SALES ANALYSIS FOR XYZ

Why revenue and profitability vary across U.S. states — insights & recommended actions

By : Prem Katyain | Date : 01-10-2025

Index

Abstract

In this project, a comprehensive sales analysis was conducted for a fictitious company spanning the years 2014 to 2017, aimed at uncovering key business insights and driving data-informed decision-making. Using a large and diverse dataset containing features such as product details, customer profiles, unit prices, revenues, quantities sold, and demographic data of U.S. states, several derived metrics—like profit margins, margin percentages, and product-wise revenue share—were created to evaluate performance at a granular level.

A time series forecasting model built using **Meta Prophet** was employed to predict sales for the upcoming six months, providing actionable foresight for strategic planning. To understand geographical disparities, state-wise sales variations were analyzed, while underperforming products were evaluated using the **Pareto principle**, leading to data-backed recommendations for product retention, rationalization, or discontinuation. Additionally, **K-Means clustering** was applied to segment customers based on their profitability and sales contribution, enabling the company to focus on retaining high-value customers while devising engagement strategies for low-value segments. This end-to-end analysis integrates forecasting, segmentation, and strategic insights—offering a holistic framework for optimizing sales and business growth.

Problem Statement

XYZ Company wants to identify their growth opportunities and ways to optimize their resources . They want to opt for data driven solutions for the same . You have to analyze the data to uncover regional sales trends , profitability and aid in decision making .

Business Questions

1. What factors are driving the inconsistent revenue and profitability across different U.S. states?
2. What are the key seasonal trends impacting sales and profitability, and how can the company forecast and plan better for these swings?
3. Which SKUs are the top contributors to revenue and profit, and how does their performance vary by region and channel?
4. Are there underperforming SKUs that dilute margins, and what should be the rationalization strategy?
5. How do different sales channels (Wholesale, Distributor, Export, Digital, etc.) contribute to profitability and customer growth?

Understanding the dataset

I had 6 tables namely :

1. Sales Orders : Order number, date , channel , price etc.
2. Customers: Customer names
3. Regions: Includes further information of different regions (e.g lat and long)
4. State Regions: State code, State and Regions
5. Products : List of Product Names
6. 2017 Budgets: Product wise estimated sales for the year 2017

I merged them into one table . Following are some key features of the final table :

1. Channel – The way the product is sold, like through a distributor or wholesale
2. Customer Names- The name of the person or company buying the product
3. Product Name - The specific item being sold.
4. Order quantity - How many units of the product were bought
5. Unit Price- The cost of one unit of the product
6. Revenue- The total money earned from the order (Quantity × Unit Price)
7. Total unit Cost- How much it cost the company to make or buy the products sold.
8. State- The US state where the customer is located
9. Region - The larger area grouping of states, like West or East
10. Budget- The planned or maximum sales allocation for that product

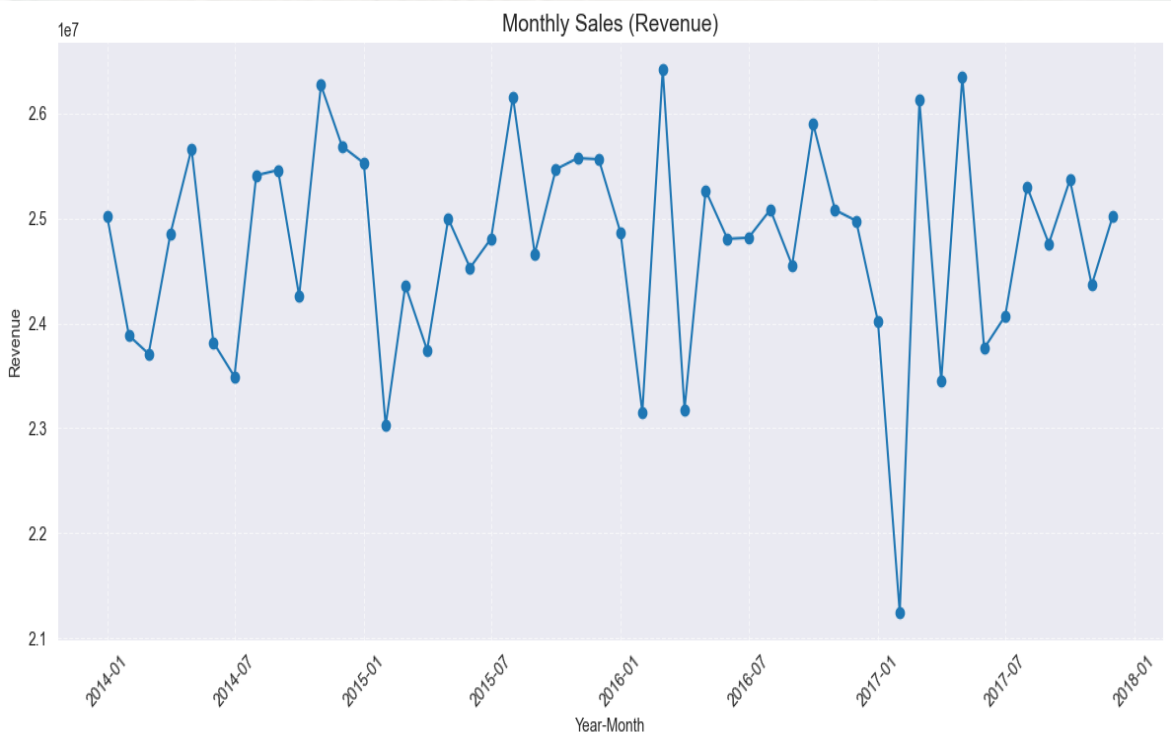
Understanding the dataset

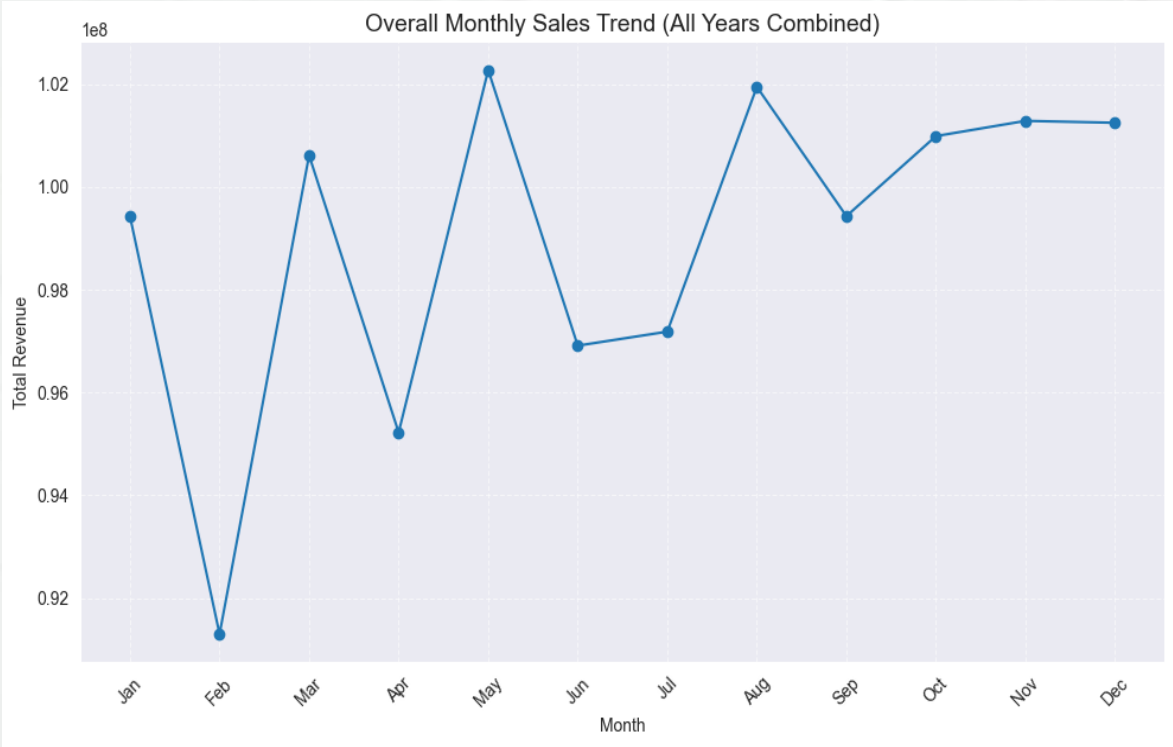
To evaluate the company's sales performance from 2015–2018, several key performance indicators were defined and analyzed. These KPIs provided a clear picture of revenue generation, profitability, and operational efficiency across different states, products, and customer segments.

- **Total Revenue & Sales Growth:** Measured overall sales trends across years, identifying periods of strong growth and seasonal fluctuations.
- **Profit & Profit Margin (%):** Calculated to assess cost efficiency and highlight products and regions delivering the highest returns.
- **Product Revenue Share:** Determined the contribution of each product to total revenue, identifying top-performing and underperforming products.
- **Cost-to-Revenue Ratio:** Evaluated the balance between total costs and revenue to measure operational efficiency.
- **Customer Contribution:** Assessed customers based on their total revenue and profit contribution to categorize them as high-value or low-value.
- **Regional Sales Performance:** Compared sales across U.S. states to uncover patterns related to population, purchasing power, and market reach.

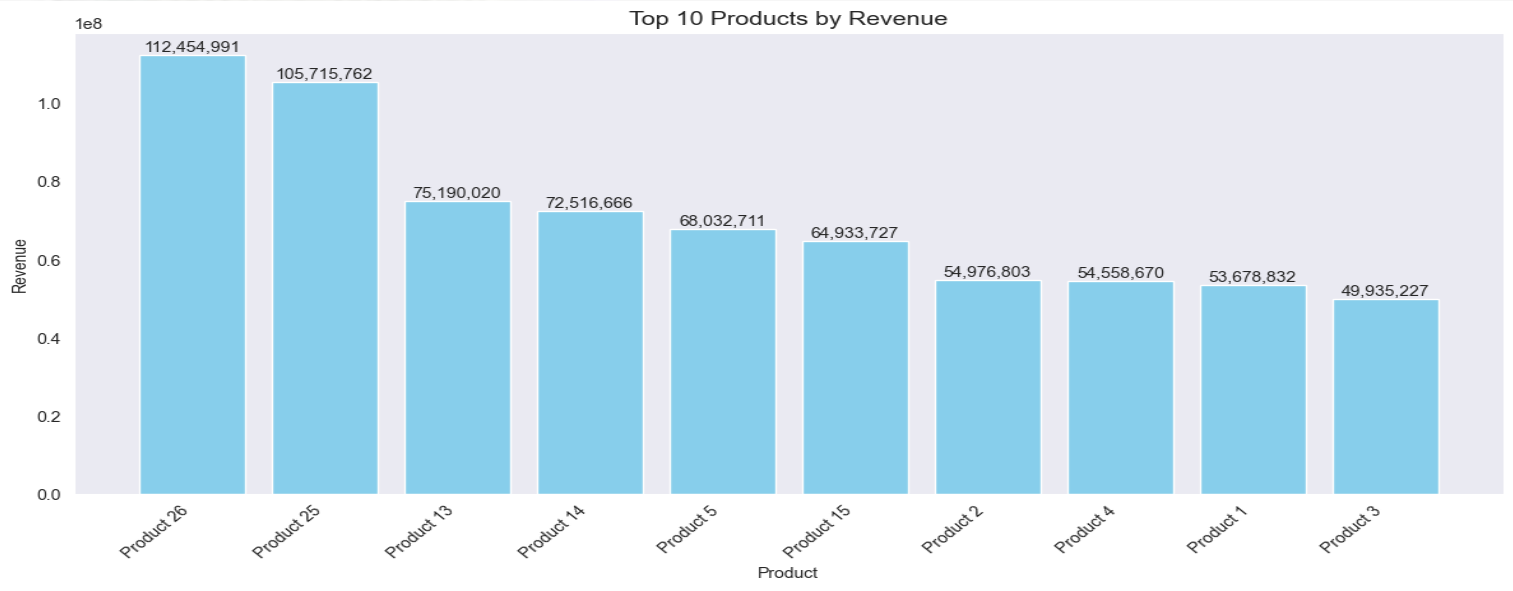
These KPIs collectively enabled the identification of core revenue drivers, profit optimization opportunities, and strategic focus areas for future growth. The insights derived from these indicators served as the foundation for deeper analyses such as forecasting, product rationalization, and customer segmentation.

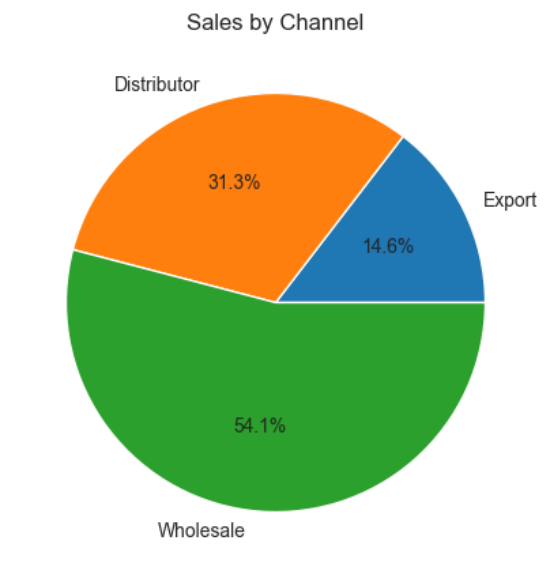
Analysis

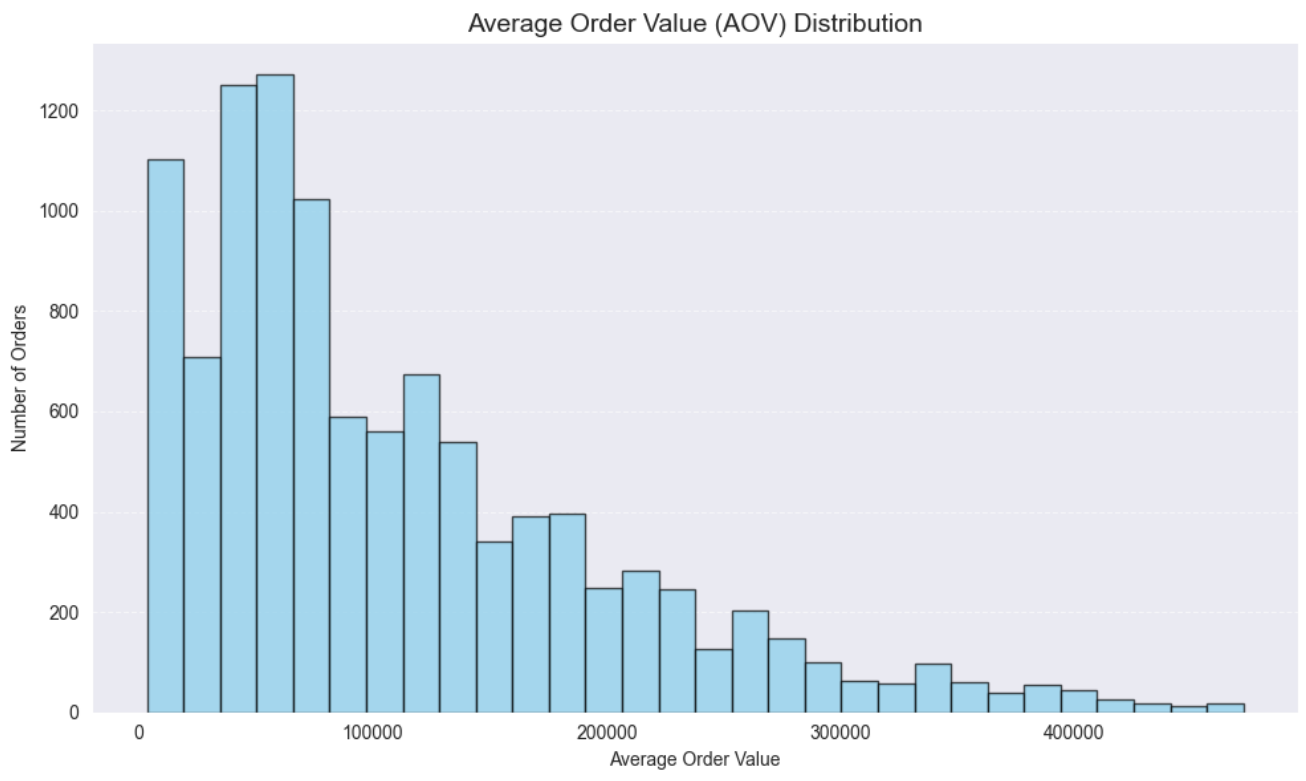




1. Seasonality can be seen
2. Second half of a year brings more revenue
3. February observes lowest revenue while May is the peak revenue month when look for all 4 years together

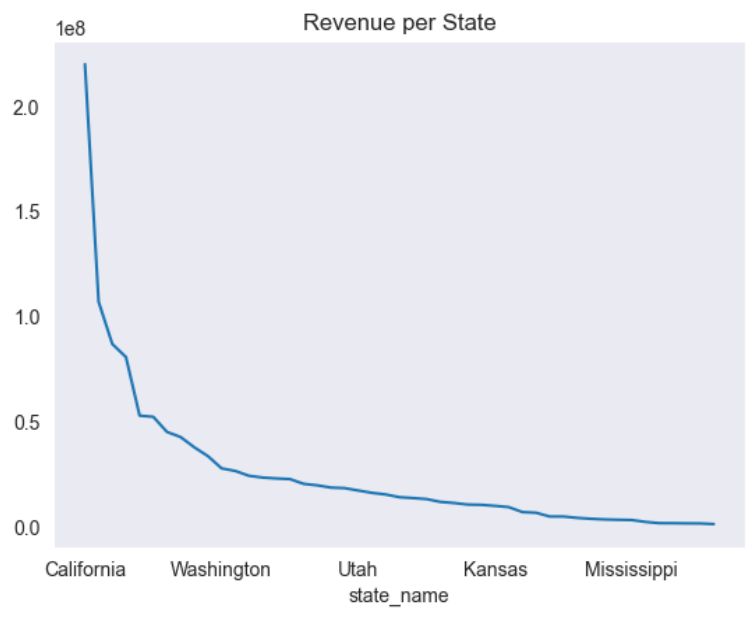


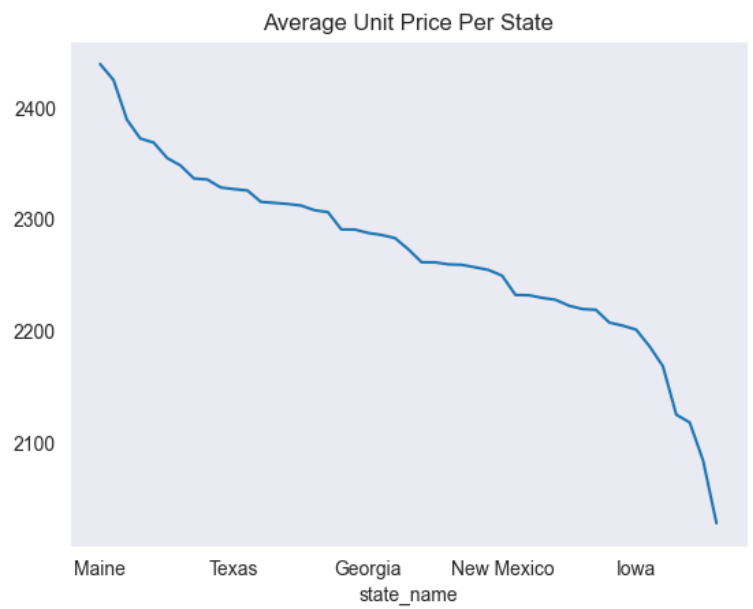


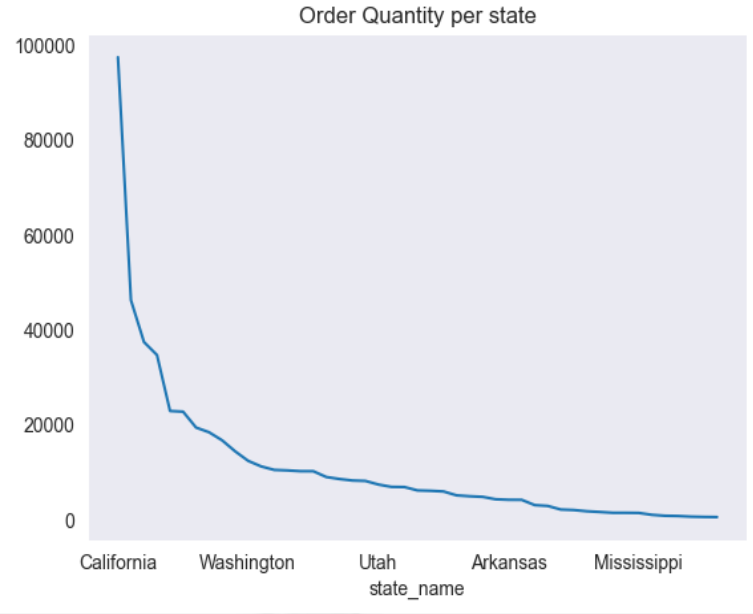


1. AOV distribution is Right Skewed
2. High frequency , Low value-transactions drives the sales
3. Product 26 and 25 are leading
4. Most % of the revenue comes from wholesale channel followed by Distributors and least from export

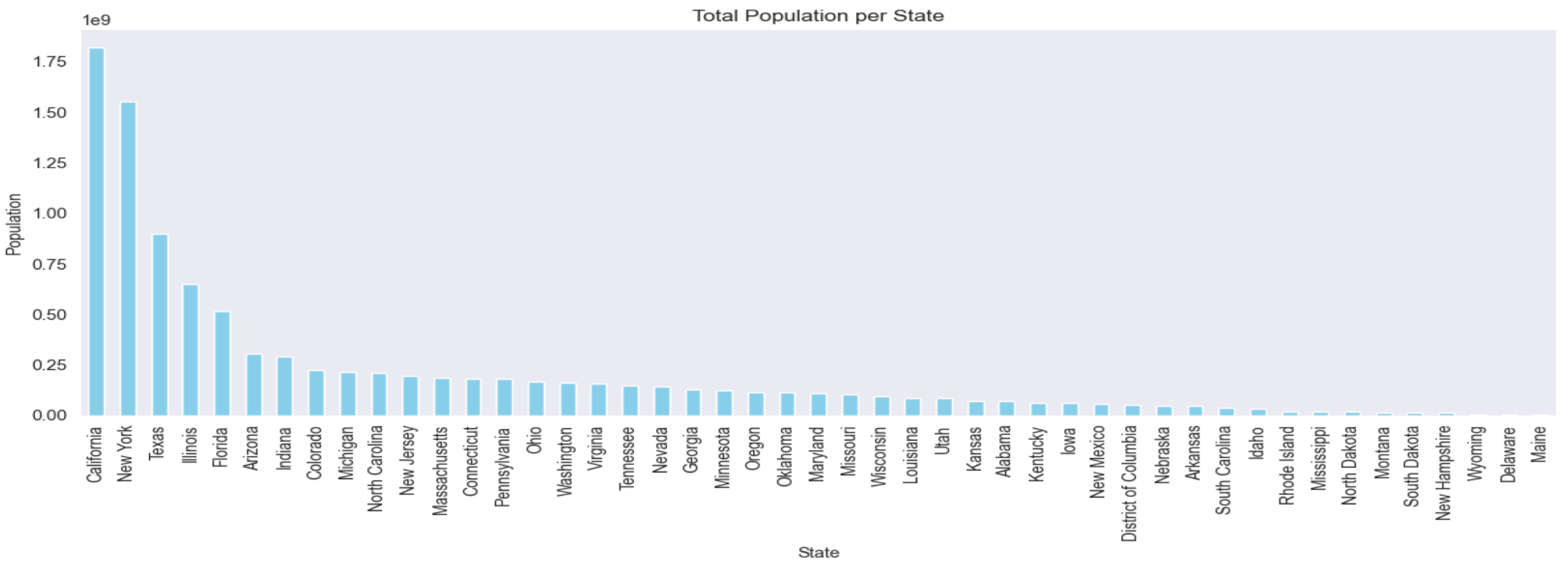
Varying Revenue Across US States

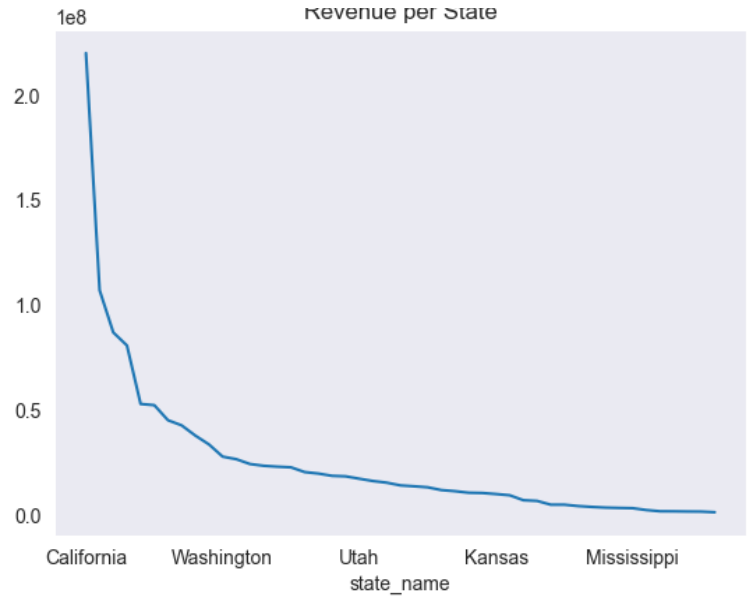






1. Revenue is not varying because of unit price
2. Order quantity variation shows the similar kind of variation as of revenue





1. As we know that revenue variation is a function of order quantity variation so , what is the reason for the variation in order quantity.
2. Diving deeper brought up Population variation as the major cause of variation in revenue across different US states.

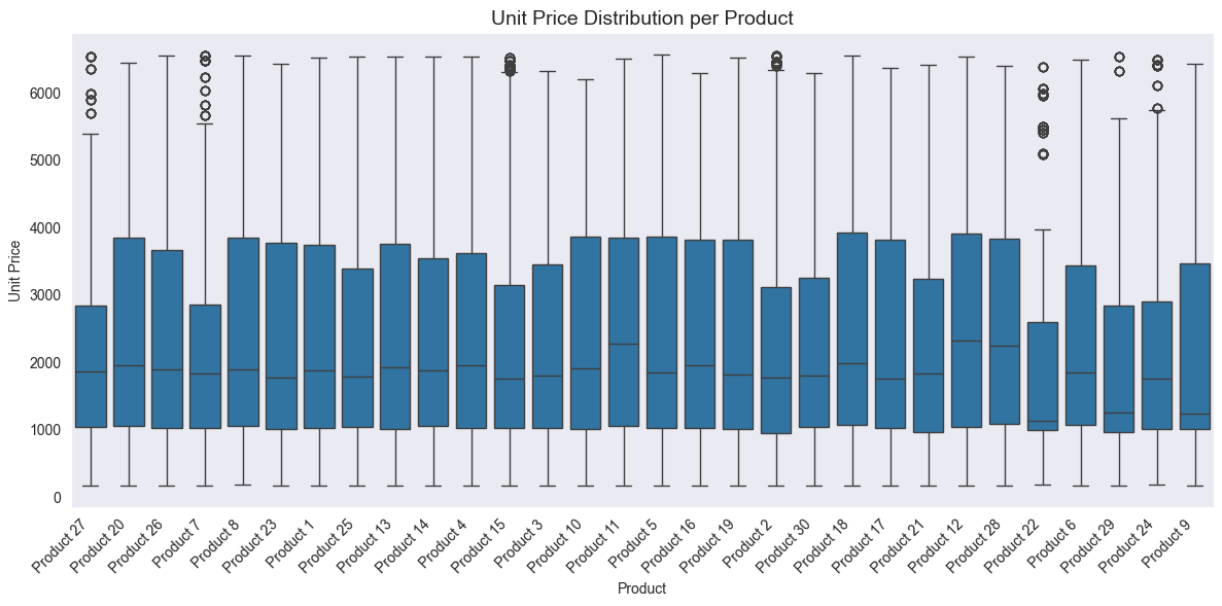
Underperforming SKU’s and Rationalization

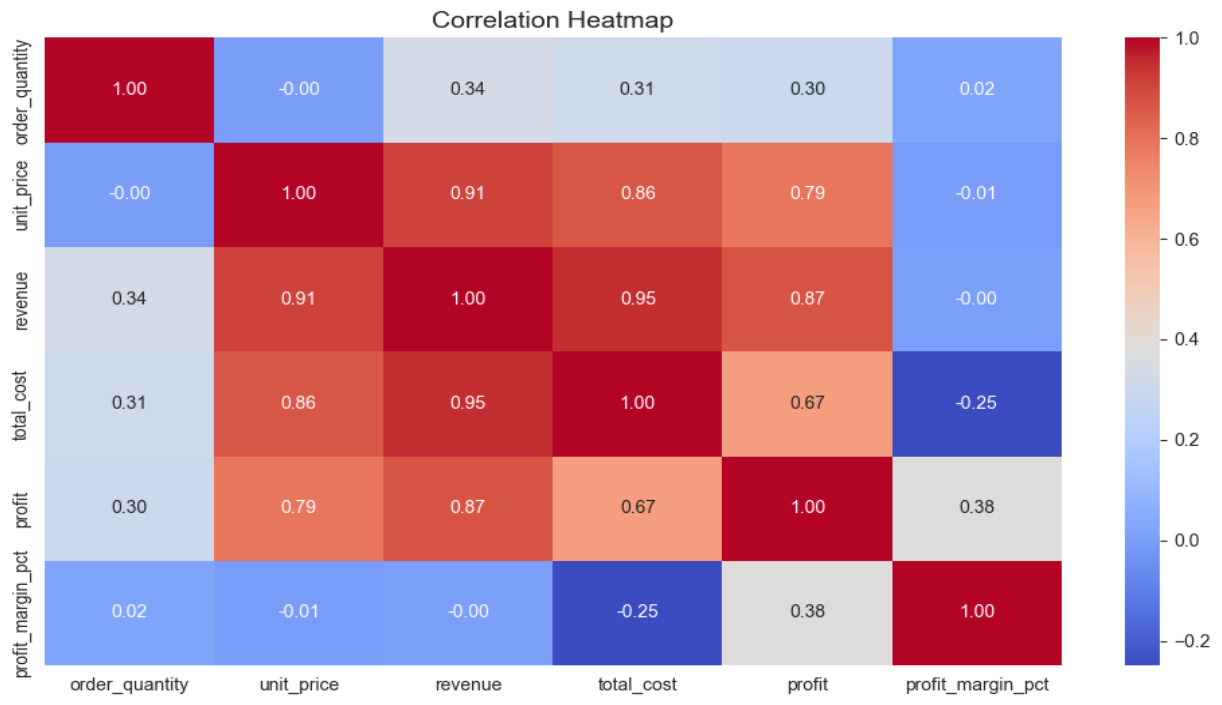
	PRODUCT_NAME	ORDER_QUANTITY	REVENUE	MARGIN	MARGIN_PCT	REVENUE_SHARE	CUM_REVENUE_SHARE	PARETO_FLAG	STRATEGY
18	Product 26	48311	112454990.9	4.252135e+07	0.375979	0.094672	0.094672	Core (Top 80%)	Retain
17	Product 25	48423	105715762.4	4.018649e+07	0.374286	0.088998	0.183670	Core (Top 80%)	Retain
4	Product 13	32726	75190019.7	2.757442e+07	0.371714	0.063300	0.246970	Core (Top 80%)	Retain
5	Product 14	31302	72516666.1	2.589333e+07	0.362352	0.061049	0.308019	Core (Top 80%)	Retain
25	Product 5	27665	68032711.2	2.547999e+07	0.379631	0.057274	0.365294	Core (Top 80%)	Retain
6	Product 15	30064	64933726.7	2.390826e+07	0.366128	0.054665	0.419959	Core (Top 80%)	Retain
11	Product 2	25286	54976803.1	1.969302e+07	0.358727	0.046283	0.466242	Core (Top 80%)	Retain
24	Product 4	23144	54558669.5	2.058514e+07	0.371080	0.045931	0.512173	Core (Top 80%)	Retain
0	Product 1	23177	53678832.2	2.059513e+07	0.381422	0.045190	0.557364	Core (Top 80%)	Retain
22	Product 3	22905	49935227.3	1.843817e+07	0.371506	0.042039	0.599402	Core (Top 80%)	Retain
9	Product 18	16990	44204569.9	1.700670e+07	0.373615	0.037214	0.636617	Core (Top 80%)	Retain
26	Product 6	17295	39022830.1	1.476922e+07	0.375598	0.032852	0.669469	Core (Top 80%)	Retain
8	Product 17	16901	38827572.0	1.455519e+07	0.377868	0.032688	0.702156	Core (Top 80%)	Retain
12	Product 20	15433	38347751.5	1.413246e+07	0.377115	0.032284	0.734440	Core (Top 80%)	Retain
7	Product 16	15905	36002496.9	1.365119e+07	0.376970	0.030309	0.764749	Core (Top 80%)	Retain
10	Product 19	14534	32965936.3	1.252999e+07	0.385206	0.027753	0.792502	Core (Top 80%)	Retain
3	Product 12	9084	23344743.0	8.043576e+06	0.369190	0.019653	0.812155	Long Tail (Bottom 20%)	Retain
2	Product 11	8577	21498879.6	7.820753e+06	0.381196	0.018099	0.830254	Long Tail (Bottom 20%)	Retain
20	Product 28	8340	20264344.3	7.873507e+06	0.381954	0.017060	0.847314	Long Tail (Bottom 20%)	Retain
28	Product 8	8070	19065131.4	6.952521e+06	0.359813	0.016050	0.863364	Long Tail (Bottom 20%)	Rationalize
13	Product 21	8195	18354006.8	7.104587e+06	0.372326	0.015452	0.878816	Long Tail (Bottom 20%)	Retain
15	Product 23	7793	18316473.4	6.956097e+06	0.378729	0.015420	0.894236	Long Tail (Bottom 20%)	Rationalize
23	Product 30	8145	17610387.2	6.862220e+06	0.376479	0.014826	0.909061	Long Tail (Bottom 20%)	Retain
19	Product 27	8513	17607358.8	6.621473e+06	0.363133	0.014823	0.923884	Long Tail (Bottom 20%)	Retain
1	Product 10	7031	16702811.9	6.047587e+06	0.372569	0.014062	0.937946	Long Tail (Bottom 20%)	Rationalize

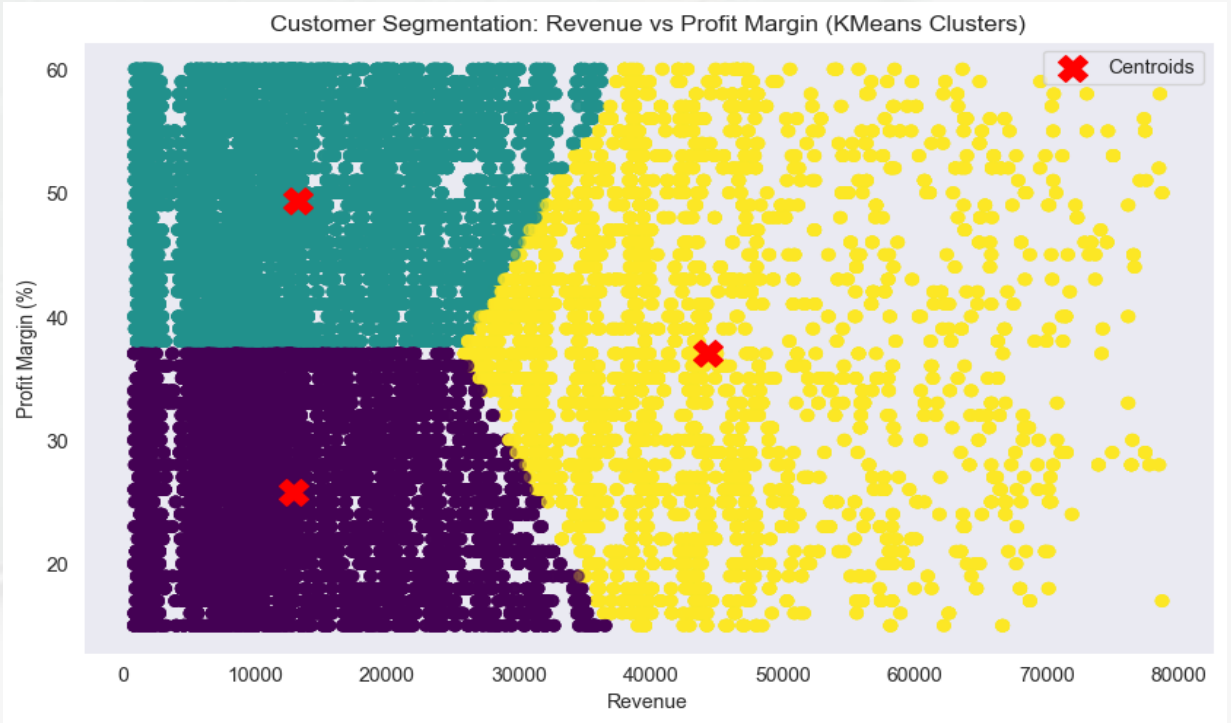
Recommended Actions

- Phase out unprofitable or low-demand SKUs to simplify inventory and focus resources on better-performing products.
- Improve marketing, pricing, packaging, or distribution to increase sales.
- Products with minimal revenue share (like Product 24, 29, 9) can be considered for discontinuation.
- Check if some SKUs are variants of the same product (size, color, flavor). Keep the most popular variant, and discontinue the rest.
- Reduce production or order frequency instead of full discontinuation.
- Keep just enough stock to fulfill sporadic demand for niche SKUs. Avoid overstocking long-tail products.
- If some long-tail SKUs have potential but low visibility, bundle them with high-selling products.

Additional Plots







Conclusion

The analysis of the company's sales data from 2015 to 2018 provided a comprehensive understanding of its overall business performance, customer behavior, and market trends across different U.S. states. By examining critical metrics such as revenue, profit, margin percentage, and product-wise revenue contribution, several key insights were uncovered that guided strategic recommendations.

The **time series forecasting model** built using *Meta Prophet* accurately predicted sales trends for the next six months, revealing a clear pattern of seasonal fluctuations and steady growth potential in specific quarters. This forecast enabled the identification of high-demand periods, empowering the company to optimize inventory levels and marketing efforts accordingly. Through **state-wise sales analysis**, it was found that sales performance varied significantly with population density and regional purchasing power—highlighting the need for region-specific sales strategies.

Product-level analysis, supported by the **Pareto technique**, demonstrated that around **20% of the products accounted for nearly 80% of the total revenue**, validating the principle's effectiveness. Based on this, products were categorized into three groups—**Retain, Rationalize, and Discontinue**—helping streamline the product portfolio for better profitability and operational focus.

The **K-Means clustering** approach provided valuable segmentation of customers into *high-value* and *low-value* groups based on their revenue and profit contributions. This segmentation guided actionable customer retention strategies aimed at rewarding loyal, high-value customers while implementing marketing efforts to engage less active ones.

Overall, the project successfully combined **data analytics, forecasting, and machine learning techniques** to deliver meaningful business insights. The study concludes that focusing on top-performing products, strengthening customer retention, and region-specific marketing, along with improving mid-range product visibility, can significantly enhance profitability. Furthermore, the integration of predictive modeling and customer analytics offers a sustainable framework for continuous performance monitoring and strategic decision-making—paving the way for consistent business growth.

Dashboard made on Power BI