Order Data Analysis

Introduction:

Our dataset includes a wide range of details, each giving us valuable insights into the different types of sales. We've got basic info like when and where sales happened, as well as more specific stuff like the type of customer and what they bought. Here are some key things we're looking at:

1. ID: Each sale has its own ID number, which helps us keep track of everything and analyze it better.

2. City, State: Tells us where the sales took place, so we can compare different regions and spot trends.

3. Product Line: This categorizes our products into groups like furniture, electronics, and home goods, so we can see which types of items are selling the most.

4. Unit Price, Net Sales: Basic details about how much stuff costs and how much money we're making.

5. Net Sales by Category, Categories Doing Well in Different States: Helps us see which product categories are bringing in the most money and which ones are popular in different areas.

6. Rating: Shows how well certain products are rated in different states, which can affect sales.

7. States: We're focusing on California, Texas, and Washington, so we can see how sales differ in different parts of the country.

All of this info together gives us a clearer picture of what's going on with our sales and helps us make better decisions.

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 states.

ANALYTICS:

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

Comparsion of segments of US states on the basis of sales
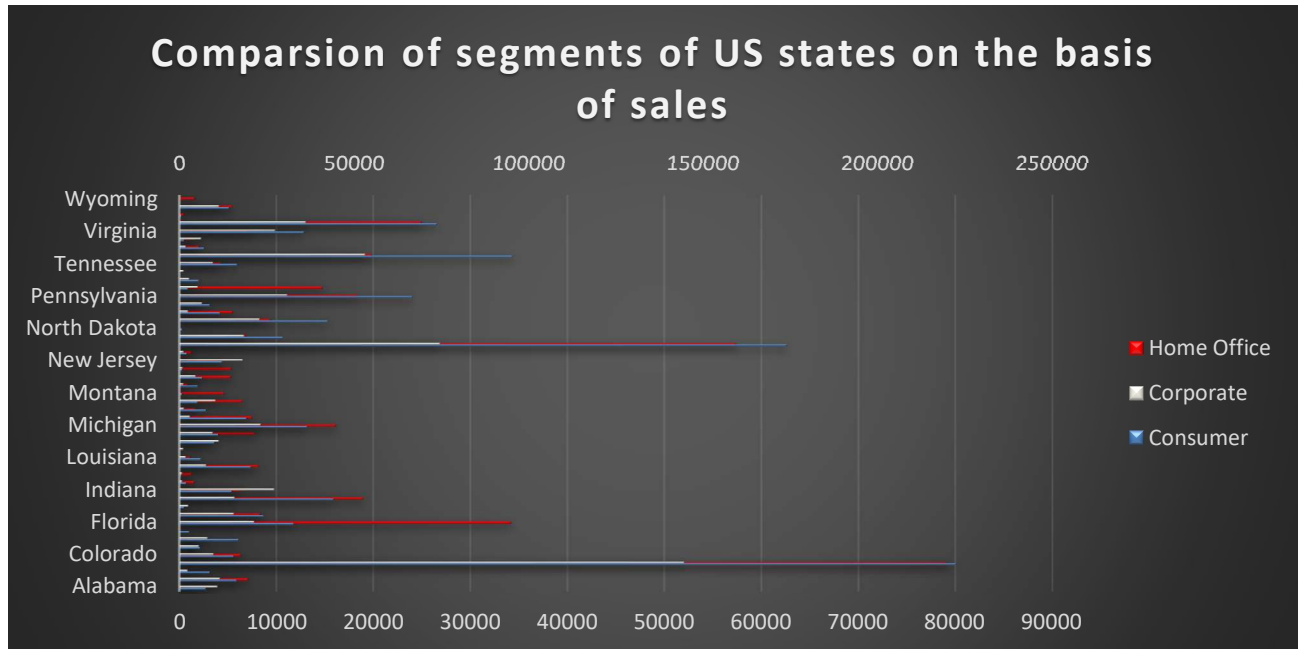
Ans: After comparing all the states in terms of segment and sales, California emerged as the state with the highest amount of sales consumer segment performed well in all the states
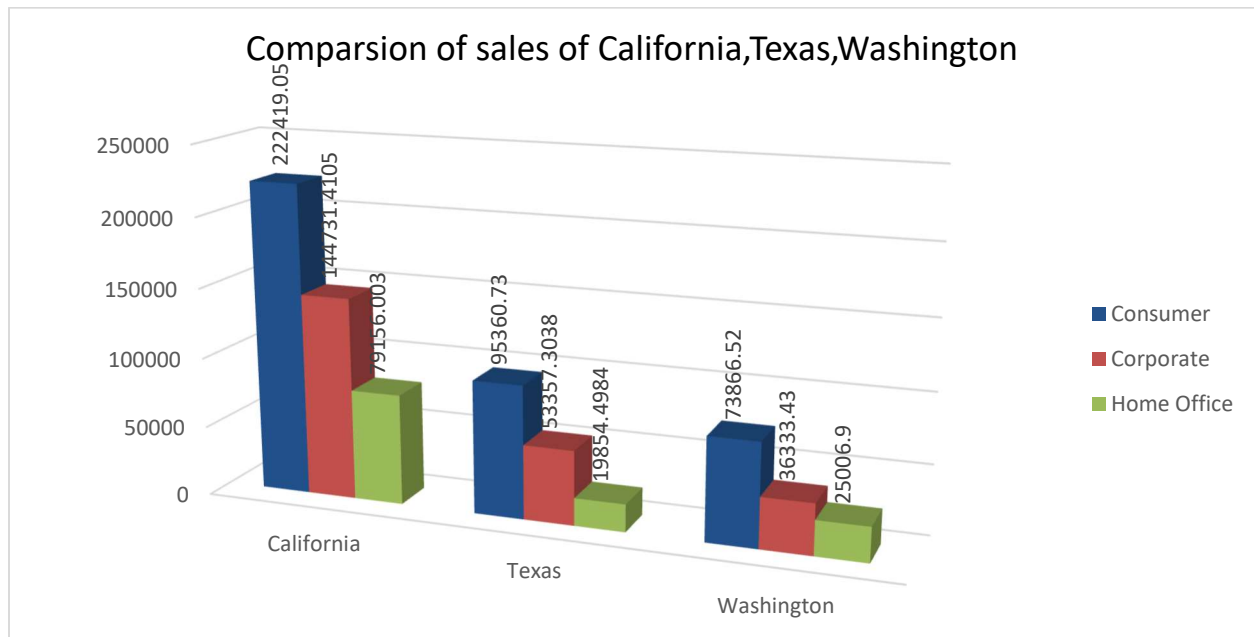
Q2. Find out top performing category in all the states.
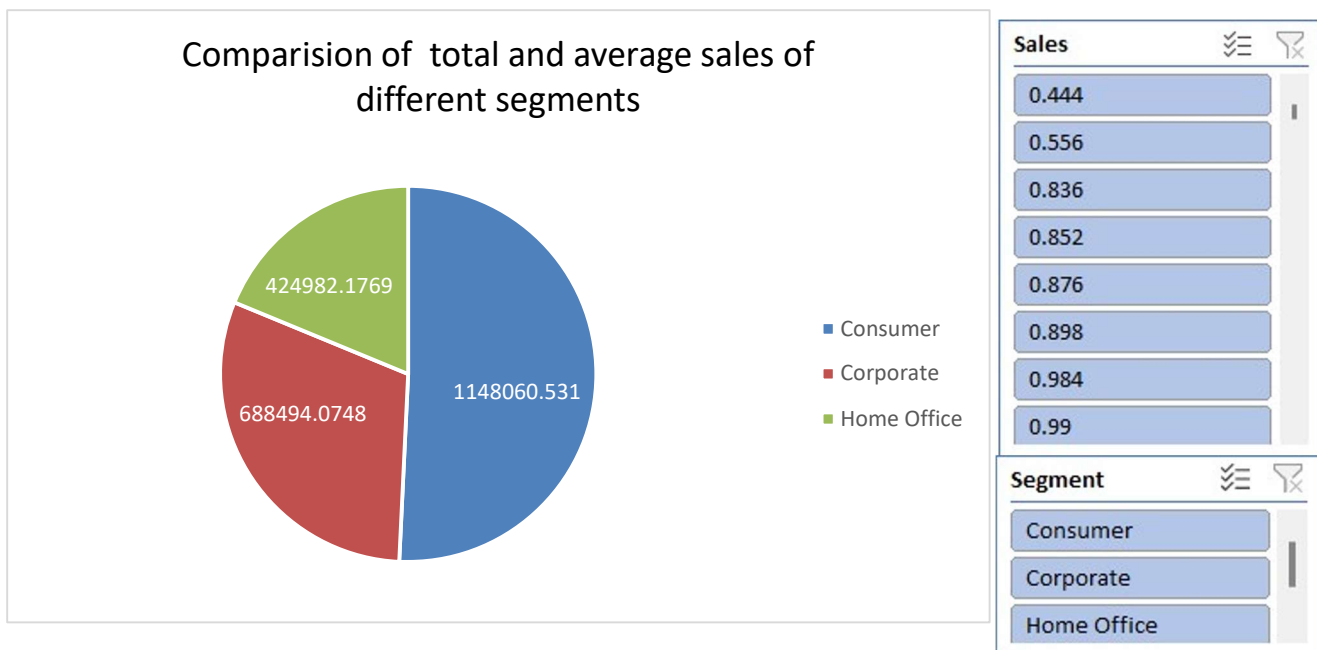 Ans. Office Supplies is the top performing category in all the states as it clearly shows from the given grap

Comparsion of performance of different states
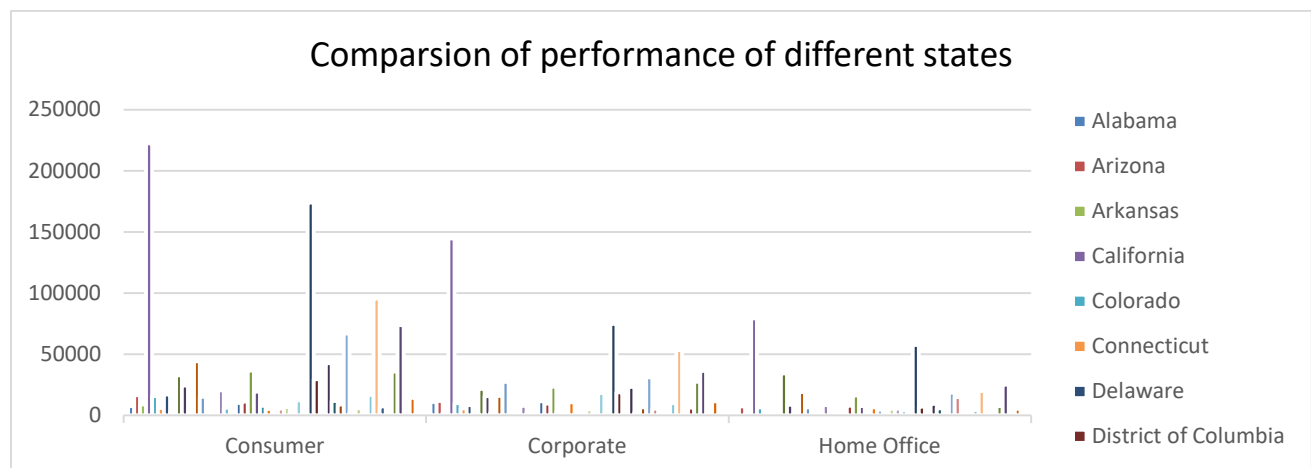
Q3. Which segment has most sales in US, California, Texas, and Washington?

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

Comparsion of sales of California,Texas,Washington

California: Consumer 222419.05, Corporate 144731.4105, Home Office 79156.003

Texas: Consumer 95360.73, Corporate 53357.3038, Home Office 19854.4984

Washington: Consumer 73866.52, Corporate 36333.43, Home Office 25006.9

Legend: Consumer, Corporate, Home Office

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

Comparision of total and average sales of different segments

Consumer: 1148060.531
Corporate: 688494.0748
Home Office: 424982.1769

Legend: Consumer, Corporate, Home Office

Sales
0.444
0.556
0.836
0.852
0.876
0.898
0.984
0.99

Segment
Consumer
Corporate
Home Office

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

Q5. Compare average sales of different category and sub category of all the states.
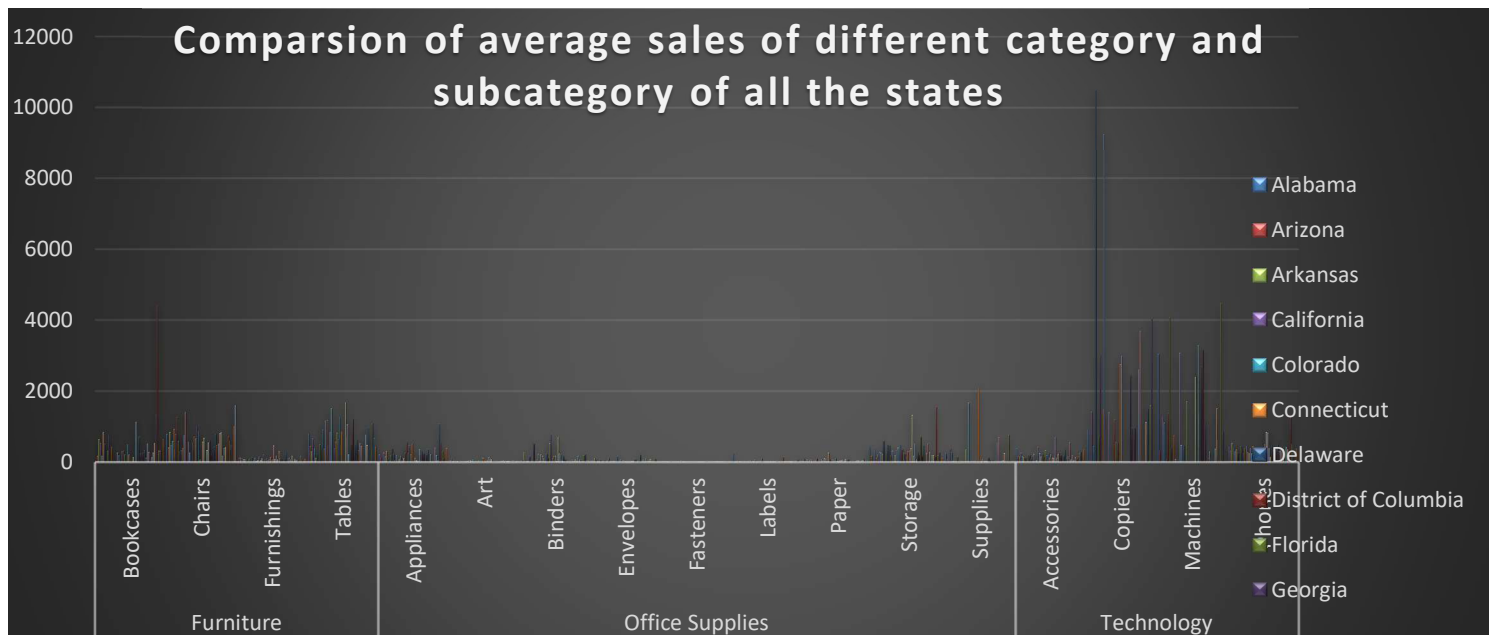
Ans: By doing analysis of the given Order Sales dataset we were able to observe that, average sales of Technology was far greater than rest of the categories.

Conclusion and Review:

After delving deep into the dataset and employing various data visualization techniques, we've unearthed a wealth of valuable insights. Through the creation of visually engaging representations such as bar graphs, pie charts, and other graphical elements, we've successfully teased out intricate patterns, discerned trends, and unearthed subtle relationships within the data that might have otherwise remained hidden.

This thorough exploration of the dataset has not only enriched our comprehension of the underlying information but has also equipped us with the knowledge needed to make well-informed decisions based on the insights gained. By harnessing the power of visual data representation, we've been able to present complex findings in a clear and accessible manner, facilitating better understanding and the formulation of actionable strategies.

Moreover, this process has underscored the pivotal role of data visualization as a potent tool for extracting meaningful information from raw data. By leveraging the visual nature of graphs, charts, and diagrams, we've transformed mere numbers and statistics into compelling narratives that not only drive understanding but also serve as the cornerstone for informed decision-making.

Regression:

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.

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

Co-relation

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

	Sales	*cost*
Sales	1	0.709412
cost	0.709412	1

Anova (single factor) :

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 of 0.999 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 variability lies within each group rather than between them.

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				

Anova (two factor) without Replication :

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

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				

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 spread in 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