Shop Data sales Analytics

Introduction:Within this dataset, you'll discover a comprehensive breakdown of our shop's sales activities over a specified timeframe. Each entry encompasses crucial details such as the date of sale, the designated salesman involved, the specific item purchased, the corresponding company, the quantity acquired, and the total expenditure incurred. This compilation serves as a rich resource for dissecting patterns, discerning customer preferences,and gauging the effectiveness of sales strategies. Whether unravelling the performance of individual products or delving into overarching market trends, this data encapsulates the dynamic landscape of our business operations in a manner accessible to all stakeholder

Questionaire:-

1. Compare all the salesmen on the basis of profit earn.

2. Find out most sold product over the period of May-September.

3. Find out which of the two product sold the most over the year Computer or Laptop?

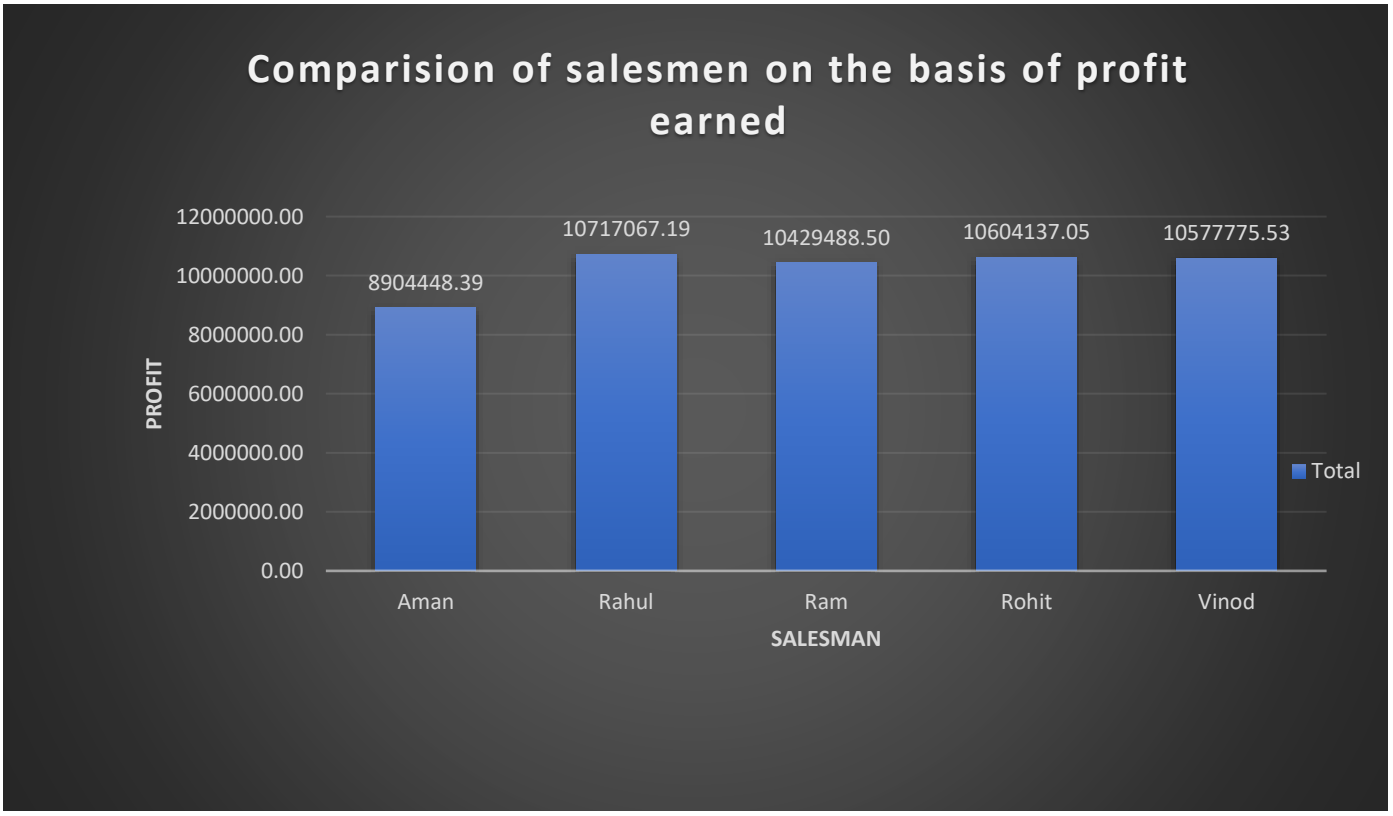
4. Which item yield most average profit?

5. Find out average sales of all the products and compare them.

Analytics:-

1 Compare all the salesmen on the basis of profit earn.

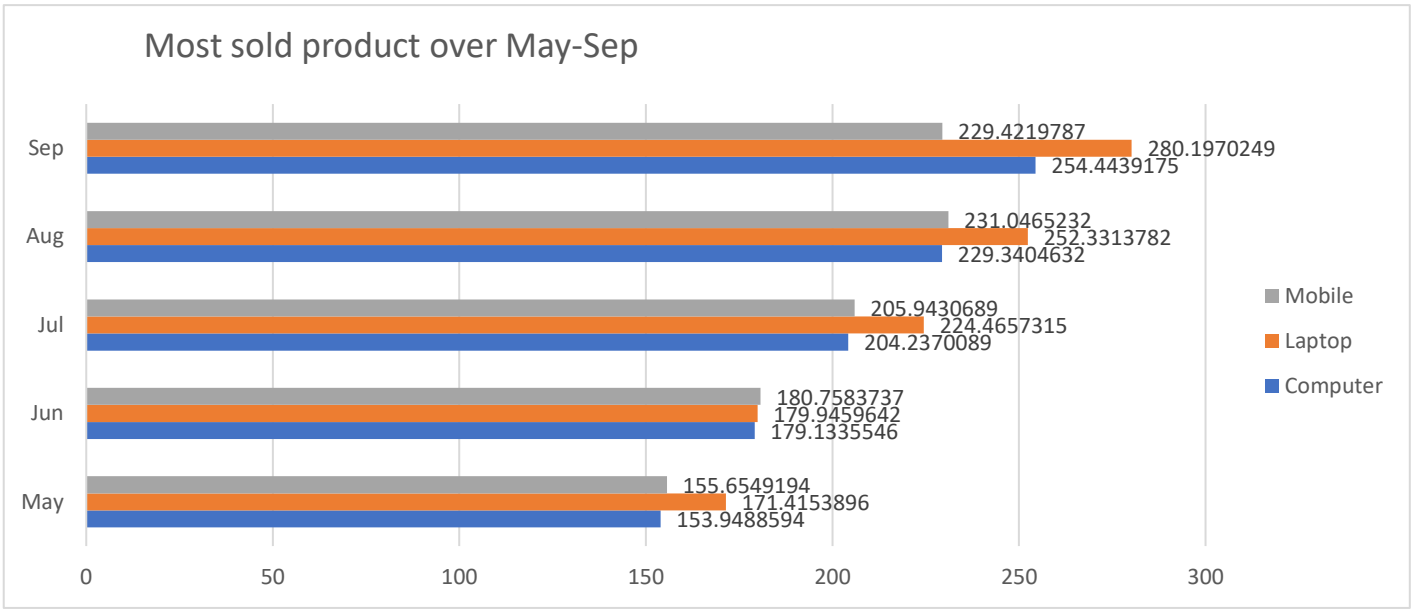
Ans



The maximum profit earned by Rahul of 10717067.

2. Find out most sold product over the period of May-September

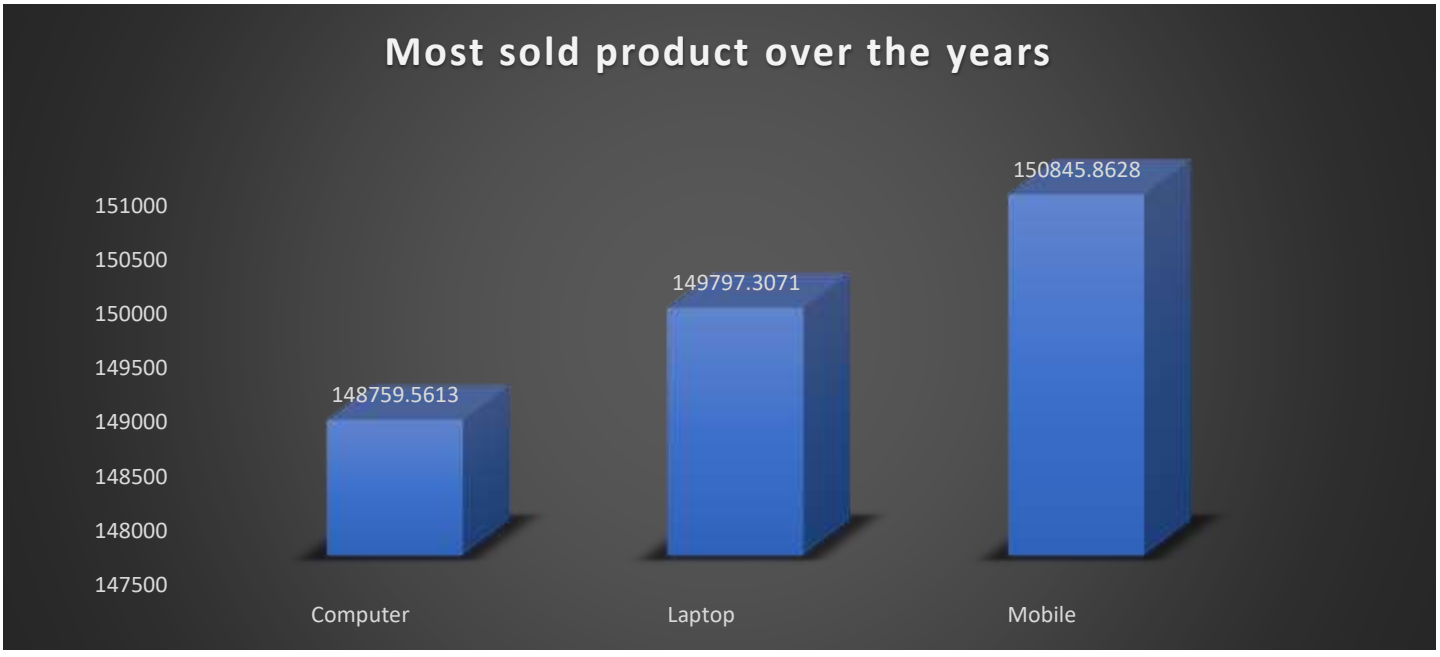
Ans:



During the May to September period, laptops emerged as our best-selling product, capturing the lion's share of customer interest. This trend reflects their enduring appeal and essential role in modern life, whether for work, education, or personal use. Our sales data meticulously captures this trend, providing valuable insights for inventory management and strategic planning

3.Find out which of the two product sold the most over the year Computer or Laptop?

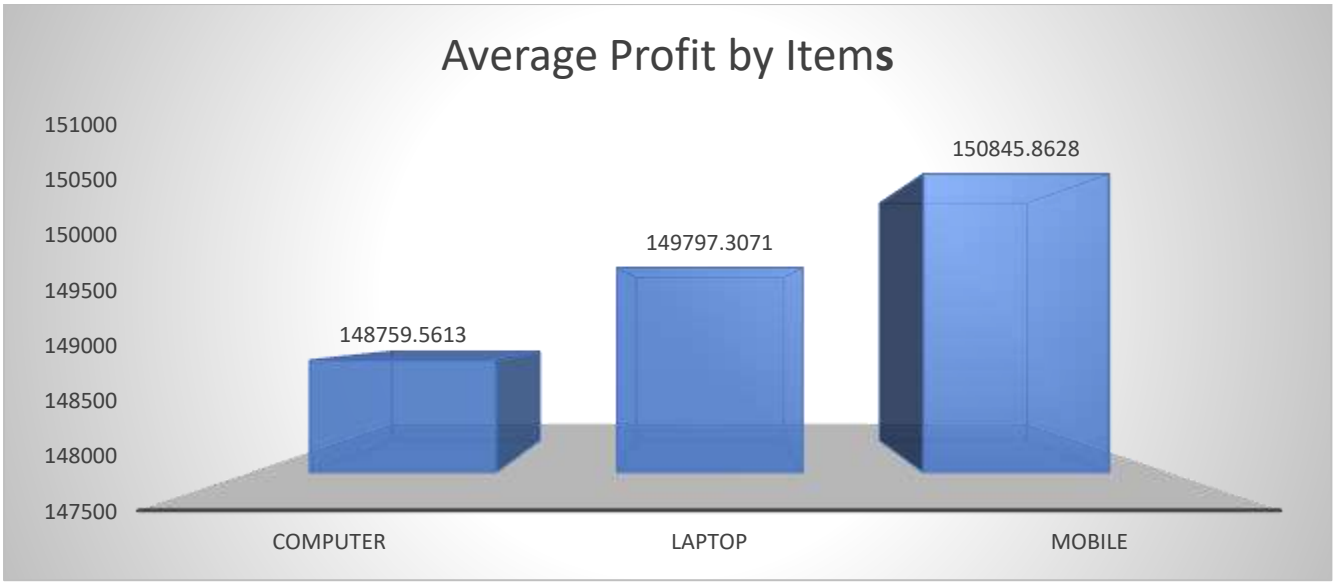
Ans:-



The two product sold the most over the year between computer or laptop

4 . Which item yield most average profit?

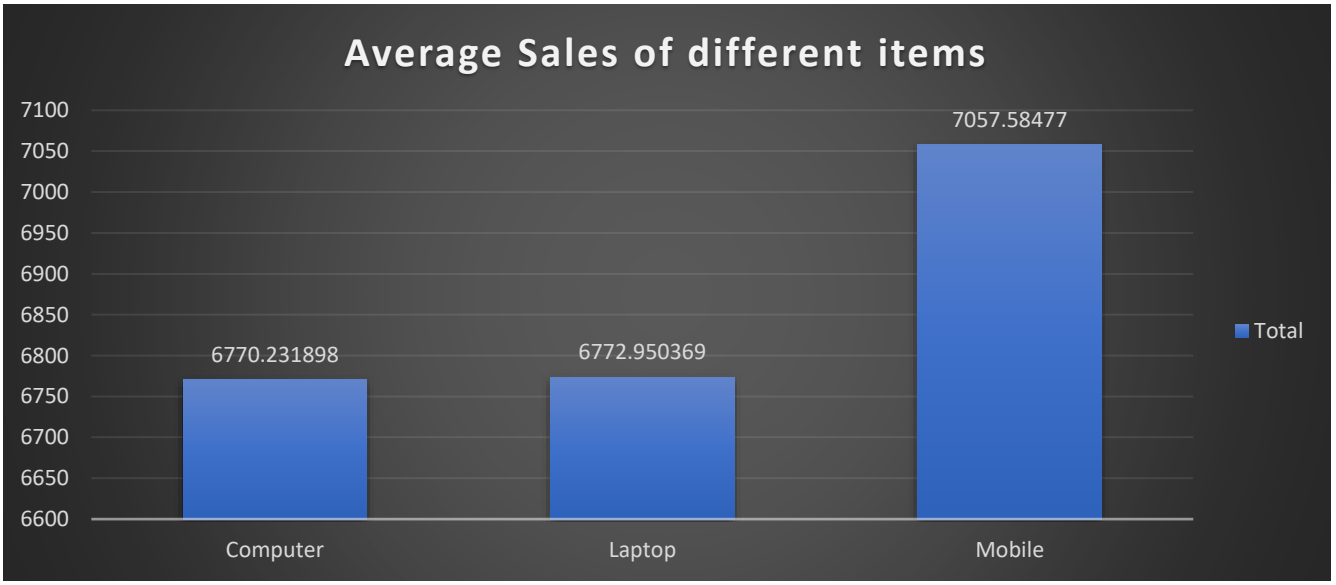
Ans:-



The item that yields the most profit between laptop, computer and mobile is Mobile.

5.Find out average sales of all the products and compare them.

Ans:



The avg sales of Mobile is 7057.

The avg salesof laptop is 6772.

The avg sale of Computer is 6770.

Conclusion and Review:

Our examination of the shop sales data sample has provided us with valuable insights into our sales performance, customer preferences, and overall business health. While the report effectively outlined the data examined and our objectives, enhancing it with more comprehensive analysis and visual representations could further clarify key findings.Nonetheless, the knowledge gained from this analysis will empower us to make informed decisions aimed at optimizing our sales processes and achieving our business objectives. It underscores the importance of ongoing analysis and refinement of our sales data

Regression:

The analysis indicates a significant correlation between sales and the variable under investigation, supported by an extremely low p-value, essentially zero. This implies that the observed relationship is highly likely genuine and not merely due to chance. The model accounts for approximately 30.41% of the variance, signifying a strong explanatory capability and a solid grasp of the factors influencing sales.Additionally, the standard error, indicating the

Correlation:

The correlation coefficient between Quantity and Amount 2 is 0.954, indicating a strong positive correlation between the two columns.

	Qty	*Amount*	*Profit*
Qty	1		
Amount	0.953708	1	
Profit	0.984563	0.951461	1

Anova (single Factor) :

The single-factor ANOVA analysis unveils significant variations among the groups, with a high F-value of 10261.03 and an ultra-low p-value close to zero, indicating a strong impact of the factor being analyzed. The degrees of freedom (df) for the between-groups factor are 3, representing the variability in means across the groups. Within the groups, the df is 11284, reflecting the variation within each group,and an error (standard error of the residuals) of approximately 848506.0368.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	342	6654.271	19.45693	66.0952
Column 2	342	2347644	6864.457	4410782

ANOVA

Multiple R	0.954077
R Square	0.910263
Adjusted R Square	0.909999
Standard Error	2.438983
Observations	342

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	20515.93	20515.93	3448.844	4.6E-180
Residual	340	2022.537	5.948639		
Total	341	22538.46			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0 %*	*Upper 95.0 %*
Intercept	-5.89533	0.451394	-13.0603	7.13E-32	-6.78321	-5.00746	-6.78321	-5.00746
X Variable 1	0.003693	6.29E-05	58.72686	4.6E-180	0.00357	0.003817	0.00357	0.003817

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8.01E+09	1	8.01E+09	3632.879	2.1E-275	3.85513
Within Groups	1.5E+09	682	2205424			
Total	9.52E+09	683				

Anova without replication:-

This ANOVA table presents the analysis of variance for a two-way ANOVA. The table is divided into three main sections: Rows, Columns, and Error.

1. Rows: This section analyzes the variation between the rows (levels) of one factor. It includes the Sum of Squares (SS), degrees of freedom (df), Mean Square (MS), F-value, andp-value. In this case, the F-value (1.003) is associated with a p-value of 0.495, indicating thatthere is no significant difference between the rows.

2. Columns: This section examines the variation between the columns (levels) of another.

Descriptive Statistics:-

The statistical summary provides insights into two datasets: one representing quantities and the other corresponding amounts. For the quantity data, typical values fall around 10 to 11 units, with a mode at 3 units. The data shows a moderate spread, with a standard deviation of around 3 to 4 units. The amount data ranges from approximately $1000 to $8888, with a mean around $4711.057 and a standard deviation of $1365.587. Both datasets exhibit slight negative skewness, indicating a slight imbalance towards lower values. These statistics offer a concise understanding of the central tendencies, variability, and distribution shapes of the datasets.

Qty		*Amount*	
Mean	10.72353	Mean	4711.057
Standard Error	0.322467	Standard Error	121.1763
Median	11.74126	Median	5031.563
Mode	3	Mode	1000
Standard Deviation	3.634024	Standard Deviation	1365.587
Sample Variance	13.20613	Sample Variance	1864828
Kurtosis	-0.62376	Kurtosis	1.333183
Skewness	-0.67898	Skewness	-0.21899
Range	12.84171	Range	7888
Minimum	3	Minimum	1000
Maximum	15.84171	Maximum	8888
Sum	1361.888	Sum	598304.2
Count	127	Count	127

Supermarket Sales Dataset Analysis

Introduction:

Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of supermarket sales. From fundamental transactional details such as Invoice ID, Date, Time, and Payment Method to more nuanced factors like Branch Location, Customer Type, Gender Demographics, Product Line, and Product Ratings, every facet has been meticulously documented.

Questionnaire:

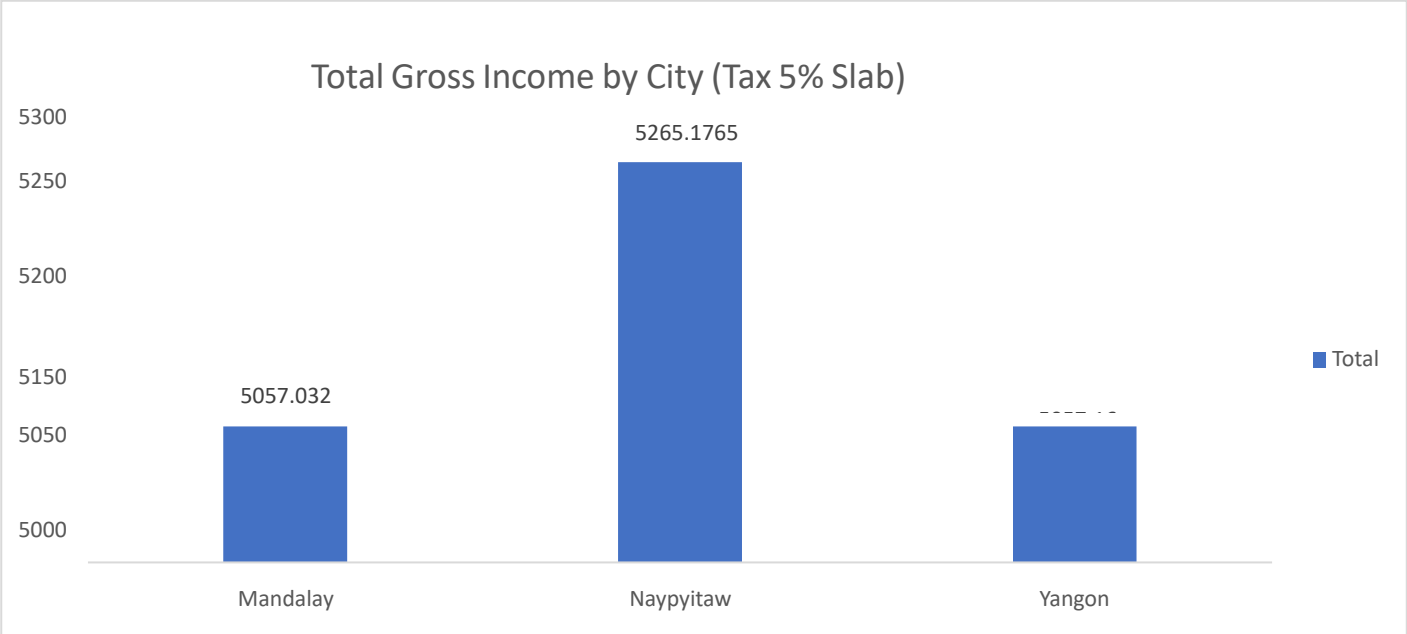
Q1. Which of the given cities having tax 5% slab performed better than all the others?
Q2. Which customer gender ordered most items from all the three branches?
Q3. Compare highest and lowest rating products on the basis of units sold.
Q4. Analyzing units sold and unit price data answer the following sub questions

a) What is the degree of freedom?
b) Co-relation of Unit price and revenue generated
c) What result you can draw from regression of the two data

Q5. What product will you suggest as per the city data analysis to each type of customer?

Analytics:

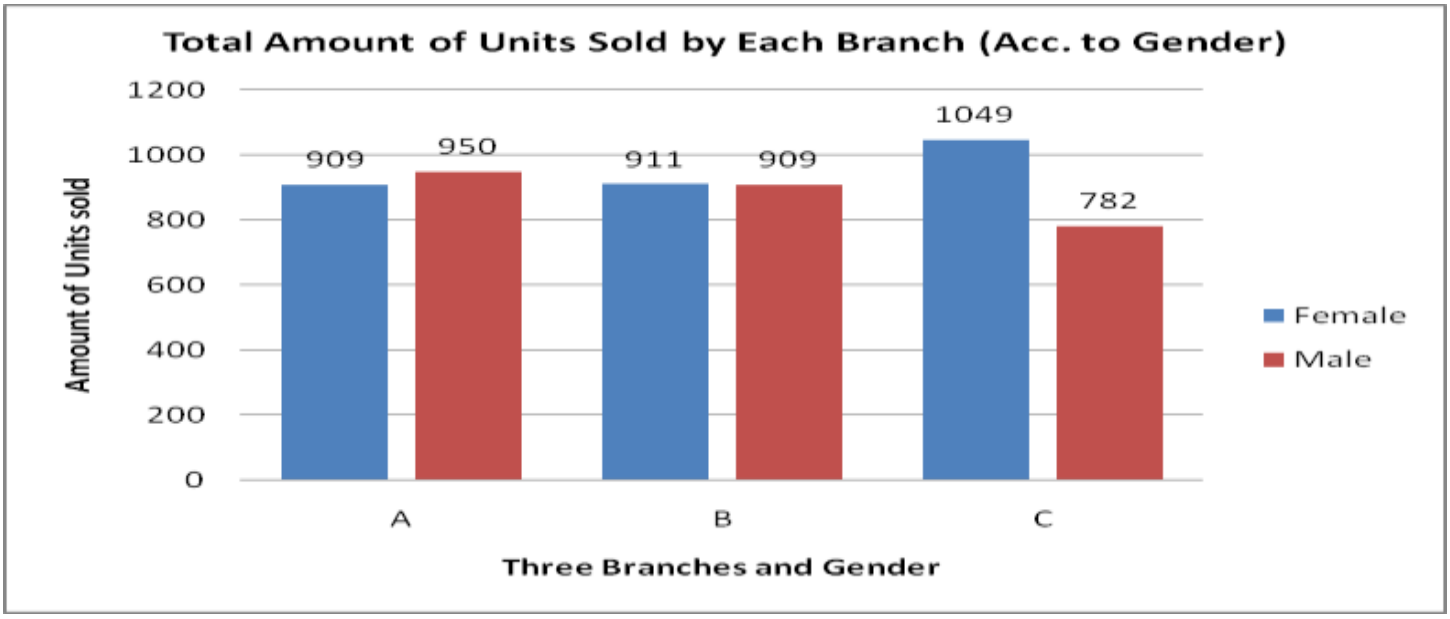
Q1. Which of the given cities having tax 5% slab performed better than all the others?
Ans



Based on the data analyzed, the city that outperformed all is **Mandalay**. This conclusion is drawn from superior performance in total sales/revenue generation compared to the other cities in the same tax slab of 5%.

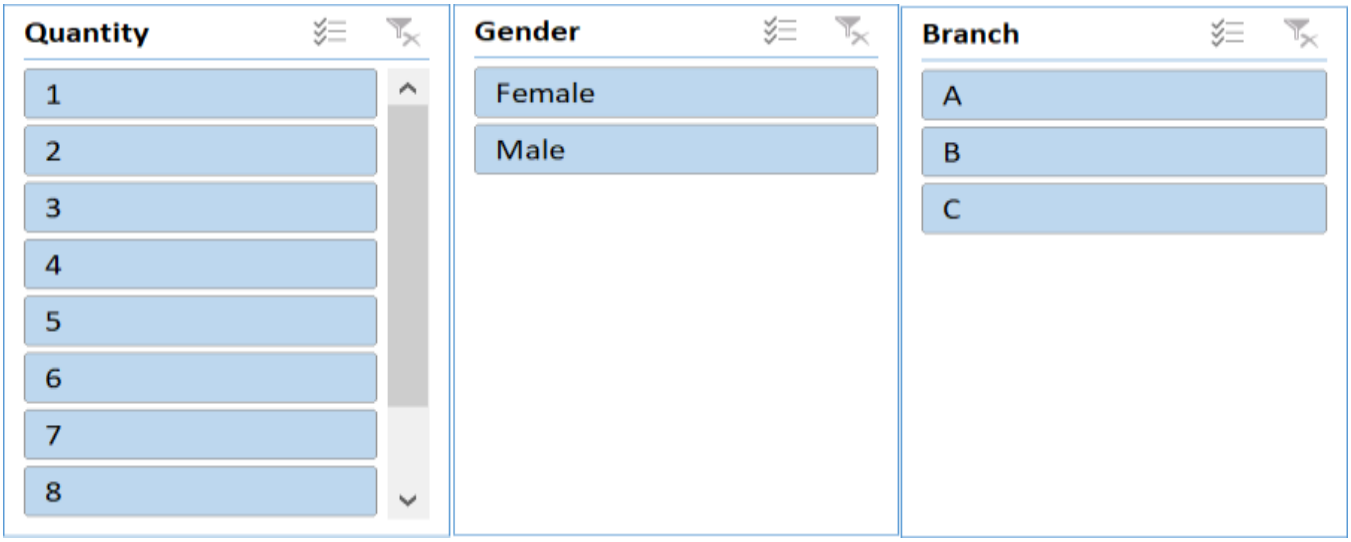
Q2. Which customer gender ordered most items from all the three branches?

Ans.



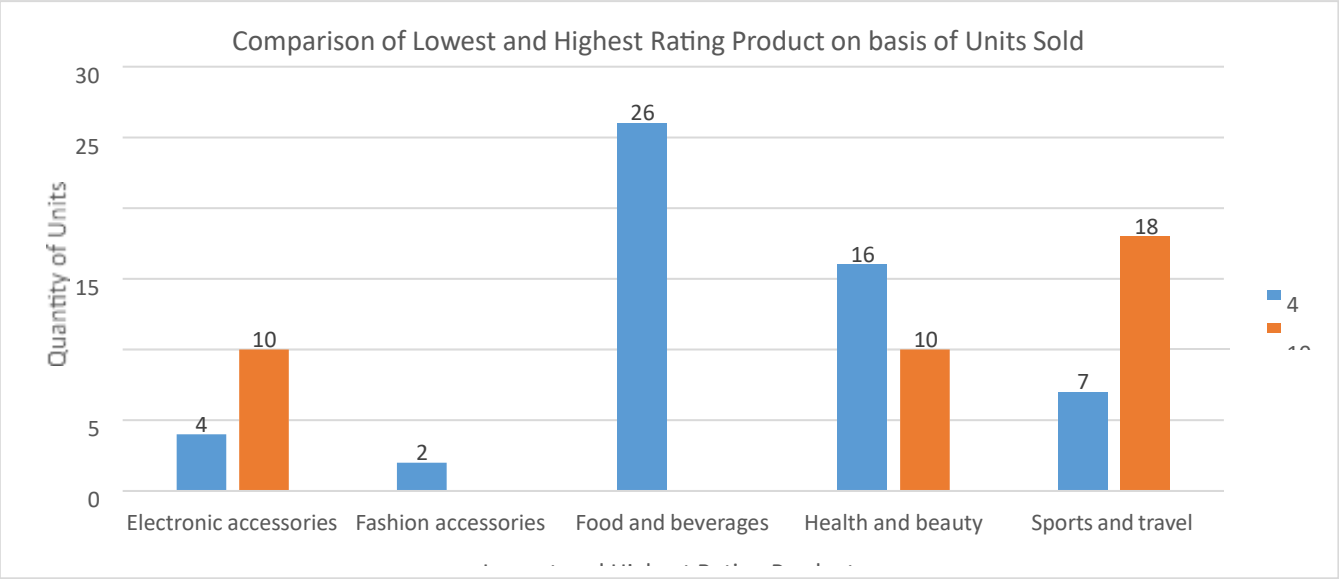
Our analysis of the Supermarket Sales Data revealed the following:

a. At Branch A, females placed the highest number of orders.
b. Branch B saw higer number of orders placed by Females
c. Meanwhile, at Branch C, males placed the most orders.



Q3. Compare highest and lowest rating products on the basis of units sold.

Ans.



Upon analyzing the Supermarket Sales Data, we discovered that product ratings ranged from a minimum of 4 to a maximum of 10.

a) Electronic Accessories with higher ratings garnered more customer purchases, indicating a preference for quality in this category.
b) Fashion accessories and food and beverages mainly comprised lower-rated products in customer purchases.
c) Health and beauty products also leaned towards lower-rated items in customer preferences.
d) However, in the Sports and Travel category, customers showed a tendency to purchase higher-rated products.

Q4. Analyzing units sold and unit price data answer the following sub questions

a) What is the degree of freedom?
b) Co-relation of Unit price and revenue generated
c) What result you can draw from regression of the two data

Ans.

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.010777564					
R Square	0.000116156					
Adjusted R Square	-0.000885732					
Standard Error	2.924724997					

Observations	1000					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	0.9917274	0.991727	0.115937	0.733555221	
Residual	998	8536.908273	8.554016			
Total	999	8537.9				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.443794599	0.215314544	25.28299	2.1E-109	5.021273429	5.86631577
Unit price	0.001189202	0.003492565	0.340495	0.733555	-0.005664411	0.008042815

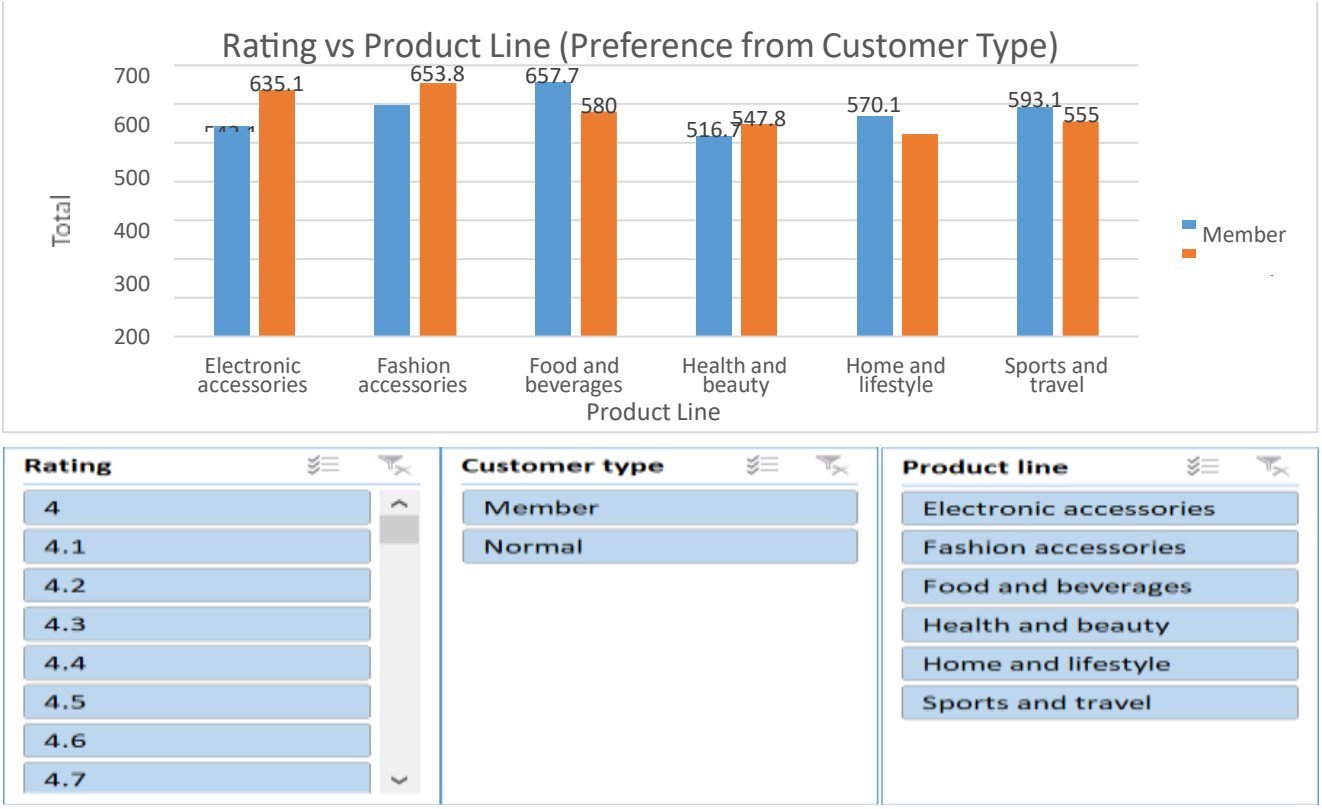
a. The degree of freedom of the analyzed data is 1.

b. The correlation between unit price and generated revenue was found to be 0.63392, indicating a moderate positive relationship. The analysis focused on the columns of unit price and total revenue, employing the CORREL function.

c. Upon examining the regression results, we aimed to discern the relationship between quantity and unit price,exploring how customers' purchasing quantity correlates with the unit price of a product.

However, from the regression analysis, it's evident that the observed trend lacks consistency. The expected outcomes derived from the trend deviate significantly from the actual outcomes.

Q5. What product will you suggest as per the city data analysis to each type of customer .



As per the city Data Analysis, **Food and Beverages** will be a good option for **Member** type customer and **Fashion Accessories** for **Normal** type of customers.

Conclusion and Reviews:-

In summary, the analysis of supermarket sales dynamics reveals valuable insights into consumer behavior and operational trends. Key findings include Mandalay's strong performance, gender-specific ordering patterns, and product recommendations based on city data. Further exploration is recommended on the relationship between product ratings and sales volume, as well as unit price correlation. Clear visuals can enhance understanding, andthe report provides actionable recommendations .

Store Data Analysis

Introduction: This dataset contains sales data from a retail store, covering various details like customer information (such as gender and age group), transaction specifics (like order ID and status), and product details (such as category and SKU). Our goal in analyzing thisdata is to understand how customers behave and what products are popular. By doing this, we can find patterns, preferences, and connections within the data. These insights can then be used by businesses to improve how they market products, manage their inventory more effectively, and make sure customers are happy with their shopping experience.

Questionnaires :

1.Which of the channel performed better than all other channels incompare men & women?

2.Compare category. Find out most sold category above 23 years of age for any gender.

3.Compare Maharashtra, Rajasthan and Tamil Nadu on the basis quantity,most items

purchased by men and women and profit earn.

4.Which city sold most of following categories:

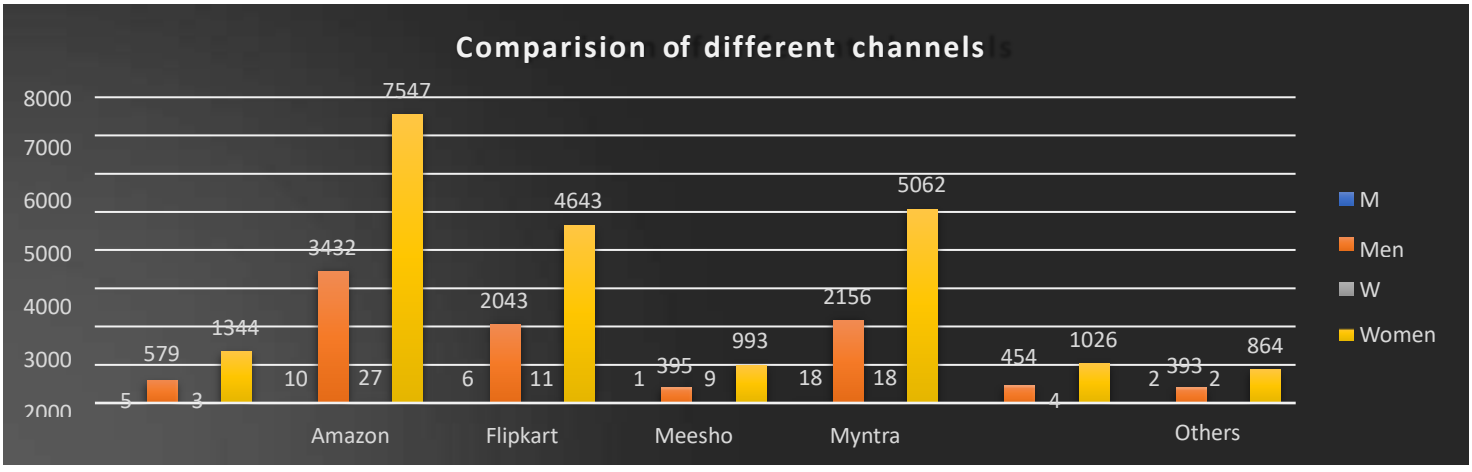
a.kurta b.set c.western wears.

5.In which month most items sold in any of the state on the basis of category

Analytics :

1.Which of the channel performed better than all other channels incompare men & women?

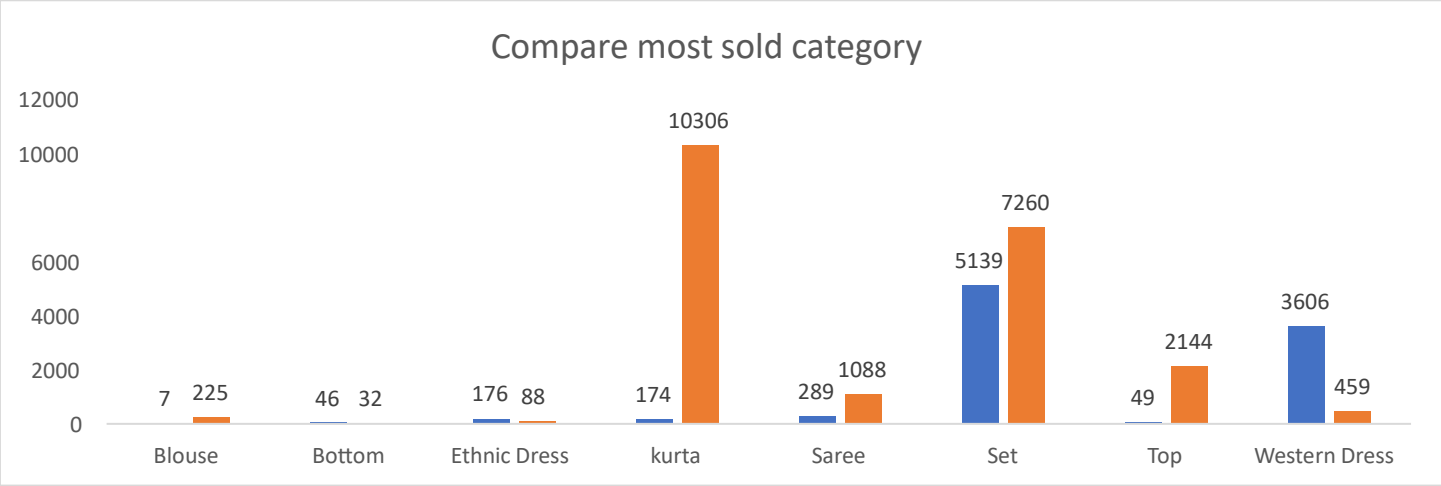
Ans



Amazon is the top seller for both men and women, with Myntra and Flipkart following closely behind. Specifically, Amazon sold nearly 3,500 units in the men's category and almost7,500 units in the women's category. Myntra, on the other hand, sold 2,000 units in the men's section.2

2.Compare category. Find out most sold category above 23 years of age for any gender.

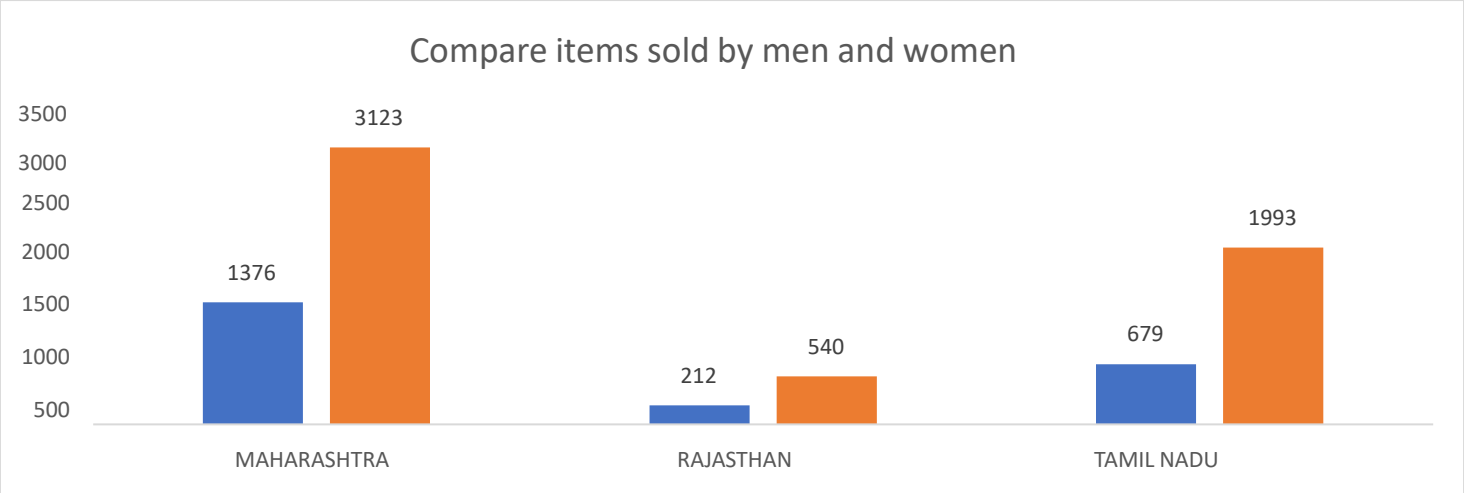
Ans



In the women's section, the most popular category among customers aged 23 years and above is Kurta, with a remarkable 8,820 units sold. Meanwhile, in the men's section, the top- selling category is Set, which saw 4,365 units sold. Interestingly, Set also ranks as the second most popular category in the women's section, indicating its broad appeal across genders.

3. Compare Maharashtra, Rajasthan and Tamil Nadu on the basis quantity,most items purchased by men and women and profit earn.

Ans:-



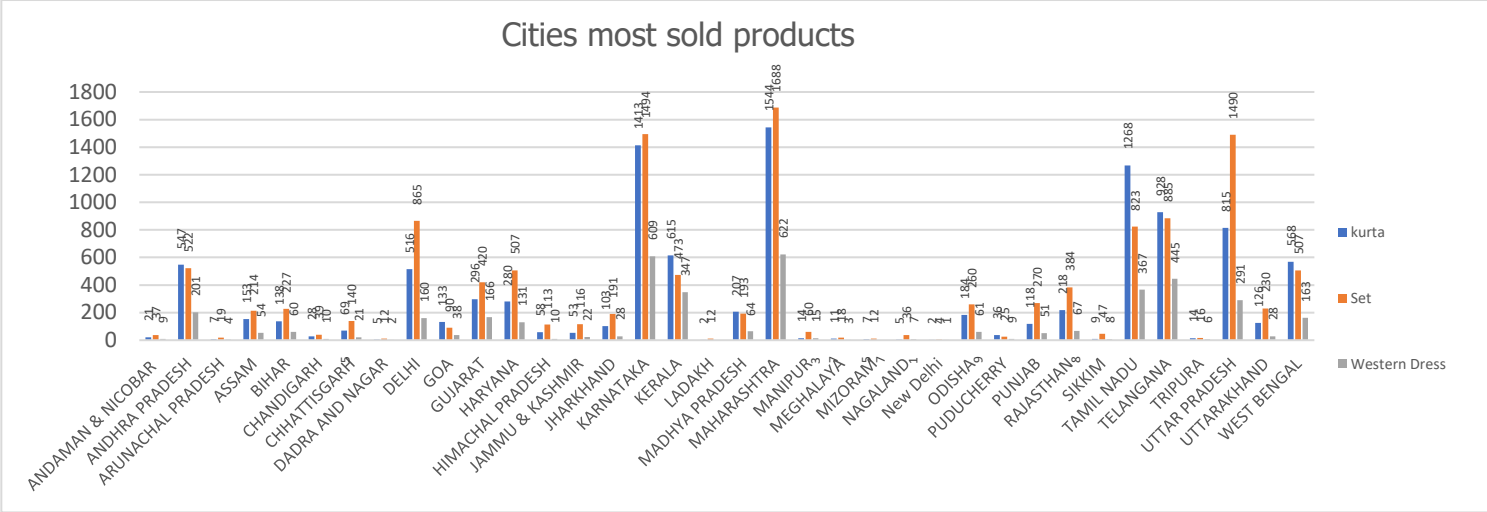
In Maharashtra, sales data indicates that the men's categorysaw a total of 1,390 units sold,while the women's category recorded a significantly higher figure of 3,144 units sold.

Moving on to Tamil Nadu, sales in the men's category amounted to 686 units, with the women'scategory showing a stronger performance at 2,023 units sold. Finally, in Rajasthan, sales werecomparatively lower, with only 21 units sold in the men's category and543 units in the women'scategory. These figures offer insights into regional sales trends, highlighting the varying consumer preferences across different states.

4.Which city sold most of following categories:

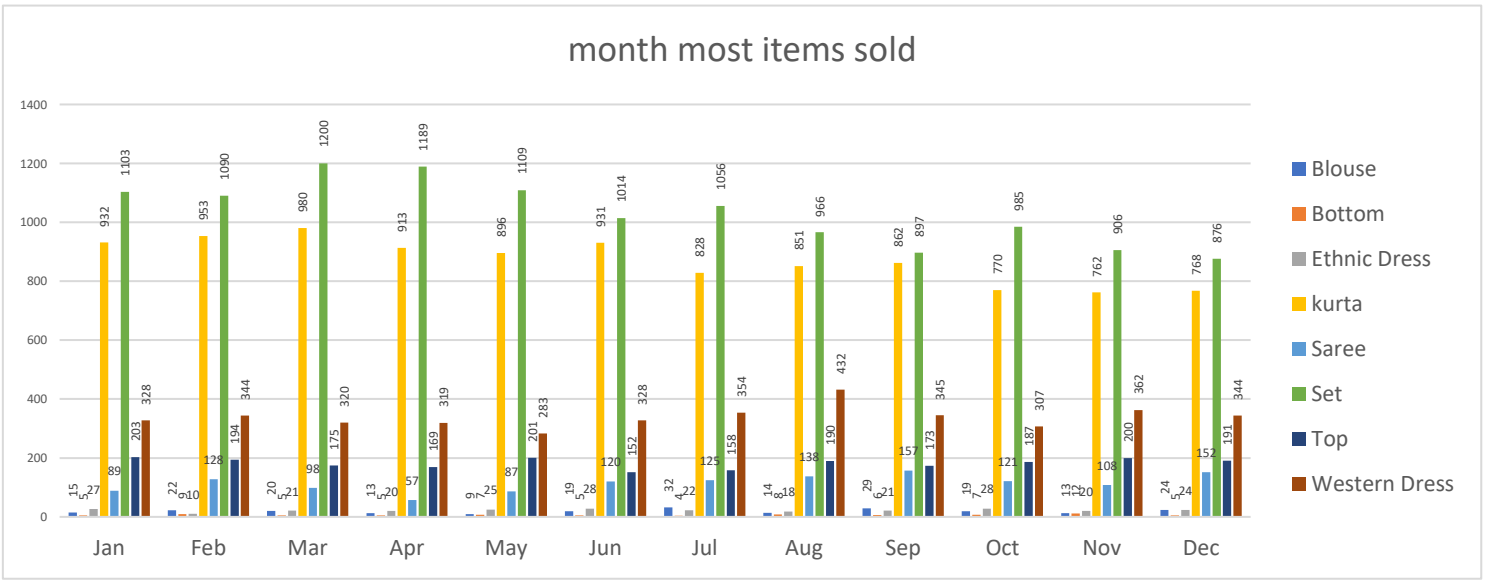
a.kurta b . set c. western wears.

Ans.



Bengaluru, Chennai, Hyderabad, Mumbai, and New Delhi stand out as the top cities forKurta, Set, and Western wear sales. These urban centers consistently show the highest demandfor these clothing categories compared to other cities, indicating strong consumer preferencesfor traditional and contemporary styles.

5.In which month most sold items in any of the state on the basis of category.



The dataset presents monthly sales data for various clothing categories over a year. The categories include Blouse, Bottom, Ethnic Dress, Kurta, Saree, Set, Top, and Western Dress. In January, the total sales were 2702, with Kurta (932) and Set (1103) being the highest-selling categories. February saw a slight increase in total sales to 2750, maintaining high numbers for Kurta (953) and Set (1090). The highest monthly total was in March, with 2819 items sold, again led by Kurta (980) and Set (1200). April's sales dropped to 2685, with a consistent trend of high sales for Kurta (913) and Set (1189). May followed with a slight decrease to 2617, while June saw a total of 2597, keeping Kurta and Set as top categories. July's sales were 2579, and August matched May's total at 2617, with a notable increase in Western Dress (432). September's total was 2490, October saw a slight

decrease to 2424, November recorded 2383, and December ended the year with 2384. Throughout the year, Kurta (10446) and Set (12391) consistently led the sales, contributing significantly to the grand total of 31047 items sold.

Conclusion and Review :

In conclusion, this dataset offers a comprehensive view of sales data from a retail store, encompassing customer demographics, transaction details, and product specifics. Our analysisaims to uncover insights into customer behavior and product popularity, with the goal of identifying patterns, preferences, and connections within the data. By leveraging these insights,businesses can refine their marketing strategies, optimize inventory management practices, and enhance the overall shopping experience for customers. Ultimately, understanding customer behavior and product trends enables businesses to make informed decisions that drive sales growth and foster customer satisfaction.

Car DataSet Analysis

Introduction: The dataset provides comprehensive information about various cars, including their make, model, color, mileage, price, and cost. Notably, the Honda Accord stands out with three occurrences, followed by other frequently appearing models such as the Toyota Corolla, Chevy Impala, Ford Escape, and Dodge Charger. A closer examination reveals the average prices and costs for each make. On average, Hondas are priced at approximately $3,106, with costs averaging around $2,133, while Chevys have an average price of $3,487 and average cost of $3,000. Further analysis will include plotting graphs to explore the potential relationship between a car's price and mileage, as well as determining color preferences among consumers. Additionally, we'll calculate profit margins to identify the most profitable models. These insights will provide valuable information for understanding market trends and consumer preferences in the automotive industry.

Questionaires:

Q1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is givingbest mileage?

Q2. Justify, buying of any Ford car is better than Honda.

Q3. Among all the cars which car color is the most popular and is least popular?

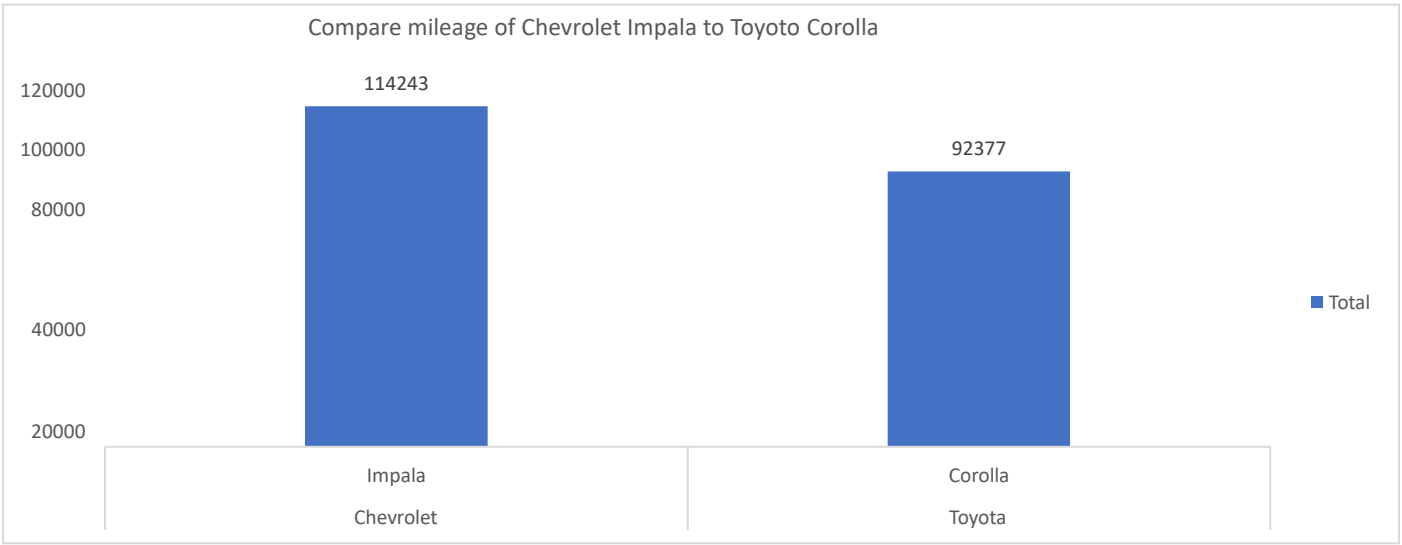
Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.

Q5. Find out all the cars, and their total cost which is more than $2000?

Analytics:

1.Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is givingbest mileage?

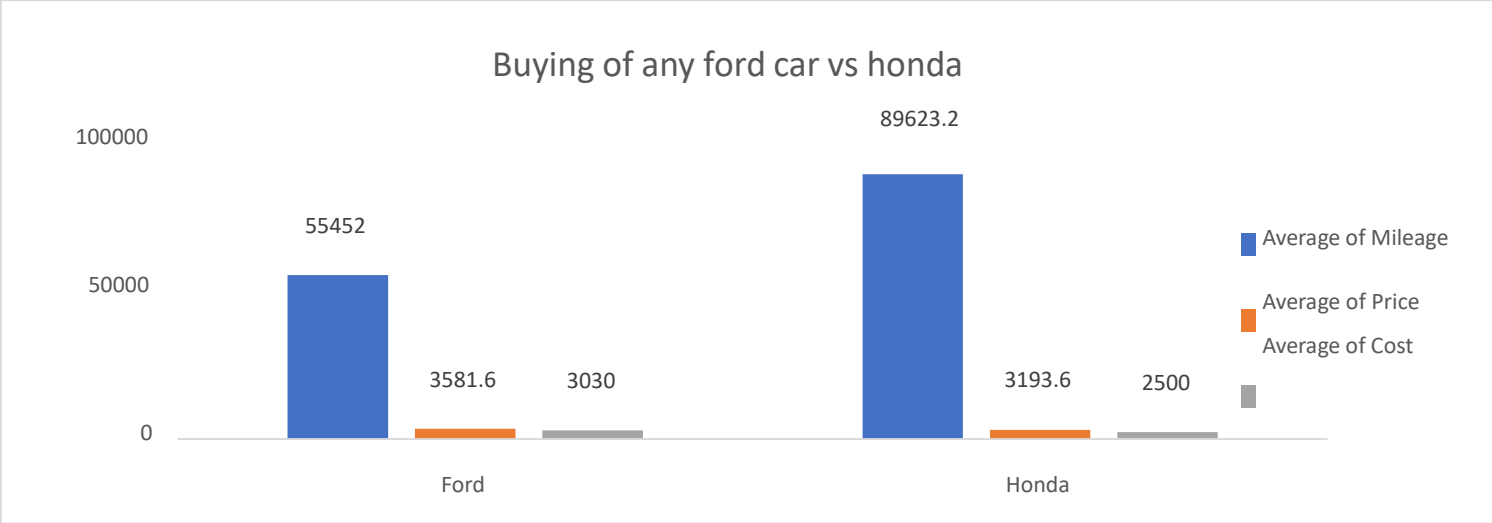
Ans



Toyota Corolla is recognized for its notable fuel efficiency, which is frequently superiorto lar vehicles such as the Chevrolet Impala.

2. Justify, Buying of any Ford car is better than Honda.

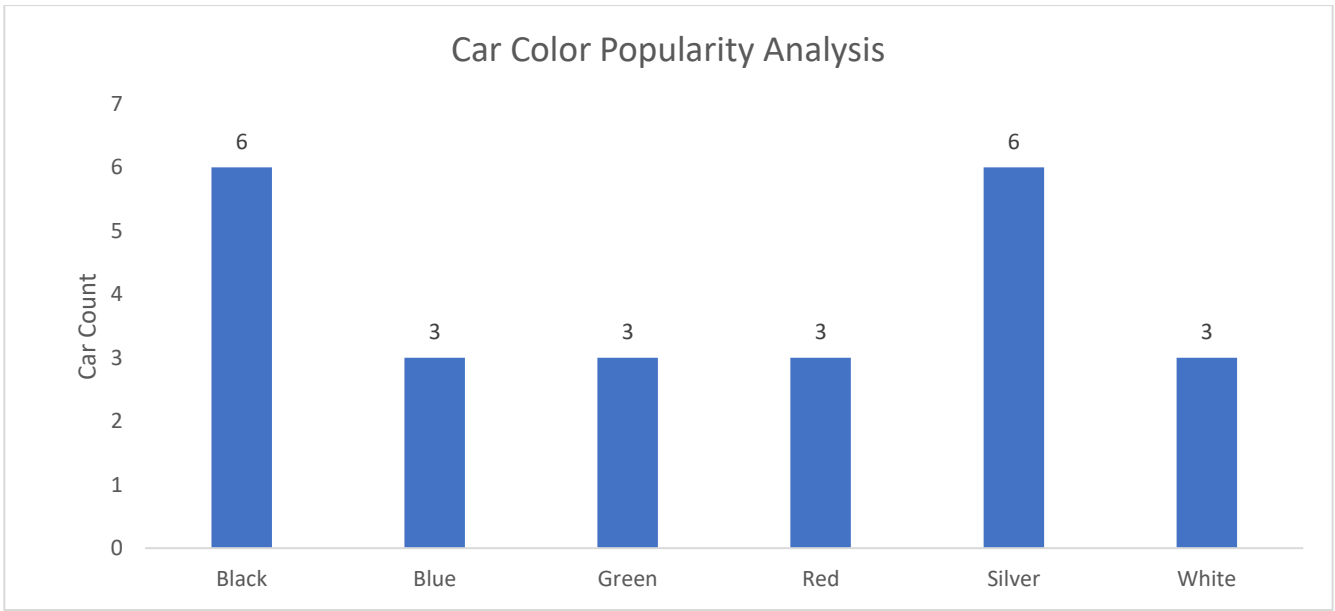
Ans.



Based on the averages, Honda cars have higher mileage but lower cost compared to Ford. Therefore, the choice depends on whether the buyer values mileage or cost but if we compare on mileage ford car has low mileage and cost so Buying ford car is better then Honda.

3. Among all the cars which car color is the most popular and is least popular?

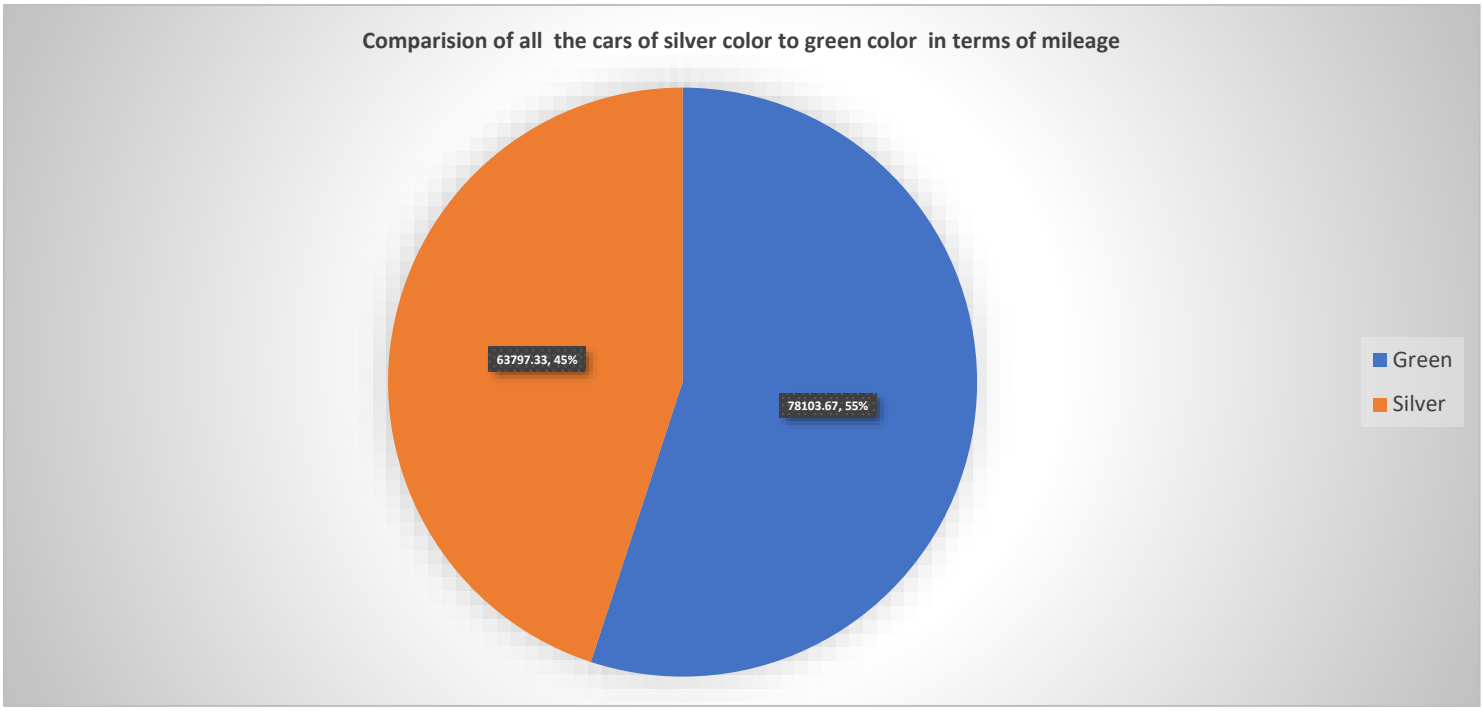
Ans.



Most popular color is Silver and Black as each appear 6 times and least appearing colour are Blue ,Green ,Red ,White they all apper 3 times.

4. Compare all the cars which are of silver color to the green color in terms of Mileage?

Ans

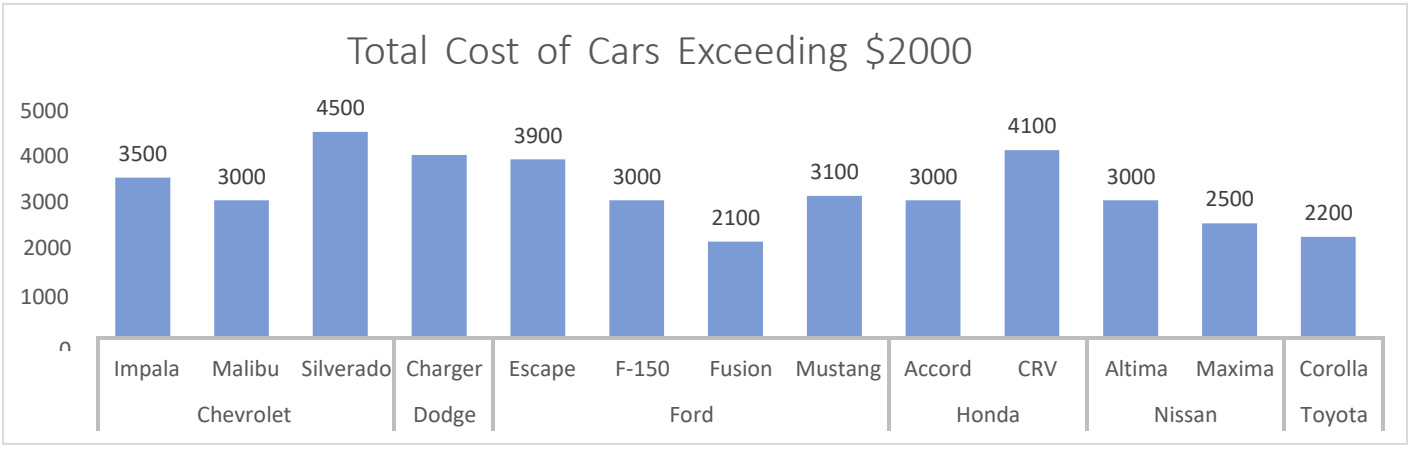


Average Mileage of Green Color Cars ≈ 78103.67 miles

Average Mileage of Silver Color Cars ≈ 63,797 miles

Q5. Find out all the cars, and their total cost which is more than $2000?

Ans.



All the car mention below cost is more than $2000

Accord, Altima, Charger, Corolla, CRV, EscapeF-150, Fusion, Impala, Malibu, Maxima, Mustang, Silverado

Conclusion and Review: -

Our analysis sheds light on what consumers look for when buying cars. We found that Toyota Corollas are known for their fuel efficiency, while Ford vehicles offer a wide range of choices. Consumers seem to prefer black and red cars. Interestingly, silver cars tend to have higher mileage. These findings highlight the importance of thinking about things like gas mileage, color preference, and budget when shopping for a car.

Regression: -

Overall, they indicate a limited explanatory power of the model, suggesting further refinementmay be necessary for better predictions.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.358764572
R Square	0.128712018
Adjusted R Square	0.087222114
Standard Error	32204.73295
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3217481630	3.22E+09	3.102249	0.09273902
Residual	21	21780041315	1.04E+09		
Total	22	24997522945			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	122108.9268	24014.1535	5.084873	4.91E-05	72168.7607	172049.093
X Variable 1	-14.51458144	8.240739406	-1.76132	0.092739	-31.6521372	2.62297432

Anova: Single Factor: -

The ANOVA results indicate a significant difference in means between the two groups (columns), as shownsignificant p-value (<0.05) for the "Between Groups" variation.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.03E+10	1	7.03E+10	123.6791	2.28E-14	4.061706
Within Groups	2.5E+10	44	5.69E+08			
Total	9.53E+100	45				

Anova: Two-Factor Without Replication:

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) v1, respectively. The error term has a degree of freedom of 22.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.23E+10	22	557756895.8	0.962803693	0.535017989	2.04777
Columns	7.03E+10	1	70315407145	121.3789272	2.01396E-10	4.30095
Error	1.27E+10	22	579304898.8			
Total	9.53E+10	45				

Descriptive Statistics: -

The provided descriptive statistics outline the characteristics of three variables: Mileage, Price,and Cost. Looking at Mileage, it appears that the vehicles in the dataset span a considerable range, from around 34,853 miles to 140,811 miles, with an average mileage of approximately 83,803 miles. Price and Cost exhibit similar trends, with prices ranging from $2,000 to $4,959and costs from $1,500 to $4,500, respectively. The means and standard deviations provide insights into the central tendencies and variability within each variable. Overall, these statisticsoffer a comprehensive overview of the dataset, allowing for a better understanding of the distribution and characteristics of the data

Mileage		*Price*		*Cost*	
Mean	83802.7917	Mean	3254.5	Mean	2756.25
Standard Error	7112.65205	Standard Error	186.751181	Standard Error	171.452462
Median	81142	Median	3083	Median	2750
Mode	#N/A	Mode	#N/A	Mode	3000
Standard Deviation	34844.7365	Standard Deviation	914.890205	Standard Deviation	839.942092
Sample Variance	1214155660	Sample Variance	837024.087	Sample Variance	705502.717
Kurtosis	-1.0971827	Kurtosis	-1.2029138	Kurtosis	-0.8126576
Skewness	0.38652215	Skewness	0.27201913	Skewness	0.47339238
Range	105958	Range	2959	Range	3000
m	34853	Minimum	2000	Minimum	1500
Maximum	140811	Maximum	4959	Maximum	4500
Sum	2011267	Sum	78108	Sum	66150
Count	24	Count	24	Count	24

Largest(1)	140811	Largest(1)	4959	Largest(1)	4500
Smallest(1)	34853	Smallest(1)	2000	Smallest(1)	1500

Correlation: -

The correlation coefficient between Column 1 and Column 2 is -0.4110586. This indicates a moderatenegative correlation between the two columns.

	Mileage	*Price*
Mileage	1	-0.4110586
Price	-0.4110586	1

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.

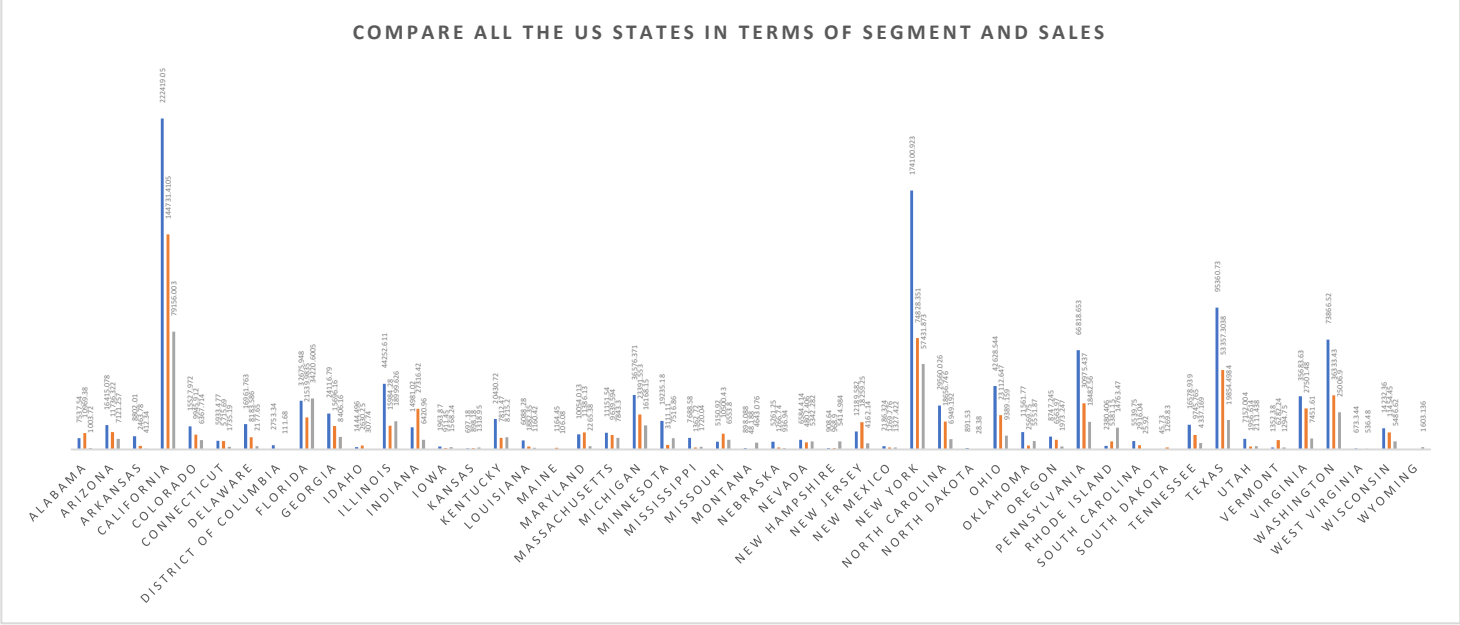
Questionnaire:

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?

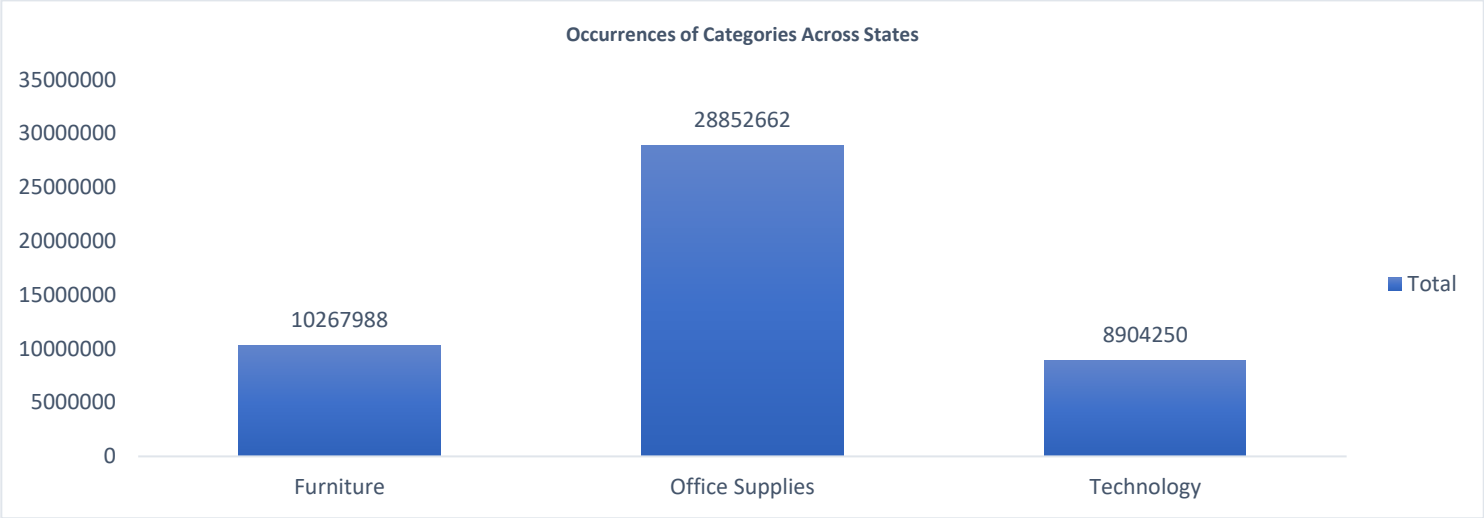
Ans



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?

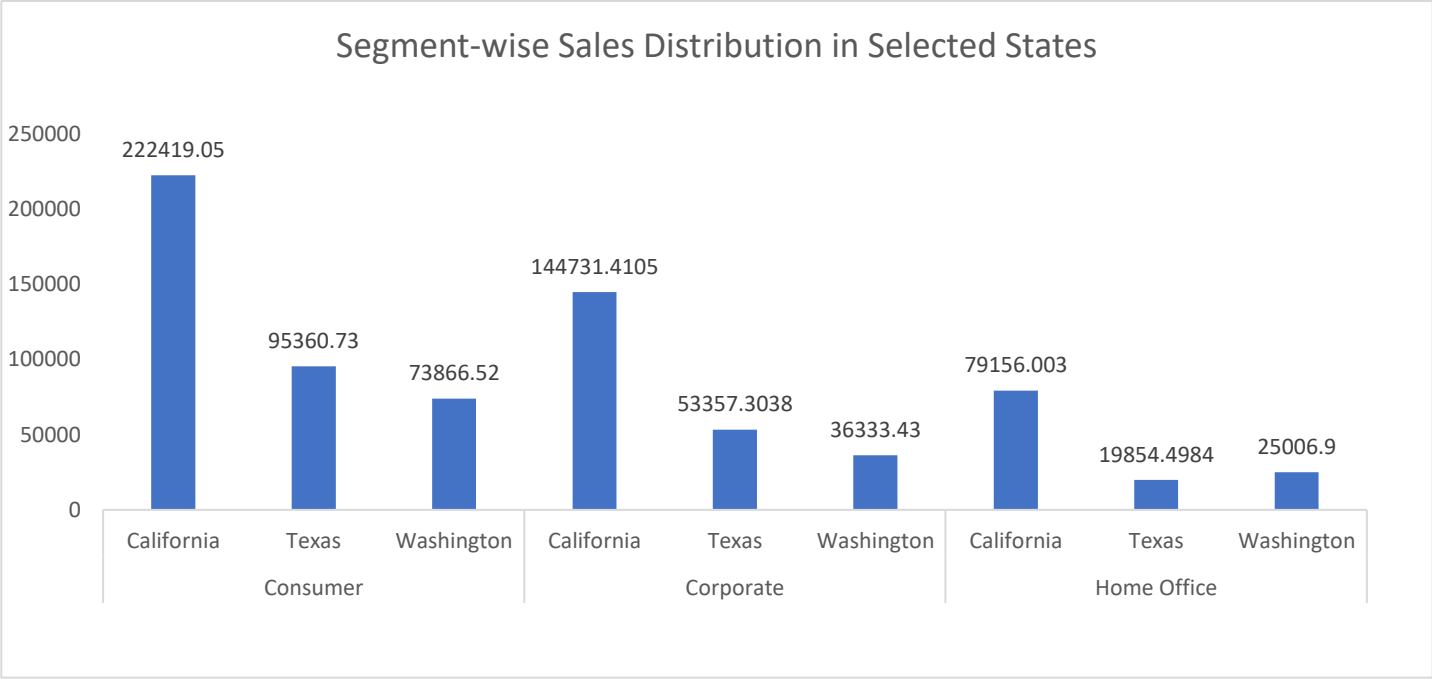
Ans



Office is the top performing category.

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

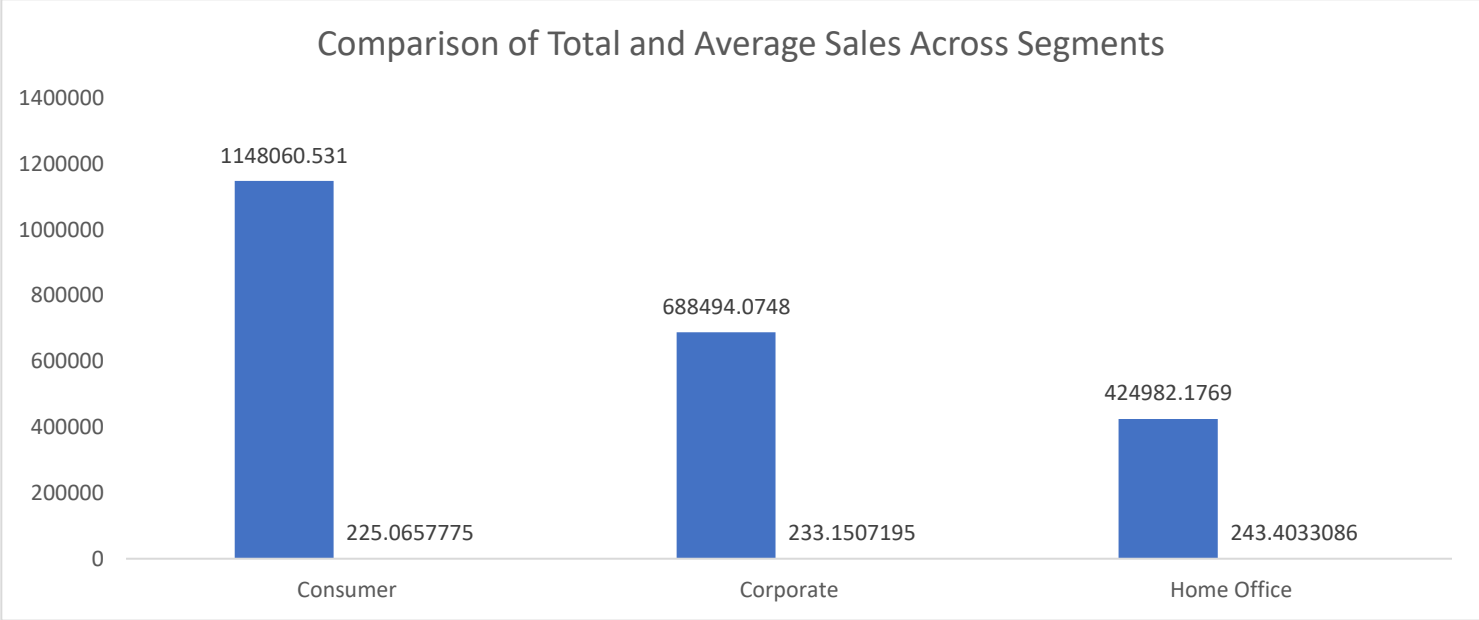
Ans



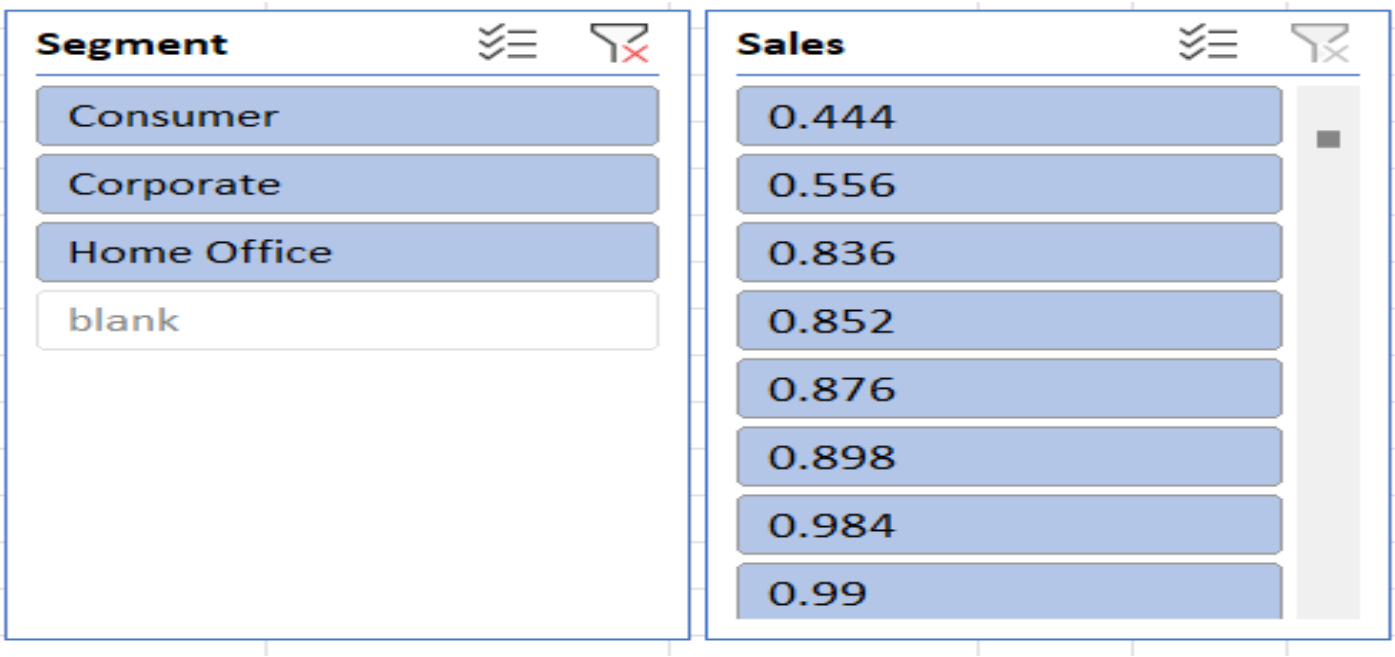
Consumer segment has the most sales in US, California, Texas, and Washington

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

Ans

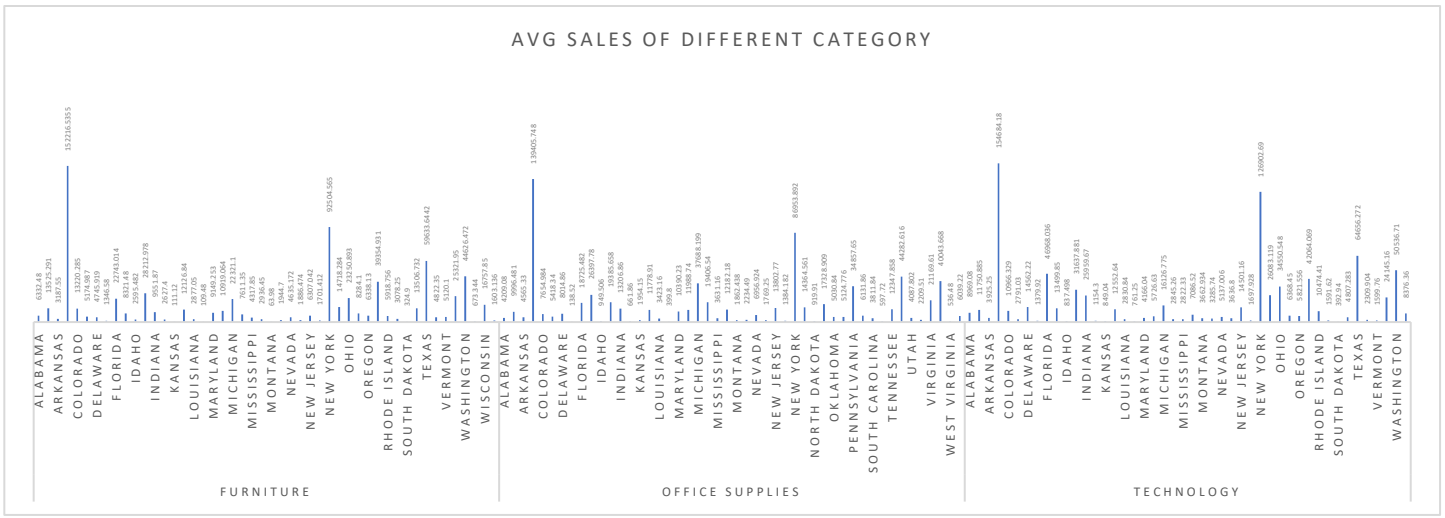


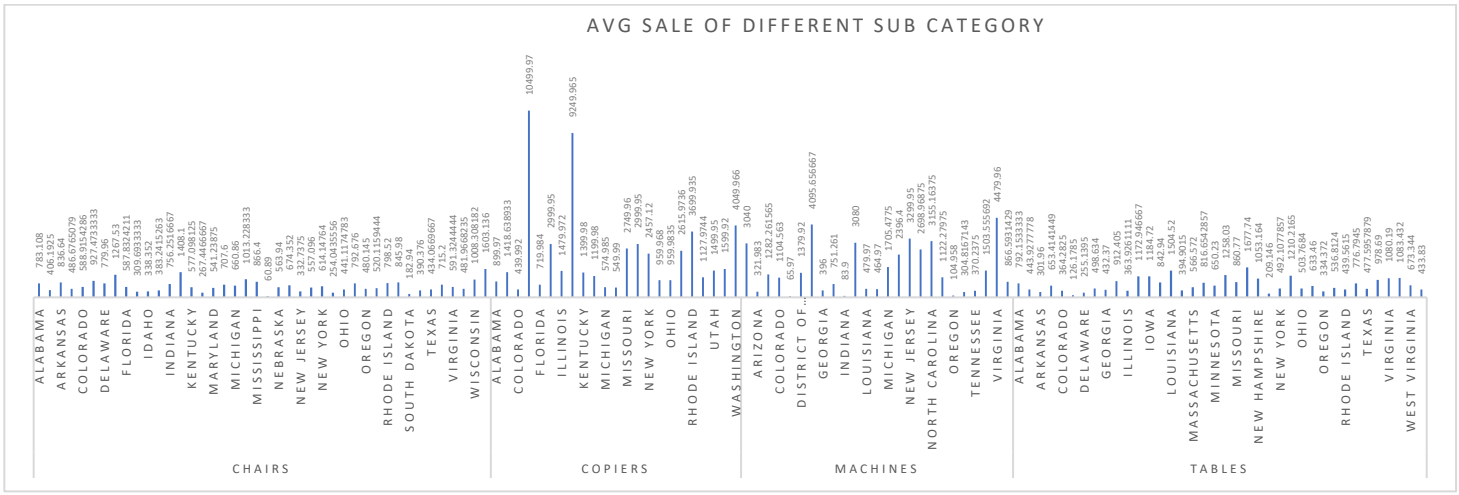
Overall, the home office segment has the highest average sales, followed by the corporate segment and then the consumer segment. However, in terms of total sales, the consumer segment has the highest, followed by the corporate segment and then the home office segment



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) and the 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

Sales Data Samples

Introduction: In the realm of business analytics, a dataset encompassing sales transactions emerges as a vital asset for deriving actionable insights. With columns detailing ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and more, it offers a comprehensive view of sales dynamics. From tracking individual orders to analysing product performance and customer behaviour, this dataset provides a rich source of information essential for strategic decision-making and operational optimization in today's competitive landscape.

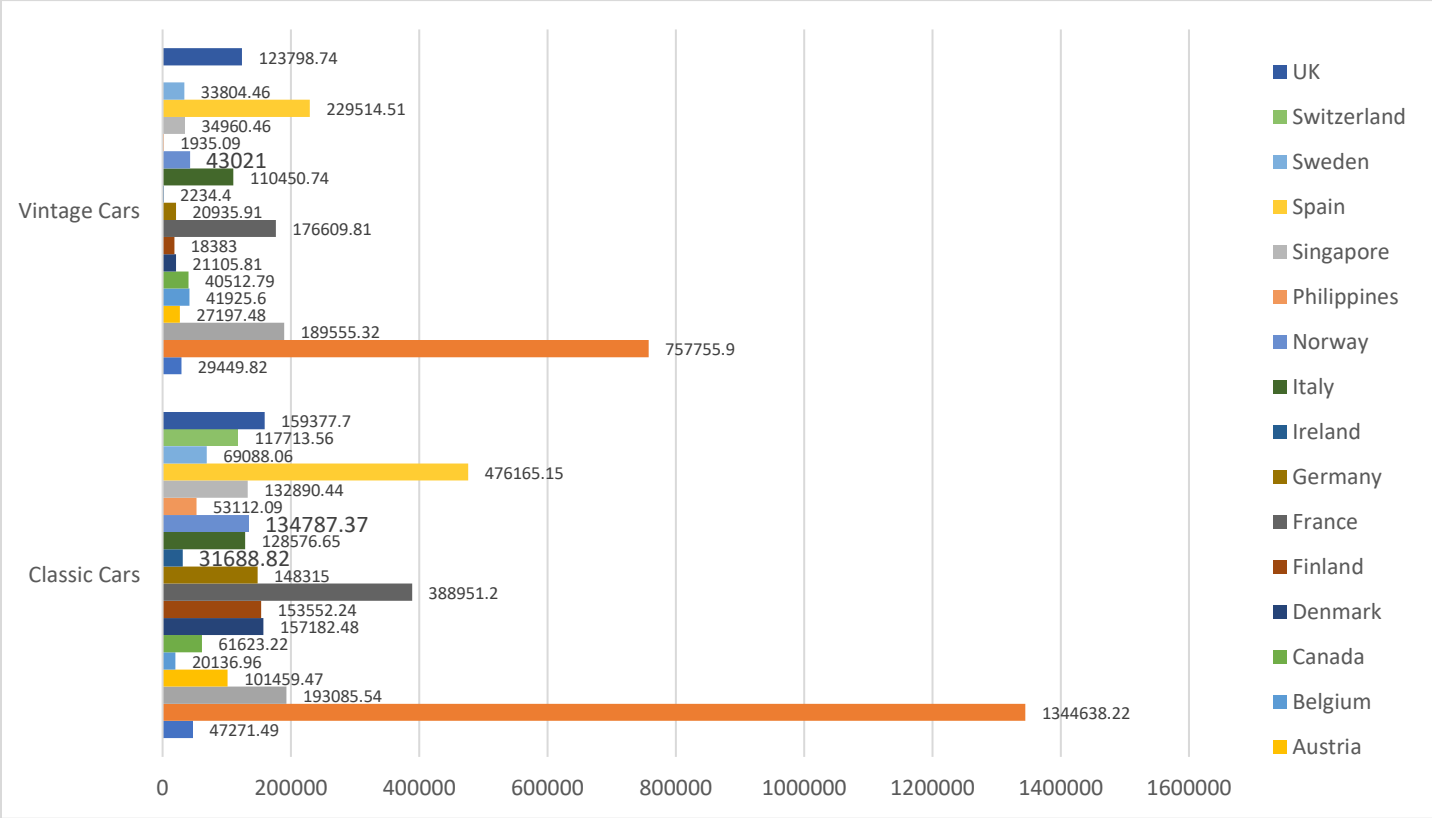
Questionaries:

1. Compare the sale of Vintage cars and Classic cars for all the countries.
2. Find out average sales of all the products? which product yield most sale?
3. Which country yields most of the profit for Motorcycles, Trucks and buses?
4. Compare sales of all the items for the years of 2004, 2005.
5. Compare all the countries based on deal size.

Analytics:

1. Compare the sale of Vintage cars and Classic cars for all the countries.

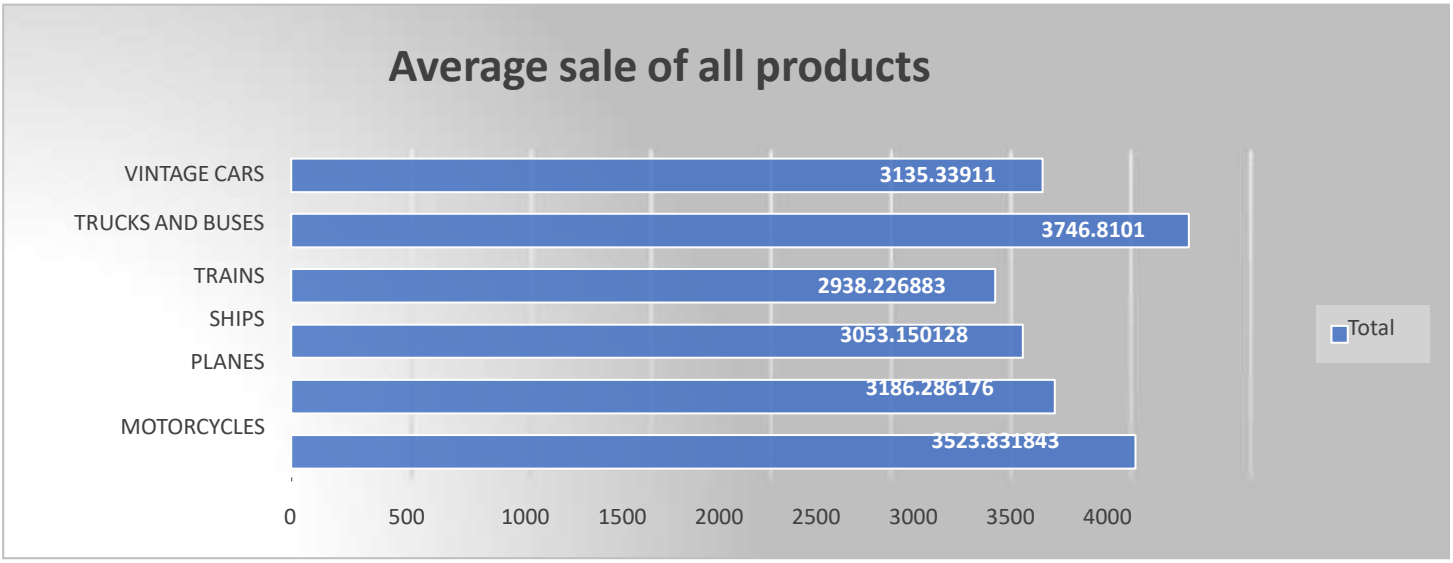
Ans:



On comparing the sales of vintage cars and classic cars we get above graph.

2. Find out average sales of all the products? which product yield most sale?

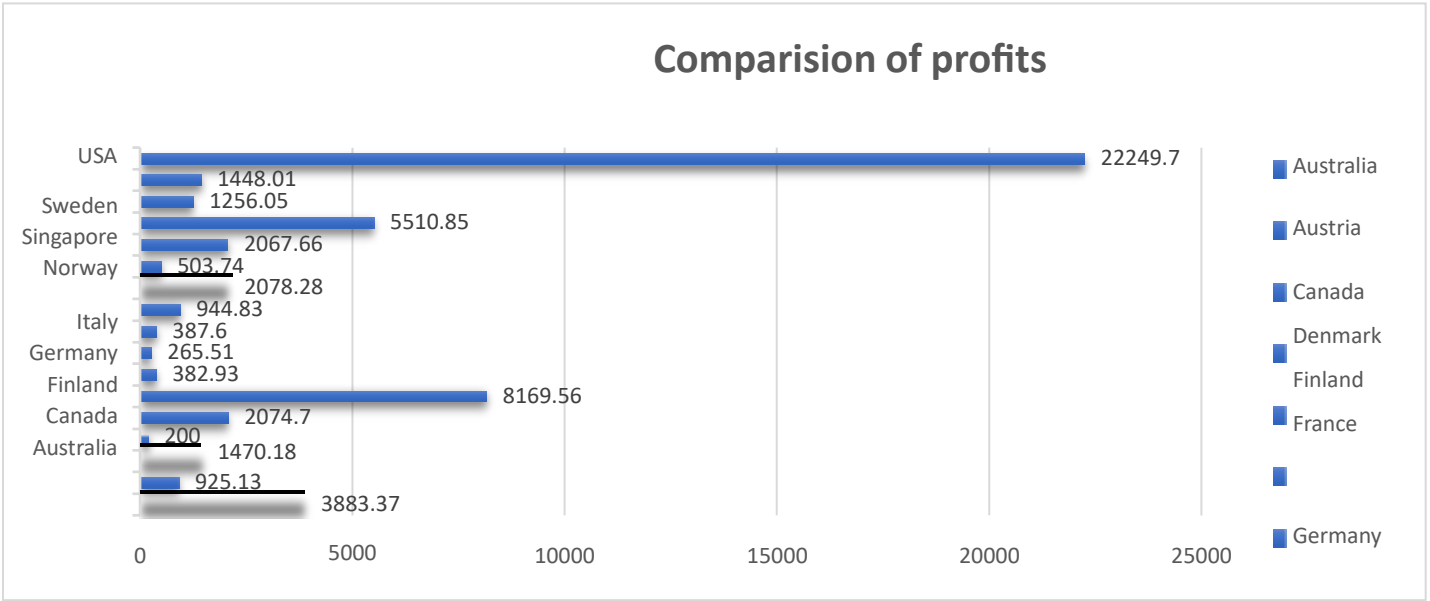
Ans:



The average sale of truck and buses 3746.8101.

3. Which country yields most of the profit for Motorcycles, Trucks and buses?

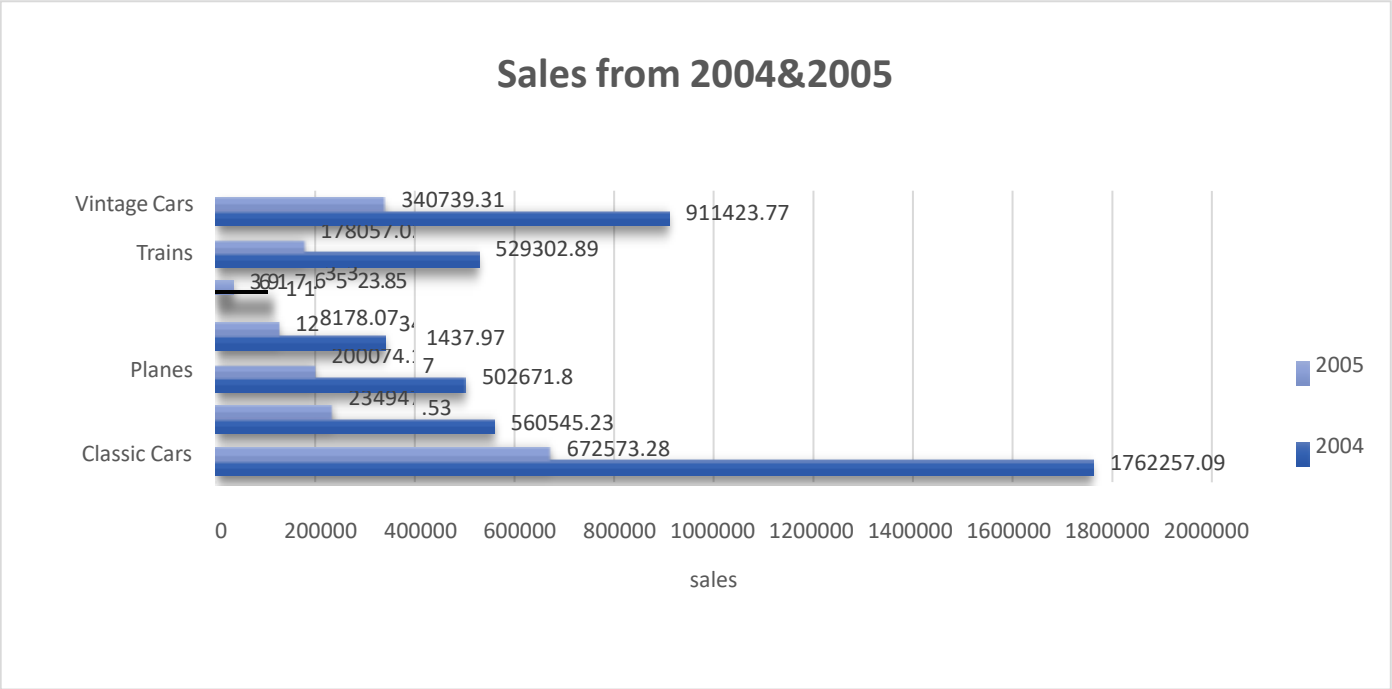
Ans:



The country Australia yields most of the profit for Motorcycles, Trucks and buses

4. Compare sales of all the items for the years of 2004, 2005.

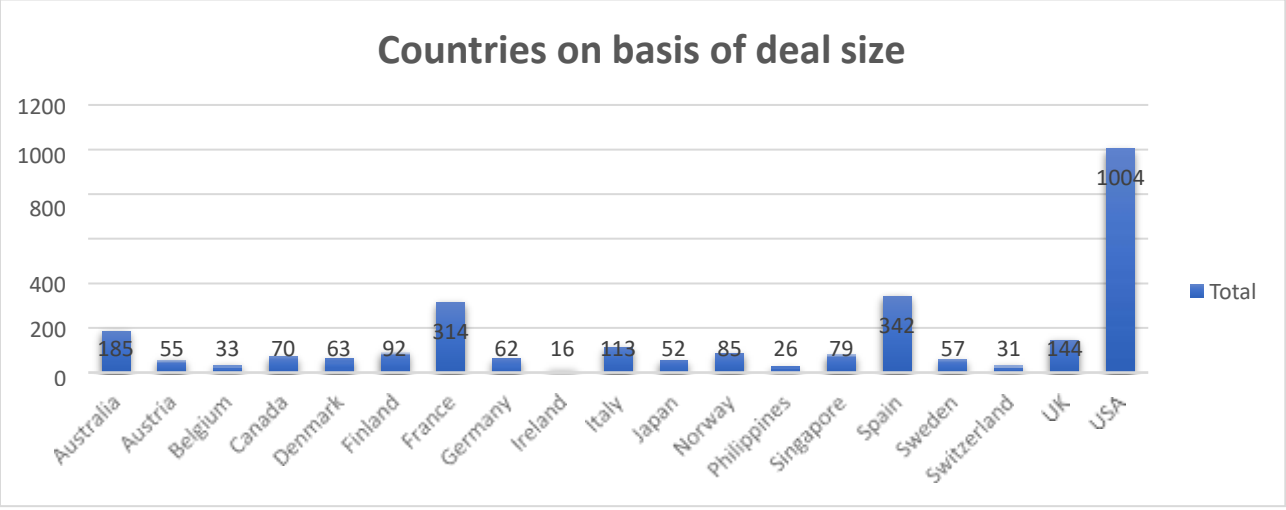
Ans:



he following is the sales of all the items for the years of 2004, 2005 and as graph represents the sales has grown down from 20024 to 2005.

5. Compare all the countries based on deal size.

Ans.



On comparing the countries on the basis of deal size,USA comes out to be highest of 1004 followed by Spain , France.

Conclusion and Review:

In conclusion, the analysis of the provided sales dataset offers a window into the intricacies of business operations, shedding light on customer preferences, product performance, and market trends. By leveraging the insights gleaned from this dataset, businesses can make informed decisions, streamline processes, and drive growth. As the landscape of data analytics continues to evolve, harnessing the power of such datasets remains instrumental in staying competitive and responsive to the ever-changing demands of the market.

Loan Dataset Analysis:

This report delves into an analysis of loan applications, aiming to extract insights into applicant demographics and loan characteristics. The dataset encompasses information such as gender, marital status, education, income, loan amount, loan term, credit history, and property area. By scrutinizing this data, we aim to discern patterns and trends regarding loan applications among different demographic groups and geographical areas..

Questionnaire:-

Q1. How many male graduates who are not married applied for Loan? What was the highest amount?

Q2. How many female graduates who are not married applied for Loan? What was the highest amount?

Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?

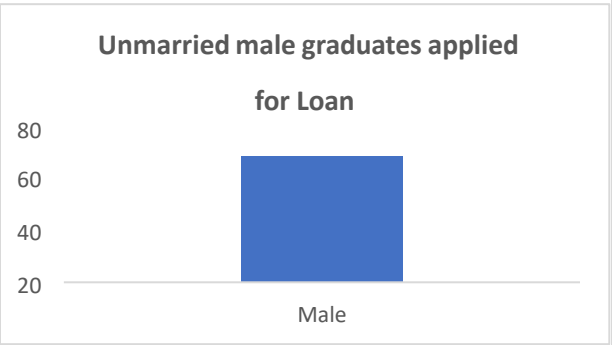
Q4. How many female graduates who are married applied for Loan? What was the highest amount?

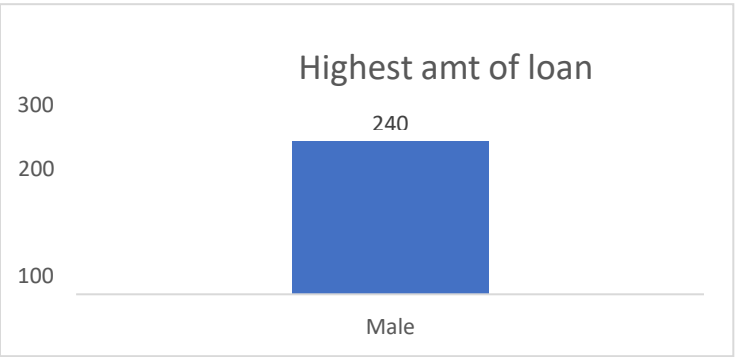
Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rular on the basis of amount.

Analytics:-

Q1. How many male graduates who are not married applied for Loan? What was the highest amount?

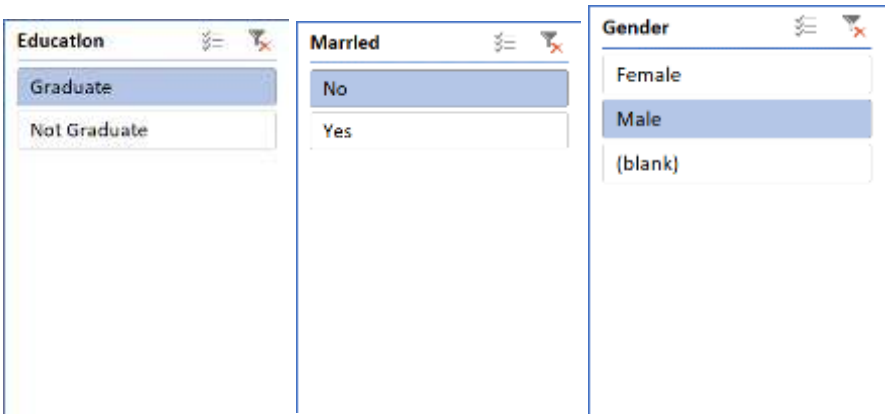
Ans.





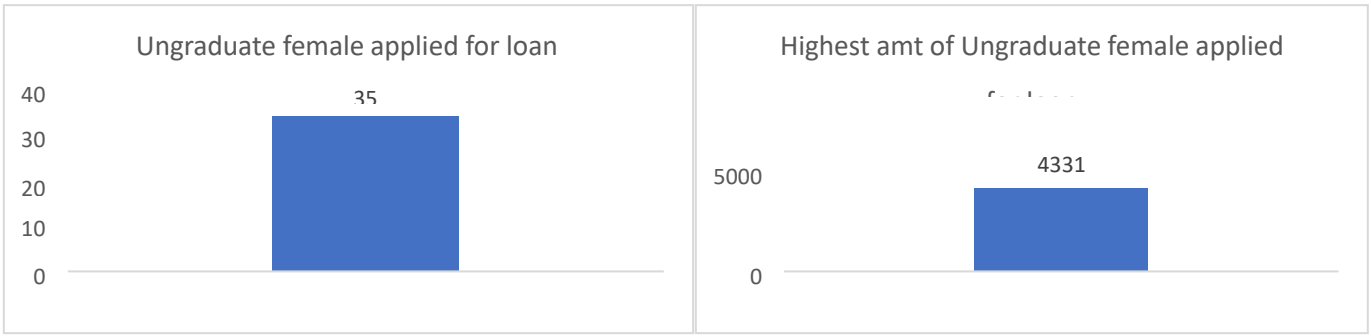
There are total 66 unmarried graduate man applied for loan.

The highest amount of the loan is 240.



Q2. How many female graduates who are not married applied for Loan? What was the highest amount?

Ans.



There are total 35 unmarried graduate female applied for loan.

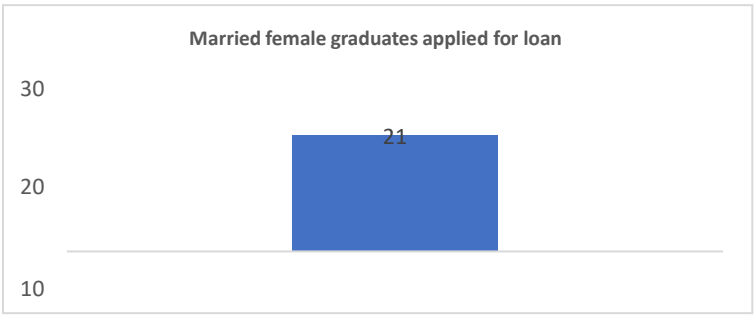
The highest amount of the loan is 4331.

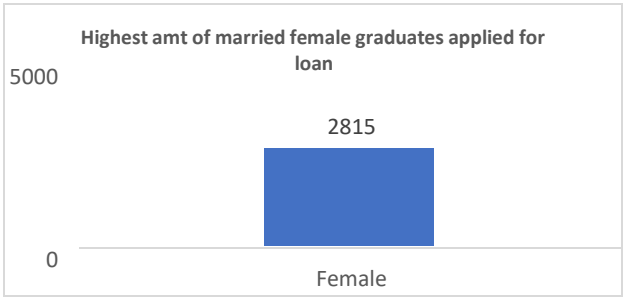
Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?

There are total 16 unmarried graduate male applied for loan.

The highest amount of the loan is 199.

Q4. How many female graduates who are married applied for Loan? What was the highest amount?



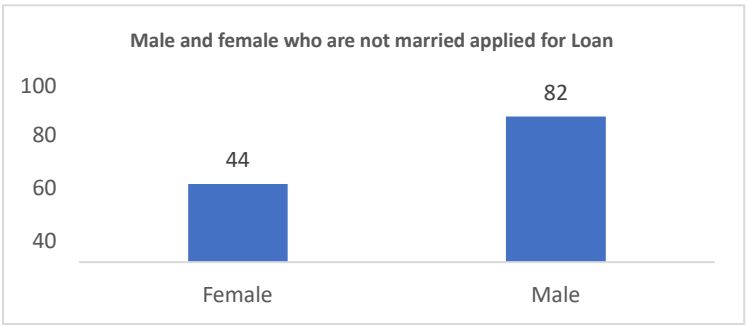


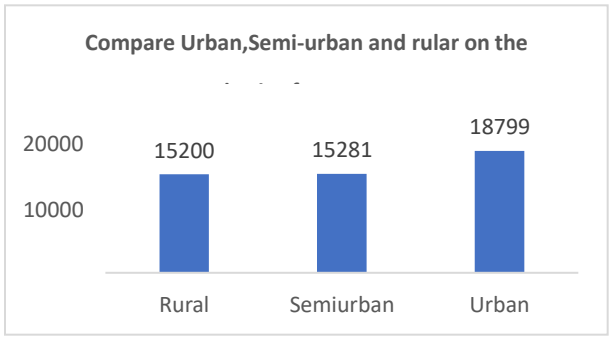
There are total 21married graduate female applied for loan.

The highest amount of the loan is 2815.

Q5. How many male and female who are not married applied for Loan? Compare Urban,Semi-urban and rular on the basis of amount.

Ans.





There are total 44 unmarried female and 82 unmarried male applied for loan.

The rural amount of the loan is 15200 and of semiurban 15281 and of urban is 18799.

Conclusion:

Our analysis, using varied visualization techniques, revealed valuable insights, enhancing comprehension and decision-making. Visualizing data clarified complex findings, facilitating actionable strategies. This highlights the pivotal role of data visualization in extracting meaningful insights and informing decisions effectively.

Regression:

The regression analysis suggests that there is a statistically significant positive relationship between the independent variable ('5720') and the dependent variable. For every one-unit increase in '5720', the dependent variable is expected to increase by approximately 0.0059 units. However, it's important to note that the model only accounts for about 21.1% of the total variance in the dependent variable.

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.45908096
R Square	0.21075532
Adjusted R Square	0.20858707
Standard Error	56.0766111
Observations	366

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	305655.205	305655.205	97.2004502	1.7676E-20
Residual	364	1144629.42	3144.58631		
Total	365	1450284.62			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*
Intercept	106.07753	4.10024098	25.8710478	1.7585E-84	98.014396	114.140665	98.014396
5720	0.0058851	0.00059692	9.85902887	1.7676E-20	0.00471125	0.00705895	0.00471125

Correlation:-

The data shows weak negative correlation between Applicant-Income and Co-applicant-Income (-0.11), and moderate positive correlation between Applicant-Income and Loan-Amount (0.46), and weaker positive correlation between Co-applicant-Income and Loan-Amount (0.14).

	ApplicantIncome	*CoapplicantIncome*	*LoanAmount*
ApplicantIncome	1		
CoapplicantIncome	-0.110334799	1	
LoanAmount	0.458768926	0.144787815	1

Anova (Single Factor) :

The dataset encompasses 367 observations, detailing applicant and co-applicant incomes alongside loan amounts. On average, applicants possess a higher income, averaging around $4805.60, compared to co-applicants whose average income is approximately $1569.58. Loan amounts vary widely, averaging $134.28. ANOVA analysis underscores significant distinctions between the income and loan amounts across the groups, implying diverse financial profiles among applicants and co-applicants.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
ApplicantIncome	367	1763655	4805.599455	24114831.09
CoapplicantIncome	367	576035	1569.577657	5448639.491
LoanAmount	367	49280	134.2779292	3964.141124

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4202537452	2	2101268726	213.2009841	5.87569E-79	3.003920577
Within Groups	10821681107	1098	9855811.573			
Total	1502421856	1100				

Anova two factor without Replication:

The ANOVA results indicate significant variation both within rows ($p = 0.441$) and between columns ($p < 0.001$). This suggests that there are meaningful differences among the row categories and column categories in the dataset,warranting further investigation into the factors influencing these variations.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1004340909	365	2751618.93	1.015674698	0.440986529	1.1881716
Columns	379216841.8	1	379216841.8	139.9761235	1.47092E-27	3.867061668
Error	988841123.7	365	2709153.763			
Total	2372398875	731				

Descriptive Statistics:

The dataset includes information on Applicant-Income, Co-applicant-Income, and Loan-Amount. The largest Applicant-Income recorded is $72,529, while the smallest is $0. For Co-applicant-Income, the largest value is $24,000, and the smallest is $0. Additionally, the Loan-Amount ranges from a maximum of $550 to a minimumof $0. Confidence levels for these variables at a 95.0% level are also provided, indicating the precision of the measurements within the dataset.

Largest(1)	72529	Largest(1)	24000	Largest(1)	550
Smallest(1)	0	Smallest(1)	0	Smallest(1)	0
Confidence Level(95.0%)	504.0756067	Confidence Level(95.0%)	239.6059543	Confidence Level(95.0%)	6.462910219

Cookie Data Analysis

Introduction:-Our dataset is all about cookies—specifically six types: Chocolate Chip,Fortune Cookie, Sugar, Oatmeal Raisin, Snickerdoodle, and White Chocolate Macadamia Nut. We've gathered a wealth of information on these cookies, including how many units were sold, their costs, the money they brought in (revenue), and the profits they made. But we're not justlooking at one place or time; we're exploring different countries and dates to see how things vary.This report goes beyond cookies; it's about understanding people's preferences, how muchthey're willing to pay, and where these treats are most popular. So, get ready to uncover somefascinating insights into the cookie world and what it means for businesses like yours.

Questionaries :

1 . Compare Malaysia and Philippines on the bases of two types of Cookies ?

2. What is the performance of Choco Chips Cookies in all Country Which Competes the best.

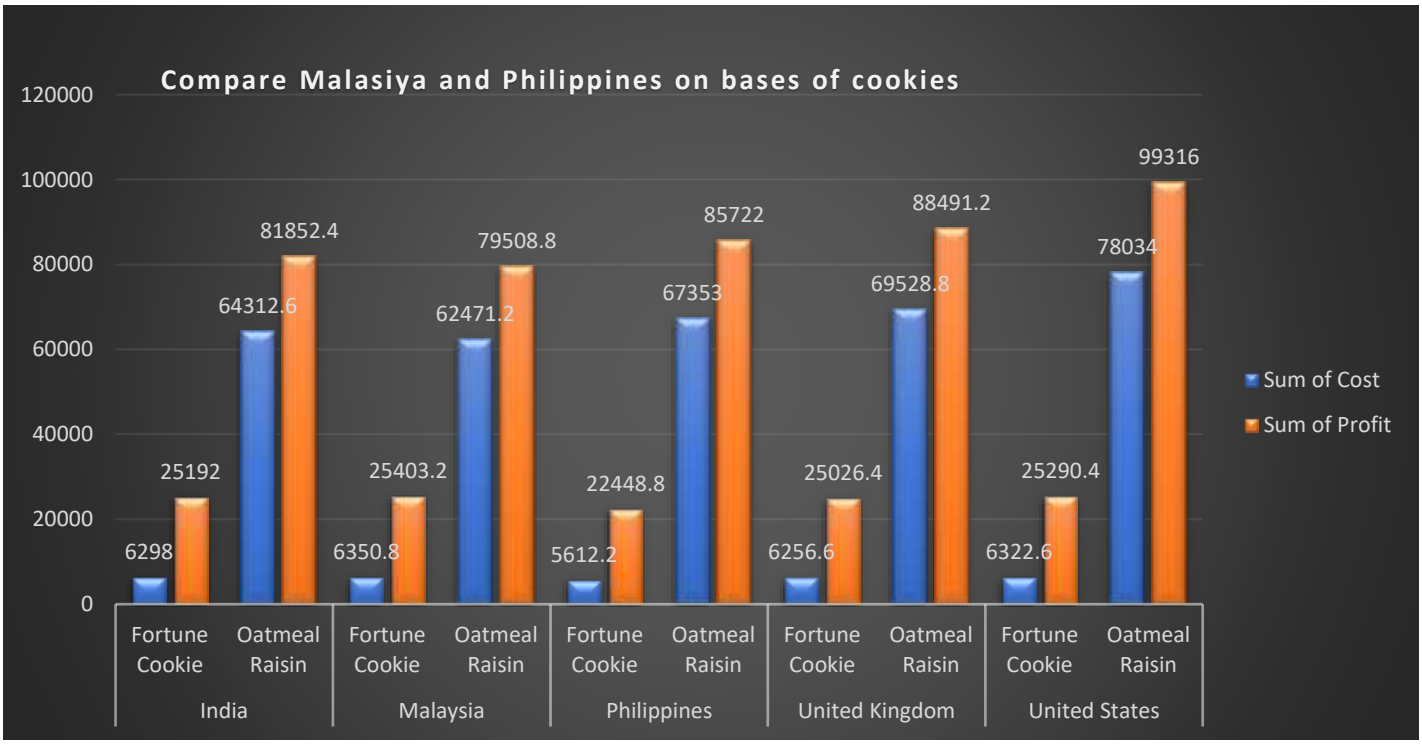
3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

4. Which Cookie is the best Selling Cookie in India and US in year 2019?

Analytics :

1 . Compare Malaysia and Philippines on the bases of two types of Cookies.

Ans:-



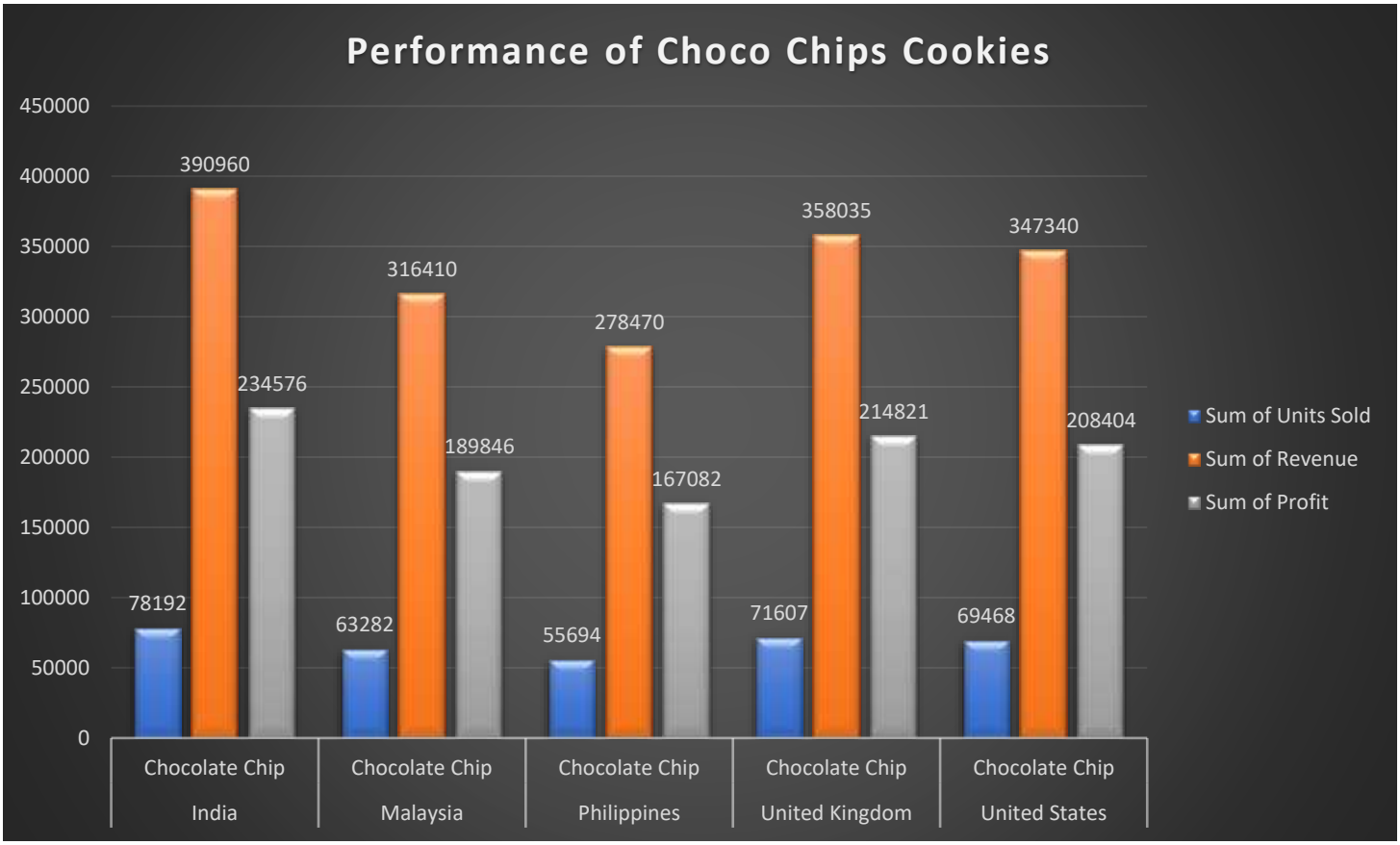
On comparing the Malaysia and Philippines the profit of oatmeal raisin in Philippines is more than in Malaysia and profit of the fortune cookie is more in Malaysia as compare to Phillipines.

The profit of the fortune cookie is 25403.2 in Malaysia and of oatmeal is 25403.

The profit of the fortune cookie is 5612.2 in Philippines and of oatmeal is 85722.

2.What is the performance of Choco Chips Cookies in all Country Which Competes the best.

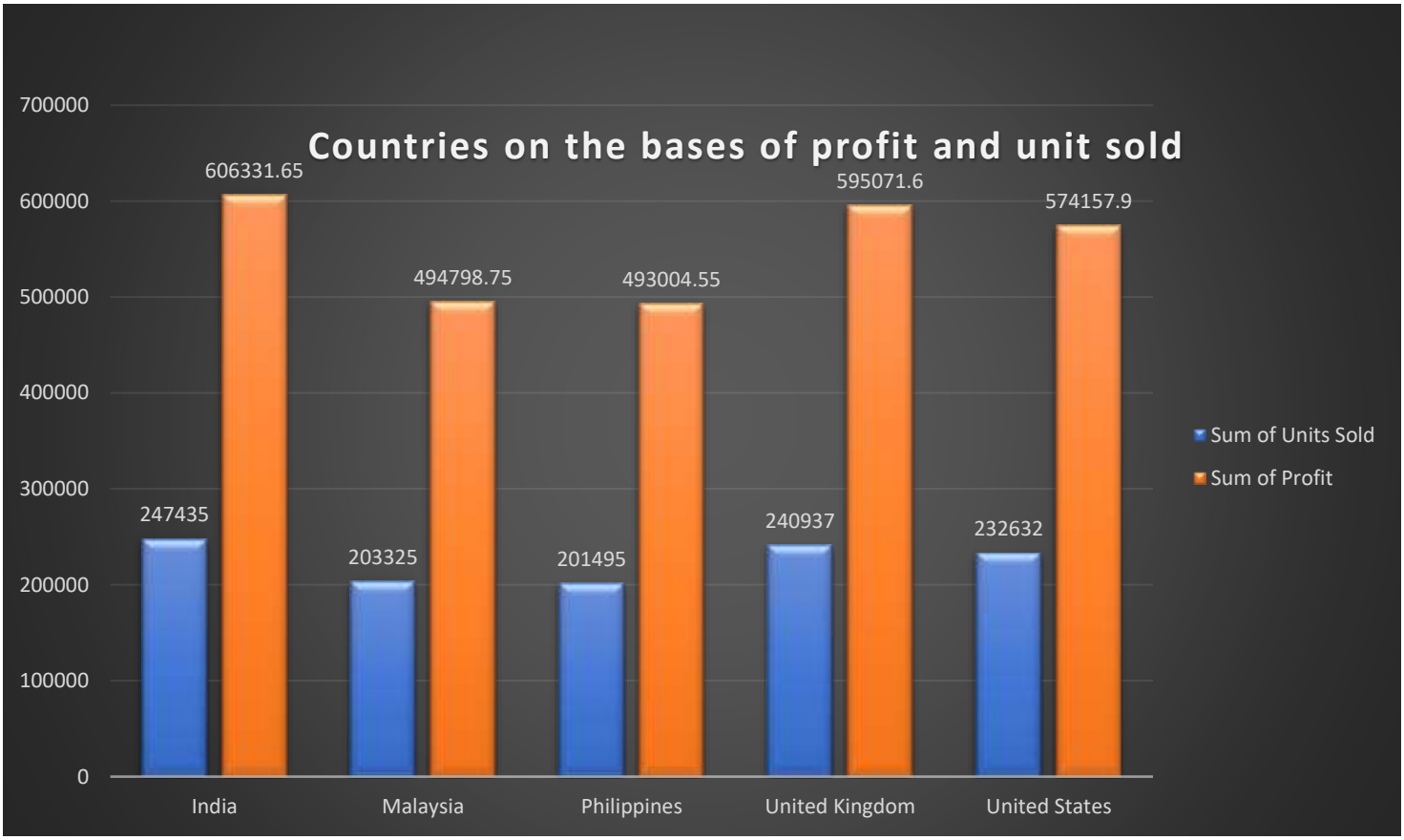
Ans.



India stands out as the foremost consumer of Choco chips worldwide, primarily due to its exceptional profitability and record-breaking sales figures. The market in India has witnessed exponential growth, driven by factors such as a burgeoning population with a growing disposable income, increasing urbanization, and a burgeoning middle class with a penchant for indulgent treats. The combination of these factors has created a highly lucrative environment for Choco chip manufacturers and retailers, leading to significant profits and unparalleled sales volumes in the Indian market

3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

Ans:-

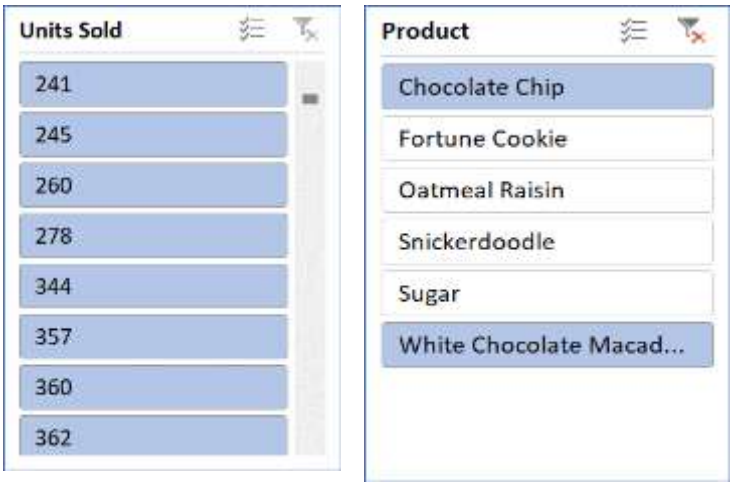


On comparing all the countries, India stands out as the leading performer globally when it comes to both profit generation and units sold in the Choco chip market.

4 .Which Cookie is the best Selling Cookie in India and US in year 2019?

Ans:-

In the year 2019, chocolate chip cookies emerged as the top-selling cookie in both India and the United States.



Conclusion and Review :

After thorough analysis of the cookie sales data, it is evident that there are notable trends and insights to be gleaned. By examining key metrics such as units sold, revenue, cost, and profit across different countries and products, we can draw valuable conclusions about market demand, pricing strategies, and overall profitability. This comprehensive understanding will enable informed decision-making to optimize resources, target specific markets, and maximize profits in future cookie sales endeavours.

Regression:

The regression model, with a significant p-value ($p < 0.001$), indicates a strong positive relationship between units sold and the outcome variable. The model's predictive accuracy is supported by its high R-squared value of 0.688, suggesting that approximately 68.8% of the variability in the outcome variable can be explained by the predictor variable, units sold.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.829304
R Square	0.687746
Adjusted R Square	0.687298
Standard Error	1462.76
Observations	700

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.29E+09	3.29E+09	1537.356	1.4E-178
Residual	698	1.49E+09	2139668		
Total	699	4.78E+09			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-74.4103	116.5304	-0.63855	0.523326	-303.202	154.3817	-303.202	154.3817
Units Sold	2.500792	0.063781	39.20914	1.4E-178	2.375567	2.626017	2.375567	2.626017

Correlation:

The correlation coefficient between units sold and revenue is 0.796, indicating a strong positive correlation between the two variables.

	Units Sold	Revenue
Units Sold	1	0.796298
Revenue	0.796298	1

Anova (Single Factor) :

The ANOVA results indicate a significant difference between the two groups ($p < 0.001$), with 1 degree of freedom. The within-group error is 7681356717, and the total R-squared value is 0.06, suggesting that the model explains 6% of the variability in the data.

SUMMARY

Groups	Count	Sum	Average	Variance
3450	699	1923505	2751.795	4154648
5175	699	2758189	3945.908	6850161

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	4.98E+08	1	4.98E+08	90.57022	7.53E-21	3.848129
Within Groups	7.68E+09	1396	5502405			
Total	8.18E+09	1397				

Anova two factor without Replication:

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) values of 48 and 3, respectively. The error term has a degree of freedom of 144.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	8.21E+08	48	17108242	5.848894	8.54E-17	1.445925
Columns	5.65E+10	3	1.88E+10	6435.486	3.8E-153	2.667443
Error	4.21E+08	144	2925039			
Total	5.77E+10	195				

Anova two factor with Replication:

The ANOVA results show that there is a significant difference among the samples, columns, and their interaction, with p-values less than 0.001. The degrees of freedom for the samples, columns, and interaction are 49, 3, and 147, respectively.

Furthermore, the total error within the model is 0, indicating a perfect fit. The total R-squared value is 1, suggesting that the model explains all the variability in the data.

ANOVA						
Sample	8.55E+08	49	17443674	65535	#NUM!	#NUM!
Columns	5.78E+10	3	1.93E+10	65535	#NUM!	#NUM!
Interaction	4.39E+08	147	2983765	65535	#NUM!	#NUM!
Within	0	0	65535			
Total	5.91E+10	199				

Descriptive Statistics:

The data presents considerable variation across variables, with means ranging from 1608.15 to 43949.81. Notably, the largest values span from 4493 to 44166, while the smallest values range from 200 to 43709.

Unit price		*Quantity*		*Tax 5%*	
Mean	55.67213	Mean	5.51	Mean	15.37937
Standard Error	0.837834	Standard Error	0.092447	Standard Error	0.370266
Median	55.23	Median	5	Median	12.088
Mode	83.77	Mode	10	Mode	39.48

Standard Deviation	26.49463	Standard Deviation	2.923431	Standard Deviation	11.70883
Sample Variance	701.9653	Sample Variance	8.546446	Sample Variance	137.0966
Kurtosis	-1.21859	Kurtosis	-1.21555	Kurtosis	-0.08188
Skewness	0.007077	Skewness	0.012941	Skewness	0.89257
Range	89.88	Range	9	Range	49.1415
Minimum	10.08	Minimum	1	Minimum	0.5085
Maximum	99.96	Maximum	10	Maximum	49.65
Sum	55672.13	Sum	5510	Sum	15379.37
Count	1000	Count	1000	Count	1000