyalty programs, personalized offers, and referral incentives will boost retention and drive repeat business. These customers can also provide valuable insights into purchasing trends, enabling Caterpillar to refine its sales strategies and expand into new, high-potential markets.

Data-Driven Strategic Decision-Making

Continuous analysis of sales performance, customer behavior, and regional demand will empower Caterpillar to make informed, real-time decisions. By aligning marketing efforts, product offerings, and resource allocation with data insights, the company can enhance profitability, streamline operations, and maintain a competitive edge in the global market.

THANK YOU

QAT

