ACROPOLIS INSTITUTE OF TECHNOLOGY & RESEARCH, INDORE

DEPARTMENT OF COMPUTER SCIENCE

CS-605 Data Analytics Lab

3rd Year 6th Semester

2023-2024

SUBMITTED BY -

Chainika Darekar

(0827CS211062)

SUBMITTED TO -

Prof. Anurag Punde

S.No.	Experiment	Remarks
1.	Data Analysis Questions: ● Data Analysis Principles ● Statistical Analytics ● Hypothesis Testing ● Regression ● Correlation ● ANOVA	
2.	Dashboards: ● Car Collection Data ● Order Data ● Cookie Data ● Loan Data ● Shop Sale Data Report ● Sales Data Sample ● Store Dataset	
3.	Reports: ● Car Collection Data ● Order Data ● Cookie Data ● Loan Data ● Shop Sales Data ● Sales Data Sample ● Store Dataset	
4.	ZOMT HISTORICAL DATA ANALYSIS (19 April 2024 - 19 May 2024)	

Assignment-1

Data Analysis Principle:

Data Analysis Principles involve systematically applying statistical and logical techniques to describe, condense, and evaluate data. Key principles include understanding the data's source, context, and quality, cleaning the data to remove errors, exploring the data using descriptive statistics and visualization techniques, modeling the data with statistical models for predictions or inferences, and interpreting results to draw meaningful conclusions and make informed decisions. This structured approach ensures accurate and actionable insights from the data.

Statistical Analysis:

Statistical Analytics employs statistical methods to collect, review, analyze, and draw conclusions from data. This involves using descriptive statistics (such as mean, median, mode, range, variance, and standard deviation) to summarize and describe data features. Inferential statistics, including hypothesis testing, confidence intervals, and regression analysis, are used to make generalizations and extend conclusions beyond the immediate dataset. Predictive analytics leverages statistical models to forecast future outcomes, while prescriptive analytics recommends specific actions based on data insights. Additionally, Statistical Analytics emphasizes data visualization techniques to enhance the interpretation and communication of data findings, ensuring that complex data is accessible and understandable. The use of advanced statistical software and algorithms also facilitates the handling of large datasets, improving the accuracy and efficiency of the analysis.

Hypothesis Testing:

Hypothesis Testing is a statistical method used to make decisions based on data from experiments or studies. It involves formulating a null hypothesis (H0), which represents no effect or difference, and an alternative hypothesis (H1), which indicates an effect or difference. The p-value measures the probability of obtaining the observed data if H0 is true; a small p-value suggests strong evidence against H0. Errors can occur in this process: Type I errors (false positives) happen when H0 is incorrectly rejected, while Type II errors (false negatives) occur when H0 is incorrectly not rejected. The significance level (α), typically set at 0.05, serves as the threshold for rejecting H0. Hypothesis testing also considers the power of a test, which is the probability of correctly rejecting H0 when H1 is true, ensuring that the test is sensitive to detecting actual effects.

Regression:

Regression analysis is a statistical technique used to understand the relationships between dependent and independent variables. Linear regression fits a linear equation to the data to model the relationship between one independent variable and a dependent variable. Multiple regression extends this by incorporating multiple independent variables to predict the dependent variable. Logistic regression is used to predict probabilities for categorical outcomes, typically binary, by modeling the log odds of the dependent variable. Polynomial regression models relationships as nth degree polynomials, allowing for more complex, non-

linear relationships between the variables. Additionally, regression analysis can be used to identify trends, make predictions, and infer causal relationships, providing valuable insights for decision-making and strategic planning.

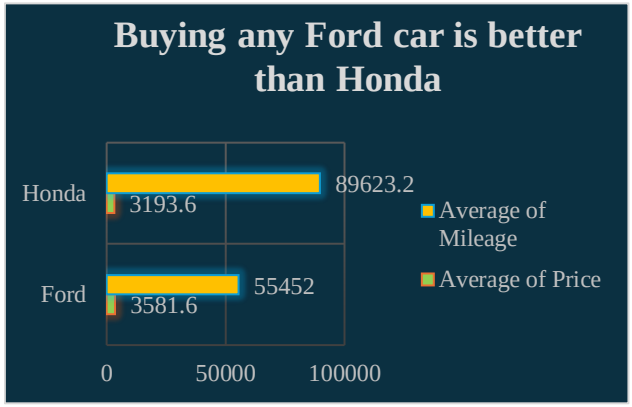
Correlation:

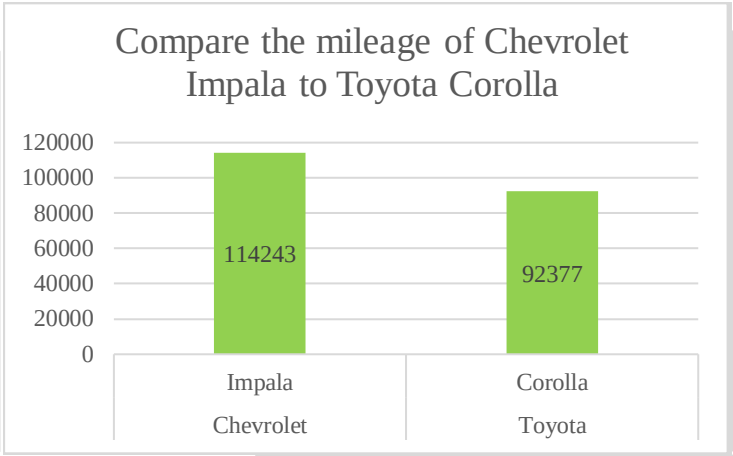
Correlation measures the strength and direction of the relationship between two variables using the correlation coefficient (r), which ranges from -1 to 1. A positive correlation indicates that both variables move in the same direction, while a negative correlation means that as one variable increases, the other decreases. No correlation suggests there is no relationship between the variables. It's important to note that correlation does not imply causation; it merely shows that a relationship exists between the variables without indicating that one causes the other.

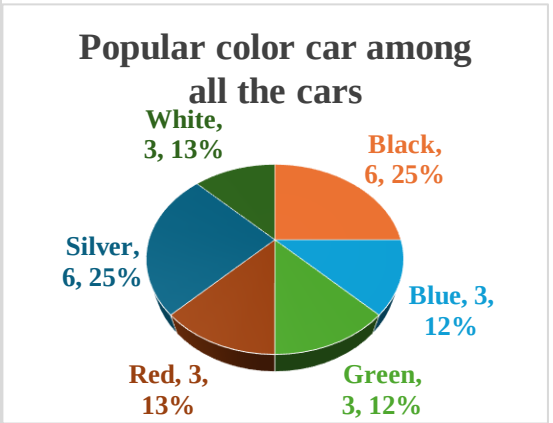
Anova:

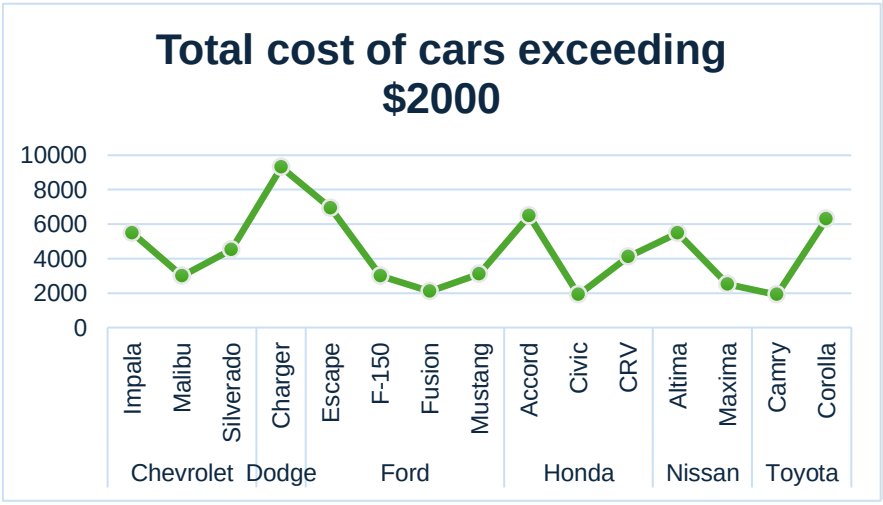
ANOVA (Analysis of Variance) is a statistical method used to compare means across multiple groups to determine if there are significant differences among them. One-way ANOVA compares means across one factor with multiple levels, while two-way ANOVA examines the influence of two categorical variables and their interaction. ANOVA relies on several assumptions: normality (data follows a normal distribution), homogeneity of variances (equal variances among groups), and independence of observations. The F-statistic, which is the ratio of variance between group means to the variance within groups, is used to determine the p-value for the test. A significant p-value indicates that at least one group mean is significantly different from the others.

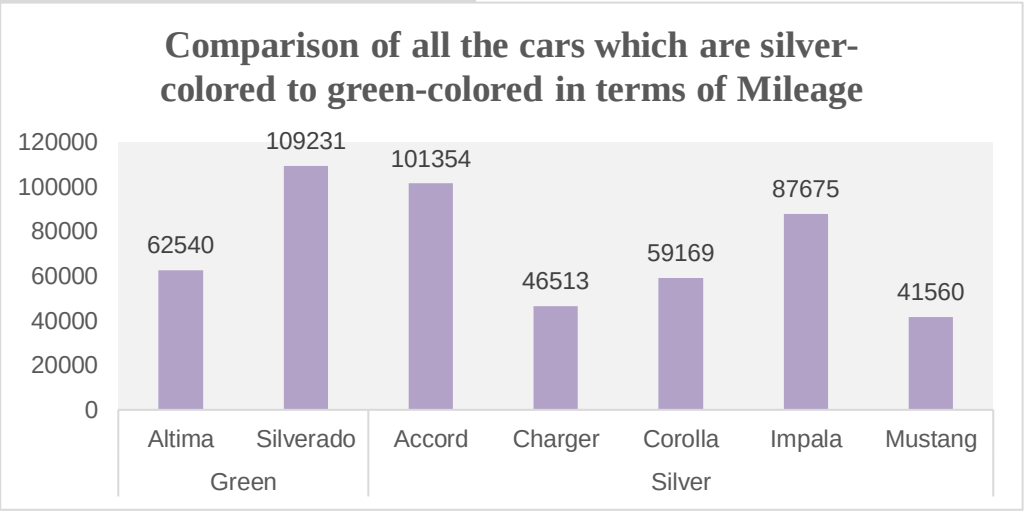
Dashboard : Car Collection Data



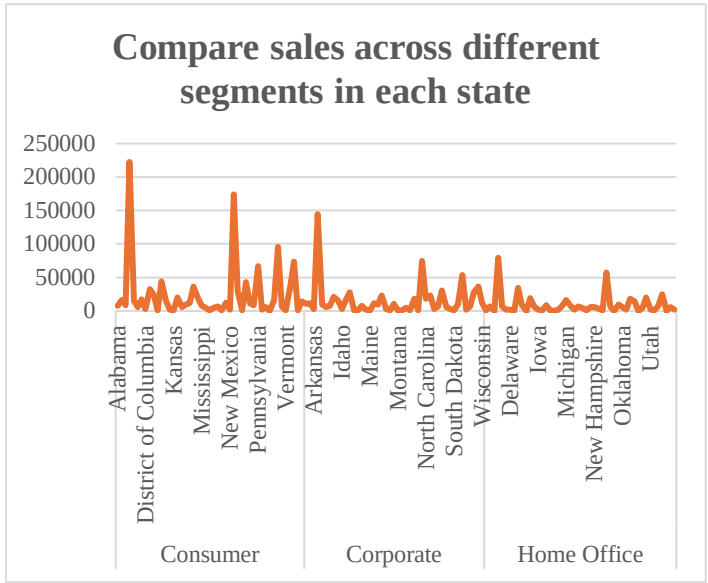


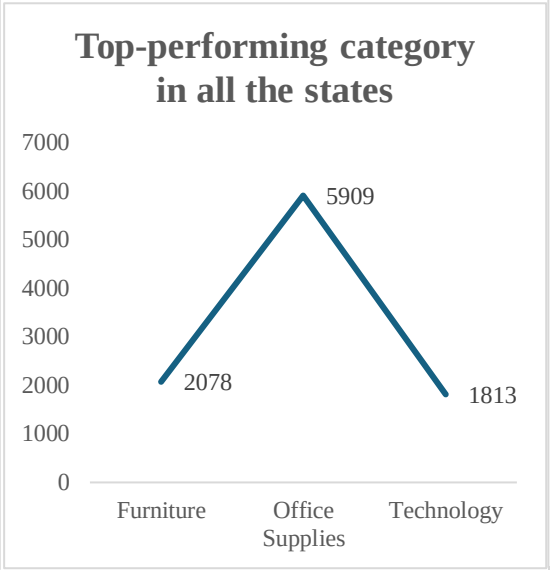


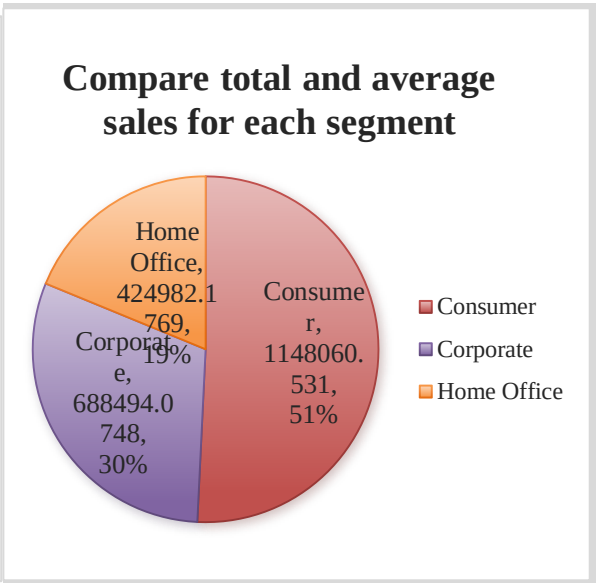


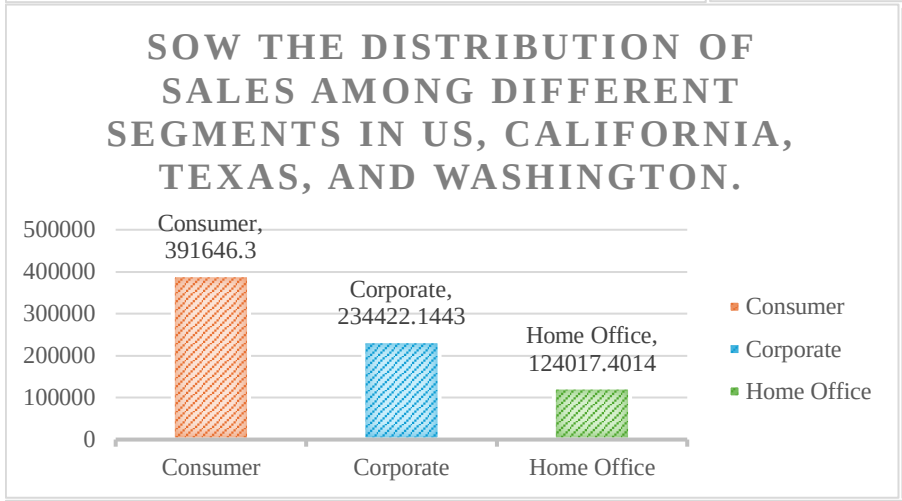


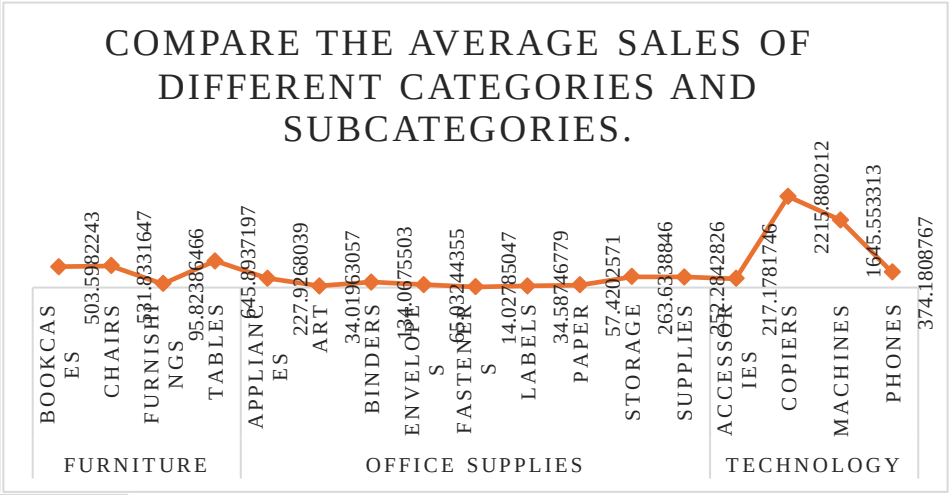
Dashboard : Order Data



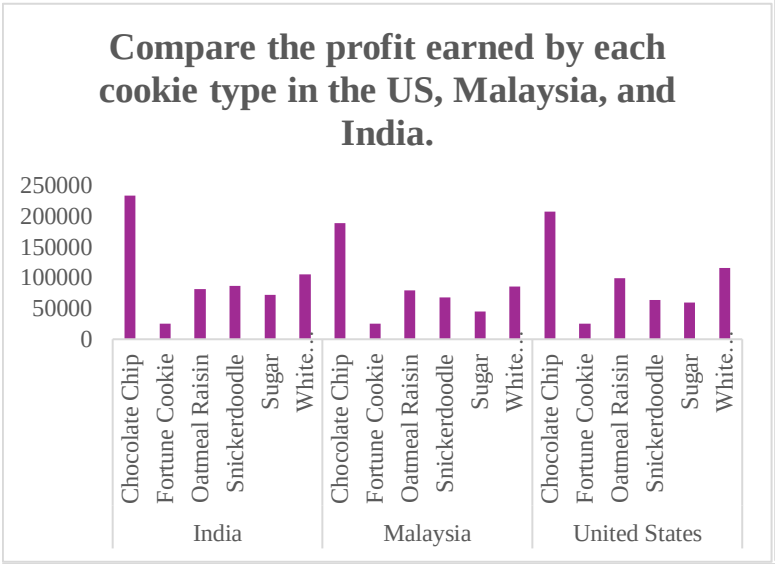


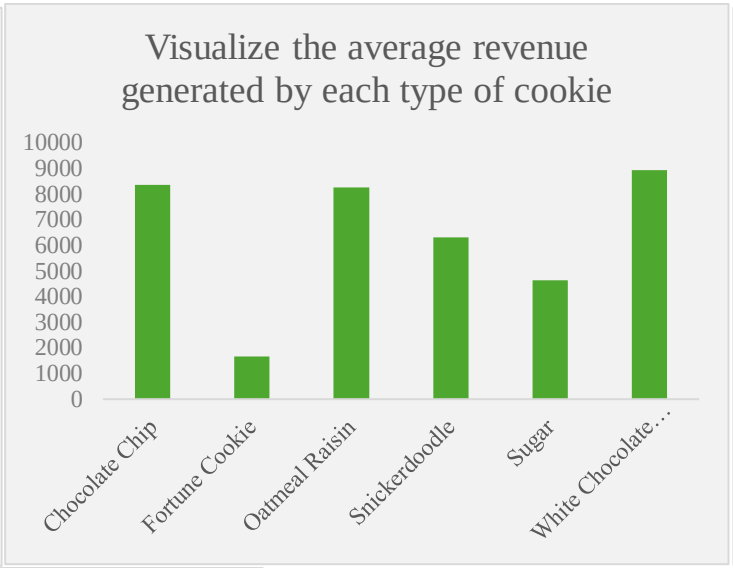


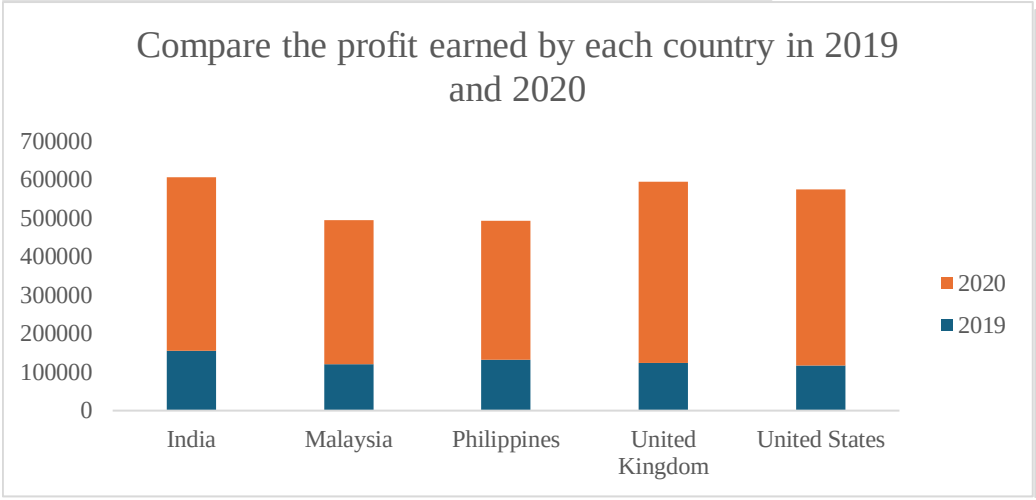


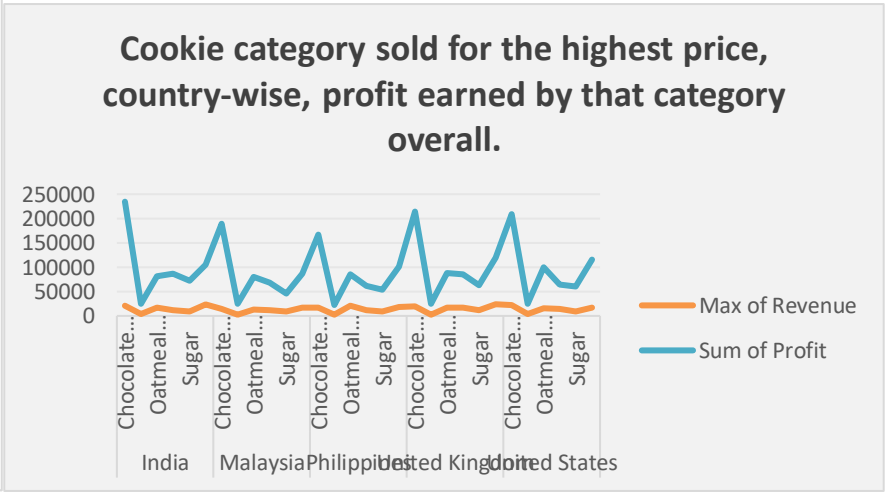


Dashboard : Cookie Data

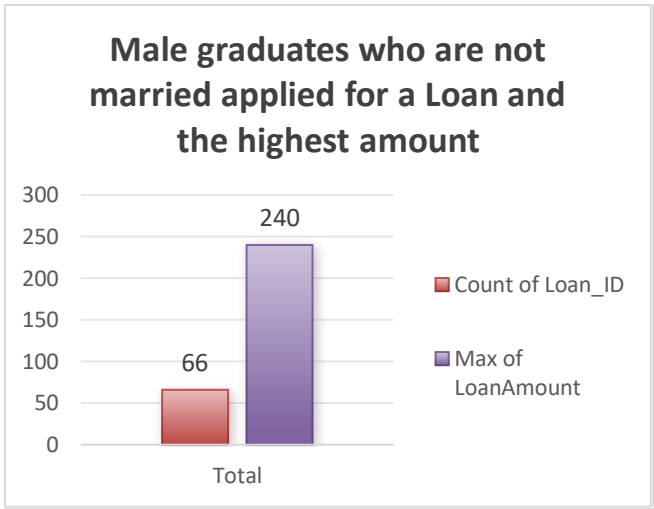


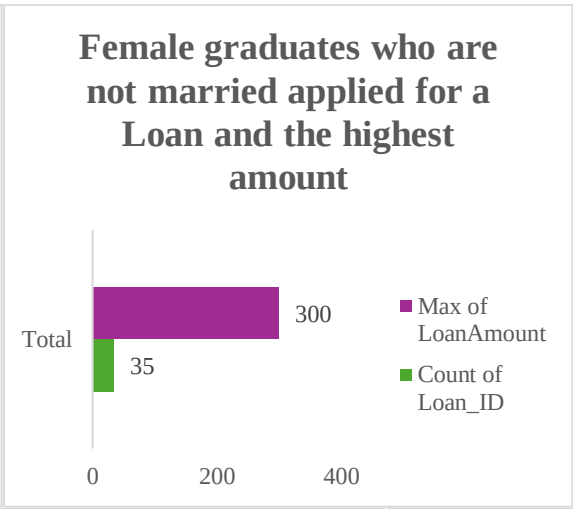


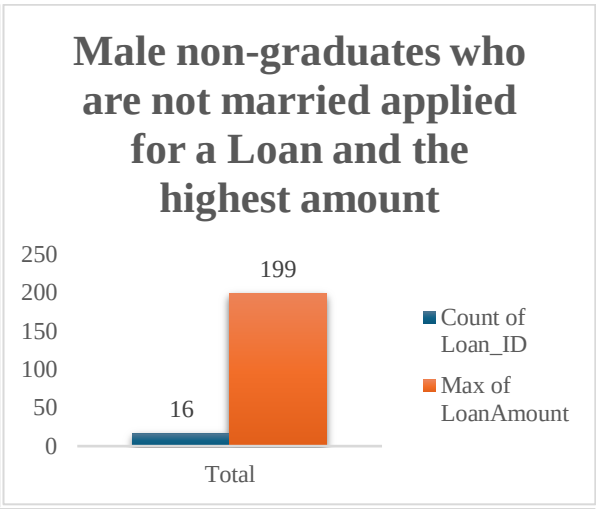


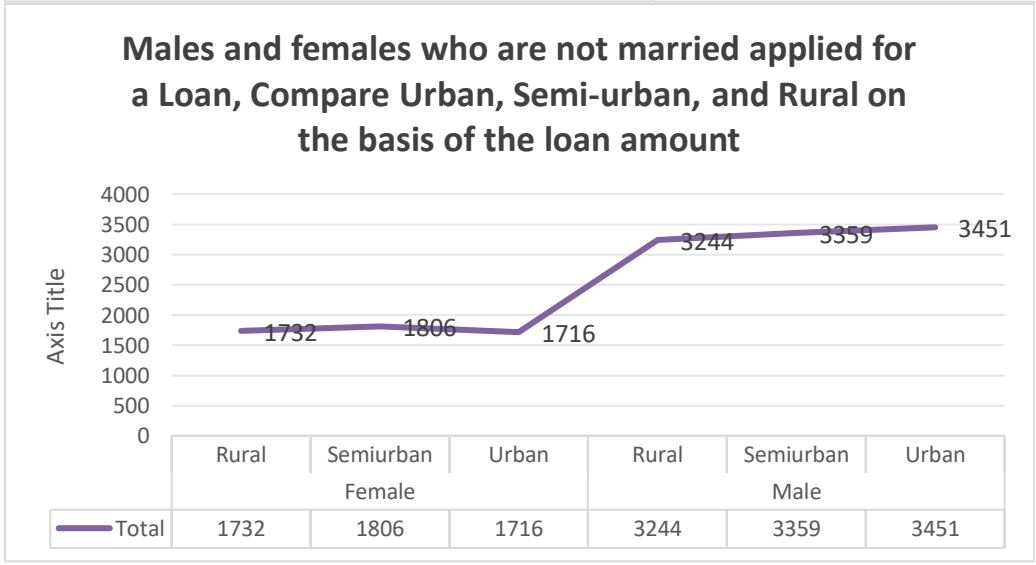


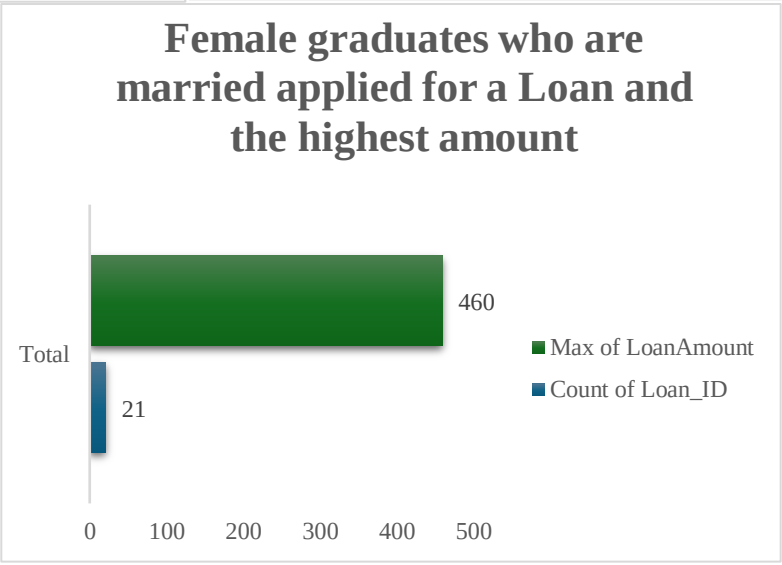
Dashboard : Loan Data



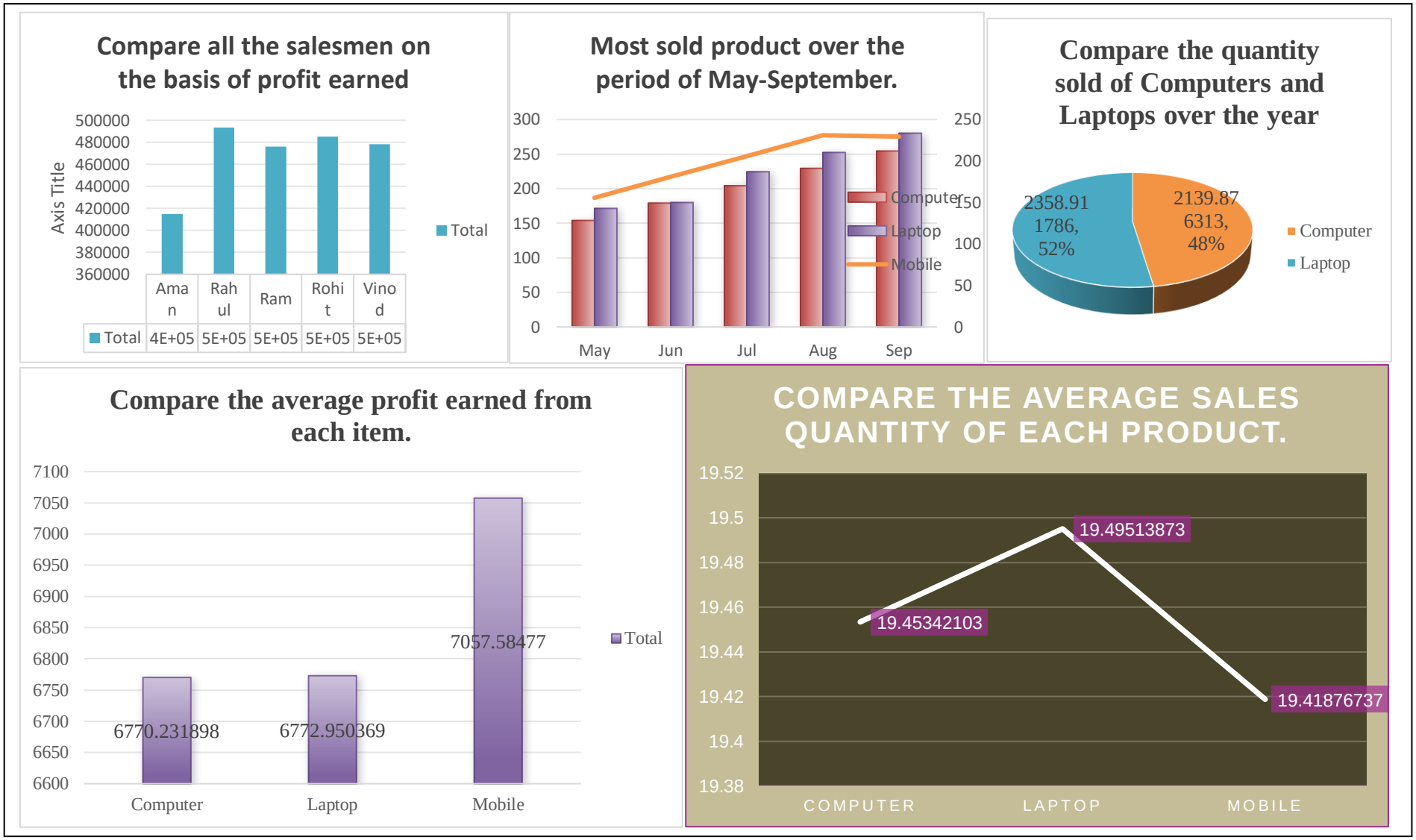




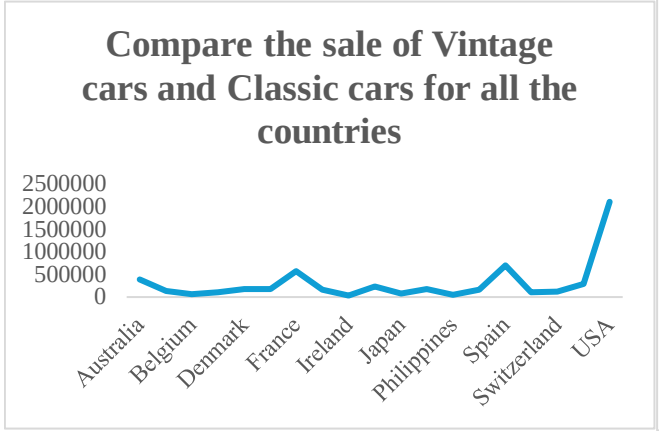


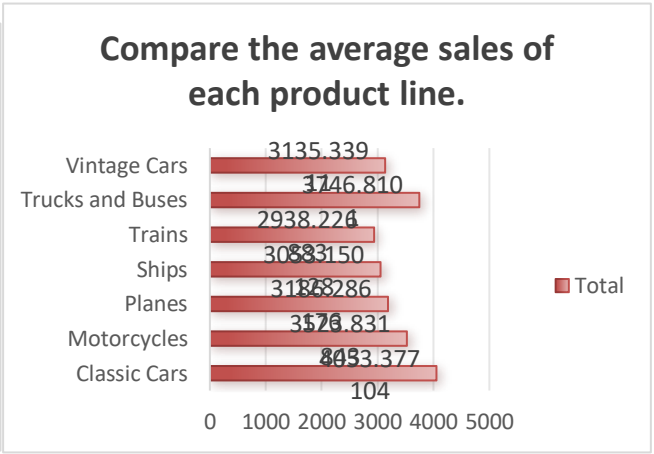


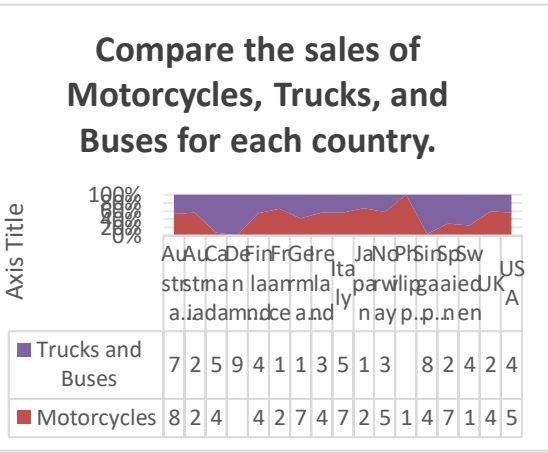
Dashboard : Shop Sales Data

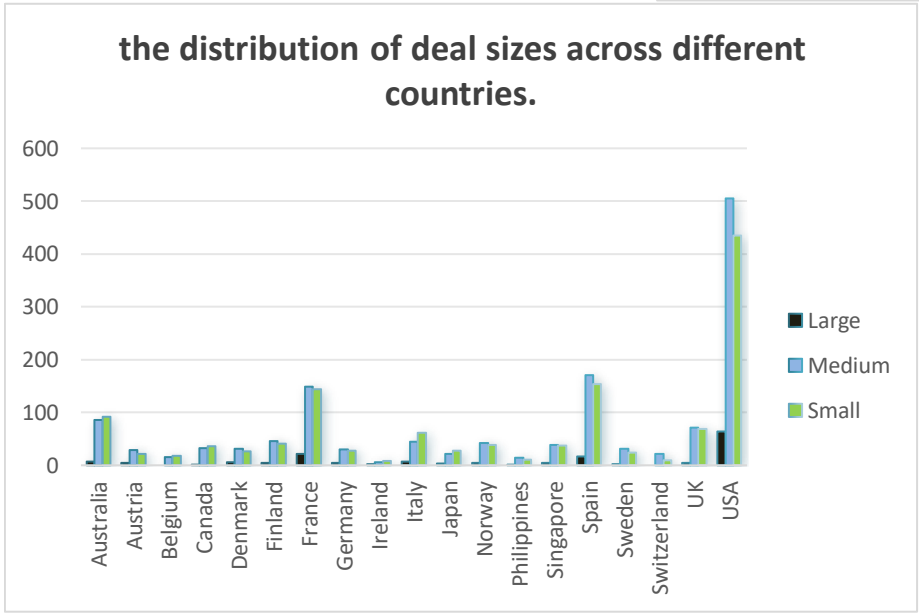


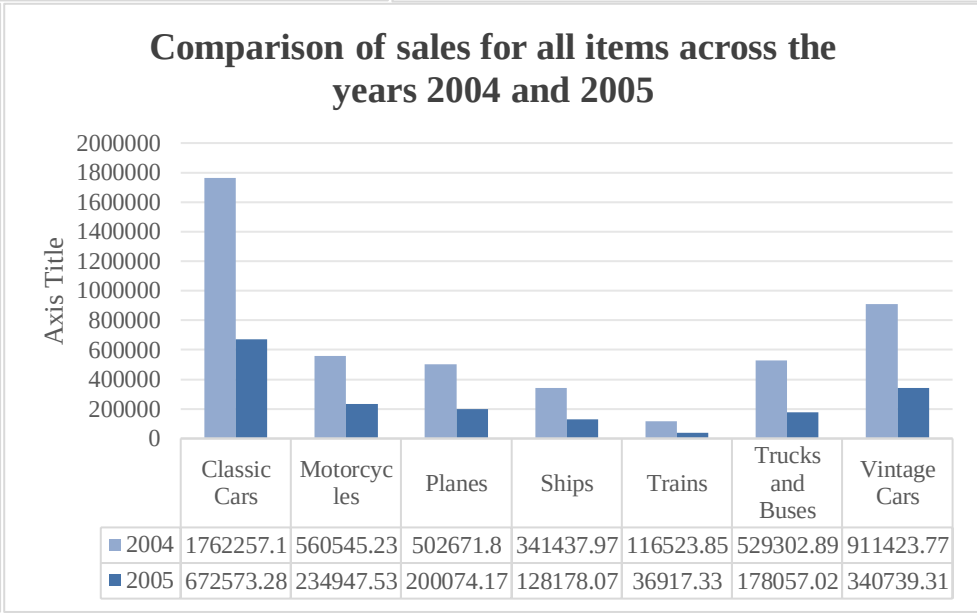
Dashboard : Sales Data Sample



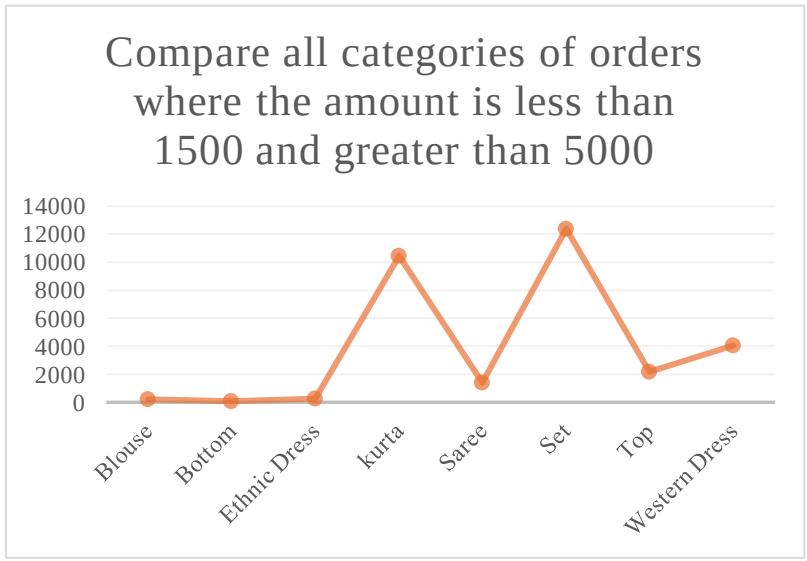


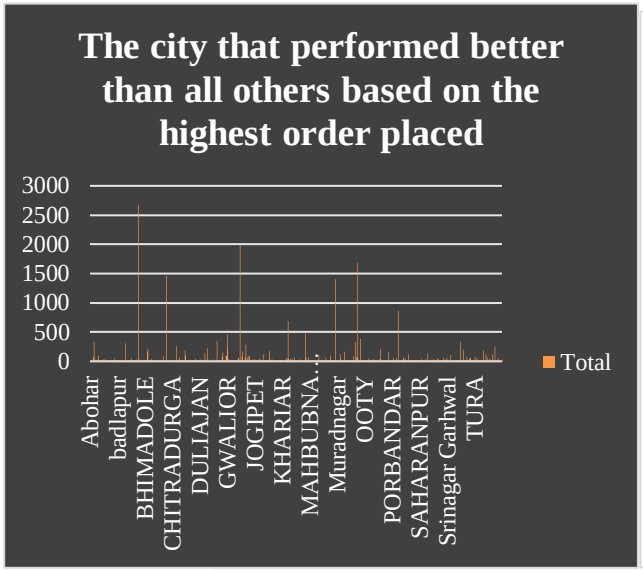


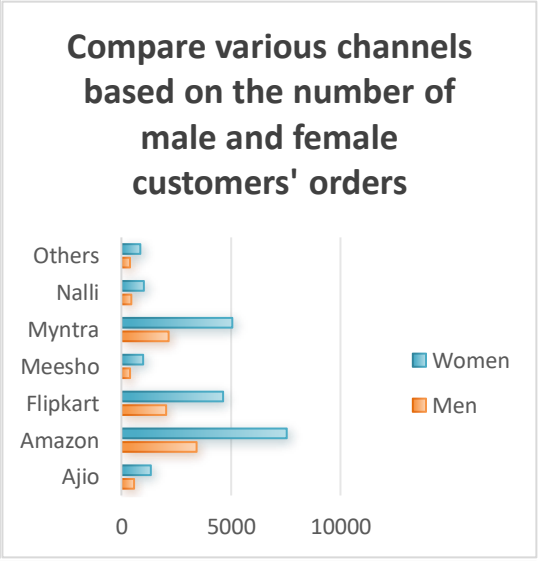


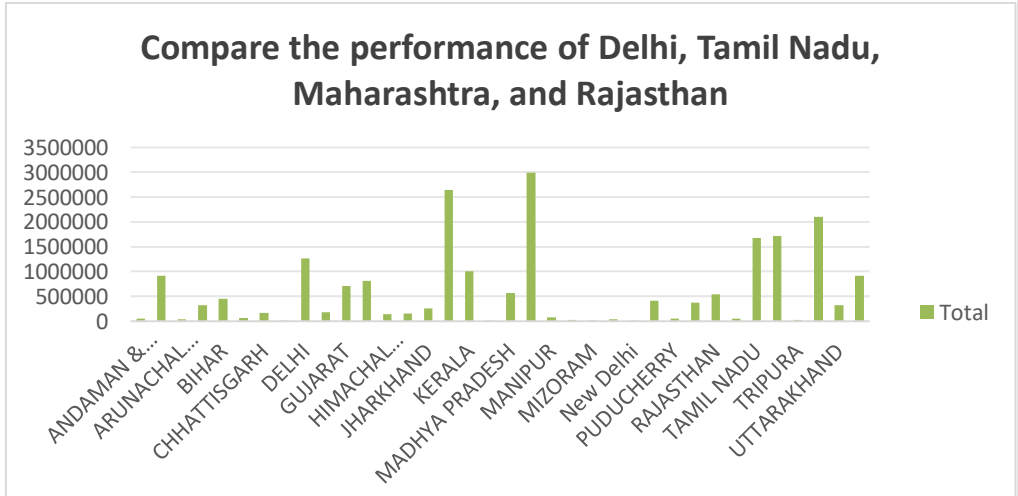


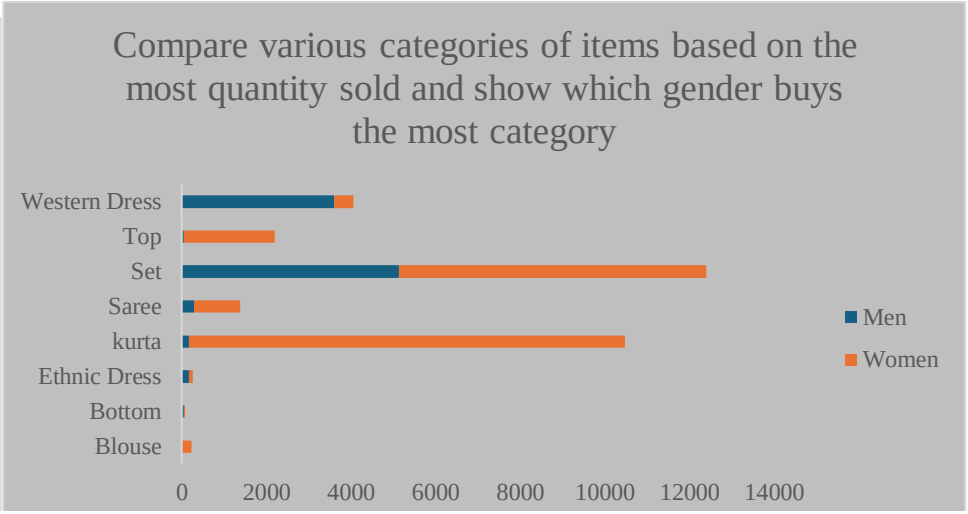
Dashboard : Store Dataset











Car Collection Data Report

Introduction

The Car Collection dataset offers a comprehensive look into various attributes of different car models, including their make, model, color, mileage, price, and cost. In this report, we aim to analyse and derive insights from this dataset to aid decision-making processes related to car purchasing and understanding market trends the dataset contains the total of 6 cars with different models namely, Honda, Chevrolet, Nissan, Toyota, Dodge, Ford.

The primary intended audience for this report includes car enthusiasts, automotive industry professionals, analysts, and individuals interested in exploring trends within the car market. The scope of this report encompasses a detailed analysis of the dataset, including statistical analyses, visualizations, and interpretation of findings.

Throughout the analysis, we have posed several key questions and performed corresponding analyses to uncover insights.

Questionnaire

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

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

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

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

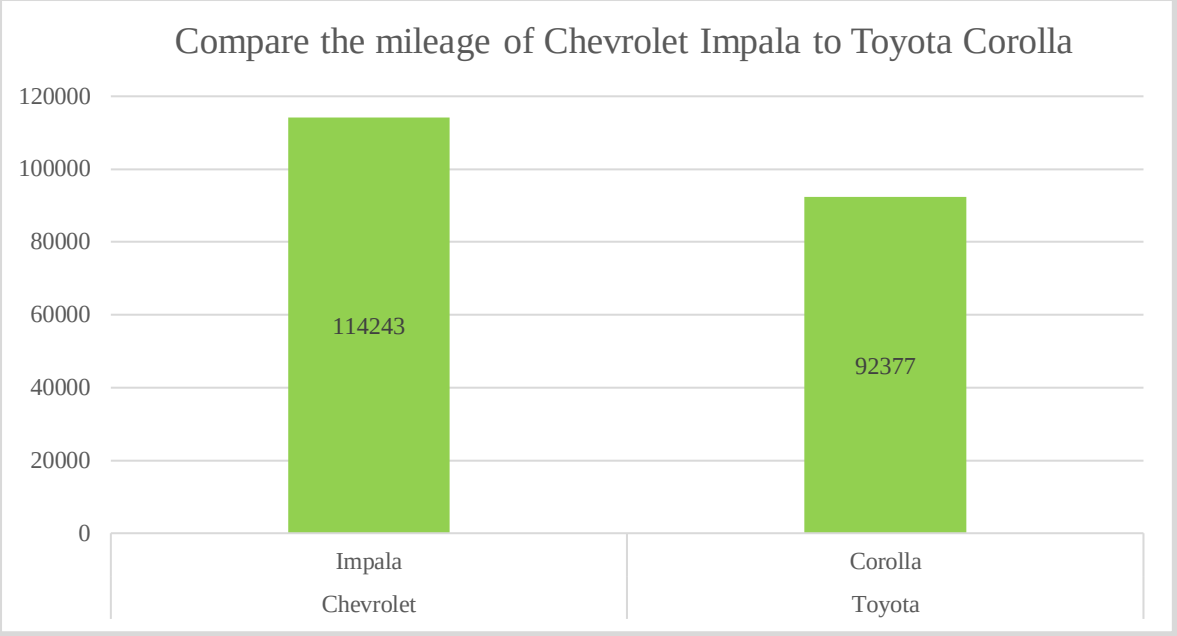
5. Find out all the cars, and their total cost which is more than $2000?

Analytics

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

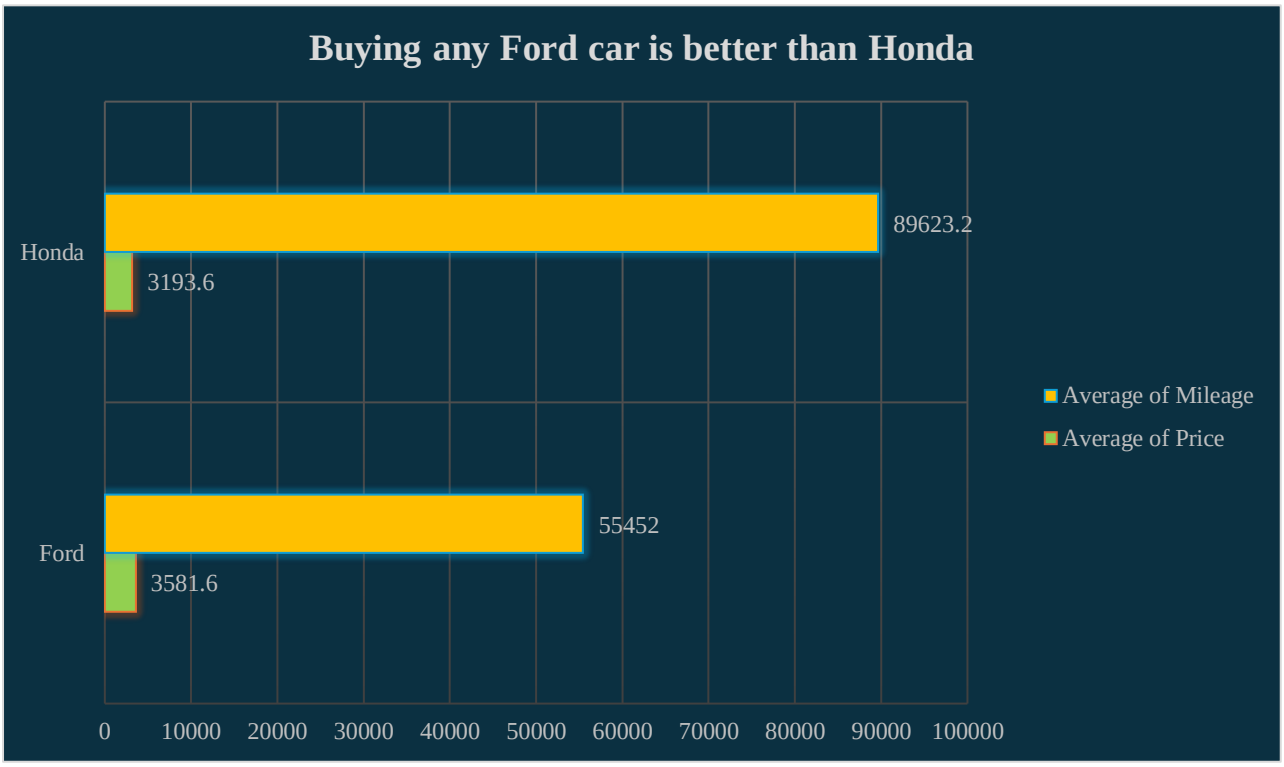
2. This analysis compares the fuel efficiency (mileage) of two popular car models, the Chevrolet Impala and the Toyota Corolla. The dataset was filtered to isolate relevant data, and a column chart was created. Based on the analysis, it was concluded that the Chevrolet Impala (114,243 miles) provides better mileage compared to the Toyota Corolla (92,377 miles).

Average of Mileage		
Make	Model	Total
Chevrolet	Impala	114243
Chevrolet Total		114243
Toyota	Corolla	92377
Grand Total		101123.4



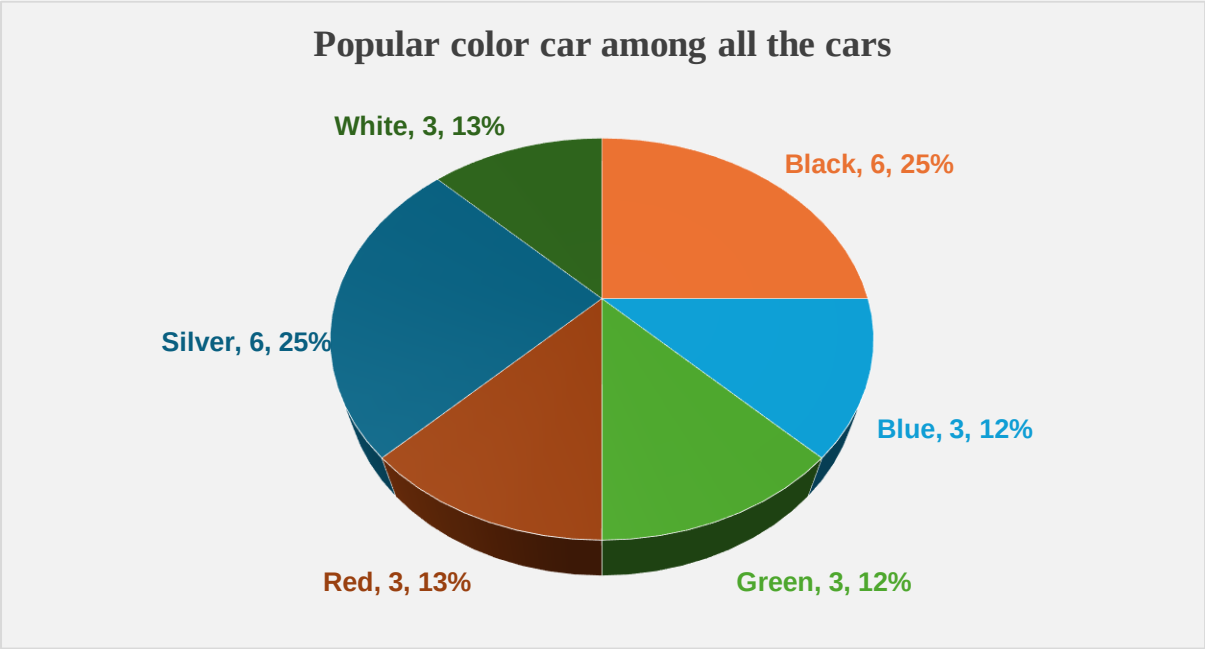
2. Justify, Buying of any Ford car is better than Honda.

Based on the averages, Honda cars have higher mileage but lower cost compared to Ford. Therefore, the choice depends on whether the buyer prioritizes mileage or cost. However, since Ford cars have lower mileage and cost, buying a Ford might be a better option than a Honda if the primary consideration is cost.



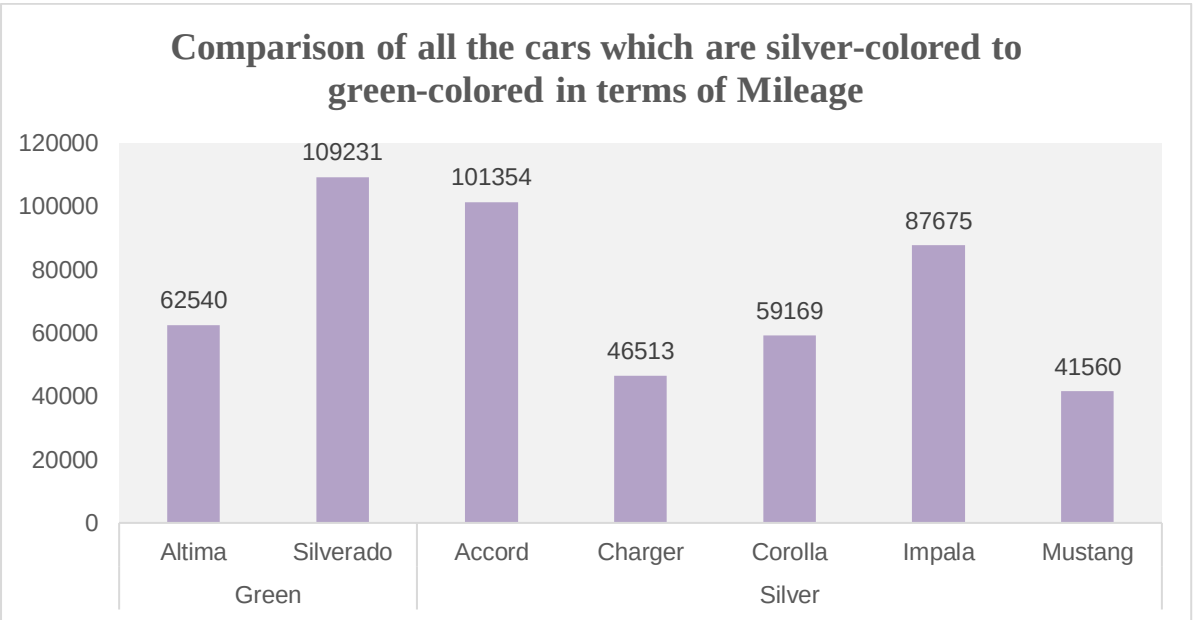
3. Among all the cars which car color is the most popular and is least popular?

This analysis aims to identify the most popular and least popular car colors among all the cars in the dataset based on the count of each make. The analysis revealed that Black and White are the most popular car colors, each accounting for 25% of the total. Conversely, Green and Blue are the least popular, each making up only 12% of the cars produced.



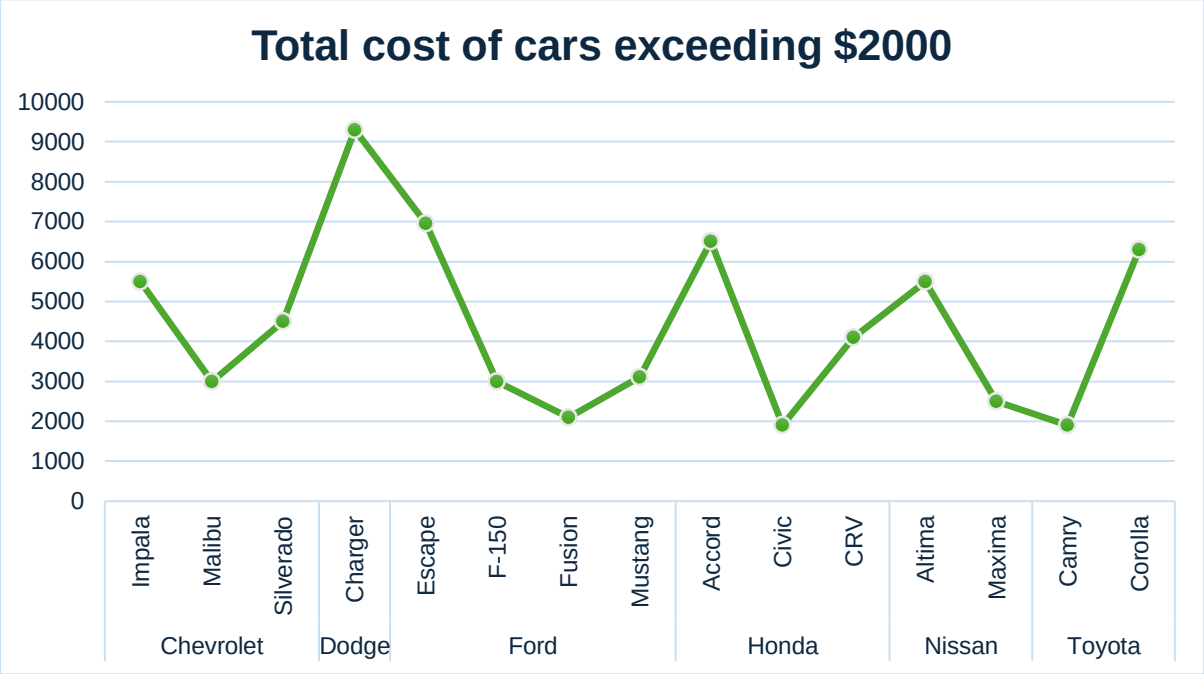
4. Compare all the cars which are of silver color to the green color in terms of Mileage.

This analysis aims to identify the mileage of cars ranging from silver to green in color. The findings reveal that there are five silver cars: Mustang, Impala, Corolla, Charger, and Accord. Among these, the Accord has the highest average mileage at 101,354 miles. Additionally, there are two green cars: Silverado and Altima, with the Silverado having the highest mileage at 109,231 miles.



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

This analysis aims to identify cars costing more than $2,000. Using a bar graph to represent the sum of these costs, the total combined cost of all cars exceeding $2,000 is shown to be $66,150.



Conclusion and Review

Mileage Comparison: The analysis comparing the mileage of the Chevrolet Impala and Toyota Corolla revealed that the Chevrolet Impala provides better fuel efficiency.

Ford vs. Honda Comparison: Contrary to the initial assumption, the analysis did not support the claim that Ford cars are better than Honda cars in terms of mileage and price. Honda cars were found to have better average mileage and lower prices compared to Ford cars.

Popular Car Colors: The analysis identified Black and White as the most popular car colors, each comprising 25% of car production. Conversely, Green and Blue were found to be the least popular colors, each accounting for only 12% of car production.

Silver vs. Green Cars Comparison: Among silver-colored cars, the Accord exhibited the highest average mileage. Among green-colored cars, the Silverado had the highest mileage.

Cars Costing More Than $2,000: The analysis determined that the total cost of cars exceeding $2,000 amounted to $66,150.

Review: The analysis provided valuable insights into various aspects of the dataset, including mileage comparisons, car color popularity, and cost considerations. However, there were discrepancies between the initial assumptions and the findings, particularly in the comparison between Ford and Honda cars. The analysis was thorough and utilized appropriate visualizations, such as column charts and bar graphs, to present the findings effectively.

Regression

Regression shows the stats for the mileage, cost, and price taking the dependent variable as mileage and independent variables as cost and price for the dataset.

Regression Statistics	
Multiple R	0.962639
R Square	0.926673
Adjusted R Square	0.91969
Standard Error	259.2716
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	17839897	8919948	132.6943	1.22E-12
Residual	21	1411657	67221.78		
Total	23	19251554			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	441.3528	288.7848	1.52831	0.141359	-159.208	1041.914	-159.208	1041.914
X Variable 1	-0.00058	0.001699	-0.34395	0.734304	-0.00412	0.002949	-0.00412	0.002949
X Variable 2	1.038413	0.070492	14.73084	1.52E-12	0.891816	1.18501	0.891816	1.18501

Anova: one factor

Anova means the Analysis of variance. The Anova one factor shows the summary of columns having count, sum, average, variance. And the source of variance with ss and df. For total of three columns mileage, price and cost the count for column1, column2, and column3 is shown below.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	24	2011267	83802.79	1.21E+09
Column 2	24	66150	2756.25	705502.7
Column 3	24	78108	3254.5	837024.1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.04E+11	2	5.22E+10	128.8822	5E-24	3.129644
Within Groups	2.8E+10	69	4.05E+08			
Total	1.32E+11	71				

Anova: two factor

A two-factor ANOVA without replication is a data analysis tool that can be used to analyse two factors. It can be used to test the main effects of either factor, here it shows the variance in the dataset for each row with ss and df.

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	70512	23504	1.2E+09
Row 2	3	99635	33211.67	2.88E+09
Row 3	3	104854	34951.33	3.31E+09
Row 4	3	79104	26368	1.77E+09
Row 5	3	76673	25557.67	1.47E+09
Row 6	3	60703	20234.33	9.19E+08
Row 7	3	91602	30534	2.41E+09
Row 8	3	135682	45227.33	5.48E+09
Row 9	3	63329	21109.67	1.09E+09
Row 10	3	143412	47804	6.21E+09
Row 11	3	96023	32007.67	2.44E+09
Row 12	3	118690	39563.33	3.64E+09
Row 13	3	94966	31655.33	2.35E+09
Row 14	3	145151	48383.67	6.41E+09
Row 15	3	145661	48553.67	6.18E+09
Row 16	3	69505	23168.33	1.21E+09
Row 17	3	49123	16374.33	4.48E+08
Row 18	3	48366	16122	4.85E+08
Row 19	3	58171	19390.33	6.72E+08
Row 20	3	107270	35756.67	3.28E+09
Row 21	3	47301	15767	5.38E+08
Row 22	3	42702	14234	3.19E+08
Row 23	3	66425	22141.67	9.74E+08
Row 24	3	140665	46888.33	6.06E+09
Column 1	24	2011267	83802.79	1.21E+09
Column 2	24	66150	2756.25	705502.7
Column 3	24	78108	3254.5	837024.1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	8.95E+09	23	3.89E+08	0.941208	0.549982	1.766805
Columns	1.04E+11	2	5.22E+10	126.3564	2.05E-19	3.199582
Error	1.9E+10	46	4.13E+08			
Total	1.32E+11	71				

Descriptive Statistics

Column1		*Column2*		*Column3*	
Mean	83802.79	Mean	2756.25	Mean	3254.5
Standard Error	7112.652	Standard Error	171.4525	Standard Error	186.7512
Median	81142	Median	2750	Median	3083
Mode	#N/A	Mode	3000	Mode	#N/A
Standard Deviation	34844.74	Standard Deviation	839.9421	Standard Deviation	914.8902
Sample Variance	1.21E+09	Sample Variance	705502.7	Sample Variance	837024.1
Kurtosis	-1.09718	Kurtosis	-0.81266	Kurtosis	-1.20291
Skewness	0.386522	Skewness	0.473392	Skewness	0.272019
Range	105958	Range	3000	Range	2959
Minimum	34853	Minimum	1500	Minimum	2000
Maximum	140811	Maximum	4500	Maximum	4959
Sum	2011267	Sum	66150	Sum	78108
Count	24	Count	24	Count	24

Correlation

	Column 1	*Column 2*
Column 1	1	
Column 2	-0.41106	1

Order Data Report

Introduction

This report delves into a comprehensive dataset capturing sales transactions within the automotive industry, encompassing various attributes such as Order ID, Order Date, Ship Date, Customer Details, Product Information, and Sales Figures. The primary objective of this analysis is to extract actionable insights to inform decision-making processes and drive business growth within the automotive sector. By examining sales data across different US states, segments, categories, and sub-categories, this report aims to identify key trends, top-performing segments, and areas of potential growth. Insights derived from this analysis will be invaluable for automotive industry stakeholders, including sales managers, marketers, and executives, seeking to optimize sales strategies, enhance customer satisfaction, and maximize revenue.

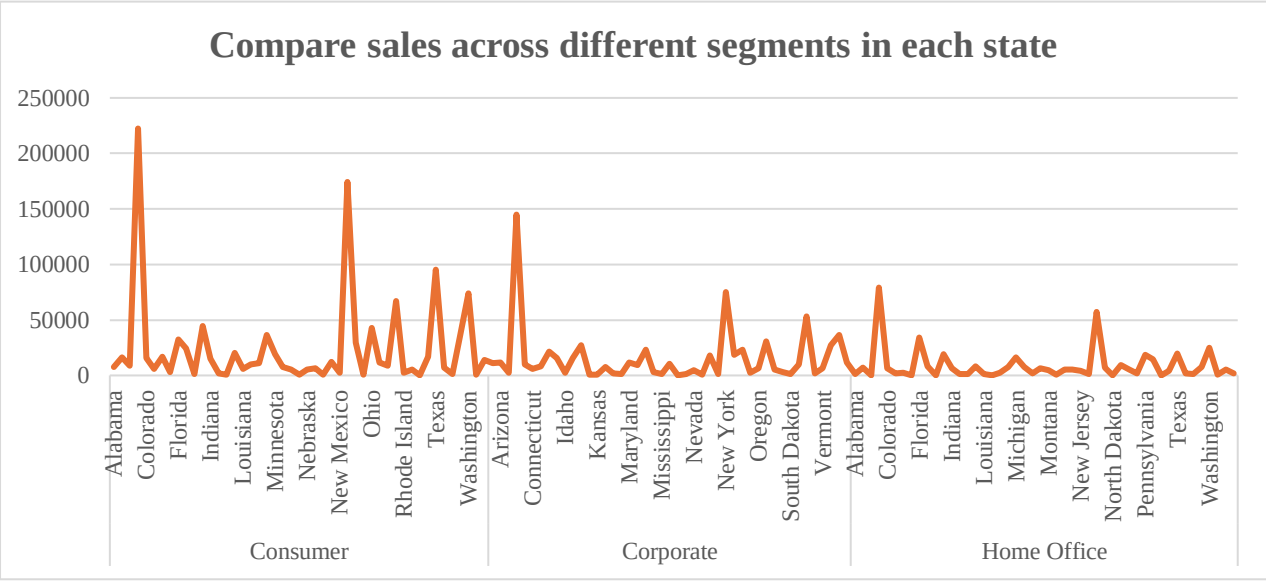
Questionnaire

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 the most sales in the US, California, Texas, and Washington?
4. Compare total and average sales for all different segments?
5. Compare the average sales of different categories and subcategory of all the states.

Analytics

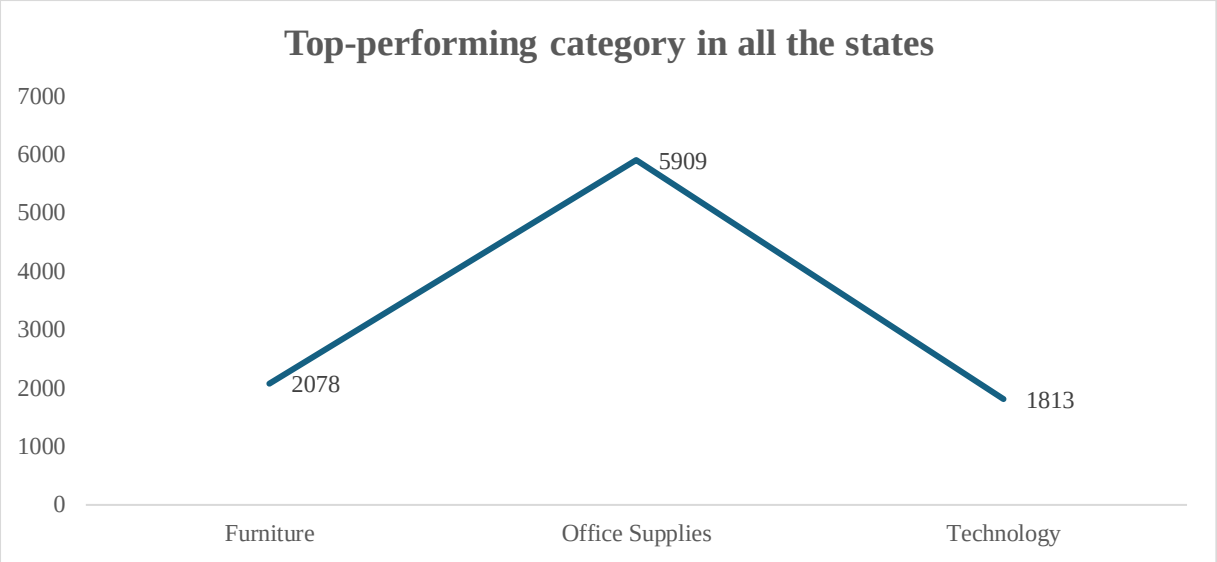
1.Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?

After comparing all the states in terms of segment and sales , California(222419.05) emerged as the state with the highest number of sales. Consumer(1148060.531) segment performed well in all the states.



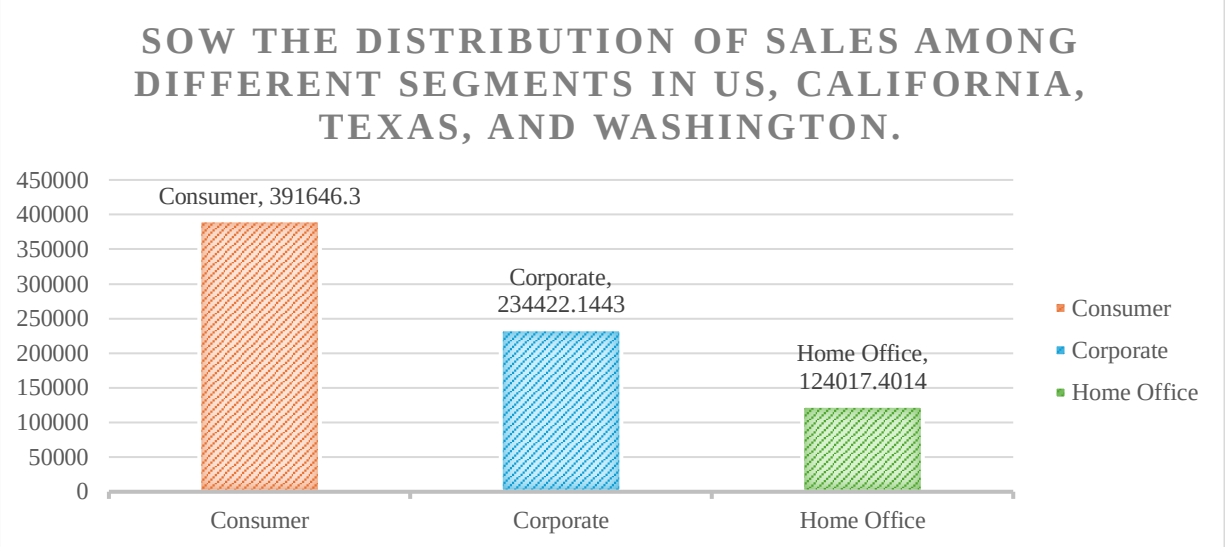
2. Find out top performing category in all the states?

Office Supplies is the top-performing category in all states, with a total sales count of 5,909, followed by Furniture with 2,078 sales and Technology with 1,813 sales.



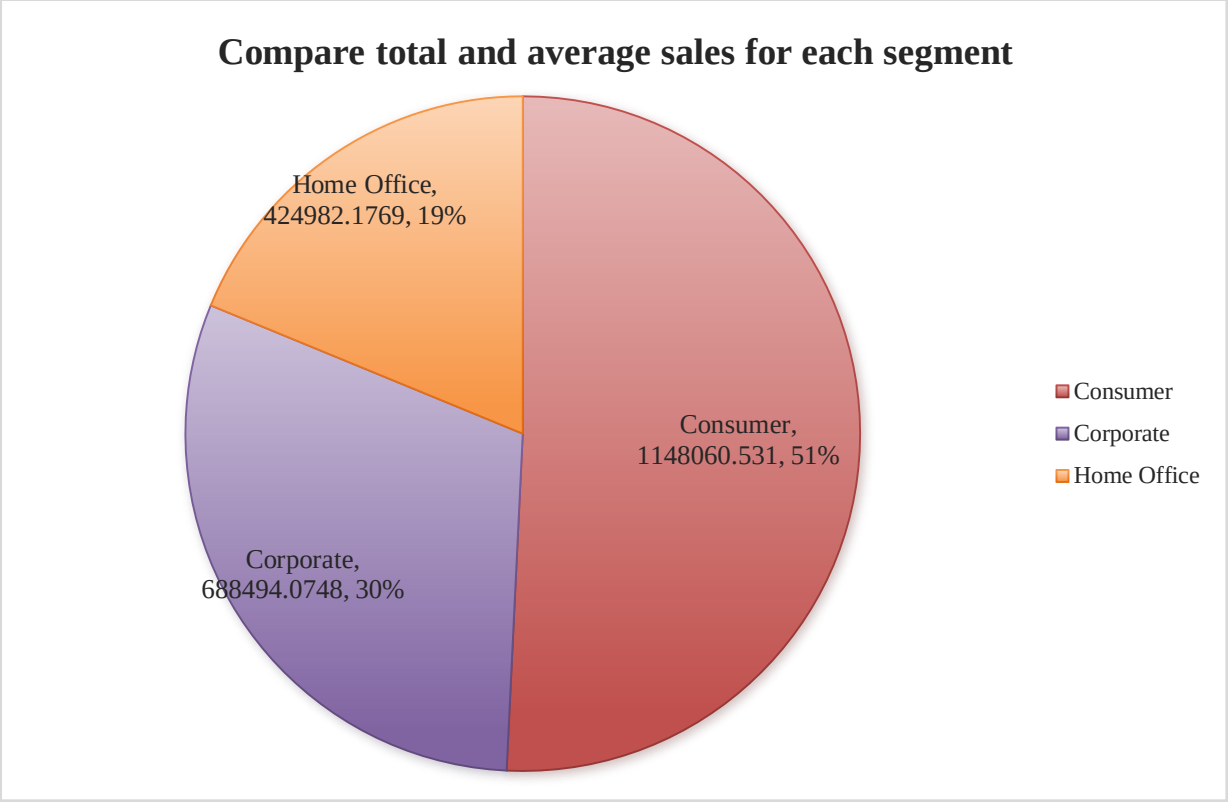
3. Which segment has most sales in US, California, Texas, and Washington?

By filtering the states for the total sales count and displaying the percentage distribution through a pie chart, it is evident that the Consumer segment has the most sales in the US, particularly in California, Texas, and Washington.



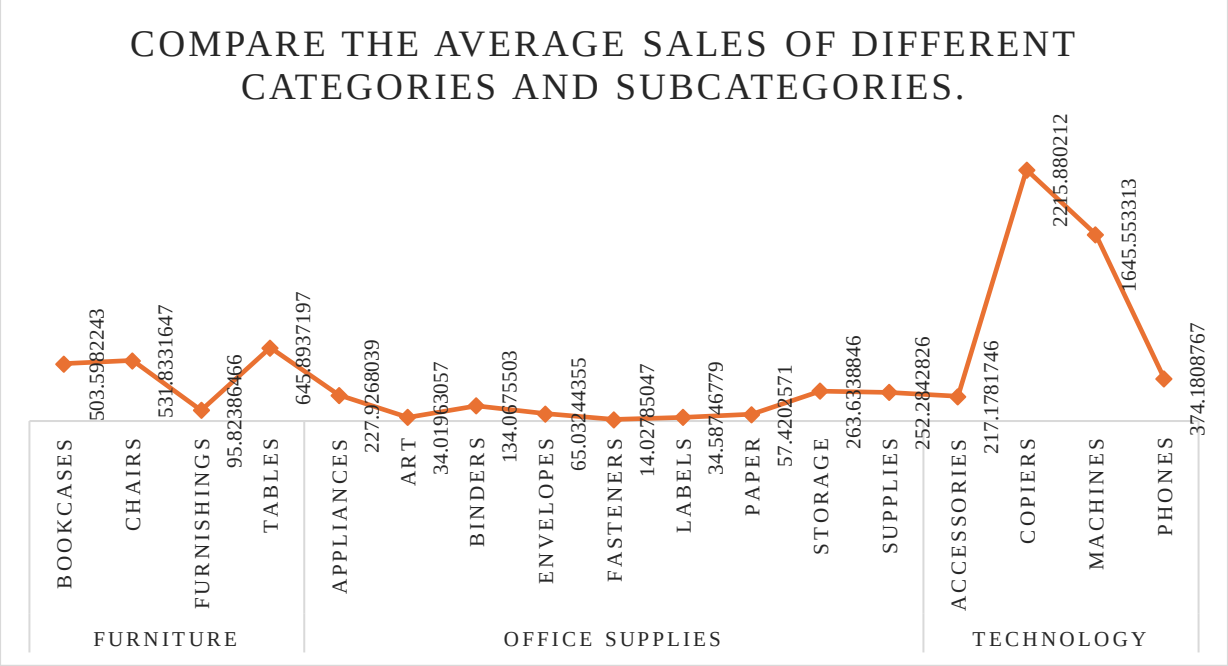
4. Compare total and average sales for all different segments?

It is clearly visible that the Consumer segment has higher average sales, totaling $1,148,060.53, while the Home Office segment has significantly lower total sales, amounting to $243.40.



5. Compare average sales of different categories and subcategory of all the states.

The analysis shows the average sales for the three categories with multiple subcategories: Furniture, Office Supplies, and Technology.



Conclusion and Review

The analysis of sales data within the automotive industry reveals several key findings. California emerges as the top-performing state in terms of sales volume, with the Consumer segment demonstrating strong performance across all states. Office Supplies is the top-performing category, followed by Furniture and Technology, indicating consumer preferences. The Consumer segment consistently dominates sales across the US, particularly in California, Texas, and Washington.

The analysis highlights the higher average sales of the Consumer segment compared to the Home Office segment. Overall, these insights provide valuable guidance for optimizing sales strategies, improving customer engagement, and driving business success within the automotive industry.

Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.000434
R Square	1.88E-07
Adjusted R Square	-0.0001
Standard Error	625.334
Observations	9789

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	721.1637	721.1637	0.001844	0.965747
Residual	9787	3.83E+09	391042.6		
Total	9788	3.83E+09			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	230.5863	12.63999	18.24261	3.83E-73	205.8093	255.3633	205.8093	255.3633
X Variable 1	-9.6E-05	0.002235	-0.04294	0.965747	-0.00448	0.004286	-0.00448	0.004286

Descriptive Statistics

	Column1
Mean	230.1162
Standard Error	6.320053
Median	54.384
Mode	12.96
Standard Deviation	625.3021
Sample Variance	391002.7
Kurtosis	307.3056
Skewness	13.05363
Range	22638.04
Minimum	0.444
Maximum	22638.48
Sum	2252607
Count	9789

Cookie Data Report

Introduction

In our cookie dataset, we have detailed information on six types of cookies: Chocolate Chip, Fortune Cookie, Sugar, Oatmeal Raisin, Snickerdoodle, and White Chocolate Macadamia Nut. This dataset includes information on the number of units sold, costs, revenue generated, and profits. Our analysis spans different countries and dates to observe variations over time and location.

This report aims to provide insights into consumer preferences, pricing tolerance, and the popularity of different cookies in various regions. By examining this data, we can uncover valuable trends and patterns that are crucial for businesses to understand customer behavior, optimize product offerings, and enhance sales strategies. Get ready to discover intriguing insights into the cookie market and what it means for your business.

Questionnaire

1. Compare the profit earn by all cookie types in US, Malaysia, and India.

2. What is the average revenue generated by different types of cookies?

3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

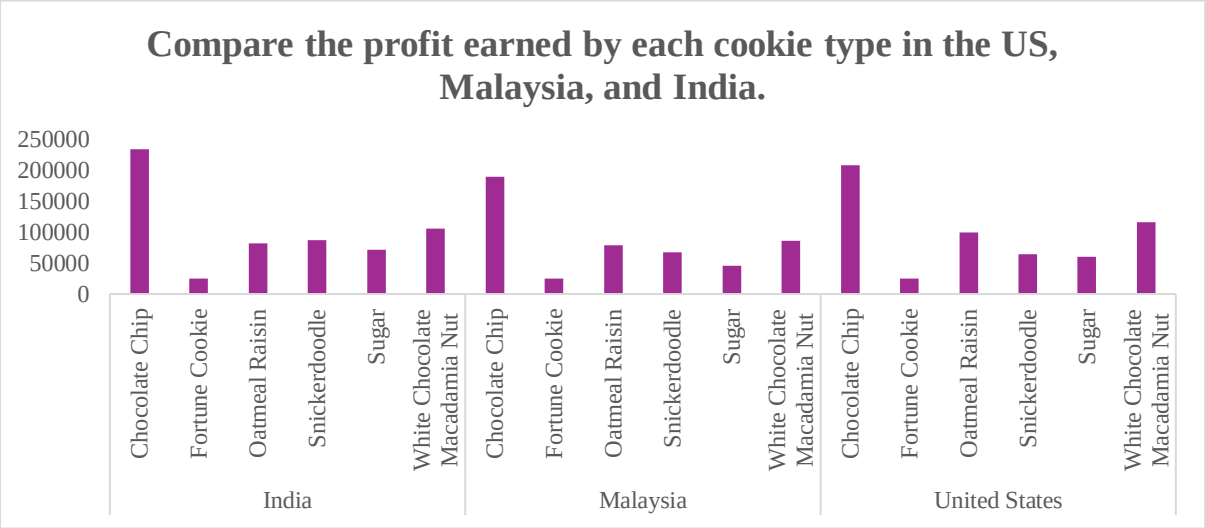
4. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?

5. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

Analytics

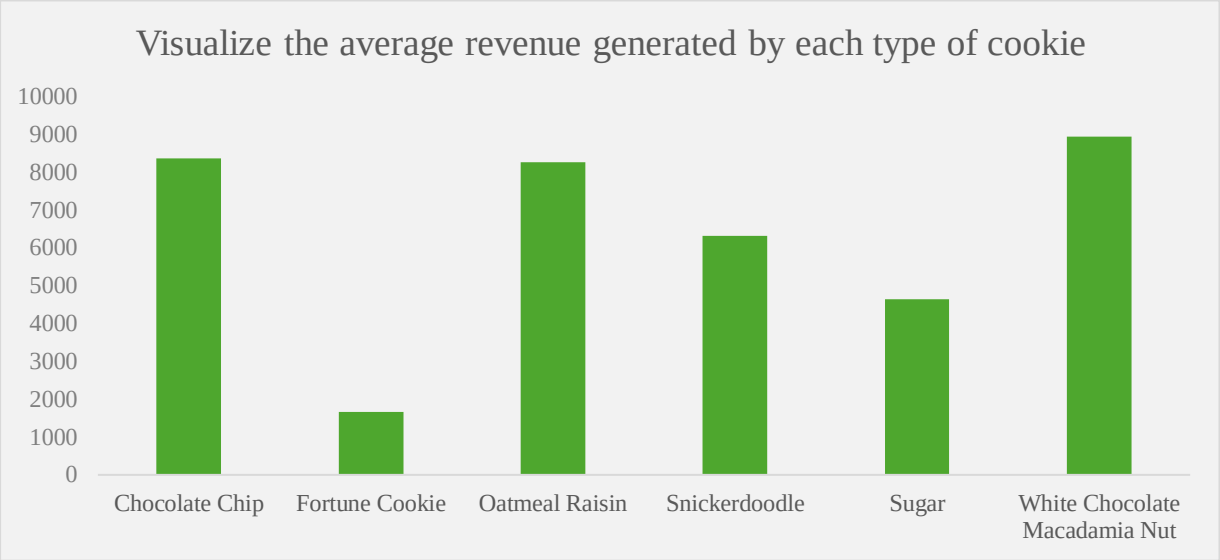
1. Compare the profit earn by all cookie types in US, Malaysia, and India.

This analysis compares the profit earned by all cookie types in US, Malaysia, and India. Max profit earned by India for chocolate chip followed by Malaysia and United States for the same.



2. What is the average revenue generated by different types of cookies?

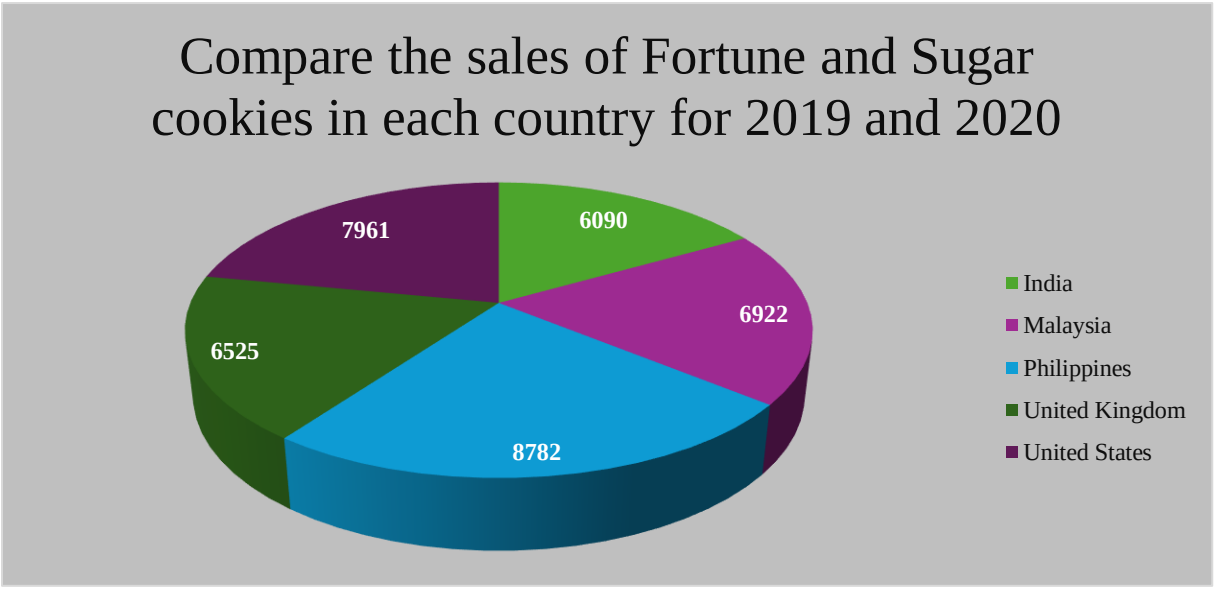
This analysis aims to provide insights into the average revenue generated by different types of cookies. It is evident that White Chocolate Macadamia Nut cookies generate the highest average revenue at $8,940.88, followed closely by Chocolate Chip cookies.



3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

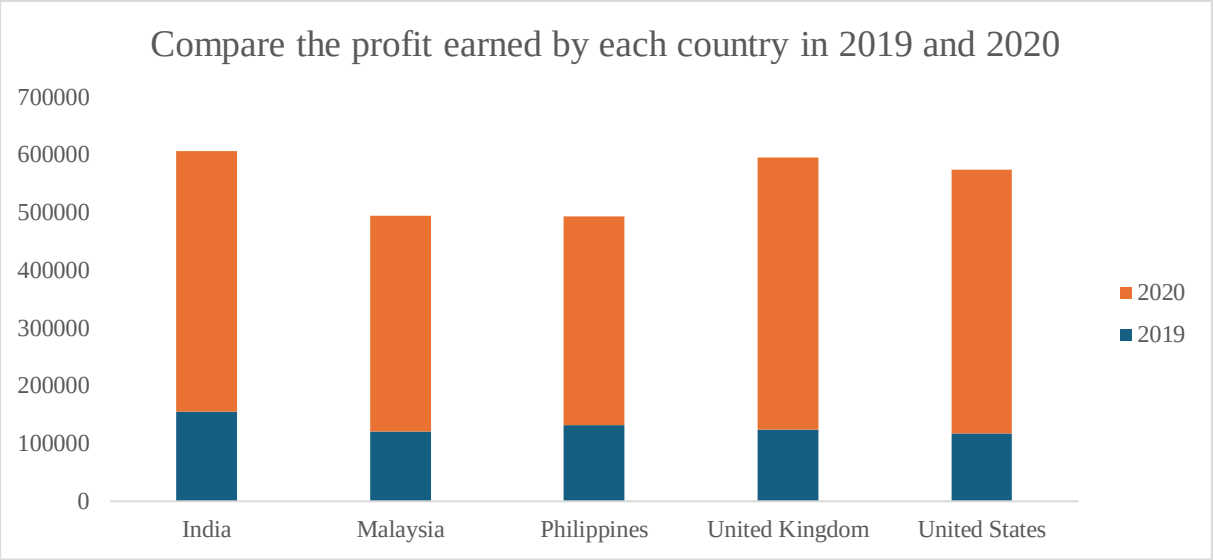
This analysis aims to compare the sales of Fortune and Sugar cookies across different countries for the years 2019 and 2020. The findings show that in 2020, India had a significant sales count for Sugar cookies, with 30,644 units sold. In 2019, the United Kingdom led in Sugar cookie sales, followed by India.

For Fortune cookies, India again had the highest sales in 2020 with 25,400 units, followed by Malaysia. In 2019, the Philippines had the highest Fortune cookie sales, with 8,782 units sold, followed by the United States.



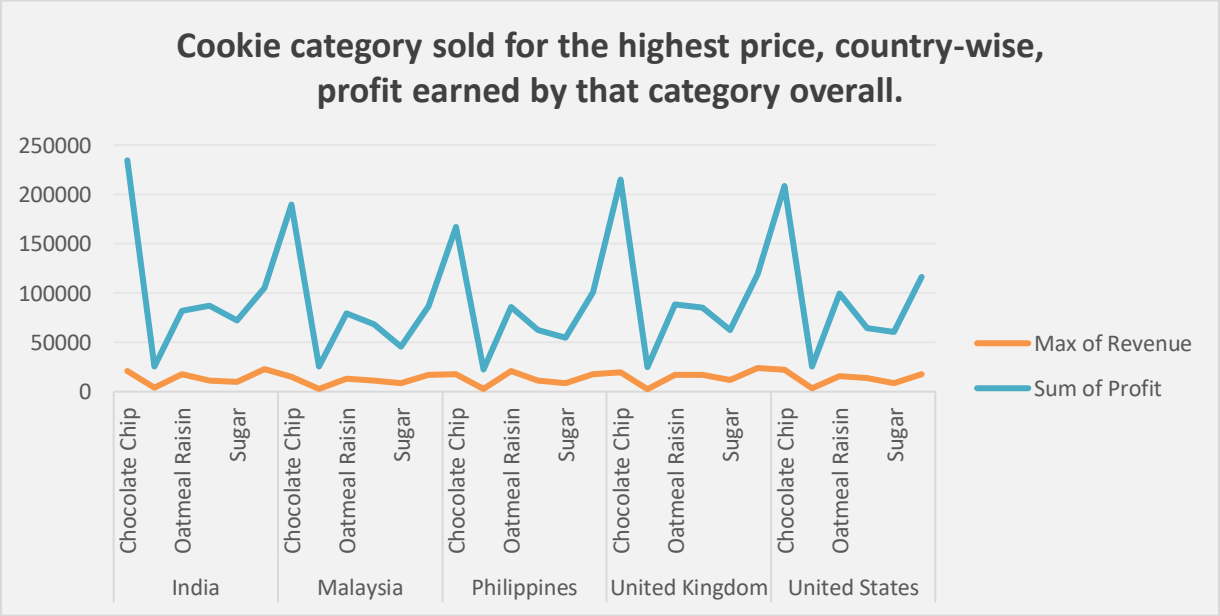
4. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?

This analysis aims to compare the profit earned by countries in the financial years 2019 and 2020. According to the graph, the United Kingdom shows the highest profit in 2020 with $471,027.55 in sales, followed by the United States with $456,839.35. In 2019, the highest profit was recorded by India with $155,515.50 in sales, followed by the Philippines with $131,474.80.



5. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

This analysis aims to find the cookie category sold for the highest price, country-wise, profit earned by that category, max of revenue is recorded by chocolate chip(23988) and sum of profit is recorded by sugar(2763364.45) for the country India followed by United Kingdom.



Conclusion and Review

The analysis provided insights into the profit earned by different cookie types in the US, Malaysia, and India. India emerged with the highest profit for chocolate chip cookies, followed by Malaysia and the United States. White chocolate macadamia nut cookies generated the highest average revenue, followed closely by chocolate chip cookies.In terms of sales, India showed significant sales of sugar cookies in 2020, while the United Kingdom had the highest sales of sugar cookies in 2019. For fortune cookies, India and Malaysia exhibited higher sales in both years, with the Philippines and the United States also contributing notable sales.

Regarding profit comparison by country for 2019 and 2020, the United Kingdom recorded the highest profit in 2020, followed by the United States. In 2019, India had the highest profit, followed by the Philippines. Chocolate chip cookies generated the highest revenue, while sugar cookies produced the highest overall profit.

The analysis presented valuable insights into the cookie industry, aiding stakeholders in understanding market dynamics and making informed decisions. The findings were effectively communicated through clear and appropriate visualizations. However, it's important to acknowledge the need for further exploration into additional factors influencing sales and profitability. Ensuring data accuracy and completeness is paramount for obtaining reliable insights.

Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	9.16E-12
Observations	700

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	4.78E+09	1.59E+09	1.9E+31	0
Residual	696	5.84E-20	8.39E-23		
Total	699	4.78E+09			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.3E-11	7.3E-13	-18.0657	4.09E-60	-1.5E-11	-1.2E-11	-1.5E-11	-1.2E-11
X Variable 1	6.56E-17	8.42E-16	0.077892	0.937936	-1.6E-15	1.72E-15	-1.6E-15	1.72E-15
X Variable 2	1	8.38E-16	1.19E+15	0	1	1	1	1
X Variable 3	-1	1.72E-15	-5.8E+14	0	-1	-1	-1	-1

Anova: one factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	700	1926955	2752.792	4149401
Column 2	700	2763364	3947.664	6842519

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5E+08	1	5E+08	90.92153	6.36E-21	3.848119
Within Groups	7.68E+09	1398	5495960			
Total	8.18E+09	1399				

Anova: two factor

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	17250	5750	6943125
Row 2	3	21520	7173.333	10805909
Row 3	3	23490	7830	12874869
Row 4	3	12280	4093.333	3518629
Row 5	3	13890	4630	4501749
Column 1	700	4690319	6700.456	21380458
Column 2	700	1926955	2752.792	4149401
Column 3	700	2763364	3947.664	6842519

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.99E+10	699	28507277	14.75112	0	1.112595
Columns	5.74E+09	2	2.87E+09	1484.458	0	3.002161
Error	2.7E+09	1398	1932550			
Total	2.84E+10	2099				

Descriptive Statistics

Column1		*Column2*		*Column3*		*Column4*	
Mean	1608.32	Mean	6700.456	Mean	2752.792	Mean	3947.664
Standard Error	32.78652	Standard Error	174.767	Standard Error	76.99166	Standard Error	98.86874
Median	1542.5	Median	5871.5	Median	2423.6	Median	3424.5
Mode	727	Mode	8715	Mode	3450	Mode	5229
Standard Deviation	867.4498	Standard Deviation	4623.901	Standard Deviation	2037.008	Standard Deviation	2615.821
Sample Variance	752469.1	Sample Variance	21380458	Sample Variance	4149401	Sample Variance	6842519
Kurtosis	-0.31491	Kurtosis	0.464596	Kurtosis	0.810043	Kurtosis	0.338621
Skewness	0.43627	Skewness	0.867861	Skewness	0.930442	Skewness	0.840484
Range	4293	Range	23788	Range	10954.5	Range	13319
Minimum	200	Minimum	200	Minimum	40	Minimum	160
Maximum	4493	Maximum	23988	Maximum	10994.5	Maximum	13479
Sum	1125824	Sum	4690319	Sum	1926955	Sum	2763364
Count	700	Count	700	Count	700	Count	700

Correlation

	Column 1	*Column 2*	*Column 3*	*Column 4*
Column 1	1			
Column 2	0.796298	1		
Column 3	0.742604	0.992011	1	
Column 4	0.829304	0.995163	0.974818	1

Loan Data Report

Introduction

The loan dataset provides comprehensive information about loan applicants, including attributes such as gender, marital status, education level, income details, loan amount, and property area. This dataset offers a rich source of insights into the dynamics of loan applications.

In this analysis, we aim to delve into the characteristics of loan applicants and explore patterns within the data. By leveraging pivot tables and charts, we seek to address specific queries regarding loan applicants' demographics, educational backgrounds, and loan amounts.Understanding the nuances of loan applications is crucial for financial institutions to make informed decisions, optimize lending processes, and tailor services to meet the diverse needs of customers. Through this analysis, we aim to uncover actionable insights that can drive strategic decision-making and enhance the efficiency of loan management systems.

Questionnaire

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

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

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

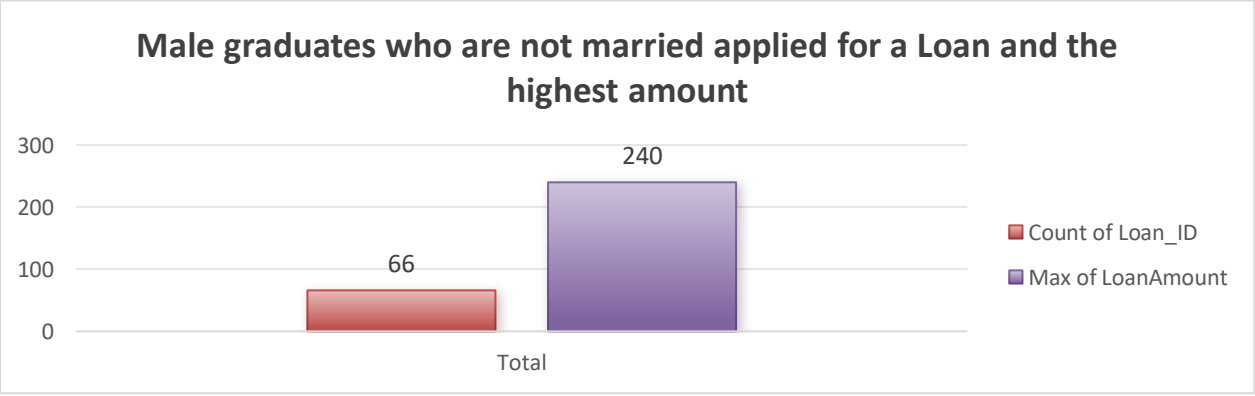
4. How many female graduates who are married applied for Loan? What was the highest amount?

5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rural based on amount.

Analytics

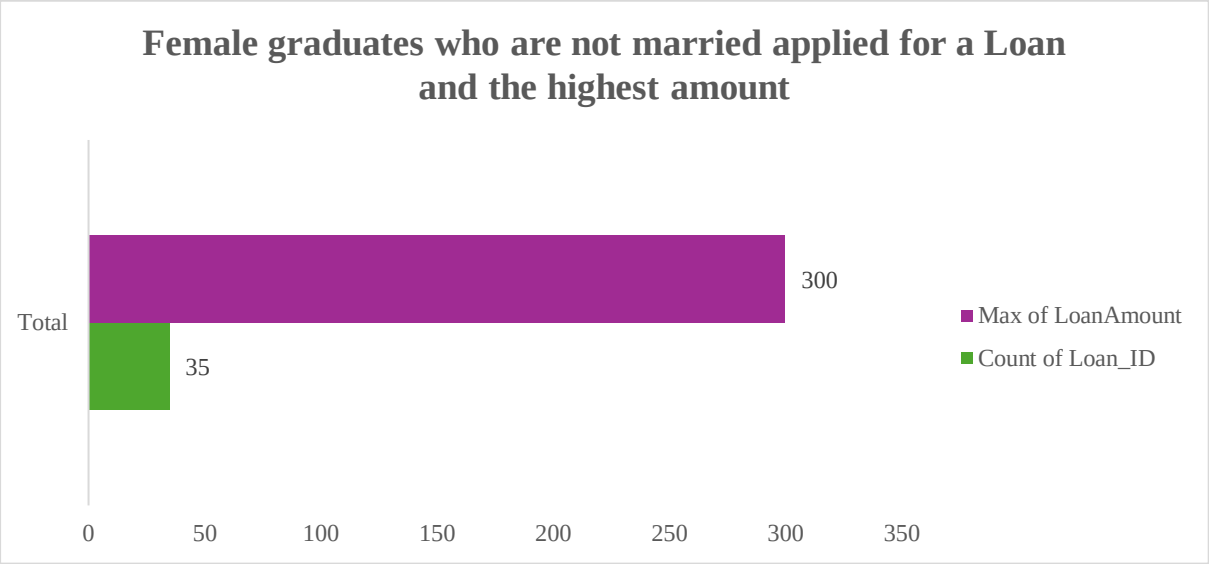
1. How many male graduates who are not married applied for Loan? What was the highest amount?

This analysis reveals that the highest loan amounts are sought by unmarried male graduates. Among the analyzed applicants, a total of 66 loans have been applied for, with the maximum loan amount being $240.



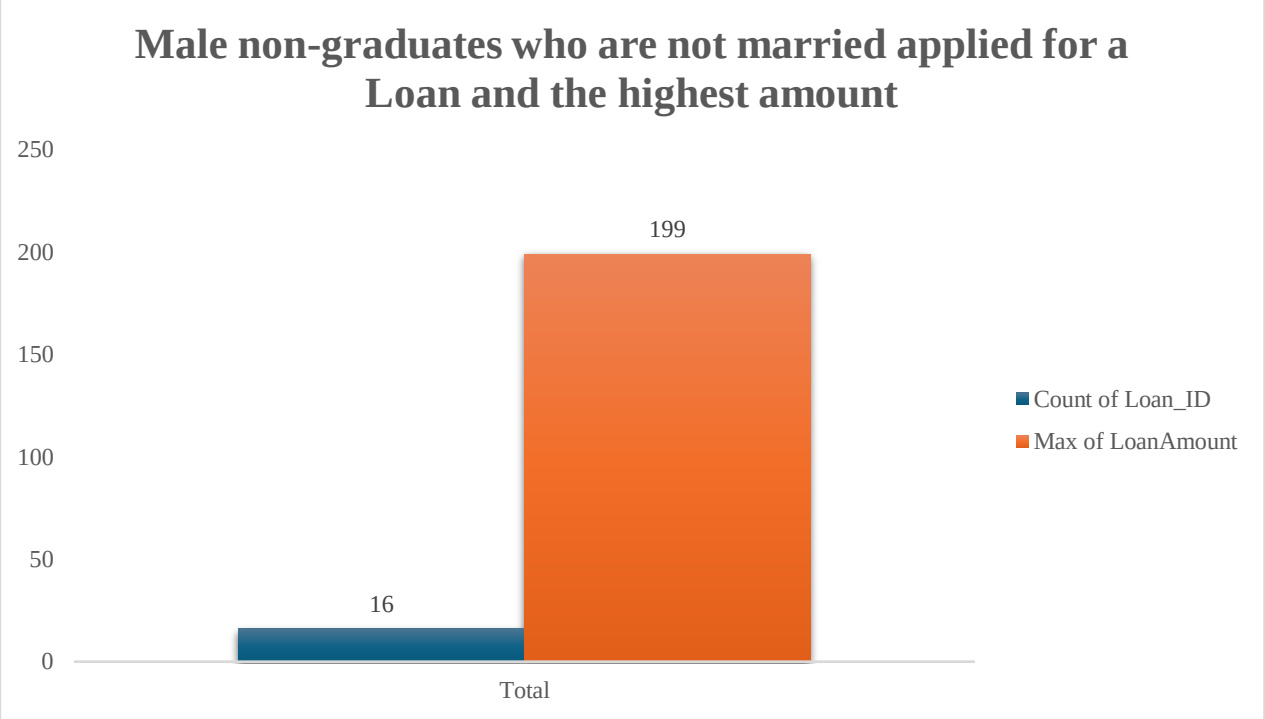
2. How many female graduates who are not married applied for Loan? What was the highest amount?

This analysis highlights that the highest loan amounts are sought by unmarried female graduates. Among the analyzed applicants, a total of 35 loans have been applied for, with the maximum loan amount being $300.



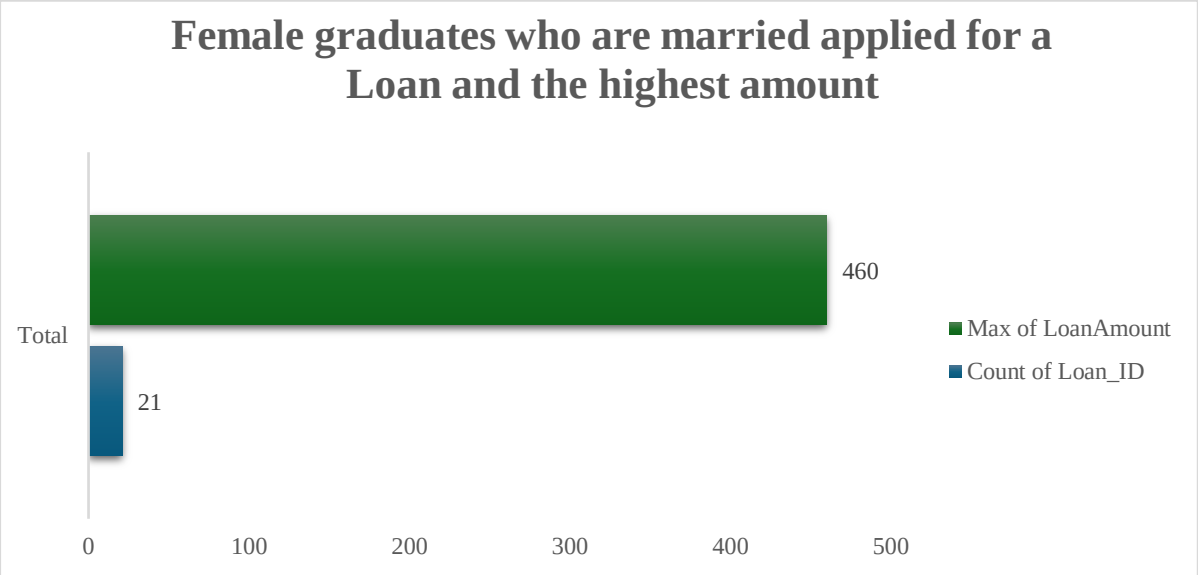
3. How many male non-graduates who are not married applied for Loan? What was the highest amount?

This analysis indicates that among non-graduate males who are not married, the highest loan amount has been sought. A total of 16 loans have been applied for in this category, with the maximum loan amount being $199.



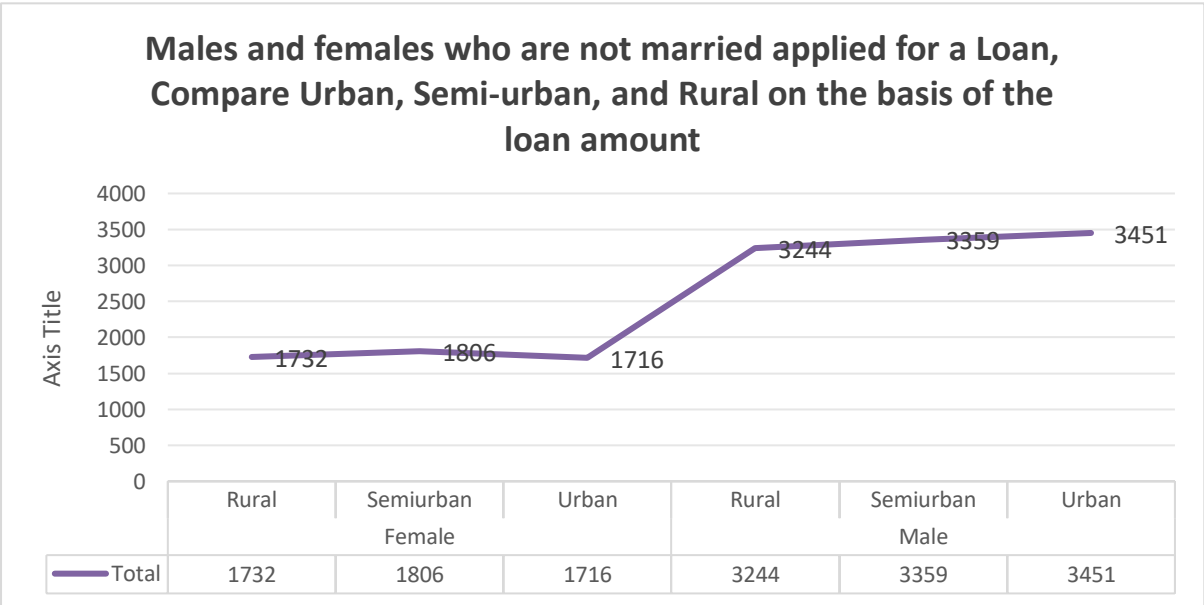
4. How many female graduates who are married applied for Loan? What was the highest amount?

This analysis reveals that among female graduates who are not married, the highest loan amount has been sought. A total of 21 loans have been applied for in this category, with the maximum loan amount being $460.



5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rural based on amount.

This analysis aims to compare the loan applications of unmarried individuals across rural, semi-urban, and urban areas, distinguishing between males and females. It shows that while females have applied for loans in rural (1,732), semi-urban (1,806), and urban (1,716) areas, the number is considerably higher for males in each category. Males have applied for loans in rural (3,244), semi-urban (3,359), and urban (3,451) areas.



Conclusion and Review

The analysis indicates clear gender disparities in loan applications. Male graduates who are not married dominated the applicant pool, followed by female graduates who are not married. Additionally, both male non-graduates who are not married and married female graduates also applied for loans, though in smaller numbers. Notably, males significantly outnumbered females across rural, semi-urban, and urban areas.

The analysis effectively illustrates gender-based trends in loan applications and provides valuable insights into borrower demographics. Further exploration into factors influencing loan decisions is recommended, along with visual enhancements to improve data presentation. Overall, the report lays a foundation for understanding loan dynamics, with potential for deeper insights.

Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.531078663
R Square	0.282044546
Adjusted R Square	0.274487121
Standard Error	50.85033905
Observations	289

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	289502.8035	96500.93	37.32019	2.25609E-20
Residual	285	736940.7397	2585.757		
Total	288	1026443.543			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	66.690952	16.26833015	4.099434	5.41E-05	34.66963005	98.71227396	34.66963	98.71227
X Variable 1	0.095771273	0.045649816	2.097955	0.03679	0.005917708	0.185624838	0.005918	0.185625
X Variable 2	0.005807787	0.000627861	9.250122	5.49E-18	0.004571955	0.007043619	0.004572	0.007044
X Variable 3	0.006772797	0.001264765	5.354983	1.76E-07	0.004283331	0.009262263	0.004283	0.009262

Anova: one factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	289	39533	136.7924	3564.04
Column 2	289	99032	342.6713	4310.645

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6124794	1	6124794	1555.565	8.4E-166	3.857654
Within Groups	2267909	576	3937.343			
Total	8392703	577				

Anova: two factor

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	2	470	235	31250
Row 2	2	486	243	27378
Row 3	2	568	284	11552
Row 4	2	438	219	39762
Row 5	2	512	256	21632
Row 286	2	473	236.5	30504.5
Row 287	2	475	237.5	30012.5
Row 288	2	518	259	20402
Row 289	2	278	139	3362
Column 1	289	39533	136.7924	3564.04
Column 2	289	99032	342.6713	4310.645

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1264619	288	4391.038	1.260472	0.024978	1.214301

Columns	6124794	1	6124794	1758.156	1.2E-124	3.87395
Error	1003290	288	3483.647			
Total	8392703	577				

Descriptive Statistics

Column1		*Column2*		*Column3*		*Column4*	
Mean	342.6713	Mean	4637.353	Mean	1528.263	Mean	136.7924
Standard Error	3.862088	Standard Error	281.8049	Standard Error	139.8588	Standard Error	3.51174
Median	360	Median	3833	Median	879	Median	126
Mode	360	Mode	5000	Mode	0	Mode	150
Standard Deviation	65.6555	Standard Deviation	4790.684	Standard Deviation	2377.599	Standard Deviation	59.69958
Sample Variance	4310.645	Sample Variance	22950653	Sample Variance	5652978	Sample Variance	3564.04
Kurtosis	8.62994	Kurtosis	141.612	Kurtosis	32.96701	Kurtosis	5.739804
Skewness	-2.64147	Skewness	10.41123	Skewness	4.510775	Skewness	1.780616
Range	474	Range	72529	Range	24000	Range	432
Minimum	6	Minimum	0	Minimum	0	Minimum	28
Maximum	480	Maximum	72529	Maximum	24000	Maximum	460
Sum	99032	Sum	1340195	Sum	441668	Sum	39533
Count	289	Count	289	Count	289	Count	289

Correlation

	Column 1	*Column 2*	*Column 3*
Column 1	1		
Column 2	-0.08435	1	
Column 3	0.445695	0.230355	1

Shop Sales Data Report

Introduction

This report delves into a comprehensive sales dataset, focusing on analyzing sales performance and product trends among salesmen. The dataset comprises attributes such as salesmen details, product information, sales quantities, and profits earned. The primary objective of this analysis is to uncover insights that can inform sales strategy formulation and enhance business performance. By examining sales data over a specified period and comparing product performance, the report aims to identify top-performing salesmen, analyze product popularity, and understand sales trends. The insights derived from this analysis will be invaluable for sales managers, marketing professionals, and executives seeking to optimize sales strategies, maximize revenue, and drive business growth. Through this analysis, we aim to provide actionable insights that can guide decision-making and contribute to overall business success.

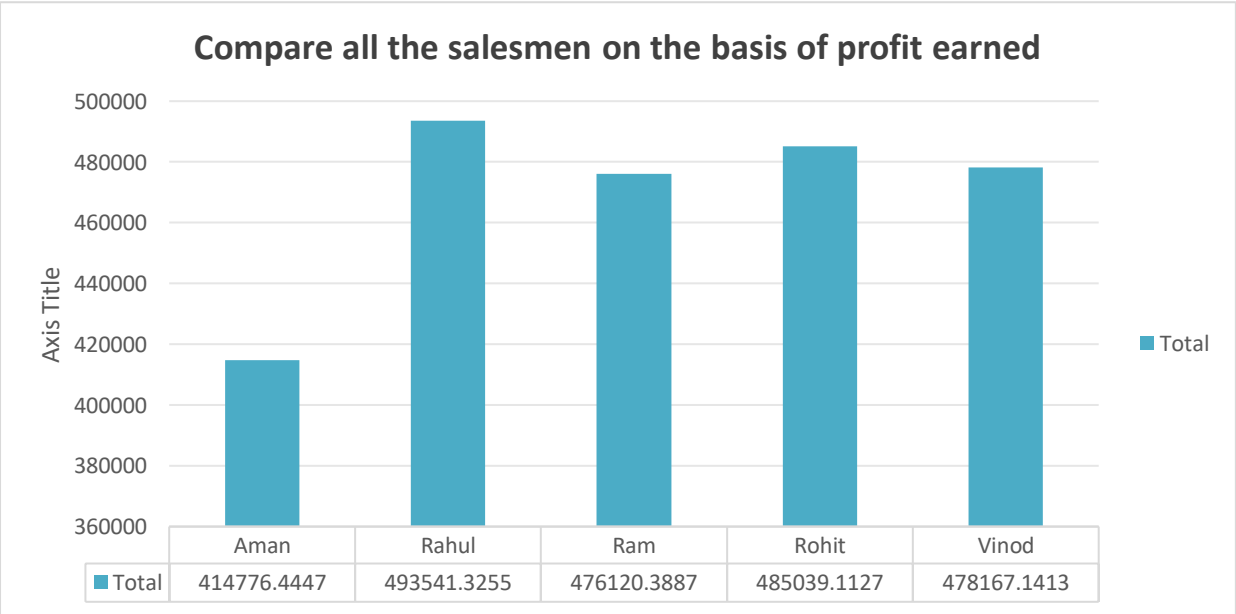
Questionaries

1. Compare all the salesmen based on 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

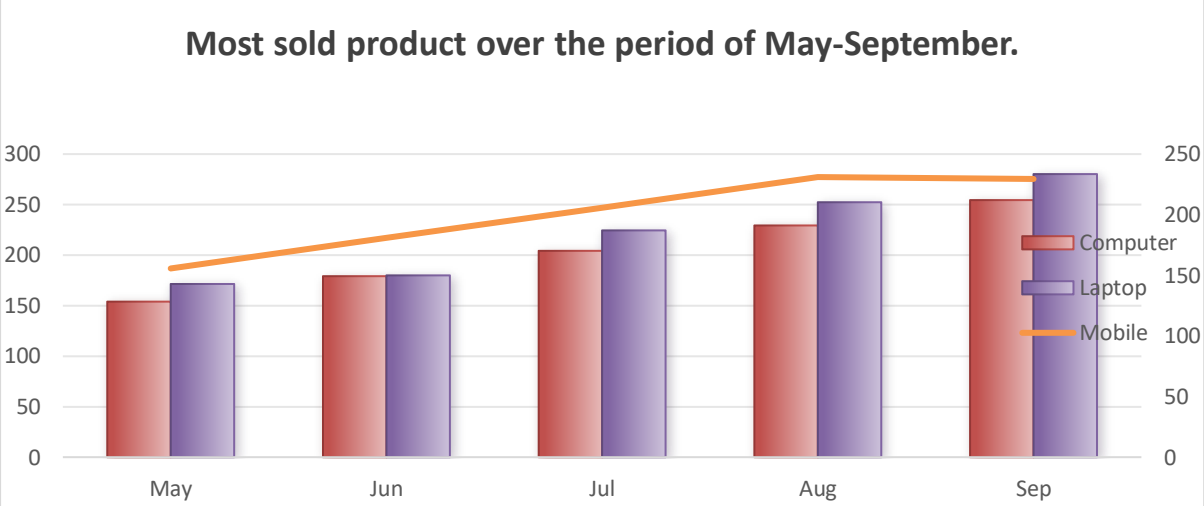
1. Compare all the salesmen on the basis of profit earn.

The comparison of all salesmen based on the profit earned reveals that Rahul has achieved the highest profit, amounting to $493,541.3255, surpassing all other salesmen.



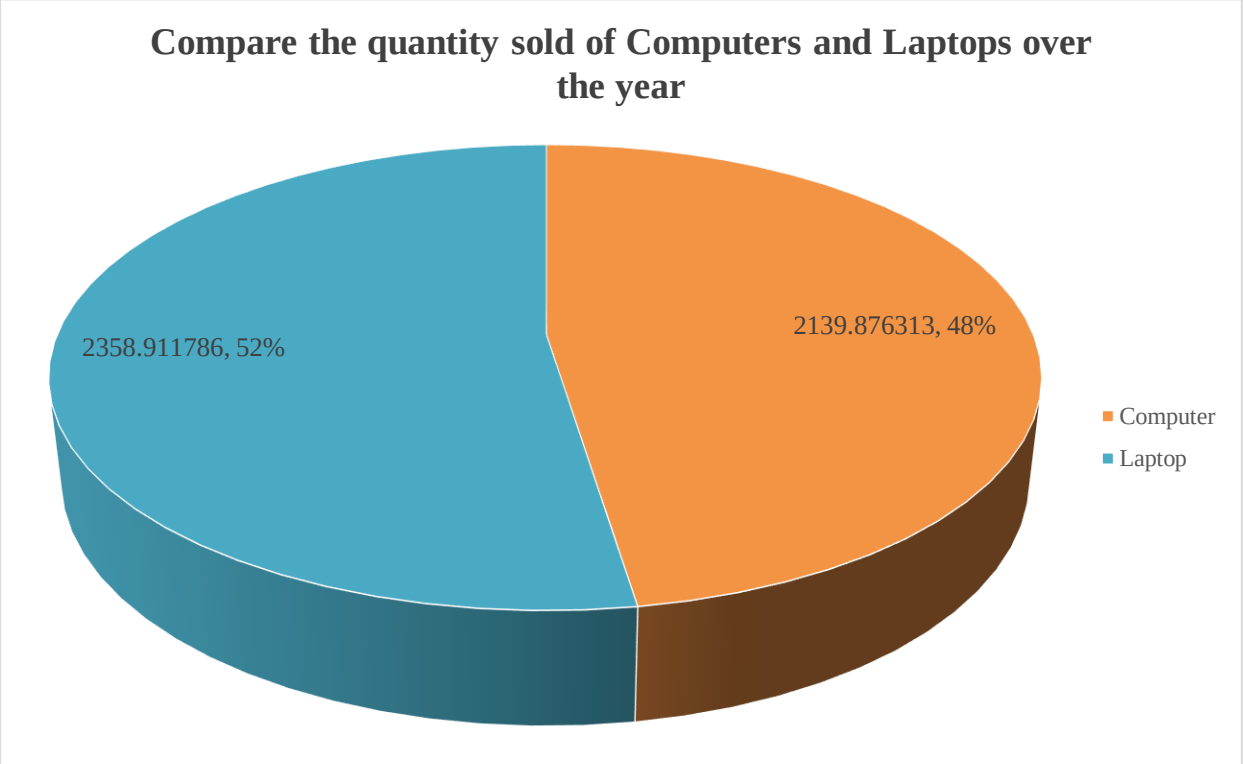
2. Find out most sold product over the period of May-September.

To identify the most sold product over the period of May-September, we need to analyze the sales data within the specified timeframe. By aggregating the quantity sold for each product across all transactions during this period, it is found that the most sold product is the Laptop, with the highest sales recorded in the month of September, totaling 280.1970249 units.



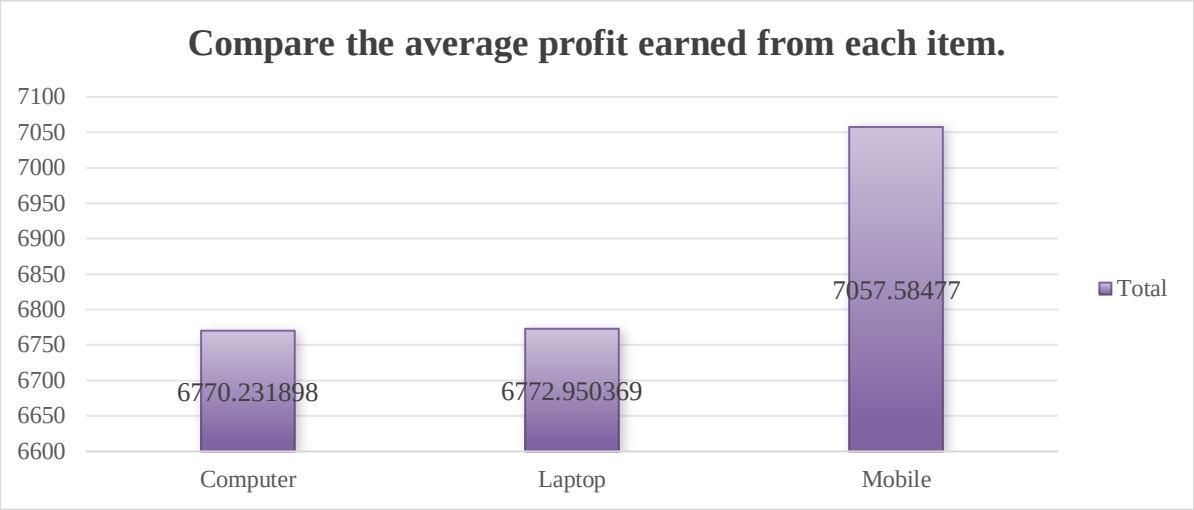
3. Find out which of the two product sold the most over the year Computer or Laptop?

Over the year, the two most sold products were computers and laptops. Computers sold a quantity of 2139.876313 units, while laptops had a higher sales quantity of 2358.911786 units.



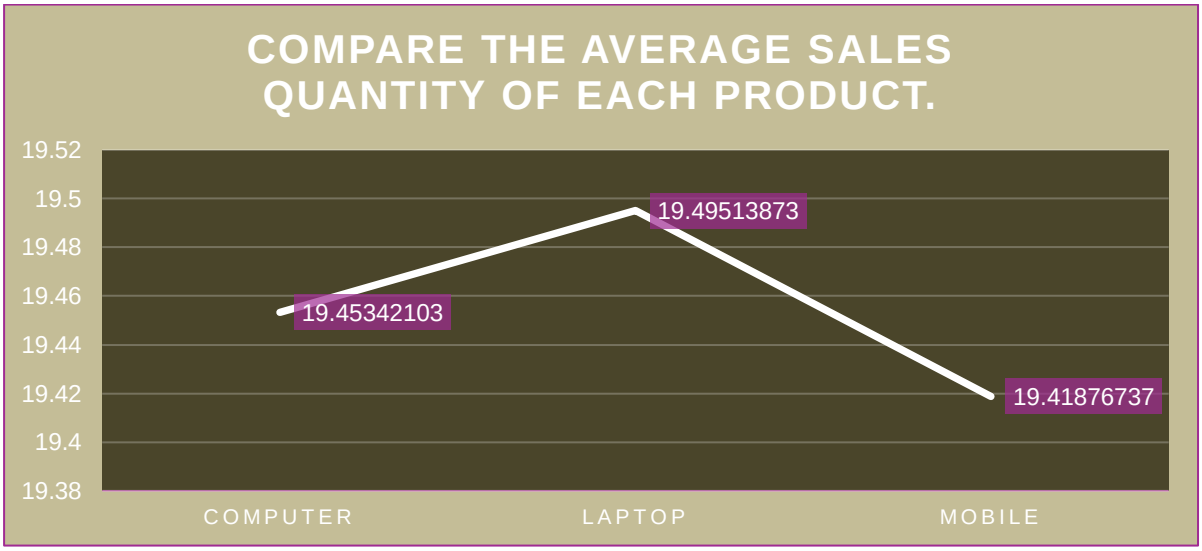
4 . Which item yield most average profit?

This analysis shows that the Mobile has the most Average profit earned among Mobile, Laptop, and Computer where Mobile has the average profit earned of 7057.58477.



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

The analysis shows that the average sales quantity of Laptop(19.49513873) is higher than the other products e.g. Mobile(19.41876737) and Computer(19.45342103).



Conclusion and Review:

The analysis provides significant insights into sales performance and product trends among salesmen. Rahul emerges as the top performer, earning the highest profit compared to all other salesmen. Additionally, the most sold product over the period of May-September is identified as the Laptop, with the highest sales recorded in September. Between computers and laptops, laptops outperform computers in terms of units sold throughout the year.

Furthermore, mobile phones exhibit the highest average profit among mobiles, laptops, and computers. Lastly, laptops demonstrate the highest average sales quantity compared to mobiles and computers.

The analysis effectively highlights sales performance and product trends, providing valuable insights for sales strategy optimization. Visualizations aid in understanding trends over time and product popularity. However, deeper insights into factors influencing sales fluctuations and product preferences could enhance the analysis. Overall, the report offers actionable insights for improving sales strategies and maximizing revenue.

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 0.910.

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.954076972
R Square	0.910262868
Adjusted R Square	0.909998936
Standard Error	630.0595983
Observations	342

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.37E+09	1.37E+09	3448.844	4.6E-180
Residual	340	1.35E+08	396975.1		
Total	341	1.5E+09			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*U*
Intercept	2068.993161	88.47952	23.38387	9.14E-73	1894.957	2243.029	1894.957	2
X Variable 1	246.4655683	4.196812	58.72686	4.6E-180	238.2106	254.7206	238.2106	2

Correlation

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

	Column 1	*Column 2*
Column 1	1	
Column 2	0.954077	1

Anova (Single Factor)

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

Anova: Single Factor

SUMMARY

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

ANOVA

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 two factor

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: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	2	1003	501.5	497004.5
Row 2	2	7804	3902	30388808
Row 3	2	3005	1502.5	4485013
Row 4	2	2304	1152	2635808
Row 5	2	7003	3501.5	24479005
Row 339	2	10252.82	5126.411	51884342
Row 340	2	10272.93	5136.467	52087770
Row 341	2	10293.05	5146.523	52291595
Row 342	2	10313.16	5156.58	52495819
Column 1	342	6654.271	19.45693	66.0952
Column 2	342	2347644	6864.457	4410782

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	7.58E+08	341	2221714	1.014883	0.445792	1.195299
Columns	8.01E+09	1	8.01E+09	3659.913	2.1E-184	3.868873
Error	7.46E+08	341	2189134			
Total	9.52E+09	683				

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

Sales Data Sample Report

Introduction

This report analyzes a comprehensive sales dataset, featuring attributes such as ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and SALES. Its aim is to extract insights to guide sales strategies and enhance business performance. The intended audience includes sales managers, marketers, and executives seeking to optimize sales operations and maximize revenue.

Through these analyses, the report aims to provide actionable insights for driving sales growth and improving overall business outcomes. The scope of the project encompasses analyzing a comprehensive sales dataset to extract valuable insights that can inform sales strategies, optimize product offerings, and enhance overall business performance. Analysts and researchers seeking insights into sales dynamics and market trends will find value in the project.

Questionnaire

1. Comparison of sales between Vintage cars and Classic cars across all countries.

2. Determination of the average sales of all products and identification of the highest-selling product.

3. Assessment of the country yielding the most profit for Motorcycles, Trucks, and Buses.

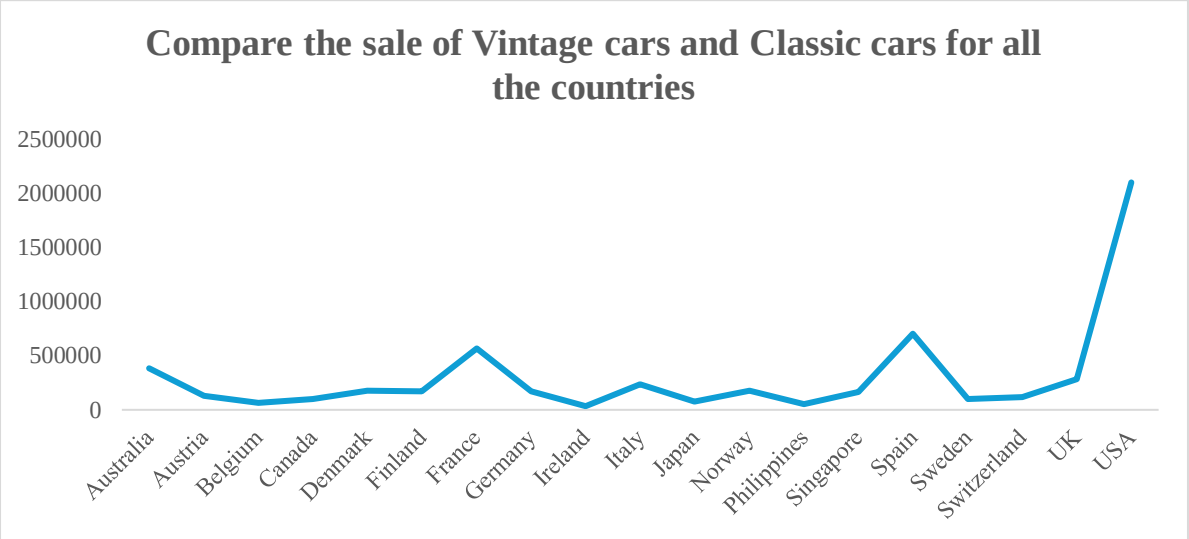
4. Comparison of sales for all items across the years 2004 and 2005.

5. Comparative analysis of all countries based on deal size.

Analytics

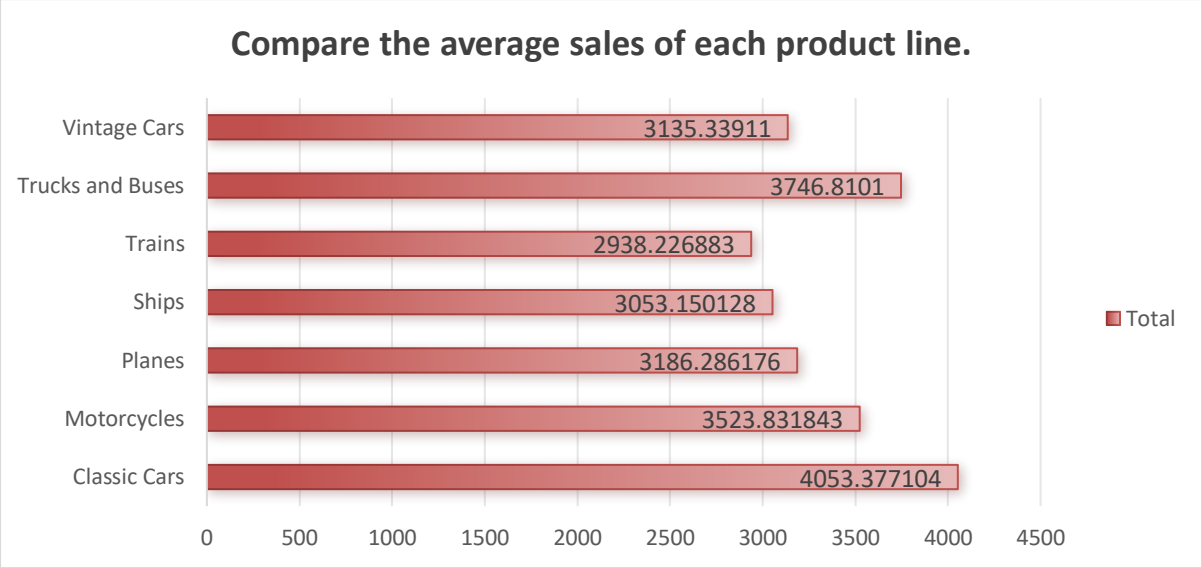
1. Comparison of sales between Vintage cars and Classic cars across all countries.

This analysis Compare the sale of Vintage cars and Classic cars for all the countries. Where USA(2102394.02) has the highest sales followed by Spain, France, and Australia.



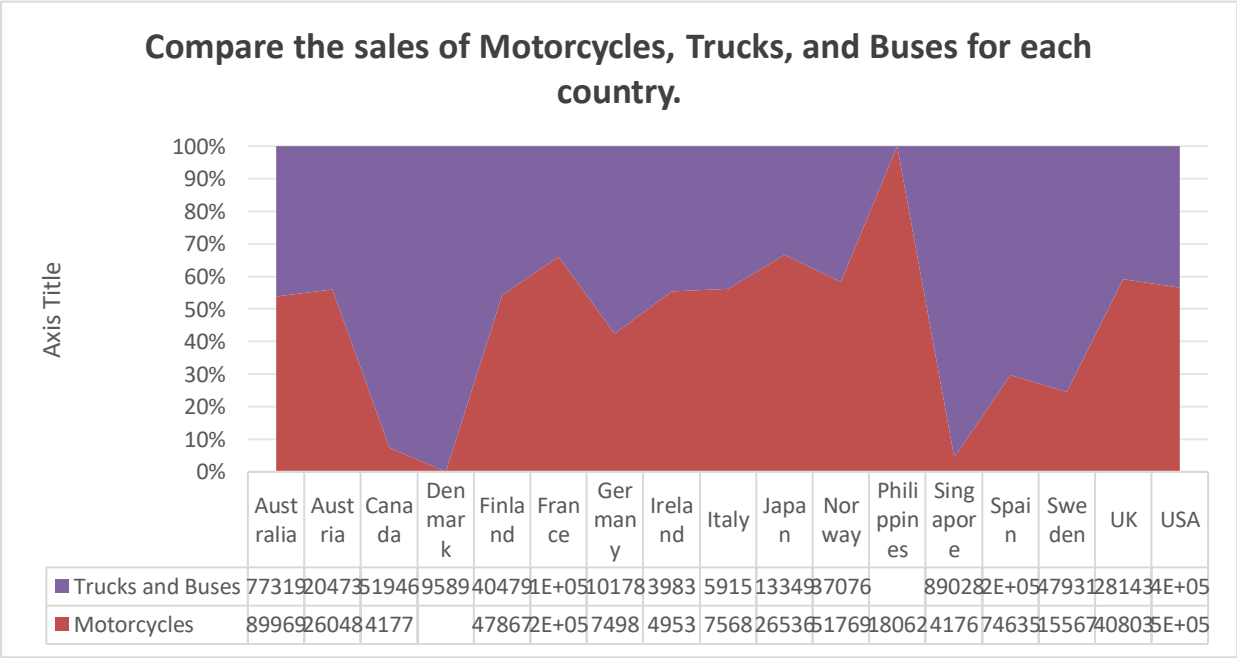
2. Determination of the average sales of all products and identification of the highest-selling product.

This analysis aims to provide the average sales of all products and identify the highest-selling product. According to the graph, Classic Cars have the highest sales, with an average of 4053.377104 units sold, followed by Trucks and Buses and Motorcycles.



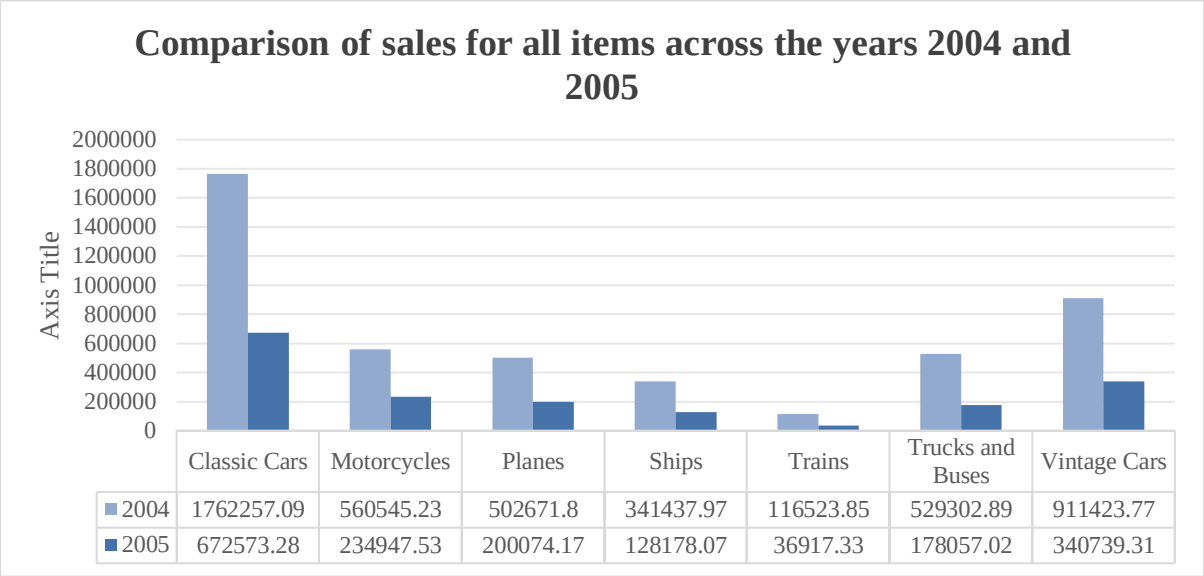
3. Assessment of the country yielding the most profit for Motorcycles, Trucks, and Buses.

This analysis aims to identify the country yielding the most profit for Motorcycles, Trucks, and Buses. According to the bar chart, the USA has the highest sales, with a total of $397,842.42 in sales for Trucks and Buses, and $520,371.70 in sales for Motorcycles, followed by France and Spain.



4. Comparison of sales for all items across the years 2004 and 2005.

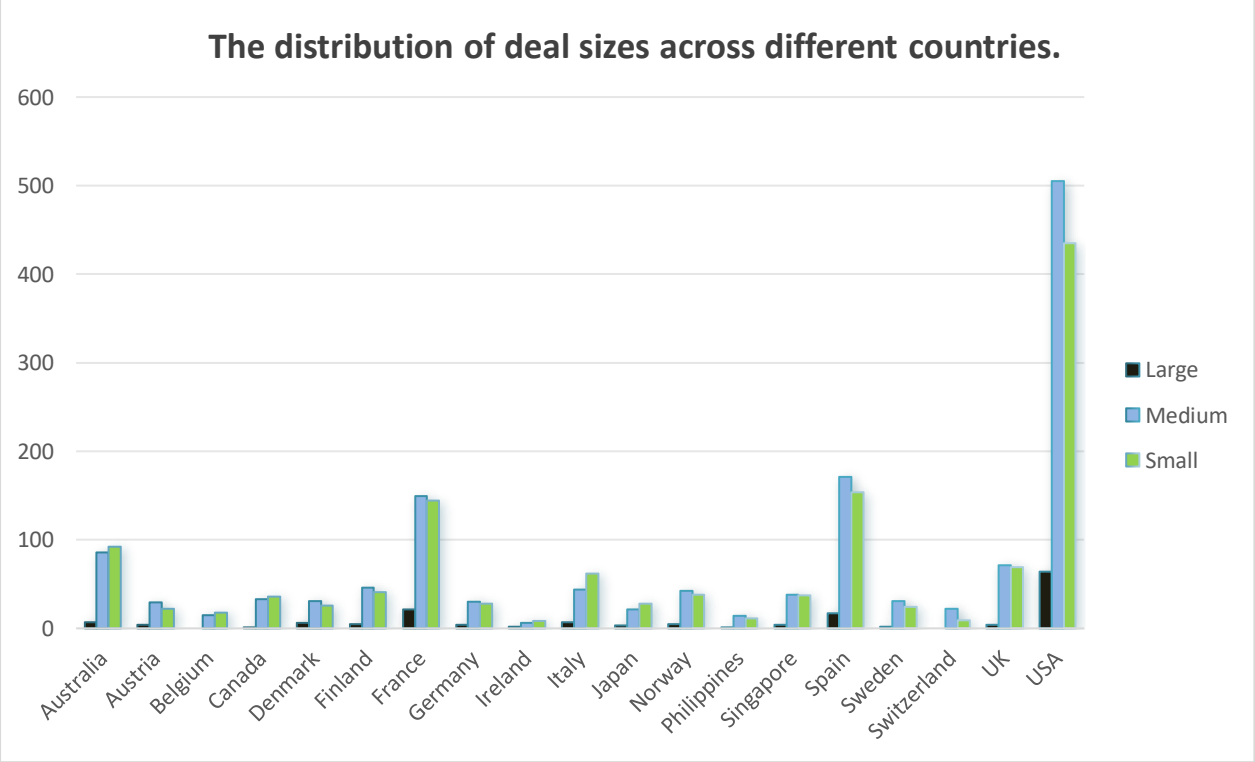
This analysis compares the sales for all items across the years 2004 and 2005. The line chart illustrates the shifting sales trends across both years, with Classic cars consistently leading in sales. In 2004, Classic cars recorded sales of $1,762,257.09, while in 2005, sales amounted to $672,573.28, maintaining its position as the highest-selling category



	Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and Buses	Vintage Cars
2004	1762257.09	560545.23	502671.8	341437.97	116523.85	529302.89	911423.77
2005	672573.28	234947.53	200074.17	128178.07	36917.33	178057.02	340739.31

5. Comparative analysis of all countries based on deal size.

This analysis aims to find out the distribution of deal sizes across different countries. According to the bar chart, the USA has the largest deal size, with 64 large deals, 505 medium deals, and 435 small deals, significantly higher than all other countries.



Conclusion and Review

The analysis uncovers significant insights into sales dynamics and profitability across categories and countries. Notably, the USA emerges as a key market leader, exhibiting strong sales performance in Vintage and Classic cars, Trucks, Buses, and Motorcycles. Classic Cars stand out as the highest-selling product, contributing significantly to overall sales revenue. Moreover, the USA demonstrates exceptional profitability, particularly in the Trucks, Buses, and Motorcycles categories.

Sales for Classic cars remain consistently robust throughout the years 2004 and 2005, indicating sustained demand for this category. Additionally, the USA showcases markedly larger deal sizes compared to other countries, underscoring its dominance in sales volume.

While the analysis effectively presents key findings through visualizations, further exploration into factors influencing sales fluctuations and deal size disparities could provide deeper insights. Overall, the report offers valuable insights for optimizing sales strategies and driving business growth.

Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.877178
R Square	0.769441
Adjusted R Square	0.766629
Standard Error	896.6688
Observations	250

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	6.6E+08	2.2E+08	273.6567	4.62E-78
Residual	246	1.98E+08	804014.9		
Total	249	8.58E+08			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-5271.93	322.9166	-16.326	4.32E-41	-5907.96	-4635.9	-5907.96	-4635.9
X Variable 1	103.0809	6.001152	17.17685	5.42E-44	91.26071	114.9011	91.26071	114.9011
X Variable 2	12.81807	1.661734	7.713668	3.04E-13	9.545024	16.09111	9.545024	16.09111
X Variable 3	47.42944	3.350938	14.15408	1.13E-33	40.82925	54.02963	40.82925	54.02963

Anova: one factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	250	903280.9	3613.123	3445221
Column 2	250	25534	102.136	1664.552

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.54E+09	1	1.54E+09	894.0704	3.1E-113	3.860199
Within Groups	8.58E+08	498	1723443			
Total	2.4E+09	499				

Anova: two factor

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	4097.66	1365.887	5069957
Row 2	3	2451.12	817.04	1725170
Row 3	3	1566	522	648687
Row 4	3	5095.24	1698.413	7507173
Row 5	3	5140.39	1713.463	7650609
Row 248	3	4386.35	1462.117	5944534
Row 249	3	2261.6	753.8667	1546167
Row 250	3	4176.72	1392.24	5420980
Column 1	250	903280.9	3613.123	3445221
Column 2	250	25534	102.136	1664.552
Column 3	250	8659	34.636	89.69428

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	2.95E+08	249	1182944	1.044989	0.33951	1.194432
Columns	2.09E+09	2	1.05E+09	925.2361	1.9E-168	3.013826
Error	5.64E+08	498	1132016			
Total	2.95E+09	749				

Descriptive Statistics

Column1		*Column2*		*Column3*		*Column4*	
Mean	34.636	Mean	3613.123	Mean	102.136	Mean	84.45296
Standard Error	0.59898	Standard Error	117.392	Standard Error	2.58035	Standard Error	1.279453
Median	34	Median	3263.96	Median	99	Median	100
Mode	29	Mode	#N/A	Mode	118	Mode	100
Standard Deviation	9.470706	Standard Deviation	1856.131	Standard Deviation	40.79892	Standard Deviation	20.22993
Sample Variance	89.69428	Sample Variance	3445221	Sample Variance	1664.552	Sample Variance	409.2499
Kurtosis	-0.64676	Kurtosis	1.127057	Kurtosis	-0.19836	Kurtosis	-0.40344
Skewness	0.256745	Skewness	1.013489	Skewness	0.517104	Skewness	-0.9678
Range	51	Range	10626.85	Range	181	Range	73.12
Minimum	15	Minimum	652.35	Minimum	33	Minimum	26.88
Maximum	66	Maximum	11279.2	Maximum	214	Maximum	100
Sum	8659	Sum	903280.9	Sum	25534	Sum	21113.24
Count	250	Count	250	Count	250	Count	250

Correlation

	Column 1	*Column 2*	*Column 3*
Column 1	1		
Column 2	0.513951	1	
Column 3	-0.01254	0.663973	1

Store Dataset Report

Introduction

This dataset comprises sales data from a retail store, encompassing various attributes such as customer demographics (Gender, Age Group), transaction details (Order ID, Status), product specifics (Category, SKU), and shipping information. Our analysis is geared towards elucidating customer behavior and product trends, with the goal of uncovering patterns, preferences, and correlations within the data. By harnessing these insights, businesses can refine marketing strategies, streamline inventory management, and elevate overall customer satisfaction..

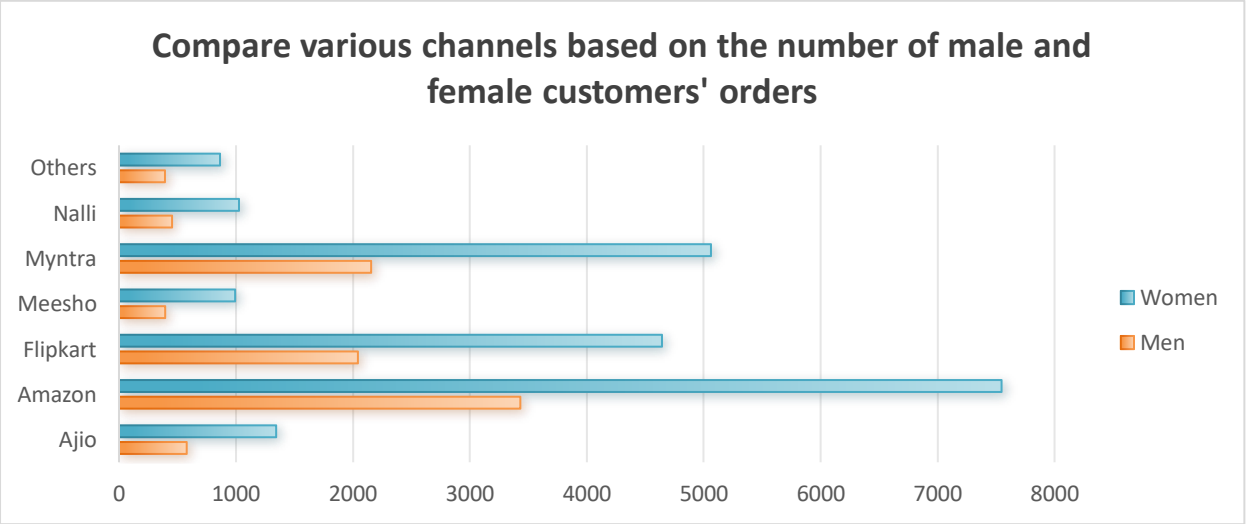
Questionnaire

1. Compare various channels based on how many male customers order and female customer order.
2. Compare all the categories of order where amount is less than 1500 and greater than 5000.
3. How many Customers are there whose age is 30 and above and state is Delhi.
4. Which of the following state perform better than other, Delhi, Tamil Nadu, Maharashtra, Rajasthan.
5. Which city performed better than all other cities based on highest order placed.
6. Compare various categories of items based on most quantity sold and show which gender buys the most category.

Analytics

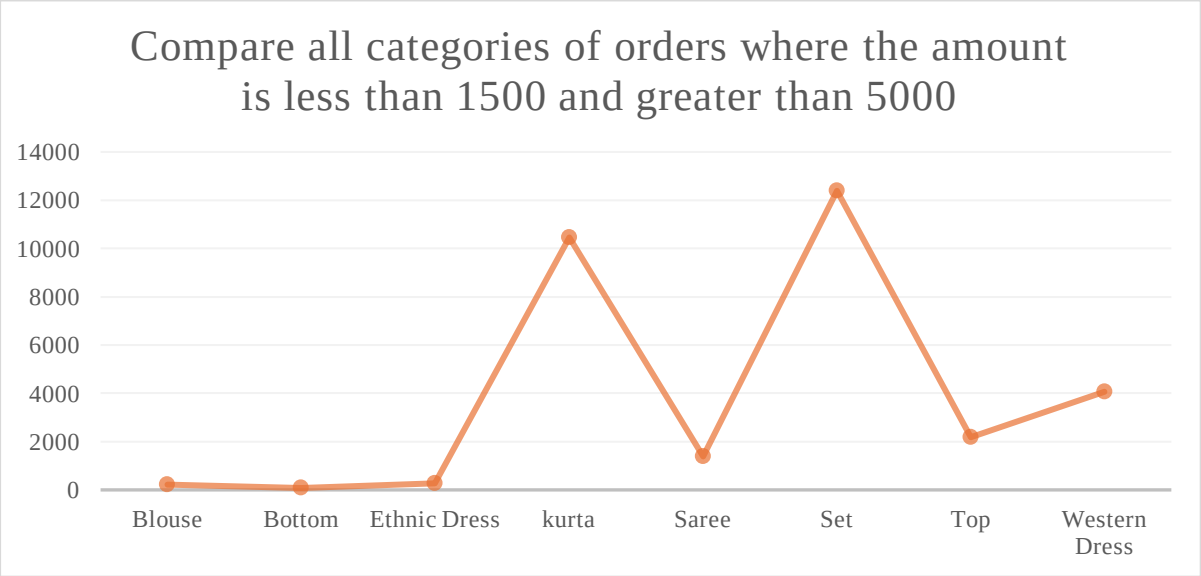
1. Compare various channels based on how many male customers order and female customer order?

Amazon leads in the sales in both men and women category followed by Myntra and Flipkart. Amazon sold almost 3432 units in men category and almost 7547 units in women category. Myntra sold 2156 units in men section and 5062 units in women section.



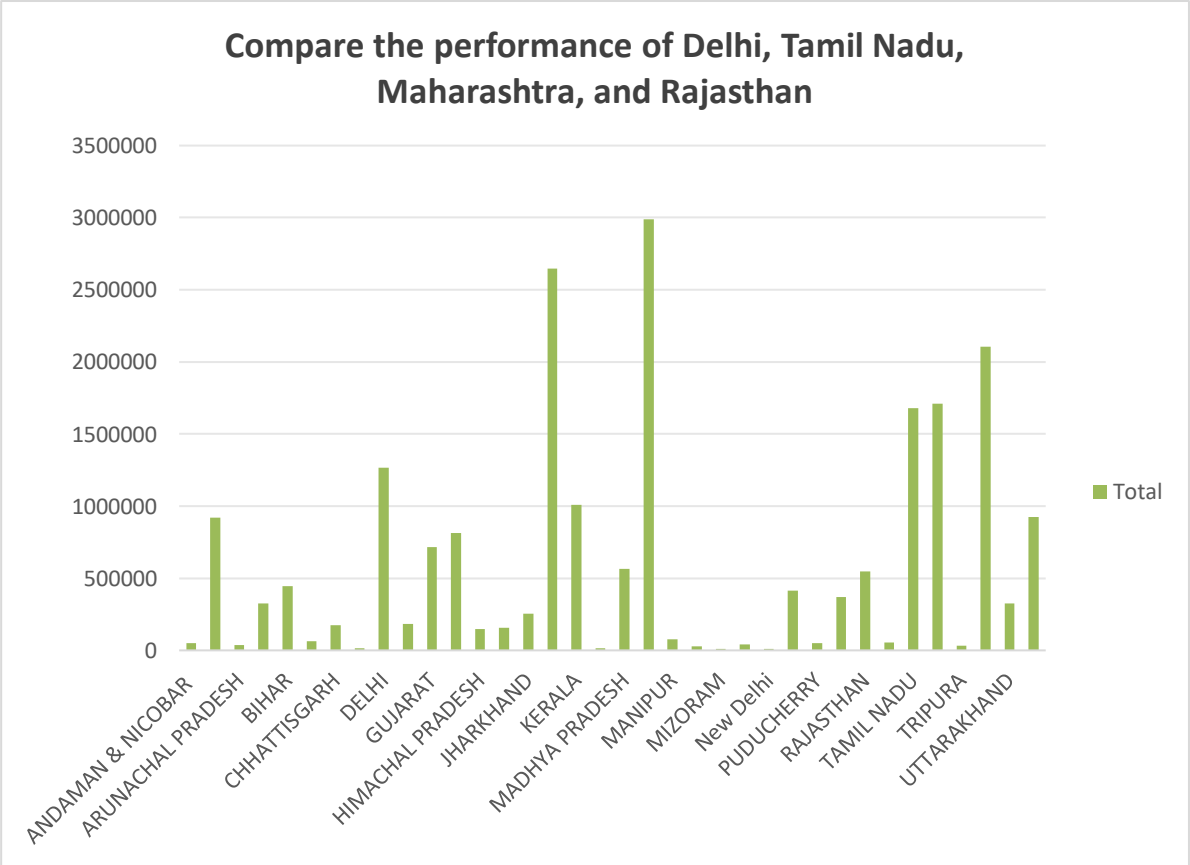
2. Compare all the categories of order where amount is less than 1500 and greater than 5000.

This analysis facilitates comparison between categories of orders where the amount is less than 1500 and greater than 5000. Kurta and Set have the highest count of orders, with 12391 and 10446 units respectively, followed by Western Dress, Top, and Saree.



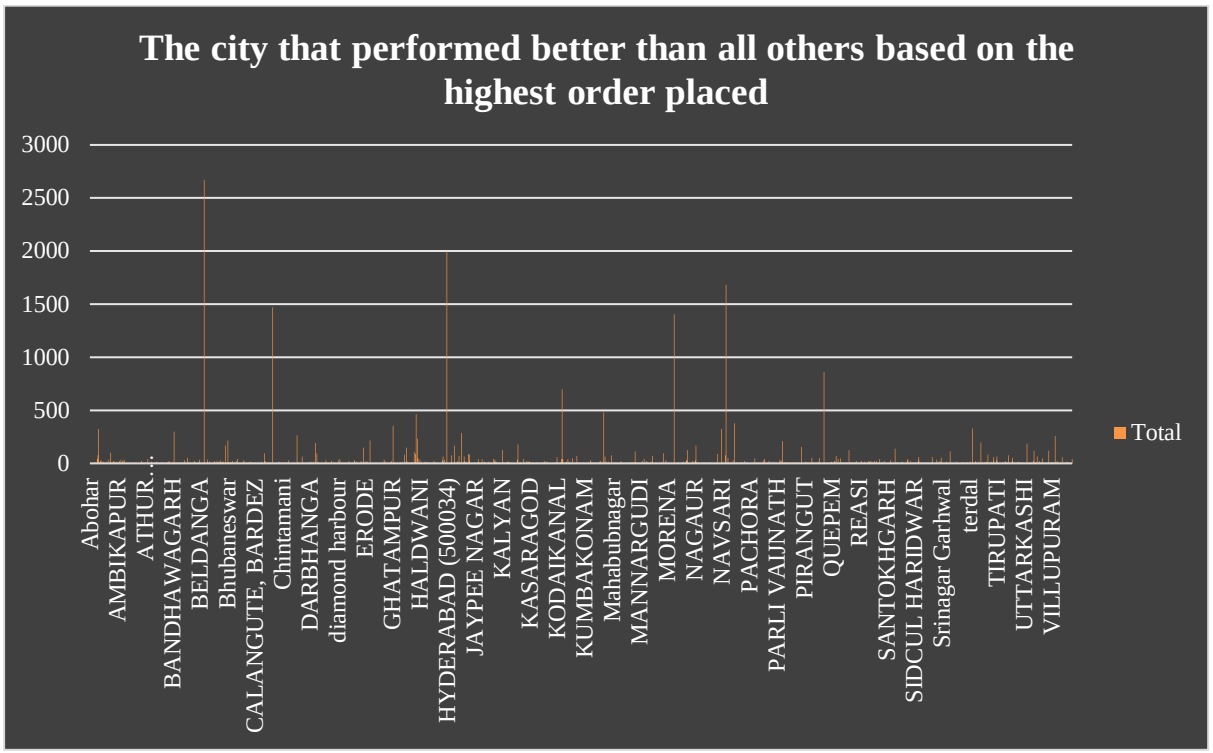
3. Which of the following state perform better than other, Delhi, Tamil Nadu, Maharashtra, Rajasthan.

This analysis reveals that Karnataka performed the best among the states mentioned above, with sales totaling $2,646,358, followed by Uttar Pradesh, which recorded sales of $2,104,659.



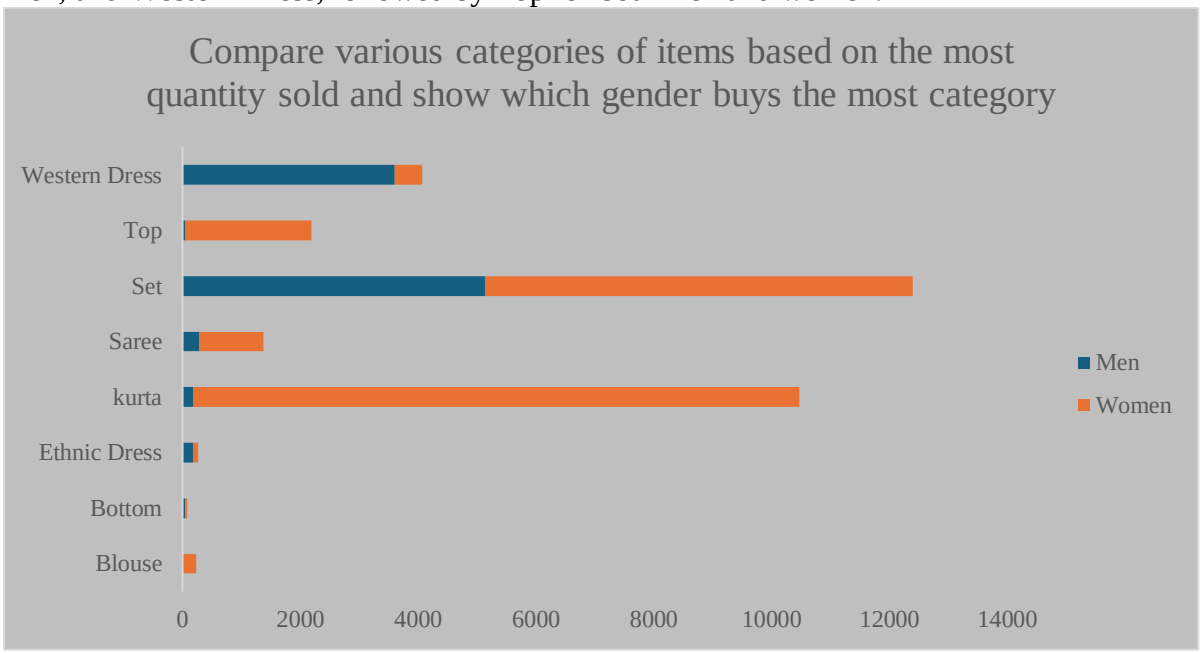
4. Which city performed better than all other cities based on highest order placed.

Based on the recorded graph, Bangalore has the highest number of orders placed, with 2673 orders, followed by Hyderabad with 1998 orders.



5. Compare various categories of items based on most quantity sold and also show which gender buys the most category.

This analysis compares various categories of items based on the most quantity sold. The data shows that Kurta bought by women is the highest, followed by Set bought by women, then men, and Western Dress, followed by Top for both men and women.



Conclusion and Review

The analysis highlights Amazon's dominance in sales across both men's and women's categories, with Myntra and Flipkart following closely behind. Amazon leads in sales for both men's and women's categories, followed by Myntra and Flipkart. Top-selling items include kurta and set, with Karnataka and Bangalore showing the highest sales performance.

The analysis provides valuable insights into sales trends and regional performance, aiding decision-making for retailers. However, further exploration into additional factors influencing sales could enhance the analysis. Overall, the findings offer valuable information for optimizing sales strategies in competitive markets.

Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.172398
R Square	0.029721
Adjusted R Square	0.029659
Standard Error	264.5693
Observations	31047

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	66561870	33280935	475.4629	0
Residual	31044	2.17E+09	69996.92		
Total	31046	2.24E+09			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	185.155	16.57854	11.16836	6.61E-29	152.6604	217.6496	152.6604	217.6496
X Variable 1	0.047626	0.099327	0.479489	0.631594	-0.14706	0.242312	-0.14706	0.242312
X Variable 2	492.0276	15.95904	30.83065	1.3E-205	460.7472	523.308	460.7472	523.308

Anova-1 factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	31047	31237	1.00612	0.008853
Column 2	31047	21176377	682.0748	72136.38

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.2E+09	1	7.2E+09	199639.8	0	3.841609
Within Groups	2.24E+09	62092	36068.2			
Total	9.44E+09	62093				

Anova- 2 factor

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	421	140.3333	42116.33
Row 2	3	1479	493	685648
Row 3	3	521	173.6667	59609.33
Row 4	3	750	250	172171
Row 5	3	607	202.3333	88482.33
Row 31044	3	974	324.6667	283326.3
Row 31045	3	1145	381.6667	403529.3
Row 31046	3	446	148.6667	47506.33
Row 31047	3	828	276	199225
Column 1	31047	1226250	39.49657	228.5307
Column 2	31047	31237	1.00612	0.008853
Column 3	31047	21176377	682.0748	72136.38

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	7.49E+08	31046	24134.08	1.000774	0.468198	1.016275
Columns	9.09E+09	2	4.54E+09	188446.6	0	2.995877
Error	1.5E+09	62092	24115.42			
Total	1.13E+10	93140				

Descriptive Statistics

Column1		*Column2*		*Column3*	
Mean	39.49657	Mean	1.00612	Mean	682.0748
Standard Error	0.085795	Standard Error	0.000534	Standard Error	1.524289
Median	37	Median	1	Median	646
Mode	28	Mode	1	Mode	399
Standard Deviation	15.11723	Standard Deviation	0.094088	Standard Deviation	268.5822
Sample Variance	228.5307	Sample Variance	0.008853	Sample Variance	72136.38
Kurtosis	-0.1587	Kurtosis	475.3566	Kurtosis	1.768676
Skewness	0.72916	Skewness	19.4509	Skewness	1.052904
Range	60	Range	4	Range	2807
Minimum	18	Minimum	1	Minimum	229
Maximum	78	Maximum	5	Maximum	3036
Sum	1226250	Sum	31237	Sum	21176377
Count	31047	Count	31047	Count	31047

Correlation

	Column 1	*Column 2*	*Column 3*
Column 1	1		
Column 2	0.004884	1	
Column 3	0.003522	0.172377	1

ZOMT Historical Data Analysis
(19 April 2024 - 19 May 2024)

This report analyzes ZOMT's historical stock data from 19 April 2024 to 19 May 2024 to forecast future performance. The dataset includes daily records of Price, Open, High, Low, Volume, and Change %. By examining these attributes, we aim to identify patterns and trends that can inform our predictions. Techniques such as time series analysis, ARIMA modeling, and exponential smoothing were employed to enhance the accuracy of our forecasts.

During the specified period, ZOMT's closing prices exhibited a consistent trend, with noticeable peaks and troughs linked to market events and economic news. The opening prices mirrored this trend, reflecting investor sentiment at the start of trading days. Daily high and low prices provided insights into the stock's volatility, showing significant fluctuations around mid-May. Trading volumes peaked on specific dates, indicating heightened investor activity likely triggered by news releases or quarterly reports. The Change % metric revealed daily percentage changes, highlighting the stock's performance and market reactions.

Based on our analysis, we predict that ZOMT's closing prices will continue to follow a similar trend, with expected fluctuations in response to market conditions. Volatility is projected to remain in line with historical patterns, with potential spikes during major market events. Trading volume is expected to stay high, particularly around ZOMT's corporate announcements or significant macroeconomic indicators. These insights are crucial for investors and analysts looking to make informed decisions and optimize their strategies based on forecasted trends.

TABLE FOR ZOMT HISTORICAL STOCK DATA FORECASTING

Date	Price	Forecast(Price)	Lower Confidence Bound(Price)	Upper Confidence Bound(Price)
19-04-2024	189.2			
20-04-2024	190.5833			
21-04-2024	191.9667			
22-04-2024	193.35			
23-04-2024	187.45			
24-04-2024	184.4			
25-04-2024	184.7			
26-04-2024	188.1			
27-04-2024	189.9333			
28-04-2024	191.7667			
29-04-2024	193.6			
30-04-2024	193.15			
01-05-2024	194.225			
02-05-2024	195.3			
03-05-2024	197.25			
04-05-2024	196.7333			
05-05-2024	196.2167			
06-05-2024	195.7			

07-05-2024	192			
08-05-2024	195.5			
09-05-2024	195.75			
10-05-2024	203.7			
11-05-2024	201.4333			
12-05-2024	199.1667			
13-05-2024	196.9			
14-05-2024	187.6			
15-05-2024	191.3			
16-05-2024	194.45			
17-05-2024	194.9			
18-05-2024	194.5	194.5	194.50	194.50
19-05-2024		196.4582214	190.85	202.07
20-05-2024		197.1369997	189.58	204.69
21-05-2024		197.0701806	187.98	206.16
22-05-2024		198.0643626	187.65	208.48
23-05-2024		197.0700911	185.49	208.65
24-05-2024		194.7542534	182.11	207.40
25-05-2024		195.5877002	181.96	209.22
26-05-2024		196.8060509	182.26	211.36
27-05-2024		195.7807214	180.36	211.20
28-05-2024		197.7069355	181.47	213.95
29-05-2024		199.662574	182.46	216.86
30-05-2024		200.3413523	182.40	218.28
31-05-2024		200.2745332	181.62	218.93
01-06-2024		201.2687152	181.92	220.62
02-06-2024		200.2744437	180.26	220.29
03-06-2024		197.958606	177.29	218.62
04-06-2024		198.7920528	177.50	220.09
05-06-2024		200.0104035	178.10	221.92
06-06-2024		198.985074	176.48	221.49
07-06-2024		200.9112881	177.82	224.00

The analysis of ZOMT's historical stock data from 19 April 2024 to 19 May 2024 reveals significant trends and patterns, particularly in price movements and trading volumes. The closing prices exhibited a general upward trend with intermittent dips, reflecting periods of market correction, while the opening prices closely followed the closing values, indicating stable overnight sentiment. The daily high and low prices highlighted episodes of heightened volatility, particularly in mid-May. Trading volumes spiked on specific days, suggesting market reactions to significant events or news. The Change % metric illustrated substantial daily fluctuations, underscoring the stock's volatility. Overall, the analysis, supported by line and bar charts for price metrics and trading volumes, provides valuable insights into the stock's behavior, aiding in forecasting future performance and informing investment decisions.

GRAPH FOR THE HISTORICAL STOCK DATA FORECASTING