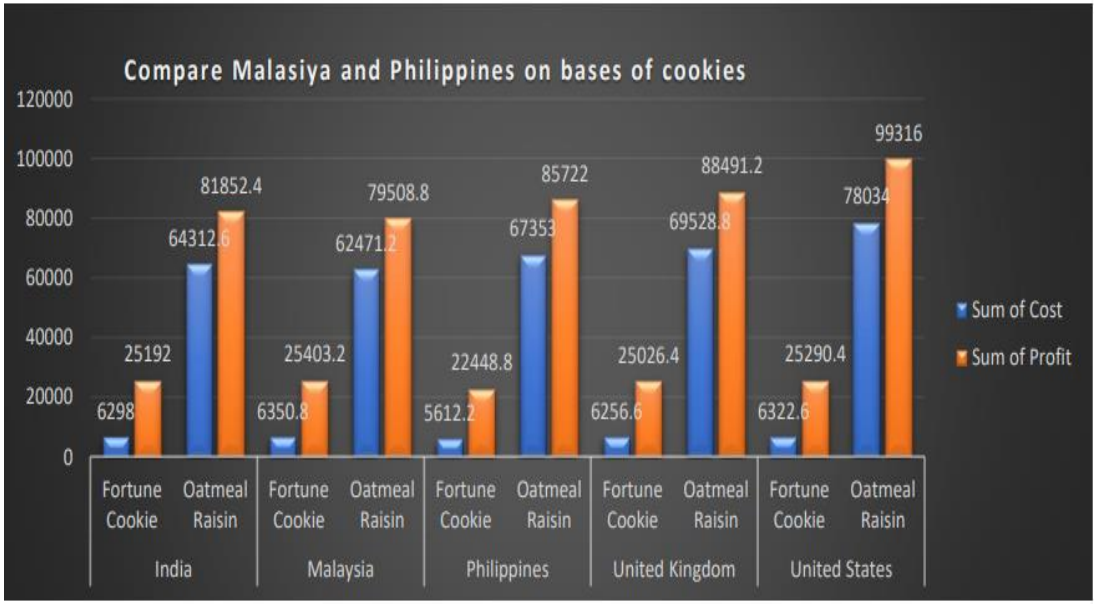
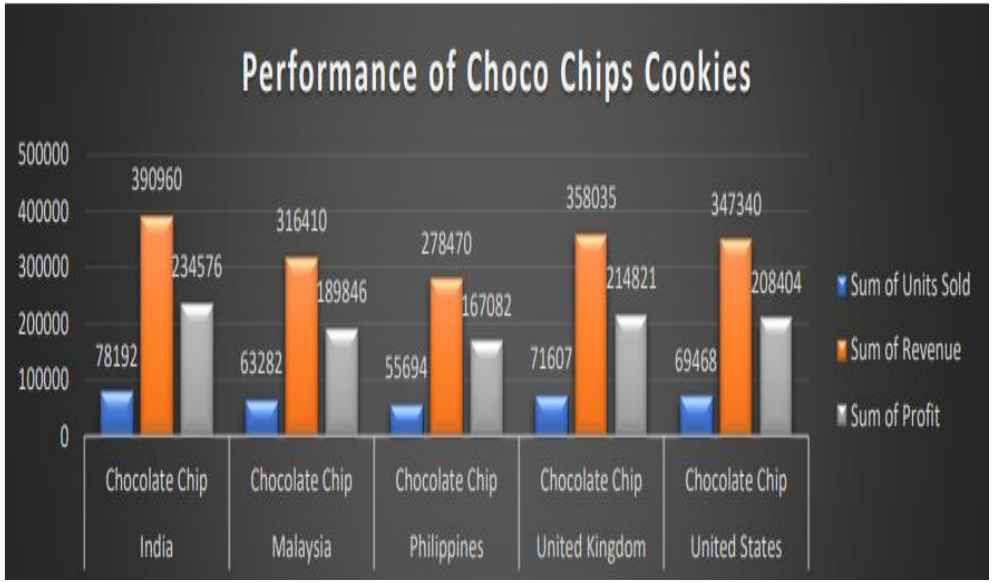
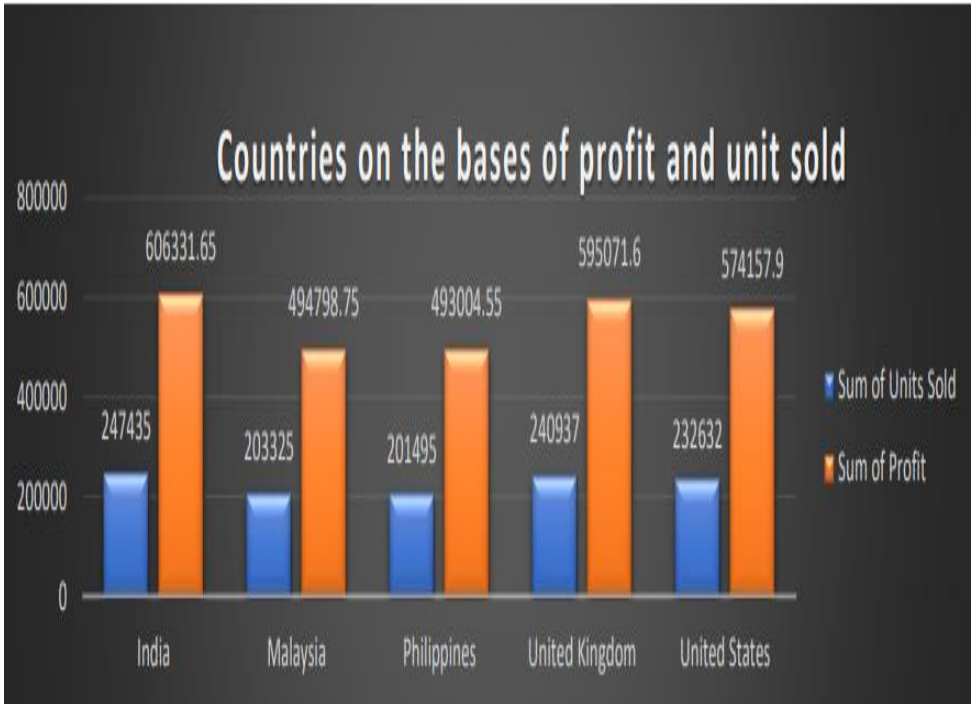
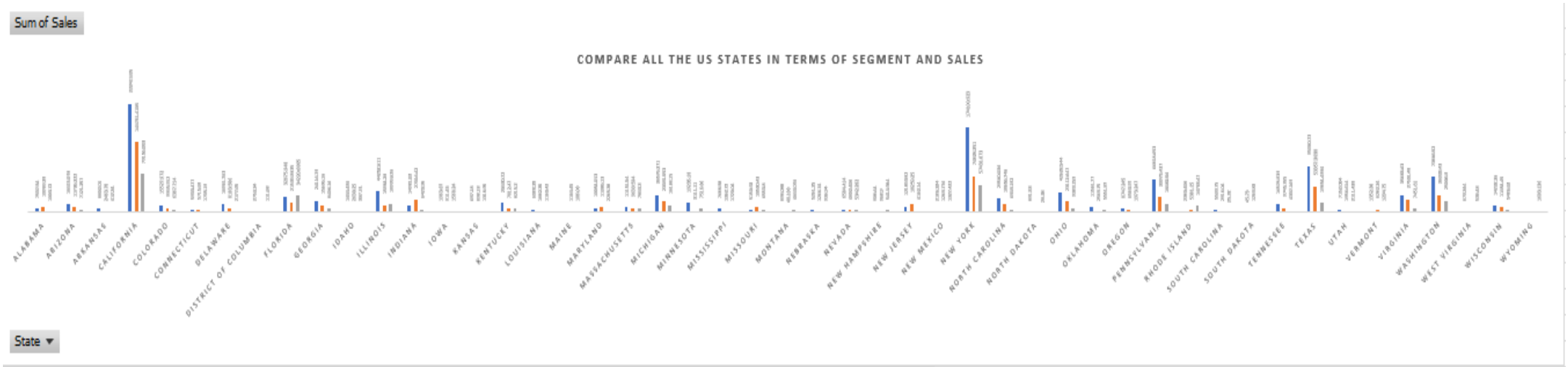
Cookies Dashboard

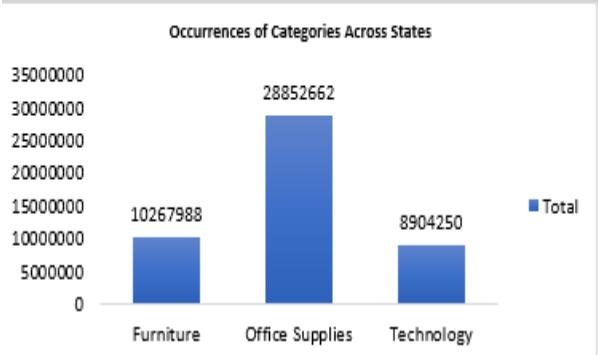


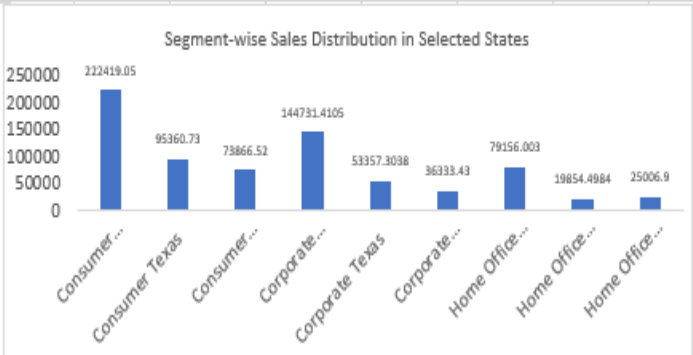


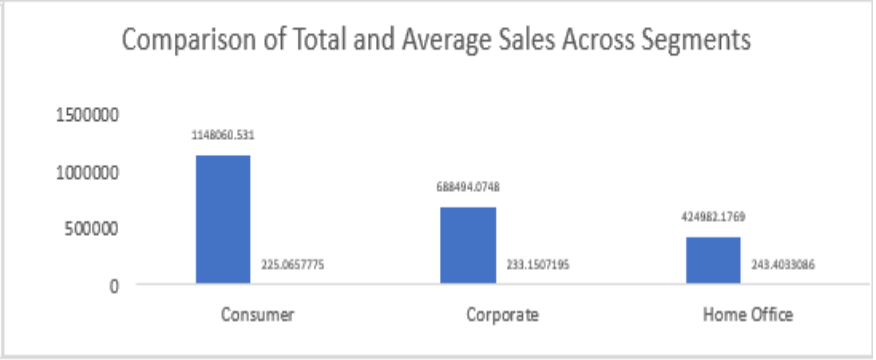


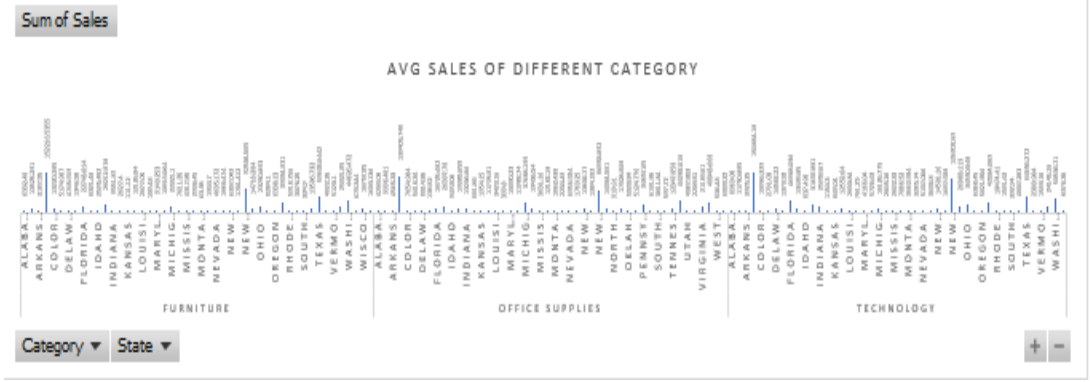
Order Data Dashboard

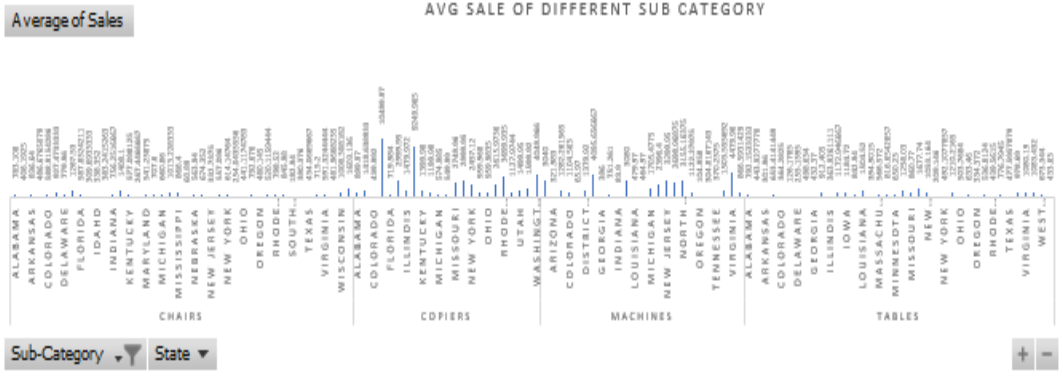




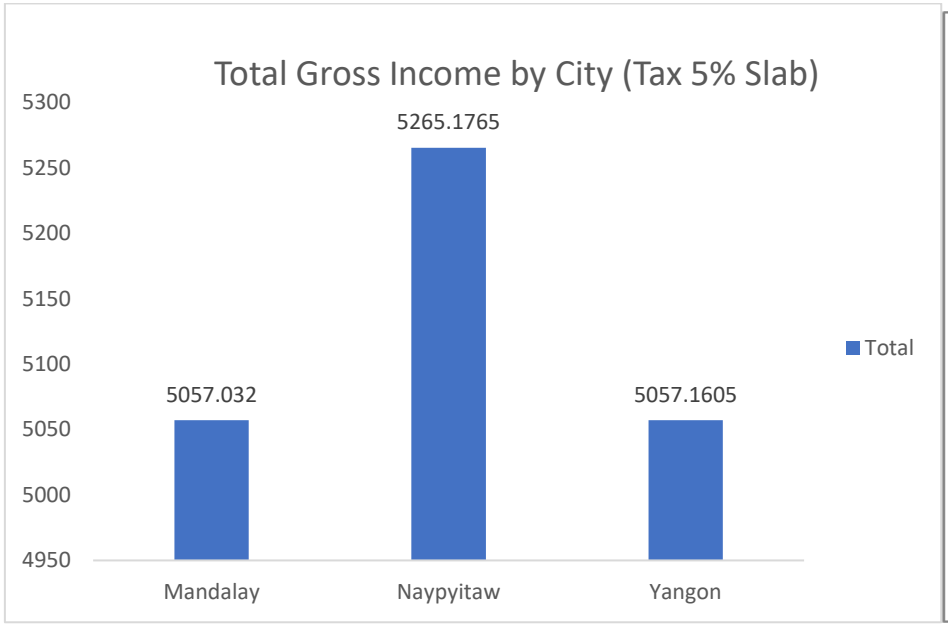


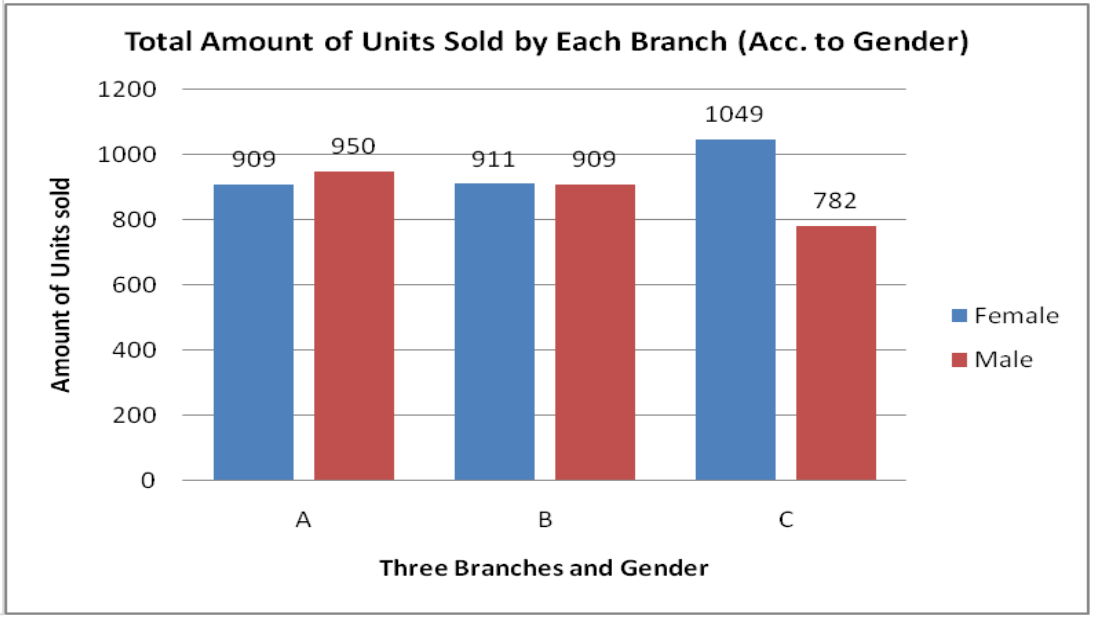


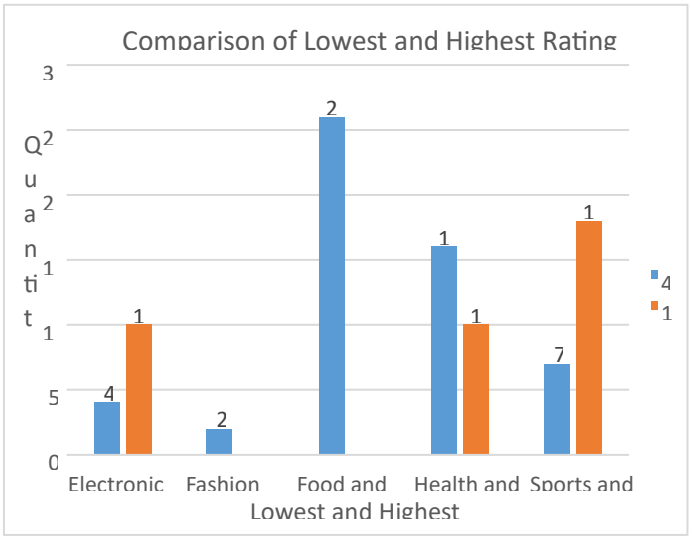


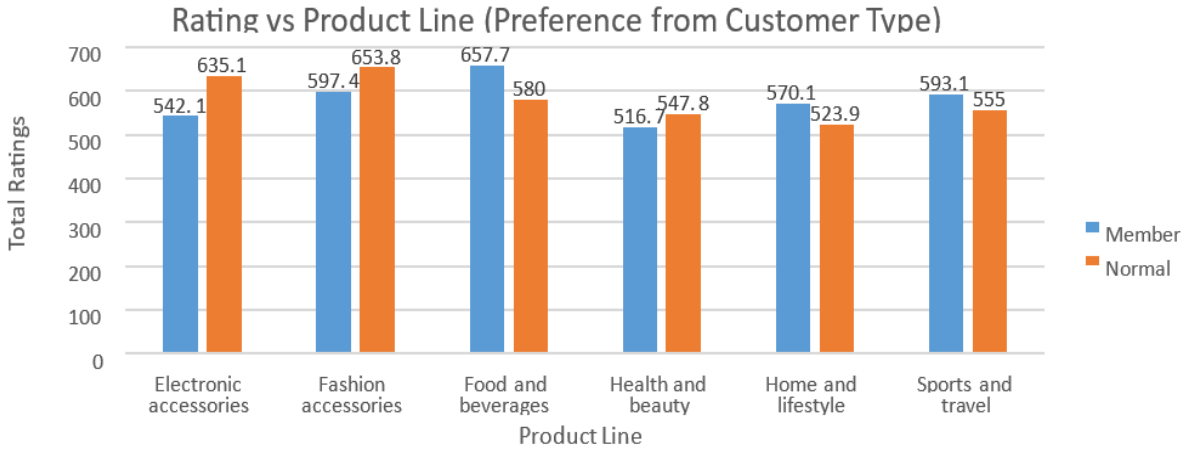


Super Market Dashboard

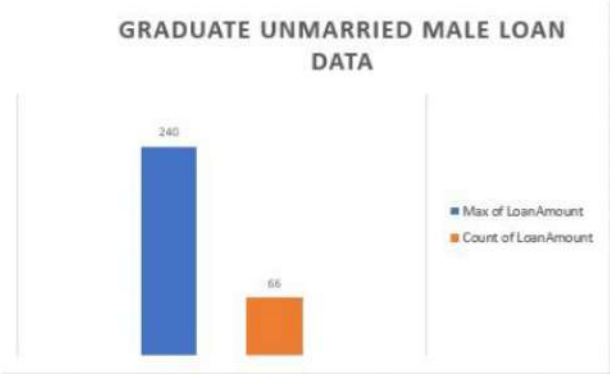


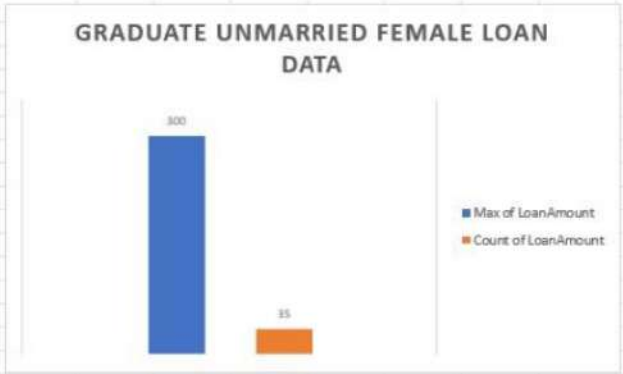


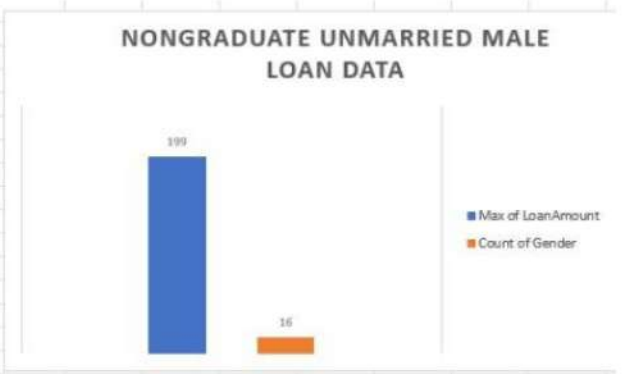


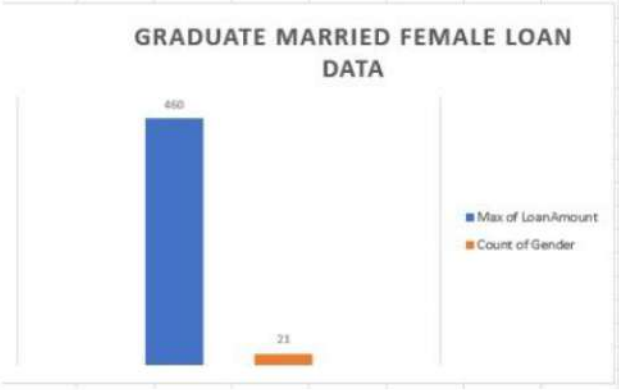


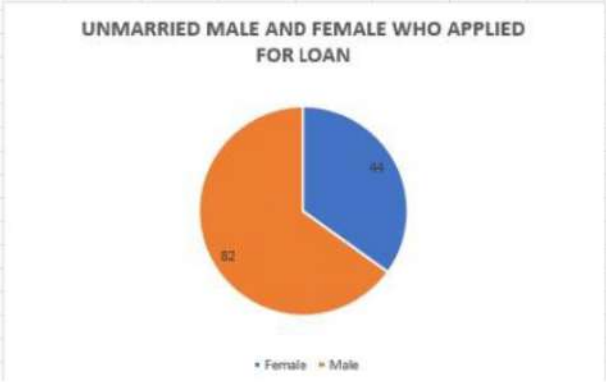
DASHBOARD FOR LOAN DATA ANALYSIS

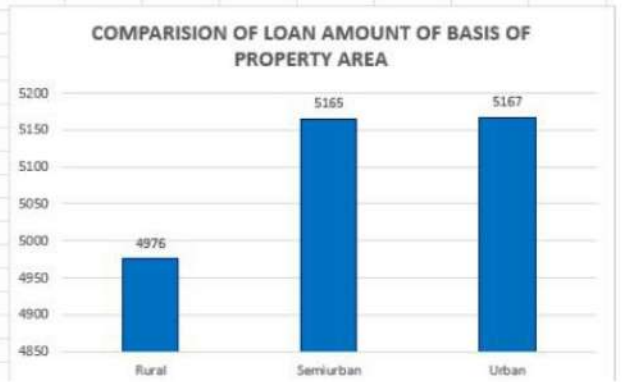




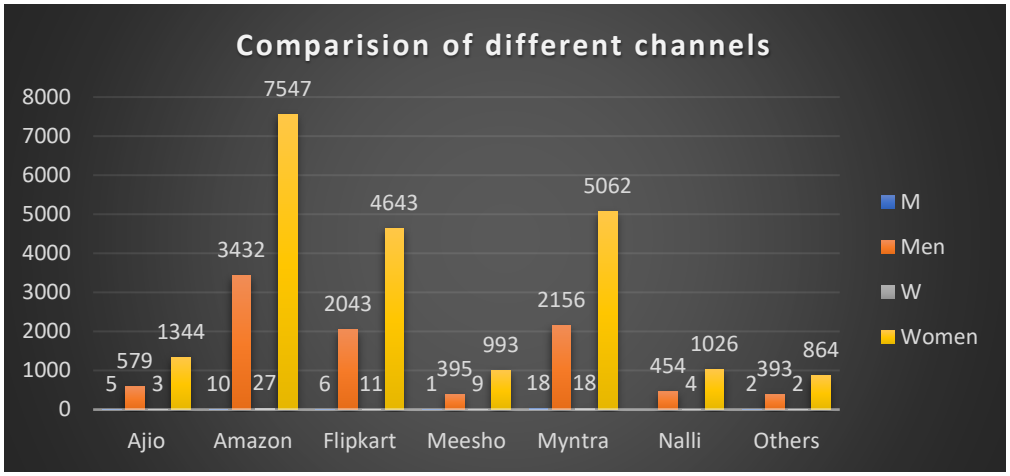


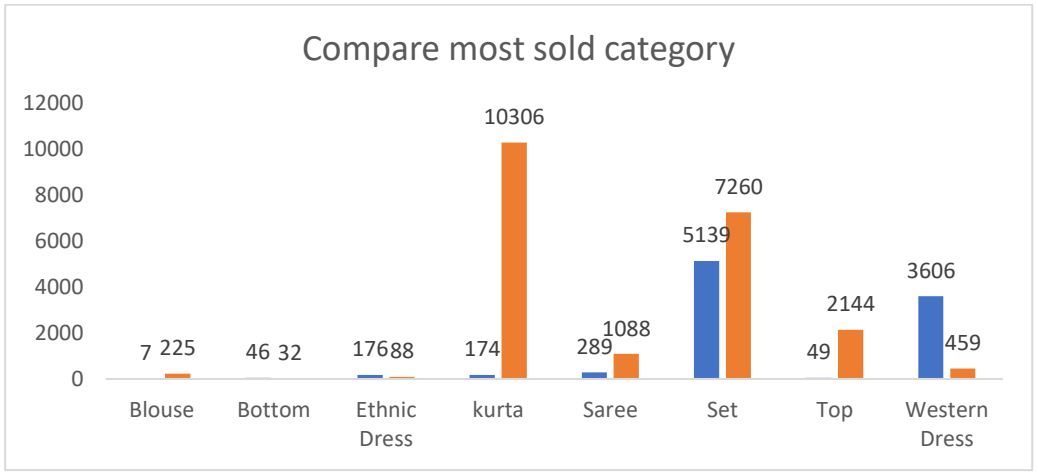


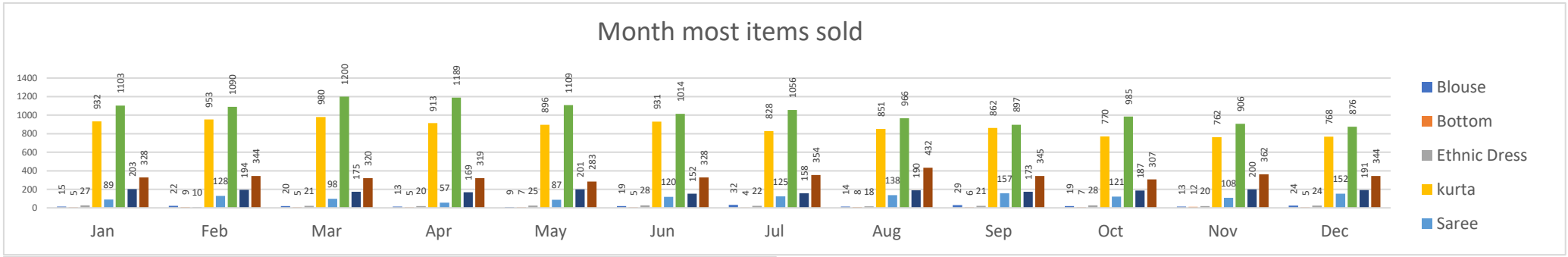


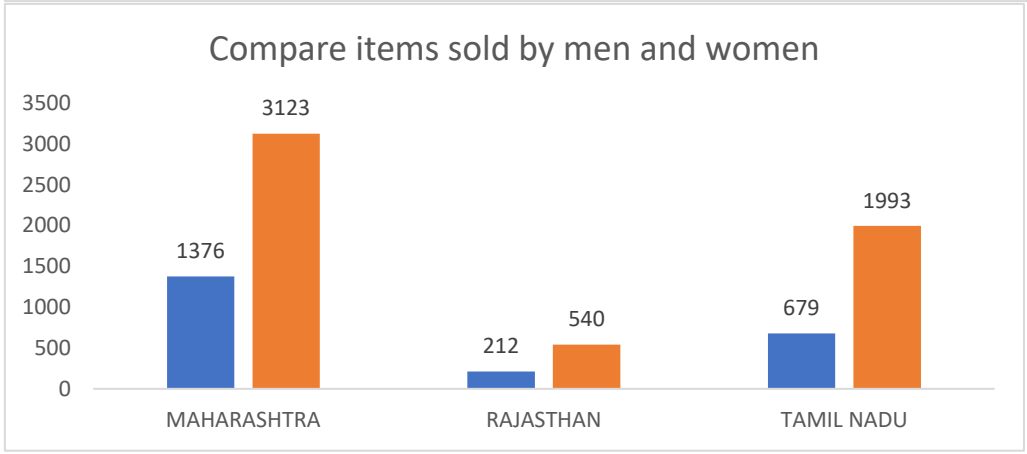


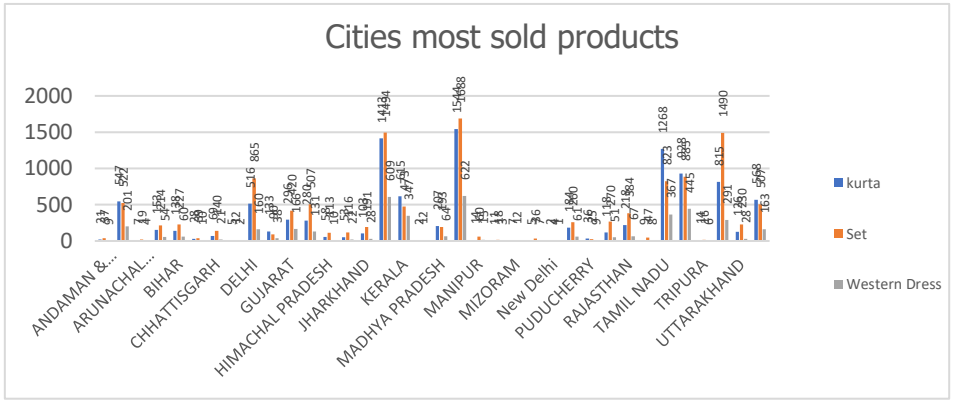
Store Dashboard





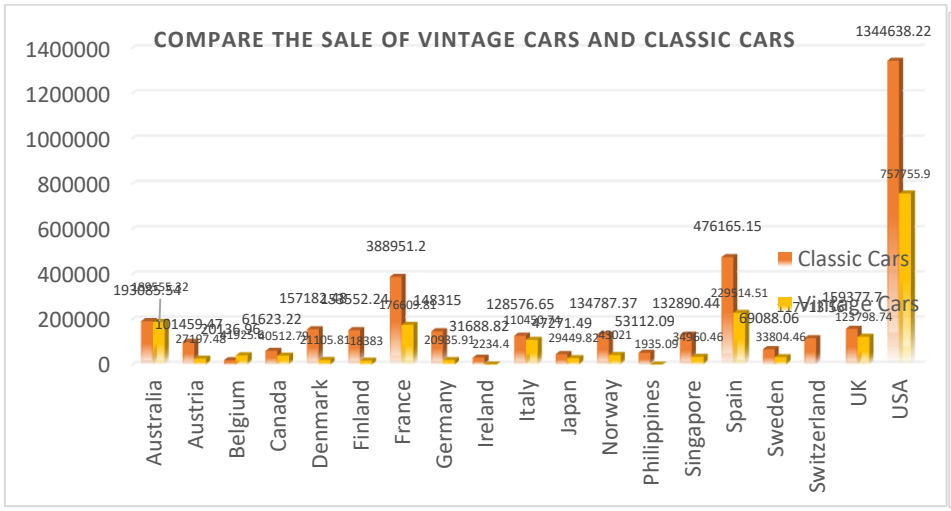


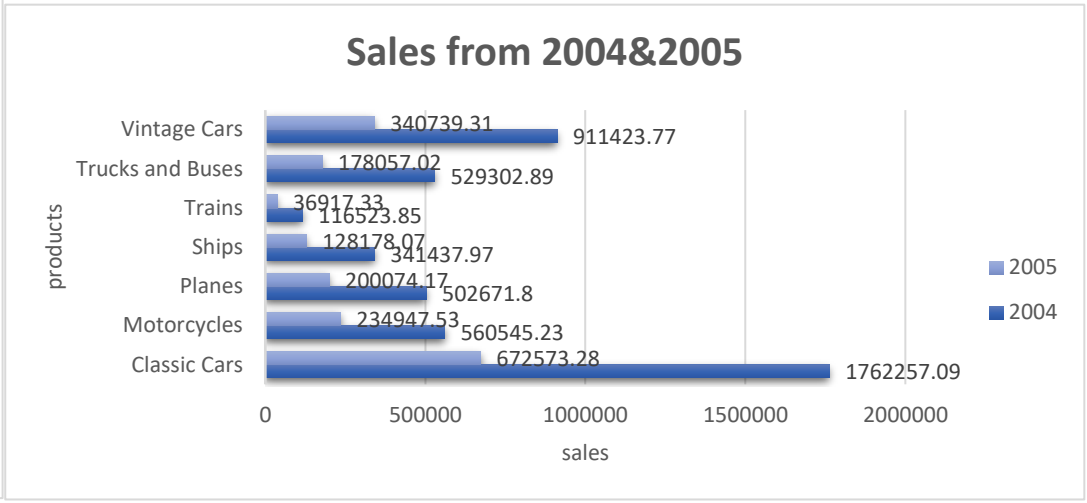


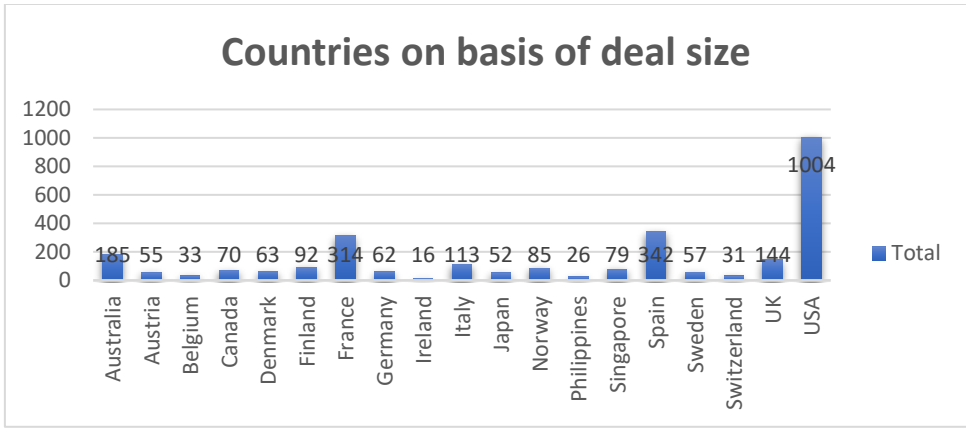


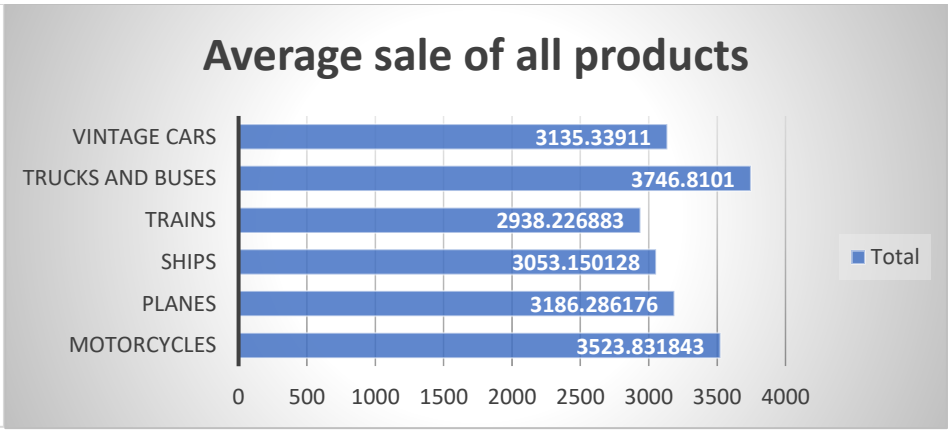
DASHBOARD FOR SHOP SALES DATA ANALYSIS

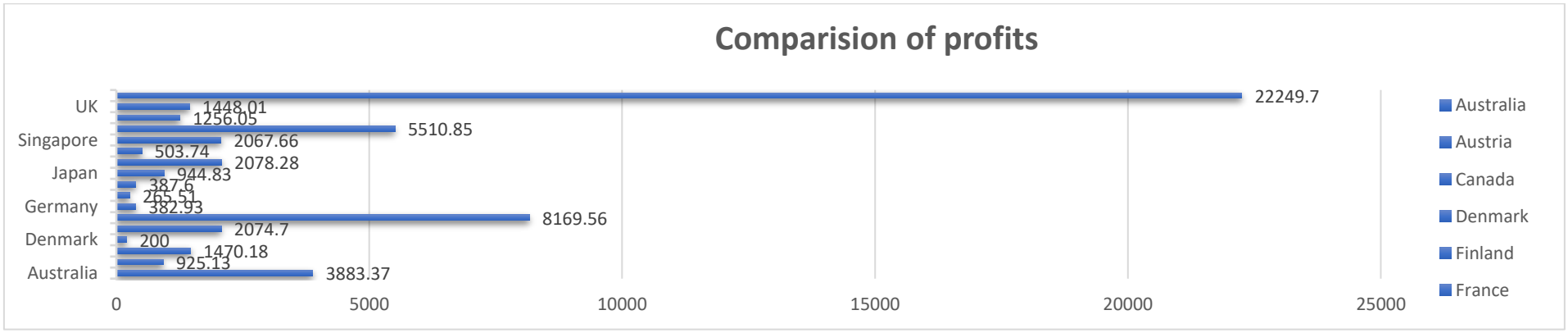
Sales Data Dashboard



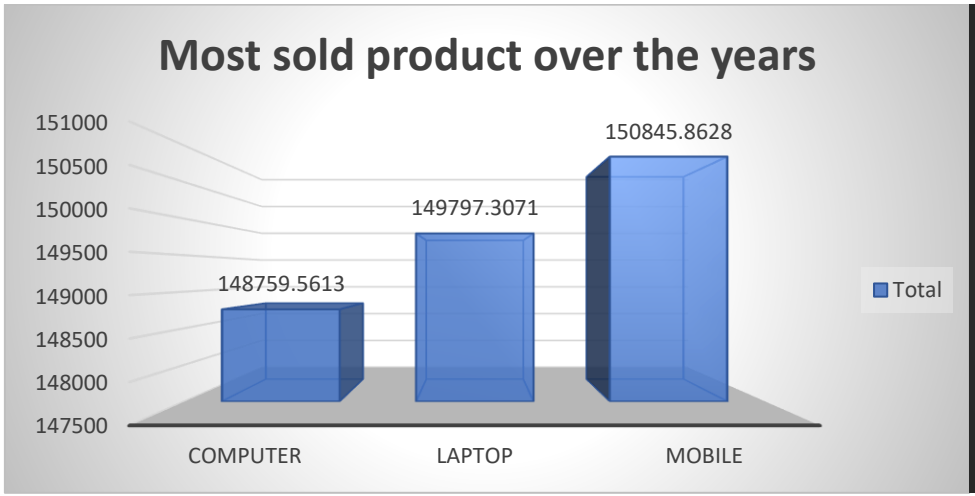


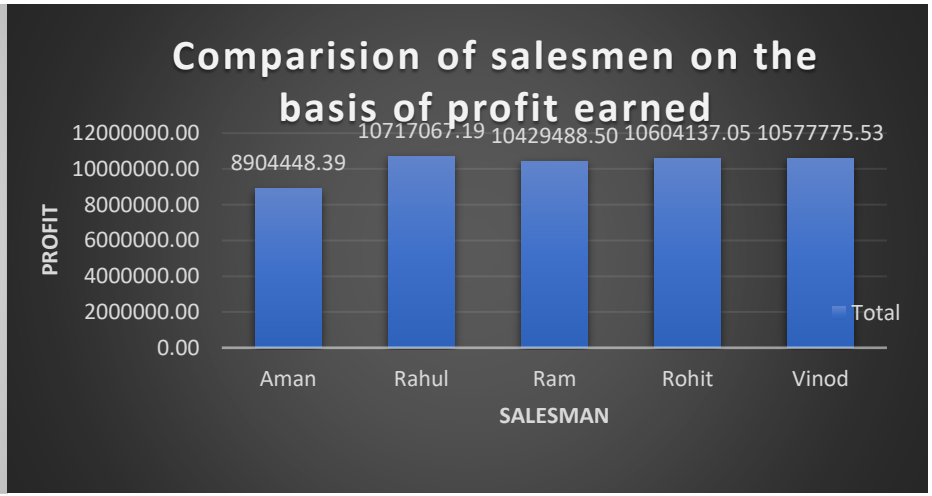


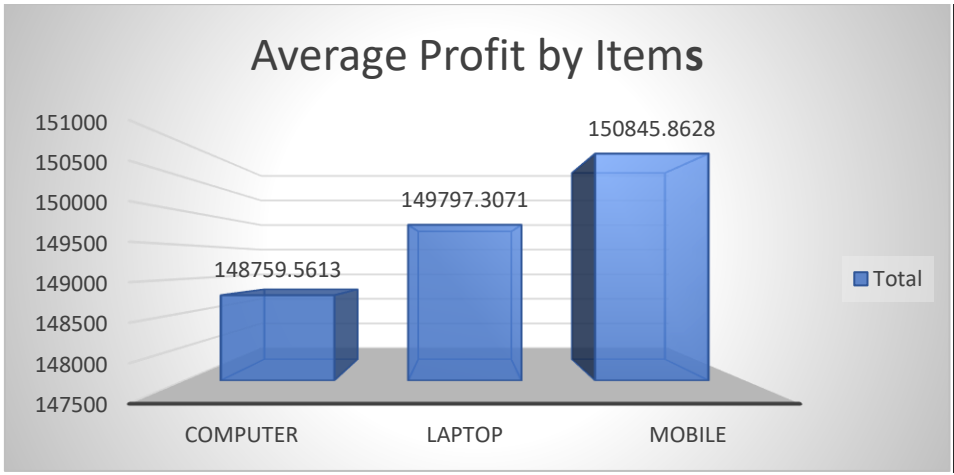


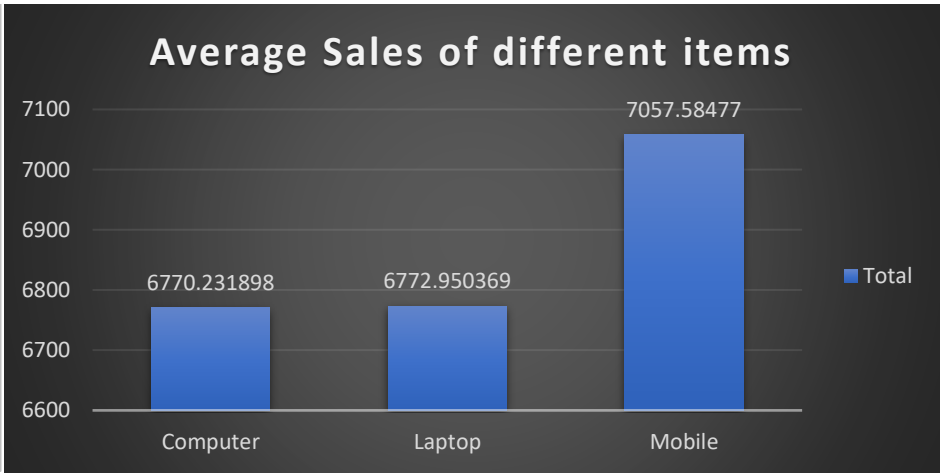


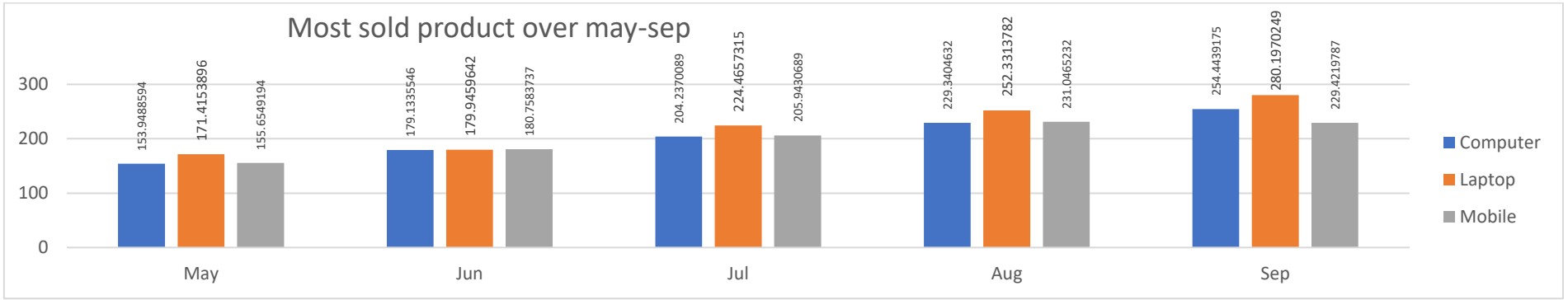
Shop_Data Dashboard











Cookie Data Report

Introduction:

In our cookie data set cookies—specifically six types: Chocolate Chip, Fortune Cookie, Sugar, oatmeal Raisin, Snickerdoodle, and White chocolate macadamia Nut.

We've got a treasure trove of data on these cookies, covering how many units were sold, their costs, the money they brought in (revenue), and the profits they made. And we're not just looking at one place or time; we're exploring different countries and dates to see how things vary.

This report isn't just about cookies; it's about understanding what people like, how much they're willing to pay, and where these treats are most popular. So, get ready to uncover some fascinating 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.

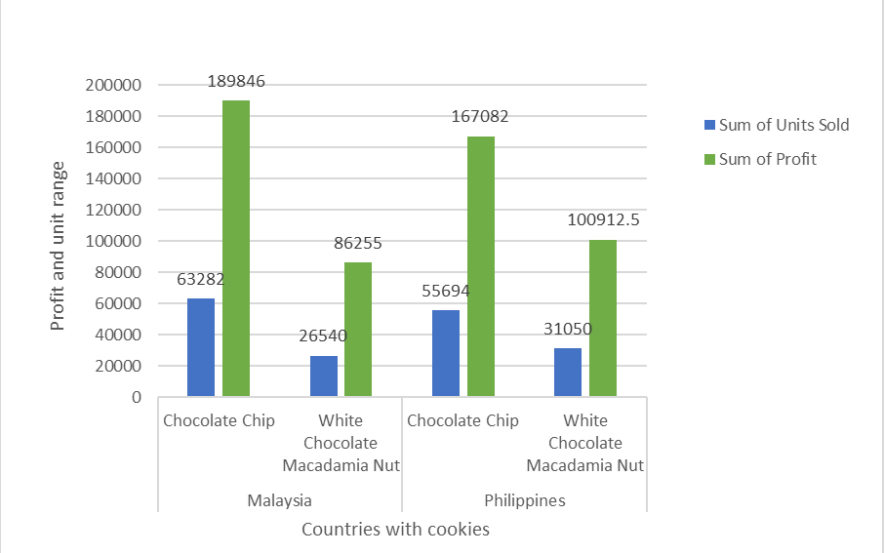
3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

4. which Cookie is the best Selling Cookie in India and US in year 2019,

Analytics :

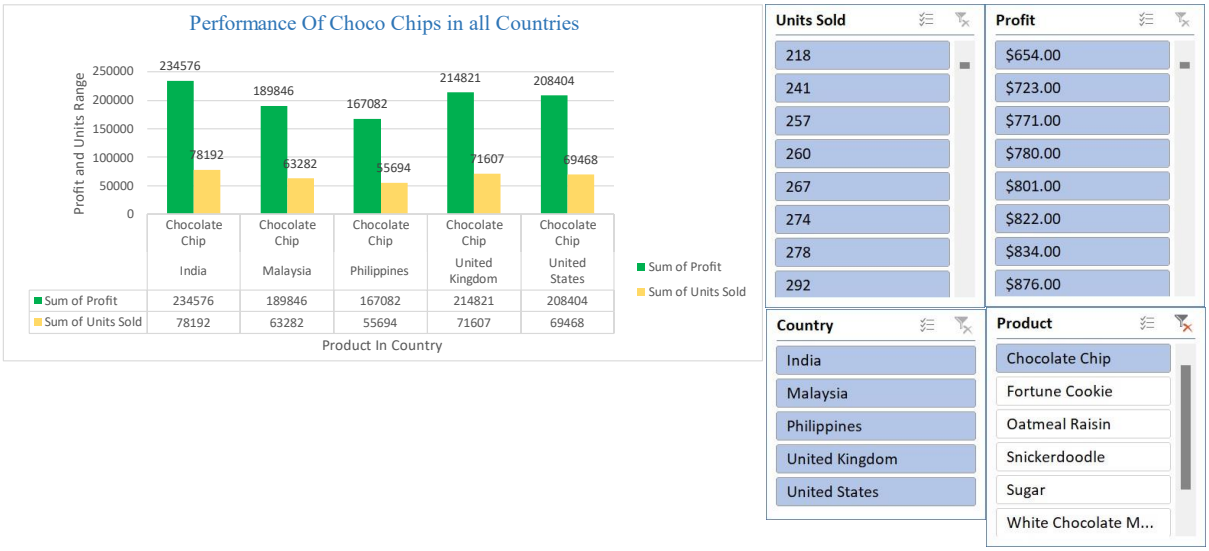
1 . Compare Malaysia and Philippines on the bases of two types of Cookies.

Ans:-The comparsion of Malaysia and Philippines on bases of Chocolate chip and White Chocolate Macadmia nut is given below:-



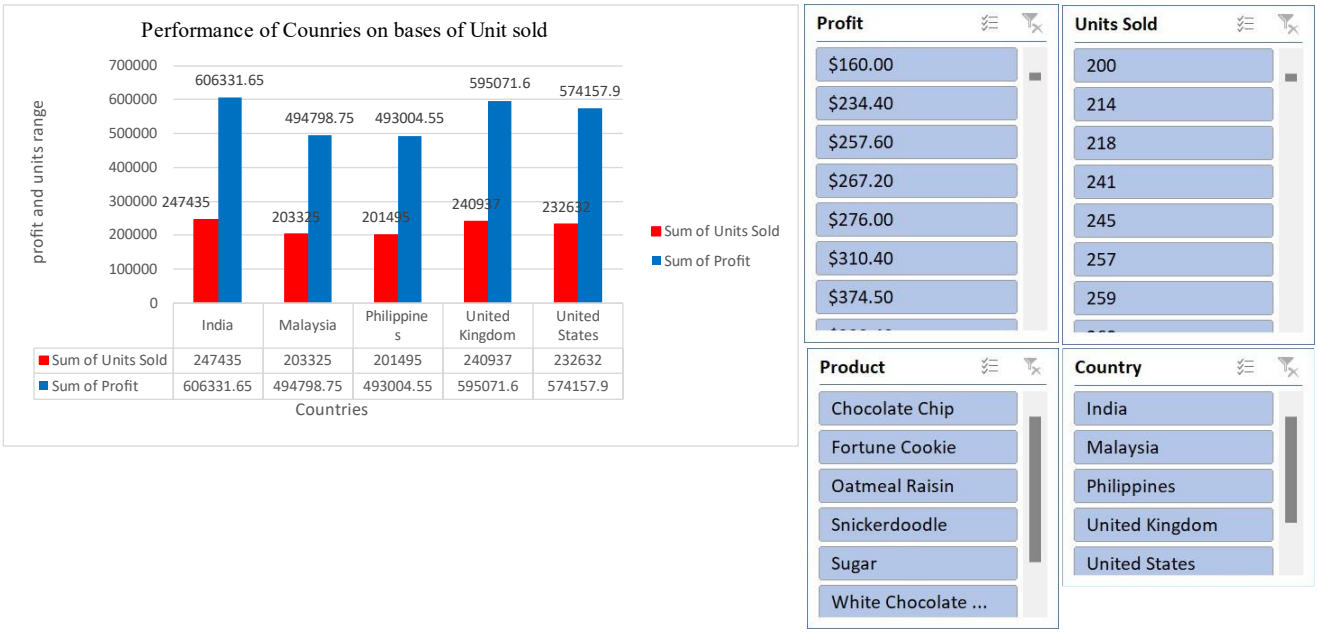
2. What is the performance of Choco Chips Cookies in all Country Which Competes the best.

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.



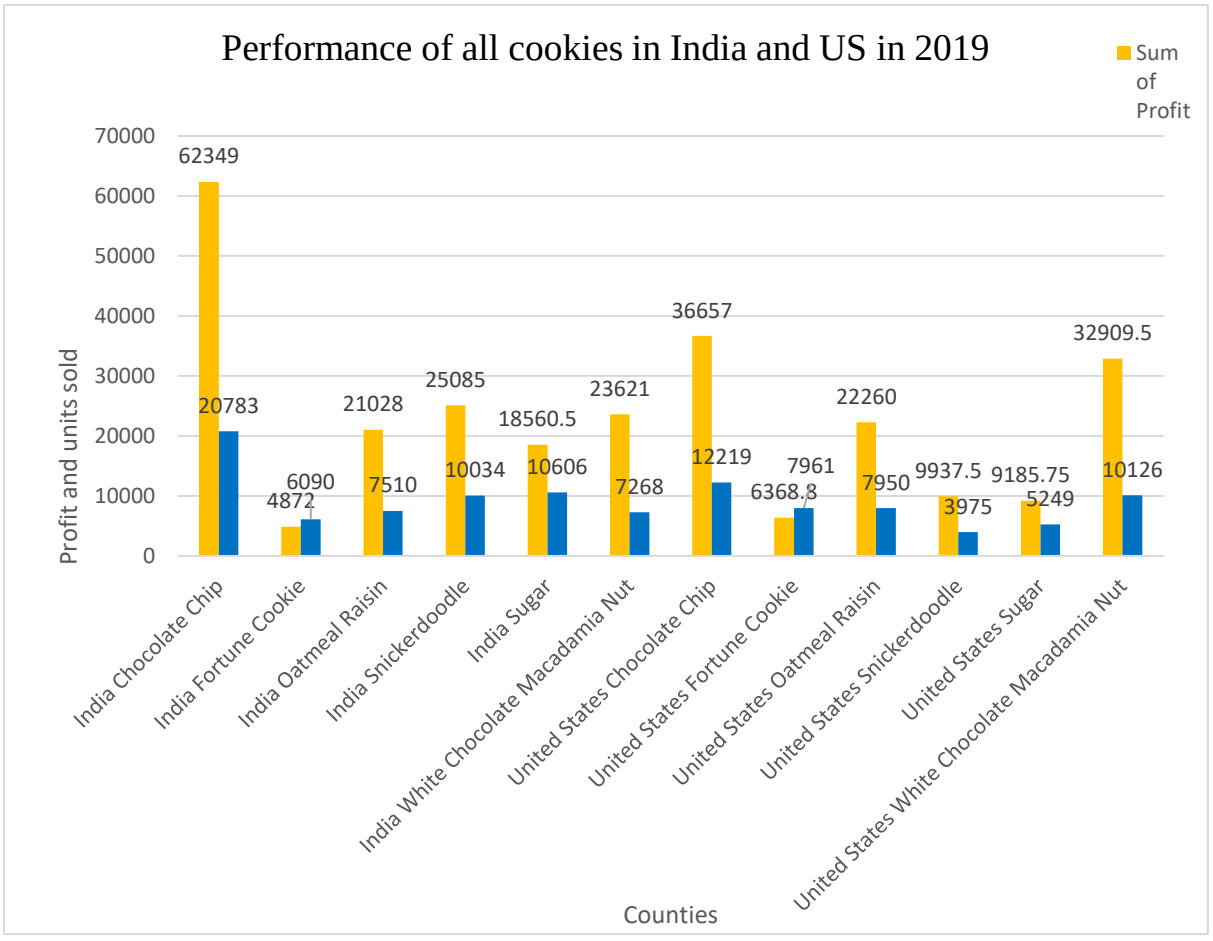
3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

Ans:- 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

Co-relation:

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

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
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.

1725		*8625*		*3450*		*5175*
Mean	1608.153	Mean	6697.702	Mean	2751.795	Mean
Standard Error	32.83303	Standard Error	174.9955	Standard Error	77.09541	Standard Error
Median	1540	Median	5868	Median	2422.2	Median
Mode	727	Mode	8715	Mode	3486	Mode
Standard Deviation	868.0597	Standard Deviation	4626.638	Standard Deviation	2038.295	Standard Deviation
Sample Variance	753527.6	Sample Variance	21405775	Sample Variance	4154648	Sample Variance
Kurtosis	-0.31828	Kurtosis	0.463405	Kurtosis	0.807696	Kurtosis
Skewness	0.436551	Skewness	0.869254	Skewness	0.931429	Skewness
Range	4293	Range	23788	Range	10954.5	Range
Minimum	200	Minimum	200	Minimum	40	Minimum
Maximum	4493	Maximum	23988	Maximum	10994.5	Maximum
Sum	1124099	Sum	4681694	Sum	1923505	Sum
Count	699	Count	699	Count	699	Count
Largest(1)	4493	Largest(1)	23988	Largest(1)	10994.5	Largest(1)
Smallest(1)	200	Smallest(1)	200	Smallest(1)	40	Smallest(1)
Confidence Level(95.0%)	64.46334	Confidence Level(95.0%)	343.5807	Confidence Level(95.0%)	151.3667	Confidence Level(95.0%)

Supermarket Sales Data Report

Introduction:

Dataset Overview:

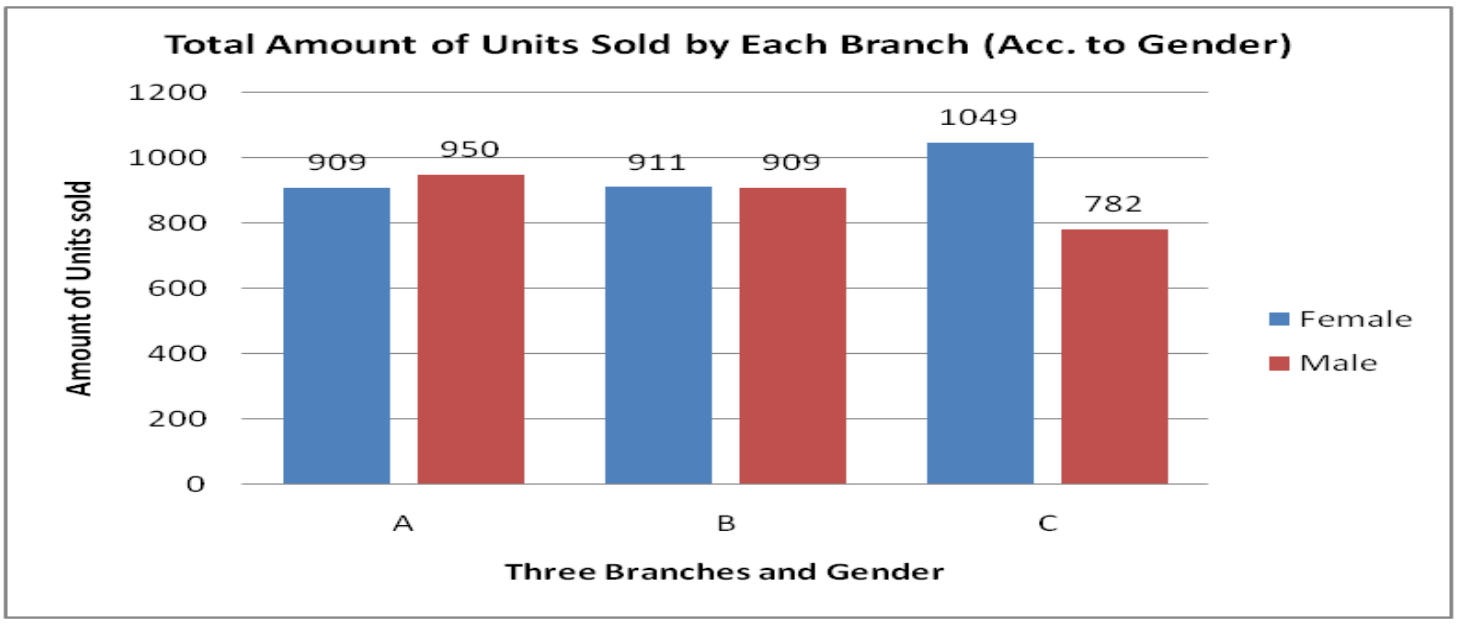
Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of supermarket sales. From fundamental transactional details such as Invoice ID, Date, Time, and Payment Method to more nuanced factors like Branch Location, Customer Type, Gender Demographics, Product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

Key Attributes:

1. Invoice ID: A unique identifier for each sales transaction, facilitating traceability and analysis.
2. Branch (A, B, C): The geographical location of the supermarket branch, allowing for regional comparisons and trend identification.
3. Customer Type (Normal, Member): Distinguishing between regular customers and members, offering insight into loyalty and engagement levels.
4. Gender (Male, Female): Demographic segmentation aiding in understanding purchasing preferences and patterns.
5. Product Line (Fashion Accessories, Electronic Accessories, Food and Beverages, Health and Beauty, Home and Lifestyle, Sports and Travel): Categorization of products facilitating analysis of sales trends across different product categories.
6. Unit Price, Quantity, Tax (5%): Fundamental transactional details crucial for revenue assessment and pricing strategies.
7. Payment Method (Credit Card, Cash, E-wallet): Reflecting evolving payment preferences and trends in consumer behavior.
8. Gross Margin Percentage, Gross Income, COGS: Performance metrics illuminating profitability and operational efficiency.
9. Rating (1 to 10): Customer feedback providing a qualitative assessment of product satisfaction and service quality.
10. City (Yangon, Mandalay, Naypyitaw): Regional segmentation enabling geographical analysis and market segmentation.

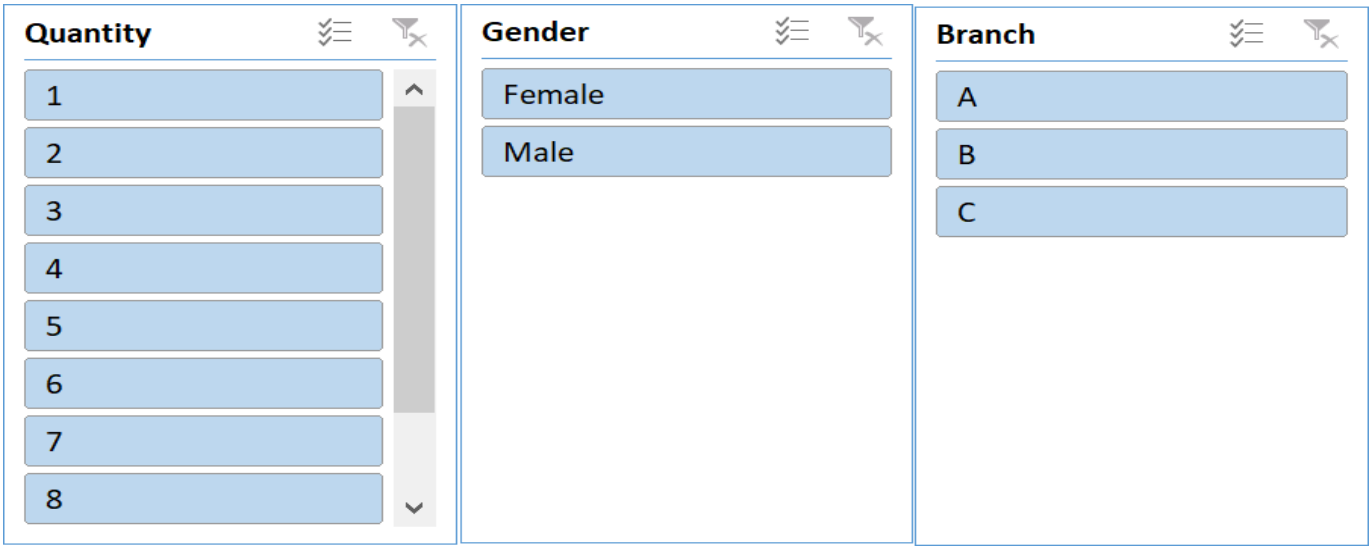
Q2. Which customer gender ordered most items from all the three branches?

Ans.

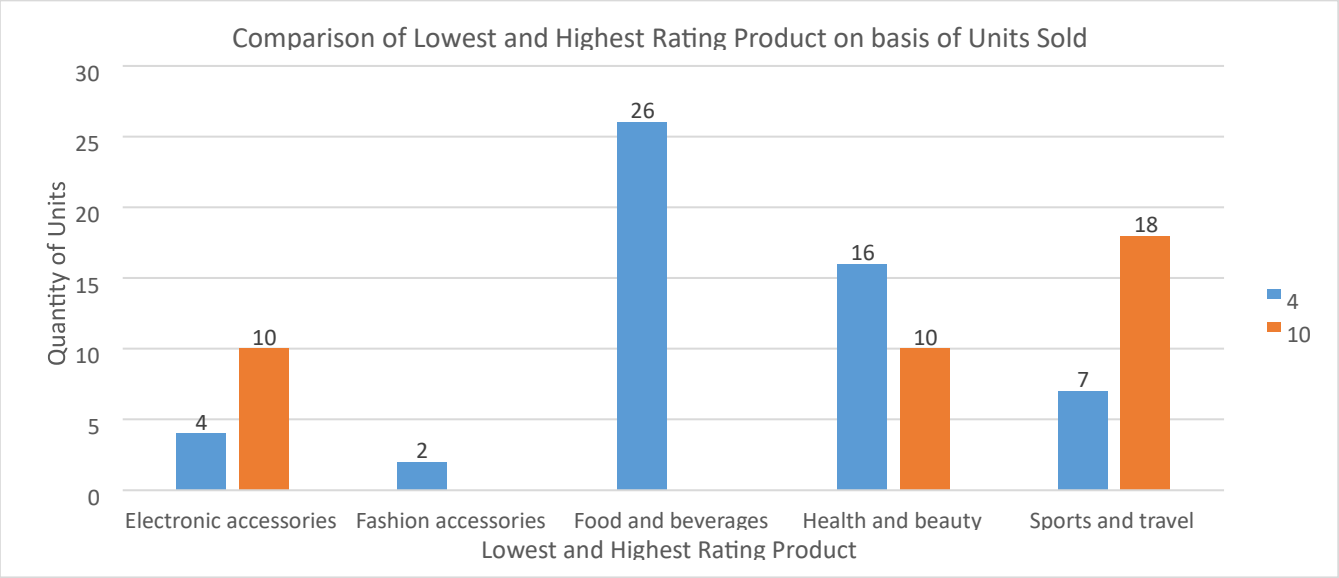


Our analysis of the Supermarket Sales Data revealed the following:

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.

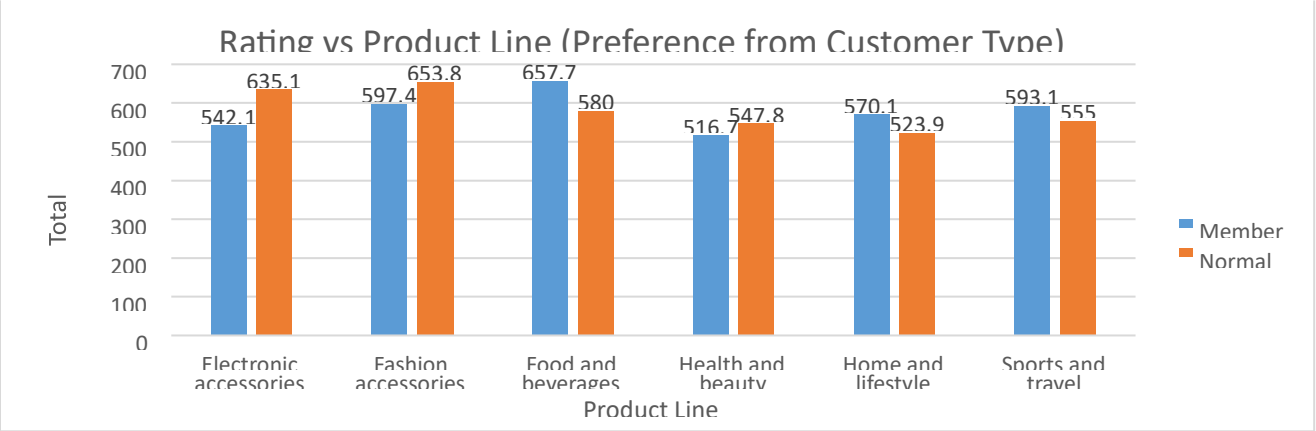
SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.010777564					
R Square	0.000116156					
Adjusted R Square	-0.000885732					
Standard Error	2.924724997					

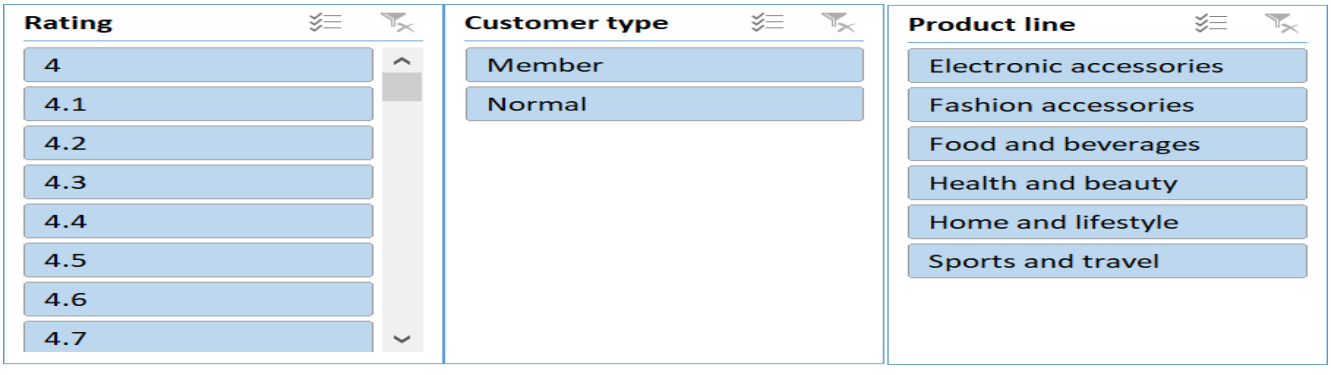
Observations	1000					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	0.9917274	0.991727	0.115937	0.733555221	
Residual	998	8536.908273	8.554016			
Total	999	8537.9				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.443794599	0.215314544	25.28299	2.1E-109	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, and the report provides actionable recommendations for targeted marketing and strategic investment

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:

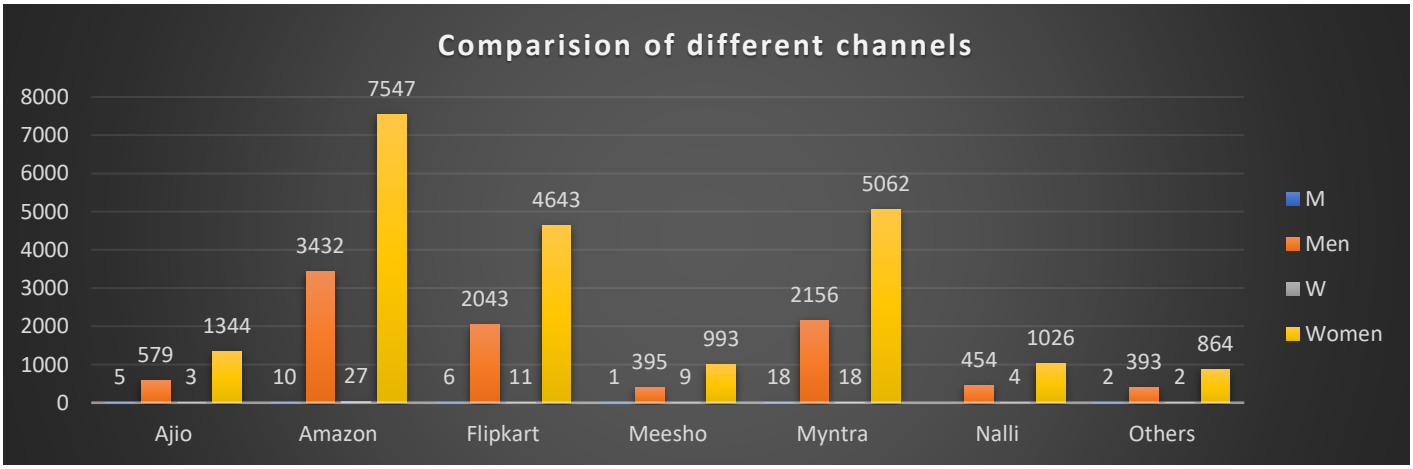
a.kurta b.set c.western wears.

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?

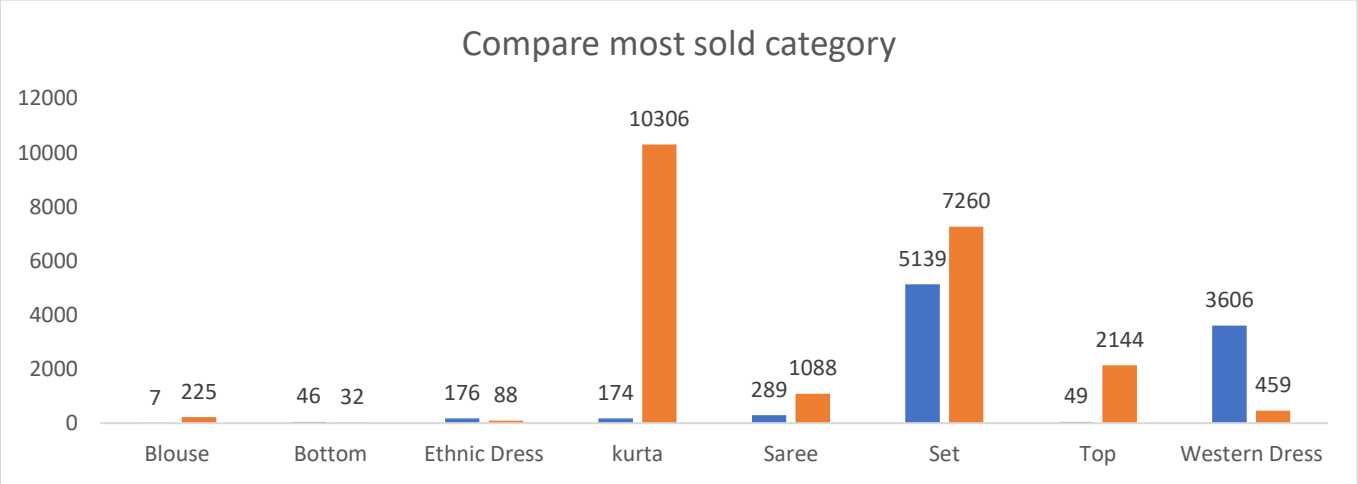
Ans



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

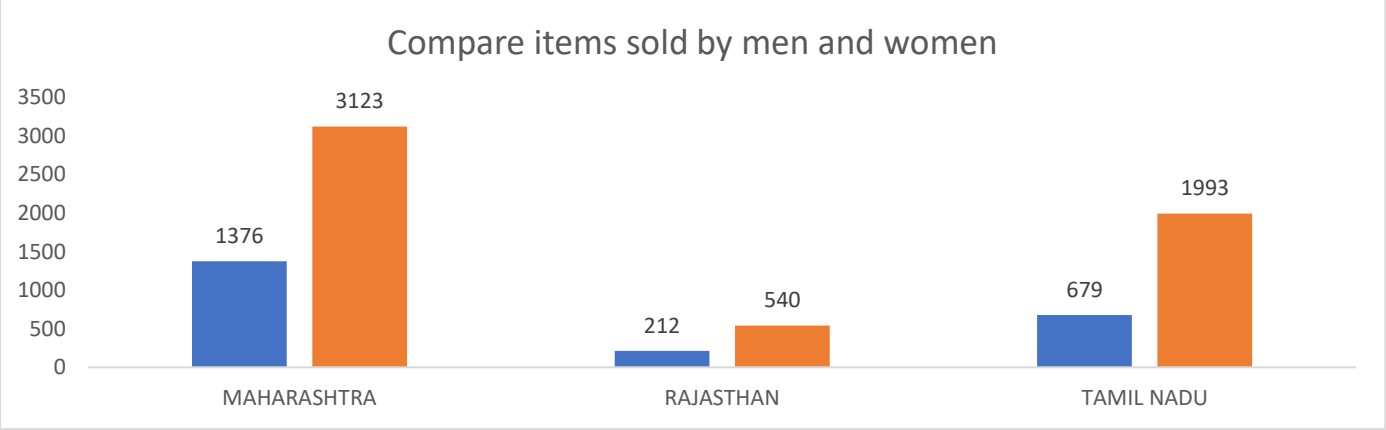
Ans



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:-

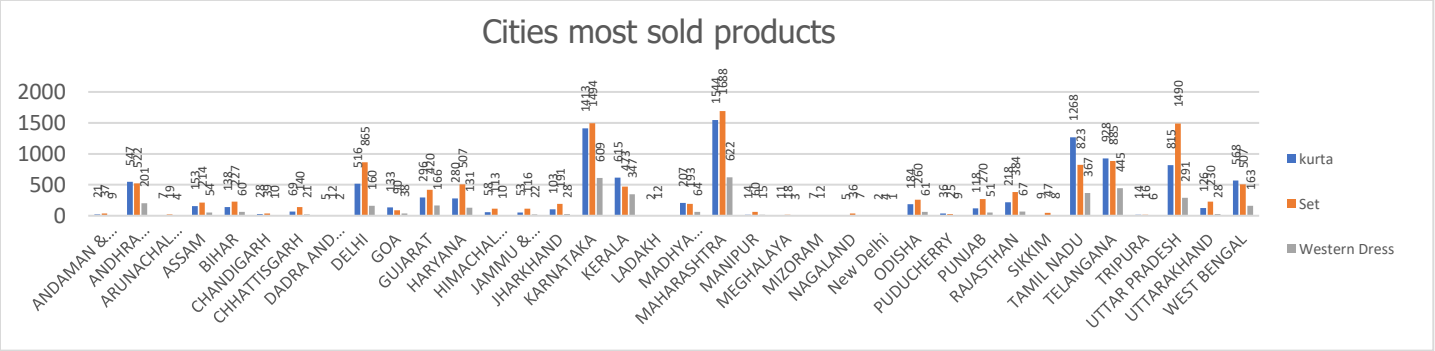


In Maharashtra, sales data indicates that the men's categorysaw a total of 1,390 units sold, while the women's category recorded a significantly higher figure of 3,144 units sold. Moving on to Tamil Nadu, sales in the men's category amounted to 686 units, with the women'scategory showing a stronger performance at 2,023 units sold. Finally, in Rajasthan, sales werecomparatively lower, with only 21 units sold in the men's category and 543 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.

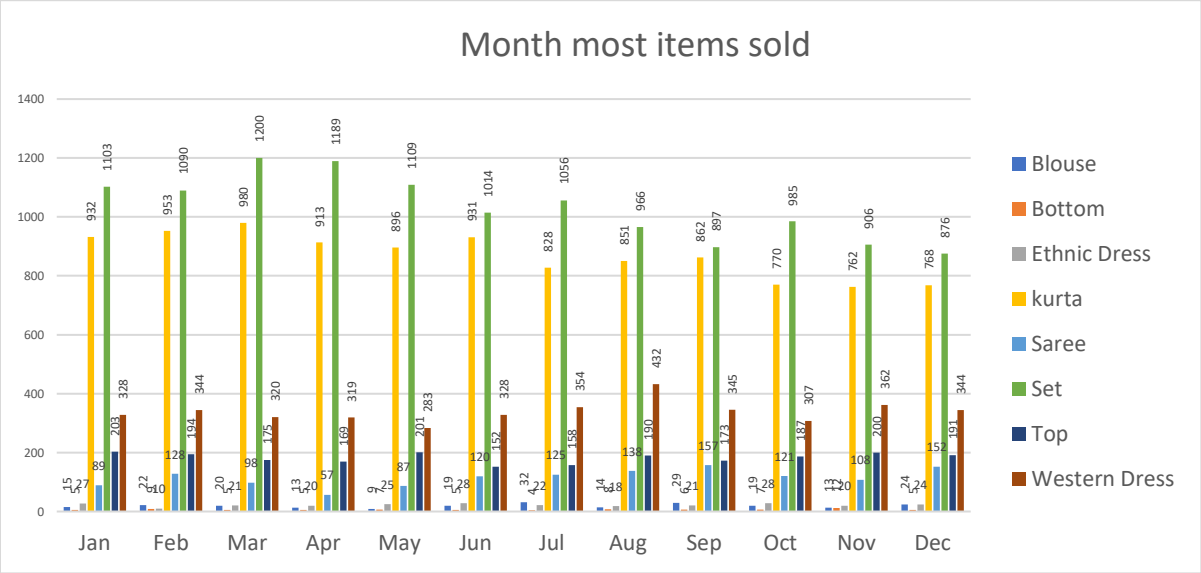
Ans



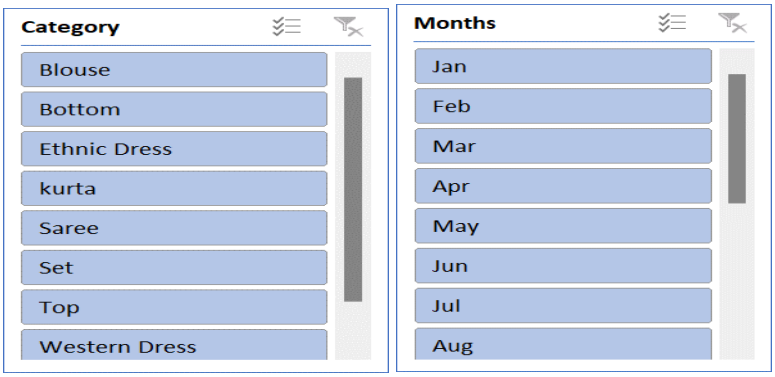
Bengaluru, Chennai, Hyderabad, Mumbai, and New Delhi stand out as the top cities forKurta, Set, and Western wear sales. These urban centers consistently show the highest demandfor these clothing categories compared to other cities, indicating strong consumer preferencesfor traditional and contemporary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

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

Ans.



The most sold items in the month is shown above in the graph . Set is the most sold item in all the months.



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.

Exploring Car Data Report

Introduction:

Dataset Overview:

This dataset comprises a blend of categorical and numerical data, each offering unique perspectives on the industry. Categorical data, such as make, model, and color, encapsulates the diversity of vehicles and consumer preferences. Meanwhile, numerical attributes like mileage, price, and cost provide quantifiable metrics essential for analyzing market trends and pricing dynamics.

Key Attributes:

1. Make: This attribute denotes the brand or manufacturer of the vehicle, offering insights into brand preferences and market share.
2. Model: The specific model of the car, providing granularity in understanding consumer choices and preferences within each brand.
3. Color: Reflects the color of the vehicle, which can influence consumer perception and aesthetic preferences.
4. Mileage: Indicates the distance traveled by the vehicle, a crucial factor influencing its value and pricing.
5. Price: Represents the listed price of the vehicle, serving as a key determinant in consumer purchasing decisions and market competitiveness.
6. Cost: Denotes the cost associated with acquiring the vehicle, which includes factors such as production costs, dealer margins, and other expenses.

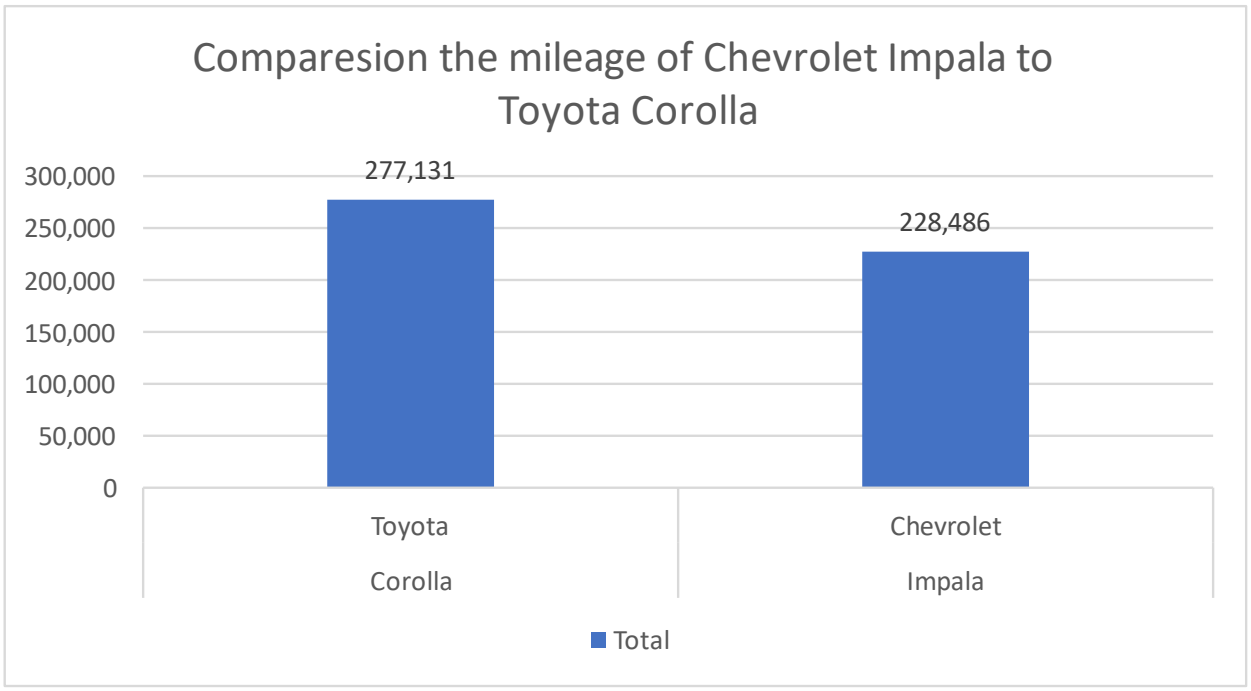
Questionnaire:

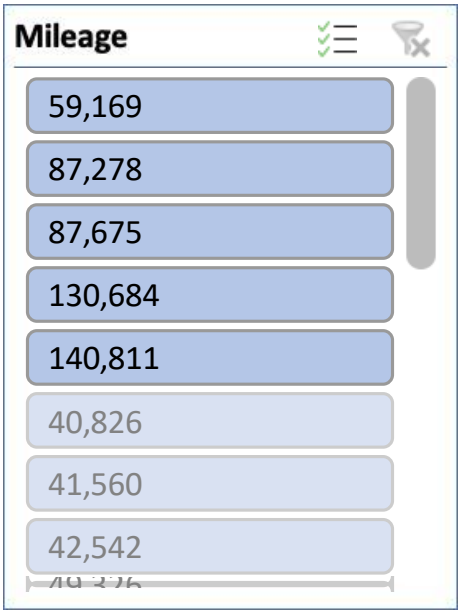
Q1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?
Q2. Justify, Buying of any Ford car is better than Honda
Q3. Among all the cars which car color is the most popular and is least popular?
Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.
Q5. Find out all the cars, and their total cost which is more than $2000?

Analytics:

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

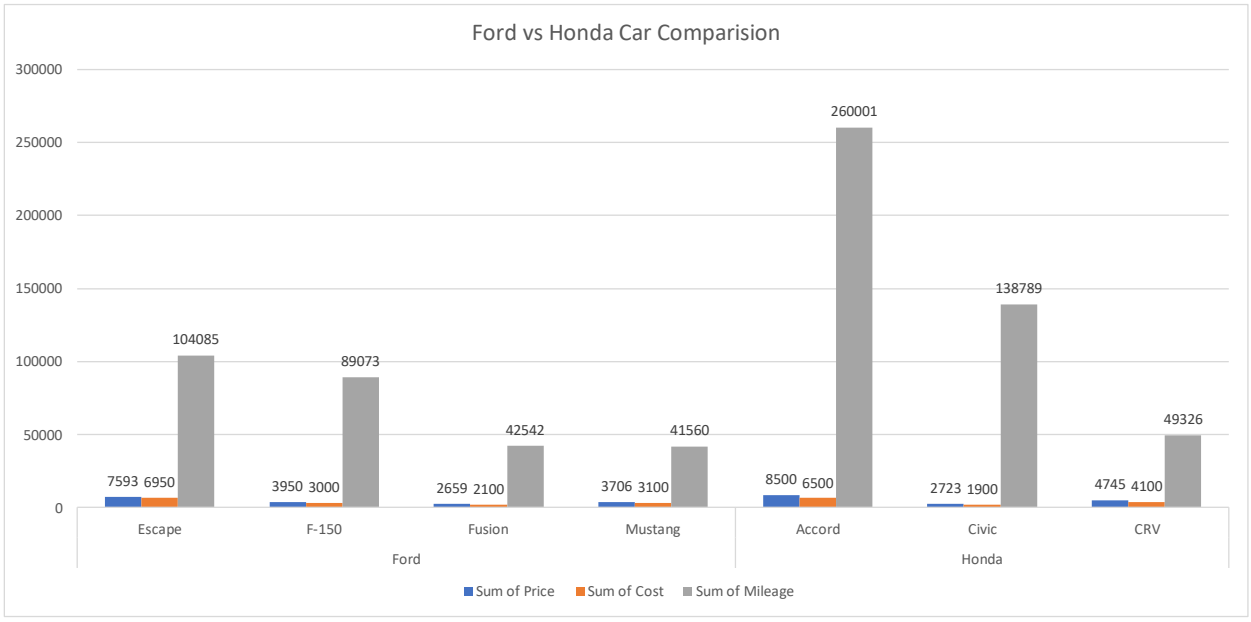
Ans. Toyota Corolla gives better mileage than Chevrolet Impala.

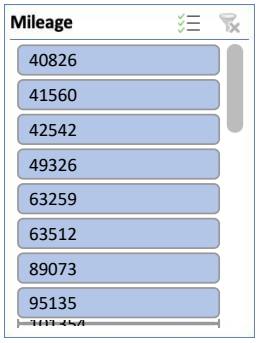


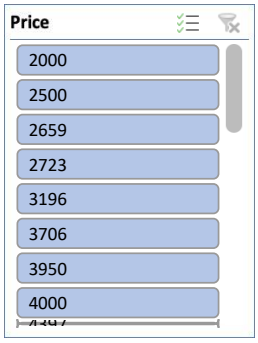


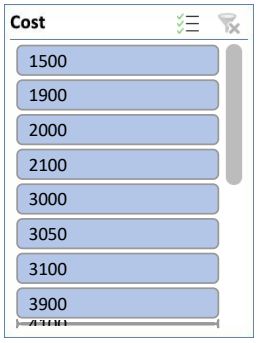
Q2. Justify, Buying of any Ford car is better than Honda.

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.



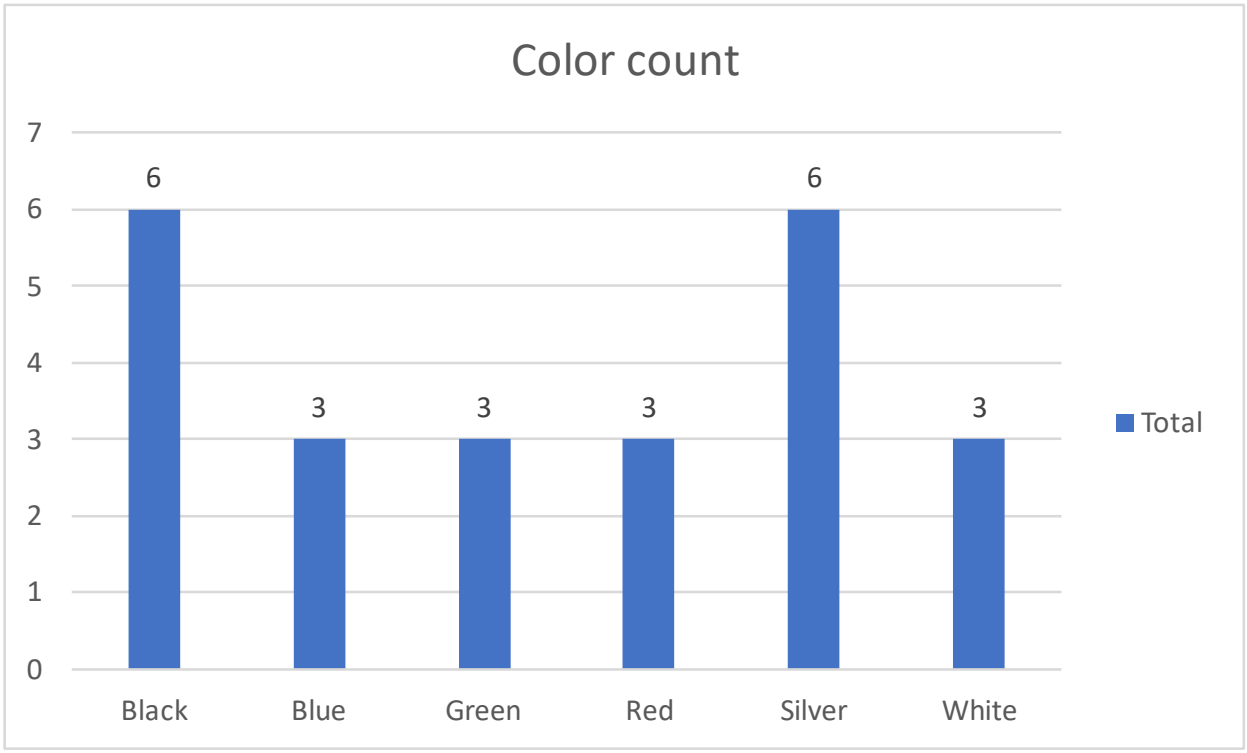


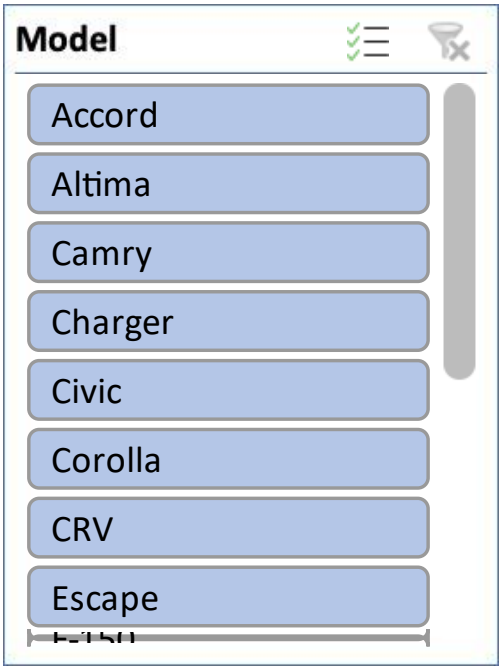


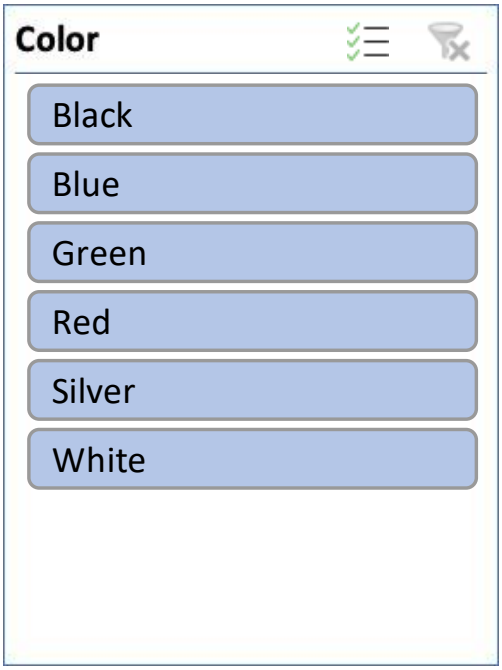


Q3. Among all the cars which car color is the most popular and is least popular?

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.

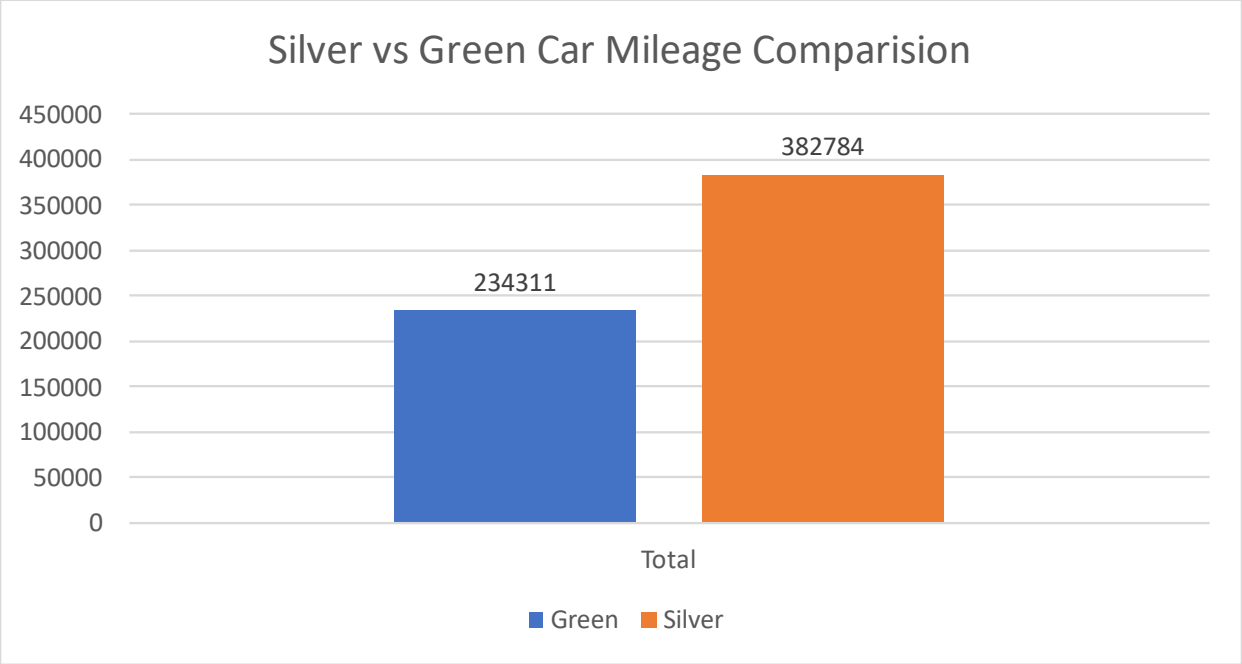


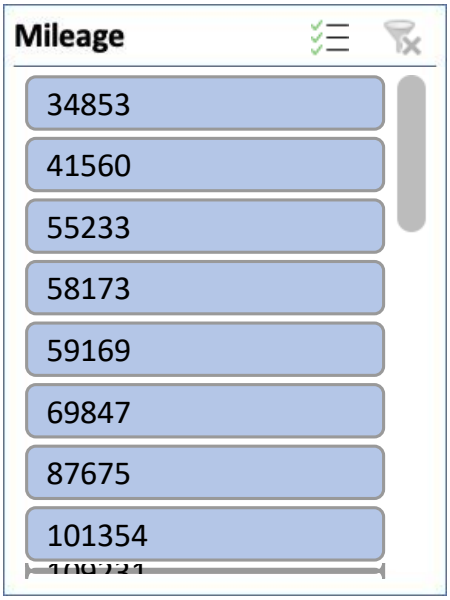




Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.

Ans. Silver color car millage is more than green color car milage if we compare there average.

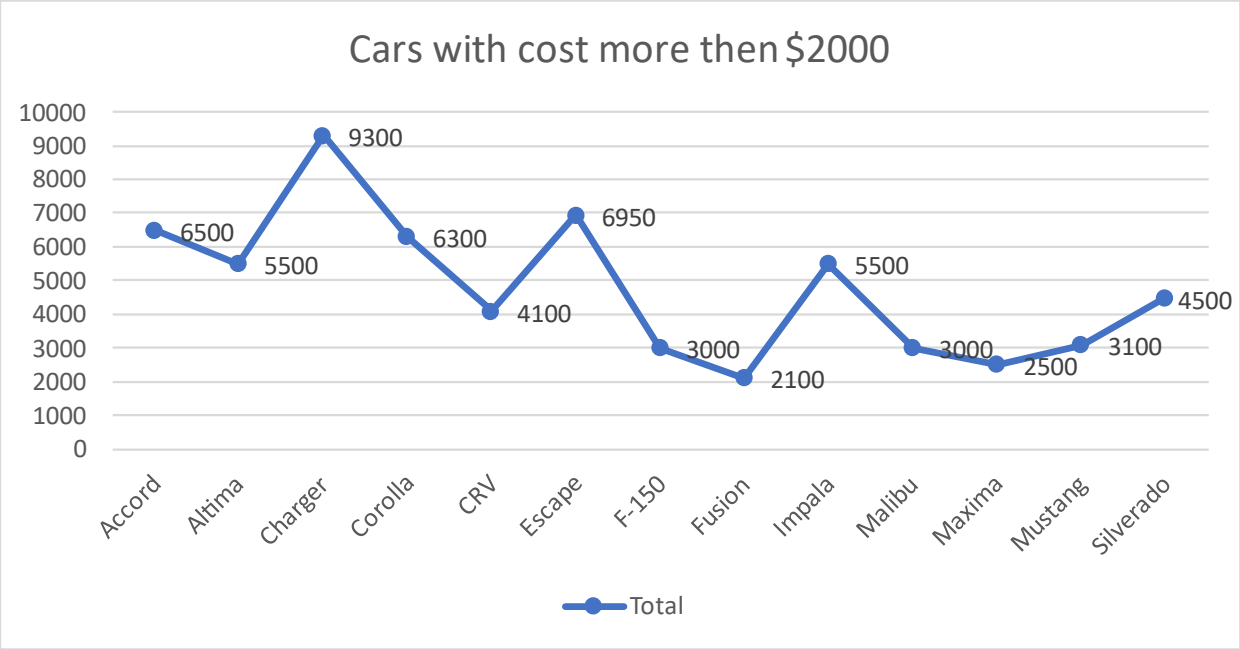


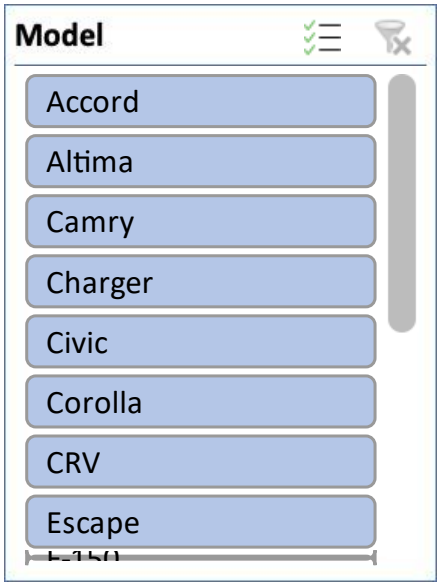


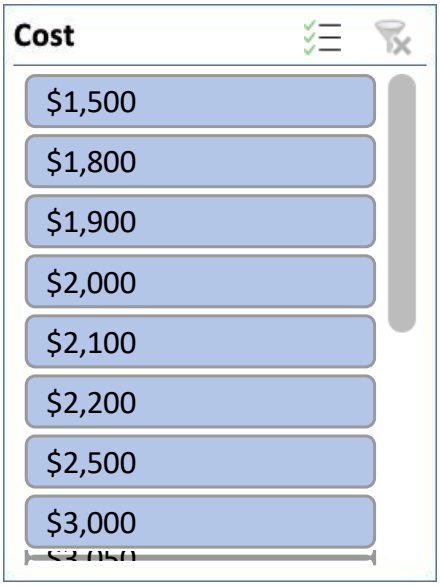
Q5. Find out all the cars, and their total cost which is more than $2000?

Ans. All the car mention below cost is more than $2000

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







Regression

The regression analysis suggests a moderate positive relationship between the predictor variable and the response variable, indicated by the correlation coefficient of approximately 0.40. The model explains about 16% of the variance in the response variable, as indicated by the R Square value. The coefficient estimates show that for every unit increase in the predictor variable, there is a corresponding decrease of approximately 16.66 in the response variable, with a p-value of 0.056, indicating a marginally significant effect.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.40404555
R Square	0.1632528
Adjusted R Square	0.1234077
Standard Error	33099.5397
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4488793099	4488793099	4.09718598	0.05586127
Residual	21	2.3007E+10	1095579531		
Total	22	2.7496E+10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	130438.919	23634.1932	5.51907645	1.7789E-05	81288.9236	179588.914	81288.9236	179588.914
3000	-16.664135	8.23265547	-2.0241507	0.05586127	-33.784879	0.45660911	-33.784879	0.45660911

Co-relational

The correlation matrix indicates a moderate negative correlation (-0.411) between Mileage and Price. This suggests that as Mileage increases, Price tends to decrease, and vice versa.

	Mileage	*Price*
Mileage	1	
Price	-0.4110586	1

Anova: Single Factor

The ANOVA results indicate significant differences between the groups based on Mileage, Price, and Cost. The F-statistic is large (128.88), with a very low p-value (5.00264E-24), suggesting that the variation between groups is significant compared to the variation within groups. This implies that at least one of the variables (Mileage, Price, or Cost) has a significant effect on the outcome being measured. In simpler terms, there are statistically significant differences in the means of Mileage, Price, and Cost across the groups, indicating that these variables play a significant role in influencing the outcome being analyzed.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Mileage	24	2011267	83802.7917	1214155660
Price	24	78108	3254.5	837024.087
Cost	24	66150	2756.25	705502.717

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.0445E+11	2	5.2227E+10	128.882161	5.0026E-24	3.12964398
Within Groups	2.7961E+10	69	405232729			
Total	1.3242E+11	71				

Anova: Two-Factor Without replication

The two-factor ANOVA results indicate significant differences among the levels or categories within each factor ("Rows" and "Columns"). Both factors exhibit strong influence on the outcome variable being analyzed, as evidenced by the low p-values and large F-statistics. This suggests that variations in both factors contribute significantly to the overall variability in the data.

Anova: Two-Factor without replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	34749383.3	23	1510842.75	47.6846408	2.2236E-14	2.01442484
Columns	2979036.75	1	2979036.75	94.023218	1.3629E-09	4.27934431
Error	728733.25	23	31684.0543			
Total	38457153.3	47				

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 $4,959 and 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 statistics offer 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
Minimum	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

Conclusion & Review

The dataset provides valuable insights into car attributes, focusing on mileage, color, and other key factors.

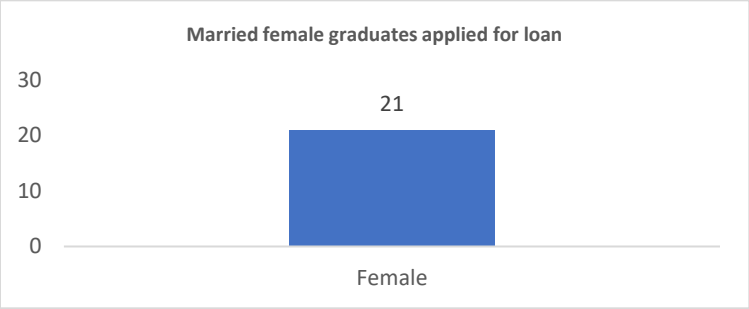
Here's a simple conclusion based on the data:

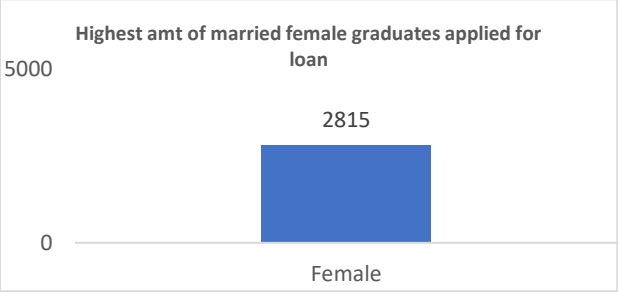
Mileage Comparison: The analysis reveals variations in mileage among different car models. Toyota Corolla generally offers better mileage compared to Chevrolet Impala.

Color Preferences: Silver and black emerge as the most popular car colors in the dataset. Blue, green, red, and white are among the least popular color choices.

Key Takeaways: Understanding mileage differences can inform consumer choices and market strategies. Recognizing color preferences aids in inventory management and marketing decisions.

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



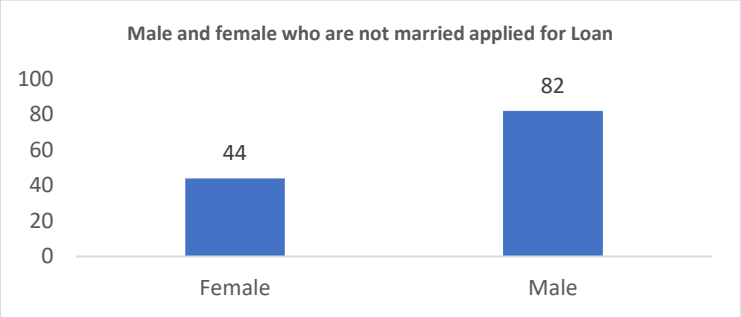


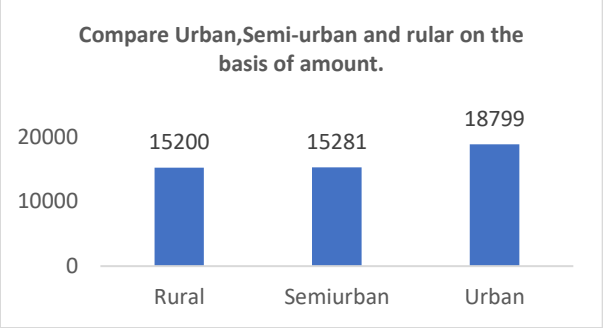
There are total 21married graduate female applied for loan.

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 minimum of $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

Shop Sales Data Report

Introduction:

This dataset encapsulates a wealth of information regarding sales transactions, providing valuable insights into the dynamics of retail operations. With columns meticulously crafted to capture key facets of each transaction, including Date, Salesman, Item Name, Company, Quantity, and Amount, analysts and businesses alike gain access to a treasure trove of actionable data.

Whether it's uncovering trends, optimizing inventory management, or refining sales strategies, this dataset serves as an invaluable resource for driving informed decision-making and unlocking new avenues for growth.

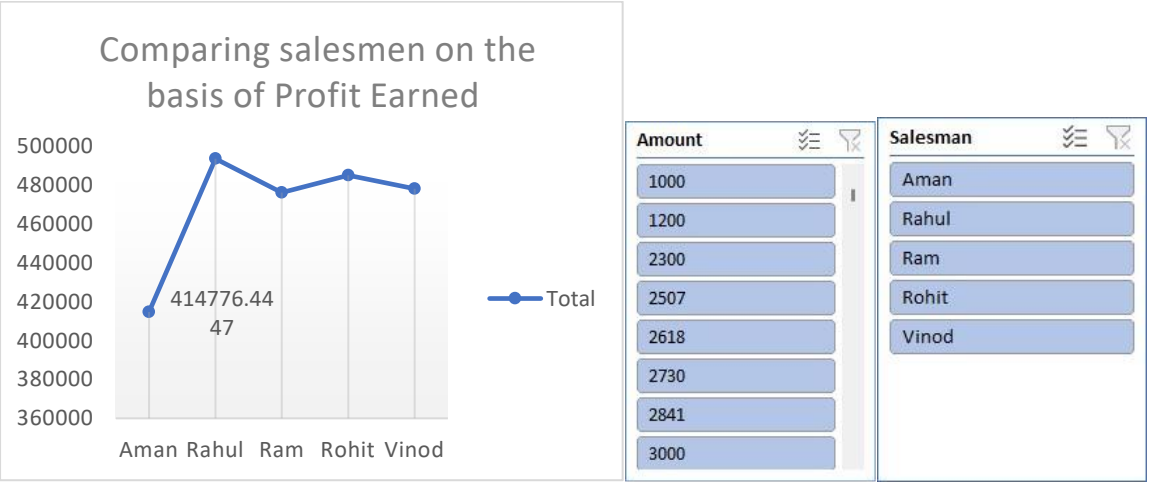
Questionnaire:

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:

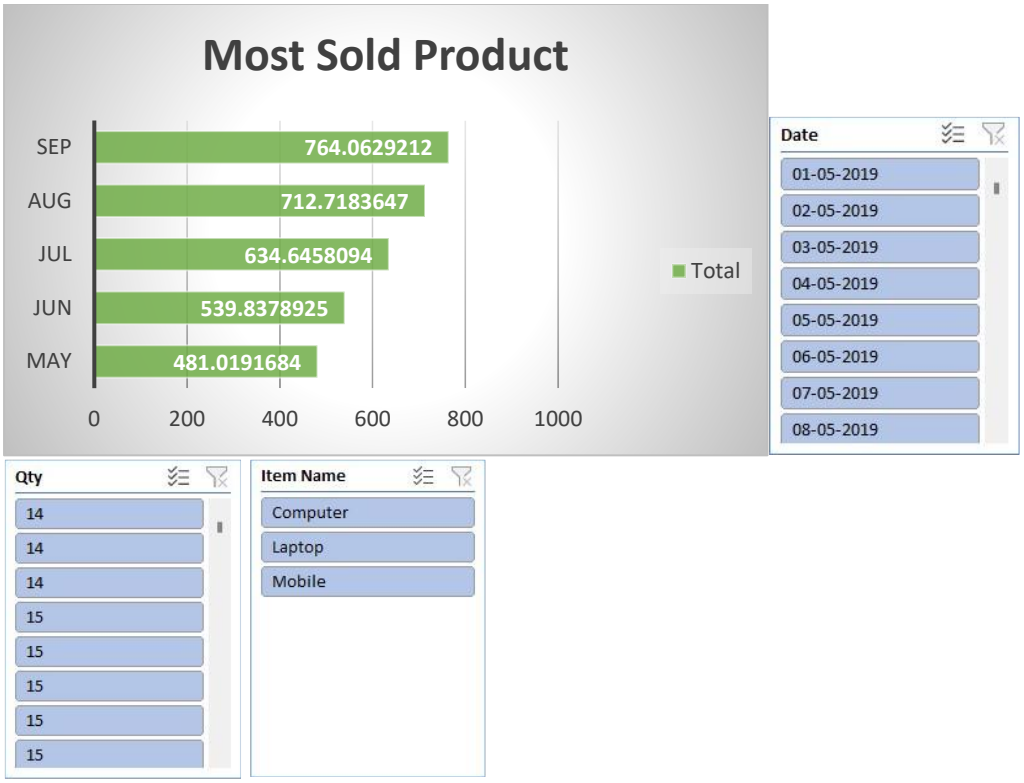
1. Compare all the salesmen on the basis of profit earn.

Ans:- The comparison of all the salesmen on the basis of profit earned is given below:



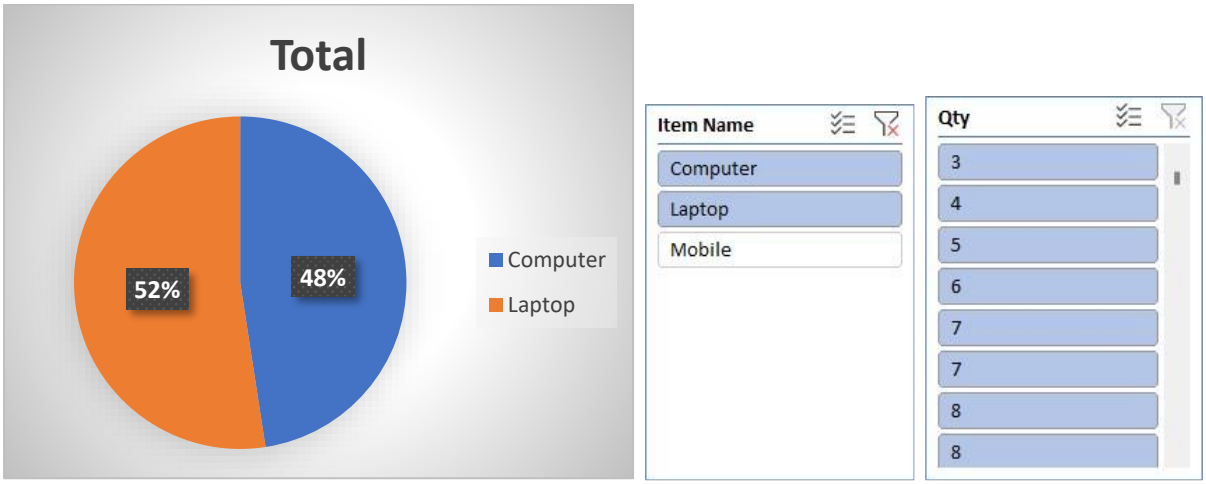
2. Find out most sold product over the period of May-September.

Ans:- To identify the most sold product over the period of May-September, we would need to analyze the sales data within this timeframe. By aggregating the quantity sold for each product across all transactions during this period and then determining which product has the highest total quantity sold, we can pinpoint the most popular item.



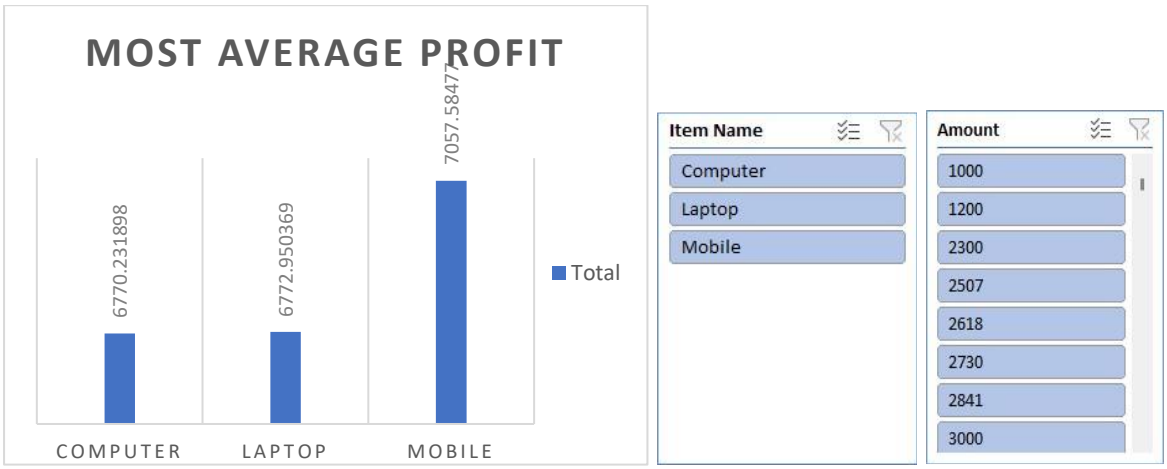
3. Find out which of the two product sold the most over the year Computer or Laptop?

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



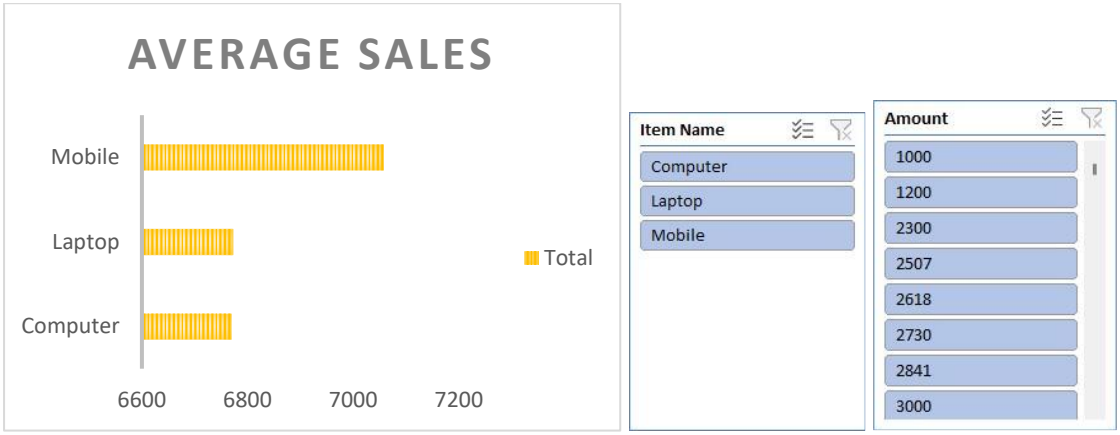
4 . Which item yield most average profit?

Ans:- The item that yields the most profit between laptop, computer and mobile is :



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

Ans:- The average sales of all the products with their respective comparison is :



Conclusion and Review :

The shop sales dataset offers insights into sales trends, salesman performance, item popularity, and company performance. Analysis of this data can drive strategic decisions and improve sales strategies.

The dataset is well-structured and provides comprehensive information on sales transactions. It allows for various analyses, but could benefit from additional variables for deeper insights.Overall, it's a valuable resource for understanding sales dynamics and informing business decisions.

Regression:

The regression model, with a significant p-value indicates a strong positive relationship between Amount and the profit earned and the outcome variable. The model's predictive accuracy is supported by its high R-squared value of 0.660.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.812617
R Square	0.660347
Adjusted R Square	0.629469
Standard Error	1215.119
Observations	13

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	31576697	31576697	21.38598	0.000753
Residual	11	16241653	14776514		
Total	12	47818350			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	244.7062	754.0557	0.32452	0.751632	-1414.96	1904.372
X Variable	0.190729	0.041243	4.624498	0.000735	0.099954	0.281505

Co-relation:

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

	Qty	*Amount*
Column 1	1	
Column 2	#DIV/0!	1

Anova (Single Factor) :

The ANOVA results indicate a significant difference between the two groups , with 1 degree of freedom.

SUMMARY				
Groups	Count	Sum	Average	Variance
Column 1	15	78.56643	5.237762	2.766871
Column 2	15	50419.05	3361.27	3416099

ANNOVA						
Source of Variance	SS	df	MS	F	P-Value	F crit
Between Group	84472135	1	84472135	49.45528	1.2E-07	4.195972
Without Group	47825420	28	170851			
Total	1.32E+08	29				

Anova two factor with Replication:

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) values of 10 respectively. The error term has a degree of freedom of 0

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	841600745	10	4160074	65535	#NUM!	#NUM!
Columns	0	0	65535	65535	#NUM!	#NUM!
Error	0	0	65535			
Total	41600745	10				

Anova two factor without Replication:

Summary	Count	Sum	Average	Variance		
4	*1*	*7800*	*7800*	*#DIV/0!*		
5	*1*	*3000*	*3000*	*#DIV/0!*		
4	1	2300	2300	*#DIV/0!*		
3	1	7000	7000	*#DIV/0!*		
3	1	1200	1200	*#DIV/0!*		
4	1	2506.667	2506.667	*#DIV/0!*		
5	1	2618.095	2618.095	*#DIV/0!*		
6	1	2729.524	2729.524	*#DIV/0!*		
7	1	2840.952	2840.952	*#DIV/0!*		
6	1	4500	4500	*#DIV/0!*		
7	1	3063.81	3063.81	*#DIV/0!*		
1000		39559.05	3596.277	4160074		

Descriptive Statistics:

Column1	
Mean	1000
Standard Error	0
Median	1000
Mode	#N/A
Standard Deviation	#DIV/0!
Sample Variance	#DIV/0!
Kurtosis	#DIV/0!
Skewness	#DIV/0!
Range	0
Minimum	1000
Maximum	1000
Sum	1000
Count	1

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 capabilityand 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 strongpositive correlation between the two columns.

	Column 1	*Column 2*
Column 1	1	
Column 2	0.954077	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 ofthe 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, and p-value. In this case, the F-value (1.003) is associated with a p-value of 0.495, indicating that there is no significant difference between the rows.

Descriptive Statistics:-

Column1		*Column2*	
Mean	19.45693	Mean	6864.457
Standard Error	0.439614	Standard Error	113.5651
Median	19.45693	Median	6984.647
Mode	3	Mode	1000

Standard Deviation	8.129896	Standard Deviation	2100.186
Sample Variance	66.0952	Sample Variance	4410782
Kurtosis	-0.99883	Kurtosis	-0.5078
Skewness	-0.09948	Skewness	-0.36449
Range	30.30852	Range	9279.851
Minimum	3	Minimum	1000
Maximum	33.30852	Maximum	10279.85
Sum	6654.271	Sum	2347644
Count	342	Count	342
Largest(1)	33.30852	Largest(1)	10279.85
Smallest(1)	3	Smallest(1)	1000
Confidence Level(95.0%)	0.864697	Confidence Level(95.0%)	223.3763