

2022 KINETIX VENTURES SALES DATA ANALYSIS AND DASHBOARD CREATION PROJECT REPORT

Introduction

In this project, I worked on a sales dataset using Microsoft Excel to perform data cleaning, transformation, analysis, and visualization. The goal was to derive actionable insights from the dataset by answering specific business questions and creating an interactive dashboard. This report outlines the entire process, from data cleaning to the final dashboard, with relevant details at each step.

Project Scope

The dataset provided from [kaggle.com](https://www.kaggle.com) contains various columns such as:

- Order ID
- Date
- Product
- Category
- Purchase Price
- Sale Price
- Profit
- Quantity
- Order Status
- Delivery Date & Time
- Delivery Cost
- Revenue
- Salesperson
- Customer Information (Name, Gender, Age, State, Country)

A	B	C	D	E	F	G	H	I	J	K
Order ID	Date	Product	Category	Purchase Price	Sale Price	Profit	Quantity	Order Status	Delivery Date	Delivery Time (Days)
1001	8/12/2022	Sandles	Footwear	22	42	0	3	Cancelled	8/18/2022	2
I	J	K	L	M	N	O	P	Q	R	S
Order Status	Delivery Date	Delivery Time (Days)	Delivery Cost	Revenue	Salesperson	Customer Name	Gender	Age	State	Country
Cancelled	8/18/2022	2	5	0	Lisa Rizi	Luis Howell	Male	31	Houston	United States

The key business questions answered in this analysis include:

1. Top three best-selling products by revenue.
2. Average delivery time (in days) for each product.
3. Total revenue by month.
4. Top five states by the quantity of items purchased.
5. The age group with the most purchases.

Data Cleaning Process

Before performing any analysis, I began with cleaning the dataset. Key cleaning tasks included:

1. Removing inconsistencies: For example, I found discrepancies in the product names, such as "sandals" with both correct and incorrect spellings. I used Excel's "Find and Replace" feature to ensure uniformity.

The screenshot shows a Microsoft Excel spreadsheet with data in columns A through F. Column A contains Order IDs from 1001 to 1020. Column B contains dates from 12/8/2022 to 8/11/2022. Column C contains months from Aug to Nov. Column D contains Product names, with cell D2 currently selected and containing the value "Sandals". Column E contains Category names, and Column F contains Purchase Prices. A "Find and Replace" dialog box is open over the spreadsheet, with the "Find what" field set to "sandles" and the "Replace with" field set to "Sandals". The "Replace All" button is highlighted with a blue border. In the background, a separate list of products and categories is displayed, showing items like Refrigerator, Sandles, Sandless, Smartphone, Footwear, Appliances, and Electronics.

2. Standardizing Gender Values: I replaced inconsistent entries in the gender column (e.g., "F," "M," and "woman") with "Male" and "Female" to maintain uniformity.

The screenshot shows two side-by-side Excel tables. Both tables have columns L, M, and Q. The left table has rows for Delivery Cost, Revenue, Sales, and Age. The right table has rows for Revenue, Sales, and Age. In the left table, the "Age" column has a dropdown menu open with "Text Filters" selected. The "Text Filters" dialog box shows checkboxes for "(Select All)", "F", "Female", "M", "Male", and "Woman". The "Search" field is empty. In the right table, the "Age" column also has a dropdown menu open with "Text Filters" selected. The "Text Filters" dialog box shows checkboxes for "(Select All)", "Female", and "Male". The "Search" field is empty. This indicates that the user is using filters to identify and standardize gender values across both tables.

3. Ensuring Data Integrity: I checked all numerical columns (like Purchase Price, Sale Price, Profit, etc.) for correct formats and missing values. All values were found to be properly formatted.

Data Transformation

After cleaning the dataset, I applied two major transformations:

1. Date to Month Conversion: I extracted the month from the "Date" column using the Excel `TEXT` function, creating a new "Month" column for monthly analysis.

The image shows two screenshots of Microsoft Excel. The top screenshot displays a small dataset with columns: Order ID, Date, Month, and Product. The formula =TEXT([@Date],"mmm") is entered in the Month column for the second row. The bottom screenshot shows a larger dataset with columns: Order ID, Date, Month, and Product. The formula =TEXT([@Date],"mmm") is also present in the Month column for the second row of this larger dataset.

	Order ID	Date	Month	Product
1	1001	8/12/2022	=TEXT([@Date],"mmm")	
2	1002	4/19/2022	0	
3	1003	5/7/2022	05	Sandles
4	1004	3/8/2022	03	T-shirt
5	1005	10/30/2022	10	T-shirt

	Order ID	Date	Month	Product
981	1980	6/4/2022	Jun	Smartphone
982	1981	3/25/2022	Mar	Washing Machine
983	1982	12/2/2022	Dec	TV
984	1983	5/22/2022	May	Sneakers
985	1984	10/21/2022	Oct	Laptop
986	1985	11/24/2022	Nov	Laptop
987	1986	8/13/2022	Aug	Watch
988	1987	9/5/2022	Sep	Sandles
989	1988	4/13/2022	Apr	Jeans
990	1989	9/15/2022	Sep	TV
991	1990	11/12/2022	Nov	Refrigerator
992	1991	9/30/2022	Sep	Sneakers

2. Age Group Categorization: I used `IFS` and `AND` functions to categorize customers into three age groups:

Seniors: Age >= 60

Middle-aged: Age between 30 and 60

Younger: Age < 30

R	S	T	U	V	W	X
Age	Age Group	State	Country			
24	=IFS([@Age]>50,"Senior",AND([@Age]>30,[@Age]<=50),"Middle Aged",[@Age]<=30,"Younger")	Boston	United States			
33		Houston	United States			
62		Houston	United States			
58		Atlanta	United States			
54		Boston	United States			
56						

Age	Age Group	State
56	Senior	Boston
35	Middle Aged	San Francisco
41	Middle Aged	Chicago
26	Younger	Boston
58	Senior	Los Angeles
29	Younger	Atlanta
36	Middle Aged	Los Angeles
37	Middle Aged	Washington, D.C.
57	Senior	Atlanta
59	Senior	San Francisco
47	Middle Aged	Chicago

Data Analysis

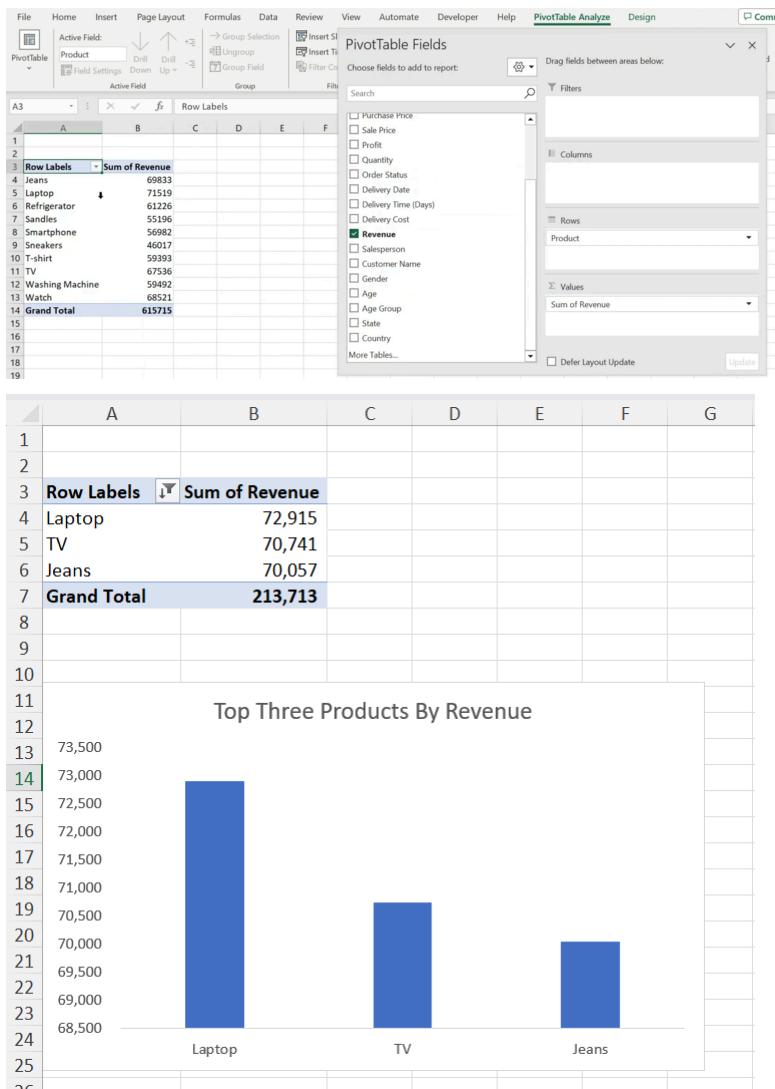
With the cleaned and transformed dataset, I proceeded with data analysis using PivotTables. The following sections outline the steps taken to answer each business question.

1. Top Three Products by Revenue

I created a PivotTable to display the top three products by revenue. A PivotChart (bar chart) was added to visualize this data.

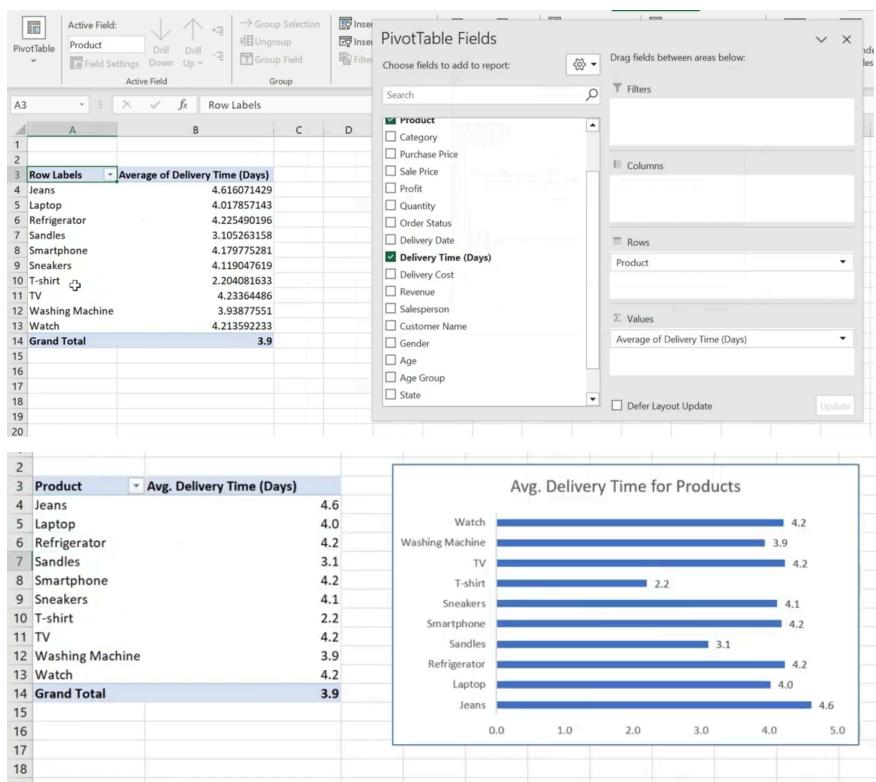
Top Products:

1. Laptop
2. Jeans
3. Tv



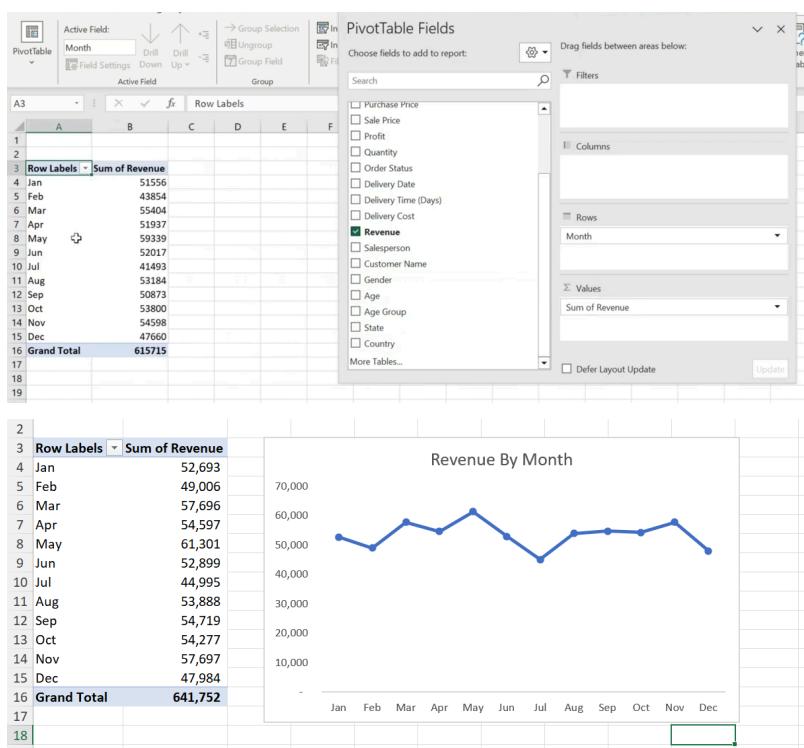
2. Average Delivery Time by Product

To calculate the average delivery time for each product, I created another PivotTable, adjusting the value field settings to display the average rather than the sum.



3. Total Revenue by Month

I analyzed revenue trends by month by utilizing the transformed "Month" column. A line chart was used to display this trend, showing monthly revenue fluctuations.

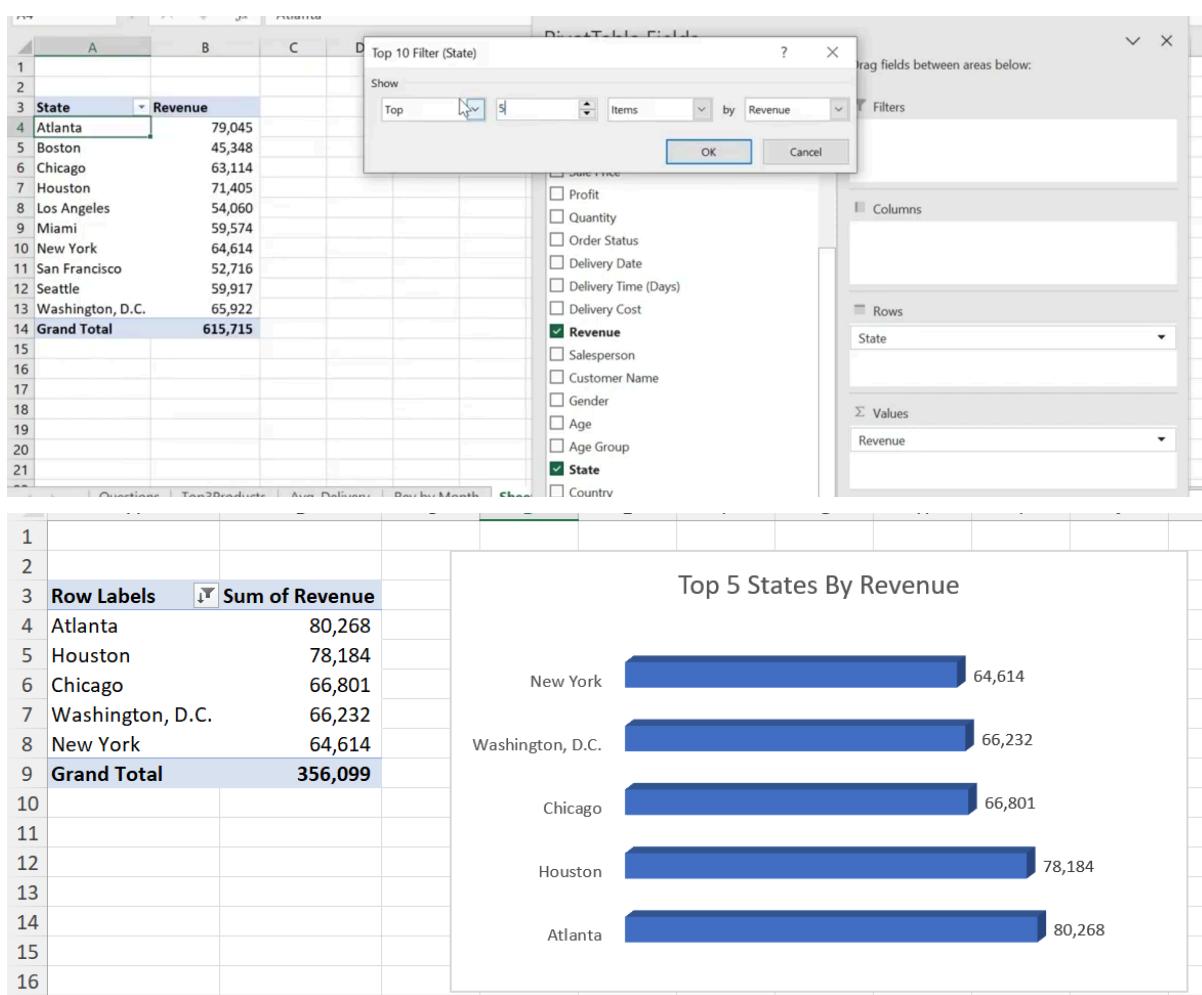


4. Top Five States by Quantity of Items Purchased

I created a PivotTable to rank states based on the quantity of items purchased. The top five states were visualized using a bar chart.

Top States:

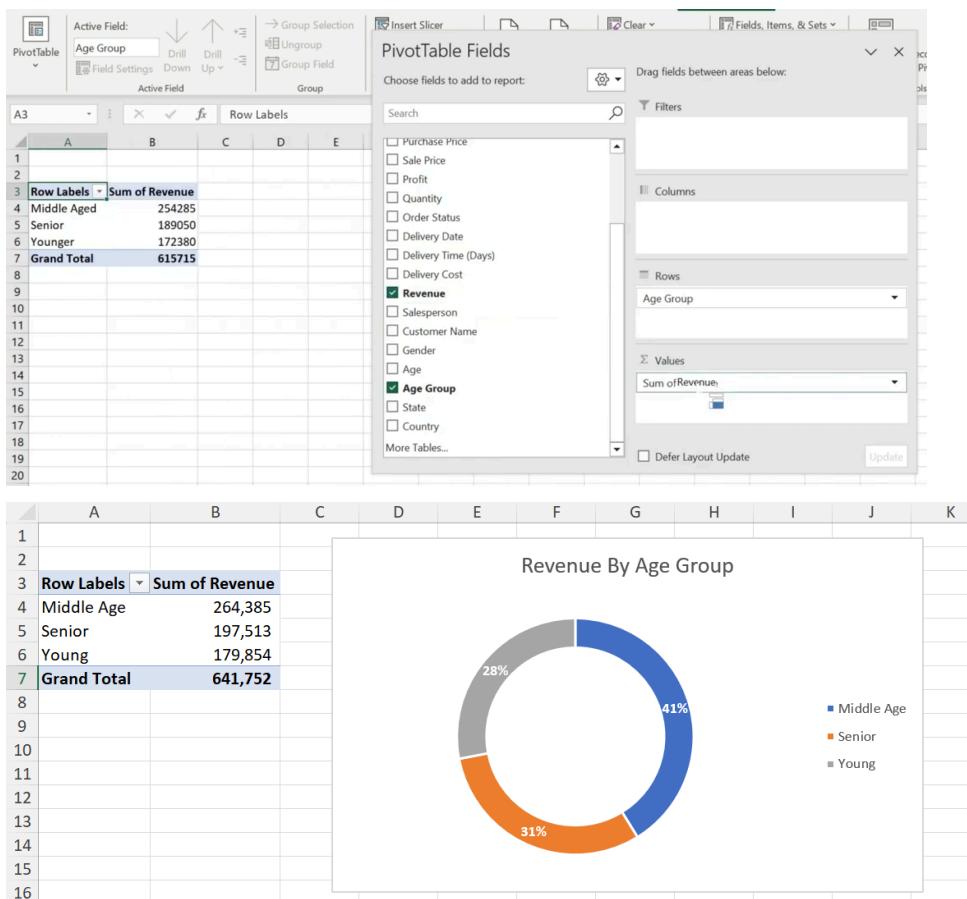
1. Atlanta
2. Houston
3. Washington D.C.
4. Chicago
5. New York



5. Age Group with Most Purchases

I grouped customers by age and calculated the total revenue generated by each group. The results were visualized using a pie chart.

- Age Group with Most Purchases: Middle-aged (30-60 years)



Dashboard Design

After addressing the business questions through thorough data analysis, I moved on to the next crucial phase of the project: designing a fully interactive and user-friendly dashboard. This dashboard is designed to provide a visual representation of the insights derived from the data, making it easier for stakeholders to interpret and act upon the findings.

The dashboard consists of the following key visualizations:

- Top Three Products by Revenue: A bar chart that highlights the top-performing products based on total revenue. This visualization allows users to quickly identify the best-selling items.
- Average Delivery Time: Another bar chart that displays the average delivery time (in days) for each product. It helps in understanding how long deliveries take on average and identifying any potential delays.
- Revenue by Month: A line chart that illustrates the revenue trends across each month, providing a clear view of how sales fluctuate over time and helping with forecasting and decision-making.
- Top Five States by Quantity of Items Purchased: A bar chart that ranks the top five states in terms of the number of items purchased. This chart is useful for identifying high-performing geographic regions.
- Age Group Purchases: A pie chart that breaks down the sales based on customer age groups, offering insights into the demographics driving the most purchases.

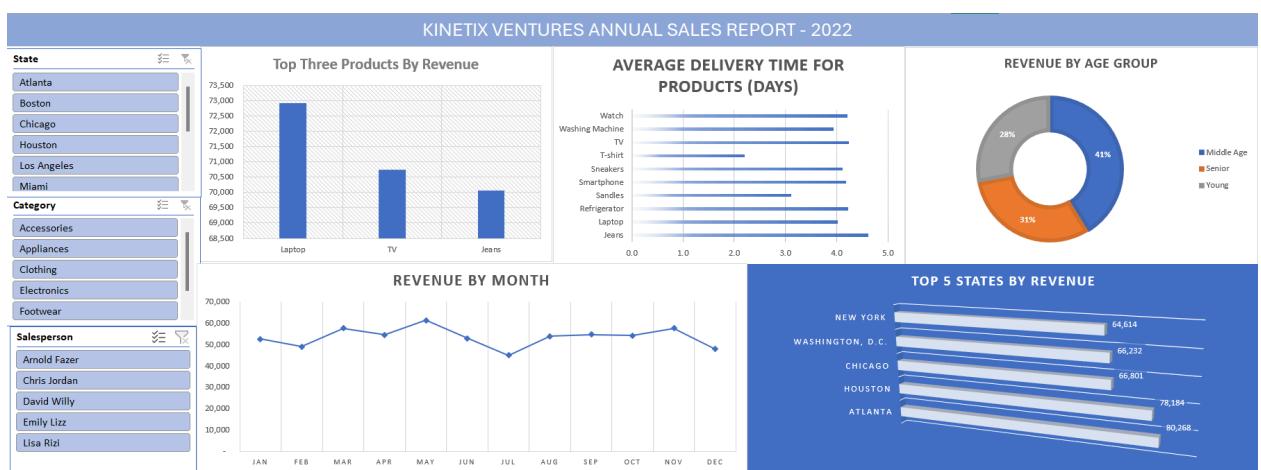
Dashboard Interactivity

To enhance the dashboard's functionality and ensure it is dynamic and adaptable to different user needs, I incorporated several slicers. These slicers add interactivity to the dashboard by allowing users to filter the data based on various criteria. The slicers included are:

- Product Category: Allows users to filter and view data based on different product categories, making it easier to analyze performance within specific product segments.

- Salesperson: This slicer provides the option to filter data by individual salesperson, enabling a detailed comparison of sales performance across the sales team.
- Gender: Allows filtering based on the customer's gender, offering insights into whether there are any gender-based trends in purchasing behavior.
- State: Filters the data by geographic location, enabling users to focus on specific states and analyze regional sales performance.

These slicers enable users to interact with the dashboard in a meaningful way, giving them the ability to explore various perspectives on the sales data. By selecting different filter combinations, users can customize their view, facilitating deeper analysis and more informed decision-making. This level of interactivity transforms the dashboard from a static presentation into a dynamic tool for ongoing business insights.



Insights and Recommendations

1. Top-Selling Products: The analysis revealed that the highest revenue-generating products are laptops, jeans, and watches. This highlights the significant demand for these items, making them prime candidates for prioritization in marketing campaigns. By focusing promotional efforts on these products, the business can capitalize on their popularity, potentially driving even greater sales and enhancing overall profitability.

Row Labels	Sum of Revenue
Laptop	72,915
TV	70,741
Jeans	70,057
Grand Total	213,713

2. **Dominance of Middle-Aged Buyers:** The middle-aged demographic, particularly individuals aged 30-60, emerged as the most active customer group in terms of purchases. This insight suggests that marketing strategies should be specifically tailored to appeal to this age bracket. Campaigns could emphasize product features, benefits, or messaging that resonate with the preferences and needs of middle-aged consumers, ensuring that the business continues to engage its largest customer base effectively.

Row Labels	Sum of Revenue
Middle Age	264,385
Senior	197,513
Young	179,854
Grand Total	641,752

3. **Regional Sales Hotspots:** The top-performing states by quantity of items purchased are Atlanta, Houston, and Washington D.C. These areas represent significant market opportunities and should be the primary focus for localized promotional activities. Targeting these states with region-specific offers, advertising, and sales initiatives can help the business strengthen its presence in these high-demand areas and further increase sales. Additionally, understanding the preferences of consumers in these regions could offer insights into tailoring the product mix to local tastes.

Row Labels	Sum of Revenue
Atlanta	80,268
Houston	78,184
Chicago	66,801
Washington, D.C.	66,232
New York	64,614
Grand Total	356,099

Conclusion

This project highlights the importance of data cleaning, transformation, and visualization in deriving actionable insights. The interactive dashboard offers a clear, dynamic way to monitor sales performance and customer trends, helping businesses like Kinetix Ventures make data-driven decisions to improve sales and customer targeting.