Order Data Report

1. INTRODUCTION:

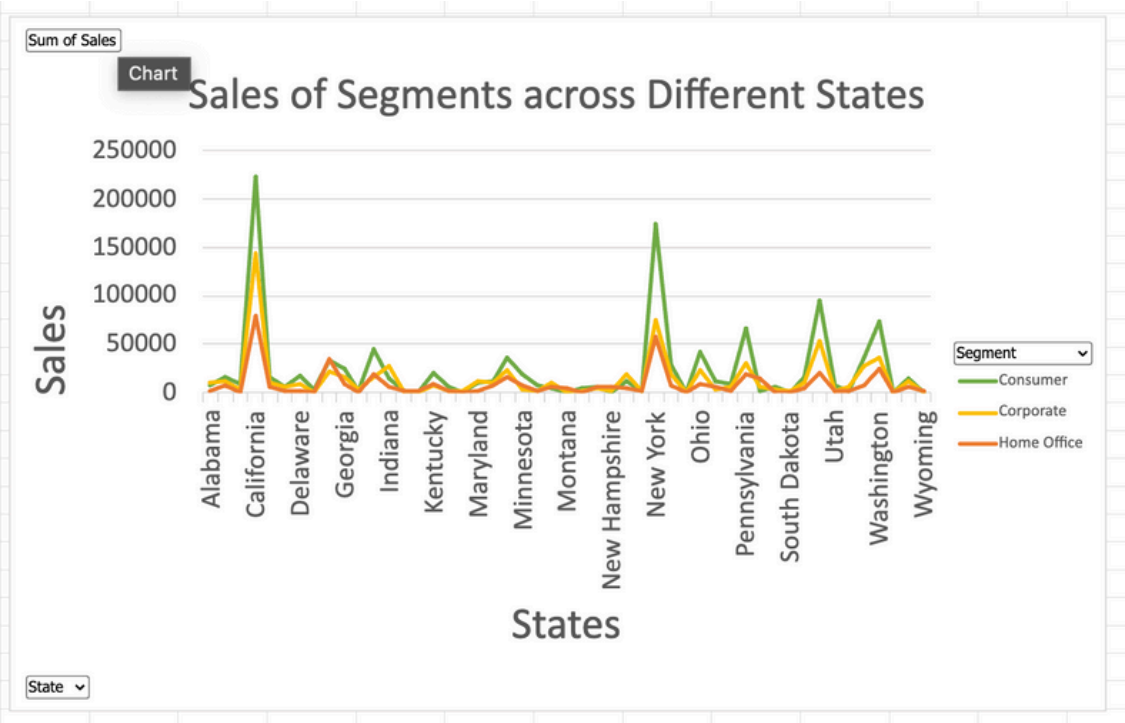
Our comprehensive dataset encompasses a vast array of variables, each elucidating distinct facets of heterogeneous category sales. Fundamental transactional attributes, such as temporal parameters and monetary exchange, are meticulously captured. Furthermore, intricate factors including clientele typology, demographic characteristics, and product classifications have been exhaustively documented, ensuring a holistic representation of the underlying phenomena.

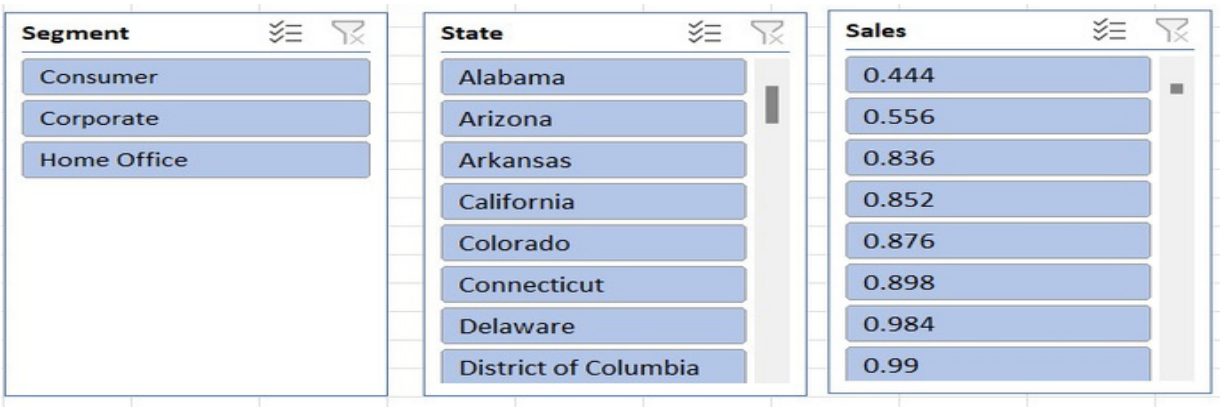
2. QUESTIONNAIRE:

1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?
2. Find out top performing category in all the states?
3. Which segment has most sales in US, California, Texas, and Washington?
4. Compare total and average sales for all different segment?
5. Compare average sales of different category and sub category of all the states.

3. ANALYTICS:

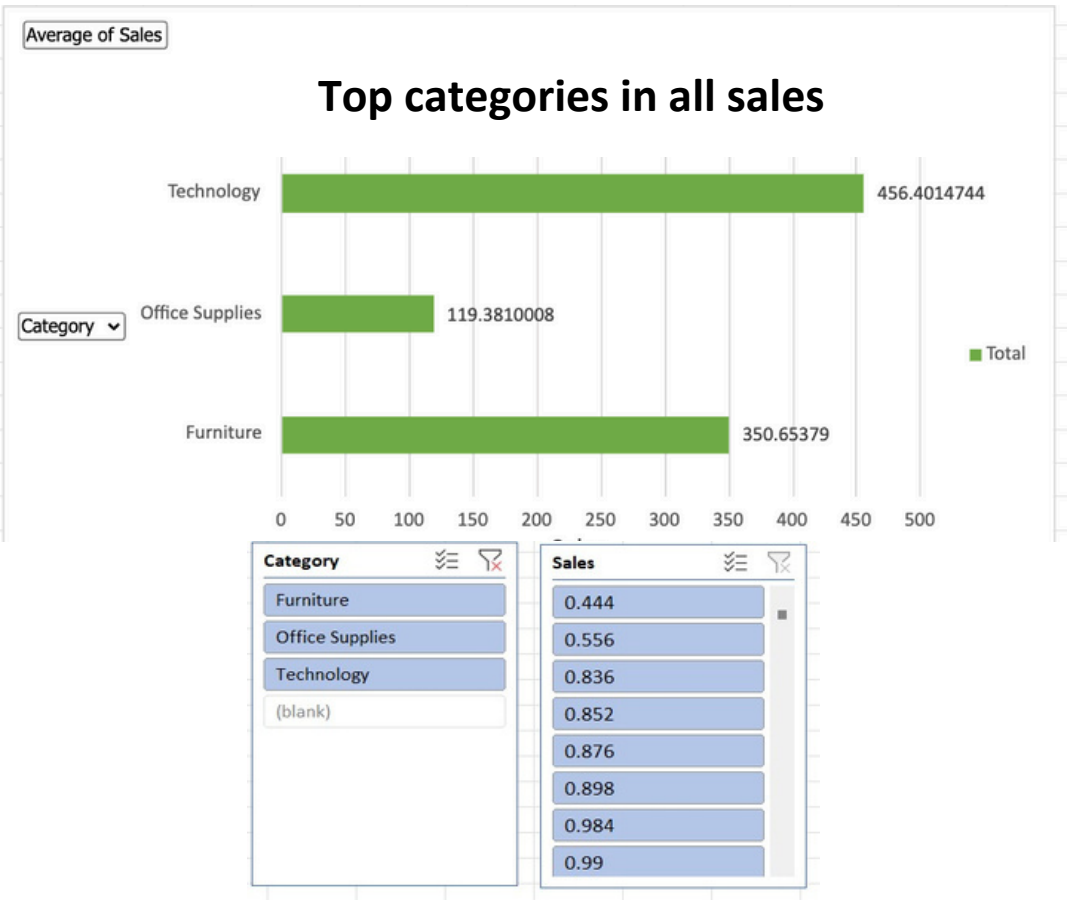
Q1. Compare all the US states in terms of Segment and Sales.Which Segment performed well in all the states?





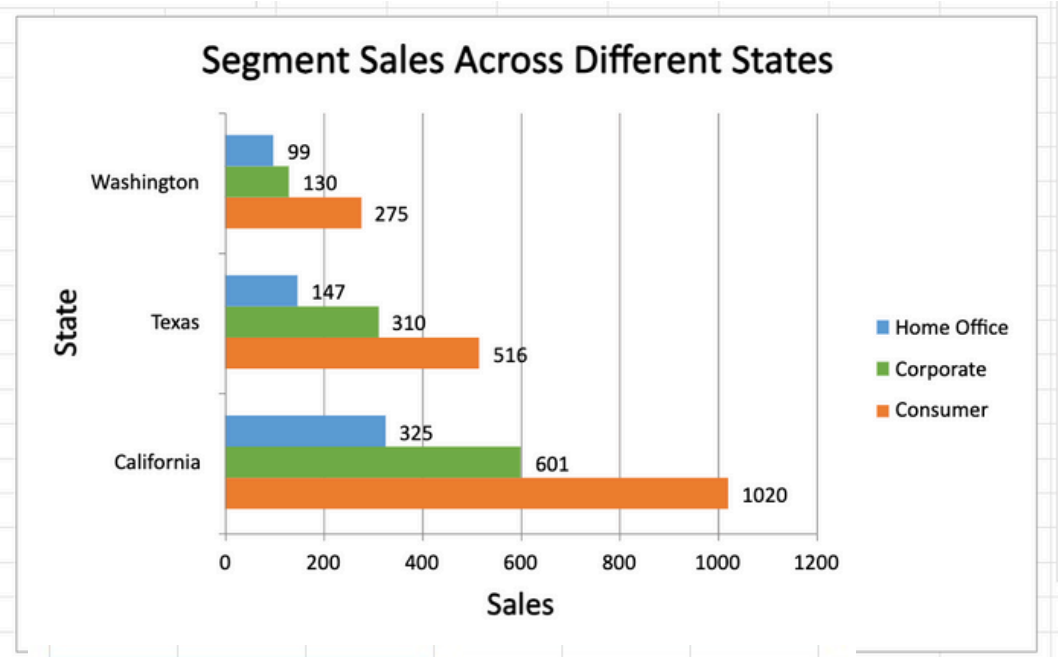
- After comparing all the states in terms of segment and sales
- California emerged as the state with the highest amount of sales Consumer segment performed well in all the states

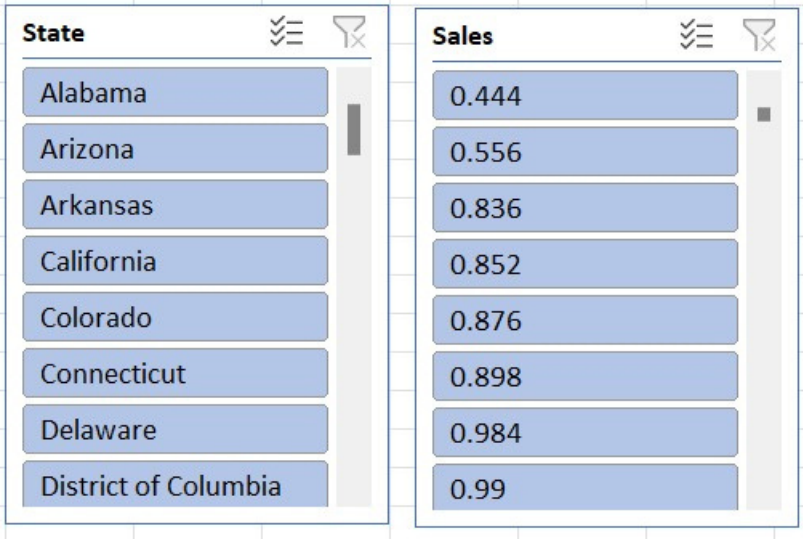
Q2. Find out top performing category in all the states?



- Office Supplies is the top performing category in all the states

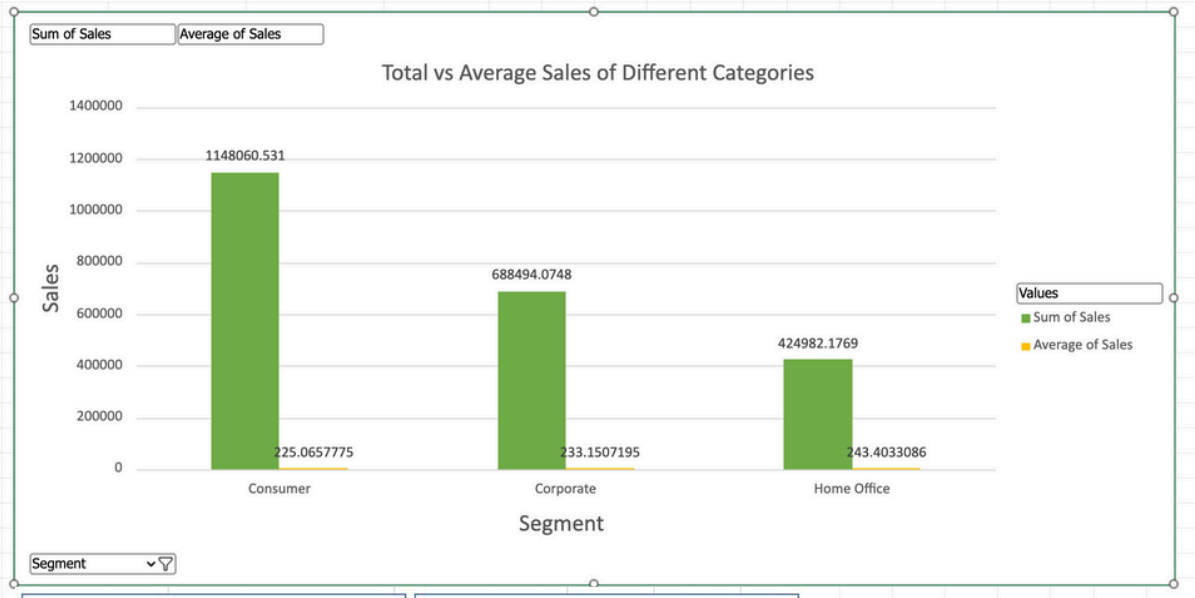
Q3. Which segment has most sales in US, California, Texas, and Washington?

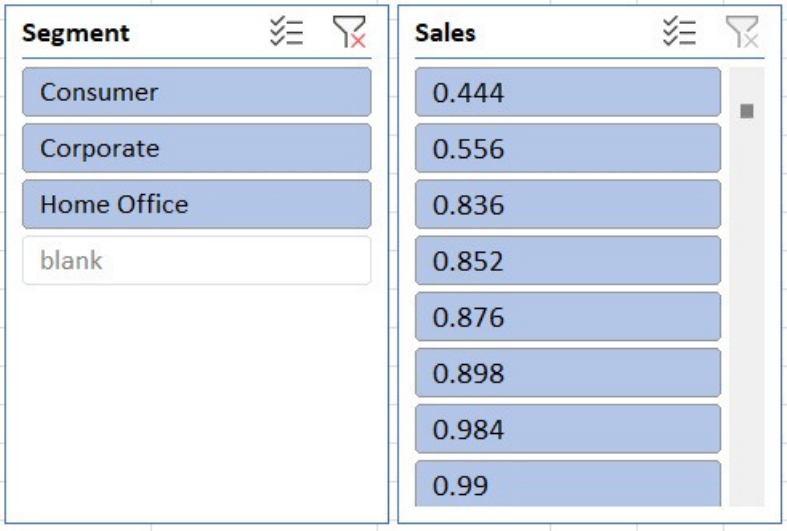




- Consumer segment has the most sales in US , California, Texas, and Washington

Q4. Compare total and average sales for all different segment?

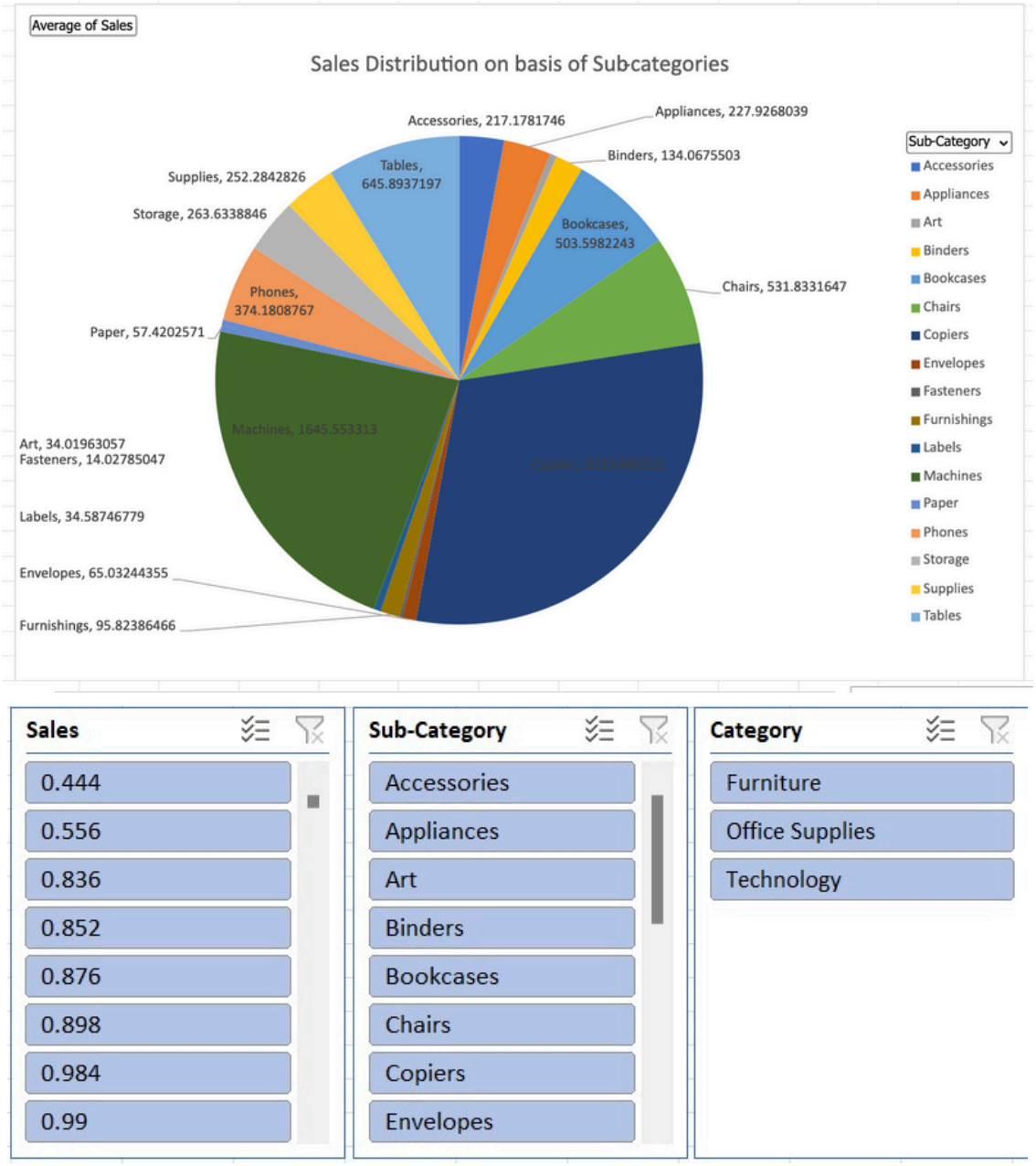




Answer:

The analysis reveals that total sales significantly exceeded the average sales for each segment.

Q5. Compare average sales of different category and sub category of all the states.



Answer.

In the analysis of Average Sales vs. Category, it is observed that the Technology category has the highest sales.

Additionally, within the Average Sales Distribution by Sub-Categories, Copiers had the highest contribution with a value of 2215.

Regression and ANOVA:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.003989285							
R Square	1.59144E-05							
Adjusted R Square	-0.000484829							
Standard Error	525.2842121							
Observations	1999							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	8769.275428	8769.275428	0.031781546	0.85852604			
Residual	1997	551019236.5	275923.5035					
Total	1998	551028005.8						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	232.5093842	22.96786412	10.12324799	1.59291E-23	187.4658976	277.5528709	187.4658976	277.5528709
Postal Code	-6.54575E-05	0.000367174	-0.178273794	0.85852604	-0.000785542	0.000654627	-0.000785542	0.000654627

This regression output provides information about the relationship between the predictor variable (Postal Code) and response variable (Sales). Let's break down each section of the output:

Regression Statistics:

Multiple R: This is the correlation coefficient, measuring the strength and direction of the linear relationship between the predictor and response variables. In this case, it's very close to zero (0.00399), indicating a very weak linear relationship.

R Square: This is the coefficient of determination, representing the proportion of the variance in the response variable explained by the predictor variable. A value close to zero (1.59144E-05) suggests that the predictor variable explains very little of the variance in the response variable.

Adjusted R Square: This modified version of R Square adjusts for the number of predictor variables in the model. A negative value (-0.000484829) suggests that the model may be overfitting or that the predictor variable does not add explanatory power to the model.

Standard Error: This represents the average deviation of the observed values from the regression line. A higher standard error (525.2842121) indicates greater variability in the data points around the regression line.

Observations: This indicates the number of data points used in the regression analysis.

ANOVA (Analysis of Variance):

df: Degrees of freedom denote the number of independent data points available for analysis.
SS: Sum of squares quantifies the total variance observed in the response variable.
MS: Mean square represents the average variance within and between groups.
F: The F-statistic evaluates the overall significance of the regression model. A low F-value compared to the critical value suggests that the model lacks significance.
Significance F: This p-value corresponds to the F-statistic. A high p-value (0.85852604) indicates that the model is not statistically significant.

In summary, this regression output suggests that the model does not meaningfully explain the variation in the response variable. The predictor variable (Postal Code) does not have a significant impact on the response variable.

Correlation:

Descriptive Statistics:

	Sales
Mean	230.7691
Standard Error	6.33014
Median	54.49
Mode	12.96
Standard Deviation	626.6519
Sample Variance	392692.6
Kurtosis	304.4451
Skewness	12.98348
Range	22638.04
Minimum	0.444
Maximum	22638.48
Sum	2261537
Count	9800

The correlation coefficient (0.024067424) has an absolute value close to zero. This indicates a very weak linear relationship between the two variables.

4. Conclusion and Reviews:

Conclusion:

In exploring sales data across various segments in different US states, our analysis has yielded valuable insights. The dataset provided a comprehensive perspective, including essential variables such as geographical location, product categorization, transactional details, and performance metrics. Through meticulous examination, we addressed pertinent inquiries and extracted actionable conclusions.

California emerged as a focal point, demonstrating the highest sales volume among the analyzed states. Particularly, the consumer segment displayed consistent performance across all states, highlighting its significance in the market landscape.

Office supplies emerged as the top performer across categories in all states, indicating a universal demand for these products. Additionally, the consumer segment exhibited dominance in sales across the US, encompassing California, Texas, and Washington.

Analysis of total versus average sales per segment revealed significant disparities, with total sales surpassing average sales universally. This underscores the presence of outlier transactions or high-value sales within each segment.

Further exploration into category and sub-category analysis revealed technology as the category with the highest average sales, suggesting a robust market demand for technological products. Subsequently, copiers emerged as the top contributor to average sales distribution within the technology category, underscoring their significance.

Reviews:

1. Thorough Analysis with Actionable Insights: The examination of sales data across US states provides a comprehensive grasp of market dynamics. The incorporation of key attributes and performance metrics facilitates a robust analysis, enabling stakeholders to derive actionable insights for strategic decision-making.

2. Clear Presentation of Findings: The presentation of findings is concise and organized, facilitating easy comprehension of complex data. The use of slicers enhances visual representation, aiding in the interpretation of results and facilitating informed decision-making.

3. Insightful Conclusions: The conclusion succinctly summarizes key findings and draws meaningful conclusions from the analysis. By highlighting overarching trends and significant observations, it provides valuable guidance for market strategies and future research endeavors.

Overall, the exploration of sales data offers valuable insights into market trends and consumer behavior, serving as a foundation for informed business strategies and market interventions.