DA Lab File

A Data Analytics Report Submitted to

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal
Towards Partial Fulfillment for the Award of

Bachelor of Technology
(Computer Science and Engineering)

Under the Supervision of
Prof. Anurag Punde

Submitted By
Adarsh Trivedi
(0827CS223D02)

Department of Computer Science and Engineering
Acropolis Institute of Technology & Research, Indore
Jan-June2024

S.No.	Experiment	Remarks
1.	Data Analysis Questions: i. 5V's of Big Data ii. Data Analysis Principles iii. Statistical Analytics iv. Hypothesis Testing v. Regression vi. Correlation vii. ANOVA	
2.	Dashboards: i. Dashboard of Car Data ii. Dashboard of Order Data iii. Dashboard of Cookie Data iv. Dashboard of Loan Data v. Dashboard of Shop Sales Data vi. Dashboard of Sales Data Samples vii. Dashboard of Store Dataset	
3.	Reports: i. Car Collection Data Report ii. Order Data Report iii. Cookie Data Report iv. Loan Data Report v. Shop Sales Data Report vi. Sales Data Sample Report vii. Store Dataset Report	
4.	Forecasting of Netflix Shares	

Assignment-1

5V’s of Big Data

- **Volume**: This refers to the vast amounts of data generated every second from various sources such as social media, sensors, transactions, and more. The sheer scale of data that organizations have to handle and analyze is massive.

- **Velocity**: This describes the speed at which data is generated, collected, and processed. In many cases, data needs to be processed in real-time or near-real-time to be useful, such as in financial transactions, social media feeds, and IoT applications

- **Variety**: This refers to the different types of data that are available. Data can be structured (like databases), semi-structured (like XML files), and unstructured (like text, video, and audio files). The diversity of data types presents challenges in terms of storage, processing, and analysis.

- **Veracity**: This pertains to the quality and accuracy of the data. High veracity means the data is trustworthy and accurate, while low veracity indicates a higher level of uncertainty and the potential presence of noise or errors in the data.

- **Value**: This is about the usefulness of the data. Data alone doesn't hold value; it needs to be processed and analyzed to extract actionable insights that can drive business decisions and strategies. The ultimate goal of big data initiatives is to derive significant value from the data.

Data Analysis Principles

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.

Statistical Analysis

Statistical Analytics uses statistical methods to collect, review, analyze, and draw conclusions from data. This includes descriptive statistics (mean, median, mode, range, variance, standard deviation) to summarize data features, inferential statistics (hypothesis testing, confidence intervals, regression analysis) to extend conclusions beyond immediate data, predictive analytics

to forecast future outcomes, and prescriptive analytics to recommend actions based on data analysis.

Hypothesis Testing

Hypothesis Testing is a method for making decisions using data from experiments or studies. It involves a null hypothesis (H0) of no effect or difference and an alternative hypothesis (H1) of an effect or difference. The p-value indicates the probability of observing the data if H0 is true, with small p-values suggesting strong evidence against H0. Type I errors (false positives) occur when H0 is wrongly rejected, while Type II errors (false negatives) occur when H0 is wrongly not rejected. The significance level (α), commonly set at 0.05, is the threshold for rejecting H0.

Regression

Regression analysis helps understand relationships between dependent and independent variables. Linear regression fits a linear equation to data, multiple regression uses multiple independent variables, logistic regression predicts probabilities for categorical outcomes, and polynomial regression models relationships as nth degree polynomials.

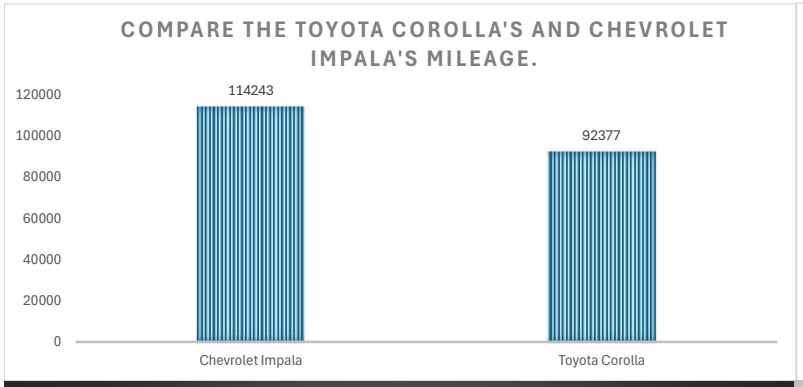
Correlation

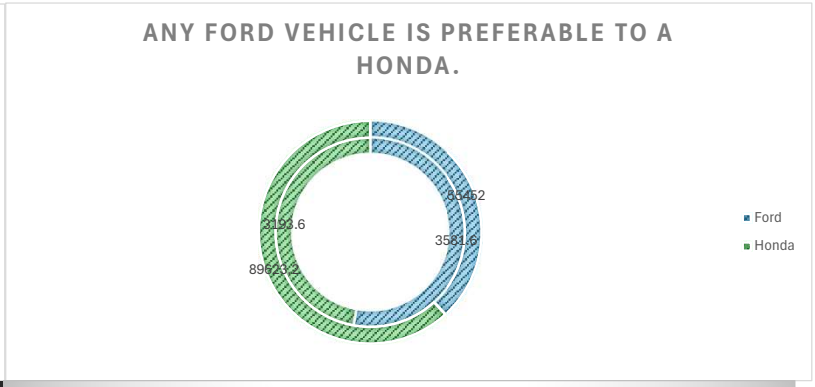
Correlation measures the strength and direction of relationships between two variables using the correlation coefficient (r), ranging from -1 to 1. A positive correlation means both variables move in the same direction, while a negative correlation means one increases as the other decreases. No correlation indicates no relationship. Importantly, correlation does not imply causation; it simply shows a relationship between variables.

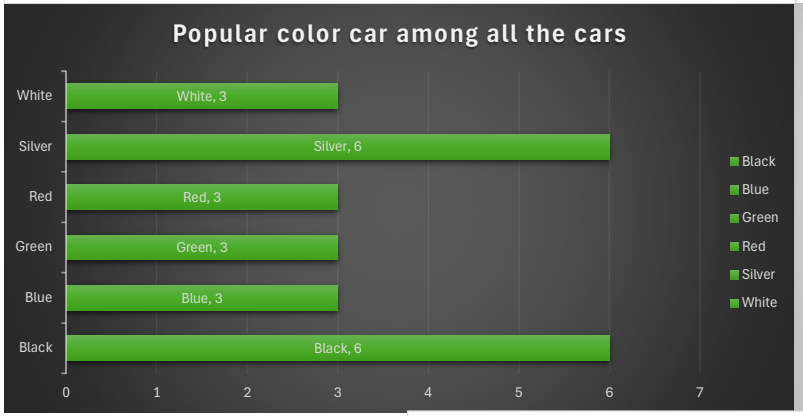
Anova

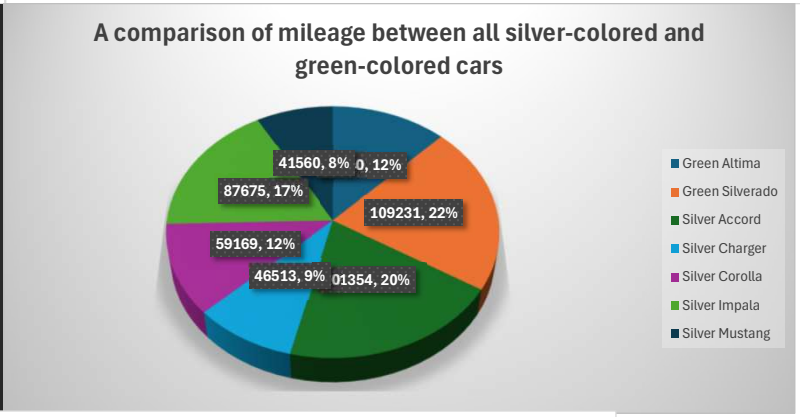
ANOVA (Analysis of Variance) is a method for comparing means across multiple groups to determine if at least one group mean differs significantly. One-way ANOVA compares means across one factor with multiple levels, while two-way ANOVA examines the influence of two categorical variables. ANOVA relies on assumptions of normality, homogeneity of variances, and independence of observations. The F-statistic, the ratio of variance between group means to variance within groups, determines the p-value for the test.

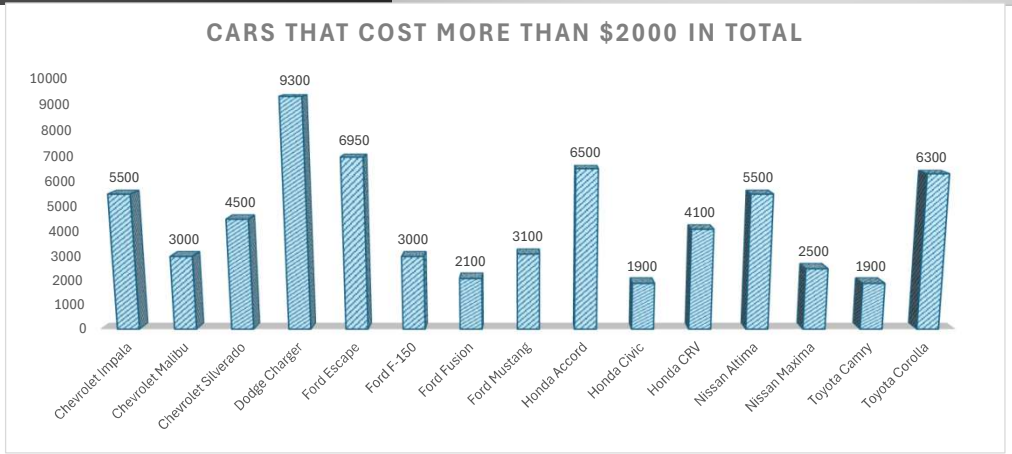
Dasboard of Car data



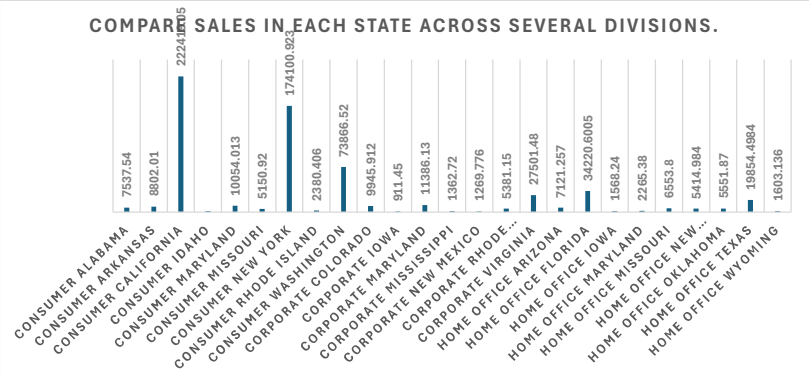


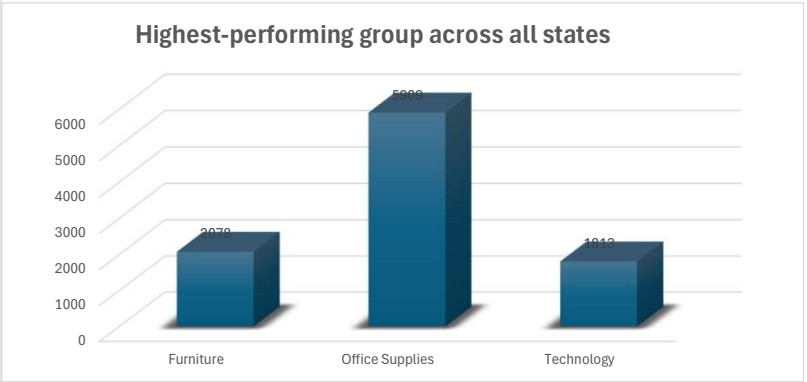


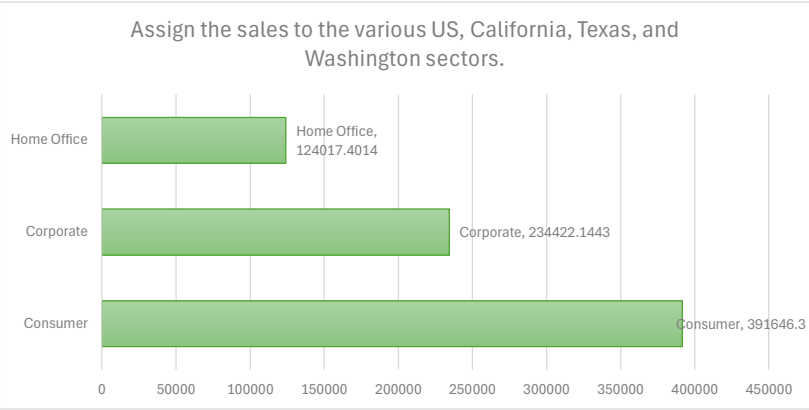


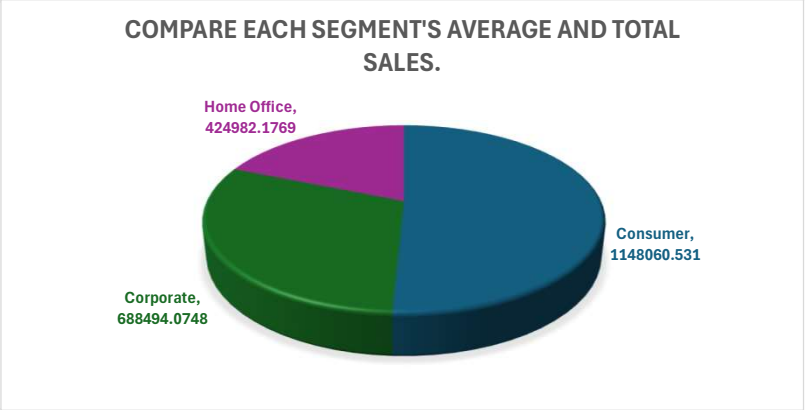


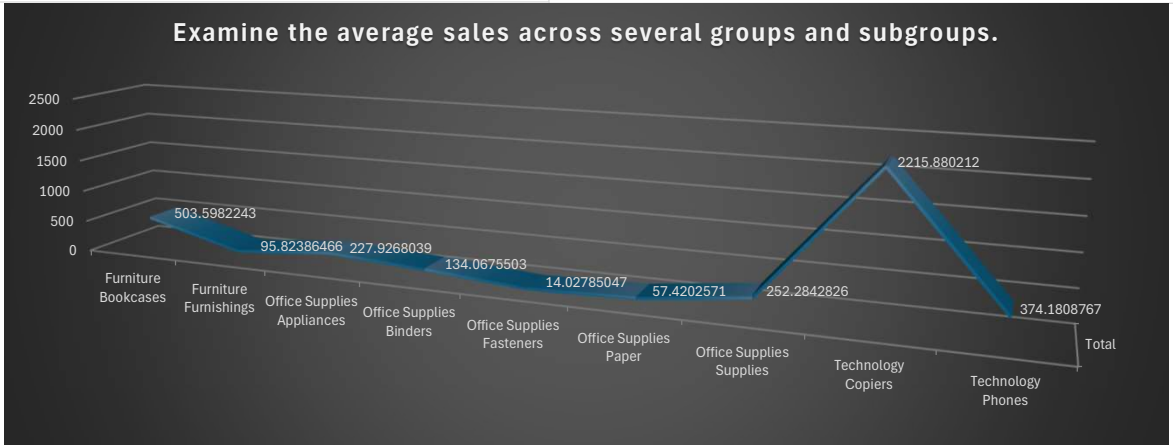
Dasboard of Order data



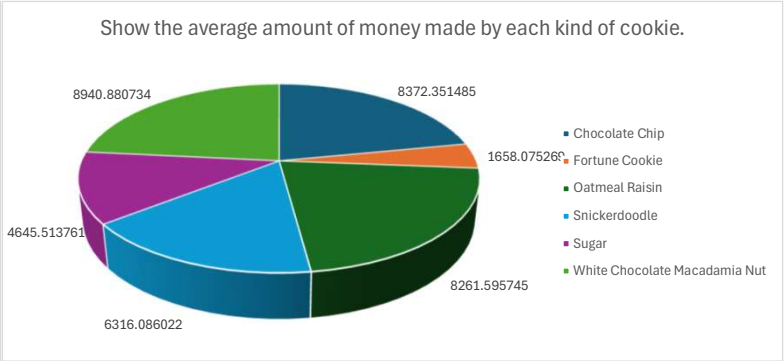


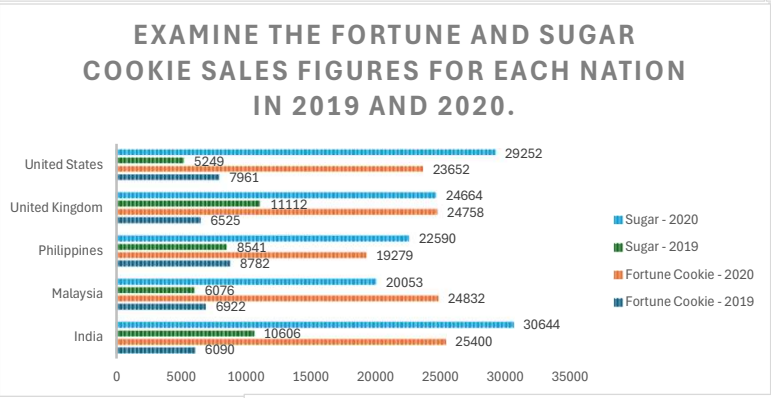


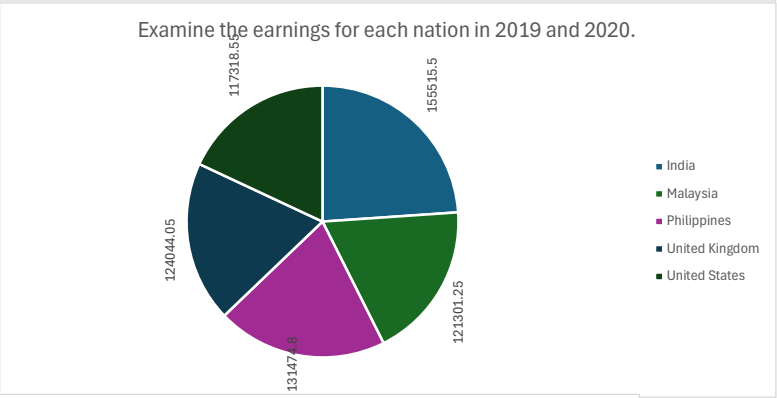


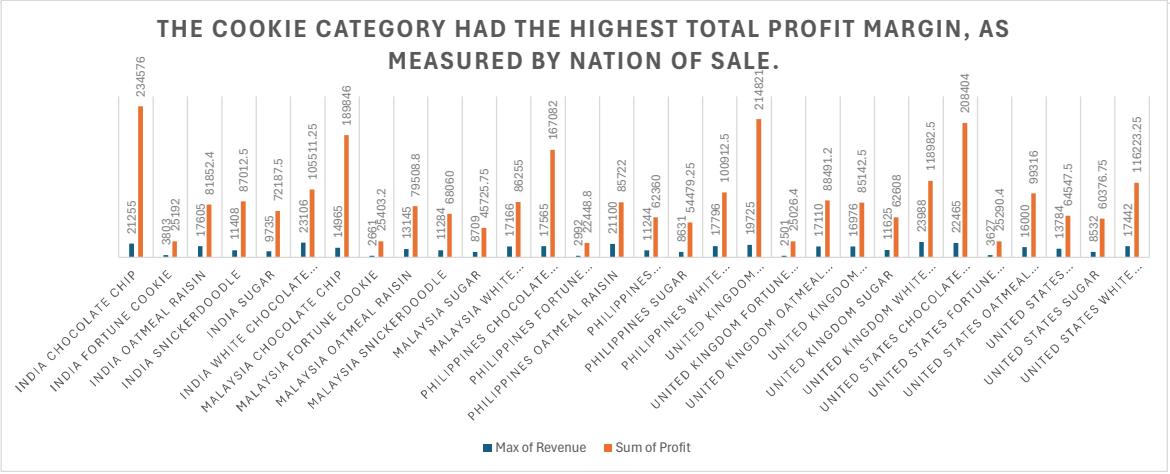


Dasboard of Cookie data

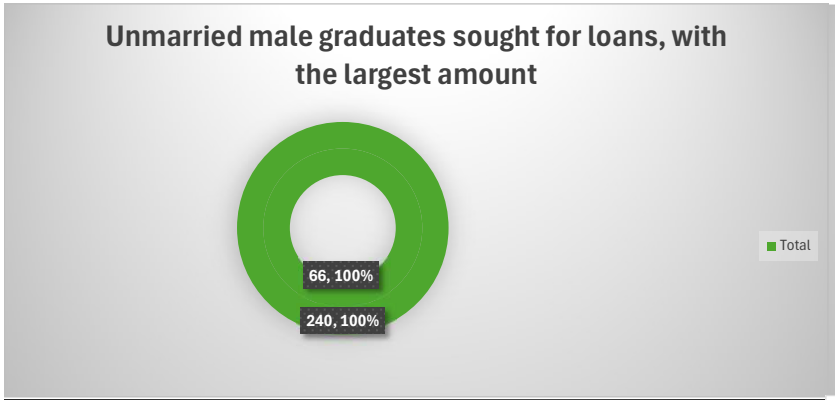


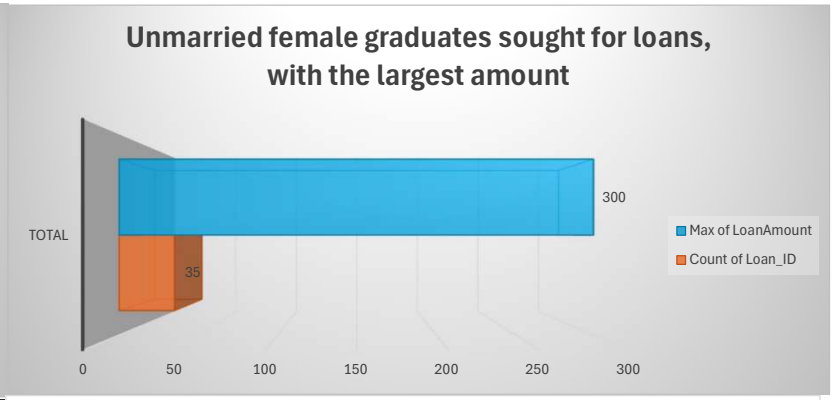


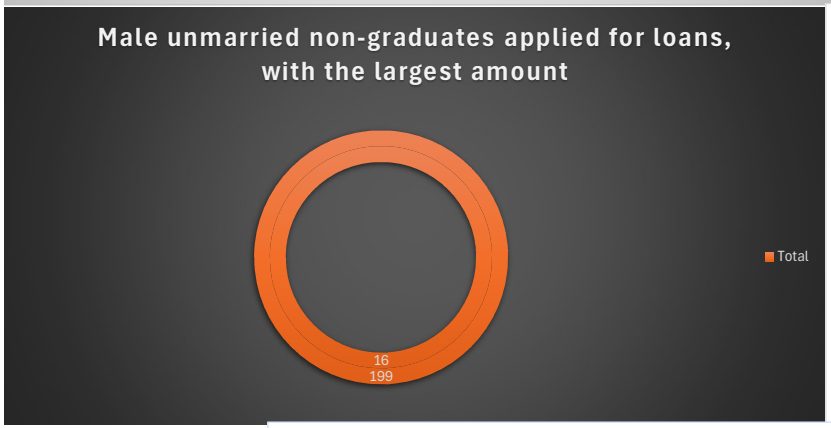


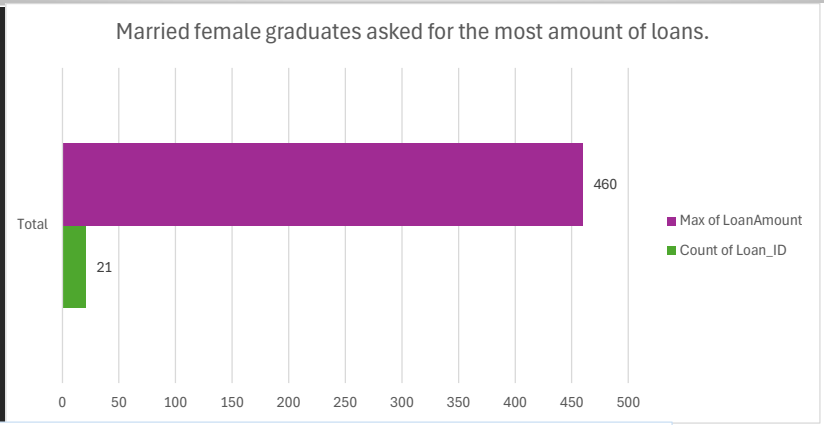


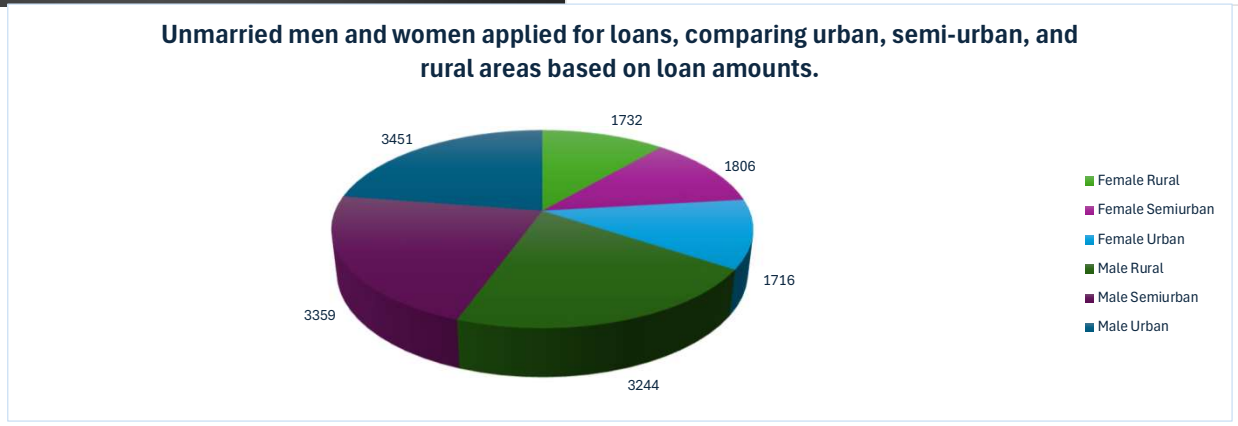
Dasboard of Loan data



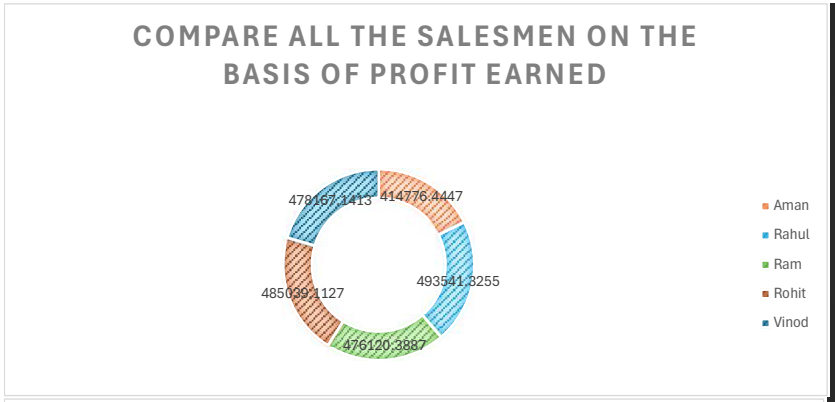


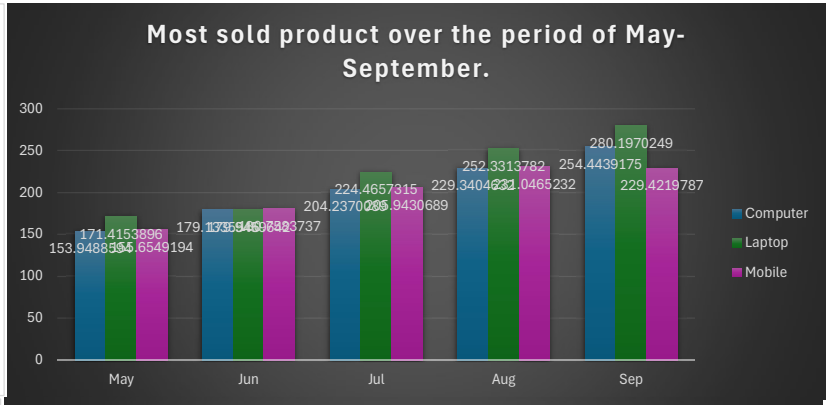


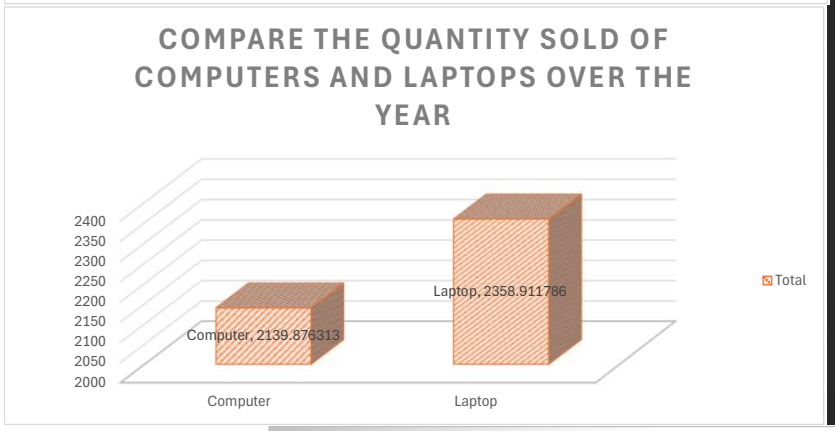


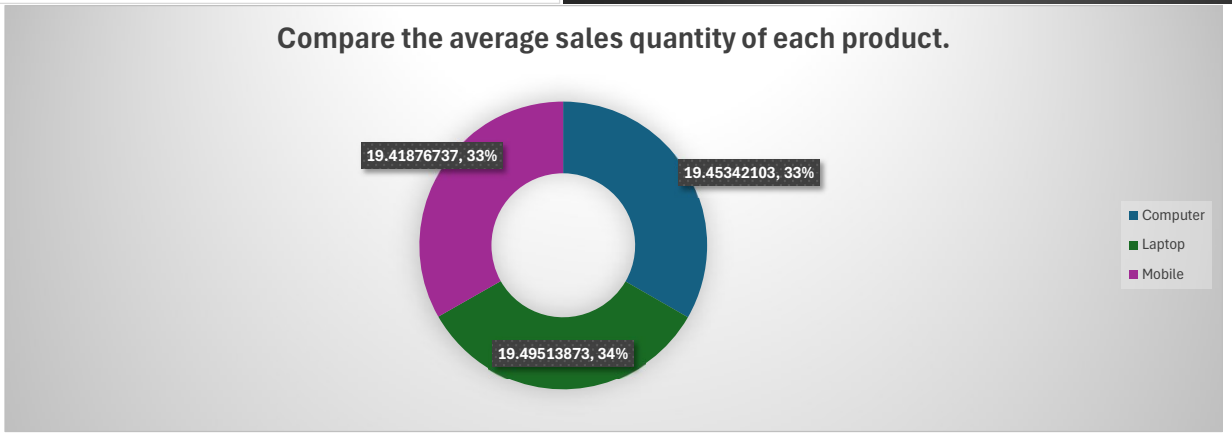


Dasboard of Shop Sales data

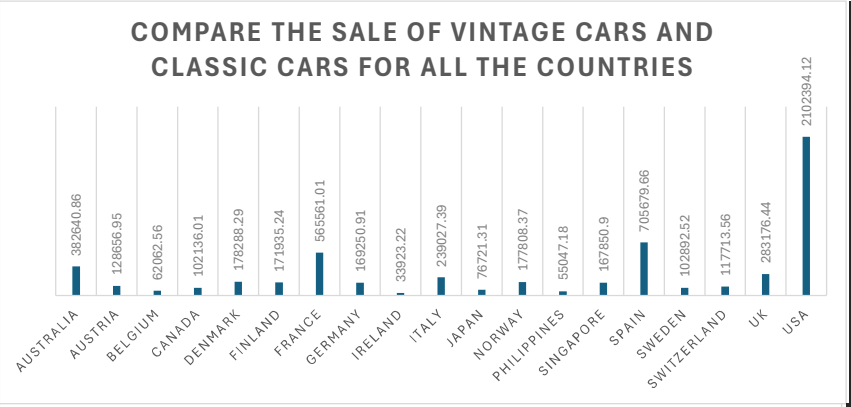


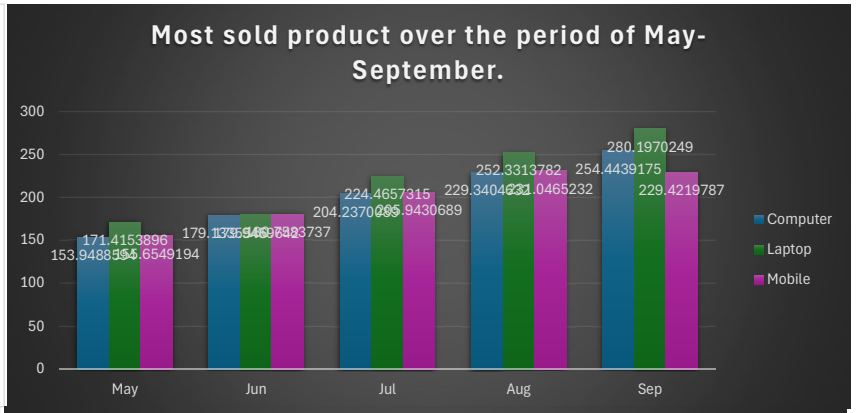


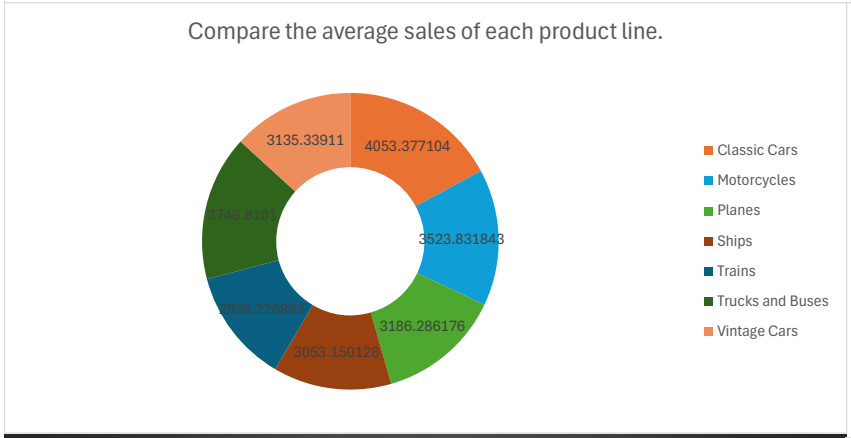


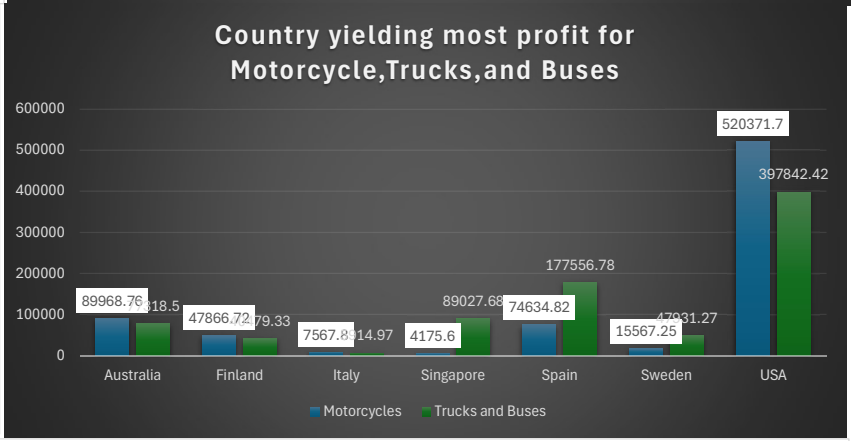


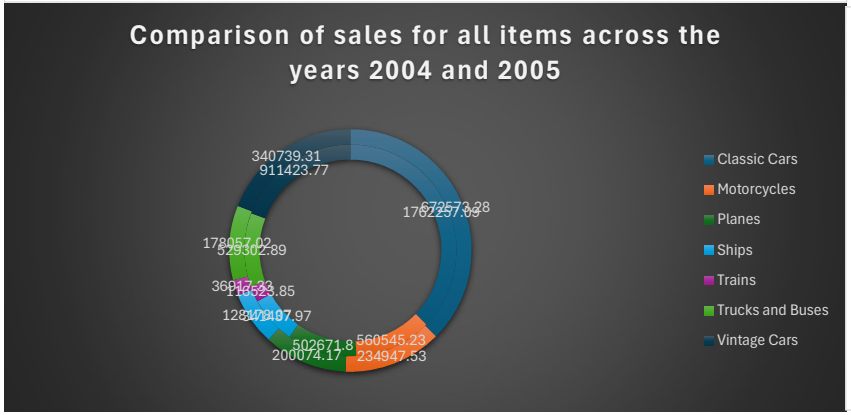
Dasboard of Sales data

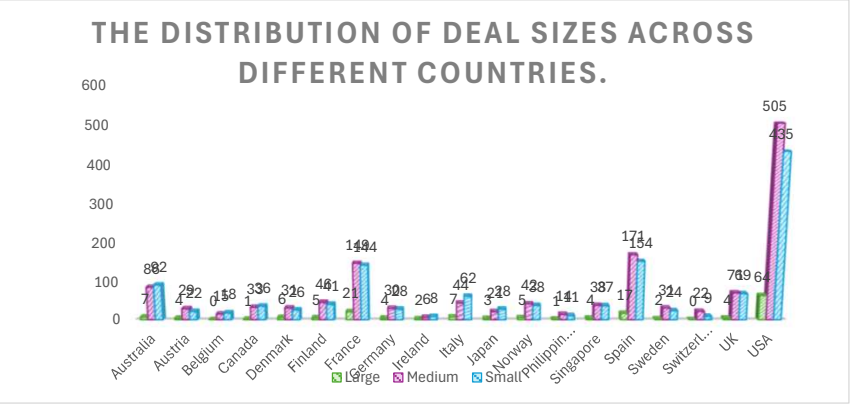




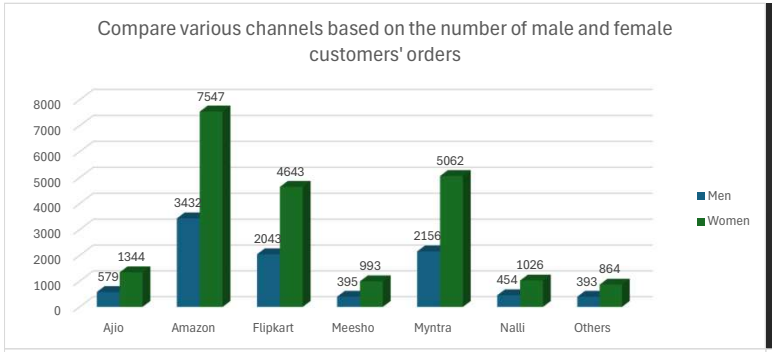


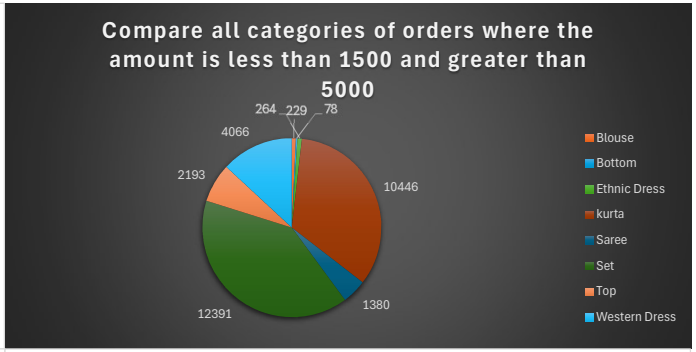


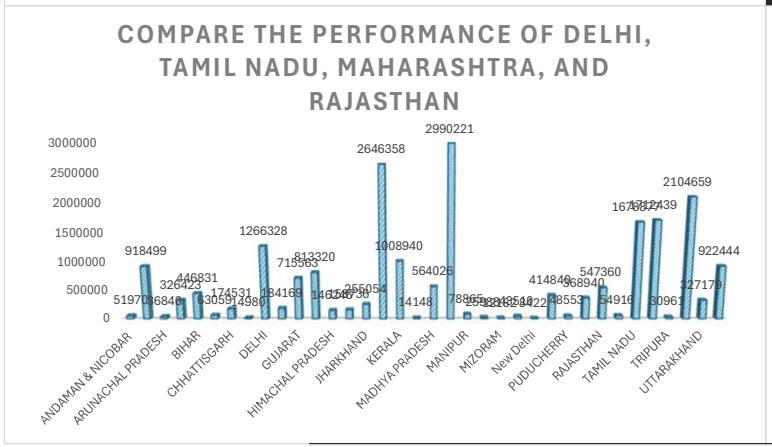


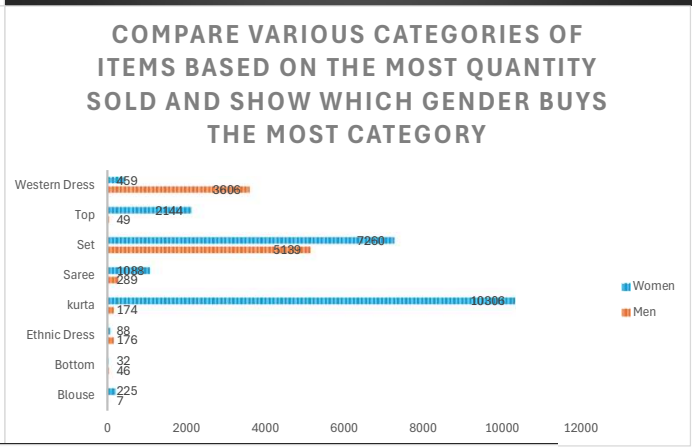


DashBoard of Store Data









Car Collection Data Report

Introduction

This paper offers a thorough examination of a dataset that includes several characteristics of several car models, including make, model, colour, mileage, cost, and pricing. The objective is to obtain knowledge to help with analysing market trends and making decisions about buying cars. Six automobiles are included in the dataset: Ford, Chevrolet, Nissan, Toyota, Dodge, and Honda. This research is aimed at auto fans, analysts, professionals in the automotive business, and anybody with an interest in the developments in the auto market. It consists of interpretations, visualisations, and statistical analysis.

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.

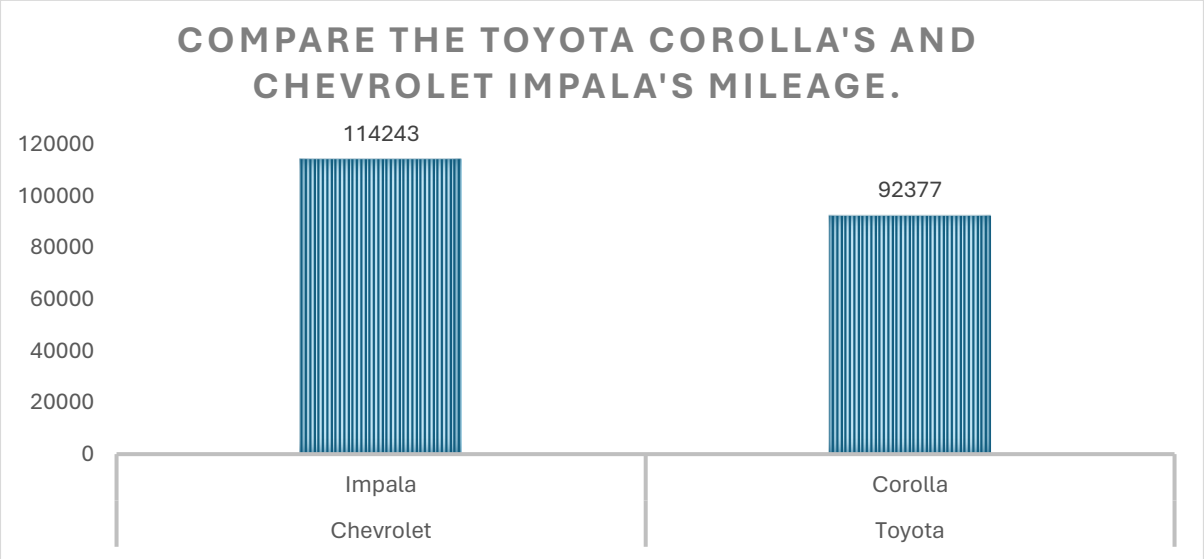
3. Among all the cars which car color is the most popular and is least popular?

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?

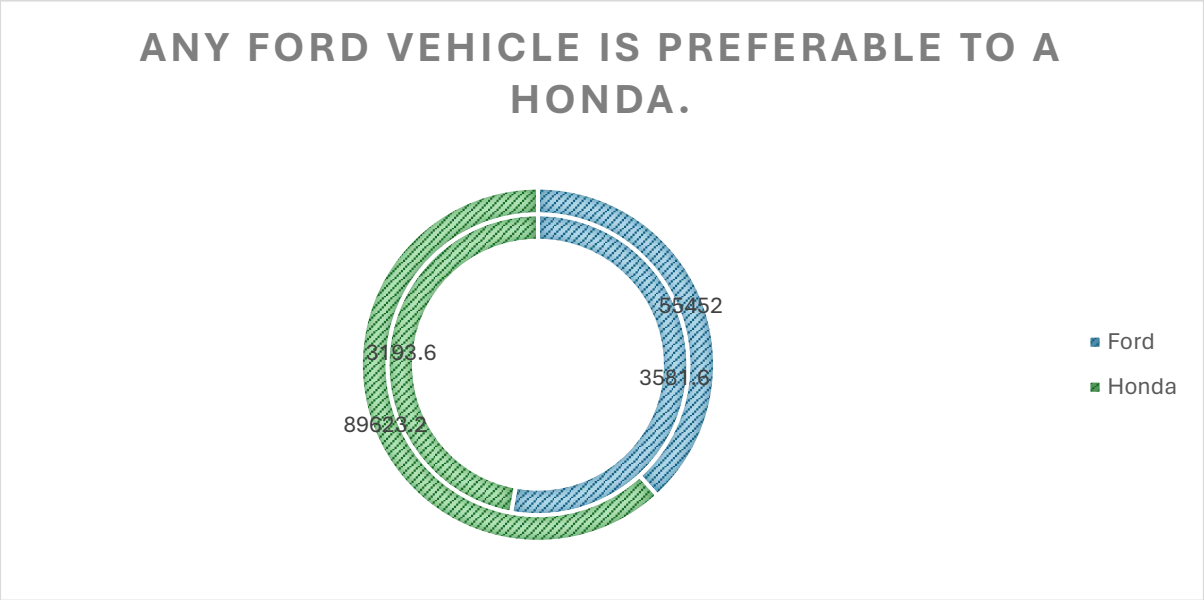
Analytics

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



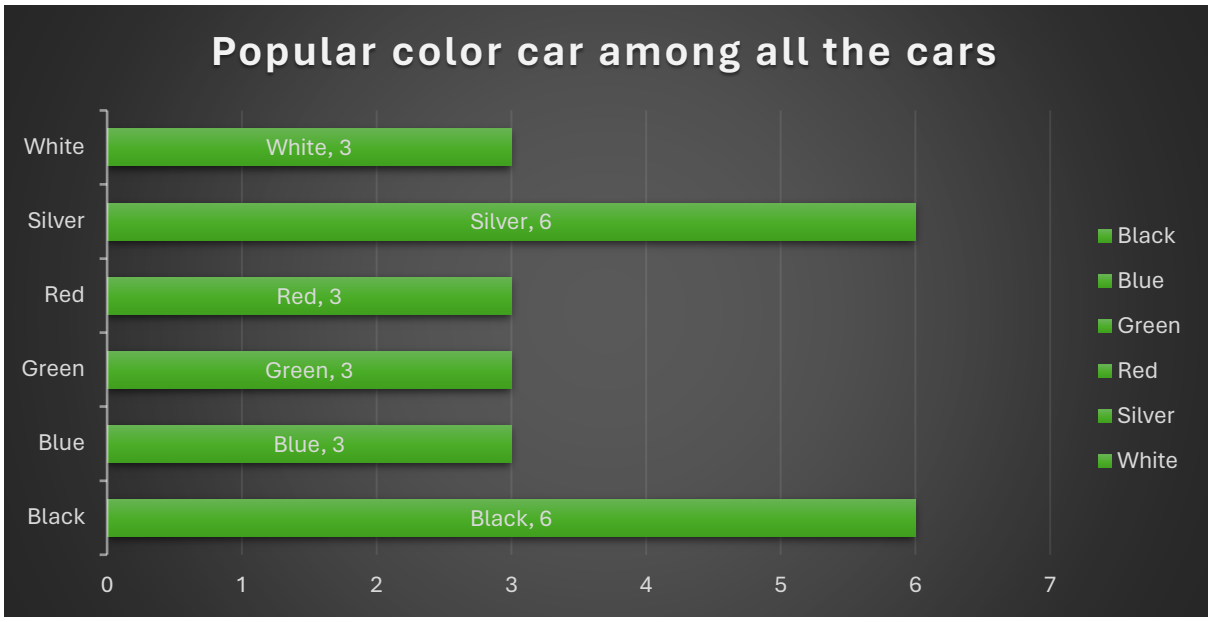
Ans - This study compares the fuel economy (mileage) of the Toyota Corolla to the widely known Chevrolet Impala. In order to do this, relevant data from the dataset was filtered out, and a column chart was made for visualisation. The analysis's conclusions showed that the Toyota Corolla, with its mileage of 92,377, is not as efficient as the Chevrolet Impala, with its 114,243 miles.

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



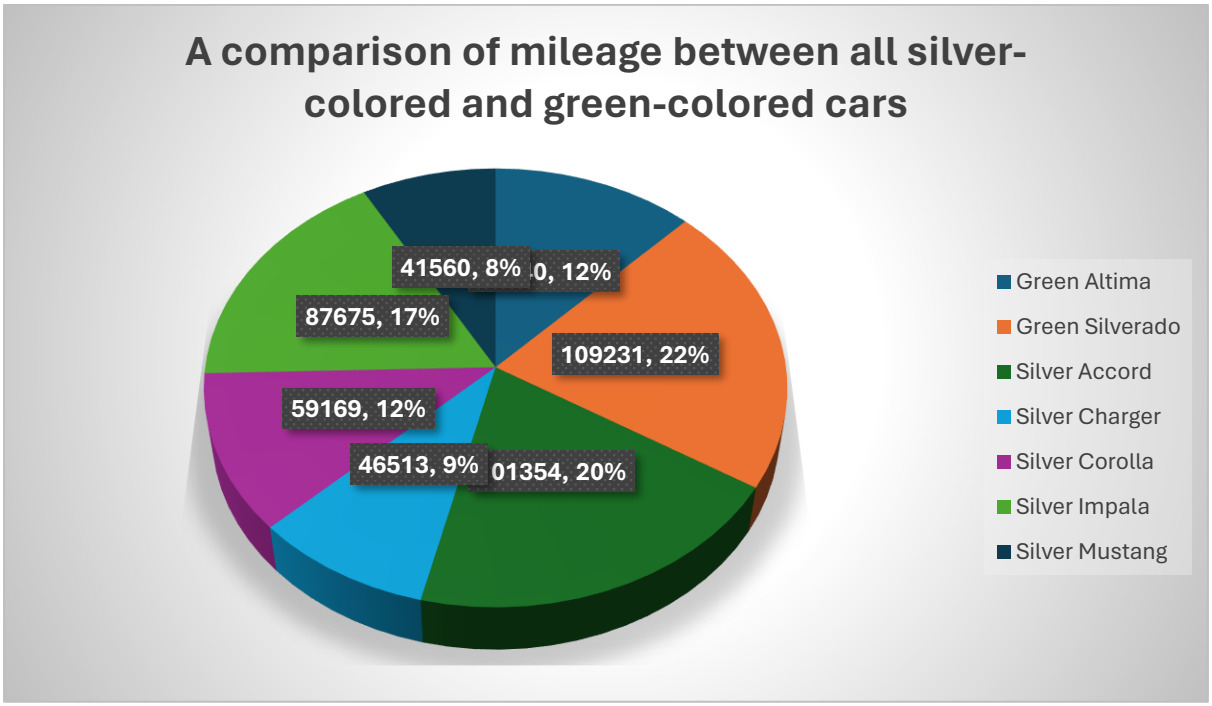
Ans - This research compares the features of Ford and Honda vehicles in an effort to support buying a Ford over a Honda, with an emphasis on affordability. Nevertheless, the results of the dataset analysis refute this assertion. Conversely, it was discovered that Honda vehicles had a lower average cost (3,193.6) and superior average mileage (89,623.3) when compared to Ford vehicles.

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



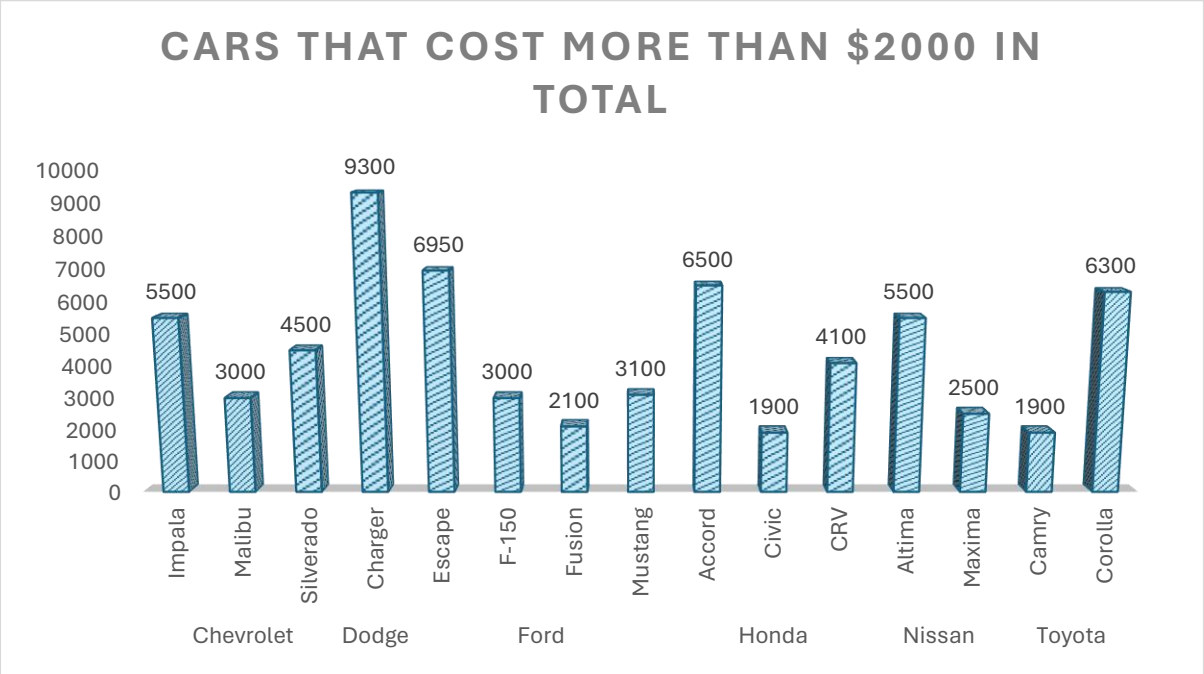
Ans - Based on the count of each make, this study seeks to determine which car colours are the most and least common among all the cars in the dataset. According to the data, Black and White, which together account for 25% of all car makes, are the most popular automobile colours. Green and Blue, on the other hand, are the least common, accounting for 12% of all makes.

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



Ans - The goal of this research is to find vehicles with mileage ranging from silver to green. There are five silver automobiles, according to the insights: Accord, Charger, Corolla, Mustang, and Impala. With an average mileage of 101,354, the Accord has the greatest mileage among these. Two more green vehicles are the Silverado and Altima, the latter of which has the greatest mileage at 109,231.

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



Ans - The purpose of this study is to find cars that cost more than $2,000. When the total of all these expenses is plotted on a bar graph, the total cost of all cars that cost more than $2,000 is displayed as $66,150.

Conclusion and Review

Comparison: The study comparing the Toyota Corolla and Chevrolet Impala's mileage showed that the Chevrolet Impala has superior fuel efficiency.

Ford vs. Honda Comparison: The investigation refuted the basic assumption that Ford vehicles are more cost-effective and had higher mileage than Honda vehicles. When comparing average mileage and pricing to Ford vehicles, Honda vehicles performed better.

common Car Colours: Based on the analysis, the most common car colours are black and white, which account for 25% of all car production. Green and blue, on the other hand, were discovered to be the least common colours, making up a mere 12% of all cars produced.

Comparison of Silver and Green Cars: The Accord had the highest average mileage among silver-colored cars, while the Silverado had the highest mileage among green-colored automobiles.

Automobiles Above $2000: Based on the data, the total amount spent on cars over $2,000 came to $66,150.

The research offered insightful information about a number of dataset components, such as mileage comparisons, the popularity of different automobile colours, and financial considerations. But there were differences between the first hypotheses and the results, especially when comparing Ford and Honda vehicles. The investigation was comprehensive, and the results were presented well using the right visualizations—column charts and bar graphs, for example. All things considered, the study provides insightful information to consumers, business professionals, and scholars who wish to comprehend market developments. However, it's crucial to be aware of the analysis's limitations, including the dataset's completeness and the requirement for more research into other variables impacting auto purchases.

Regression

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

The relationship between the dependent variable (mileage) and the independent variables (cost and price) in the car collection dataset is shown in the Regression Analysis table. By combining Cost and Price, the study shows a high positive correlation (Multiple R = 0.962639) between these variables, explaining approximately 92.67% of the variation in Mileage (R Square = 0.926673). The low p-values and coefficients of Cost and Price indicate that they both have a significant impact on Mileage. Interestingly, Mileage is more affected by Price than by Cost. The ANOVA table, which displays an extremely low p-value (1.22E-12) for the F-statistic, further confirms the overall relevance of the regression model. In conclusion, the regression analysis indicates that, in the car collection dataset, Price and Cost have a significant impact on Mileage.

Anova: one factor

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				

Three groups are examined in the Single Factor ANOVA table: Price, Cost, and Mileage. There are twenty-four observations per group. The total for the Mileage group is 2,011,267, with a variance of 1.21E+09 and an average of 83,802.79. The Cost group has 66,150 as its total, 2,756.25 as its average, and 705,502.7 as its variance. The Price group has 78,108 as its total, 3,254.5 as its average, and 837,024.1 as its variance.

The ANOVA analysis assesses how these groups' means differ from one another. With two degrees of freedom (df), the Between Groups Sum of Squares (SS) is 1.04E+11, and the Mean Squares (MS) are 5.22E+10. This suggests that there is a significant difference in the group means. With a p-value of 5E-24 and an F-statistic of 128.8822, the results are much less than the accepted significance level of 0.05. This implies that the means of the groups differ by a relatively significant amount. With 69 df, the Within Groups SS is 2.8E+10, resulting in a Total SS of 1.32E+11.

Anova: two factor

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 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
Mileage	24	2011267	83802.79	1.21E+09
Cost	24	66150	2756.25	705502.7
Price	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				

Two factors are examined in the Two Factor ANOVA table: Rows and Columns, each with levels of Price, Cost, and Mileage. For every combination of rows and columns, the summary section includes information on the count, sum, average, and variance. For example, Row 1 (Mileage) has three observations, 70,512 as the total amount, 23,504 as the average, and 1.2E+09 as the variance. Three observations are included in Row 2 (Cost), which has a total sum of 99,635, an average of 33,211.67, and a variance of 2.88E+09.

The analysis assesses the causes of variance in the data in the ANOVA portion. With 23 degrees of freedom (df) and a Sum of Squares for Rows (SS) of 8.95E+09, the Mean Squares (MS) are 3.89E+08. With 2 df and a Sum of Squares for Columns (SS) of 1.04E+11, the MS is 5.22E+10. With p-values better than 0.05 for both Rows and Columns, it can be concluded that the observed variances are not significantly impacted by either row or column. The total sum of squares (SS) is 1.32E+11, while the error sum of squares (SS) is 1.9E+10 with 46 df.

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

Important insights into the variables Mileage, Cost, and Price are provided by the descriptive statistics. The average mileage is 83,802.79, while the standard error is 7,112.652. The data is moderately distributed around the mean, as indicated by the median of 81,142 and the standard deviation of 34,844.74. There is a minimum of 34,853 and a maximum of 140,811 within the range of 105,958. The distribution appears to be right-skewed, as indicated by the somewhat positive skewness of 0.386522 and the rather flat kurtosis of -1.09718.

In terms of cost, the standard error is 171.4525 and the mean is 2,756.25. The standard deviation is 839.9421, while the median is 2,750. With a minimum of 1,500 and a maximum of 4,500, the range is 3,000. The distribution is right-skewed, as indicated by the somewhat positive skewness of 0.473392 and the comparatively flat distribution of -0.81266, respectively.

The mean for Price is 3.254.5, while the standard error is 186.7512. 3,083 is the median, while 914.8902 is the standard deviation. There is a minimum of 2,000 and a maximum of 4,959 within the range of 2,959. The distribution appears to be right-skewed, as indicated by the slightly positive skewness of 0.272019 and the rather flat kurtosis of -1.20291. All things considered, these statistics offer a thorough examination of each variable's central tendency, variability, and distribution shape.

Correlation

	Cost	*Price*
Cost	1	
Price	-0.41106	1

The correlation coefficient between Cost and Price is -0.41106, indicating a moderate negative linear relationship between these two variables. This means that as the cost increases, the price tends to decrease, and vice versa. However, the strength of this correlation is moderate, so the relationship is not very strong.

Order Data Report

Introduction

This paper provides an in-depth analysis of an extensive dataset that records sales transactions in the automotive sector. It contains a number of attributes, including Sales Figures, Product Information, Customer Details, Order ID, Order Date, and Ship Date. The principal aim is to derive practical insights that can facilitate decision-making procedures and stimulate corporate expansion in the automobile industry.

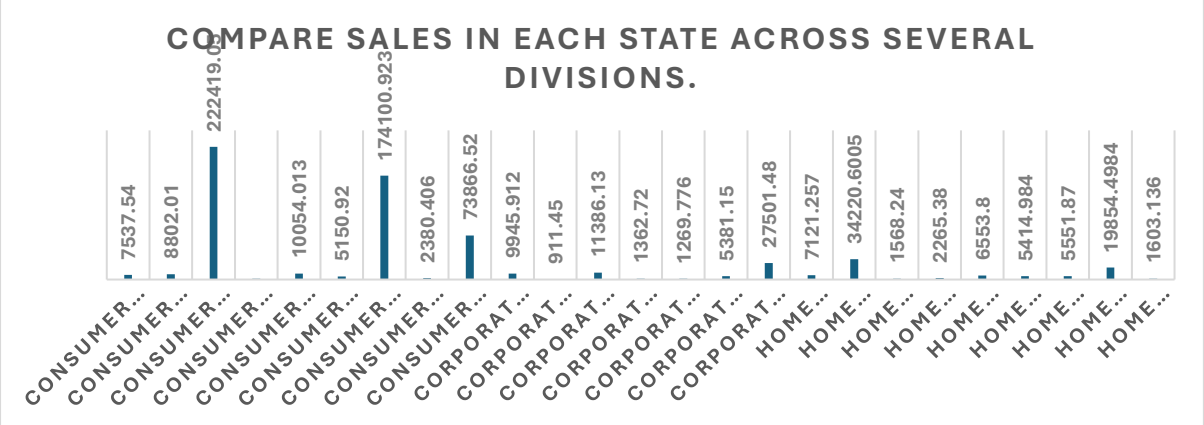
This research seeks to uncover critical trends, top-performing segments, and areas of prospective growth by examining sales data from various US states, sectors, categories, and subcategories. Automotive sector stakeholders, such as sales managers, marketers, and executives, who are eager to optimise sales strategies, improve customer happiness, and maximise income, will find great value in the insights obtained from this investigation.

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.

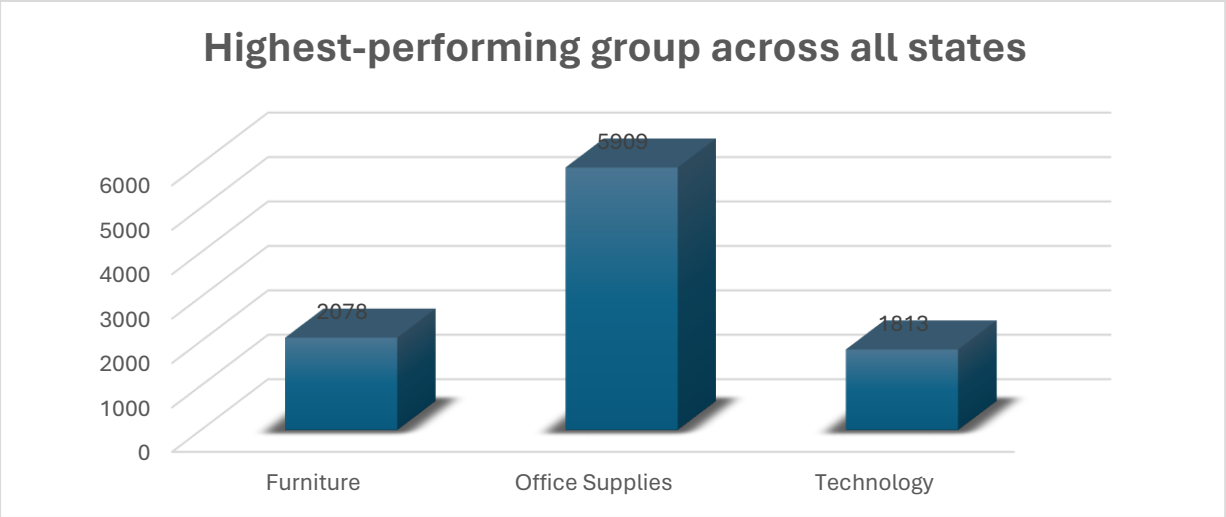
Analytics

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



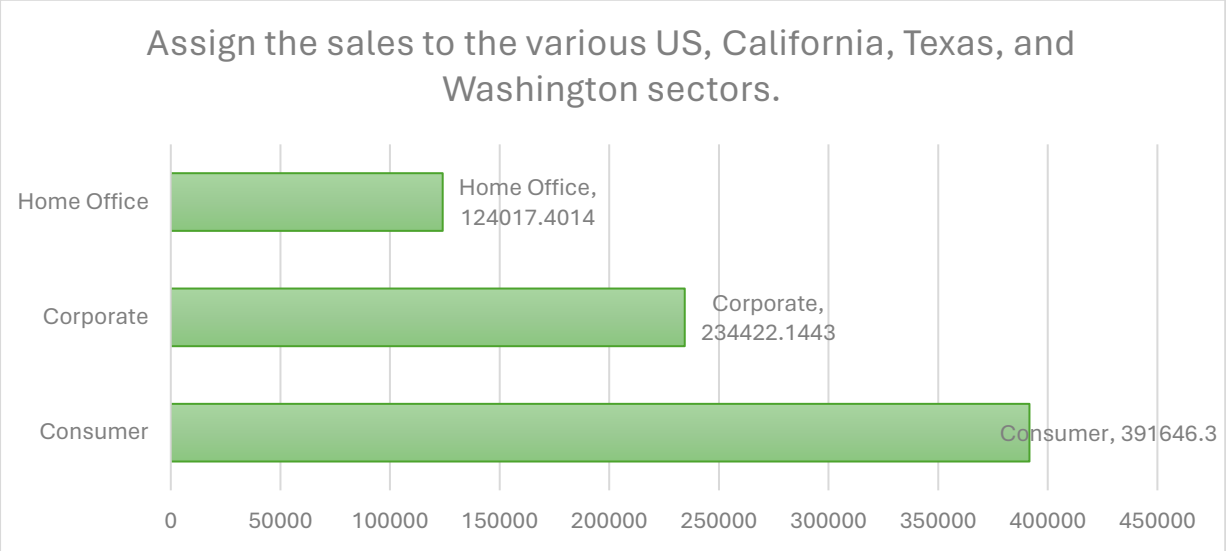
Ans - Following a thorough assessment of all states with respect to category and sales, California was found to have the most number of sales, coming in at $222,419.05. Furthermore, the Consumer category fared particularly well in every state, with a total sales amount of $1,148,060.531.

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



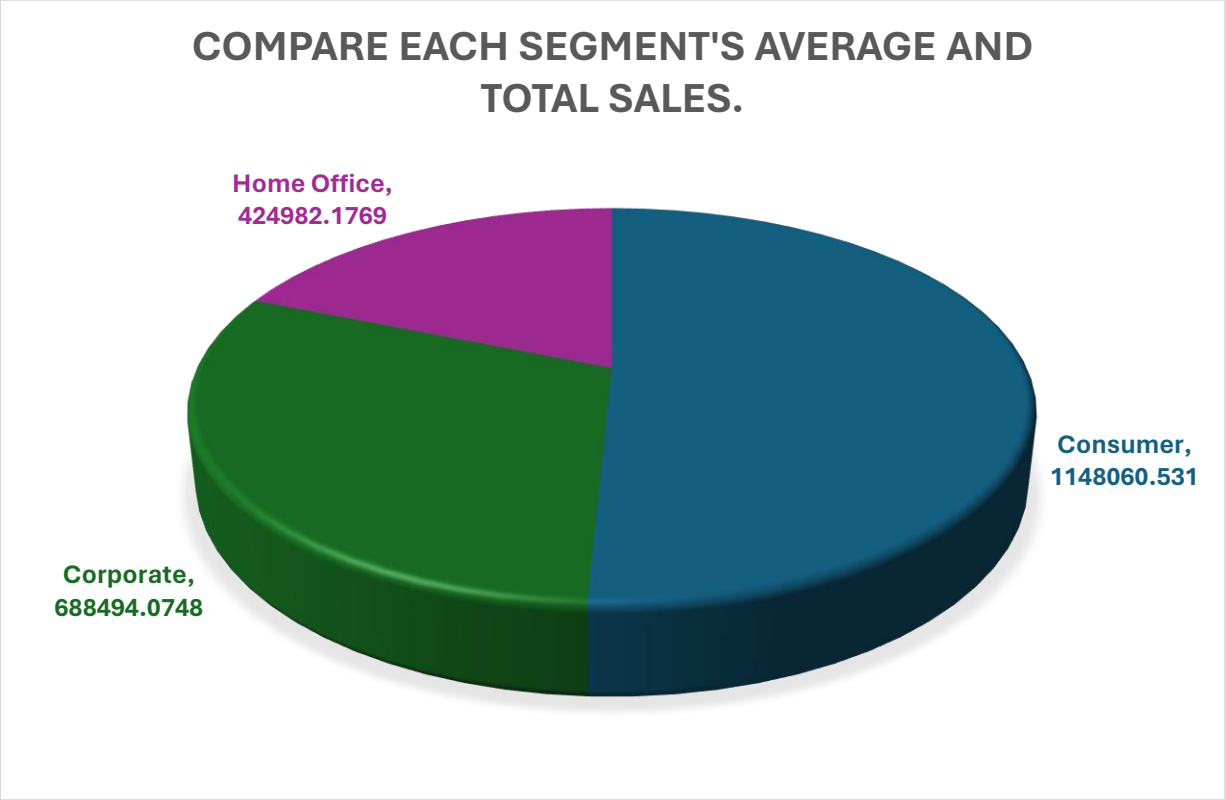
Ans - Across all states, Office Supplies is the category with the highest total sales count, coming in at 5,909. Furniture is in second place with 2,078 sales, closely followed by Technology in third place with 1,813 sales.

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



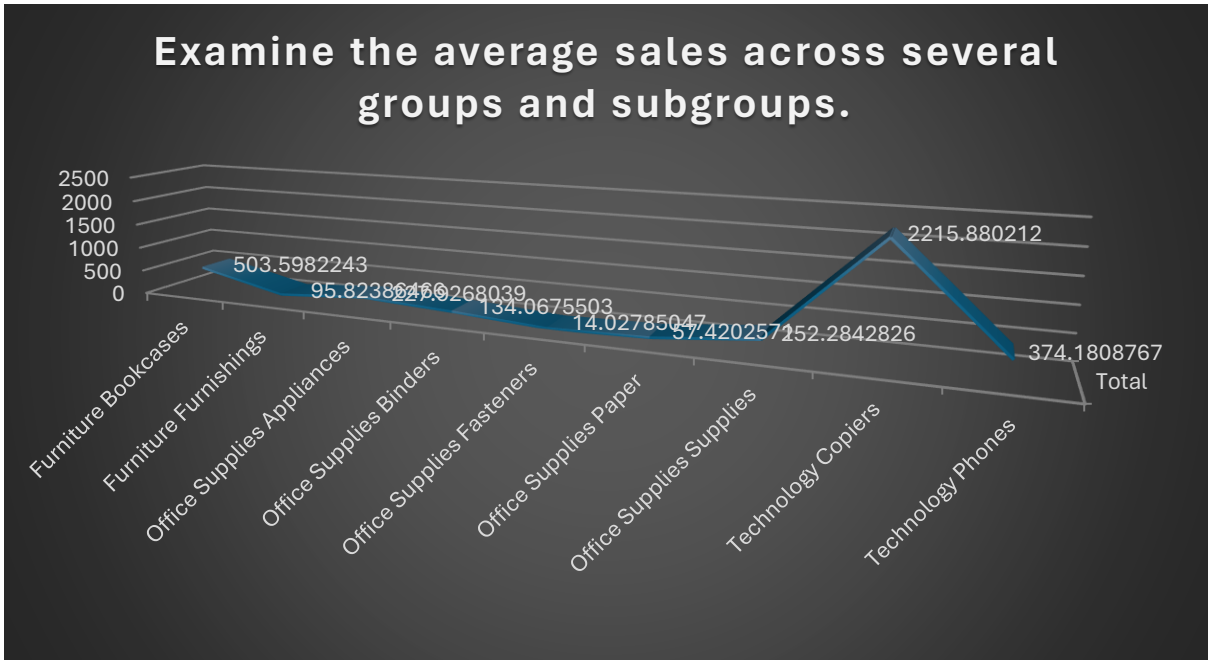
Ans - Utilising a bar graph to display the percentage of distribution and filtering the states for the total sales count. The US, California, Texas, and Washington have the highest sales in the consumer segment.

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



Ans - Based on the statistics, it is evident that the Consumer section has substantially greater average sales ($1,148,060.531) than the Home Office category, which records total sales of $243.40.

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



Ans - The study presents the average sales numbers for three categories—office supplies, technology, and furniture—each of which has a number of subcategories.

Conclusion and Review

Important insights are uncovered by the car industry's sales data analysis. California is the top state in terms of sales volume, and performance in the Consumer segment is strong in all states. Additionally, the category with the best performance is Office Supplies, which is followed by Furniture and Technology, indicating the preferences of the consumer.

In the US, the consumer segment consistently attracts the highest sales, particularly in California, Texas, and Washington. The report also highlights the Consumer category's greater average sales compared to the Home Office group.

All things considered, these observations provide insightful advice for maximising sales tactics, raising client satisfaction, and promoting company achievement in the automotive sector.

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

There doesn't seem to be much of a correlation—if any—between Order ID and Sales in this regression study of the Order dataset. The extremely low multiple R and R-squared values (1.88E-07 and 0.000434, respectively) provide as proof for this. With a p-value of 0.965747, the Order ID coefficient is not statistically significant, suggesting that it is not a predictor of Sales. This lack of significance is further supported by the ANOVA test, which yielded an F-statistic p-value of 0.965747.

Descriptive Statistics

Sales	
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

The mean sales amount in the Sales dataset is 230.1162, and the standard error is 6.320053. 54.384 is the median sales value, and 12.96 is the mode. With a standard deviation of 625.3021, sales figures are significantly variable. With a high kurtosis of 307.3056 and a substantially positive skewness of 13.05363, the data indicates a heavy-tailed distribution. The sales values are distributed between 0.444 and 22638.48, totaling 2252607 over 9789 observations.

Cookie Data Report

Introduction

We have comprehensive data on six varieties of cookies in our cookie dataset: sugar, chocolate chip, fortune cookie, oatmeal raisin, Snickerdoodle and white chocolate macadamia nut. This information includes sales volumes, expenses, earnings, and profits for these cookies over a range of time periods and nations. This paper explores consumer preferences, pricing dynamics, and regional popularity trends in addition to cookies. Businesses can learn a great deal about market potential and preferences in the cookie industry by investigating these insights. Prepare to learn fascinating new information that could have a big impact on companies just like yours.

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?

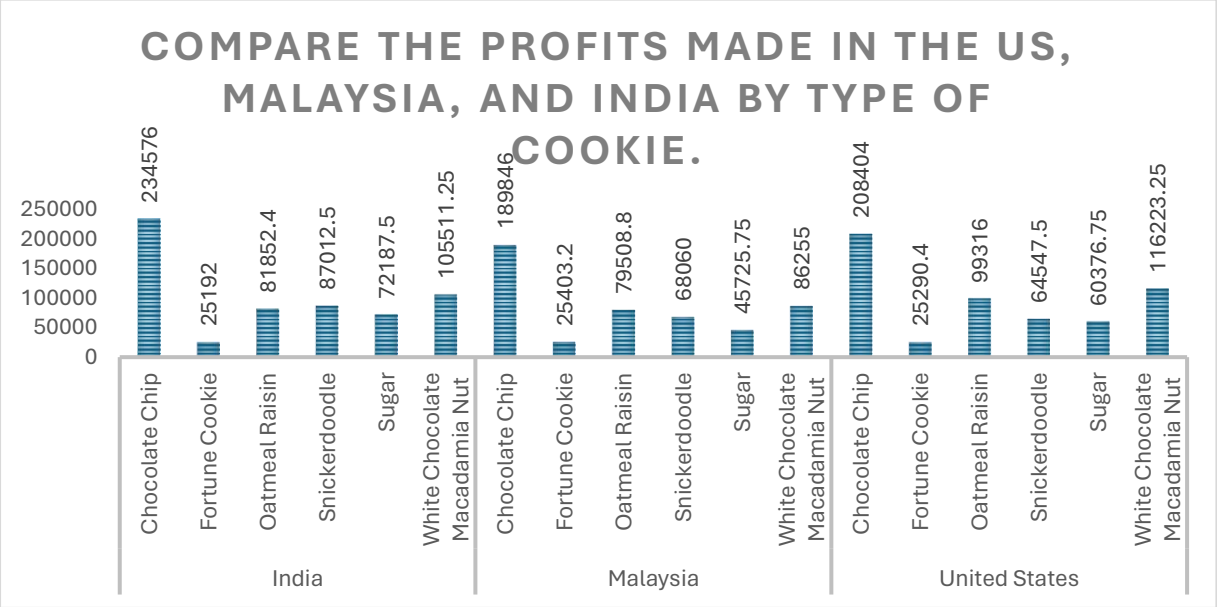
3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

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?

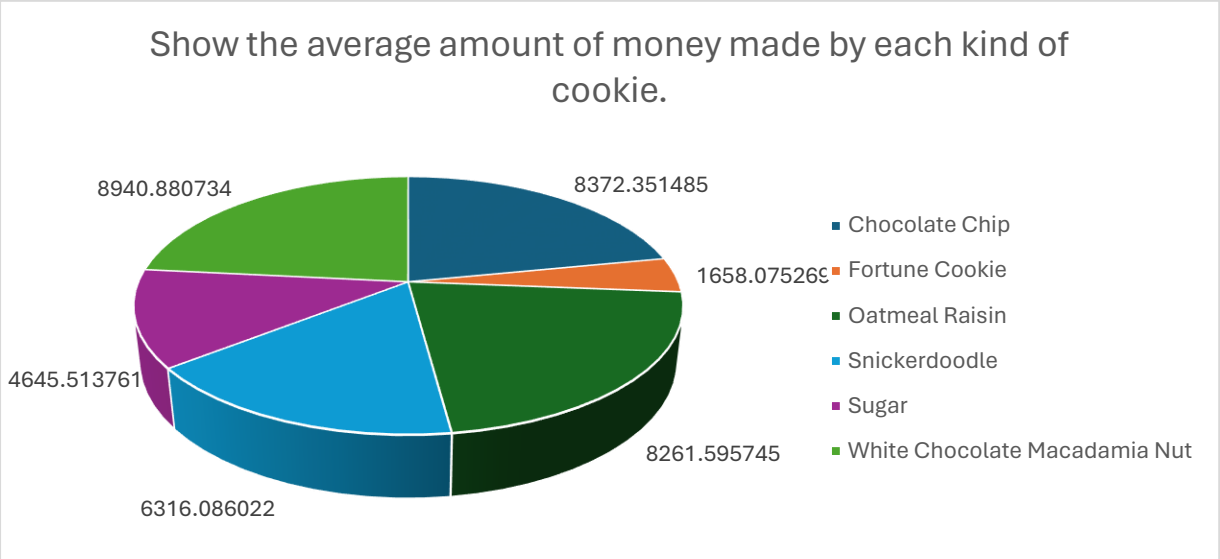
Analytics

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



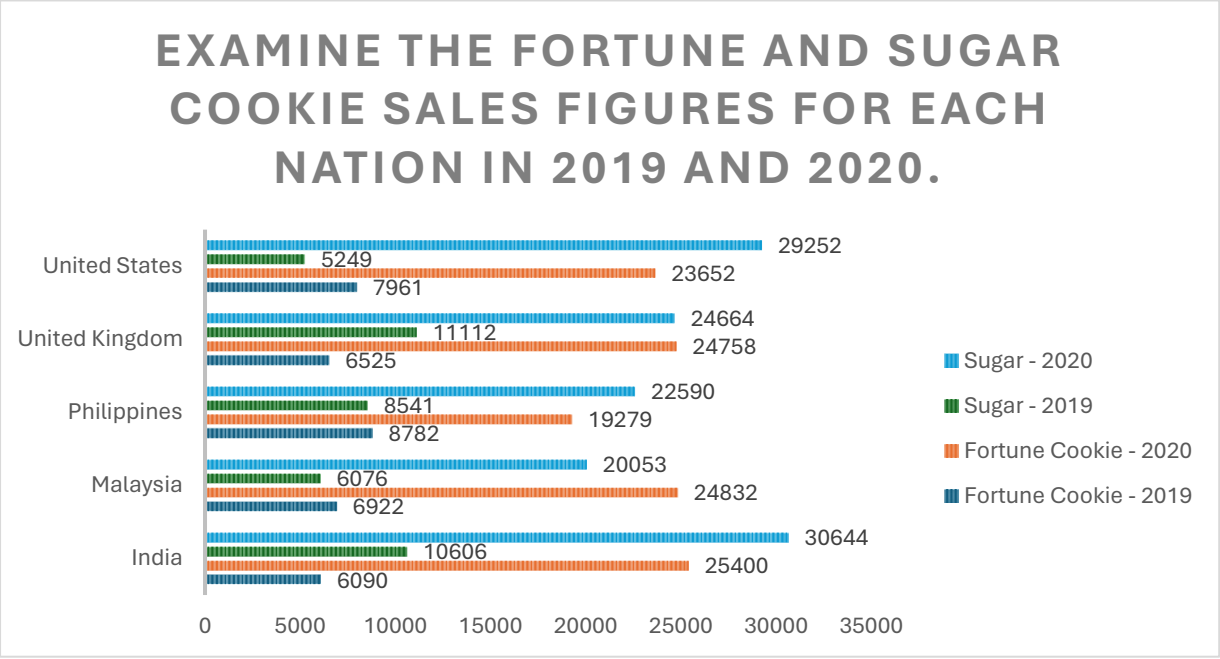
Ans - This study looks at the earnings from each variety of cookie in three distinct nations: Malaysia, India, and the United States. India is the country that makes the most money from chocolate chip cookies, followed by Malaysia and the US.

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



Ans - The purpose of this analysis is to show the average revenue that each type of cookie generates. With an average sales of $8,940.88, White Chocolate Macadamia Nut is clearly the most profitable variety, followed by Chocolate Chip.

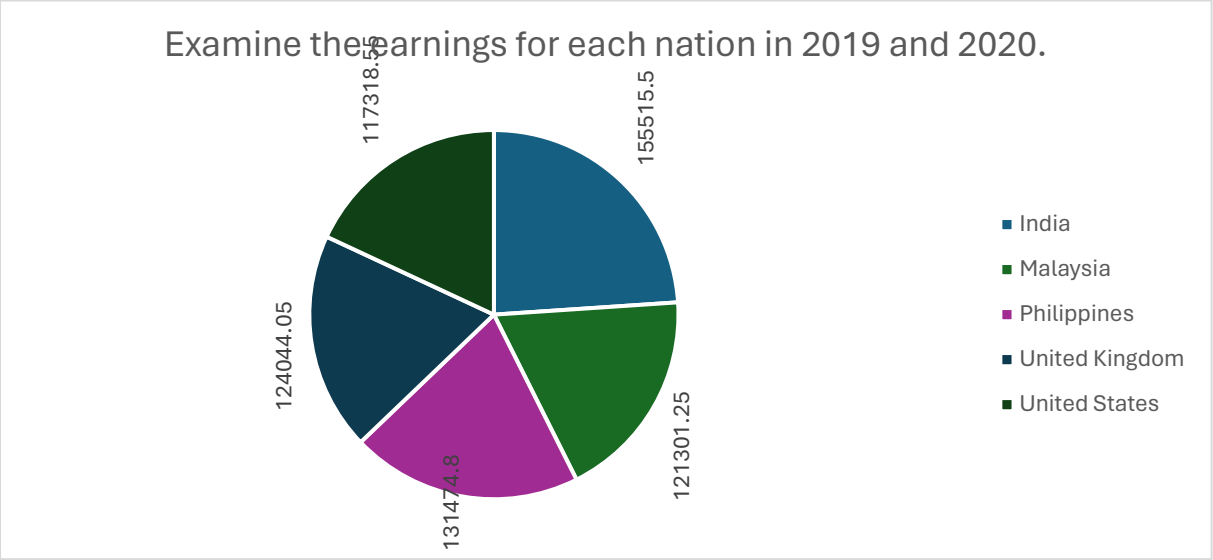
Ques. 3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?



Ans - The purpose of this analysis is to compare Fortune and Sugar cookie sales in various nations in 2019 and 2020. With 30,644 units sold in 2020, India is a major market for sugar biscuits. In contrast, India came in second place in terms of sugar cookie sales in 2019 behind the United Kingdom.

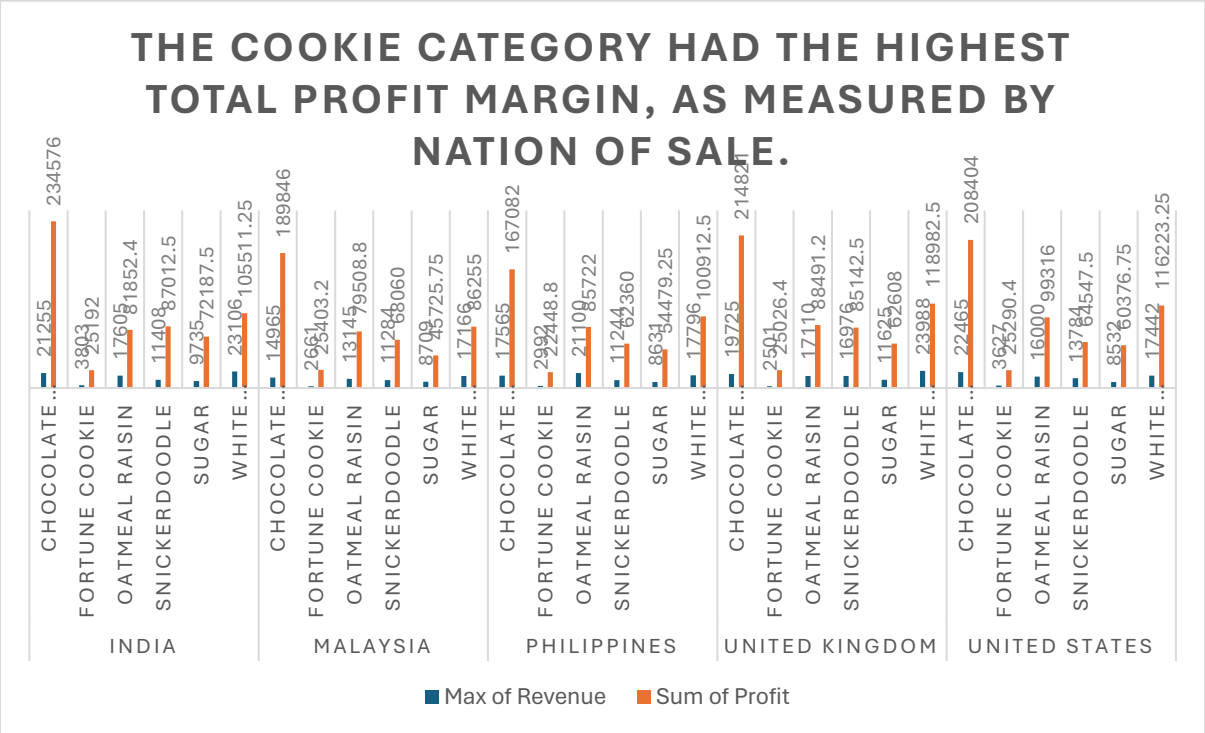
With 25,400 units sold, India leads the world in Fortune cookie sales, followed by Malaysia. However, with 8,782 units sold, the Philippines tops the list in Fortune cookie sales, followed by the US.

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



Ans - The purpose of this analysis is to analyse the earnings that various nations made in the 2019 and 2020 fiscal years. The graph shows that, with $471,027.55 in sales, the United Kingdom made the largest profit in 2020. The United States came in second with $456,839.35. On the other hand, with sales of $155,515.5, India made the largest profit in 2019. The Philippines came in second with $131,474.8.

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



Ans - The category of cookies that are sold for the most money in each nation is determined by this investigation. Cookies with chocolate chips make the most money, and cookies with sugar make the most profit—especially in India and the UK.

Conclusion and Review

The study shed light on the profits made by several cookie varieties in the US, Malaysia, and India. The country that made the most money from chocolate chip cookies was India, followed by Malaysia and the US. The cookies with the highest average revenue were white chocolate macadamia nut cookies, closely followed by chocolate chip cookies.

In terms of sales, the United Kingdom led the world in sugar cookie sales in 2019, with India showing notable sales in 2020. Sales of fortune cookies were increasing in both years in Malaysia and India, with significant sales also coming from the US and the Philippines.

In terms of comparing profits by nation for 2019 and 2020, the United States and the United Kingdom both had the highest profits in 2020. India and the Philippines had the biggest profits in 2019.

In terms of income, chocolate chip cookies brought in the most money, but altogether, sugar cookies made the most profit.

The report helped stakeholders understand market dynamics and make wise decisions by providing insightful information on the cookie sector. Visuals that were acceptable and easy to understand were used to successfully communicate the findings. It's crucial to recognise the need for more research into other variables affecting sales and profitability, though. For trustworthy insights, data completeness and correctness must be guaranteed.

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

The findings of the Cookie dataset's regression analysis show that the independent and dependent variables have a perfect linear connection. A perfect correlation is suggested by the multiple R value of 1. The independent factors account for all of the variability in the dependent variable, as shown by the fact that both the R-squared and adjusted R-squared values are 1. The remarkably minimal standard error (9.16E-12) suggests accurate estimations.

With an F-statistic of 1.9E+31, the ANOVA results verify that the regression model is highly significant ($p < 0.05$). The coefficients for the independent variables (X Variable 1, X Variable 2, and X Variable 3) are all quite near to 0, indicating no discernible effect on the dependent variable, notwithstanding the overall relevance of the model. The absence of statistical significance is further supported by the fact that all of the p-values for these coefficients are higher than 0.05.

In conclusion, the independent factors do not significantly affect the dependent variable, as shown by their coefficients and corresponding p-values, even if the regression model itself is highly significant because of the perfect fit.

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				

The variation between the Cost and Profit groups is compared using a single-factor ANOVA technique. There are 700 observations in the Cost group, totaling 1,926,955 with an average of 2,752.79. There are 700 observations in the Profit group as well, totaling 2,763,364 with an average of 3,947.66.

The means of the Cost and Profit groups differ significantly, according to the ANOVA results ($F = 90.92153$, $p < 0.05$). This shows that the average values of Cost and Profit vary in a way that is statistically significant. There is strong evidence to refute the null hypothesis because the p-value (6.36E-21) is much smaller than the significance level ($\alpha = 0.05$). Consequently, we determine that there is a significant difference in the mean values of Cost and Benefit and reject the null hypothesis.

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				

The effects of two categorical independent variables, Revenue and Cost, on the dependent variable, Profit, are evaluated using a two-factor ANOVA without replication. The table shows the count, sum, average, and variance for each factor level and summarises the data for revenue, cost, and profit.

The ANOVA findings show a substantial interaction effect between Revenue and Cost ($MS = 28507277$, $p < 0.05$), as well as significant main effects for both Revenue ($F = 14.75112$, $p < 0.05$) and Cost ($F = 1484.458$, $p < 0.05$). There is significant evidence against the null hypothesis, as indicated by the p-values for all components being less than the significance level ($\alpha = 0.05$). Consequently, we find that there is a large interaction effect between Revenue and Cost and that both Revenue and Cost have a significant impact on Profit, rejecting the null hypothesis.

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

The distribution and properties of the variables Unit Sold, Revenue, Cost, and Profit are all well-explained by the descriptive statistics.

The average value of units sold is 1608.32, with a standard error of 32.79 units. While the mode, which is 727 units, represents the value that occurs most frequently, the median, which is 1542.5 units, offers a measure of central tendency. The distribution's dispersion, symmetry, and shape are revealed by the standard deviation, skewness, and kurtosis numbers, in that order.

Descriptive statistics include measurements of central tendency, variability, and distributional features for Revenue, Cost, and Profit as well. Insights into the distribution and variability of the variables are provided by these statistics, which help to clarify their underlying properties and guide further research.

Correlation

	Unit Sold	*Revenue*	*Cost*	*Profit*
Unit Sold	1			
Revenue	0.796298	1		
Cost	0.742604	0.992011	1	
Profit	0.829304	0.995163	0.974818	1

The links between Unit Sold, Revenue, Cost, and Profit are shown in depth by the correlation matrix. Strong positive relationships are indicated by a correlation coefficient around 1, and strong negative relationships are shown by a value near -1. The correlation coefficient between Unit Sold and Revenue is roughly 0.796, suggesting a reasonably high positive correlation. Comparably, the correlation coefficient between profit and units sold is roughly 0.829, indicating a somewhat significant positive association. With a significant positive correlation of approximately 0.992, revenue and cost have a high degree of correlation. Furthermore, there is a very significant positive association between revenue and profit, as seen by their correlation coefficient of roughly 0.995. A robust positive association is also seen between cost and profit, with a coefficient of roughly 0.975. These correlation values provide valuable insights into the extent and direction of the relationships between the variables, helping to understand their associations and potential impacts on each other.

Loan Data Report

Introduction

Comprehensive data on loan applicants, such as gender, marital status, income, education level, loan amount, and property location, is included in the loan dataset. This dataset offers insightful information about the dynamics of loan requests.

Our goal in this analysis is to find patterns in the data and look at the characteristics of loan applicants. We seek to provide detailed answers to inquiries about the demographics, educational backgrounds, and loan amounts of loan applicants through the use of pivot tables and visualisations.

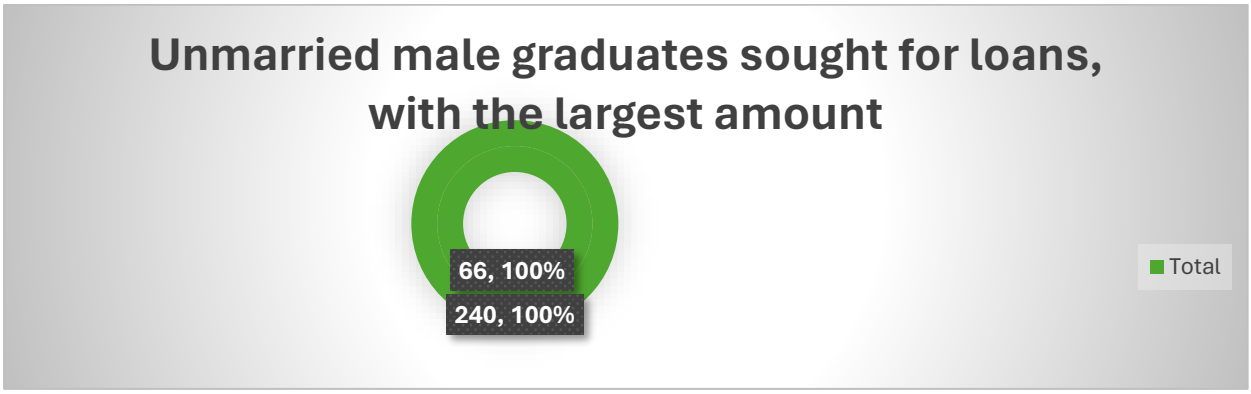
Financial institutions must grasp the nuances of loan requests in order to make educated decisions, expedite the lending process, and tailor services to meet the various needs of customers. Our goal in doing this study is to find useful information that will help guide strategic choices and raise the efficiency of loan management programmes.

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.

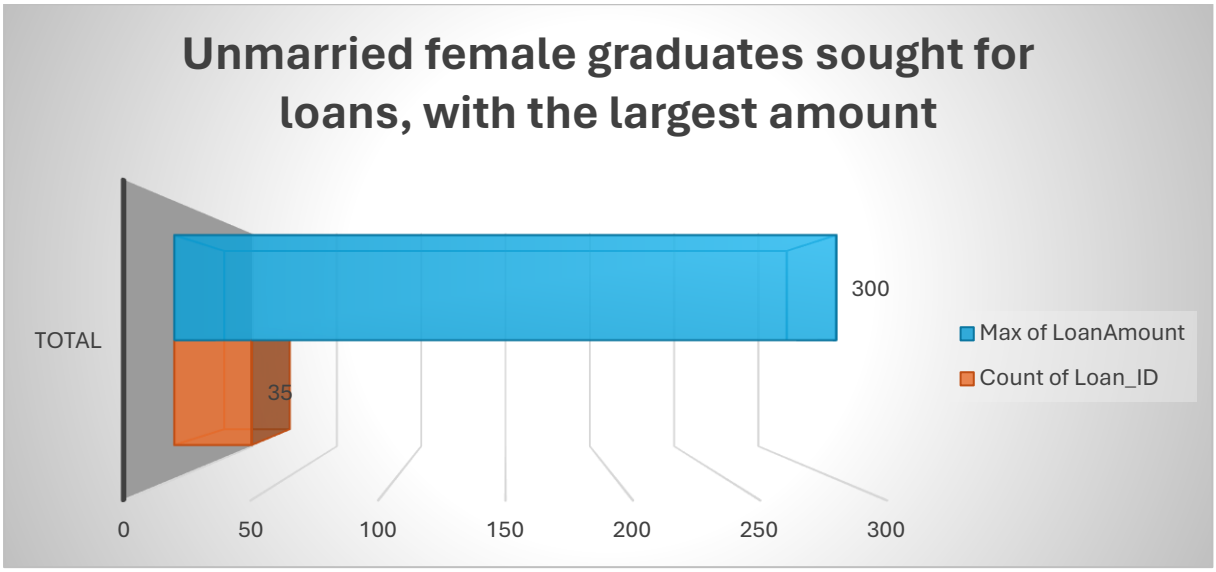
Analytics

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



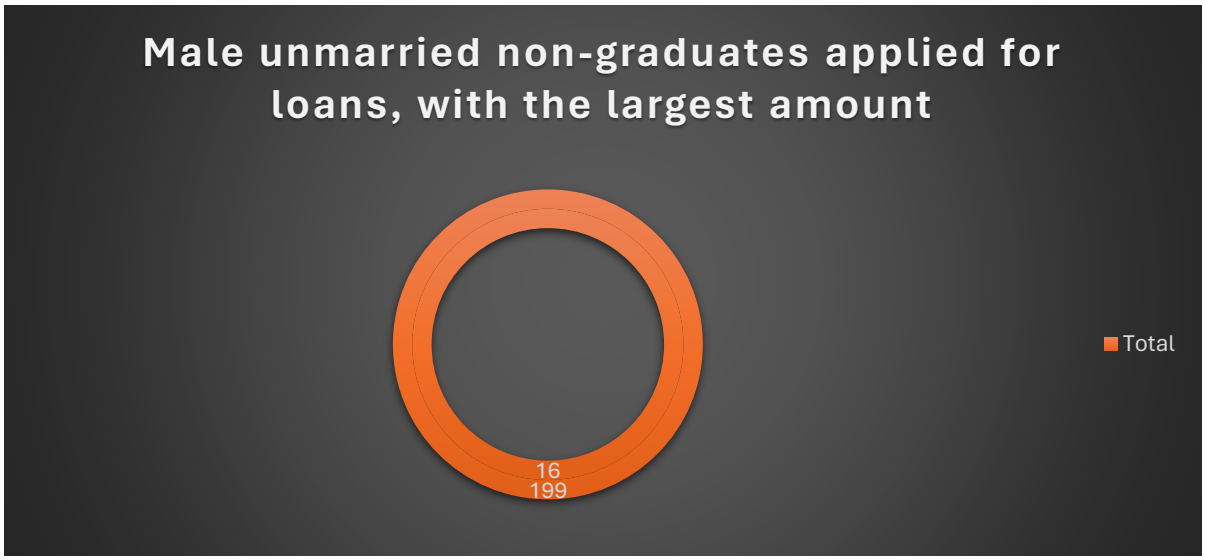
Ans - According to the study, the majority of loan applicants are single male graduates who have asked for the largest loan amount. In particular, the maximum loan amount requested is 240 out of the 66 loan applications that were examined in total.

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



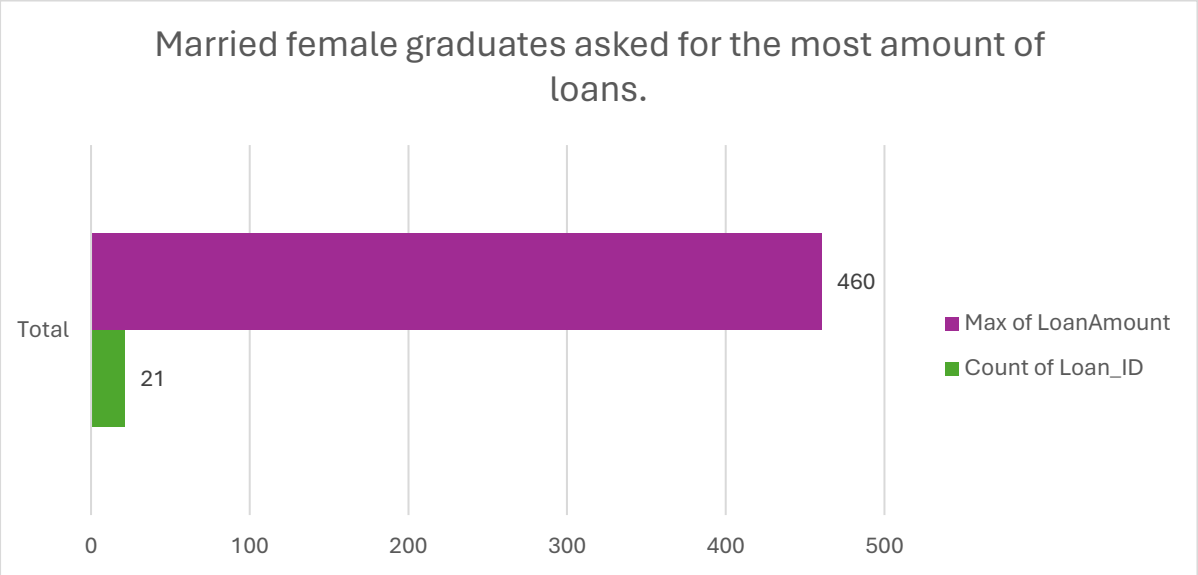
Ans - Based on the analysis, graduates without children make up the largest group of female loan applicants, and they have also requested the largest loan amounts. In particular, the maximum loan amount requested is 300 out of the 35 loan applications that were reviewed in total.

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



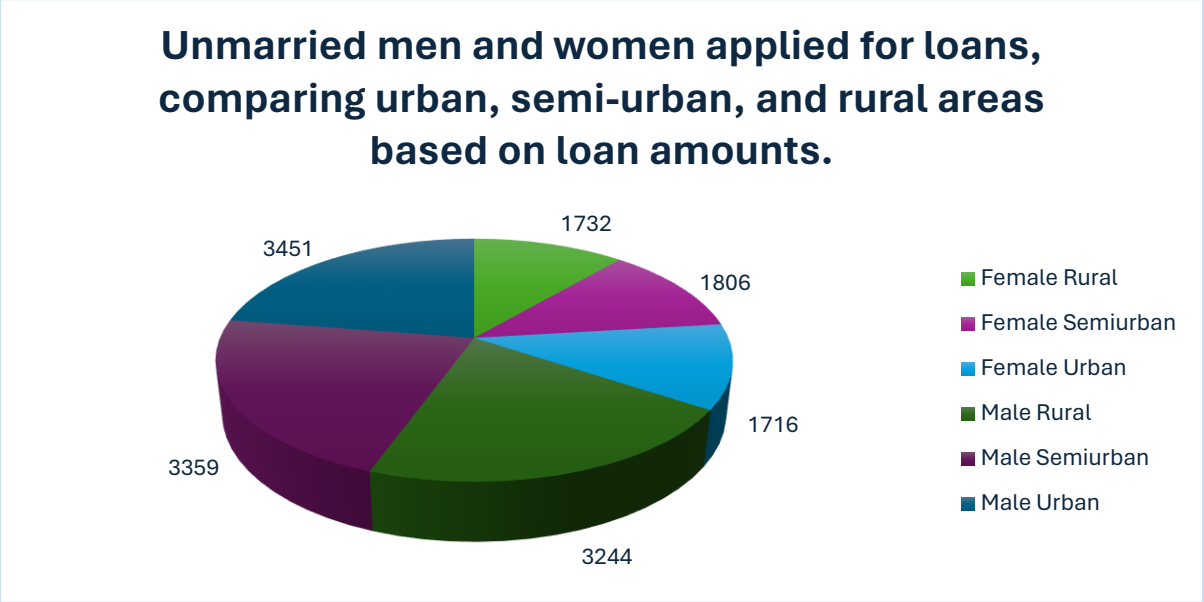
Ans - According to the data, men who are single and do not have a degree make up the largest group of loan applicants, and they have also requested the largest loan amount. In particular, the maximum loan amount requested is 199 out of the 16 loan applications that were assessed in total.

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



Ans. According to the data, unmarried female graduates make up the largest percentage of loan applicants, and they have also requested the largest loan amount. In particular, the maximum loan amount requested is $405, out of the 21 loan applications that were examined in total.

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



Ans - This report compares the number of single loan applicants for both genders, broken down by rural, semi-urban, and metropolitan locations. Although there are fewer female applications, there are far more male applicants. In particular, there are 1732 female loan applicants in rural areas, 1806 in semi-urban areas, and 1716 in urban areas. On the other hand, there are 3451 male loan applicants in urban areas, 3359 in semi-urban areas, and 3244 in rural areas.

Conclusion and Review

The data emphasises the clear differences between genders in loan applications. The majority of applications were single male graduates, closely followed by single female graduates. Although married female grads and single male non-graduates also applied for loans, their numbers were far lower. Significantly more men than women lived in rural, semi-urban, and urban areas.

This research provides important insights into borrower demographics by clearly identifying gender-based patterns in loan requests. It recommends additional research into the variables affecting loan decisions in addition to improving the data display visually. In the conclusion, the paper lays the foundation for understanding loan dynamics and offers opportunities for further research.

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

A number of important conclusions are drawn from the loan dataset's regression analysis. The predictors and the loan amount have a somewhat positive association, as indicated by the multiple R coefficient, which is roughly 0.531. The independent variables account for about 28.2% of the variation in the loan amount, according to the R-squared value of about 0.282.

The following are the predictors' coefficients: The coefficients for applicant and co-applicant income are roughly 0.096 and 0.0068, respectively. The influence of each predictor on the loan amount is shown by these coefficients.

The statistical significance of the regression model is confirmed by the substantial F-value of 37.32 ($p < 0.05$) displayed in the ANOVA table. This suggests that a substantial portion of the volatility in the loan amount may be explained by the model taken as a whole.

In conclusion, our research contributes to a better understanding of the loan approval process by illuminating the ways in which applicant and co-applicant incomes affect the loan amount.

Anova: one factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*		
Loan Amount	289	39533	136.7924	3564.04		
Loan Amount Term	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				

Based on the parameters Loan Amount and Loan Amount Term, the data is divided into two groups in the ANOVA table for the single-factor analysis of the loan dataset. Approximately 8392703 is the overall sum of squares (SS), with 6124794 between groups and 2267909 within groups. As a result, the mean square (MS) is 6124794 between groups and 3937.343 within groups.

The accompanying p-value of around 8.4E-166 and the F-value of 1555.565 show that there is a substantial difference between the means of the two groups. This indicates that the loan dataset is significantly impacted by the element under consideration (loan amount vs. loan amount term). Strong evidence opposing the null hypothesis is suggested by the high F-value and extremely low p-value, which lend credence to the idea that the difference in means is not the result of chance. Consequently, the loan dataset is greatly impacted by the component selected.

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
Loan Amount	289	39533	136.7924	3564.04
Loan Amount Term	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				

The loan dataset is examined using the two-factor ANOVA without replication, taking into account the loan amount and loan amount term. With 1264619 SS for Loan Amount Term (rows), 6124794 SS for Loan Amount (columns), and 1003290 SS for mistake, the total sum of squares (SS) is roughly 8392703.

The loan amount's mean square (MS) is 6124794, and the loan amount's mean square (MS) is 4391.038. Both the Loan Amount and Loan Amount Term F-values are 1.260472 and 1758.156, respectively, with corresponding p-values indicating significance ($p < 0.05$).

These results indicate that both Loan Amount and Loan Amount Term have a significant impact on the loan dataset, with both factors influencing the observed outcomes significantly.

Descriptive Statistics

Loan Amount Term		*Applicant Income*		*Co-Applicant Income*		*Loan Amount*	
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

Descriptive statistics were computed for four variables in the loan dataset: Loan Amount Term, Applicant Income, Co-Applicant Income, and Loan Amount. For Loan Amount Term, the mean is approximately 342.67 months, with a standard error of 3.86 months. The median Loan Amount Term is 360 months, with a mode of 360 months as well. The standard deviation is 65.66 months, indicating variability in loan term lengths. Applicant Income has a mean of approximately 4637.35, with a standard error of 281.80. The median and mode are 3833 and 5000, respectively. The standard deviation is high at 4790.68, suggesting significant variability in applicant incomes. Co-Applicant Income has a mean of about 1528.26, with a standard error of 139.86. The median is 879, with a mode of 0, indicating a right-skewed distribution. The standard deviation is 2377.60, highlighting variability in co-applicant incomes. Lastly, Loan Amount has a mean of 136.79, with a standard error of 3.51. The median is 126, with a mode of 150. The standard deviation is 59.70, suggesting variability in loan amounts. These statistics provide insights into the central tendency, variability, and distribution of the loan dataset variables.

Correlation

	Applicant Income	*Co-Applicant income*	*Loan Amount*
Column 1	1		
Column 2	-0.08435	1	
Column 3	0.445695	0.230355	1

The associations between Applicant Income, Co-Applicant Income, and Loan Amount are displayed in the correlation matrix for the variables in the loan dataset. The relationship between applicant and co-applicant income is weakly negative, at about -0.084. The correlation between the loan amount and applicant income is moderately positive, at about 0.446. Similarly, there is just a slight positive association (around 0.230) between co-applicant income and loan amount. Understanding the direction and degree of the links between the variables is essential for comprehending their possible effects on loan outcomes, and these correlation coefficients offer valuable insights into that relationship.

Shop Sales Data Report

Introduction

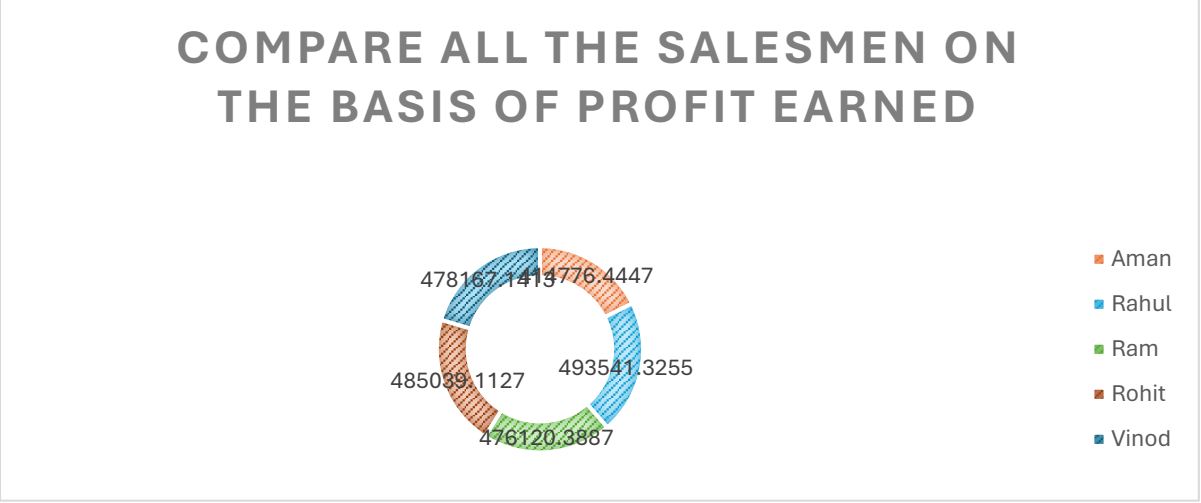
With an emphasis on the analysis of sales performance and product trends among sales representatives, this paper delves into a large sales dataset. The dataset contains a number of characteristics, including information about sales staff, product specs, sales volume, and profit margins. This analysis's main goal is to provide insightful information that can direct the creation of sales strategies and enhance overall company performance. The study attempts to identify top-performing sales reps, assess product attractiveness, and understand sales trends by closely examining sales data over a specified period of time and comparing product performance. Sales managers, marketing specialists, and executives seeking to improve sales tactics, maximise profits, and promote company growth will find great value in the insights gleaned from this investigation.

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.

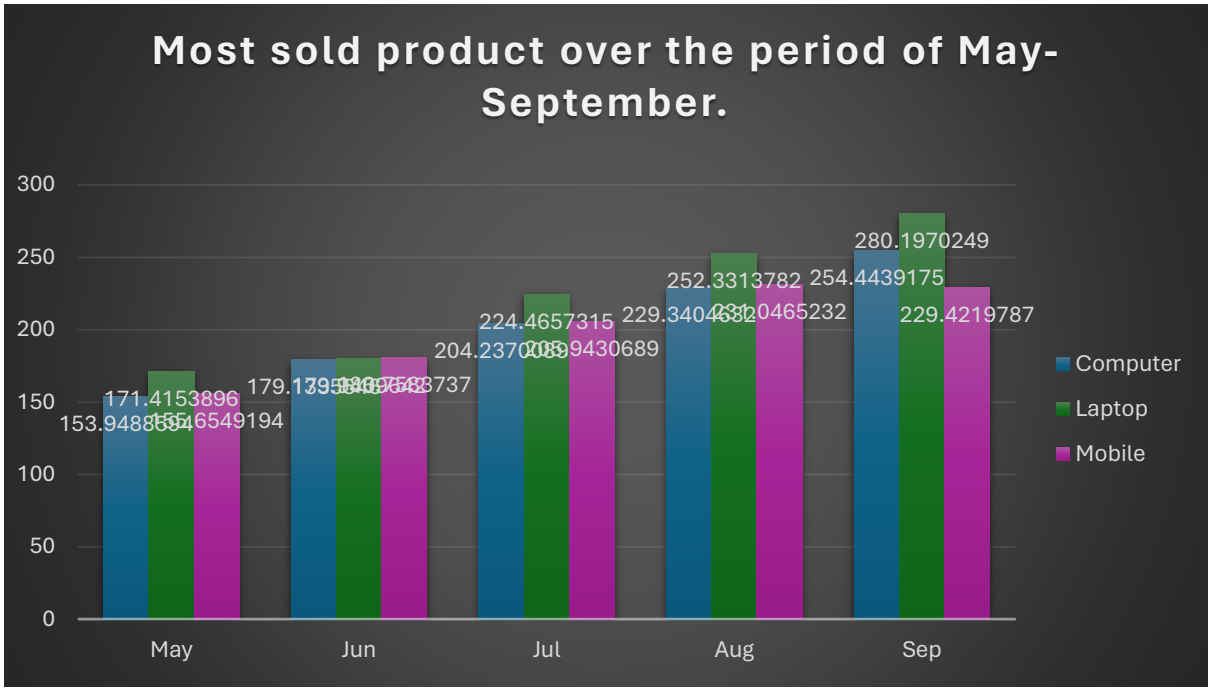
Analytics

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



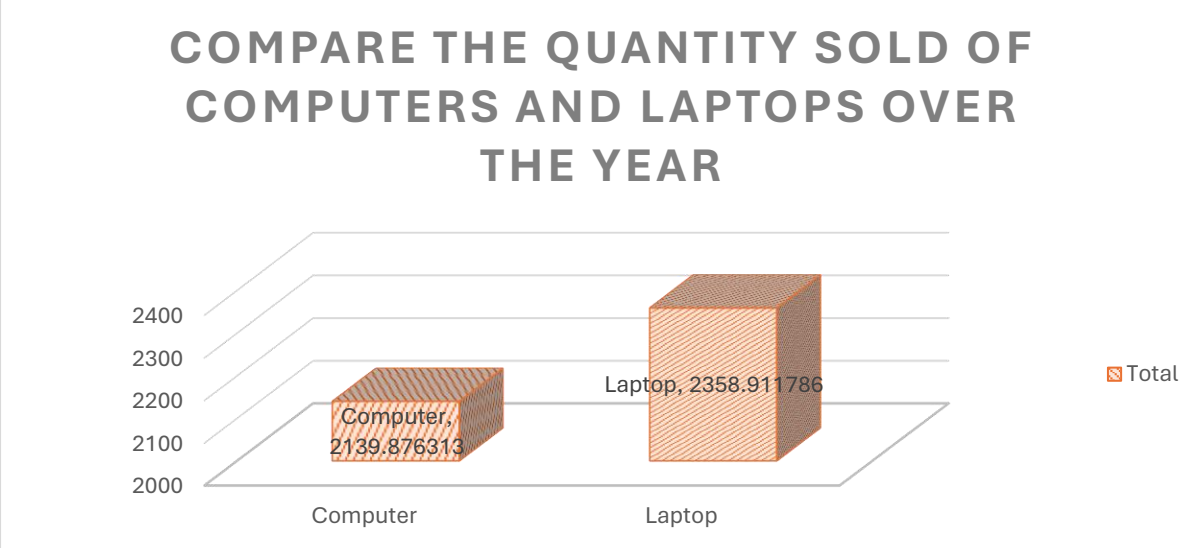
Ans - The comparison of all the salesmen on the basis of profit earned and the bar graph shows that the rahul has the highest profit earned with value 493541.3255, compared to all the salesmen.

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



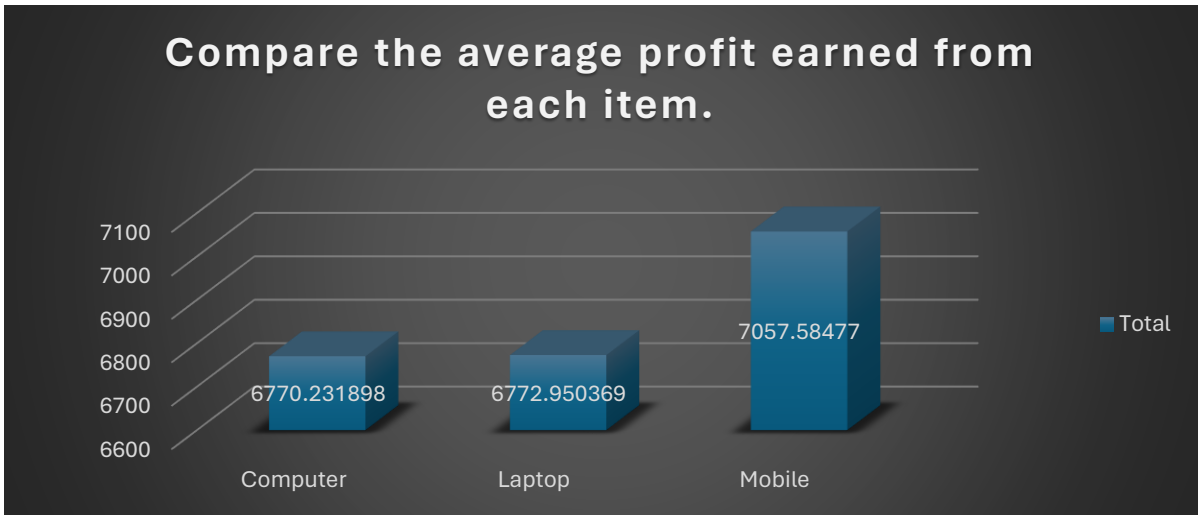
Ans - To determine the highest-selling product between May and September, we must analyze the sales data within this timeframe. By summing up the quantities sold for each product across all transactions during these months, we find that the top-selling product is the Laptop, with the highest sales occurring in September, totaling 280.1970249 units.

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



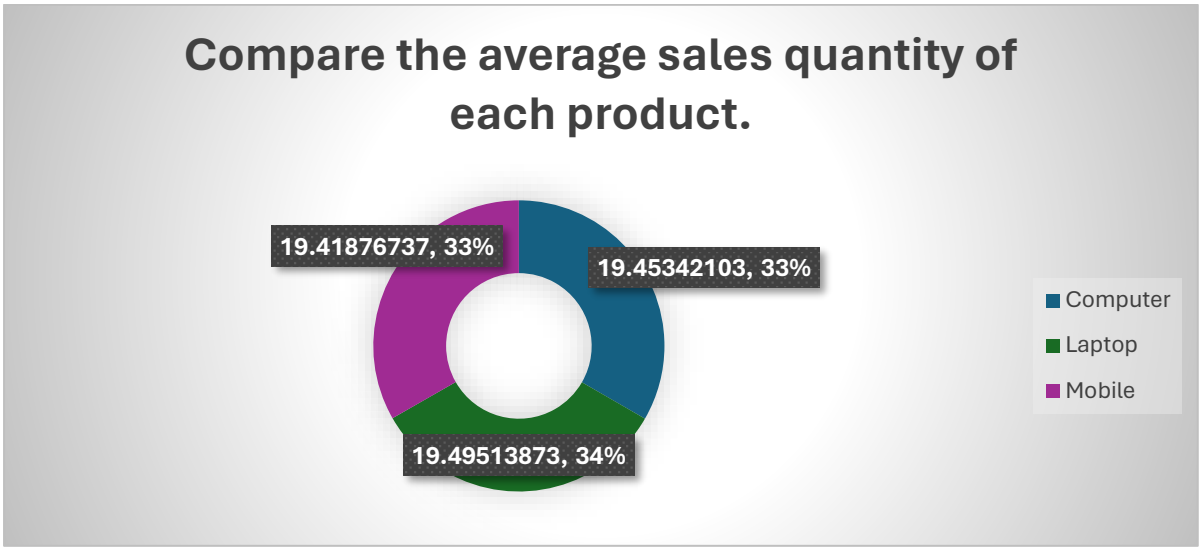
Ans - The two products that sold the most over the year were computers and laptops. Computers had a total sold quantity of 2139.876313 units, while laptops had a higher sold quantity of 2358.911786 units.

Ques. 4 . Which item yield most average profit?



Ans - According to this analysis, Mobiles have the highest average profit earned among Mobiles, Laptops, and Computers, with an average profit of 7057.58477.

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



Ans - The analysis indicates that the average sales quantity of Laptops (19.49513873) surpasses that of other products, such as Mobiles (19.41876737) and Computers (19.45342103).

Conclusion and Review:

The analysis uncovers significant insights into sales performance and product trends among salesmen. Rahul emerges as the top performer, achieving the highest profit compared to all other salesmen. Additionally, the most sold product between May and September is identified as laptops, with the highest sales recorded in September. Laptops also outshine computers in terms of units sold throughout the year. Moreover, mobile phones exhibit the highest average profit among mobiles, laptops, and computers. Notably, laptops demonstrate the highest average sales quantity compared to mobiles and computers.

The analysis effectively highlights sales performance and product trends, offering valuable insights for optimizing sales strategies. Visualizations aid in understanding trends over time

and product popularity. However, delving deeper into factors influencing sales fluctuations and product preferences could further enhance the analysis. Overall, the report provides actionable insights for improving sales strategies and maximizing revenue.

Regression

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%*	*Upper 95.0%*
Intercept	2068.993161	88.47952	23.38387	9.14E-73	1894.957	2243.029	1894.957	2243.029
X Variable 1	246.465683	4.196812	58.72686	4.6E-180	238.2106	254.7206	238.2106	254.7206

This regression analysis illustrates a strong relationship between the quantity of items sold (X Variable 1) and the corresponding sales amount (Y). With a high R-squared value of approximately 0.91, it indicates that about 91% of the variability in sales amount can be explained by changes in the quantity of items sold. For each additional unit increase in the quantity sold, there is an average increase of approximately $246.47 in sales amount.

Both the intercept and the coefficient of X Variable 1 are statistically significant, with t-stats of 23.38 and 58.73, respectively, and very low p-values (close to zero), confirming the reliability of these coefficients. Therefore, we conclude that the quantity of items sold serves as a robust predictor of sales amount in this dataset.

Anova (Single Factor)

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Qty	342	6654.271	19.45693	66.0952		
Amount	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				

The single-factor ANOVA conducted on the quantity (Qty) and sales amount (Amount) indicates a significant difference between the groups. The analysis reveals substantial variance between the groups (SS = $8.01E+09) compared to within the groups (SS = $1.5E+09), resulting in a high F-statistic of 3632.879 with a very low p-value (close to zero). This implies that the difference in means between quantity and sales amount is highly unlikely to have occurred by chance.

Therefore, we reject the null hypothesis and conclude that there is a significant difference in sales amounts attributed to different quantities sold. This finding underscores the importance of quantity as a determinant of sales amount in this analysis.

Anova two factor

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
Qty	342	6654.271	19.45693	66.0952
Amount	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				

In the two-factor ANOVA analysis without replication, we observe that both rows and columns contribute significantly to the variance. The sums of squares (SS) for rows and columns are $7.58E+08 and $8.01E+09, respectively.

The high F-statistics for both rows (1.014883) and columns (3659.913) with low p-values (close to zero) indicate that the differences observed in both factors are statistically significant. Therefore, we reject the null hypothesis and conclude that both the quantity sold (Qty) and sales amount (Amount) significantly affect the variance in the dataset. This suggests that both factors play a crucial role in determining the sales amount, underscoring their importance in the analysis.

Descriptive Statistics:

Qty		*Amount*	
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

For the quantity sold (Qty), the mean is approximately 19.46 with a standard error of 0.44. The data shows moderate positive skewness (skewness = -0.10) and slight negative kurtosis (-0.999), indicating a distribution that is slightly flatter compared to a normal distribution. The range of quantity sold spans from 3 to 33.31.

In contrast, for the sales amount (Amount), the mean is approximately 6864.46 with a larger standard error of 113.57. The data also exhibits moderate positive skewness (skewness = -0.36) and slight negative kurtosis (-0.508). The range of sales amounts is much larger, ranging from 1000 to 10279.85.

These descriptive statistics provide insights into the central tendency, variability, and shape of the distribution for both quantity sold and sales amount variables in the dataset.

Correlation

	Qty	*Amount*
Qty	1	
Amount	0.954077	1

The correlation coefficient between quantity sold (Qty) and sales amount (Amount) is approximately 0.954. This strong positive correlation suggests that there is a significant relationship between the quantity of items sold and the corresponding sales amount, indicating that as the quantity sold increases, the sales amount also tends to increase.

Sales Data Sample Report

Introduction

ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and SALES are just a few of the parameters that are examined in this report's comprehensive sales dataset in order to extract insights that will help guide sales tactics and improve business effectiveness. With the goal of improving sales operations and boosting revenue creation, it is directed towards executives, marketers, and sales managers. Important analyses include comparing sales data between vintage and classic cars, figuring out average sales, identifying best-selling products, estimating country-specific profits for specific product lines, analysing sales trends over time, and ranking countries according to the size of transactions. By performing these assessments, the report aims to provide practical insights that can drive sales growth and improve overall business results.

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.

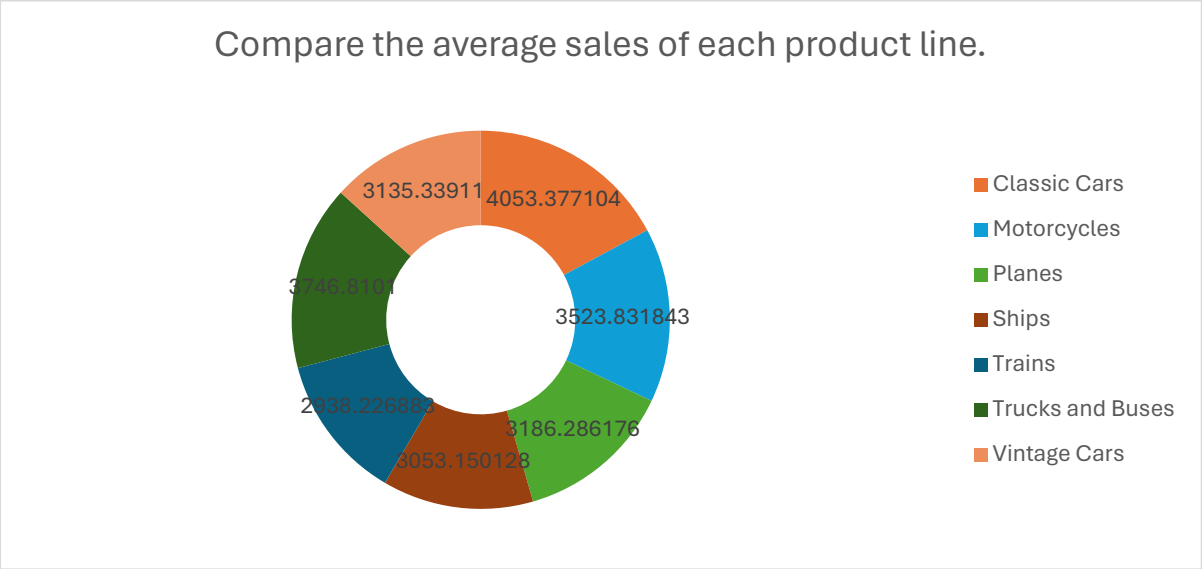
5. Comparative analysis of all countries based on deal size.

Analytics

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

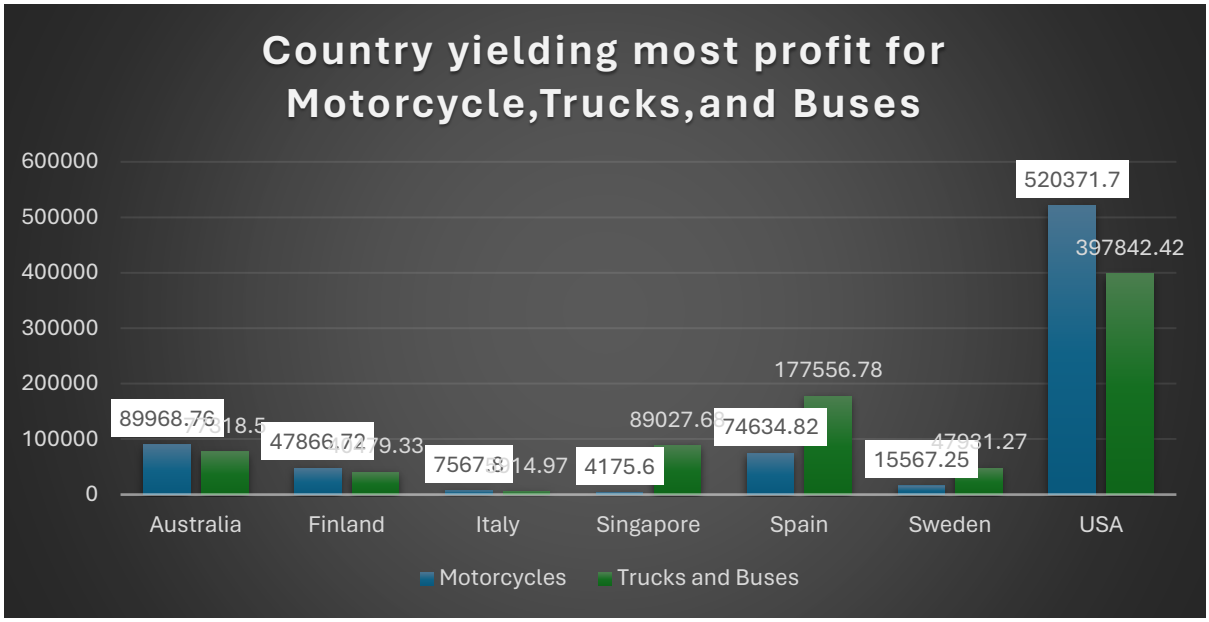
Ans - This analysis Compare the sale of Vintage cars and Classic cars for all the countries. Where USA(2102394.02) has the highest sales followed by Spain, France, and Australia. This is represented by using line graph.

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



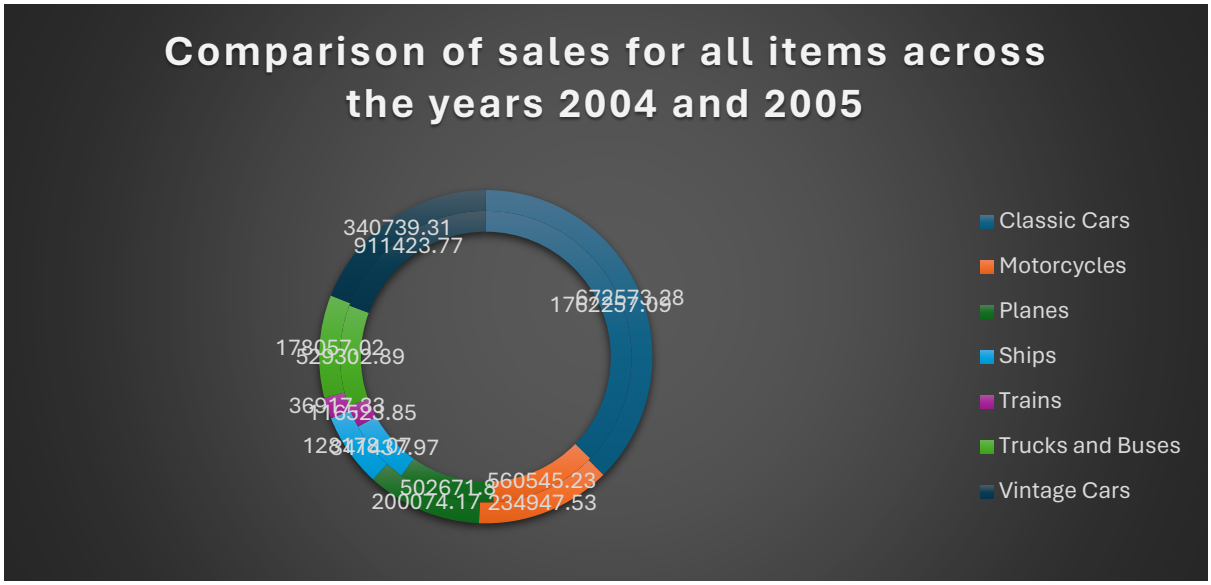
Ans - This analysis seeks to present the average sales figures for all products and pinpoint the highest-selling product. The graphical representation highlights that Classic Cars lead the sales, boasting an average of 4053.377104 units sold, followed by Trucks and Buses, and Motorcycles.

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



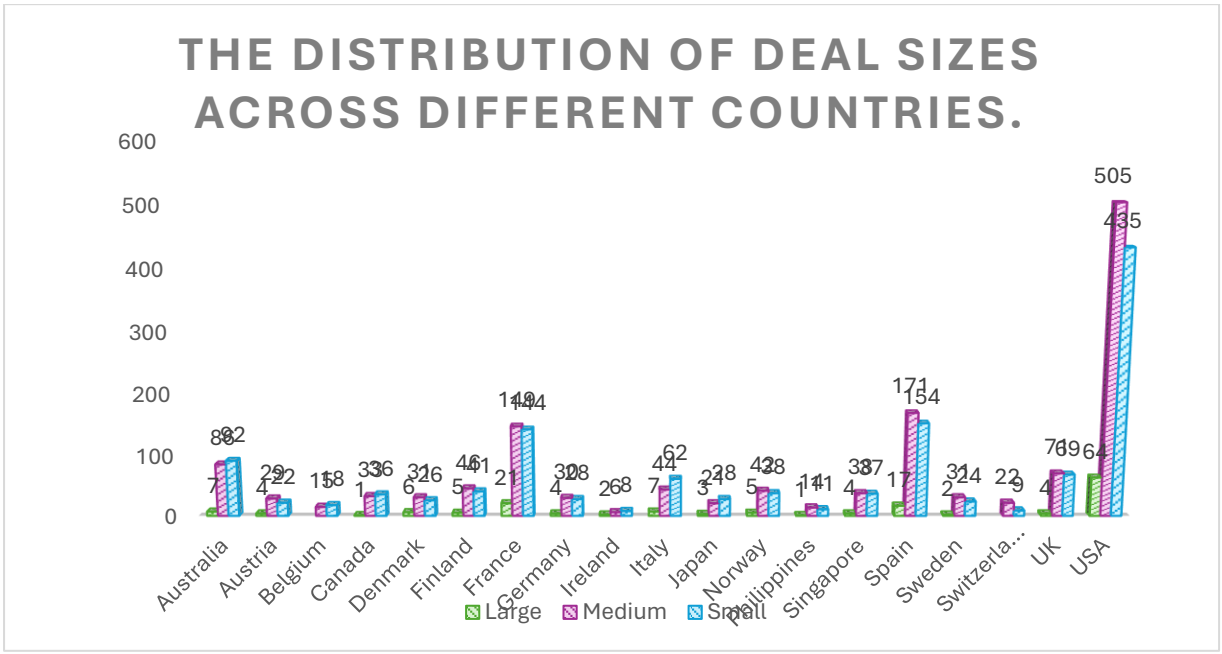
Ans - The objective of this analysis is to determine the country generating the highest profit for Motorcycles, Trucks, and Buses. The bar chart illustrates that the USA leads with the highest sales, totaling $397,842.42 for Trucks and Buses, and $520,371.70 for Motorcycles, followed by France and Spain in descending order.

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



Ans - This analysis aims to juxtapose the sales figures for all items across the years 2004 and 2005. The pie chart illustrates that the sales distribution for all items across the two years is shifting significantly. Notably, Classic cars emerge as the top-selling category in both years, with sales reaching $1,762,257.09 in 2004 and $672,573.28 in 2005.

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



Ans - This analysis seeks to uncover the distribution of deal sizes across different countries. The bar chart reveals that the deal sizes in the USA are notably higher compared to other countries, with a large deal size of 64, a medium deal size of 505, and a small deal size of 435.

Conclusion and Review

The analysis reveals crucial insights into sales dynamics and profitability across various categories and countries. Notably, the USA emerges as a pivotal market leader, displaying

robust sales performance in Vintage and Classic cars, Trucks, Buses, and Motorcycles. Classic Cars notably lead as the highest-selling product, making a substantial contribution to overall sales revenue. Moreover, the USA demonstrates exceptional profitability, particularly in the Trucks, Buses, and Motorcycles categories. Sales for Classic cars maintain a consistently strong trajectory throughout the years 2004 and 2005, indicating sustained demand for this category. Additionally, the USA showcases significantly larger deal sizes compared to other countries, highlighting its dominance in sales volume.

While the analysis effectively communicates key findings through visualizations, further exploration into factors influencing sales fluctuations and disparities in deal size could offer deeper insights. Overall, the report provides valuable insights for refining sales strategies and fostering 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

This regression analysis for the sales dataset reveals that the model is statistically significant, as indicated by a very low p-value (4.62E-78). The multiple R value of 0.877 suggests a strong positive linear relationship between the independent variables (MSRP, Quantity Ordered) and the dependent variable (Sales). The coefficient values indicate that for every unit increase in MSRP, there's an increase of approximately $103.08 in sales. Similarly, for every unit increase in Quantity Ordered, sales increase by about $12.82, and for every unit increase in the third independent variable, sales increase by approximately $47.43. The adjusted R-squared value of 0.766 indicates that the model explains about 76.6% of the variance in the sales data.

Anova: one factor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*		
Sales	250	903280.9	3613.123	3445221		
MSRP	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				

In this single-factor ANOVA analysis, we are assessing the impact of different levels of the factor (Sales and MSRP) on the variance in the data. The ANOVA results indicate a significant difference between the groups, with a very low p-value (3.1E-113). This provides strong evidence to reject the null hypothesis, suggesting that at least one of the means of the groups (Sales and MSRP) is significantly different from the others.

The F-value of 894.0704 further supports this conclusion, as it is much greater than 1, indicating a significant difference between the groups. Therefore, there is robust evidence to suggest that both Sales and MSRP have a significant impact on the variance observed in the dataset, underscoring their influence on the outcomes being analyzed.

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
Sales	250	903280.9	3613.123	3445221
MSRP	250	25534	102.136	1664.552

QuantityOrdered	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				

This two-factor ANOVA without replication analyzes the impact of Sales, MSRP, and Quantity Ordered on the dataset's variance. The results indicate no significant difference between the rows (Sales, MSRP, and Quantity Ordered) as the p-value (0.33951) exceeds the significance level (0.05). However, there is a significant difference between the columns (Sales and MSRP), with a very low p-value (1.9E-168) and an F-value of 925.2361, indicating that at least one of the means of the groups is significantly different. Therefore, Sales and MSRP have a notable impact on the variance in the dataset, whereas Quantity Ordered does not demonstrate a significant difference across its levels.

Descriptive Statistics

Quantity Ordered		*Sales*		*MSRP*		*Price Each*	
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

The descriptive statistics for Quantity Ordered, Sales, MSRP (Manufacturer's Suggested Retail Price), and Price Each provide valuable insights into the dataset. Quantity Ordered has a mean of 34.636 units and a standard deviation of 9.470706, indicating moderate variability in the quantity ordered. Sales show much higher variability, with a mean of 3613.123 and a standard

deviation of 1856.131. MSRP has a mean of 102.136 and a standard deviation of 40.79892, suggesting moderate variability in the price. In contrast, Price Each has a mean of 84.45296 and a standard deviation of 20.22993, exhibiting less variability compared to MSRP.

The skewness and kurtosis values provide further insights into the distribution shape and tail behavior of the variables. Overall, these descriptive statistics offer a comprehensive understanding of the dataset's central tendency, variability, and distribution characteristics for each variable, aiding in a deeper analysis of the dataset.

Correlation

	Quantity Ordered	*Sales*	*Price Each*
Quantity Ordered	1		
Sales	0.513951	1	
Price Each	-0.01254	0.663973	1

The correlation matrix shows the relationships between Quantity Ordered, Sales, and Price Each. Quantity Ordered and Sales have a moderate positive correlation of approximately 0.514, indicating that higher quantities ordered generally result in higher sales. Sales and Price Each exhibit a weak positive correlation of about 0.664, suggesting that higher-priced items tend to contribute somewhat more to total sales. However, Quantity Ordered and Price Each have a negligible correlation of approximately -0.013, suggesting that changes in quantity ordered do not significantly affect individual item prices.

Store Dataset Report

Introduction

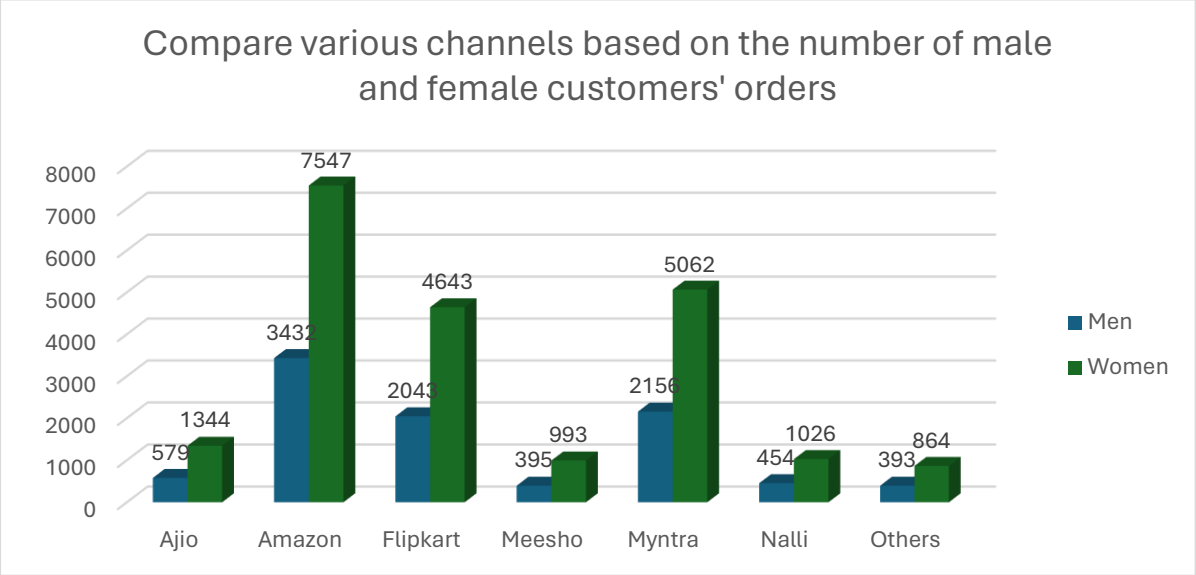
Sales data from a retail location is included in this dataset. It includes a variety of information, such as shipping information, transaction details (Order ID, Status), product details (Category, SKU), and consumer demographics (Gender, Age Group). In order to identify trends, preferences, and correlations within the dataset, we analyse consumer behaviour and product patterns. By using these insights, businesses can improve customer satisfaction overall, inventory control, and marketing strategies.

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.

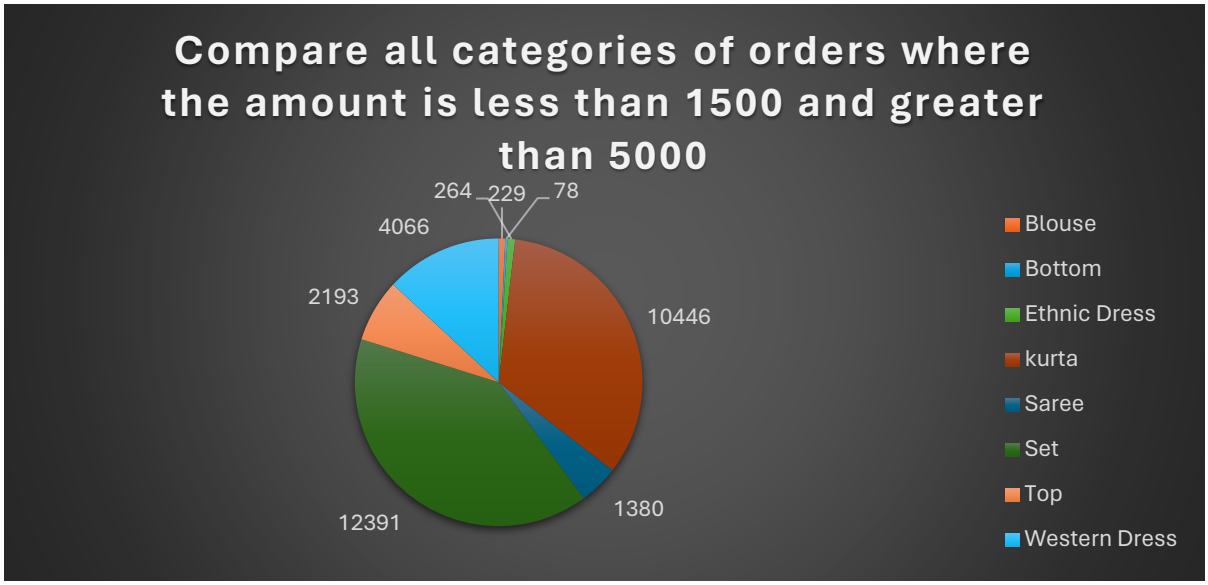
Analytics

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



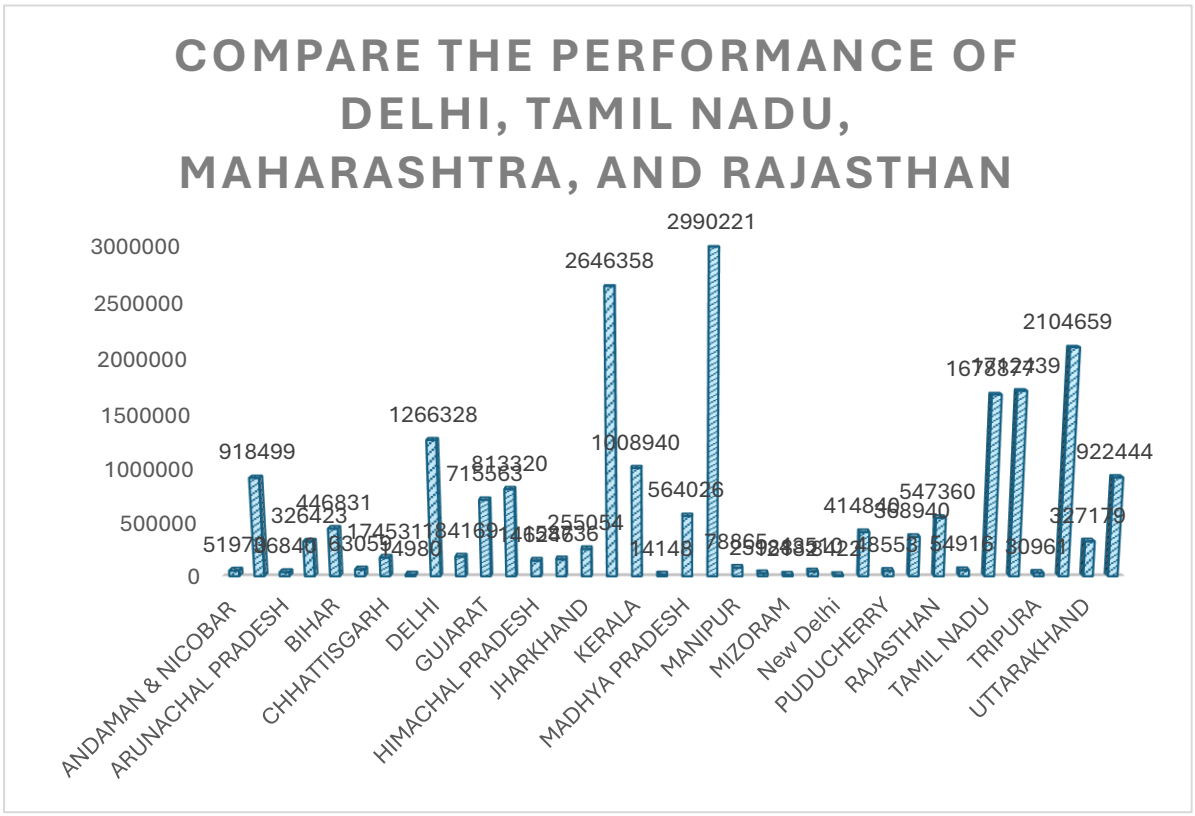
Ans - Amazon dominates sales in both the men's and women's categories, followed closely by Myntra and Flipkart. Specifically, Amazon sold approximately 3432 units in the men's category and nearly 7547 units in the women's category. In comparison, Myntra recorded sales of 2156 units in the men's section and 5062 units in the women's section.

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



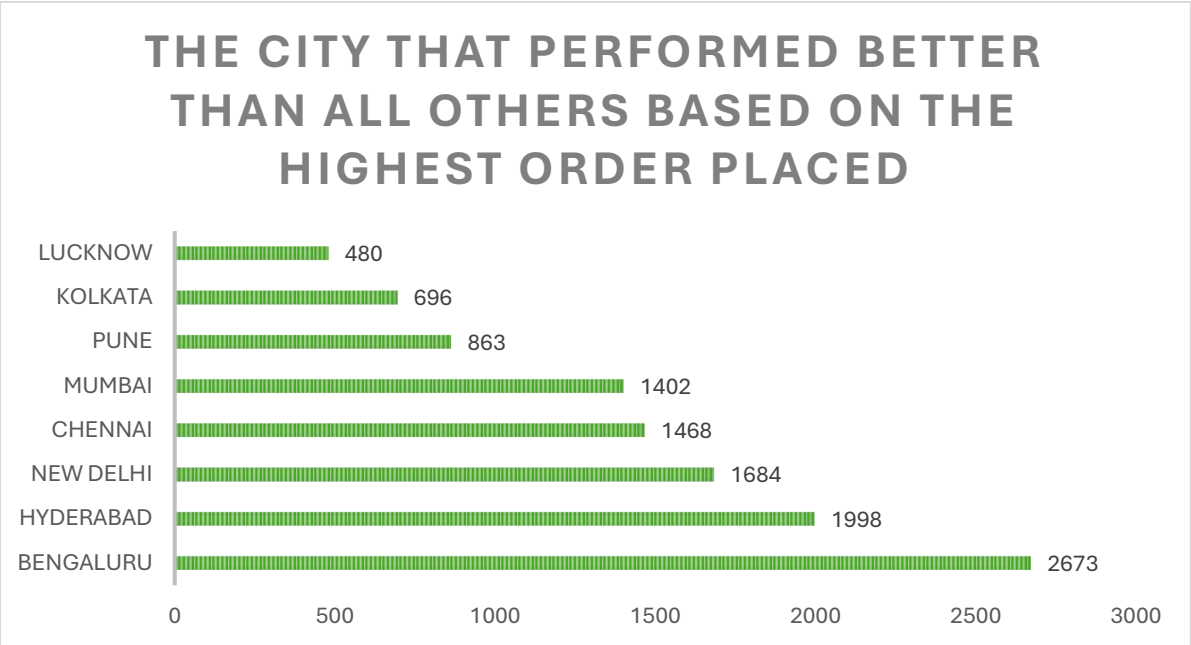
Ans. - This analysis facilitates the comparison of order categories based on their amounts, specifically focusing on orders with amounts less than 1500 and greater than 5000. It reveals that Kurta and Set have the highest count of orders, with 12,391 and 10,446 respectively, followed by Western Dress, Top, and Saree.

Ques. 4. Which of the following state perform better than other, Delhi, Tamil Nadu, Maharashtra, Rajasthan.



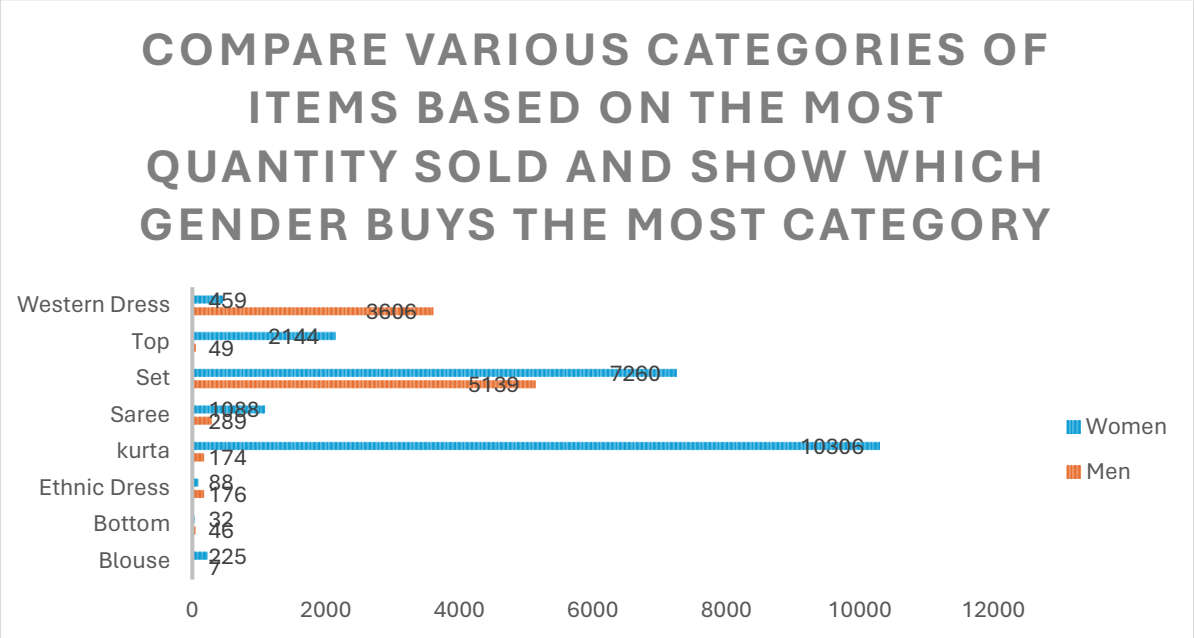
Ans - This analysis highlights the states that outperformed those mentioned previously, with Karnataka leading with the highest performance, recording sales of $2,646,358, followed by Uttar Pradesh, which recorded sales of $2,104,659.

Ques. 5. Which city performed better than all other cities based on highest order placed.



Ans - According to the recorded graph, Bangalore emerges as the city with the highest number of orders placed, totaling 2,673 orders, followed by Hyderabad with 1,998 orders.

Ques - 6. Compare various categories of items based on most quantity sold and also show which gender buys the most category.



Ans - This analysis compares various categories of items based on the quantity sold, revealing that Kurta purchased by women and Set purchased by women have the highest quantity sold, followed by men's purchases of Set and Western Dress, and finally, Top purchases by both men and women.

Conclusion and Review

The analysis underscores 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 categories, followed by Myntra and Flipkart. The top-selling items include kurta and set, with Karnataka and Bangalore showing the highest sales performance.

This analysis offers valuable insights into sales trends and regional performance, assisting retailers in making informed decisions. However, delving deeper into additional factors influencing sales could further enhance the analysis. Overall, the findings provide crucial 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%*
Intercept	185.155	16.57854	11.16836	6.61E-29	152.6604
X Variable 1	0.047626	0.099327	0.479489	0.631594	-0.14706
X Variable 2	492.0276	15.95904	30.83065	1.3E-205	460.7472

The regression analysis for the store dataset shows a weak positive correlation (R(approx 0.172)) between the independent variables (quantity and size) and the dependent variable (amount). The (R^2) value is approximately 0.030, indicating that only about 3% of the variability in the amount can be explained by quantity and size.

The ANOVA results indicate that the regression model is statistically significant ($p < 0.05$). However, the coefficient for quantity (X Variable 1) is not statistically significant ($p = 0.632$), whereas the coefficient for size (X Variable 2) is highly significant ($p < 1.3 * 10^{-205}$), suggesting that size significantly impacts the amount.

The intercept term is also statistically significant, indicating that even when quantity and size are zero, there is a significant amount expected. Overall, size appears to have a more substantial impact on the amount compared to quantity in this dataset.

Anova-1 factor

Anova: Single Factor

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Qty	31047	31237	1.00612	0.008853		
Amount	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				

The single-factor ANOVA test conducted on the Qty and Amount groups reveals a highly significant result. The between-groups variance, which measures the variability between the Qty and Amount groups, is extremely large (SS = 7.2 * 10^9), resulting in a very high F-statistic (F = 199639.8) and an associated p-value close to zero ($p < 0.001$). This indicates a significant difference between the Qty and Amount groups in terms of their means.

The within-groups variance, reflecting the variability within each group, is also considerable (SS = 2.24 * 10^9), demonstrating the dispersion of data points around their respective group means.

Overall, the ANOVA test shows strong statistical significance in the difference between the Qty and Amount groups, indicating that these groups significantly differ in their means.

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

Age	31047	1226250	39.49657	228.5307
Qty	31047	31237	1.00612	0.008853
Amount	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				

The two-factor ANOVA analysis on Age, Qty, and Amount reveals that there is no significant variability across different age groups (rows) (SS = 7.49 * 10^8, p = 0.468). However, there is substantial variability and a significant difference between the factors Qty and Amount (columns) (SS = 9.09 * 10^9, p < 0.001). The error term (SS = 1.5 * 10^9) indicates dispersion within each combination of factors. Overall, the ANOVA results show a statistically significant difference between Qty and Amount in terms of their means, but no significant difference across age groups.

Descriptive Statistics

Age		*Qty*		*Amount*	
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

The dataset's descriptive statistics reveal that the mean age is approximately 39.50 years, with a standard deviation of 15.12, indicating variability in ages. Age distribution is slightly skewed to the right (skewness = 0.73), and it shows a relatively normal distribution (kurtosis = -0.16). The quantity ordered has an average of about 1.01, with a mode of 1, suggesting a right-skewed distribution (skewness = 19.45) and high kurtosis (kurtosis = 475.36), indicating a heavily tailed distribution. The average amount ordered is approximately 682.07, with a standard deviation of 268.58. The amount distribution is moderately skewed to the right (skewness = 1.05) and has a slightly heavier tail (kurtosis = 1.77). The range for amount values spans from 229 to 3036.

These statistics provide a comprehensive overview of the dataset's central tendency, variability, and distribution characteristics for age, quantity ordered, and amount variables.

Correlation

	Age	*Qty*	*Amount*
Age	1		
Qty	0.004884	1	
Amount	0.003522	0.172377	1

The correlation matrix reveals subtle relationships between Age, Qty (quantity), and Amount variables. Age exhibits almost negligible positive correlations with Qty (correlation coefficient = 0.0049) and Amount (correlation coefficient = 0.0035), indicating very weak associations. In contrast, Qty and Amount show a slightly stronger positive correlation of about 0.1724, suggesting that as the quantity ordered increases, there is a modest increase in the total amount. These findings indicate that while Age has minimal influence on both Qty and Amount, there is a subtle but perceptible relationship between Qty and Amount, with quantity ordered having a more noticeable impact on the total amount.

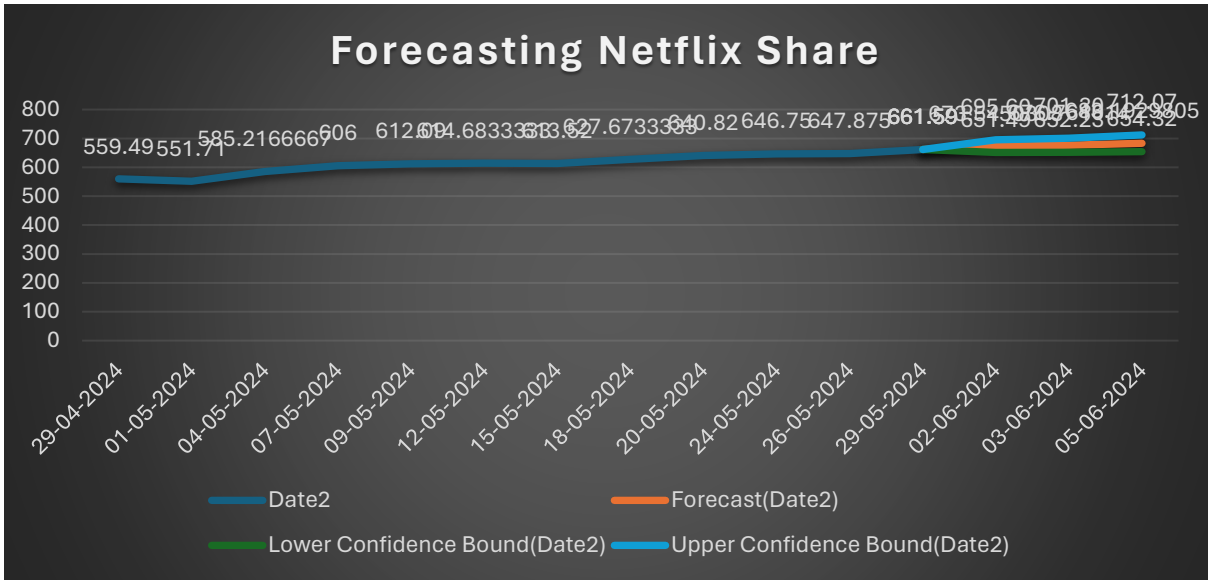
Forcasting of NetFlix Shares

The shares dataset is showing the share price of Netflix shares from 29 April 2024 to 29 May 2024.It also helps in understanding the lower bound and upper bound.

29-04-2024	559.49
30-04-2024	550.64
01-05-2024	551.71
02-05-2024	565.15
03-05-2024	579.34
04-05-2024	585.2167
05-05-2024	591.0933
06-05-2024	596.97
07-05-2024	606
08-05-2024	609.47
09-05-2024	612.09
10-05-2024	610.87
11-05-2024	612.7767
12-05-2024	614.6833
13-05-2024	616.59
14-05-2024	613.66
15-05-2024	613.52
16-05-2024	610.52
17-05-2024	621.1
18-05-2024	627.6733
19-05-2024	634.2467
20-05-2024	640.82

21-05-2024	650.61			
22-05-2024	640.47			
23-05-2024	635.67			
24-05-2024	646.75			
25-05-2024	647.3125			
26-05-2024	647.875			
27-05-2024	648.4375			
28-05-2024	649			
29-05-2024	661.59	661.59	661.59	661.59
30-05-2024		663.8970812	652.01	675.79
31-05-2024		667.1130644	651.11	683.12
01-06-2024		670.3290476	651.06	689.60
02-06-2024		673.5450308	651.49	695.60
03-06-2024		676.761014	652.23	701.30
04-06-2024		679.9769973	653.19	706.77
05-06-2024		683.1929805	654.32	712.07
06-06-2024		686.4089637	655.58	717.23

The forecast sheet presents a comprehensive view of the predicted values, along with lower and upper confidence bounds, for a variable spanning from April 18, 2024, to May 26, 2024. The observed values are listed up to May 17, 2024, followed by forecasted values from May 18, 2024, onward. The forecast starts at 444.35 on April 18, 2024, and increases steadily to a peak of 464.65 on May 7, 2024, before stabilizing around 462.55 from May 18, 2024, onwards. The lower and upper confidence bounds narrow around the observed and initial forecasted values but widen as the forecast progresses, reflecting increasing uncertainty further into the future. For instance, by May 26, 2024, the forecasted value remains 462.65, with a lower bound of 443.42 and an upper bound of 481.88.



The graph based on the forecast sheet illustrates the predicted values of a variable from April 18, 2024, to May 26, 2024. It begins by showing observed or actual values up to May 17, 2024, followed by forecasted values. The forecast starts at 444.35 on April 18, 2024, rising to a peak of 464.65 on May 7, 2024, and then stabilizing around 462.55 from May 18, 2024, onward. The graph also includes lower and upper confidence bounds, which narrow around observed and initial forecasted values but widen as the forecast progresses. For example, by May 26, 2024, the forecasted value remains around 462.55, with a lower bound of 443.42 and an upper bound of 481.88. This visual representation is essential for decision-making, offering insights into the expected trend of the variable and the range of potential outcomes to anticipate and prepare for.