Acropolis Institute of Technology and Research, Indore

Department of Computer Science and Engineering

B. Tech. VI Semester

JAN-JUNE 2024

DATA ANALYTICS LAB REPORT

CS-605

Submitted To: Submitted By:

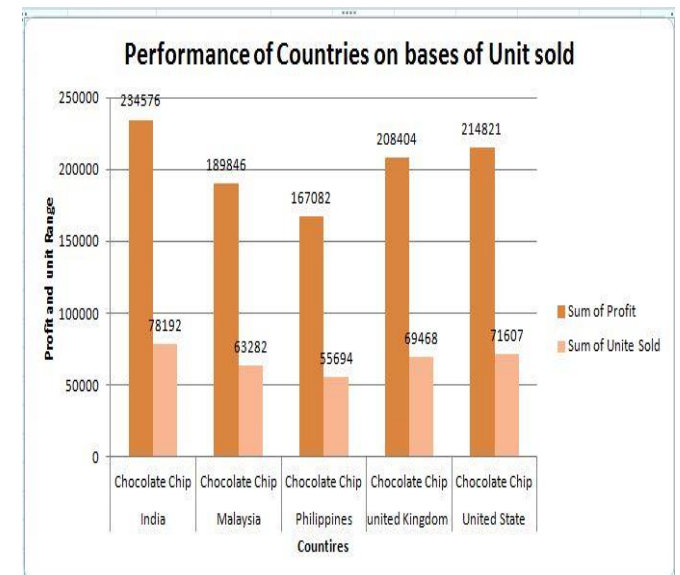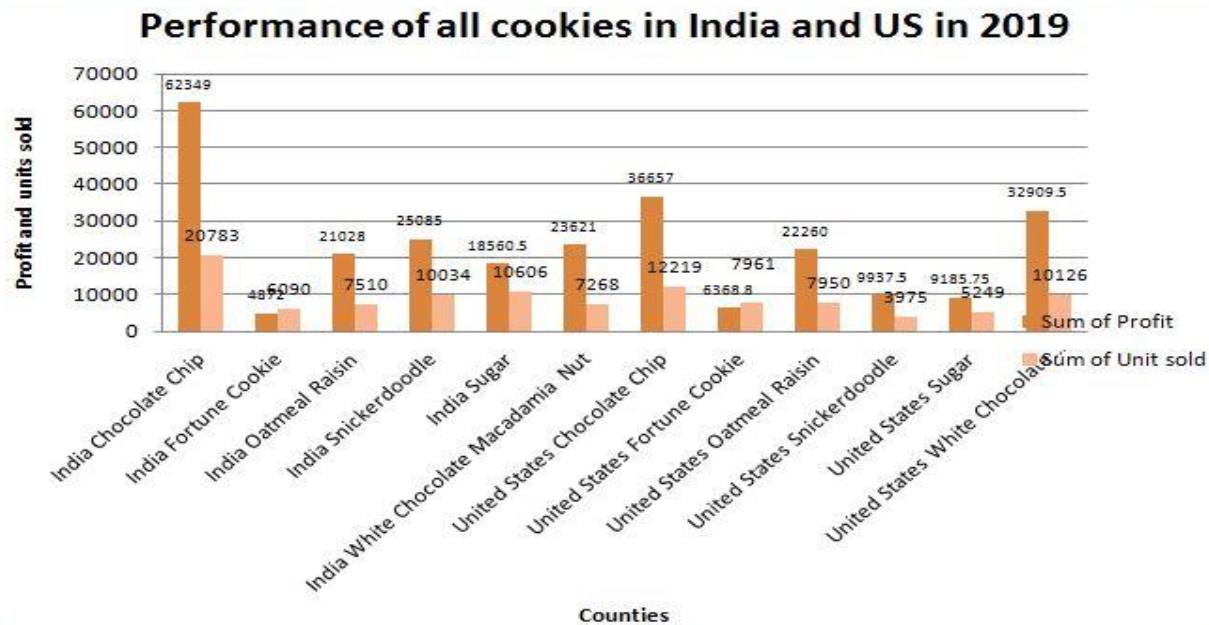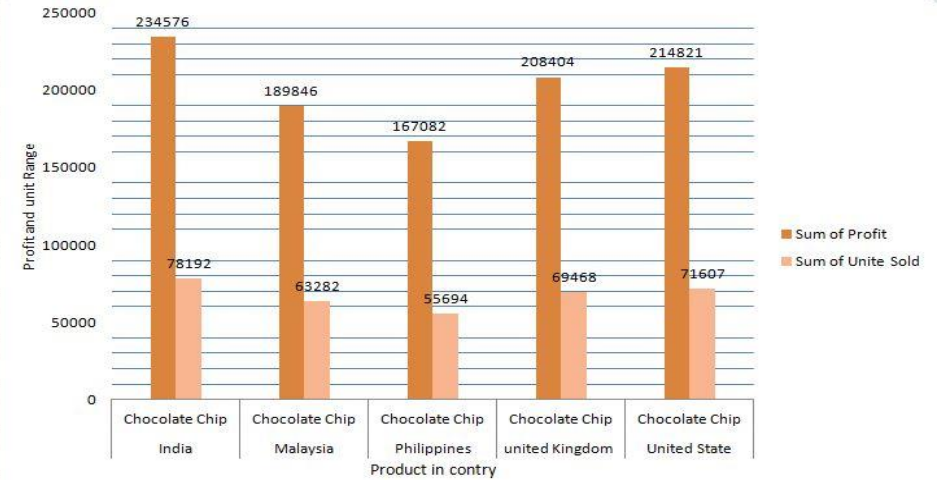Prof. Anurag Punde   Gourav Devrale

0827CS211085

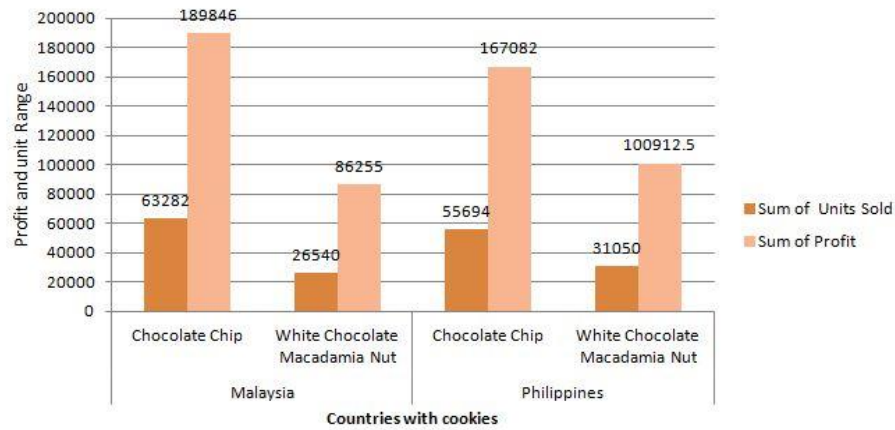# Table of Contents

Dashboards

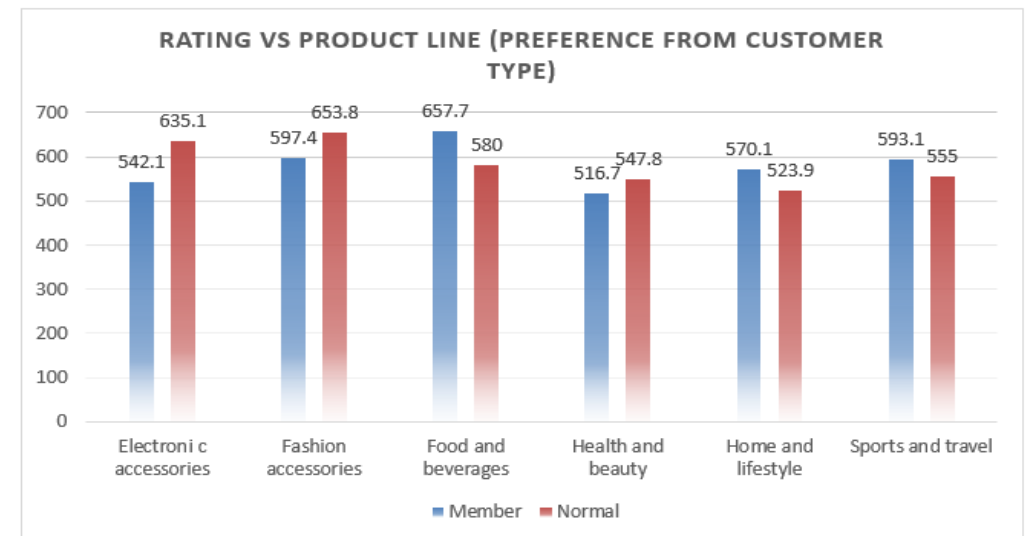| S. No.	Date of Experiment	Experiment	Date of Submission	Signature
1.1		Cookie Data Dashboard		
1.2		Supermarket Dashboard		
1.3		Store Data Dashboard		
1.4		Car Collection Dashboard		
1.5		Order Data Dashboard		
1.6		Loan Data Dashboard		
1.7		Shop Sales Data Dashboard		
1.8		Sales Data Dashboard		

Reports

S. No.	Date of Experiment	Experiment	Date of Submission	Signature
2.1		Cookie Data Report		
2.2		Supermarket Report		
2.3		Store Data Report		
2.4		Car Collection Report		
2.5		Order Data Report		
2.6		Loan Data Report		
2.7		Shop Sales Data Report		
2.8		Sales Data Sample Report		

DASHBOARD FOR COOKIE DATA ANANLYSIS

Performance of all cookies in India and US in 2019

Performance of Countries on bases of Unit sold

DASHBOARD FOR SUPER MARKET DATA ANANLYSIS

Total Amount of Sales made by each city

- YAGON: 4998.042
- NAYPYITAW: 4631.8125
- MANDALAY: 5124.651

City Name

Total Amount of Units Sold by Each Branch (According to Gender)

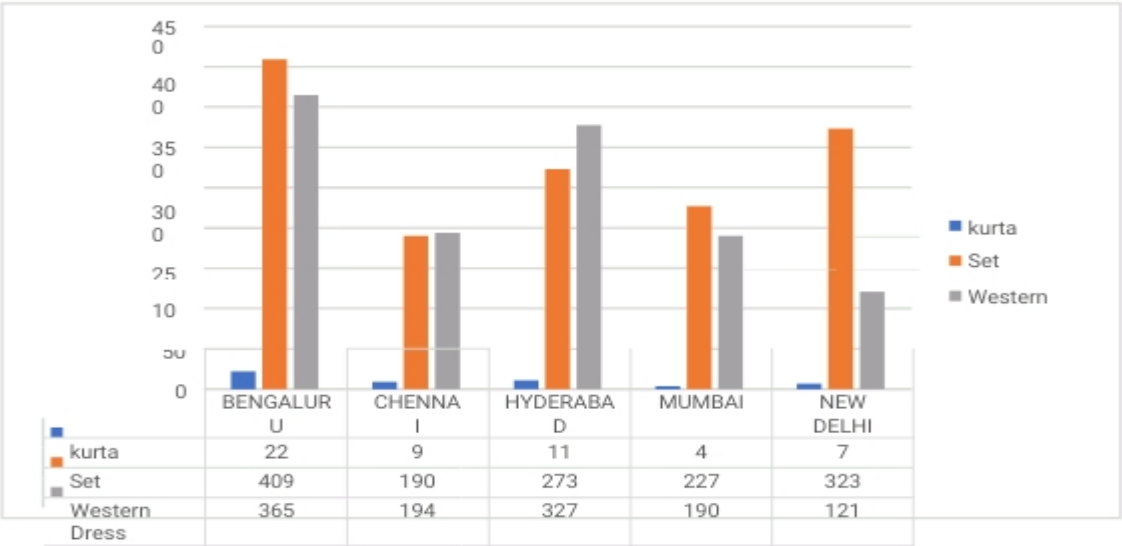| Branch	Female	Male
A	909	950
B	911	909
C	1049	782

■ Female ■ Male

Comparison of Lowest and Highest Rating Product on basis of Units Sold

Lowest and Highest Rating Product

- SPORTS AND T... : 18 / 7
- HEALTH AND BEAUTY: 10 / 16
- FOOD AND BEVERAGES: 26
- FASHION ACCESSORIES: 2
- ELECTRONIC ACCESSORIES: 10 / 4

Chart Area

■ 10 ■ 4

RATING VS PRODUCT LINE (PREFERENCE FROM CUSTOMER TYPE)

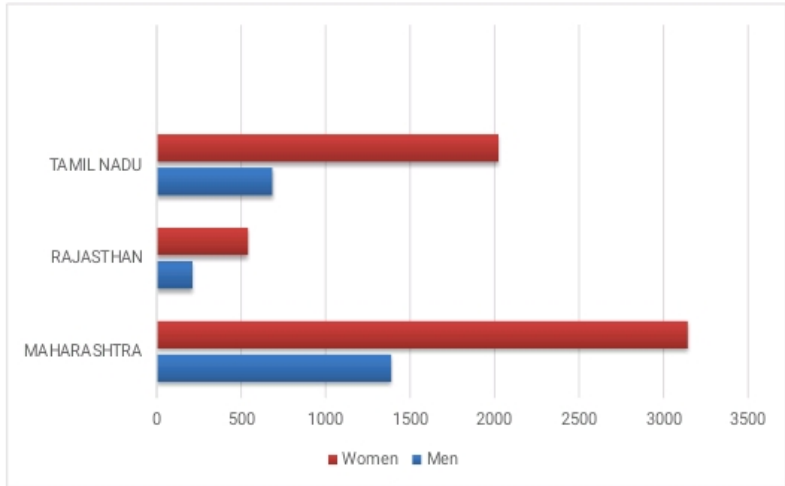| Product Line	Member	Normal
Electronic accessories	542.1	635.1
Fashion accessories	597.4	653.8
Food and beverages	657.7	580
Health and beauty	516.7	547.8
Home and lifestyle	570.1	523.9
Sports and travel	593.1	555

■ Member ■ Normal

DASHBOARD FOR STORE DATA ANANLYSIS

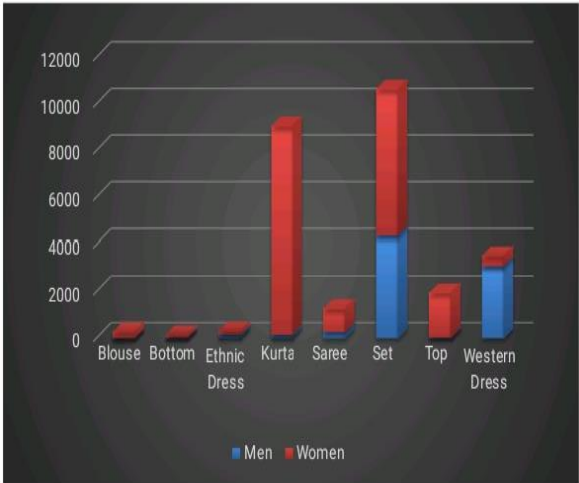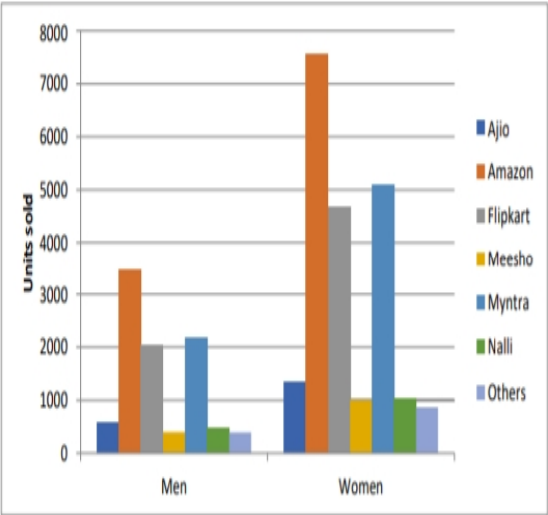|	BENGALURU	CHENNAI	HYDERABAD	MUMBAI	NEW DELHI
kurta	22	9	11	4	7
Set	409	190	273	227	323
Western Dress	365	194	327	190	121
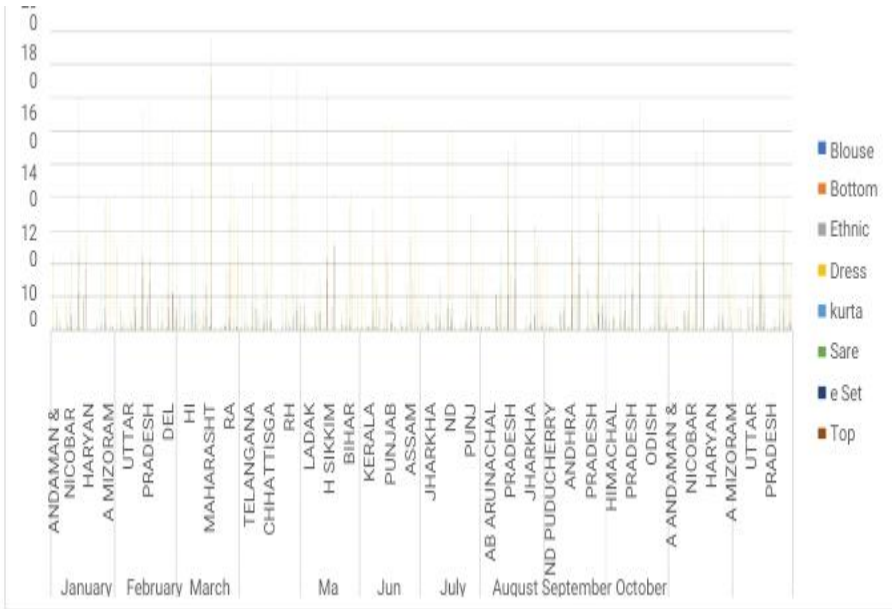
DASHBOARD FOR CAR DATA ANALYSIS

Corolla vs Chevrolet Mileage

Bar chart: Toyota Corolla = 277,131; Chevrolet Impala = 228,486. Legend: Total

Ford car Vs Honda Car Comparision

Grouped bar chart for Ford (Escape, F-150, Fusion, Mustang) and Honda (Accord, Civic, CRV). Legend: Sum of Price, Sum of Cost, Sum of Mileage

Green Vs Silver Color Car Mileage Comparision

Bar chart: Total — Green, Silver. Legend: Green, Silver

Color count

Bar chart: Black = 6, Blue = 3, Green = 3, Red = 3, Silver = 6, White = 3. Legend: Total

Car with cost greater then $2000

Line chart: Accord 6500, Altima 5500, Charger 9300, Corolla 6300, CRV 4100, Escape 6950, F-150 3000, Fusion 2100, Impala 5500, Malibu 3000, Maxima 2500, Mustang 3100, Silverado 4500. Legend: Total

DASHBOARD FOR ORDER DATA ANALYSIS

Sum of Sales — Total pie chart with Segment and State filters (Consumer Alabama, Consumer Arizona, Consumer Arkansas)

Total column chart — Furniture, Office Supplies, Technology

Column chart by state (California, Texas, Washington) — Consumer, Corporate, Home Office

Sum of Sales / Average of Sales by Segment — Consumer 1148060.531 (225.0657775), Corporate 688494.0748 (233.1507195), Home Office 424982.1769 (243.4033086)

Sub category / category column chart — Bookcases, Chairs, Furnishings, Tables, Appliances, Art, Binders, Envelopes, Fasteners, Labels, Paper, Storage, Supplies, Accessories, Copiers, Machines, Phones

DASHBOARD FOR LOAN DATA ANANLYSIS

Graduate UnMarried Male Loan Data

350

0

Male

■ Max of LoanAmount ■ Count of LoanAmount

Graduate UnMarried Female Loan Data

400
350
300
250
200
150
100
50
0

Female

■ Max of LoanAmount ■ Count of LoanAmount

Non Graduate Unmarried Male Loan

25

20
19

15

10

50

0
16

Male

■ Max of LoanAmount ■ Count of

Graduate Married Female Loan Data

500
450
400
350
300
250
200
150
100
50
0

Female

■ Max of LoanAmount ■ Count of Gender

35%

65%

■ Female ■ Male ■

Comparision of Loan Amount on basis of Property Area

6000

5000

4000

3000

2000

1000

0

Rural   Semiurban   Urban

□ Series 1 □ Series 2

DASHBOARD FOR SHOP SALES DATA ANALYSIS

DASHBOARD FOR SALES DATA ANALYSIS

COMPARE THE SALE OF VINTAGE CARS AND CLASSIC CARS FOR ALL THE COUNTRIES.

average sales of all the products

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

Sales from 2004&2005

countries based on deal size.
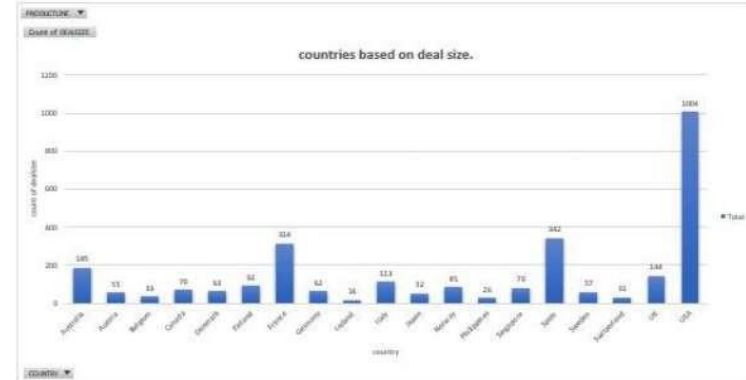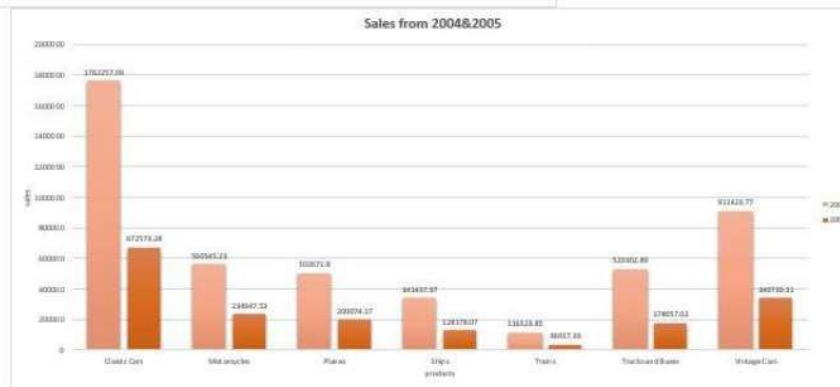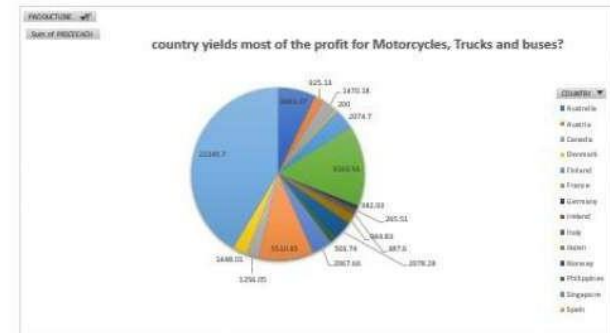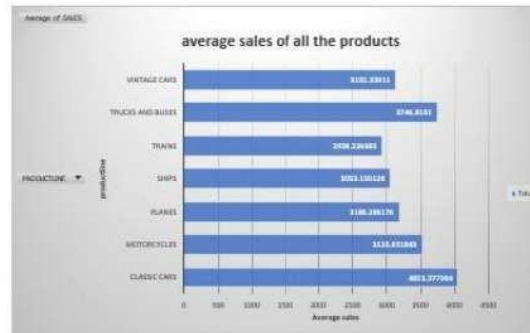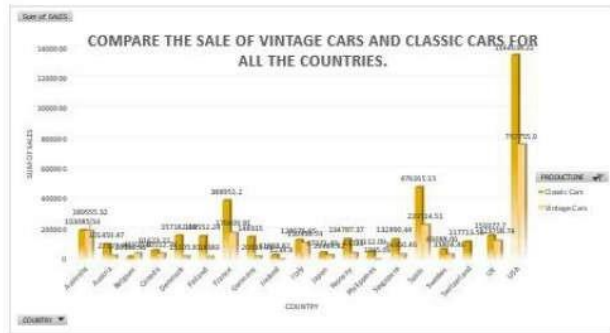
Cookie Data Report

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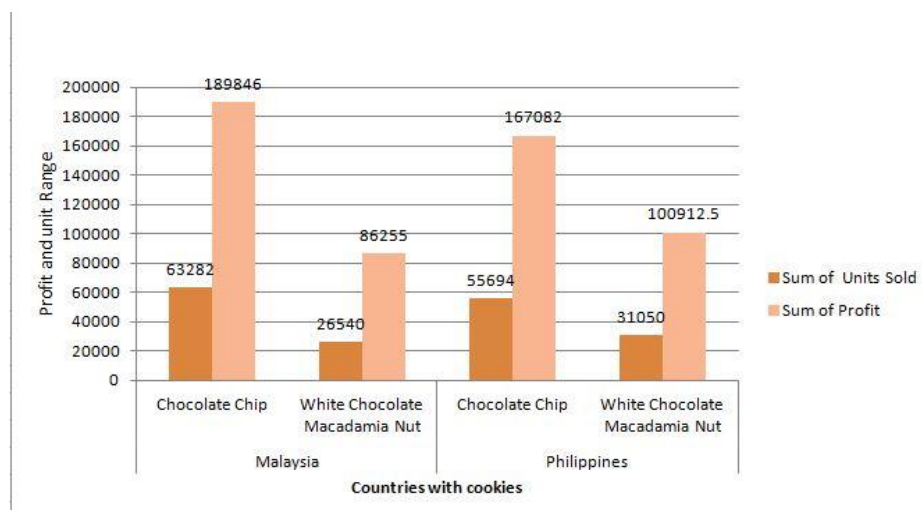
2. What is the performance of Choco Chips Cookies in all Country Which Competes the best.

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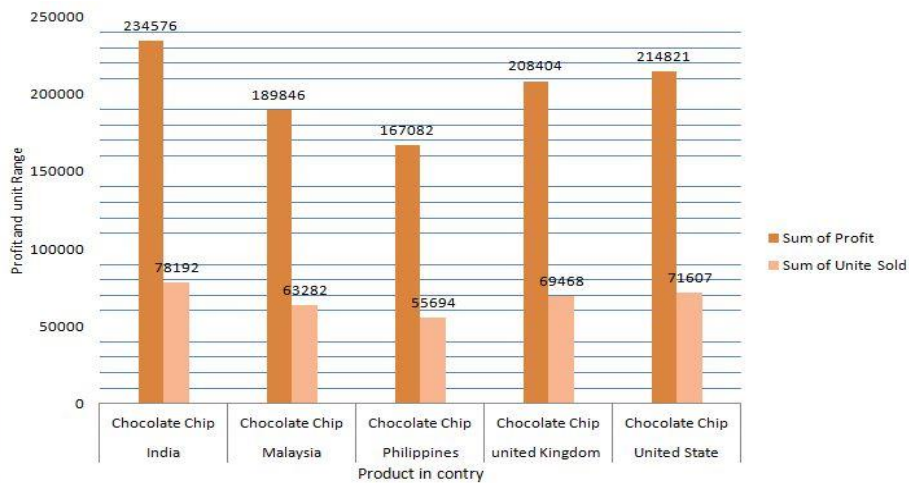
3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

Ans:- India stands out as the leading performer globally when it comes to both profit generation and units sold in the Choco chip market.

Performance of all cookies in India and US in 2019

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.
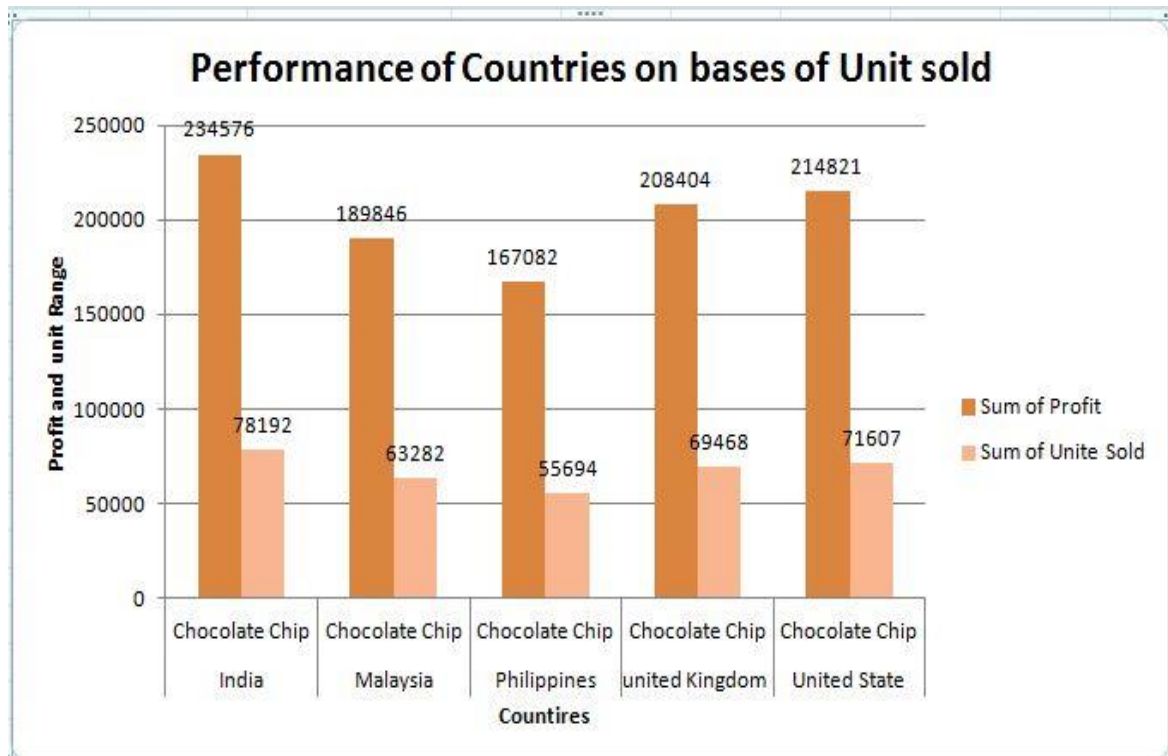
Conclusion and Review:

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.

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
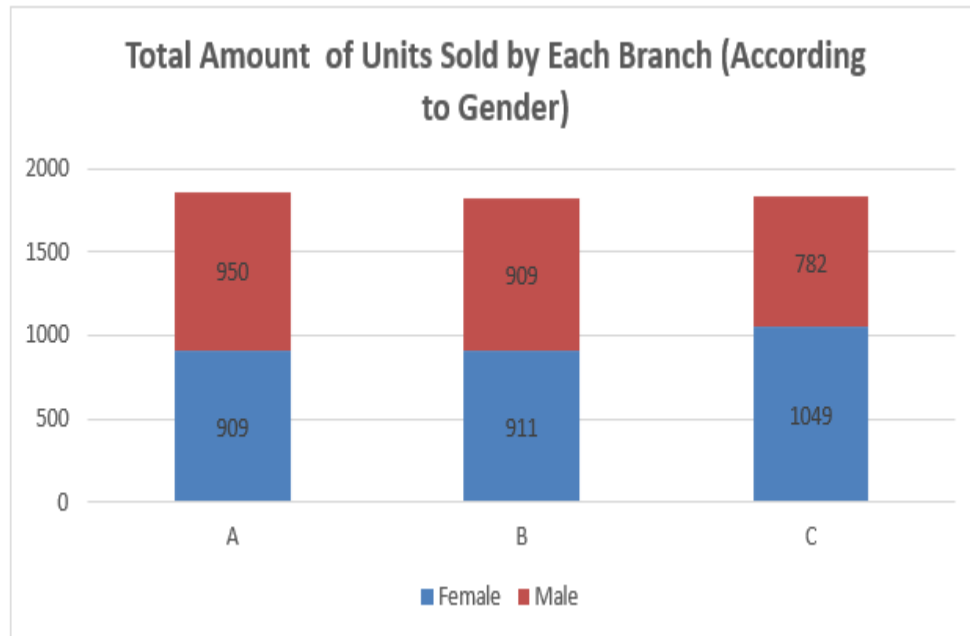
Analytics:

Q1. Which of the given cities having tax 5% slab performed better than all the others?
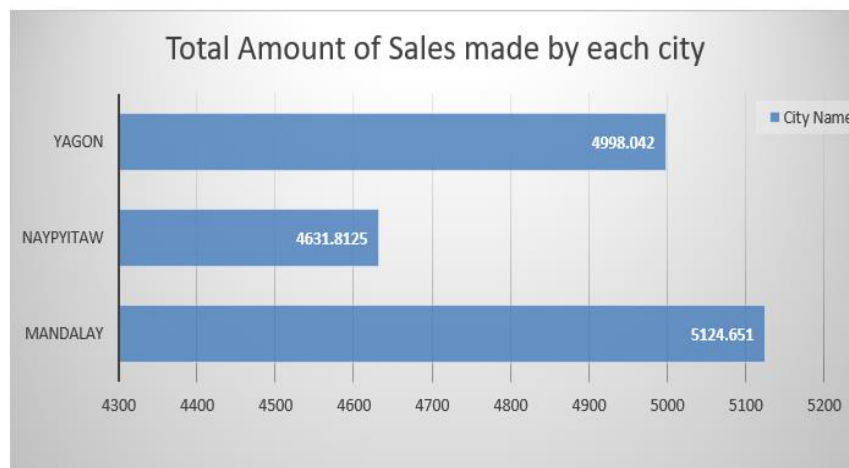
Based on the data analyzed, the city that outperformed all is **Mandalay**. This conclusion is drawn from superior performance in total sales/revenue generation compared to the other cities in the same tax slab of 5%.

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

Total Amount of Units Sold by Each Branch (According to Gender)

	A	B	C
Male	950	909	782
Female	909	911	1049

■ Female ■ Male

Quantity	Gender	Branch
1	Female	A
2	Male	B
3		C
4		
5		
6		
7		
8		

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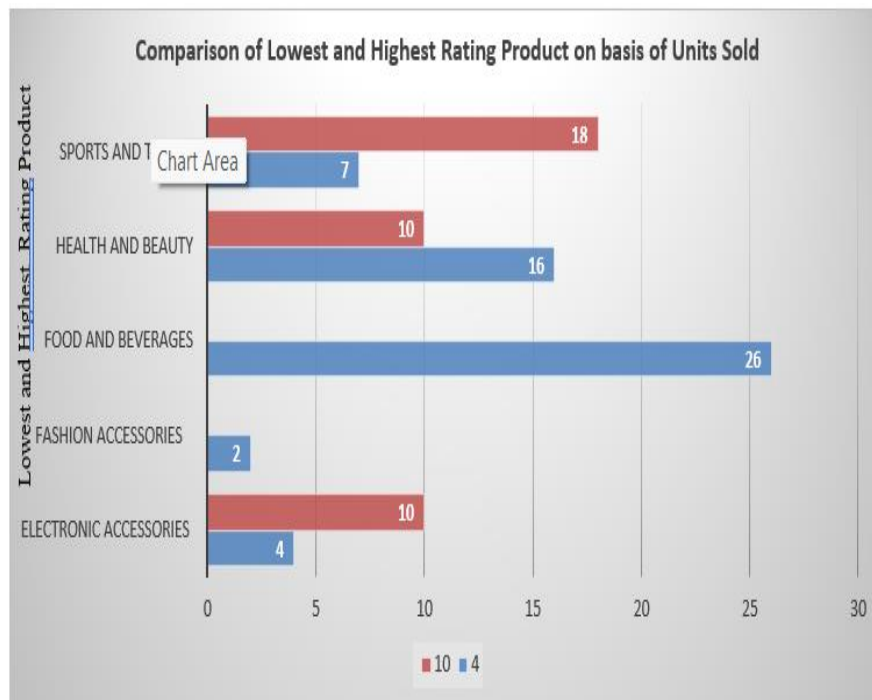
Answer. Upon analyzing the Supermarket Sales Data, we discovered that product ratings ranged from a minimum of 4 to a maximum of 10.

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

| 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

Answer:

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.

   With a degree of freedom of 1, the trendline equation stands as

   Quantity = 0.0012x + 5.4438. Despite this equation, the coefficient of determination (R2) is merely 0.0001,   highlighting the inconsistency in customer buying patterns solely based on unit price.

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

RATING VS PRODUCT LINE (PREFERENCE FROM CUSTOMER TYPE)
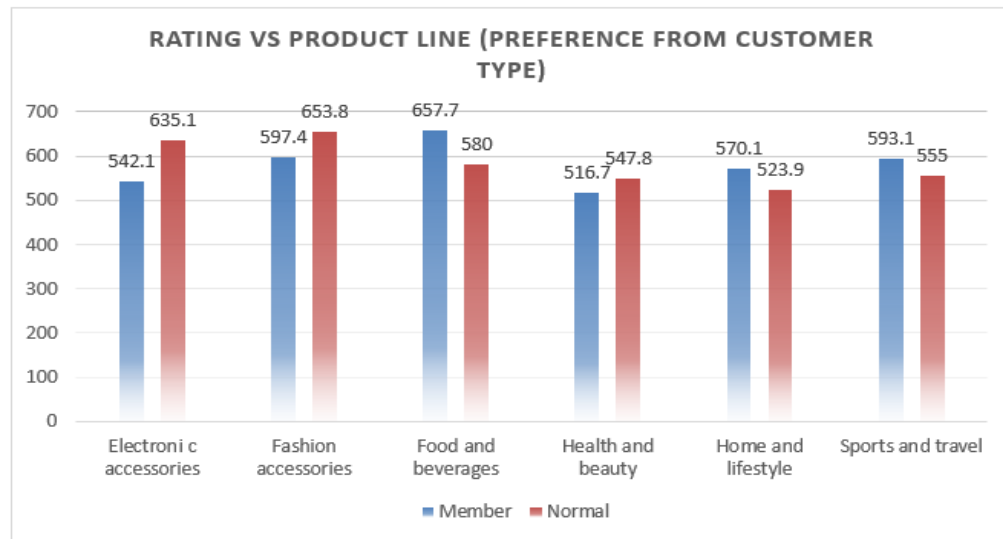
Rating	Customer type	Product line
4	Member	Electronic accessories
4.1	Normal	Fashion accessories
4.2		Food and beverages
4.3		Health and beauty
4.4		Home and lifestyle
4.5		Sports and travel
4.6		
4.7		

Answer. 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

The comprehensive analysis of supermarket sales dynamics provides valuable insights into consumer behavior, operational trends, and performance metrics. Here's a summary of the findings and reviews:

1. City Performance:

Mandalay emerged as the top-performing city among those with a 5% tax slab. Its superior sales/revenue generation signifies a potentially lucrative market for supermarket businesses.

2. Gender-based Ordering:

 Female customers showed a higher propensity to order items from Branch A, while males dominated in Branch C. Branch B saw equal orders from both genders. This gender-specific trend highlights the importance of targeted marketing strategies.

3. Rating and Units Sold:

Further analysis is needed to compare products with the highest and lowest ratings based on units sold. Understanding the correlation between product ratings and sales volume can inform inventory management and marketing decisions.

4. Unit Price and Revenue Relationship:

The regression analysis revealed a weak correlation ($R^2 = 0.0001$) between unit price and quantity sold. This suggests that customers' purchasing decisions may not be significantly influenced by unit price alone, indicating the need for deeper insights into consumer preferences and behavior.

5. Product Recommendations:

Based on city data analysis, Food and Beverages are recommended for member-type customers, while Fashion Accessories are suggested for normal customers. These recommendations align with the observed preferences and purchasing patterns in respective cities.

Reviews:

The report provides a thorough exploration of supermarket sales dynamics, covering various aspects such as city performance, gender-based ordering trends, and product recommendations.

The inclusion of regression analysis enhances the depth of insights, though further interpretation of the results could strengthen the analytical rigor.

Clear visuals, such as graphs and charts, would enhance the presentation of findings and aid in understanding complex relationships.

Overall, the report offers valuable insights for supermarket stakeholders, highlighting areas for strategic focus and improvement in marketing and operational strategies.

Store Data Report

Introduction:

This dataset encompasses sales data from a retail store, featuring a range of attributes including customer demographics (Gender, Age Group), transaction details (OrderID, Status),product specifics (Category, SKU), and shipping information. With a focus on understanding customer behaviour and product trends, our analysis aims to uncover patterns, preferences, and correlations within the data. By leveraging these insights, businesses can optimize marketing efforts, enhance inventory management, and improve customer satisfaction.

Questionnaire:

1. which of the channel performed better than all other channels in compare 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 of 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 in compare men & women?

<u>Ans</u>: Amazon leads in the sales in both men and women category followed by Myntra and Flipkart. Amazon sold almost 3500 units in men category and almost 7500 units in women category. Myntra sold 2000 units in men section.

Graduate UnMarried Male Loan Data
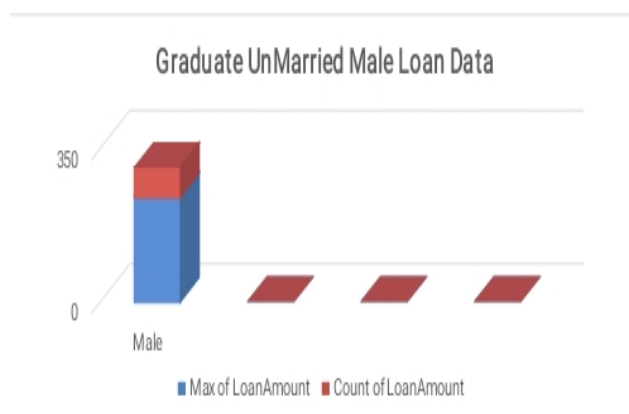
2. Compare category. Find out most sold category above 23 years of age for any gender.

Ans: In the above 23 years of age group Kurta is most sold category in women section with 8820 units sold. Set is most sold category in men section with 4365 units sold also set is the second most sold category in women section.

The table of items sold is given below:

Item	Men	Women	Grand Total
Blouse	6	190	196
Bottom	40	28	68
Ethnic Dress	150	77	227
kurta	156	8820	8976
Saree	261	941	1202
Set	4365	6204	10569
Top	45	1825	1870
Western Dress	3078	380	3458
Grand Total	8101	18465	26566

The graph is as follows:

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

Ans: In Maharashtra: Sales in men category=1390, Sales in women category= 3144

In Tamil Nadu: Sales in men category=686, Sales in women category= 2023

In Rajasthan: Sales in men category=21, Sales in women category=543

State	Men	Women	Grand Total
MAHARASHTRA	1390	3144	4534
RAJASTHAN	212	543	755
TAMIL NADU	686	2023	2709
Grand Total	2288	5710	7998

ship-state

New Delhi
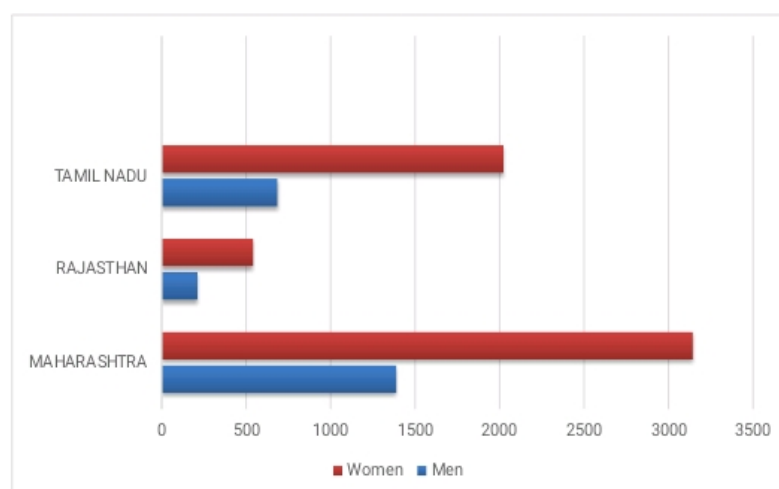ODISHA
PUDUCHERRY
PUNJAB
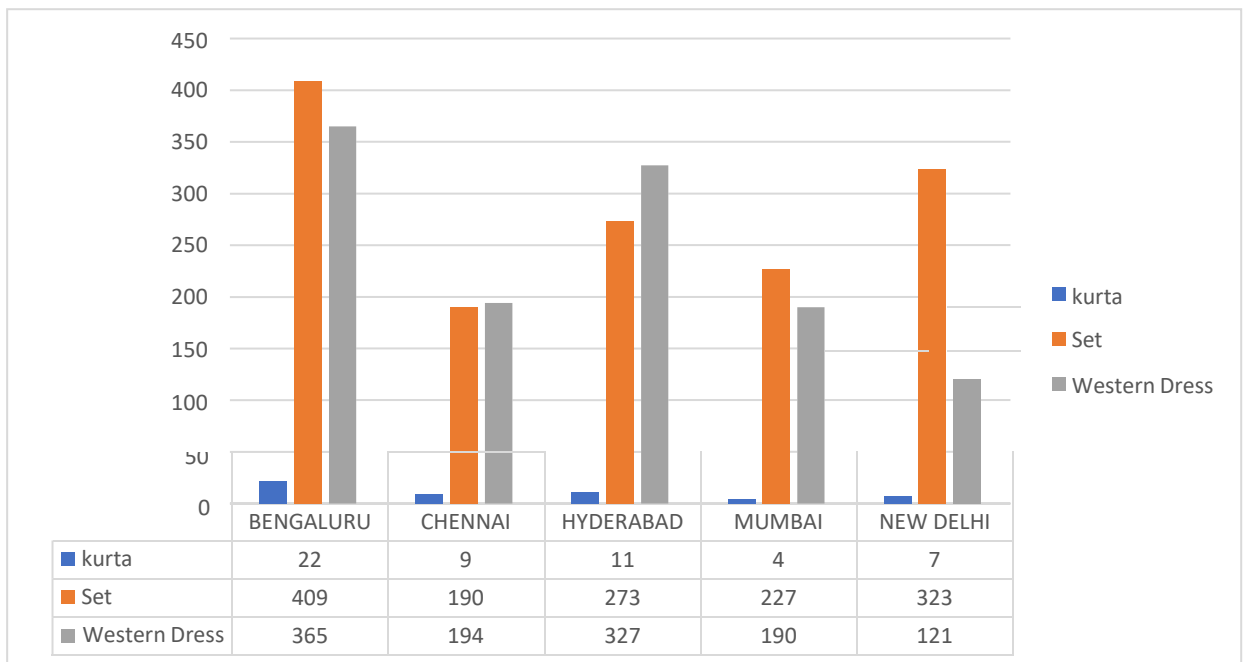RAJASTHAN
SIKKIM
TAMIL NADU
TELANGANA

4. Which city sold most of following categories
   a. Kurta
   b. Set
   c. Western wears

Ans: Bengaluru, Chennai, Hyderabad, Mumbai and New Delhi are the cities sold most of kurtas, Sets and western wears.

	BENGALURU	CHENNAI	HYDERABAD	MUMBAI	NEW DELHI
kurta	22	9	11	4	7
Set	409	190	273	227	323
Western Dress	365	194	327	190	121

City	kurta	Set	Western Dress	Grand Total
BENGALURU	964	938	422	2324
CHENNAI	666	451	217	1334
HYDERABAD	713	687	370	1770
MUMBAI	437	515	207	1159
NEW DELHI	479	792	142	1413
Grand Total	3259	3383	1358	8000

Category
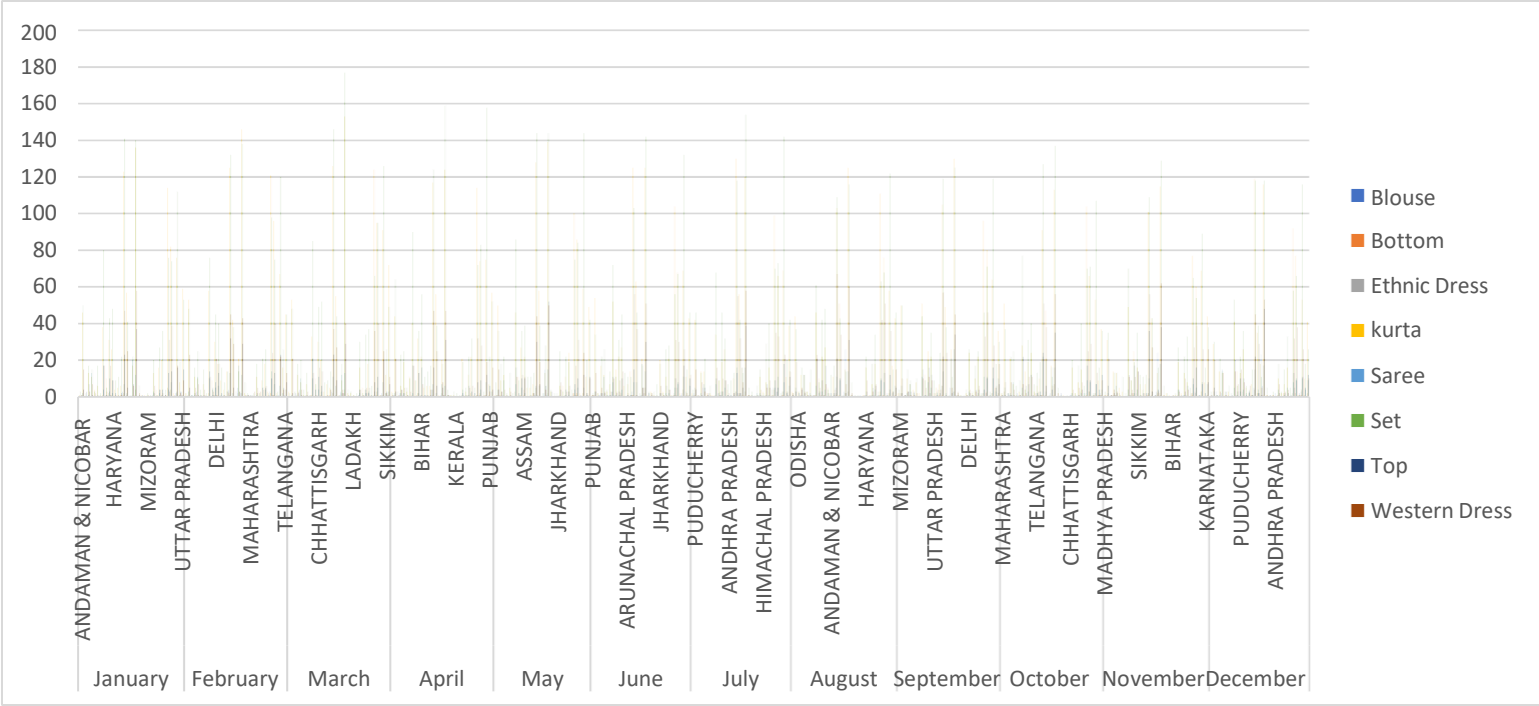
- Blouse
- Bottom
- Ethnic Dress
- kurta
- Saree
- Set
- Top
- Western Dress

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

Ans: The graph for most items sold in any of stats on basis of category is as follows:

Conclusion and Review:

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

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.

Comparesion the mileage of Chevrolet Impala to Toyota Corolla

Mileage

59,169
87,278
87,675
130,684
140,811
40,826
41,560
42,542
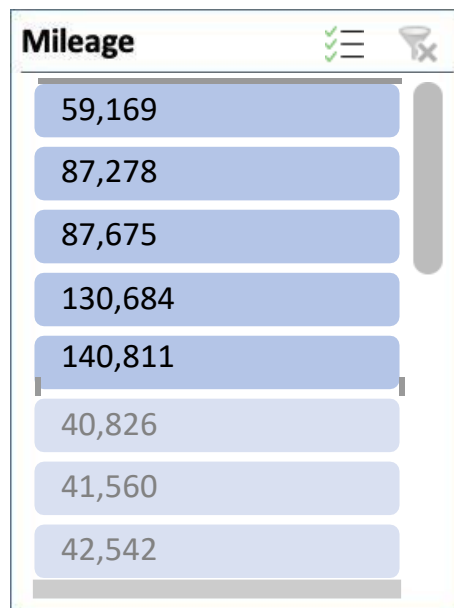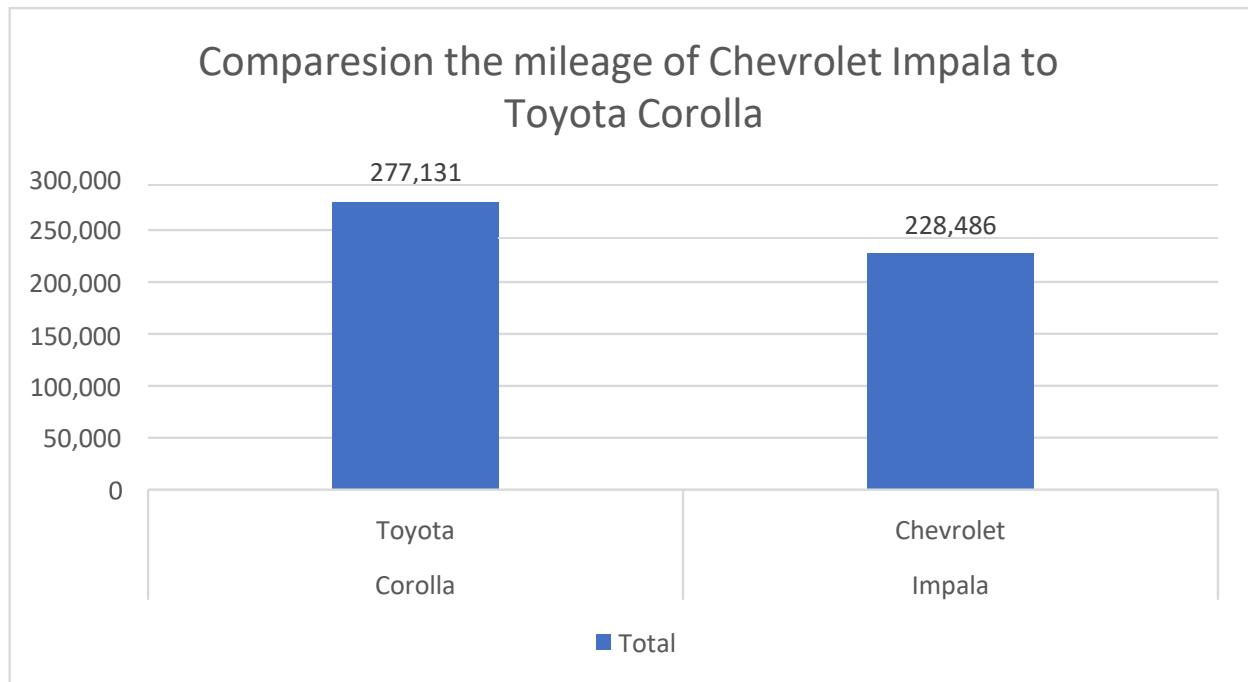
Q2. Justify, Buying of any Ford car is better than Honda.

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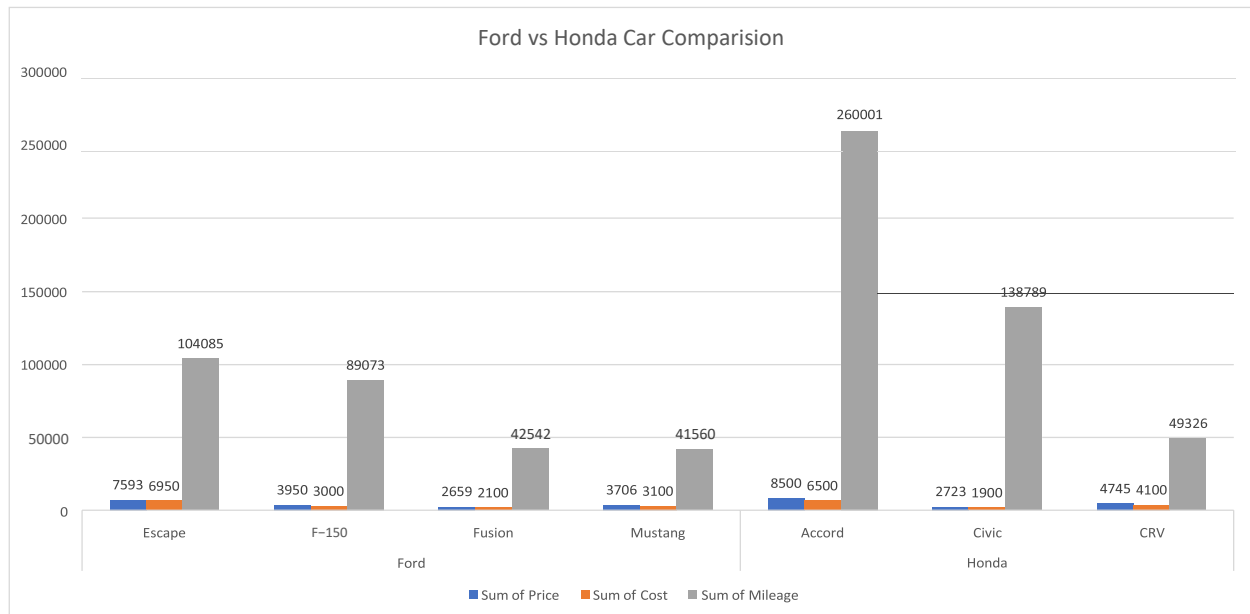
Q3. Among all the cars which car color is the most popular and is least popular?

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.

Color count

Black	6
Blue	3
Green	3
Red	3
Silver	6
White	3

Legend: ■ Total
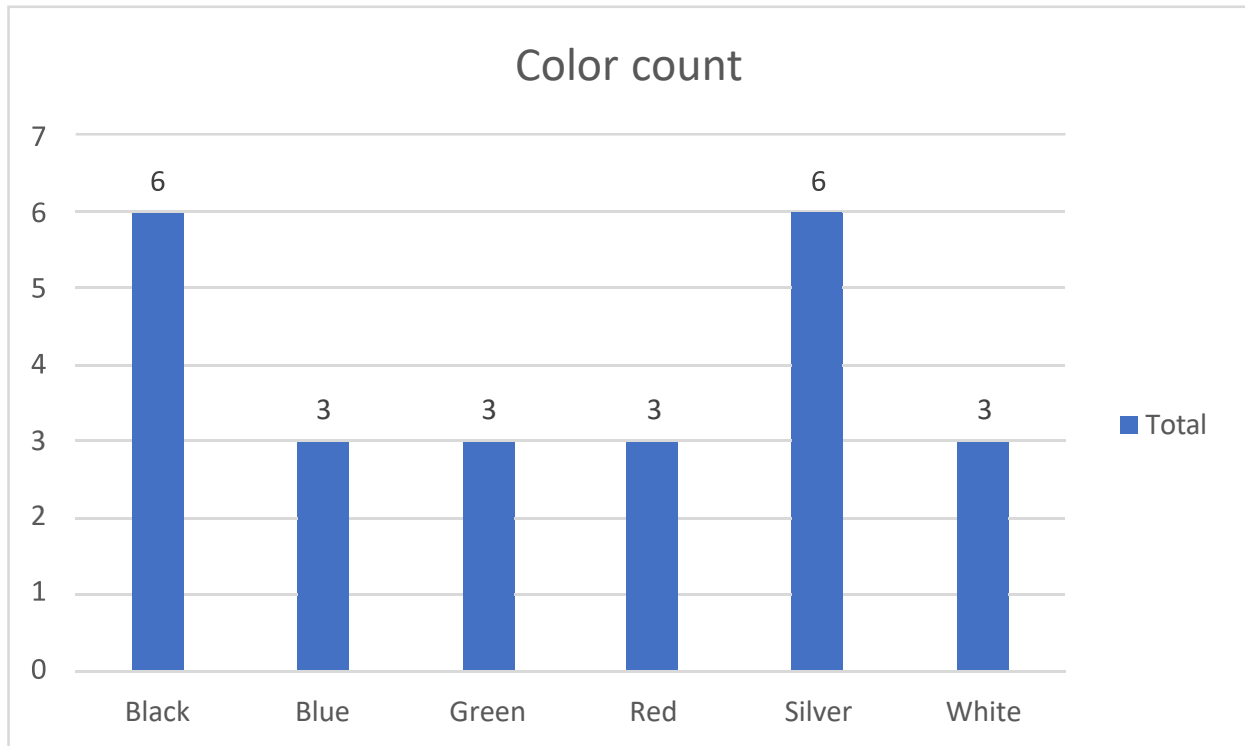
Model

- Accord
- Altima
- Camry
- Charger
- Civic
- Corolla
- CRV
- Escape

Color

- Black
- Blue
- Green
- Red
- Silver
- White

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.

Silver vs Green Car Mileage Comparision

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
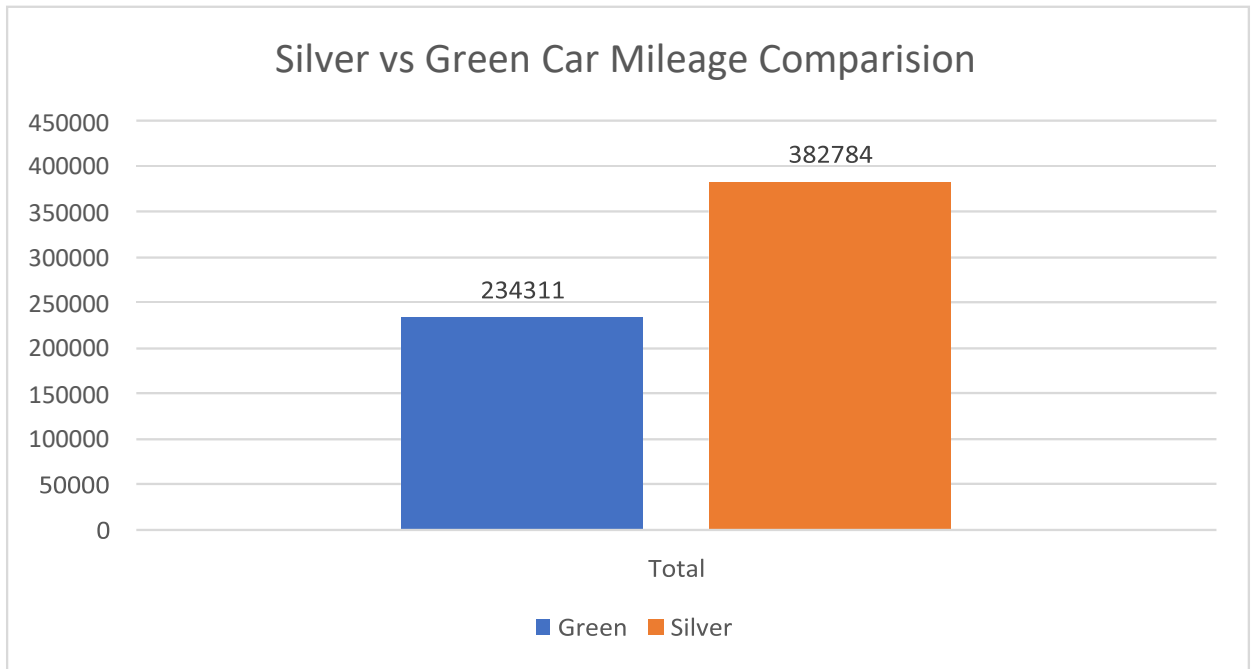
Cars with cost more then $2000

Car	Total
Accord	6500
Altima	5500
Charger	9300
Corolla	6300
CRV	4100
Escape	6950
F-150	3000
Fusion	2100
Impala	5500
Malibu	3000
Maxima	2500
Mustang	3100
Silverado	4500

Model	Cost
Accord	$1,500
Altima	$1,800
Camry	$1,900
Charger	$2,000
Civic	$2,100
Corolla	$2,200
CRV	$2,500
Escape	$3,000

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.

EXPLORING SALES OF DIFFERENT SEGMENT IN US STATES

INTRODUCTION:

Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of different category sales. From fundamental transactional details such as Date, Time, sales, states to more nuanced factors like Customer Type, Demographics, category and sub category, every facet has been meticulously documented.

Key Attributes:

1. ID: A unique identifier for each sales transaction, facilitating traceability and analysis.
2. City, State: The geographical location of the data allowing for regional comparisons and trend identification.
3. Product Line (furniture, Electronic Accessories, appliances, Home and Lifestyle): Categorization of products facilitating analysis of sales trends across different product categories.
4. Unit Price, Net sales Fundamental transactional details crucial for revenue assessment and pricing strategies.
5. Net sales of different category, category performing well in different states: Performance metrics
6. Rating: different product performing well in different state
7. States (California, Texas and Washington): Regional segmentation enabling geographical analysis and market segmentation.

2. 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.
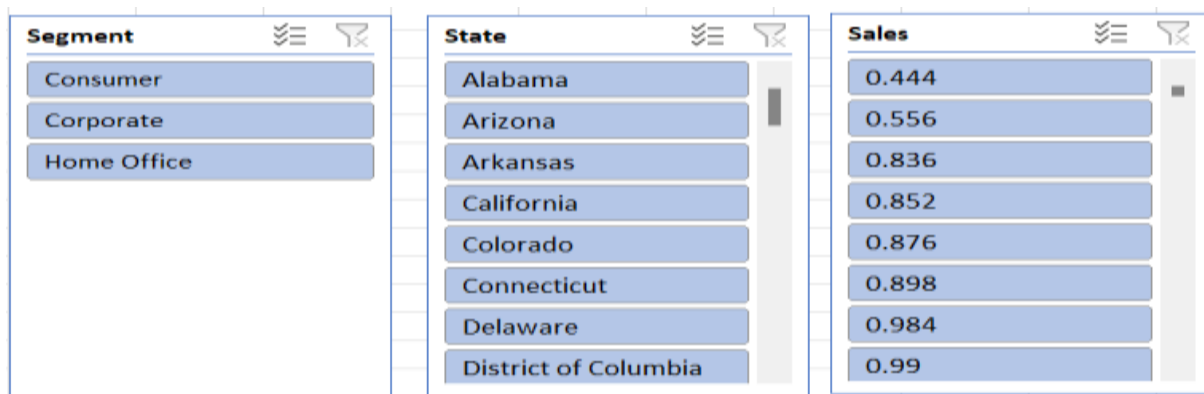
3. ANALYTICS:

Q1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?
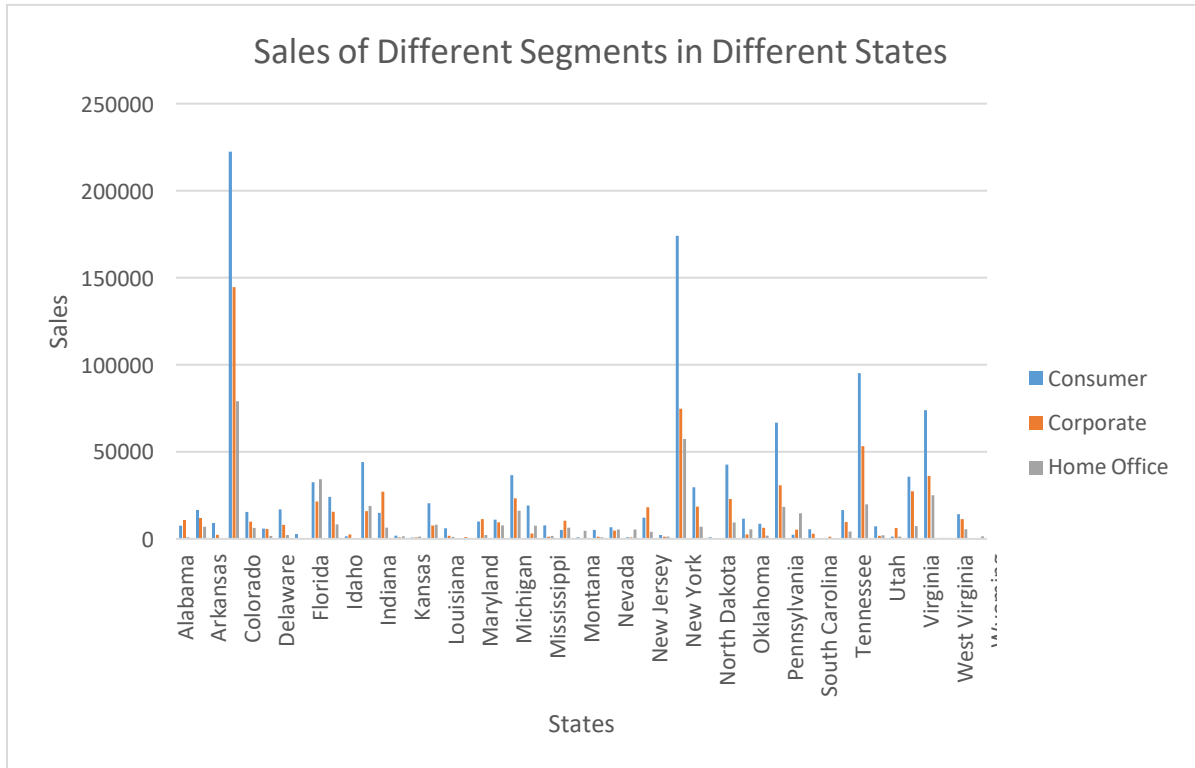
- After comparing all the states in terms of segment and sales, California emerged as the state with the highest amount of sales
- Consumer segment performed well in all the states

Slicers:

Q2. Find out top performing category in all the states?
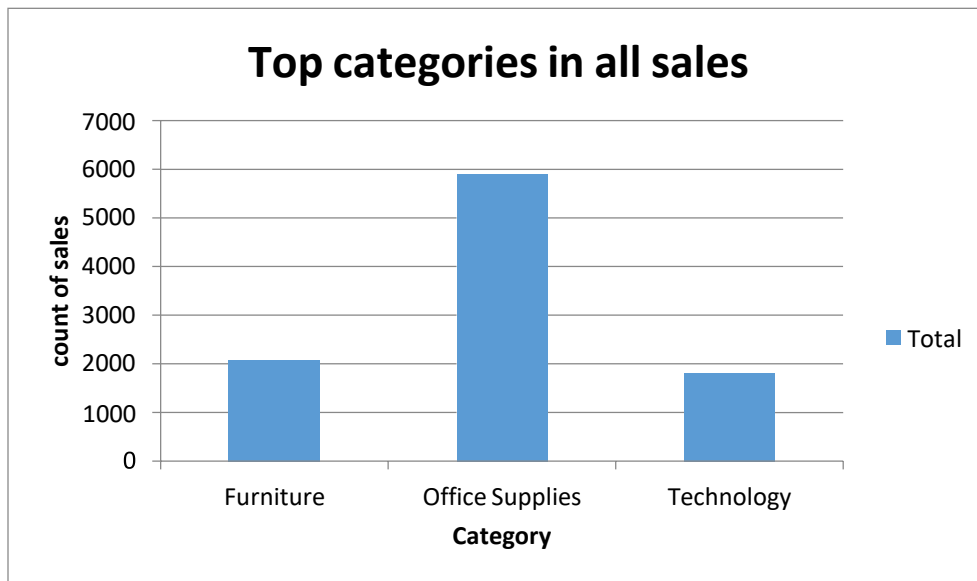
Ans. Office Supplies is the top performing category in all the states

Slicers:

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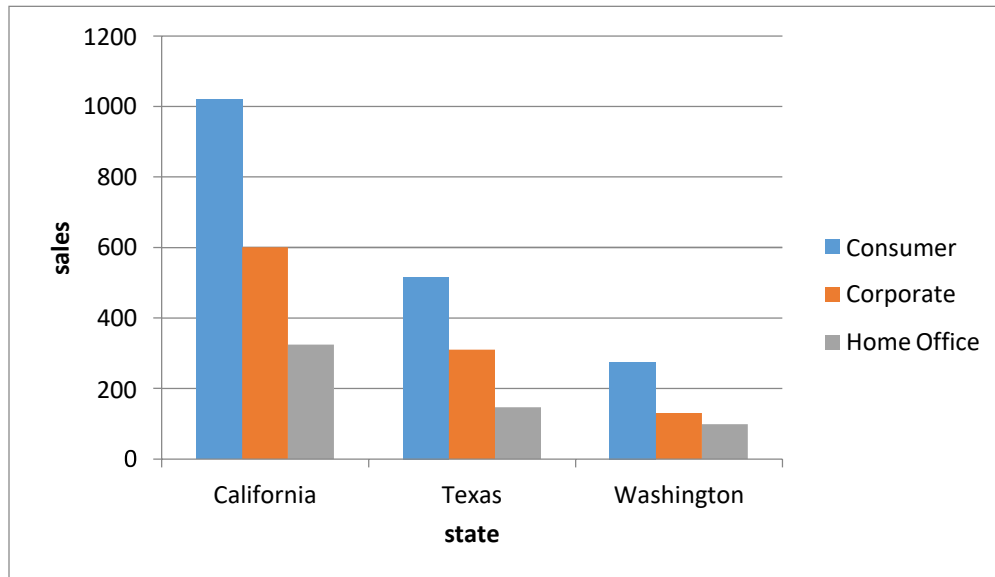
Ans. Consumer segment has the most sales in US, California, Texas, and Washington

Slicers:

Q4. Compare total and average sales for all different segment?
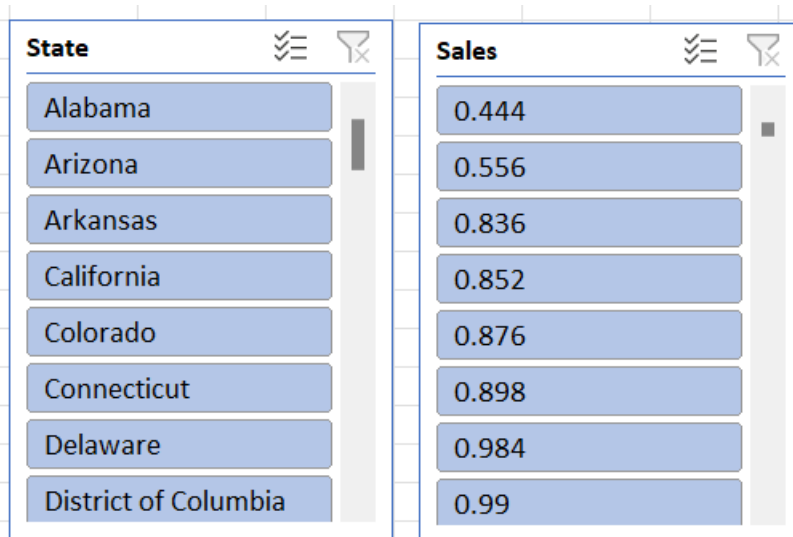
Total vs Average Sales of Different Segments
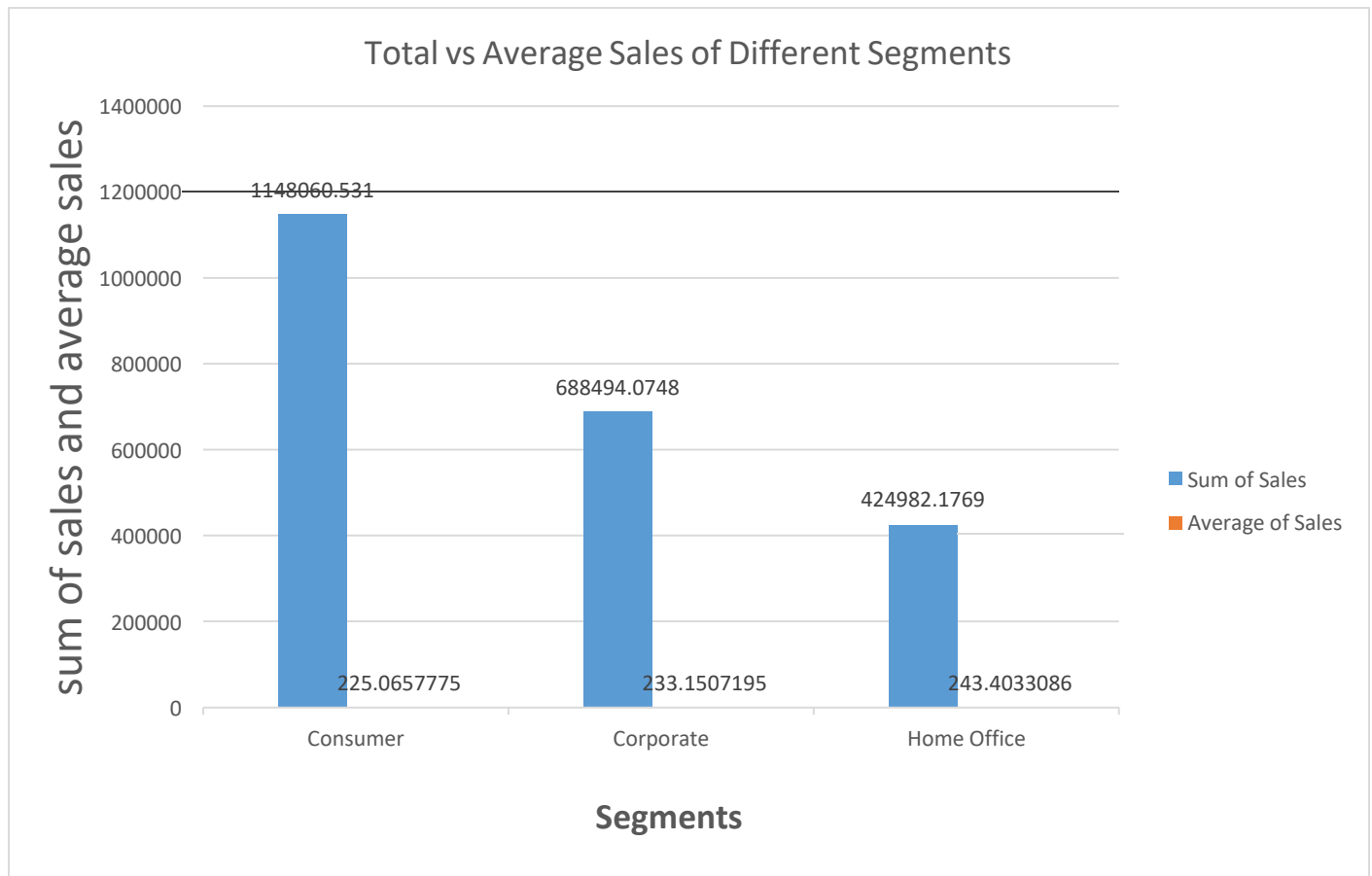
Ans. By Analysis of the given data set we can found that in all the three segments the total sales were greater than the average sales.

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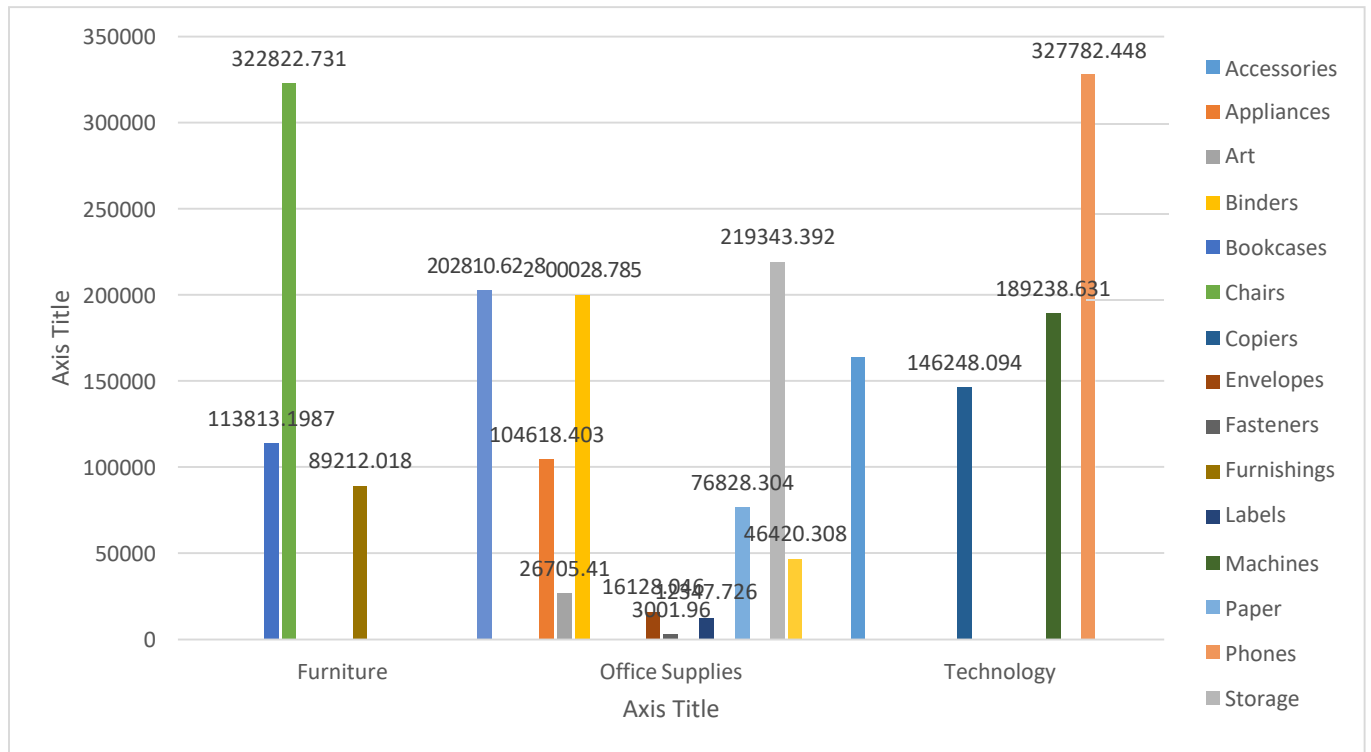
Ans. By doing analysis of the given Order Sales dataset we were able to observe that, average sales of Technology was far greater than rest of the categories

Regression and ANOVA:

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

This regression analysis aims to examine the relationship between two variables: an independent variable represented by "Postal Code" and a dependent variable (not explicitly mentioned in the output). Here's an explanation of the key components:

1. Regression Equation:

The regression equation is of the form:

$Y = 232.38 + 0.000167458*$ (Postal Code)

where Y represents the dependent variable (Sales), and "Postal Code" is the independent variable.

2. Interpretation of Coefficients:

The intercept coefficient (232.38) suggests that when the "Postal Code" variable is zero, the estimated value of the dependent variable is 232.38. However, the interpretation of this intercept may not be meaningful since postal codes are unlikely to be zero.

The coefficient for "Postal Code" (0.000167458) suggests that for every one-unit increase in the postal code, the estimated value of the dependent variable increases by approximately

0.000167458 units. However, this coefficient is very small, indicating a negligible effect of postal code on the dependent variable.

3. Statistical Significance:

The p-value associated with the coefficient for "Postal Code" is 0.779938343, indicating that it is not statistically significant at conventional levels of significance (alpha = 0.05). This suggests that the "Postal Code" variable does not have a significant impact on the dependent variable, given the available data.

4. Goodness of Fit:

- The R-squared value (0.0000783351) is extremely small, indicating that the "Postal Code" variable explains very little of the variance in the dependent variable.

- The Adjusted R-squared value (-0.000924595) is negative, which can happen when the model is over fit or when the independent variable is not relevant. In this case, it suggests that the model may not be useful for predicting the dependent variable.

5. ANOVA:

- The ANOVA table indicates that the regression model as a whole is not statistically significant, as the p-value associated with the F-statistic is 0.779938343.

6. Standard Error:

- The standard error (596.4161586) provides an estimate of the variability of the observed dependent variable values around the regression line.

7. Observations:

- The analysis is based on a sample of 999 observations.

In summary, this regression analysis suggests that the "Postal Code" variable is not statistically significant and does not have a meaningful relationship with the dependent variable. Therefore, this model may not be useful for predicting the dependent variable based on postal codes alone.

Correlation:

The absolute value of the correlation coefficient (0.024067424) is close to zero. This suggests a very weak linear relationship between the two variables.

Descriptive Statistics:

Sales	
Mean	230.7691
Standard Error	6.33014
Median	54.49
Mode	12.96
Standard Deviation	626.6519
Sample Variance	392692.6
Kurtosis	304.4451
Skewness	12.98348
Range	22638.04
Minimum	0.444
Maximum	22638.48
Sum	2261537
Count	9800

4. 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.

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

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

Loan Data Report

Introduction:

Dataset Overview:

Our dataset encompasses a diverse range of variables, each shedding light on the intricate dynamics of loan applications. From fundamental applicant details such as Gender, Marital Status, and Education to more nuanced factors like Employment Status, Loan Amount, and Residential Type, every aspect has been meticulously recorded.

Key Attributes:

1. Gender: A demographic identifier providing insights into the gender distribution among loan applicants.
2. Marital Status (Married, Not Married): Categorization based on marital status aiding in demographic segmentation.
3. Education (Graduate, Non-graduate): Classification based on educational background for further analysis.
4. Employment Status (Employed, Unemployed): Distinction between employed and unemployed applicants, crucial for risk assessment.
5. Loan Amount: The principal amount applied for, providing a measure of financial need and capacity.
6. Residential Type (Urban, Semi-urban, Rural): Geographic classification enabling analysis across different residential 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?
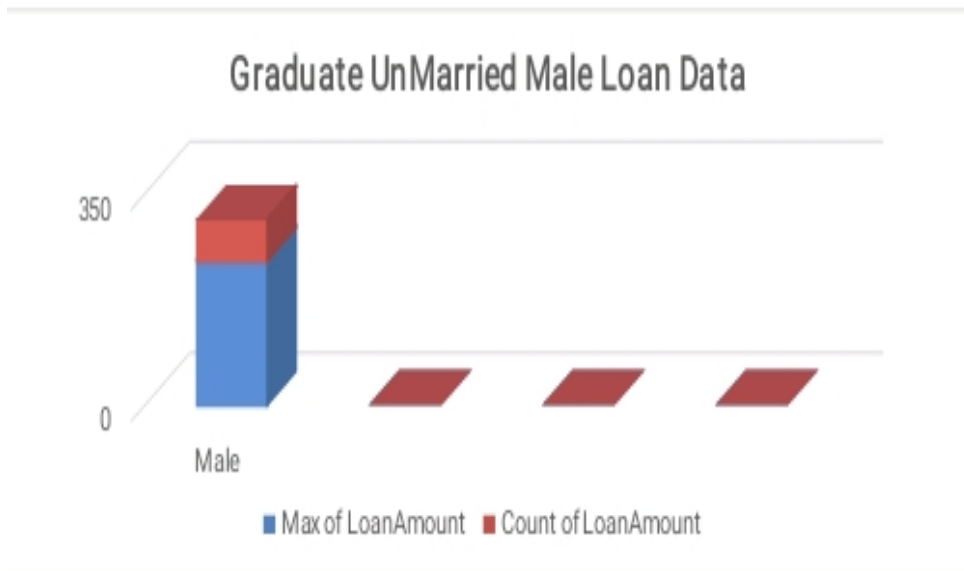
Graduate UnMarried Male Loan Data

350

0

Male

■ Max of LoanAmount   ■ Count of LoanAmount
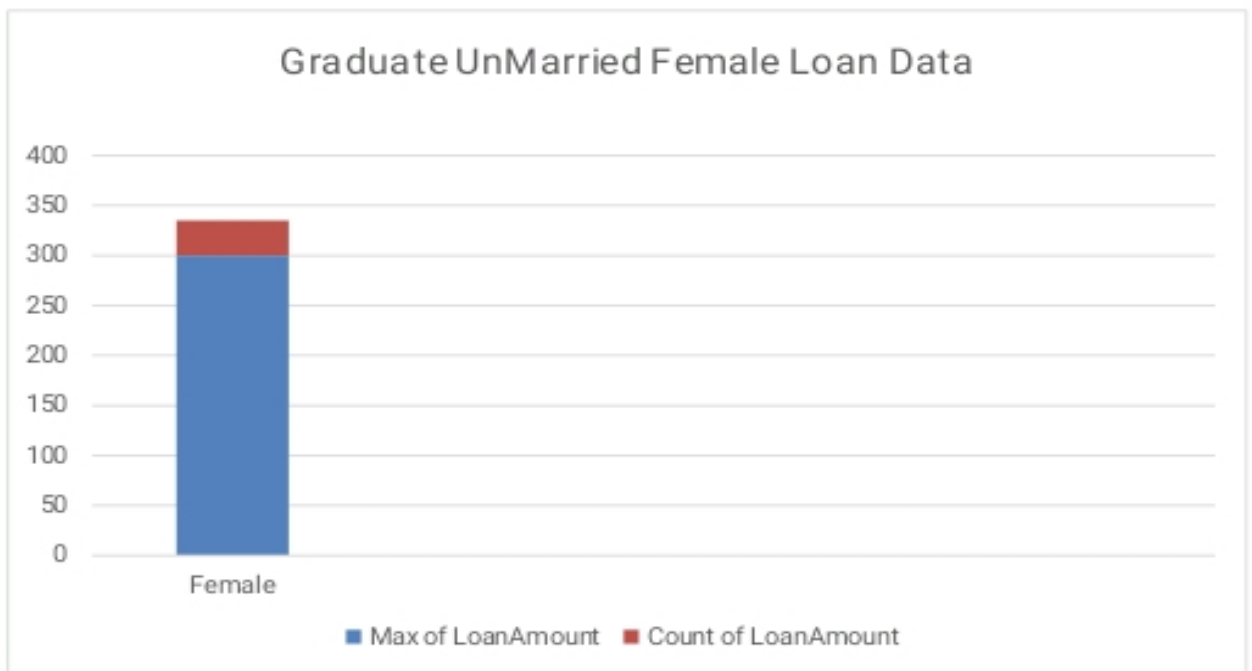
Gender		Married		Education	
Female		No		Graduate	
Male		Yes		Not Graduate	
(blank)		(blank)		(blank)	

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

Graduate UnMarried Female Loan Data

■ Max of LoanAmount ■ Count of LoanAmount

Gender ✔≡ ▼x	Married ✔≡ ▼x	Education ✔≡ ▼x
Female	No	Graduate
Male	Yes	Not Graduate
(blank)	(blank)	(blank)

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

NonGraduate UnMarried Male Loan Data

Gender	Married	Education
Male	No	Not Graduate
(blank)	(blank)	(blank)

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

Graduate Married Female Loan Data

■ Max of LoanAmount ■ Count of Gender
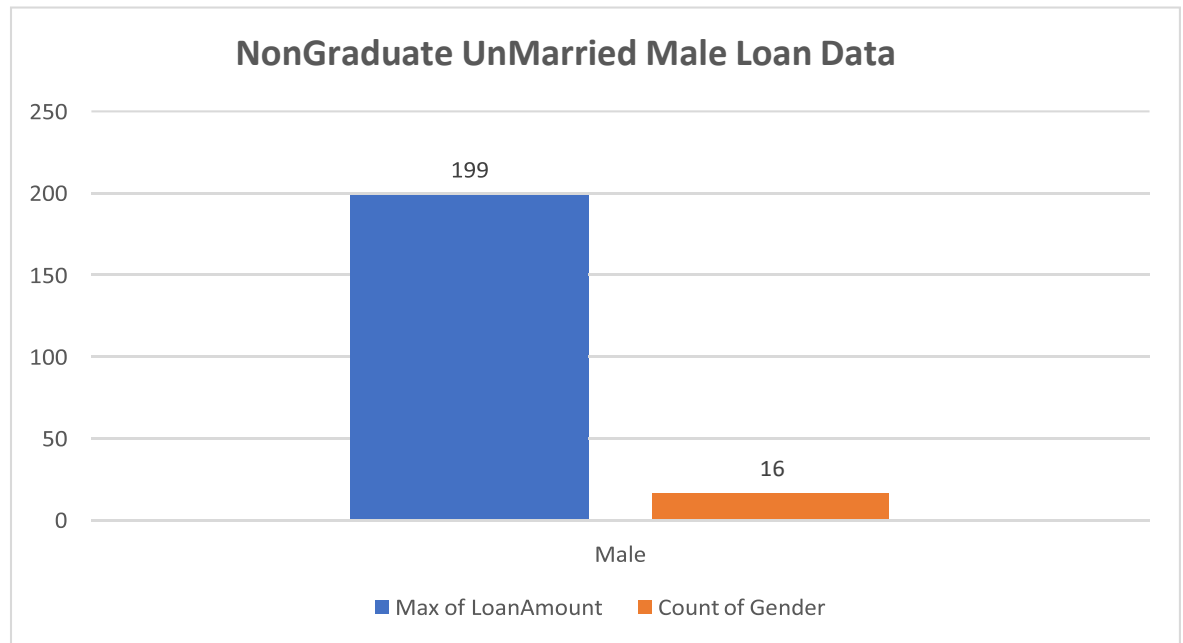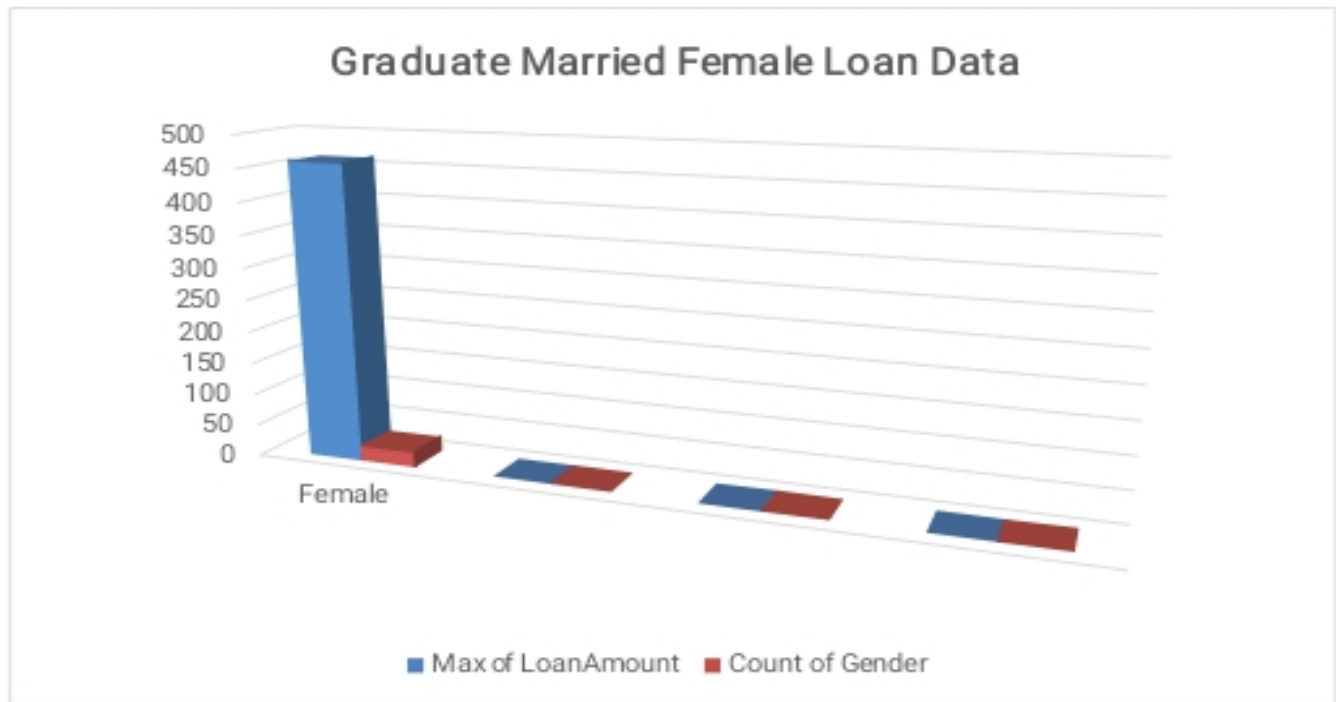
Gender	Married	Education
Female	Yes	Graduate

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

Comparision of Loan Amount on basis of Property Area

6000		
5000		
4000		
3000		
2000		
1000		
0		
Rural	Semiurban	Urban

☐ Series 1 ☐ Series 2
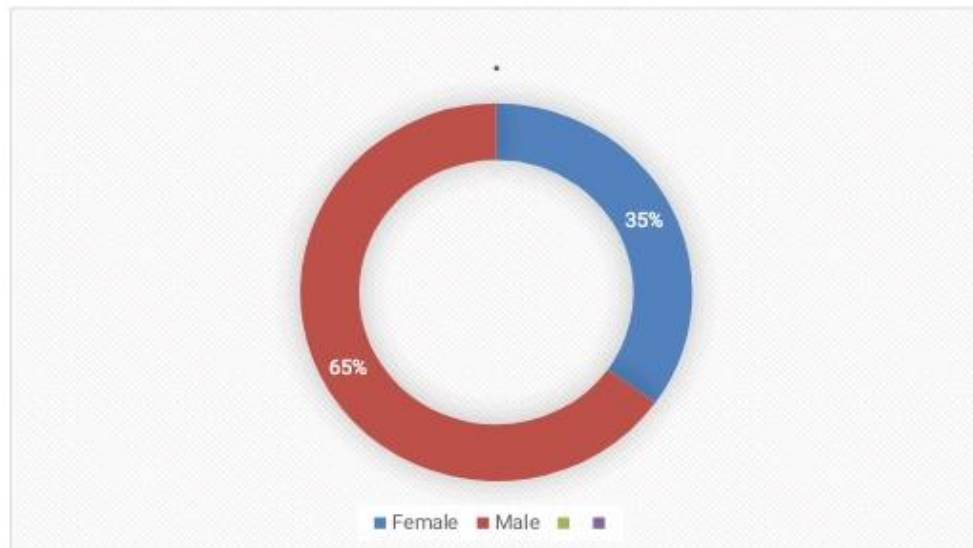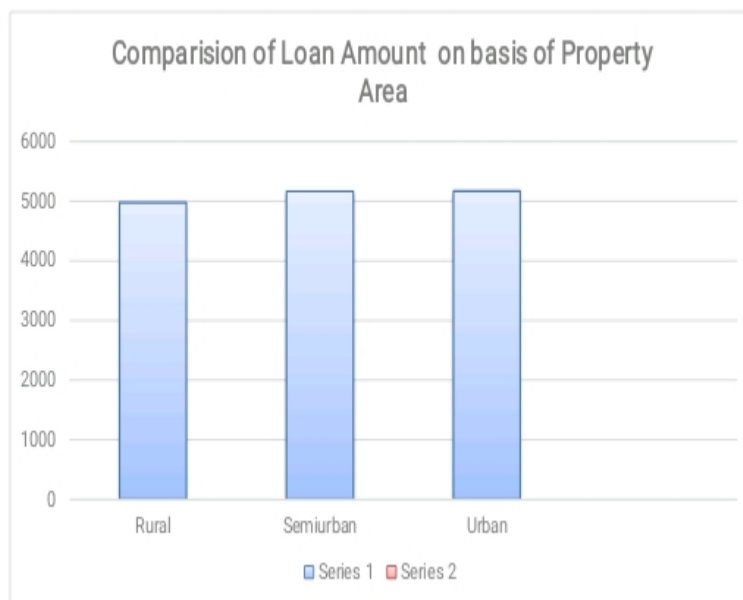
Property_Area

- Rural
- Semiurban
- Urban
- (blank)

LoanAmount

- 28
- 40
- 46
- 50
- 55
- 59
- 60
- 64

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	98.01439
5720	0.0058851	0.00059692	9.85902887	1.7676E-20	0.00471125	0.00705895	0.004711

Co-Relation:

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	176365 5	4805.59945 5	24114831.0 9
CoapplicantIncom e	367	576035	1569.57765 7	5448639.49 1
LoanAmount	367	49280	134.277929 2	3964.14112 4

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	4202537452	2	2101268726	213.200984 1	5.87569E- 79	3.00392057 7
Within Groups	1082168110 7	1098	9855811.57 3			
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%)	504.0756 067	Confidence Level(95.0%)	239.6059 543	Confidence Level(95.0%)	6.462910 219

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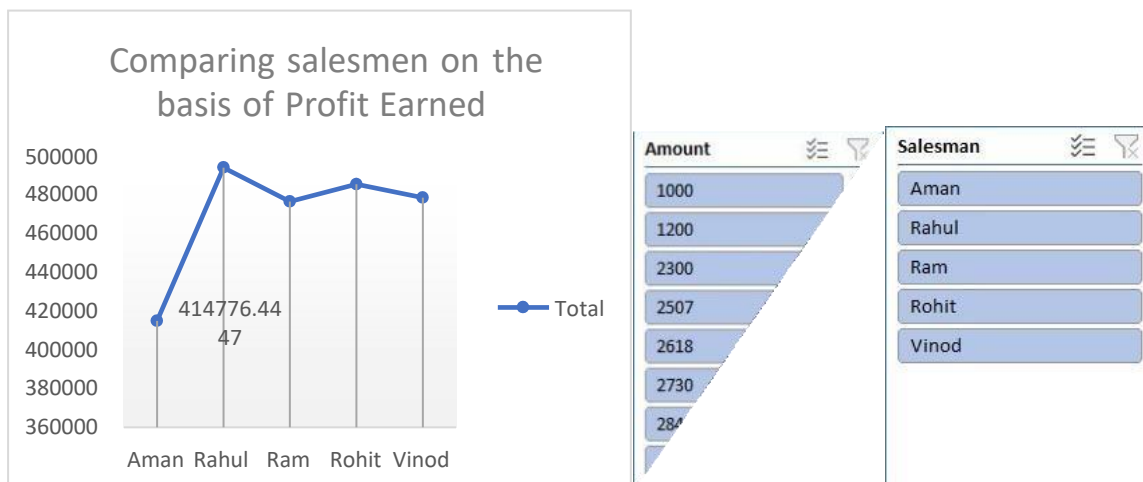
2. Find out most sold product over the period of May-September.

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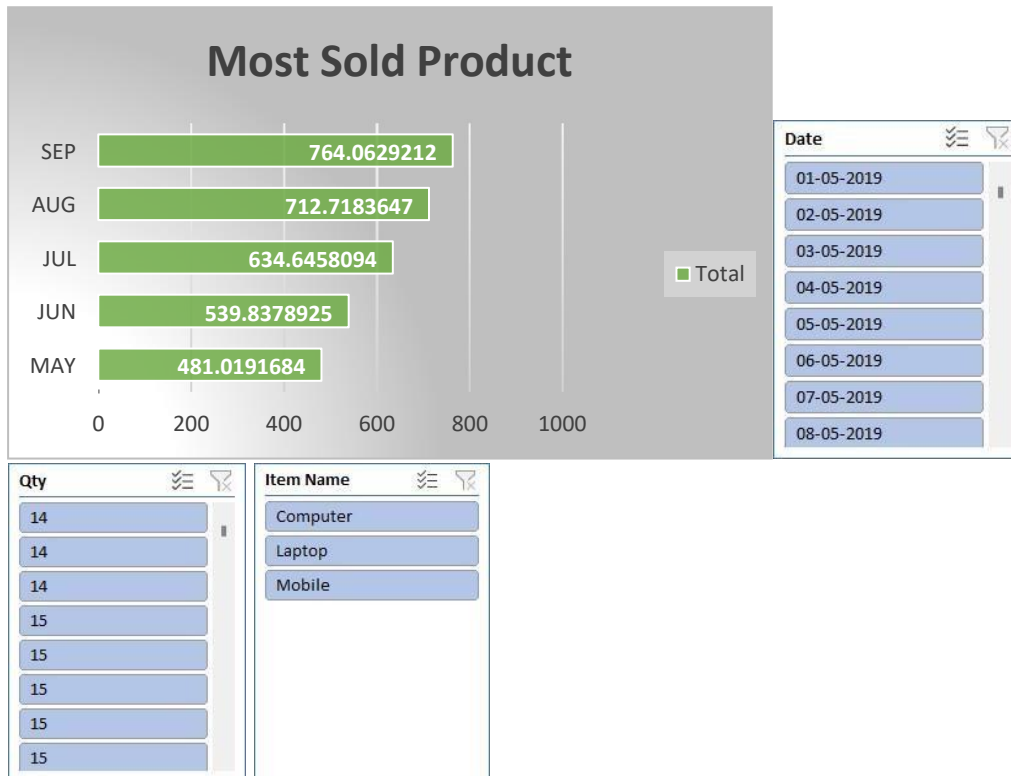
3. Find out which of the two product sold the most over the year Computer or Laptop?

Ans:- The two product sold the most over the year between computer or laptop :
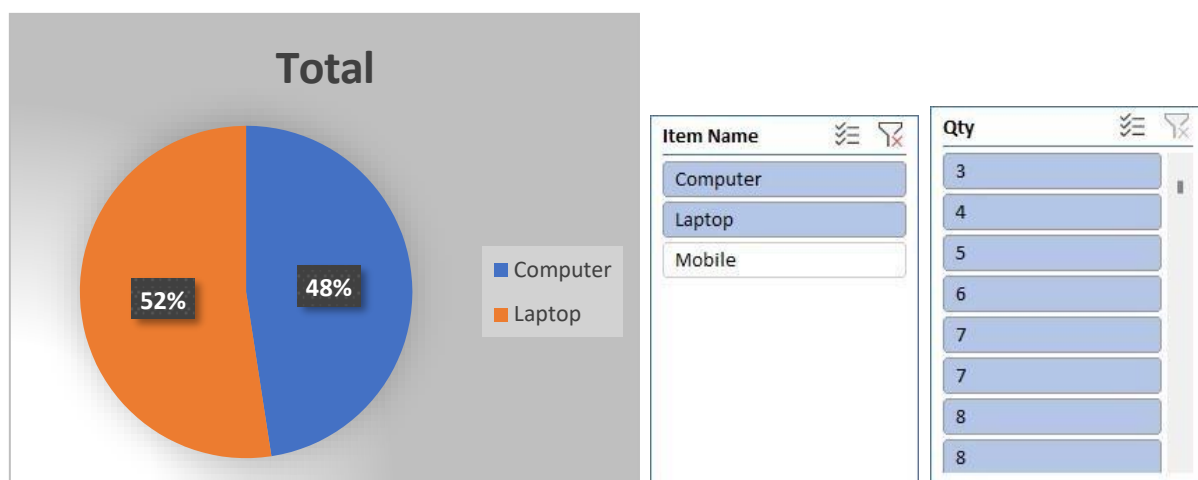
4 . Which item yield most average profit?

Ans:- The item that yields the most profit between laptop, computer and mobile is :
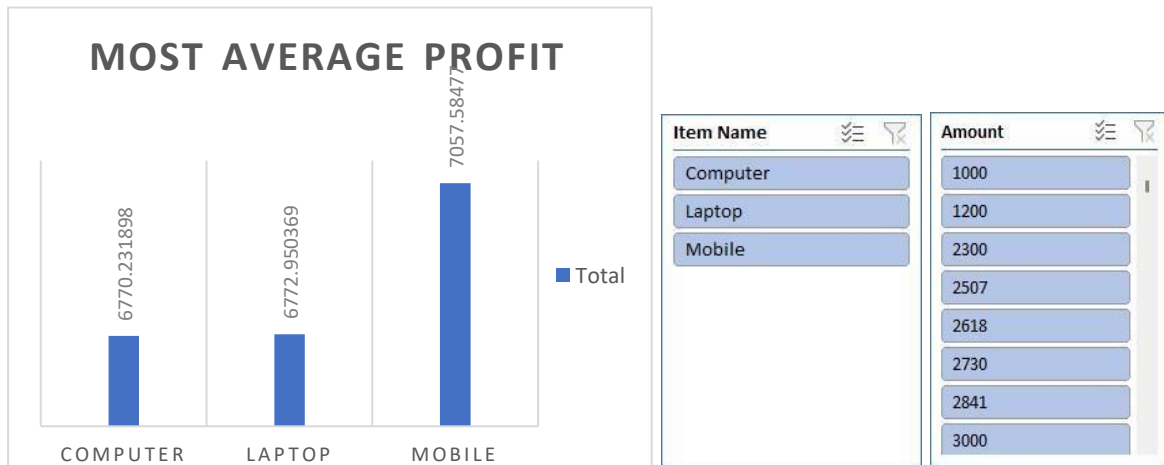
5. Find out average sales of all the products and compare them.

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.

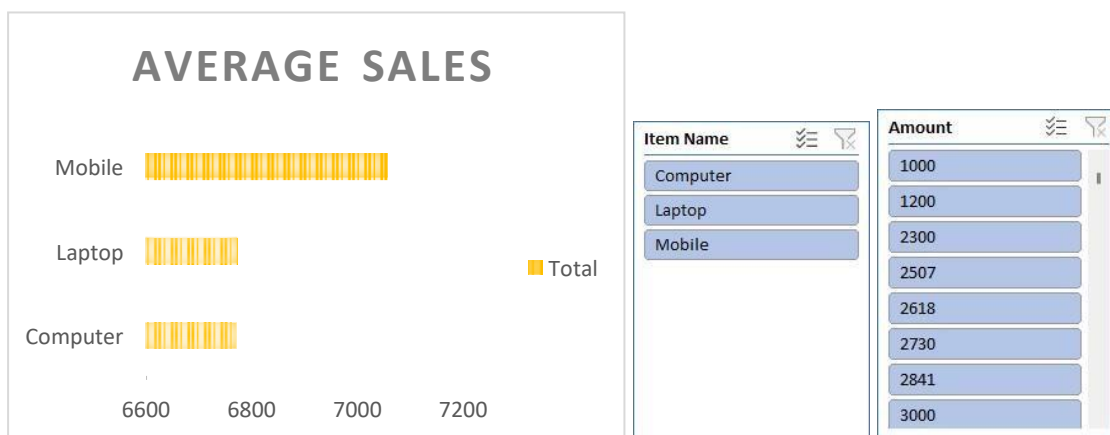The dataset is well-structured and provides comprehensive information on sales transactions. It allows for various analyses, but could benefit from additional variables for deeper insights.Overall, it's a valuable resource for understanding sales dynamics and informing business decisions.

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

Sales Data Samples Report

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.
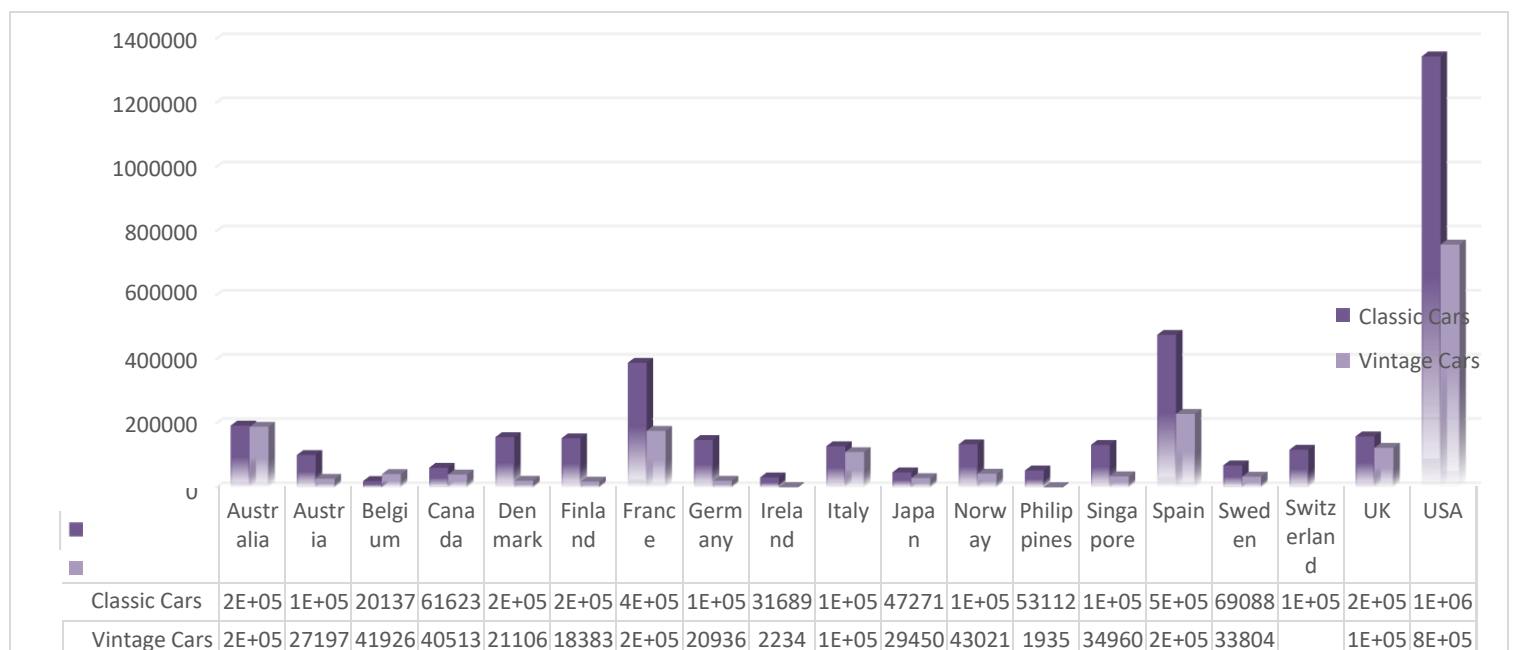
Questionnaire:

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:-The comparsion of sale of Vintage cars and Classic cars for all the countries is given below:-

|	Austr alia	Austr ia	Belgi um	Cana da	Den mark	Finla nd	Franc e	Germ any	Irela nd	Italy	Japa n	Norw ay	Philip pines	Singa pore	Spain	Swed en	Switz erlan d	UK	USA
Classic Cars	2E+05	1E+05	20137	61623	2E+05	2E+05	4E+05	1E+05	31689	1E+05	47271	1E+05	53112	1E+05	5E+05	69088	1E+05	2E+05	1E+06
Vintage Cars	2E+05	27197	41926	40513	21106	18383	2E+05	20936	2234	1E+05	29450	43021	1935	34960	2E+05	33804		1E+05	8E+05

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

Ans:

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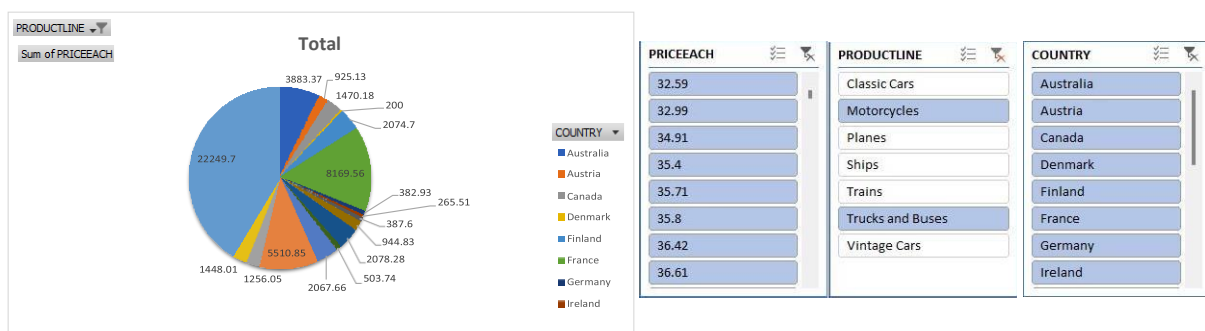
4. Compare sales of all the items for the years of 2004, 2005.

Ans: - The 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. The comparison of all the countries based on deal size are:

Regression and Anova

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.657840928					
R Square	0.432754687					
Adjusted R Square	0.432553607					
Standard Error	1387.45926					
Observations	2823					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	4142995200	4142995200	2152.157001	0	
Residual	2821	5430546866	1925043.199			
Total	2822	9573542065				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1470.590019	111.4099971	-13.19980305	1.20143E-38	-1689.043329	-1252.13671
PRICE EACH	60.05936566	1.294624334	46.39134619	0	57.52085944	62.59787188

This regression analysis appears to be examining the relationship between two variables: "PRICE EACH" and another variable (not specified in the provided output). Here are the results:

1. **Regression Equation:** The regression equation can be written as: $Y = -1470.59$ (PRICE EACH)$+60.06$where:

   - Y represents the dependent variable Quantity.

   - X represents the independent variable "PRICE EACH".

2. **Interpretation of Coefficients:**

   - The intercept coefficient (-1470.59) suggests that when the "PRICE EACH" variable is zero, the estimated value of the dependent variable is -1470.59. However, depending on the context, this interpretation might not make sense practically.

   - The coefficient for "PRICE EACH" (60.06) suggests that for every one-unit increase in "PRICE EACH", the estimated value of the dependent variable increases by 60.06 units.

3. **Statistical Significance:**

   - The p-value associated with the coefficient for "PRICE EACH" is 00, indicating that the coefficient is statistically significant at conventional levels of significance (typically $\diamond = 0.05 \alpha = 0.05$).

   - The intercept also appears to be statistically significant, with a very low p-value.

4. **Goodness of Fit:**

   - The R-squared value (0.433) indicates that approximately 43.3% of the variance in the dependent variable is explained by the independent variable "PRICE EACH".

   - The adjusted R-squared value (0.433) adjusts the R-squared value for the number of predictors in the model.

5. **ANOVA:**

   - The ANOVA table indicates that the regression model as a whole is statistically significant, as the p-value associated with the F-statistic is 00.

6. **Standard Error:**

   - The standard error (1387.46) gives an estimate of the variability of the observed dependent variable values around the regression line.

7. **Observations:**

   - The analysis is based on a sample of 2823 observations.

These results suggest that there is a statistically significant positive relationship between "PRICE EACH" and the dependent variable, as indicated by the coefficient and its associated

p-value. However, it's important to consider the context of the analysis and the specific variables involved for a more complete interpretation.

CORELATION:

The correlation coefficient you calculated (0.657840928) represents the strength. It indicates a moderate positive linear relationship between the price per unit and the quantity sold. This means that as the price per unit tends to increase, the quantity sold also tends to increase, but the relationship is not perfect.

Descriptive Statistics:

SALES	
Mean	3553.889072
Standard Error	34.66589212
Median	3184.8
Mode	3003
Standard Deviation	1841.865106
Sample Variance	3392467.068
Kurtosis	1.792676469
Skewness	1.161076001
Range	13600.67
Minimum	482.13
Maximum	14082.8
Sum	10032628.85
Count	2823

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.