

# EXPLORING SALES OF DIFFERENT SEGMENT IN US STATES

## Introduction:

Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of different category sales. From fundamental transactional details such as Date, Time, sales, states to more nuanced factors like Customer Type, Demographics, category and sub category, every facet has been meticulously documented.

## Key Attributes:

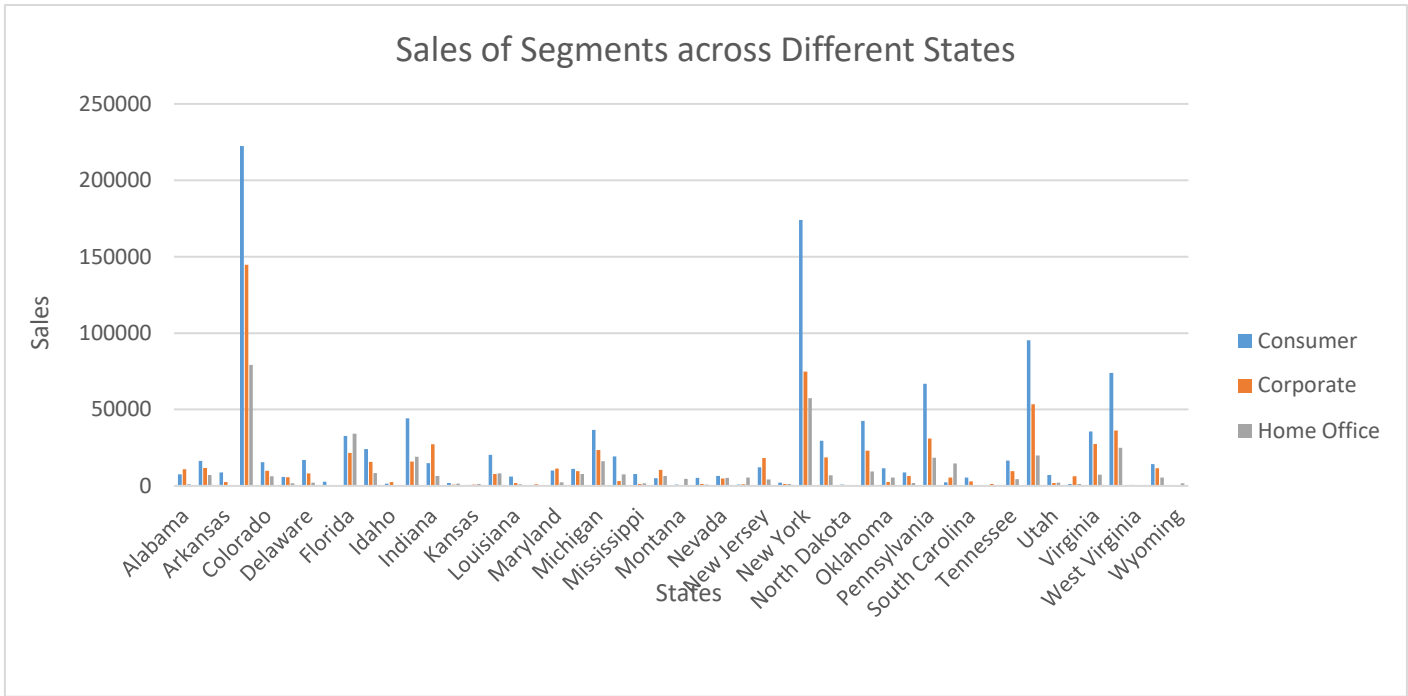
1. ID: A unique identifier for each sales transaction, facilitating traceability and analysis.
2. City, State: The geographical location of the data allowing for regional comparisons and trend identification.
3. Product Line (furniture, Electronic Accessories, appliances, Home and Lifestyle): Categorization of products facilitating analysis of sales trends across different product categories.
4. Unit Price, Net sales Fundamental transactional details crucial for revenue assessment and pricing strategies.
5. Net sales of different category, category performing well in different states: Performance metrics
6. Rating: different product performing well in different state
7. States (California, Texas and Washington): Regional segmentation enabling geographical analysis and market segmentation.

## Questinnaire:

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.

# Analytics:

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

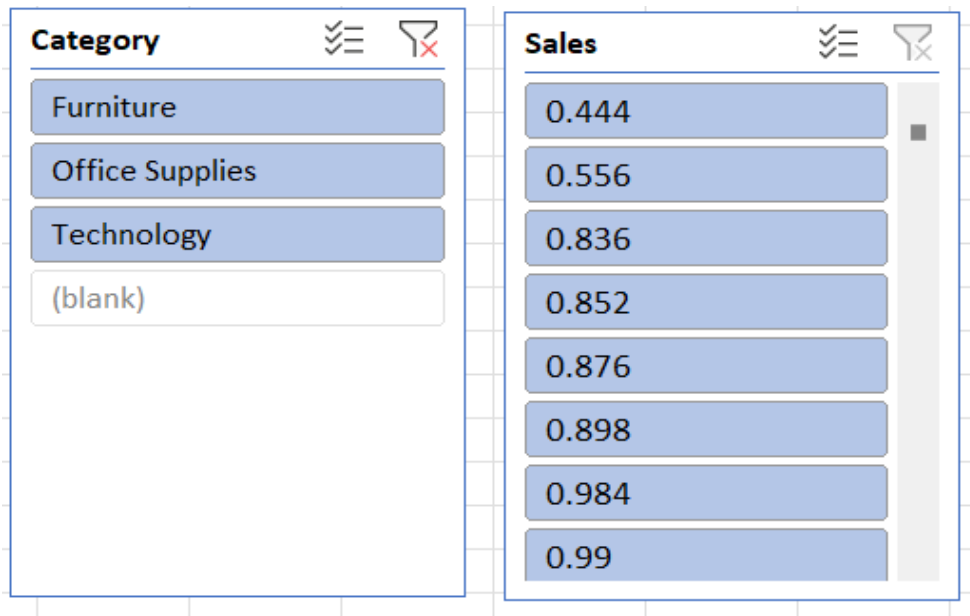
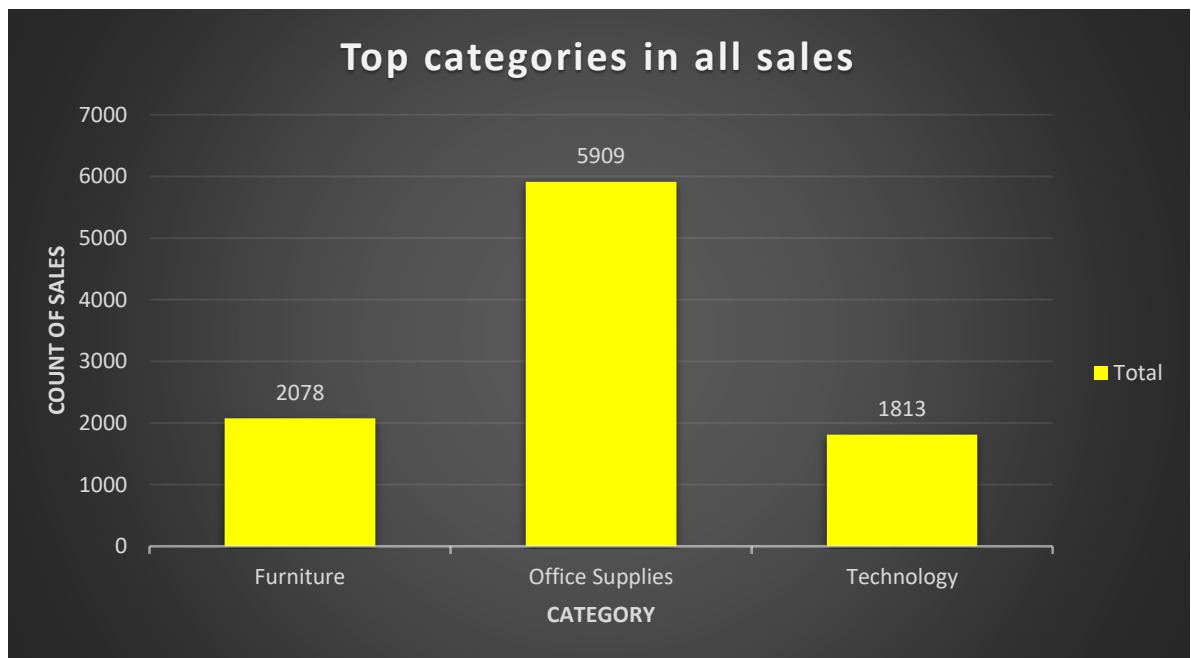


Segment
Consumer
Corporate
Home Office

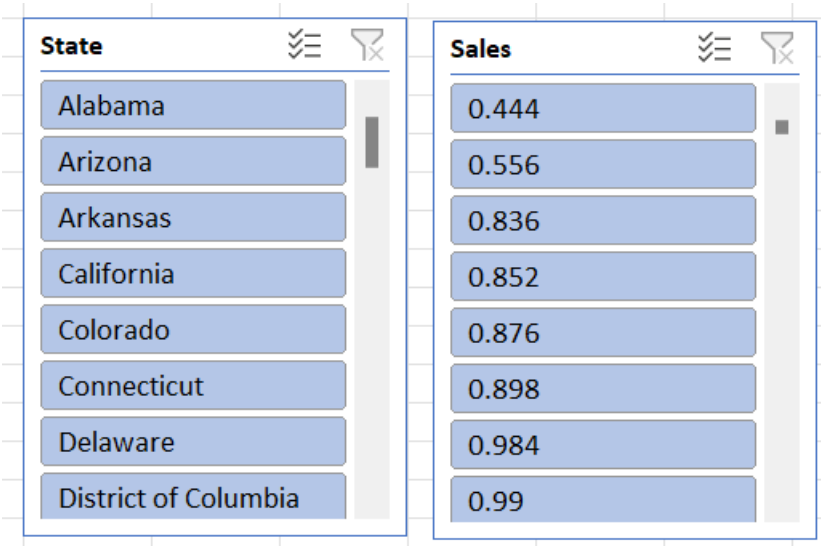
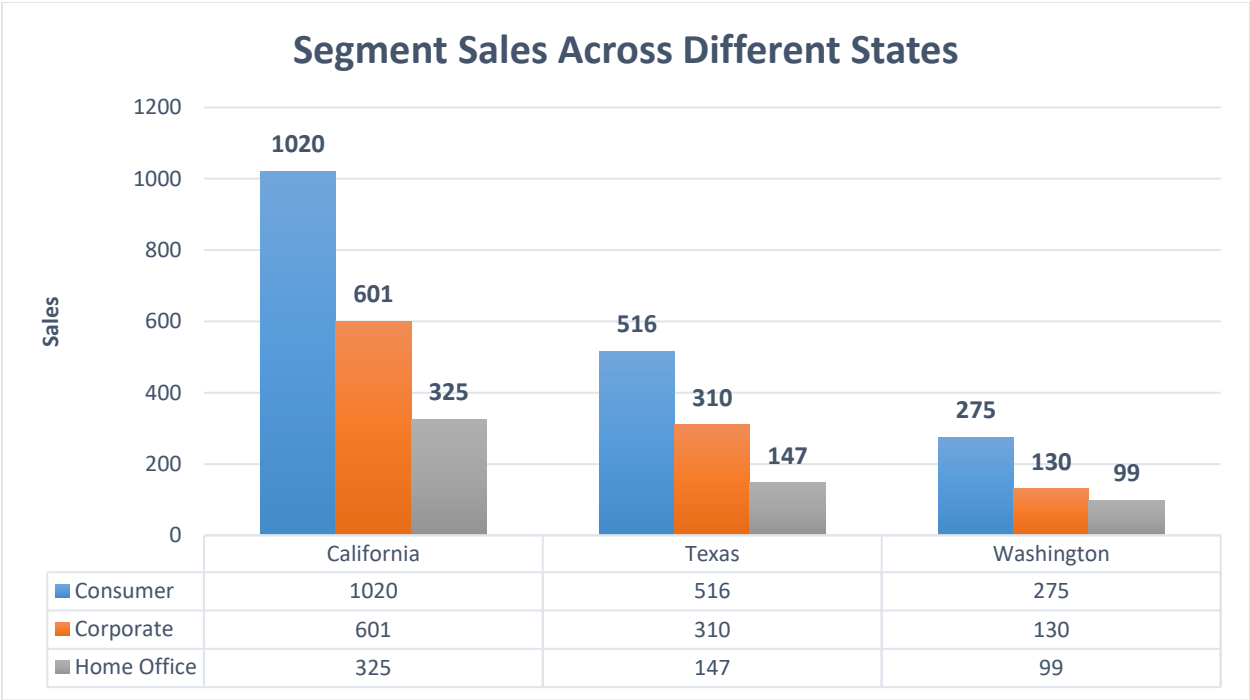
State
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia

Sales
0.444
0.556
0.836
0.852
0.876
0.898
0.984
0.99

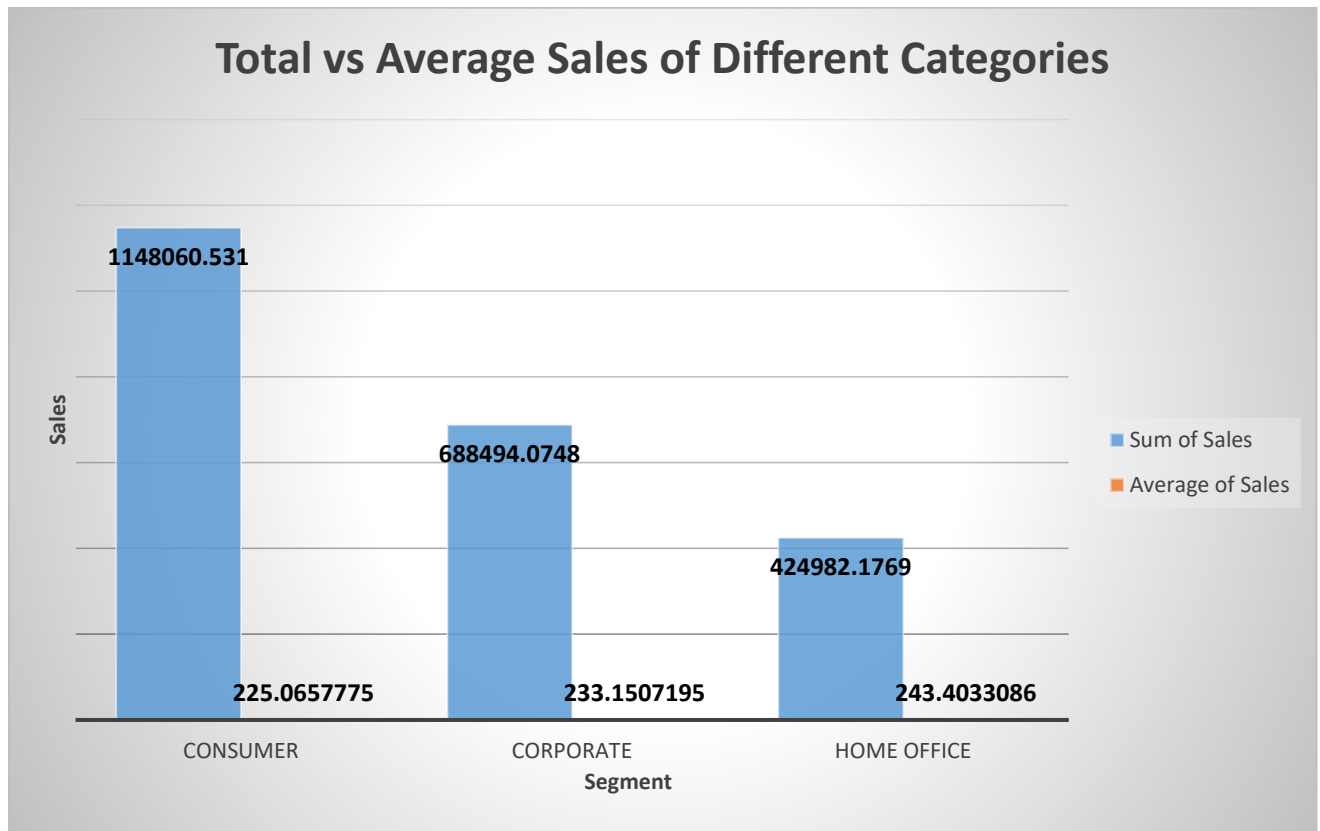
**Q2.** Find out top performing category in all the states?



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

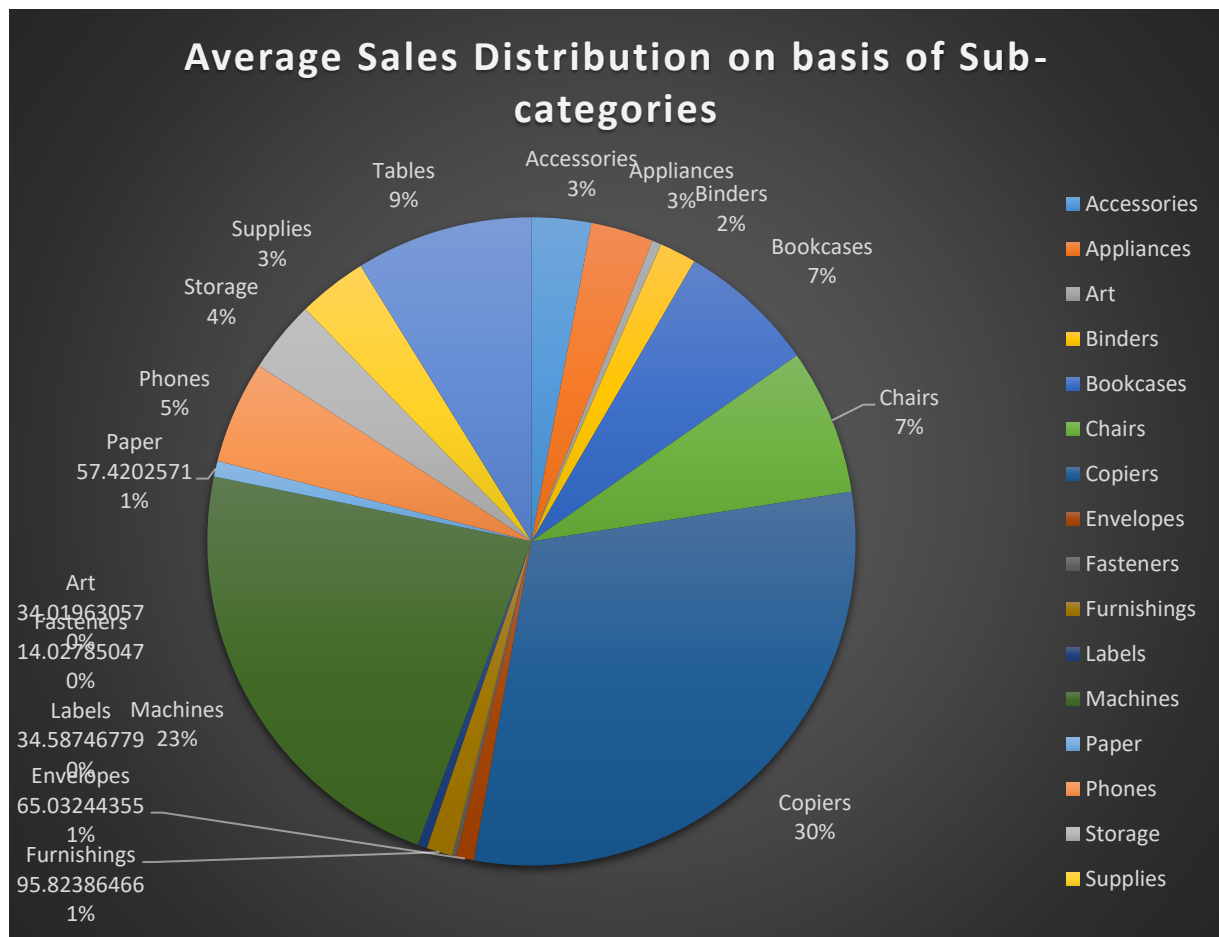
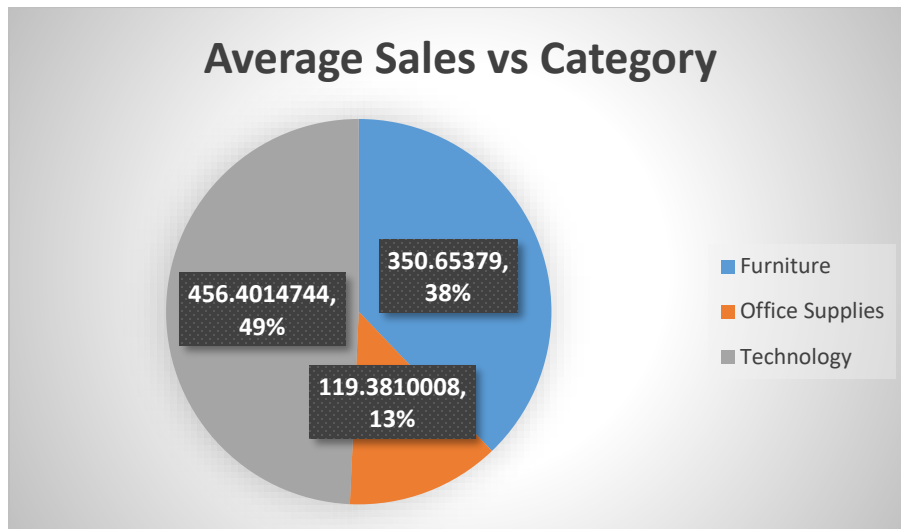


**Q4.** Compare total and average sales for all different segment?



Segment	Sales
Consumer	0.444
Corporate	0.556
Home Office	0.836
blank	0.852
	0.876
	0.898
	0.984
	0.99

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



Sales	Sub-Category	Category
0.444	Accessories	Furniture
0.556	Appliances	Office Supplies
0.836	Art	Technology
0.852	Binders	
0.876	Bookcases	
0.898	Chairs	
0.984	Copiers	
0.99	Envelopes	

**Q6.** Find out state wise mode for Customer and Segment. California, Illinois, New York, Texas, Waashington

Ans.

For California: Customer- William Brown, Segment-Consumer

For Illinois: -Customer- William Brown, Segment-Consumer

For New York: Customer- William Brown, Segment-Consumer

For Texas: Customer- William Brown, Segment-Consumer

For Washington: Customer- William Brown, Segment-Consumer

## Regression and ANOVA:

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.003989285							
R Square	1.59144E-05							
Adjusted R Square	-0.000484829							
Standard Error	525.2842121							
Observations	1999							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
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 the response variable (which is not explicitly mentioned in the output). Let's break down each section of the output:

### **\*\*Regression Statistics:\*\***

- **\*\*Multiple R:\*\*** This is the correlation coefficient, which measures 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, which represents the proportion of the variance in the response variable that is explained by the predictor variable. A value close to zero (1.59144E-05) indicates that the predictor variable explains very little of the variance in the response variable.



Adjusted R Square: This is a modified version of R Square that 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 is not adding any 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 represent the number of independent pieces of information available in the data.

SS: Sum of squares measures the total variation in the response variable.

MS: Mean square is the average variation within groups or between groups.

F: The F-statistic tests the overall significance of the regression model. A low F-value relative to the critical value indicates that the model is not significant.

Significance F: This is the p-value associated with the F-statistic. A high p-value (0.85852604) suggests that the model is not statistically significant.

## Coefficients:

Intercept: This is the y-intercept of the regression line, representing the predicted value of the response variable when the predictor variable is zero. The coefficient (232.5093842) indicates the average value of the response variable when the predictor variable is zero.

Postal Code: This is the coefficient for the predictor variable. It represents the change in the response variable for a one-unit change in the predictor variable. The coefficient (-6.54575E-05) suggests a very small negative effect of the Postal Code on the response variable, but it is not statistically significant given the high p-value (0.85852604).

Overall, based on this regression output, the model does not appear to provide a meaningful explanation of the variation in the response variable, and the predictor variable (Postal Code) does not appear to have a significant effect on the response variable.

## Correlation:

The absolute value of the correlation coefficient (0.024067424) is close to zero. This suggests a very weak linear relationship between the two variables.

## Descriptive Statistics:

---

<i>Sales</i>	
<hr/>	
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

---

## Conclusion:

### Conclusion:

In delving into the sales data across various segments in different US states, our analysis has unearthed valuable insights. The dataset provided a comprehensive view, encompassing crucial variables such as geographical location, product categorization, transactional details, and performance metrics. Through meticulous scrutiny, we addressed pertinent questions and gleaned actionable conclusions.

California emerged as a focal point, exhibiting the highest sales volume among the states analyzed. Notably, the consumer segment showcased consistent performance across all states, underscoring its significance in the market landscape.

Across categories, office supplies emerged as the top performer in all states, indicating a universal demand for these products. Furthermore, the consumer segment demonstrated dominance in terms of sales across the US, including California, Texas, and Washington.

Analysis of total versus average sales per segment revealed significant disparities, with total sales outweighing average sales across the board. This highlights the presence of outlier transactions or high-value sales within each segment.

Delving deeper into category and sub-category analysis, technology emerged as the category with the highest average sales, suggesting a strong market demand for technological products. Subsequently, copiers emerged as the top contributor to average sales distribution, emphasizing their importance within the technology category.

## Reviews:

1. Thorough Analysis with Actionable Insights: The exploration of sales data across US states provides a comprehensive understanding of market dynamics. The inclusion of key attributes and performance metrics ensures 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 structured, facilitating easy comprehension of complex data. The use of slicers enhances the 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.