Regional Sales Analysis

Contents

Problem Statement

Sales teams often lack a clear, data-driven understanding of regional performance, making it difficult to identify opportunities and optimize resources. This project aims to analyze and visualize regional sales data to uncover trends, evaluate profitability, and support strategic decision-making.

What's the Business Question?

➢ Inconsistent revenue and profit performance across U.S. regions

➢ Lack of visibility into seasonal swings, top SKUs, and channel profitability

➢ Goal: Leverage 5 years of historical data to pinpoint growth levers and optimize strategy

Setup, Data Ingestion and Data Profiling

```
#importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
sheets=pd.read_excel('/content/Regional Sales Dataset.xlsx',sheet_name=None)
```

```
# Assign to named DataFrames
df_sales       = sheets['Sales Orders']
df_customers   = sheets['Customers']
df_products    = sheets['Products']
df_regions     = sheets['Regions']
df_state_reg   = sheets['State Regions']
df_budgets     = sheets['2017 Budgets']
```

Pre-processing and Feature Engineering

Set header row for state-region table

```
# The first row of df_state_reg contains the real column names.
new_header        = df_state_reg.iloc[0]      # grab the first row
df_state_reg.columns = new_header             # set columns
df_state_reg       = df_state_reg[1:].reset_index(drop=True)  # drop header row and reset
```

Merge Sales, Customer , Products, Region, State_Region & Budgets tables

```
# 4.1 Merge with Customers
df = df_sales.merge(
    df_customers,
    how='left',
    left_on='Customer Name Index',
    right_on='Customer Index'
)

# 4.2 Merge with Products
df = df.merge(
    df_products,
    how='left',
    left_on='Product Description Index',
    right_on='Index'
)

# 4.3 Merge with Regions
df = df.merge(
    df_regions,
    how='left',
    left_on='Delivery Region Index',
    right_on='id'
)

# 4.4 Merge with State Regions
df = df.merge(
    df_state_reg[["State Code","Region"]],
    how='left',
    left_on='state_code',
    right_on='State Code'
)
```

Drop redundant columns

Standardize column names to lowercase

```
# Convert all column names to lowercase for consistency and easier access
df.columns = df.columns.str.lower()
```

Rename columns to more sensible names

```
df = df.rename(columns={
    'ordernumber'       : 'order_number',    # snake_case for consistency
    'orderdate'         : 'order_date',      # date of the order
    'customer names'    : 'customer_name',   # customer who placed it
    'product name'      : 'product_name',    # product sold
```

Note: No missing values or duplicate rows were found in the dataset

Create profit and profit_margin_pct columns

```
# 1. Calculate total cost for each line item
df['total_cost'] = df['quantity'] * df['cost']

# 2. Calculate profit as revenue minus total_cost
df['profit'] = df['revenue'] - df['total_cost']

# 3. Calculate profit margin as a percentage
df['profit_margin_pct'] = (df['profit'] / df['revenue']) * 100

# 4. Extract full month name from order_date for labeling (e.g., 'January', 'February')
df['order_month_name'] = df['order_date'].dt.month_name()

# 5. Extract month number from order_date for correct sorting (1-12)
df['order_month_num'] = df['order_date'].dt.month
```

Final Dataset Structure – Ready for Analysis:

	order_number	order_date	customer_name	channel	product_name	quantity	unit_price	revenue	cost	state	state_name	us_region	lat	lon	budget	total_cost	p
0	SO - 000225	2014-01-01	Rhynoodle Ltd	Wholesale	Product 27	6	2499.1	14994.6	1824.343	GA	Georgia	South	32.08354	-81.09983	NaN	10946.058	4048
1	SO - 0003378	2014-01-01	Thoughtmix Ltd	Distributor	Product 20	11	2351.7	25868.7	1269.918	IN	Indiana	Midwest	39.61366	-86.10665	NaN	13969.098	1189
2	SO - 0005126	2014-01-01	Amerisourc Corp	Wholesale	Product 26	6	978.2	5869.2	684.740	CA	California	West	37.66243	-121.87468	NaN	4108.440	176
3	SO - 0005614	2014-01-01	Colgate-Pa Group	Export	Product 7	7	2338.3	16368.1	1028.852	IN	Indiana	Midwest	39.16533	-86.52639	NaN	7201.964	916
4	SO - 0005781	2014-01-01	Deseret Group	Wholesale	Product 8	8	2291.4	18331.2	1260.270	CT	Connecticut	Northeast	41.77524	-72.52443	NaN	10082.160	824
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
64099	SO - 0007573	2018-02-28	Dazzlesphe Corp	Wholesale	Product 26	12	1815.7	21788.4	980.478	PA	Pennsylvania	Northeast	40.15511	-74.82877	NaN	11765.736	1002
64100	SO - 0007706	2018-02-28	Yombu Corp	Export	Product 21	6	864.3	5185.8	579.081	IL	Illinois	Midwest	42.11030	-88.03424	NaN	3474.486	171
64101	SO - 0007718	2018-02-28	Bath Group	Distributor	Product 13	11	3953.0	43483.0	2648.510	FL	Florida	South	26.24453	-80.20644	NaN	29133.610	1434
64102	SO - 0008084	2018-02-28	Linklinks Ltd	Distributor	Product 20	7	3959.7	27717.9	2930.178	NY	New York	Northeast	42.91002	-78.74182	NaN	20511.246	720
64103	SO - 0008654	2018-02-28	SAFEWAY Ltd	Distributor	Product 15	8	998.3	7986.4	848.555	OR	Oregon	West	45.48706	-122.80371	NaN	6788.440	119

64104 rows × 20 columns

Identifiers: order_number, order_date, customer_name, channel, product_name
Financials: quantity, unit_price, revenue, cost, profit, profit_margin_pct

Calendar: order_month_name, order_month_num, order_month
Geography: state (code), state_name, us_region, lat, lon
Planning: budget (2017)

Exploratory Analysis:

Monthly Sales Trend Over Time

Insights:

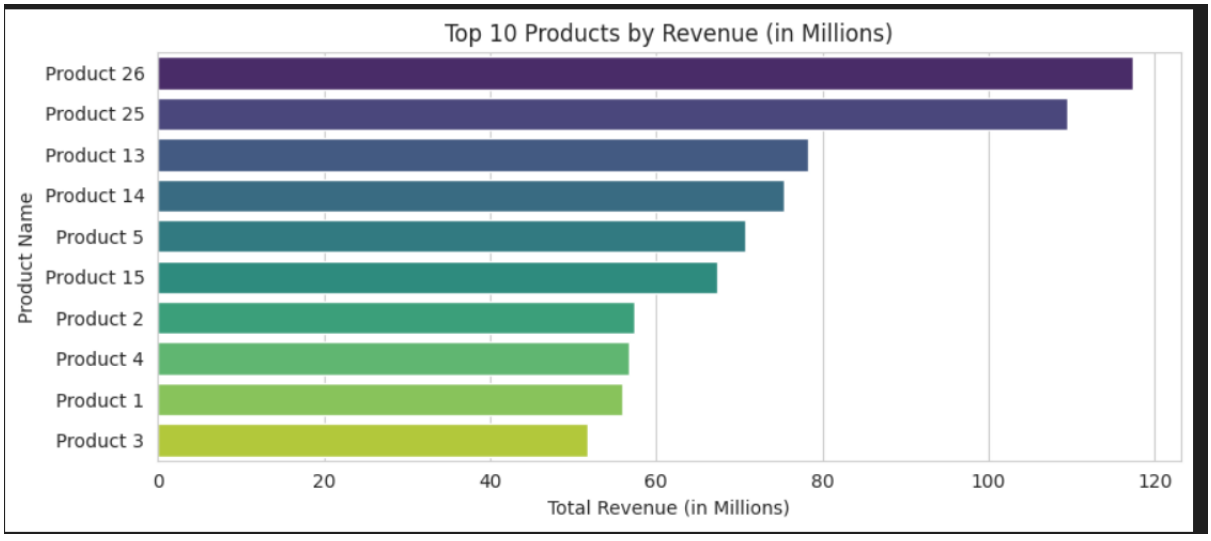
Consistent sales cycle: $24M to $26M.

Seasonal peaks: Late spring/early summer (May-June).

Annual low: January.

Notable outlier: Sharp revenue drop in early 2017.

Top 10 Products by Revenue:



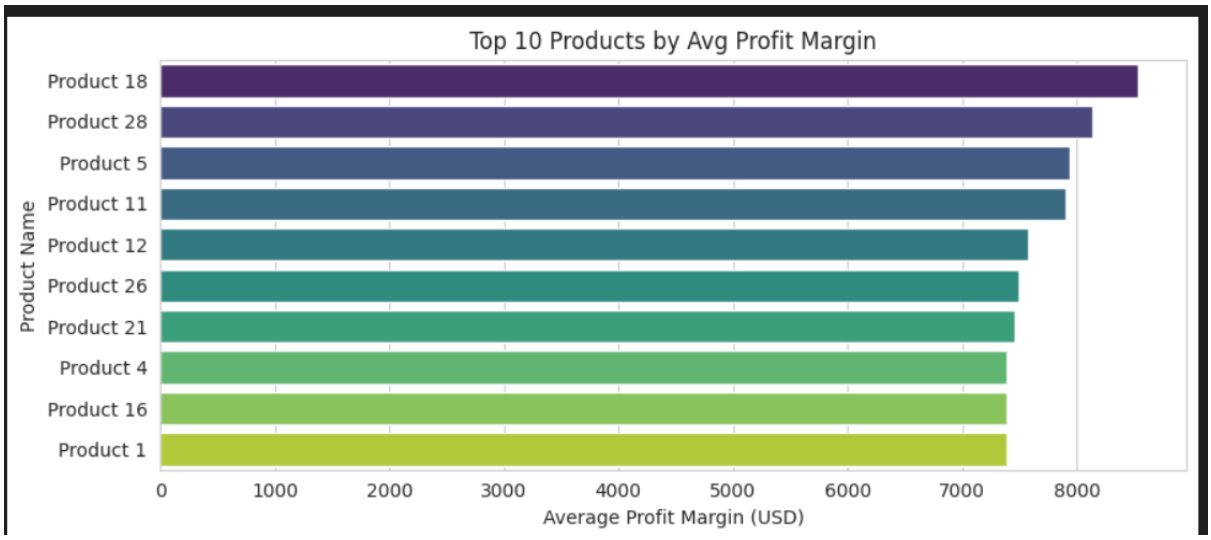
Insights:

Revenue leaders: Products 26 & 25 dominate.

Mid-range: Products 5, 13, 14, 15 show similar revenue.

Strategy: Grow mid-tier, improve lower performers.

Top 10 Products by Avg Profit Margin



Insights:

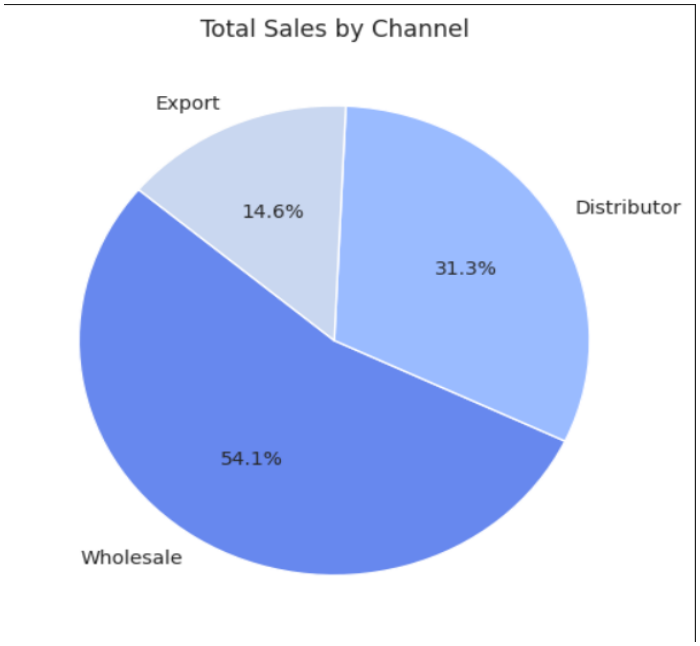
Top: Products 18 & 28 lead at ~$8.1–8.4K

Next tier: Products 5, 11, 12 & 26 at ~$7.5–7.8K.

Entry-level: Products 1, 4, 16 & 21 around $7.3K.

Takeaway: Top 10 all exceed $7.3K—consistent high margins

Sales by Channel:



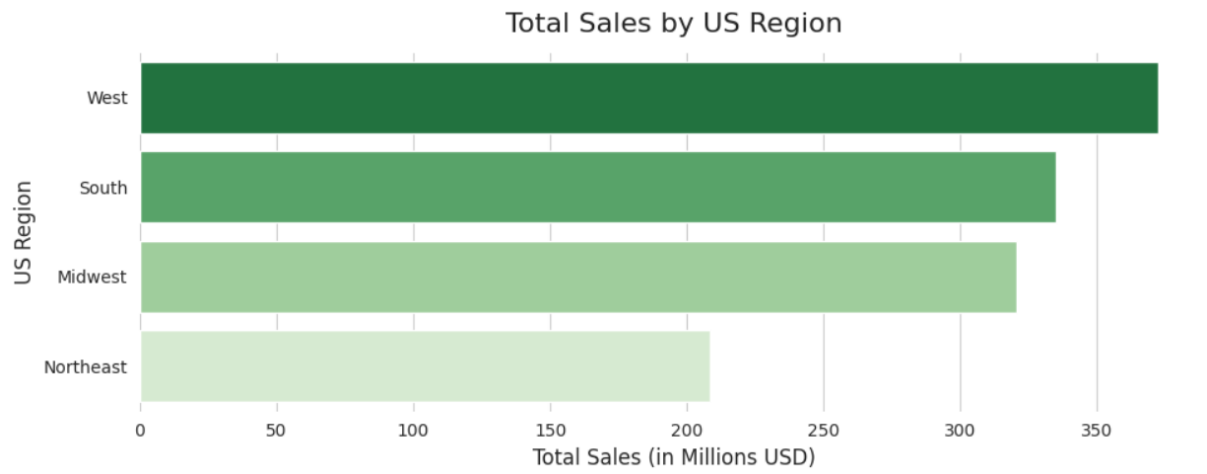
Insights:

Wholesale dominates: Generates the majority of total sales at 54.1%.

Distributor is significant: Contributes a substantial 31.3% to total sales.

Export is a smaller portion: Accounts for 14.6% of the total sales.

Total Sales by US Region:



Insights:

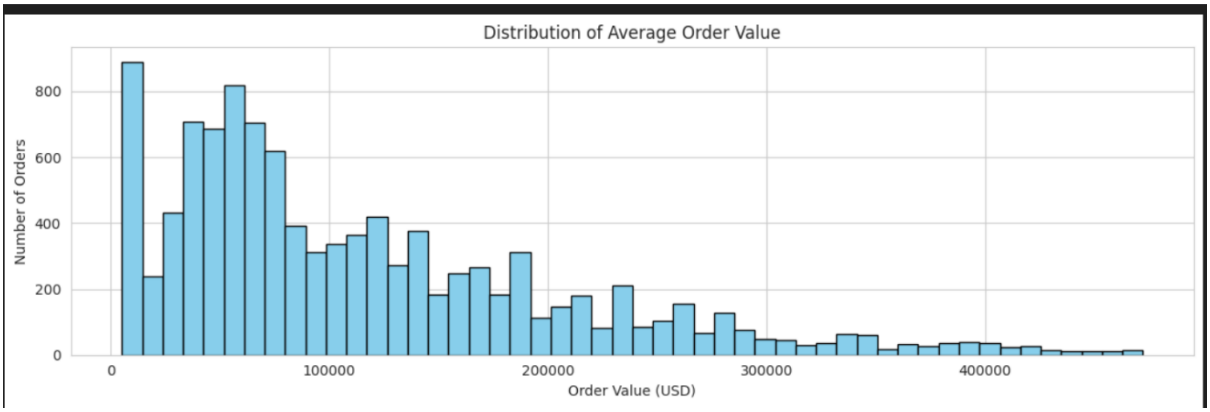
West: Highest sales, strong market influence.

South: Major sales contributor, key market area.

Midwest: Steady sales performance, moderate market size.

Northeast: Lowest sales, suggests need for deeper market understanding

Average Order Value (AOV) Distribution:



Insights:

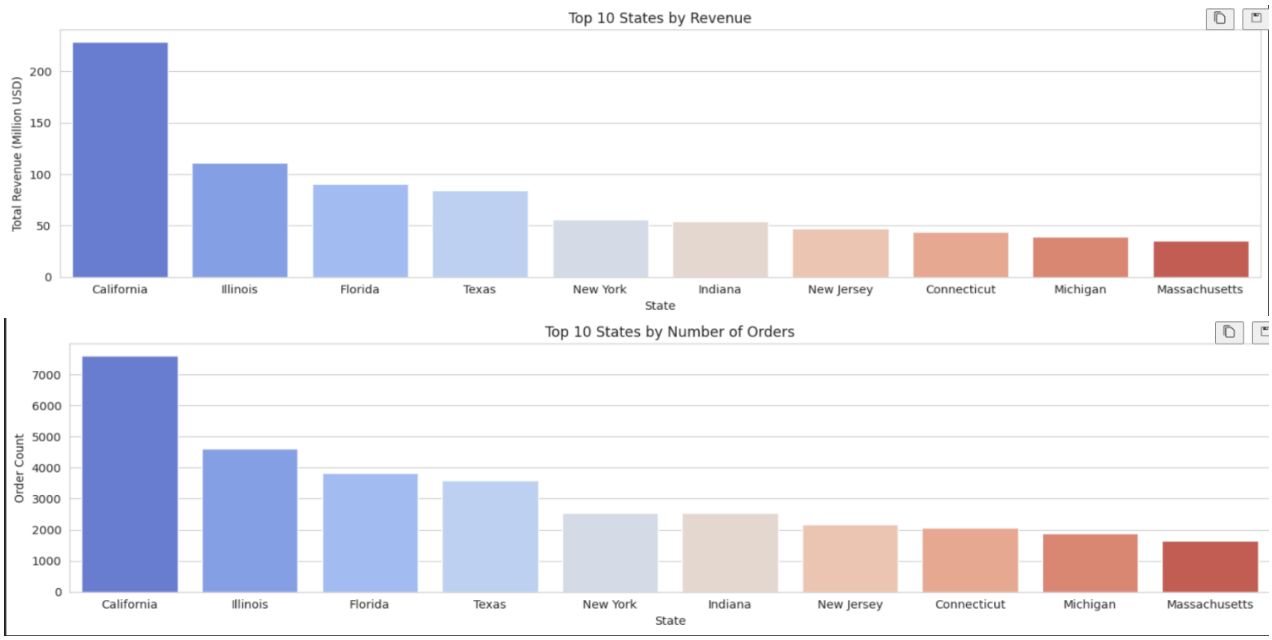
Low average order values are frequent.

Distribution is right-skewed (long tail of high-value orders).

Multiple order value clusters exist.

Higher order values are less common.

Top State Performance: Revenue vs Orders:



Insights:

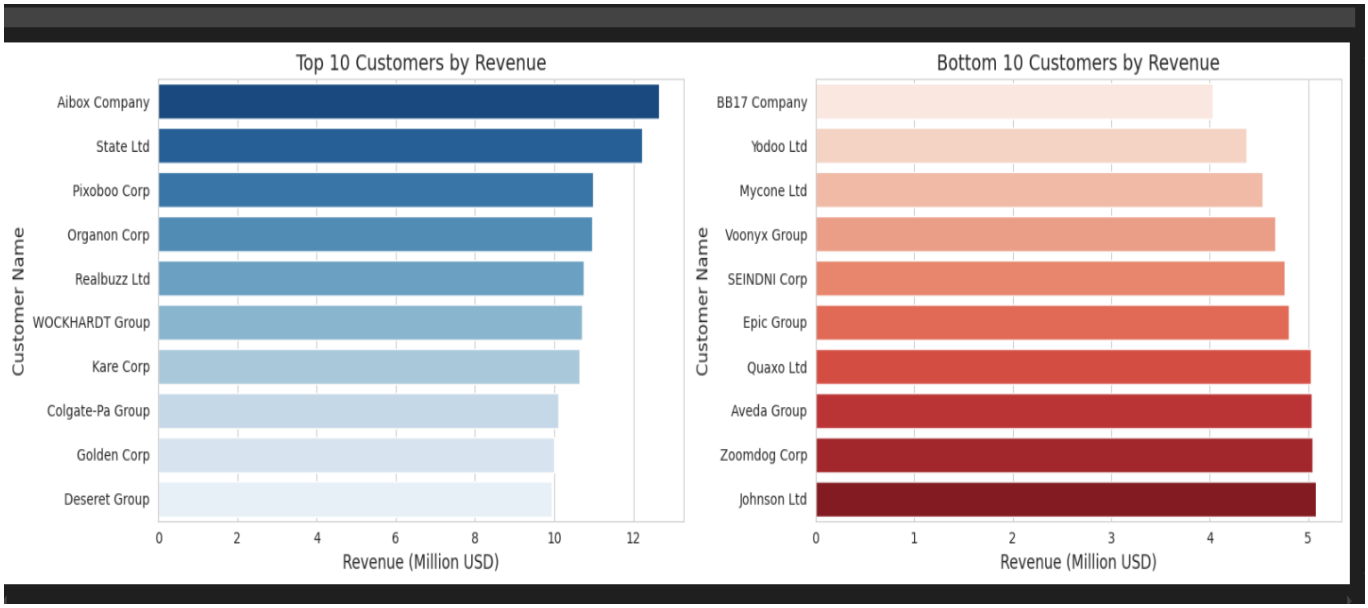
California tops revenue & orders.

IL, FL, TX: High in both.

Revenue & orders linked.

Other top states: Lower contribution

Top and Bottom 10 Customer by Revenue:



Insights:

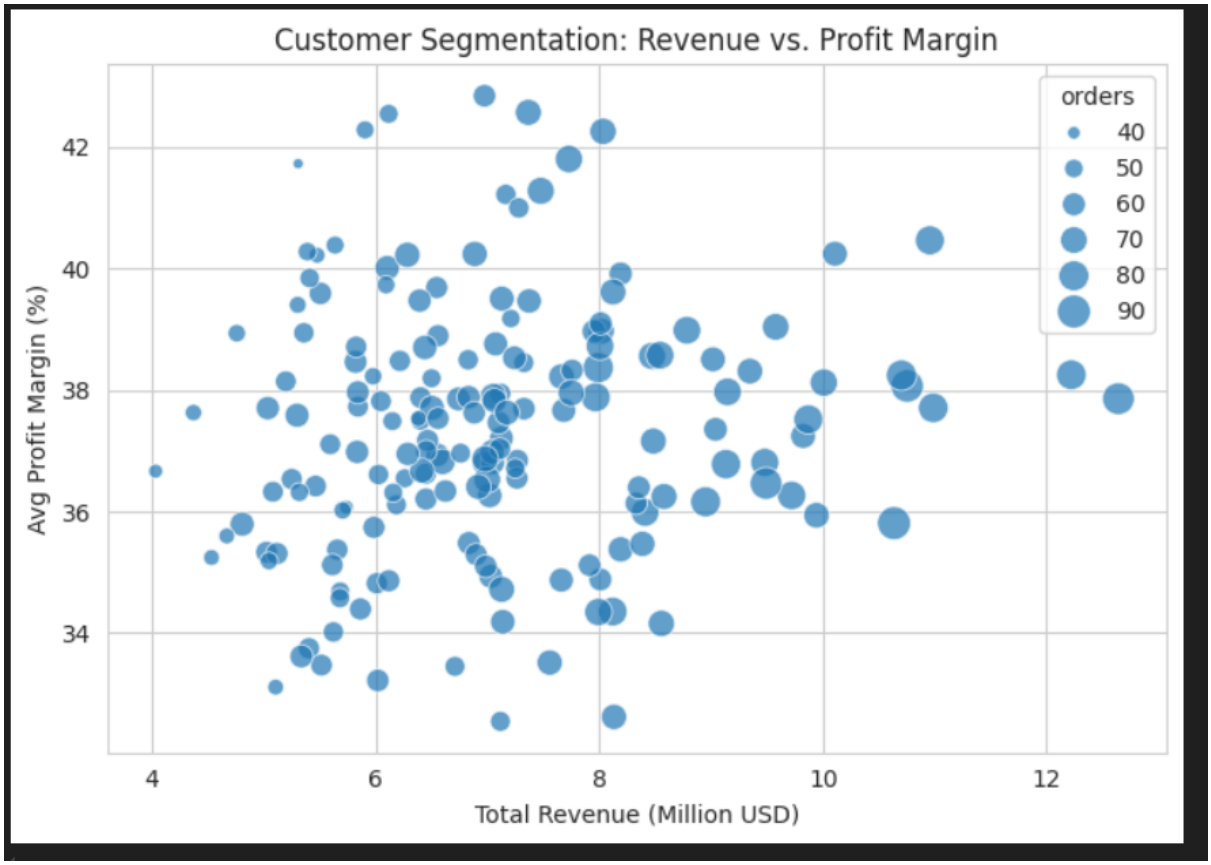
Aibox Company leads significantly as the top revenue generator.

Bottom 10 customers generate substantially less revenue (around $4-5M).

Revenue concentration: Top customers drive a disproportionate share.

Large gap: Exists between the revenue of top and bottom tier customers.

Customer Segmentation: Revenue vs. Profit Margin:



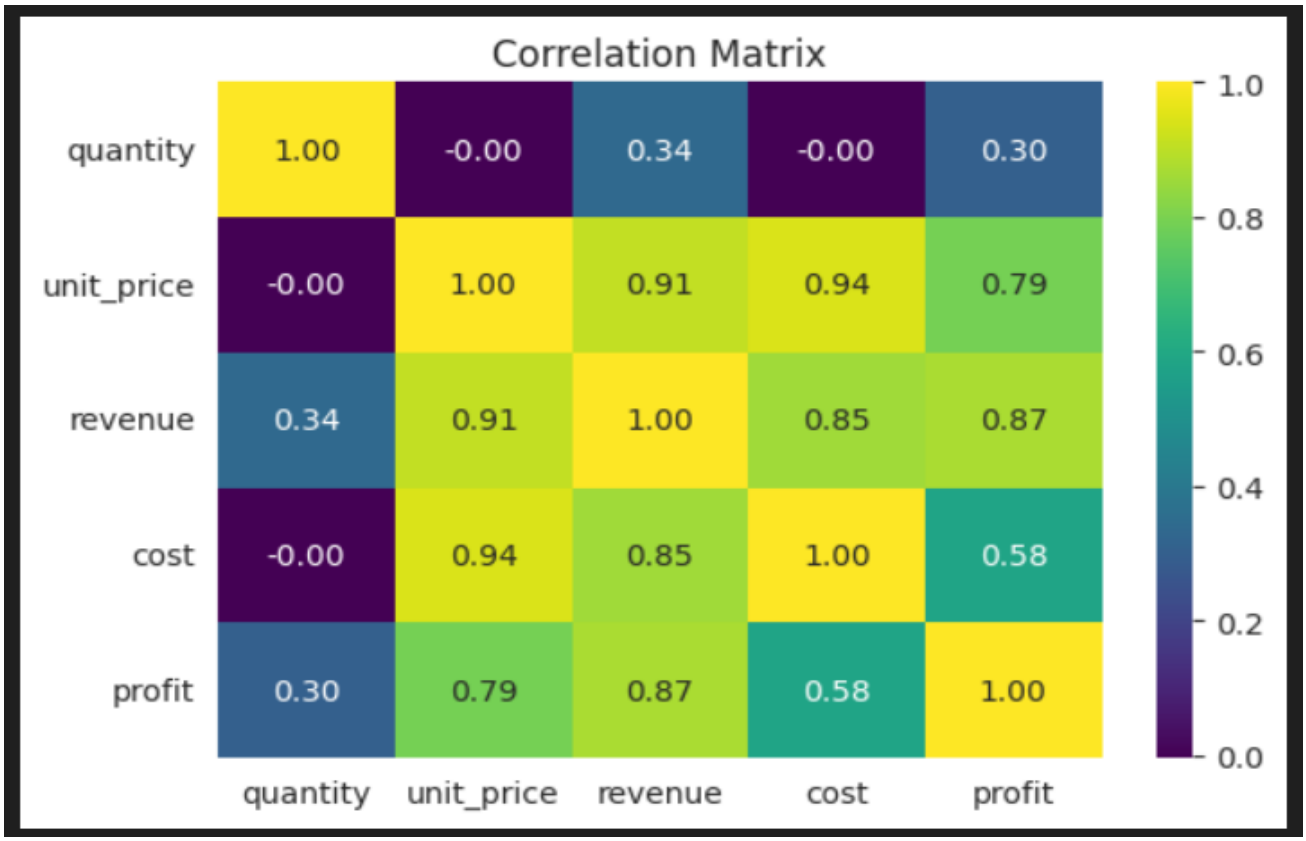
Insights:

Those Uniform 35–40 % margins confirm strong, consistent pricing and cost control.

>$10 M clients with <36 % margins reveal discounting hotspots—re evaluate large-account terms.

$6–9 M clients with >40 % margins are high-value candidates for targeted upsell.

Correlation Heatmap of Numeric Features:



Insights:

- Unit price is the primary driver, showing very strong correlations with cost (0.94), revenue (0.91) and profit (0.79).
- Revenue & profit maintain a high link (0.87), underscoring direct profitability gains.
- Quantity's impact is minimal (≤ 0.34 vs. financials), indicating volume plays a secondary role.
- Cost vs. profit correlation (0.58) is moderate, suggesting margin improvement focus should center on pricing.

Key Insights:

Pronounced Seasonality: January revenues average $124 M, dipping to $95 M in April.

SKU Concentration: Products 26 & 25 together drive ~25 % of total sales.

Channel Trade-Off: Wholesale captures 54 % of volume; Export leads with ~38 % average margin.

Geographic Dominance: California alone logs 7.6K orders ($230 M); the West region shows the largest swings.

Aibox Company and State Ltd are the most valuable customers in terms of Revenue.

Recommendations:

Seasonal Promotions: Launch recovery campaigns in April and amplify January offers to smooth revenue swings.

SKU Optimization: Double down on top products 26 & 25 and re-evaluate pricing or phase out low-margin SKUs.

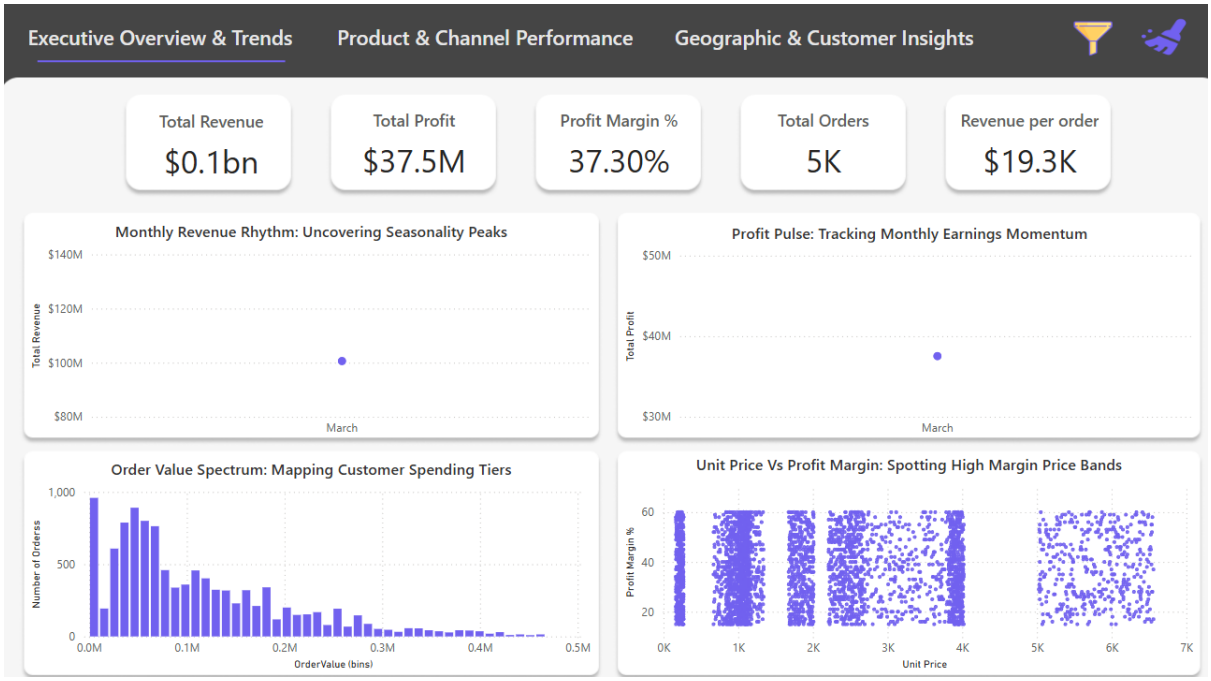
Channel Expansion: Incentivize Export partnerships for high margins and introduce volume deals in Wholesale.

Regional Investment: Replicate California's success in other regions and boost marketing in the Northeast & Midwest.

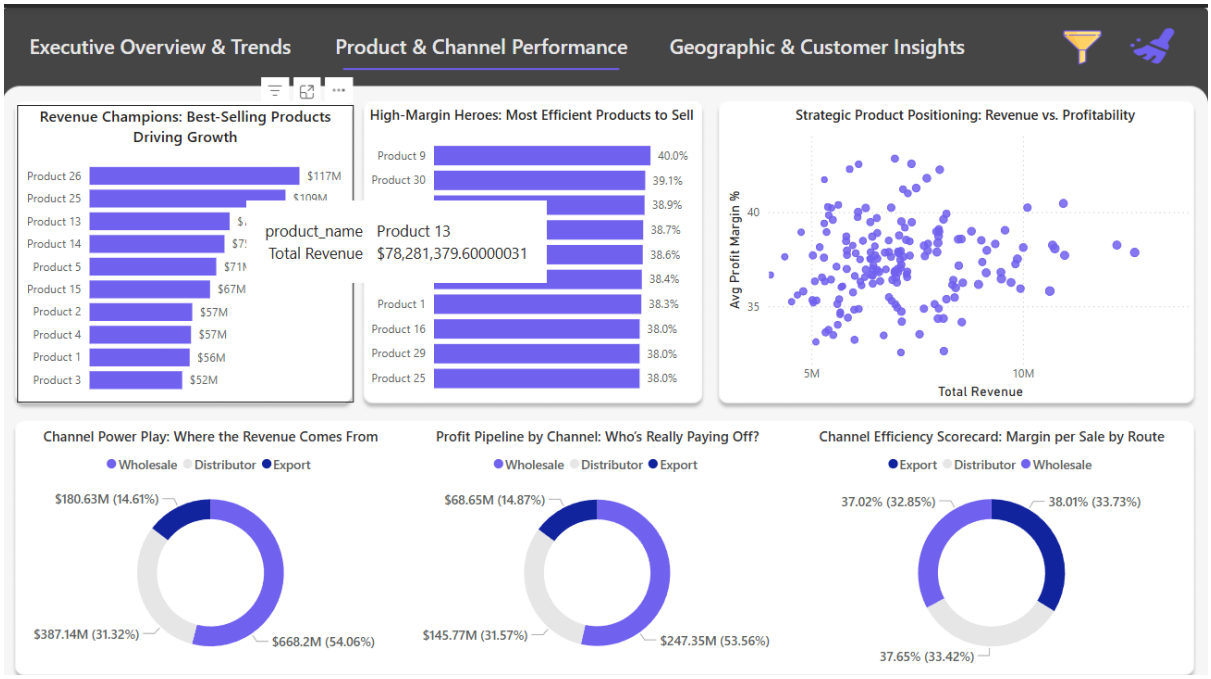
Margin Monitoring: Flag orders below 80 % margin and analyse cost drivers to uplift underperforming segments.

Dashboard

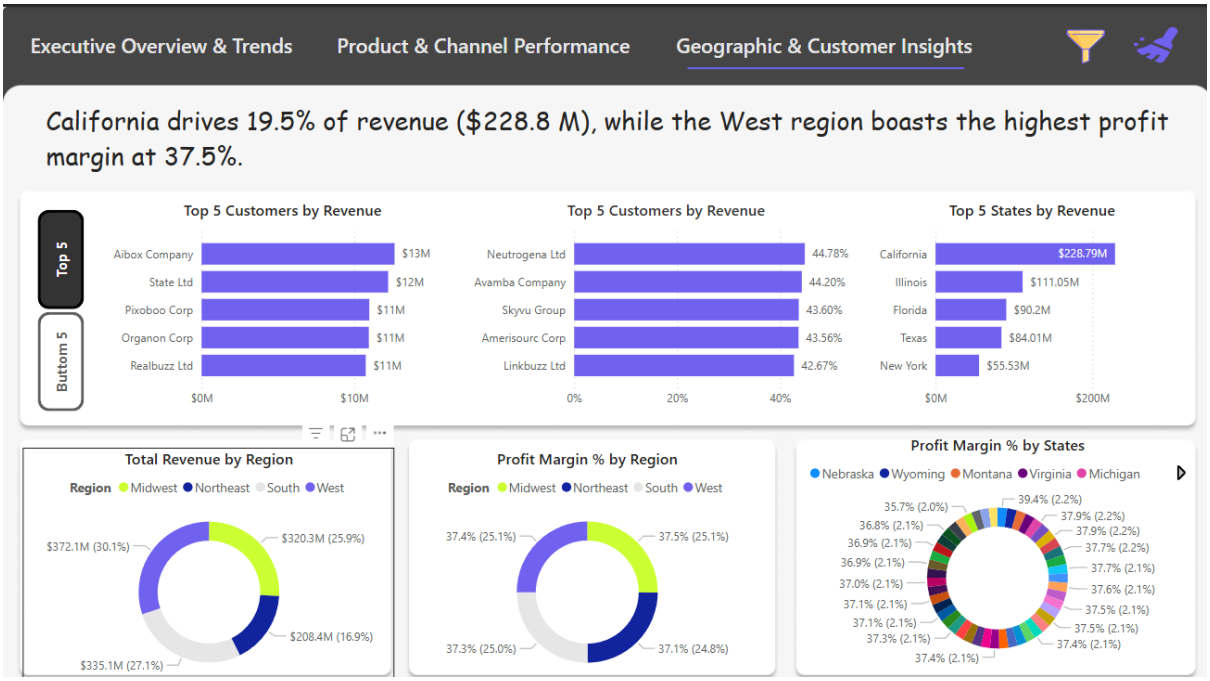
Page 1 - Performance Summary:



Page 2 - Customer Segmentation:



Page 3 - Revenue Scenarios:



Conclusions:

Completed end-to-end EDA and interactive Power BI dashboard, surfacing seasonality, SKU, channel & regional insights.

Insights inform sales policies and operational planning (store & warehouse prep aligned with annual trends).

Stakeholders can self-serve real-time analysis and confidently onboard new datasets for additional use cases.