Order Dataset Analysis

Introduction:

Our dataset comprises a plethora of variables, each offering unique insights into them 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**.**

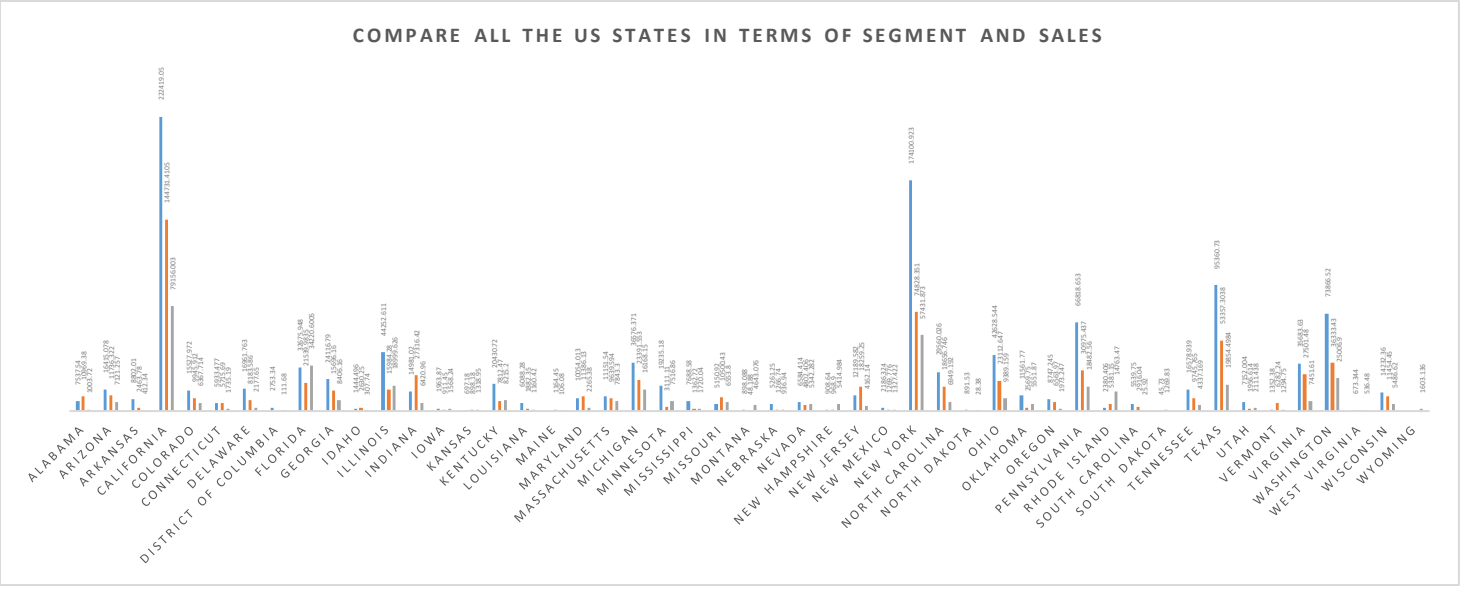
Questionnaire:

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.

Analytics:

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

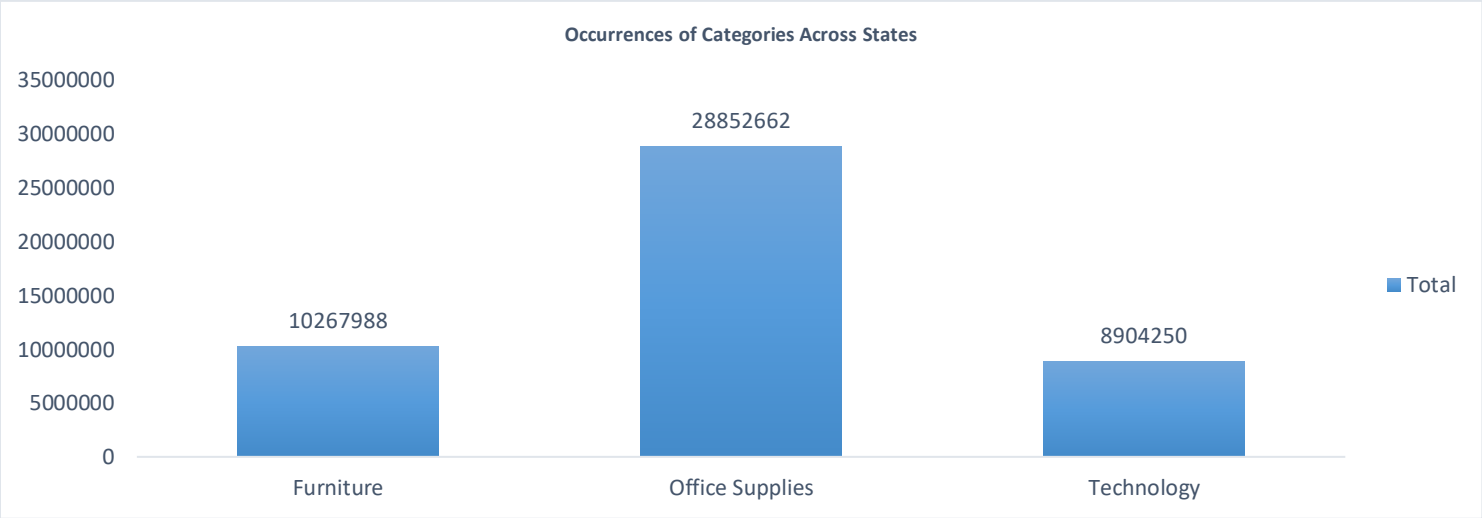
Ans



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?

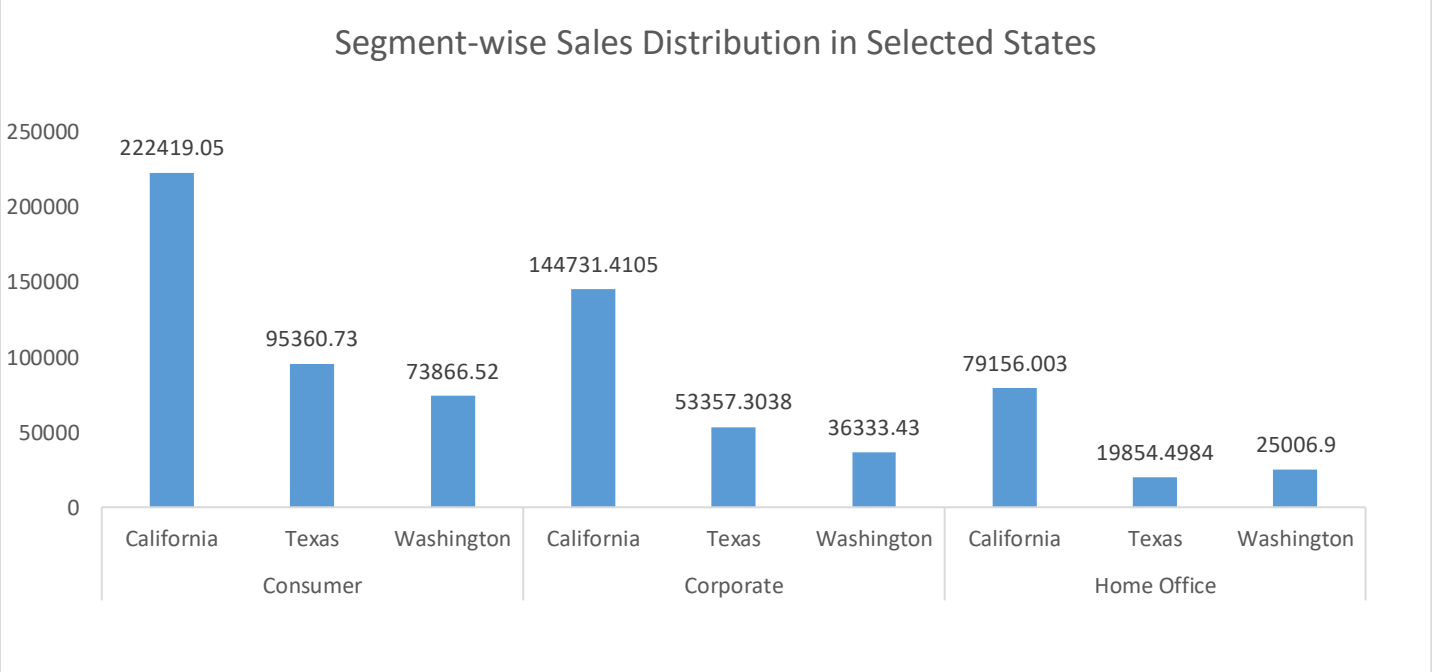
Ans



Office is the top performing category.

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

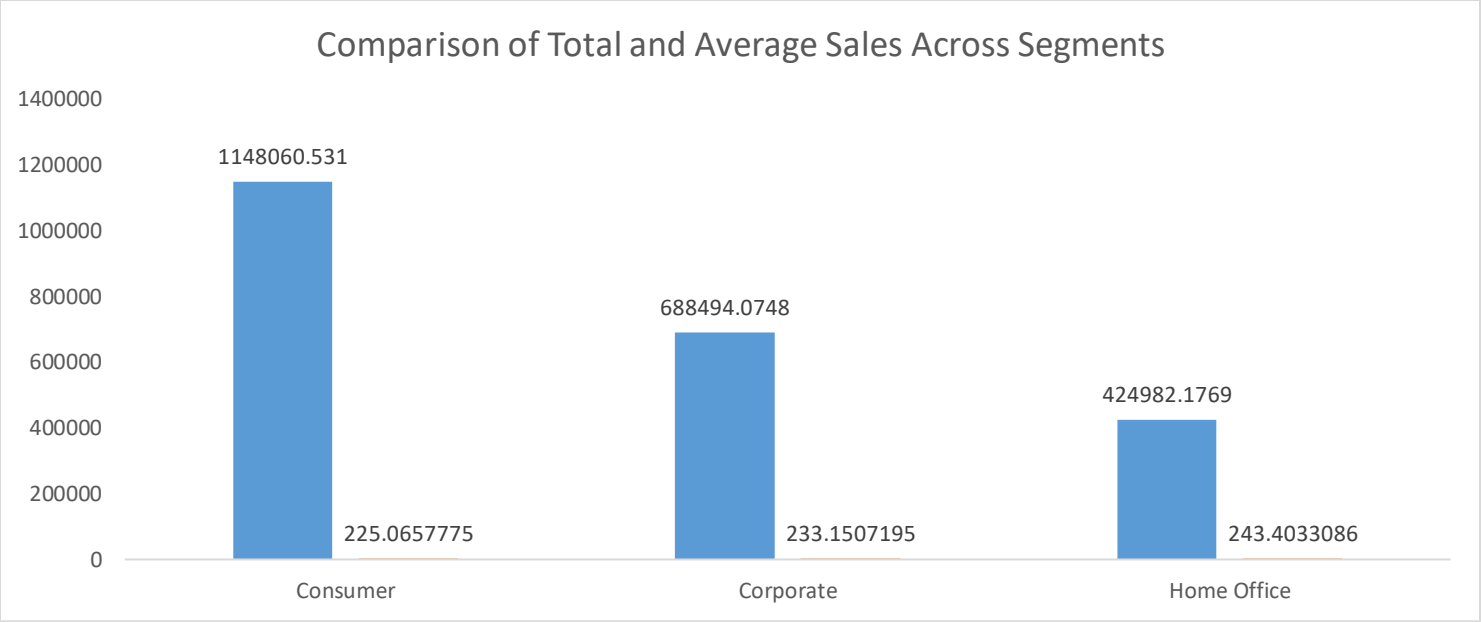
Ans



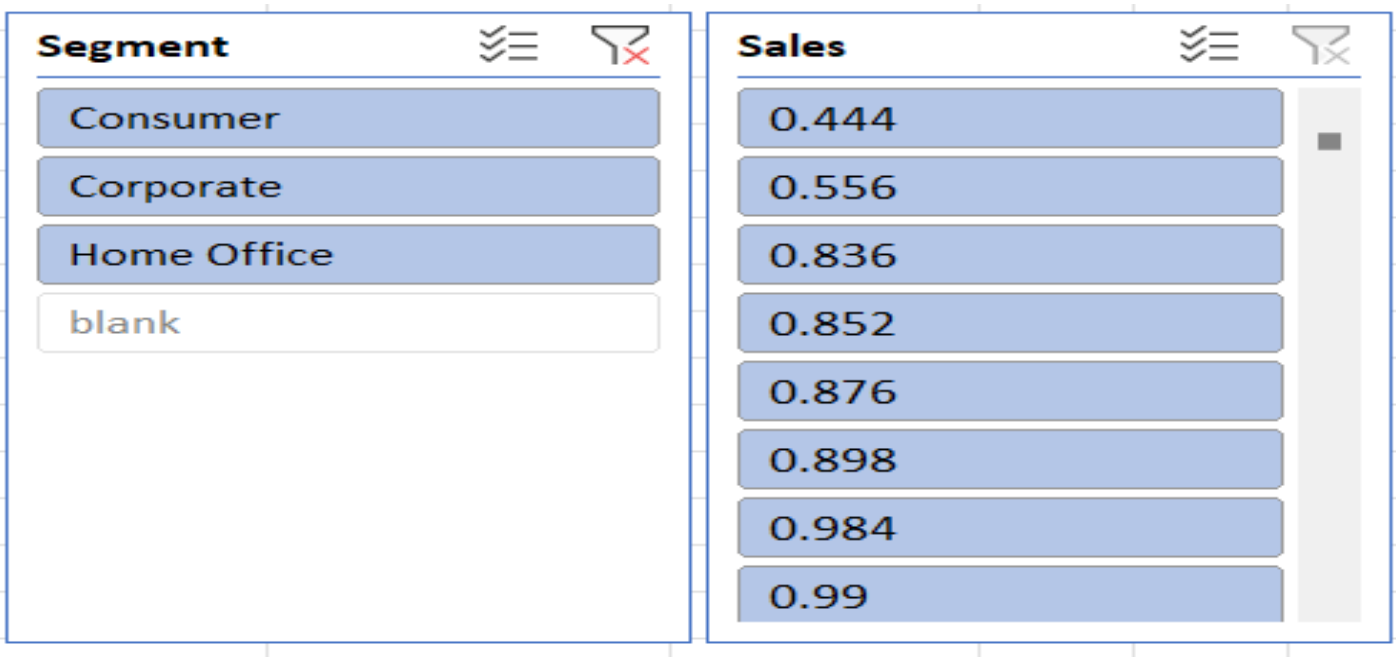
Consumer segment has the most sales in US, California, Texas, and Washington

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

Ans

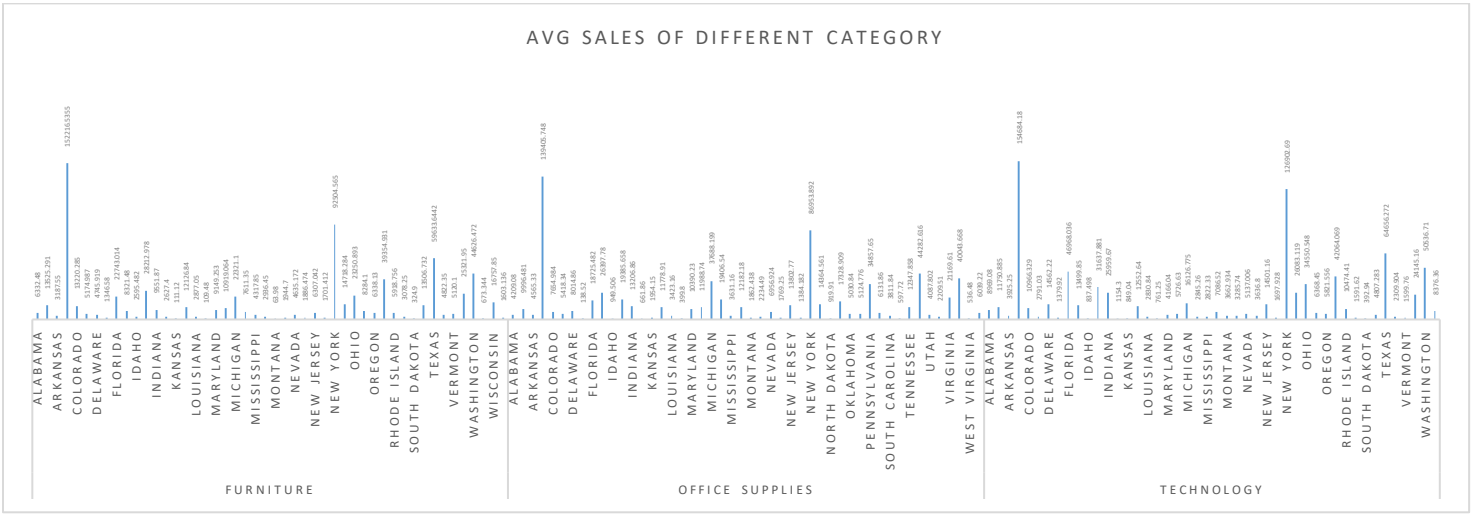


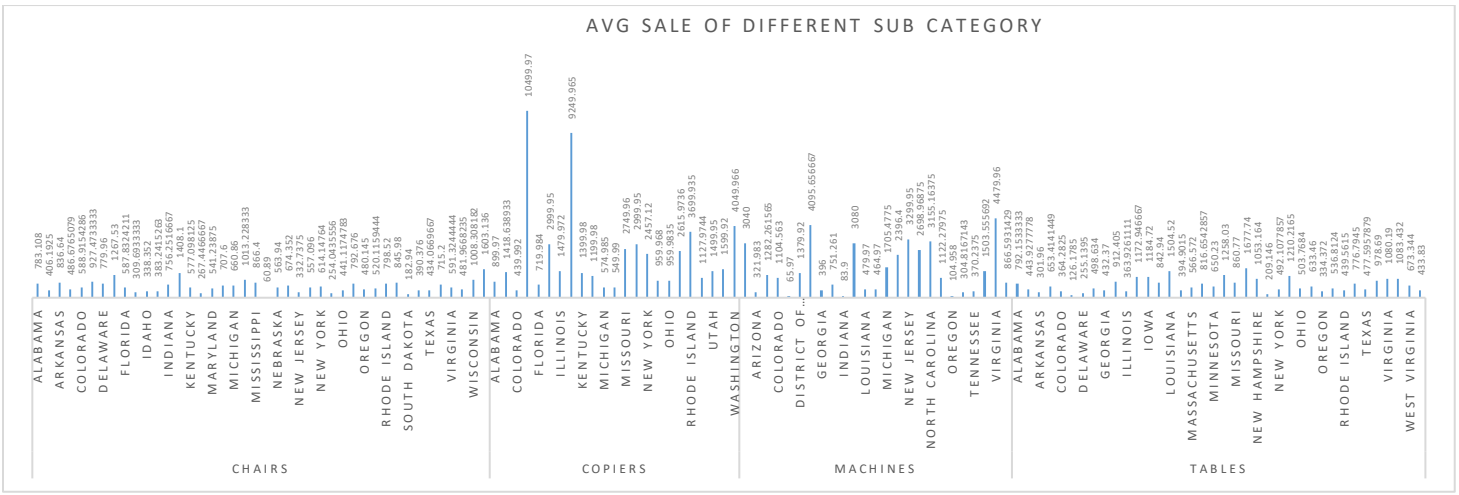
Overall, the home office segment has the highest average sales, followed by the corporate segment and then the consumer segment. However, in terms of total sales, the consumer segment has the highest, followed by the corporate segment and then the home office segment



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

Ans





The sales data for different categories within the furniture, office supplies, and technology segments paint a varied picture. On average, furniture items sell for $350.65, with bookcases leading the category at $503.60, followed by chairs at $531.83, tables at $645.89, and furnishings at $95.82. In contrast, office supplies have an average sales figure of $119.38, with appliances leading the category at $227.93, followed by storage at $263.63, supplies at $252.28, and binders at $134.07. Technology items have the highest average sales at $456.40, with copiers topping the list at $2215.88, followed by machines at $1645.55, accessories at $217.18, and phones at $374.18. These figures illustrate the diverse consumer preferences and spending patterns across these product categories.

Conclusion:-

Our comprehensive analysis of the provided dataset through various data visualization techniques has yielded valuable insights. Through the creation of bar graphs, pie charts, and other visual representations, we've been able to discern patterns, trends, and relationships within the data that might have otherwise remained obscured.

Regression:

The regression analysis reveals a moderately strong relationship between the independent variable (cost) andthe dependent variable, with a coefficient of determination (R-squared) of 0.503. The coefficient for the costvariable is highly significant, with a t-statistic of 99.63, indicating that changes in cost significantly affect thedependent variable. However, the intercept's coefficient is not statistically significant, suggesting that its impact on the dependent variable may not be meaningful.

SUMMARY OUTPUT				
Regression Statistics				
Multiple R	0.008850713			
R Square	7.83351E-05			
Adjusted R Square	-0.000924595			
Standard Error	596.4161586			
Observations	999			
ANOVA				
	Df	*SS*	*MS*	*F*
Regression	1	27783.3433	27783.3433	0.078106235
Residual	997	354645097.6	355712.2343	
Total	998	354672880.9		
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	232.3779806	37.2042048	6.246013907	6.22491E-10
Postal Code	0.000167458	0.000599189	0.279474927	0.779938343

Correlation

The correlation matrix indicates a strong positive correlation of 0.71 between sales and cost, suggesting that as the cost increases, sales tend to increase as well. This correlation coefficient reflects a moderately strong linear relationship between the two variables. Both sales and cost exhibit mutual influence on each other

	cost
1	0.709412
0.709412	1

Descriptive Statistics:

The data on sales reveals a wide variation, with a mean value of $230.77 and a significant standard deviation of $626.65, indicating a diverse range of sales figures. The skewness of 12.98 suggests a pronounced asymmetry in the distribution, potentially indicating outliers or skewed data points. With a maximum sales value of $22,638.48 and a minimum of $0.44, the range illustrates the considerable spreadin sales amounts within the dataset.

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