Report on Order Data Analysis

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: Performancemetrics
6..Rating: different product performing well in different state
7.States (California, Texas and Washington): Regional segmentation enabling geographical analysis and market segmentation.

. Questionnaires:

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?

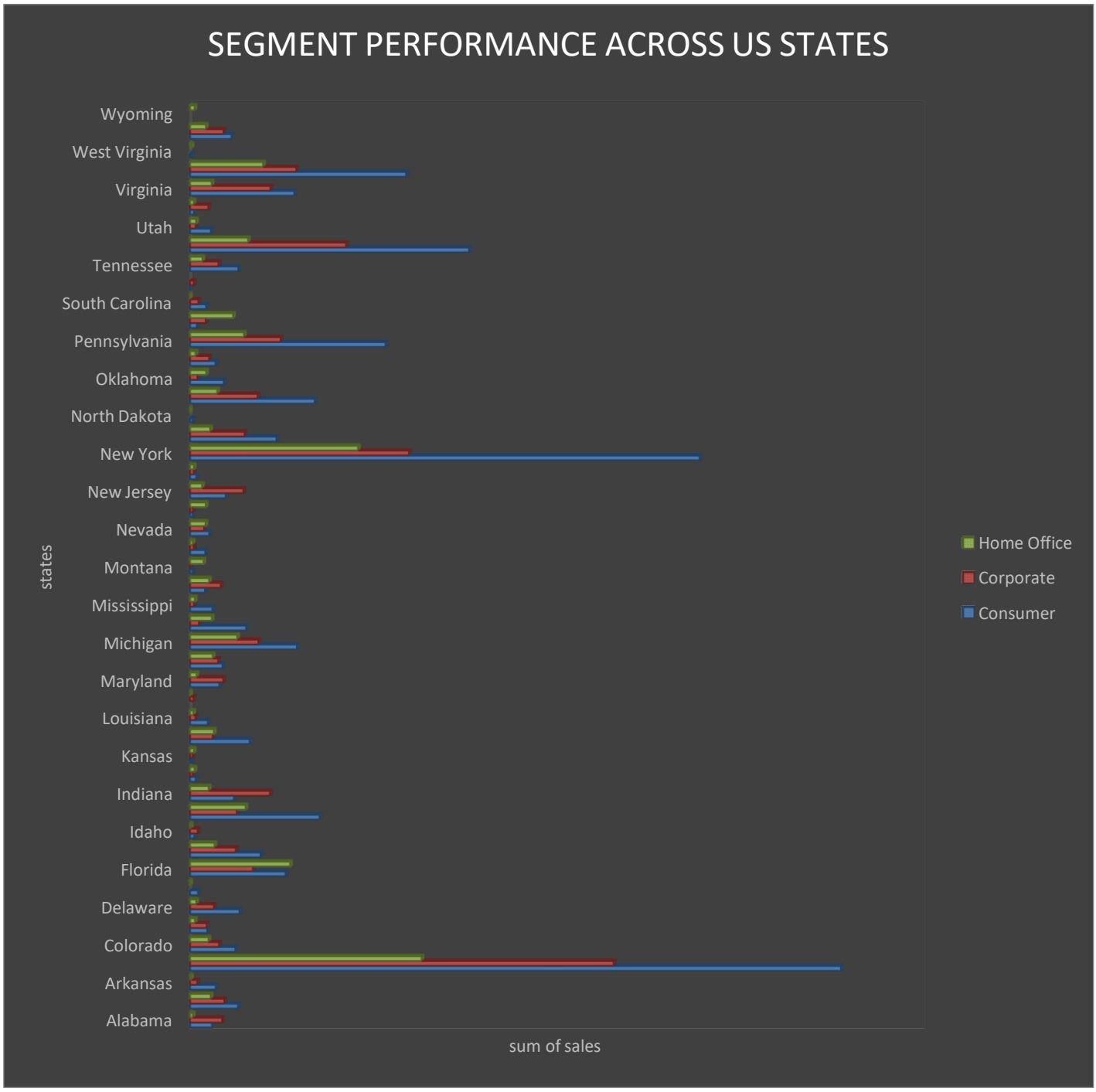
3. Which segment has most sales in US, California, Texas, and Washington?

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

5. Compare average sales of different category and sub category of all the state

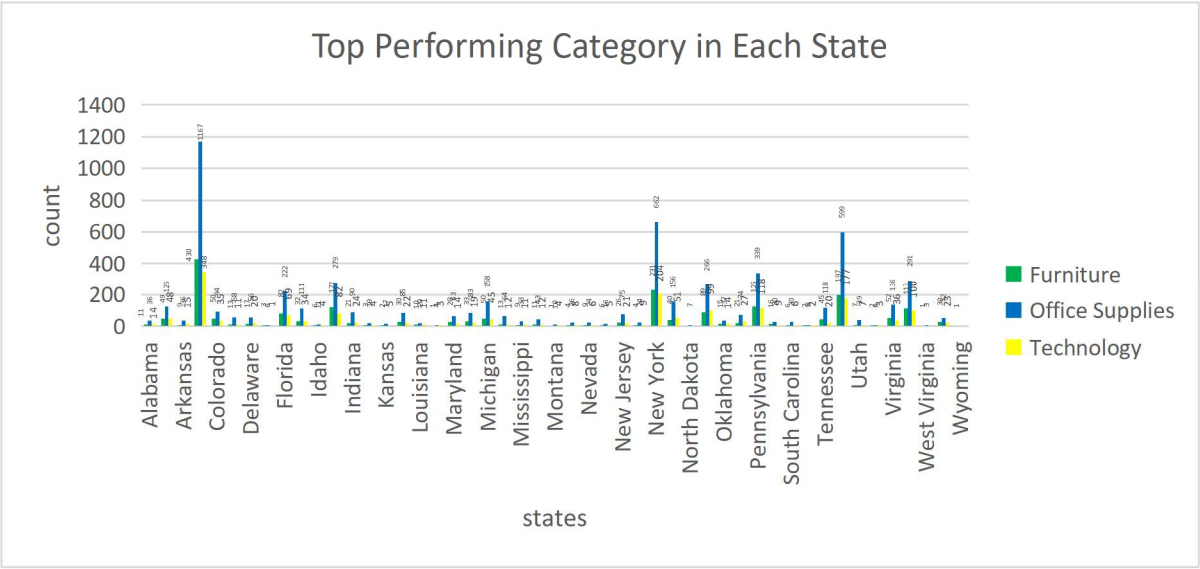
Analytics:

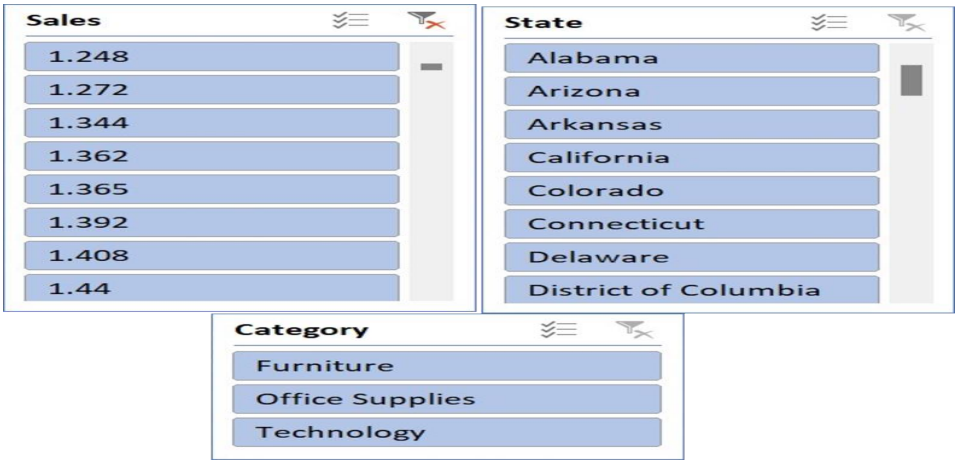
Q1. Compare all the US states in terms of Segment and Sales. Which Segment performedwell in all the states?



Ans: After comparing all the states in terms of segment and sales, California emerged asthe state with the highest amount of sales. Consumer segment performed well in all the states

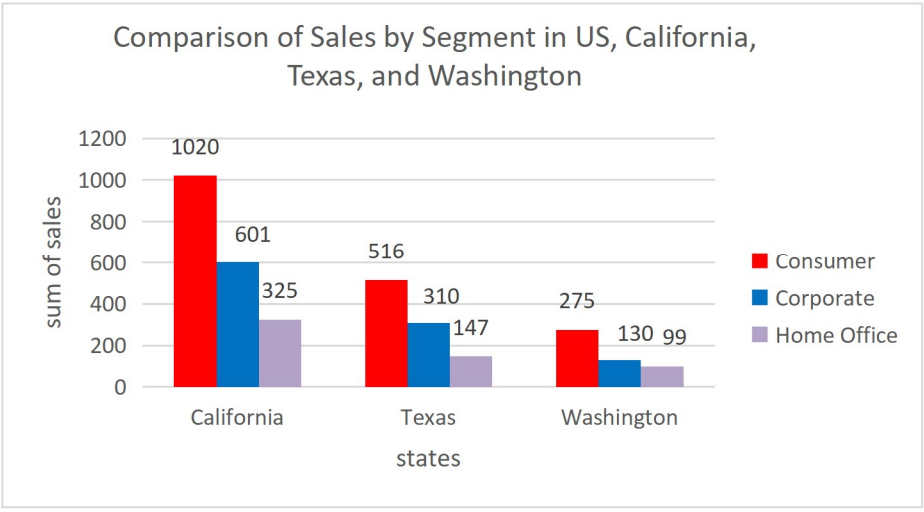
Q2. Find out top performing category in all the states





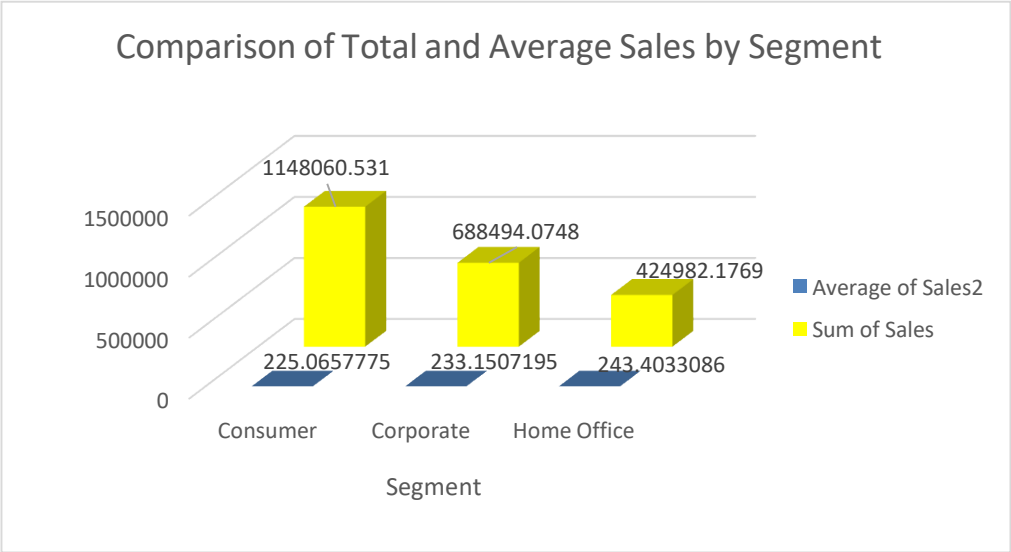
Ans. Office Supplies is the top performing category in all the states as it clearly shows from the given graph

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



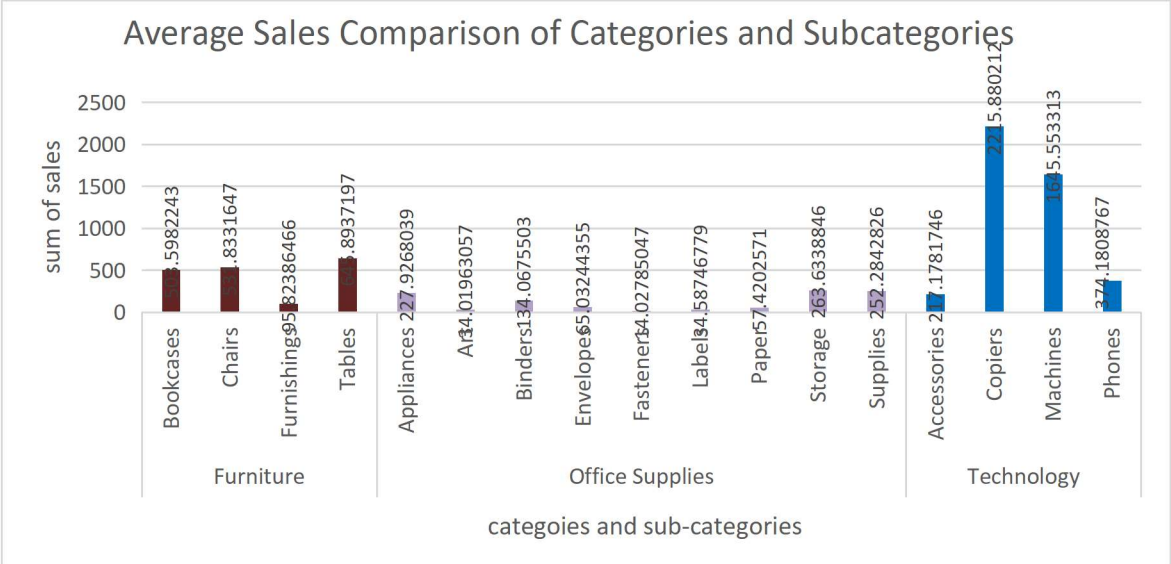
Ans. Consumer segment has the most sales in US, California, Texas, and Washington.

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



Ans. By Analysis of the given data set we can found that in all the three segments the total sales were greaterthan the average sales

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



Ans. By doing analysis of the given Order Sales dataset we were able to observe that, average sales of Technology was far greater than rest of the categories

Conclusion and Review:

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.

Our deep dive into the dataset has not only enhanced our understanding of the underlying information but has also empowered us to make informed decisions based on the insights gained. By visually depicting the data, we've been able to communicate complex findings in a clear and accessible manner, facilitating better comprehension and actionable strategies.

Furthermore, this process has underscored the importance of data visualization as a powerful tool for extracting meaningful information from raw data. By harnessing the visual nature of graphs and charts, we've transformed numbers and statistics into compelling narratives that drive understanding and inform decision-making.

Regression:

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

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

Correlation

	Sales	*cost*
Sales	1	0.709412
cost	0.709412	1

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

Anova (single factor) :

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Sales	9800	2261537	230.7691	392692.6
cost	9800	2261411	230.7562	197630.9

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.807262	1	0.807262	2.73E-06	0.99868	3.841933
Within Groups	5.78E+09	19598	295161.7			
Total	5.78E+09	19599				

The ANOVA analysis compares the variability between two groups, sales and cost, revealing a minimal difference between them with a small sum of squares (SS) of 0.81. The F-statistic of 2.735 and p-value of0.999 suggest that this difference is not statistically significant, indicating that the means of sales and cost are likely equal. The within-groups variation is considerably higher, suggesting that most of the variabilitylies within each group rather than between the

Anova (two factor) without Replication :-

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.94E+09	9799	197630.9	65535	#NUM!	#NUM!
Columns	0	0	65535	65535	#NUM!	#NUM!
Error	0	0	65535			
Total	1.94E+09	9799				

The ANOVA table illustrates significant variation attributed to rows, represented by a sum of squares (SS) of 1,936,585,107 and 9,799 degrees of freedom (df), resulting in a mean square (MS) of 197,630.89. The F-statistic is notably high at 65535, indicating a substantial influence of row factors on the observed variance. However, the p-value is reported as #NUM!, suggesting a potential issue with the calculation or data. Similarly, for columns, no variation is observed, with an SS and MS of 0, and the F-statistic equals 65535

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 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**.**