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
Aman Mehra(0827CS201026)

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 TCS Shares	

Assignment-1

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.

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.

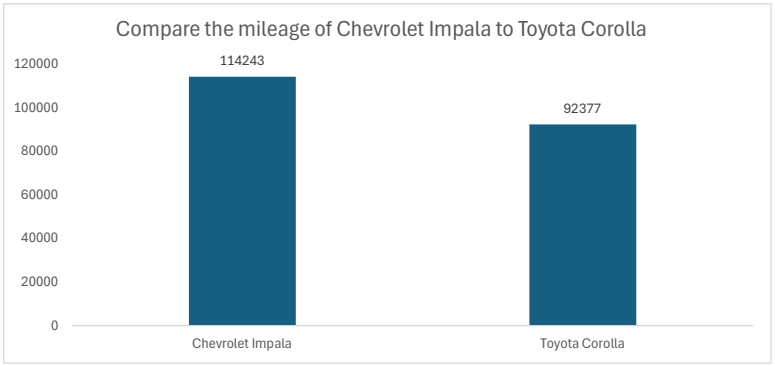
- **Velocity**: This describes the speed at which data is generated, collected, and processed. In many cases, data needs to be processed in real-time or near-real-time to be useful, such as in financial transactions, social media feeds, and IoT applications

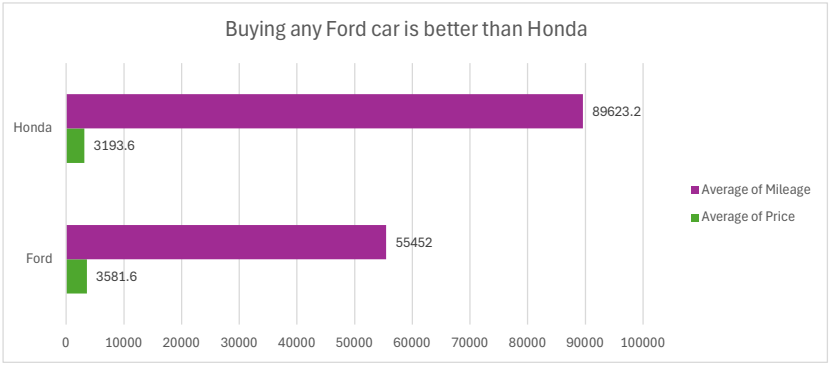
- **Variety**: This refers to the different types of data that are available. Data can be structured (like databases), semi-structured (like XML files), and unstructured (like text, video, and audio files). The diversity of data types presents challenges in terms of storage, processing, and analysis.

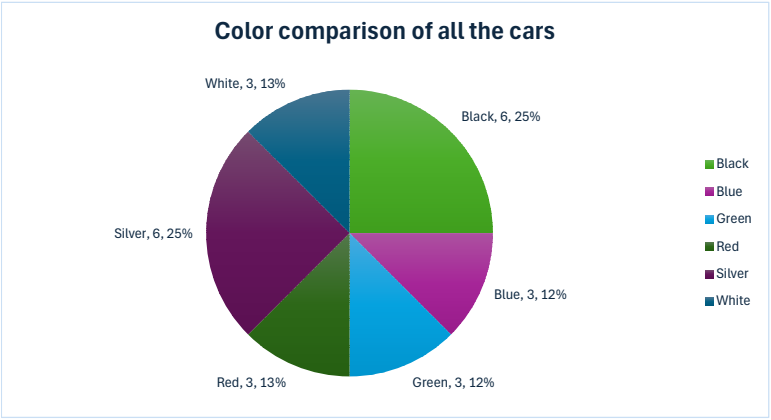
- **Veracity**: This pertains to the quality and accuracy of the data. High veracity means the data is trustworthy and accurate, while low veracity indicates a higher level of uncertainty and the potential presence of noise or errors in the data.

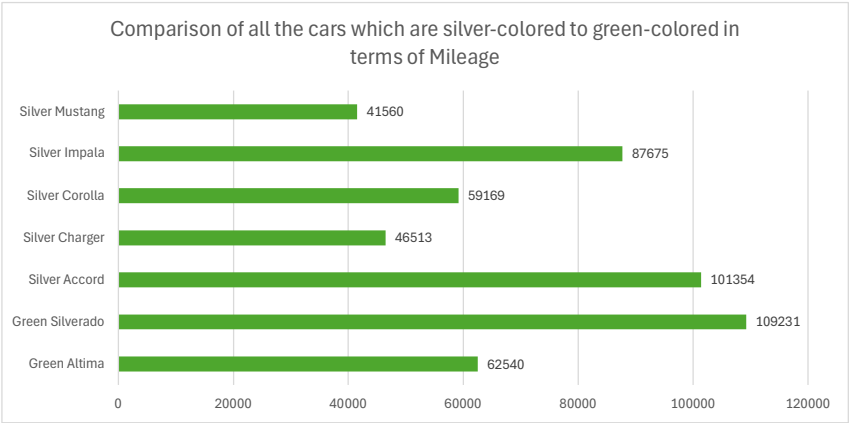
- **Value**: This is about the usefulness of the data. Data alone doesn't hold value; it needs to be processed and analyzed to extract actionable insights that can drive business decisions and strategies. The ultimate goal of big data initiatives is to derive significant value from the data.

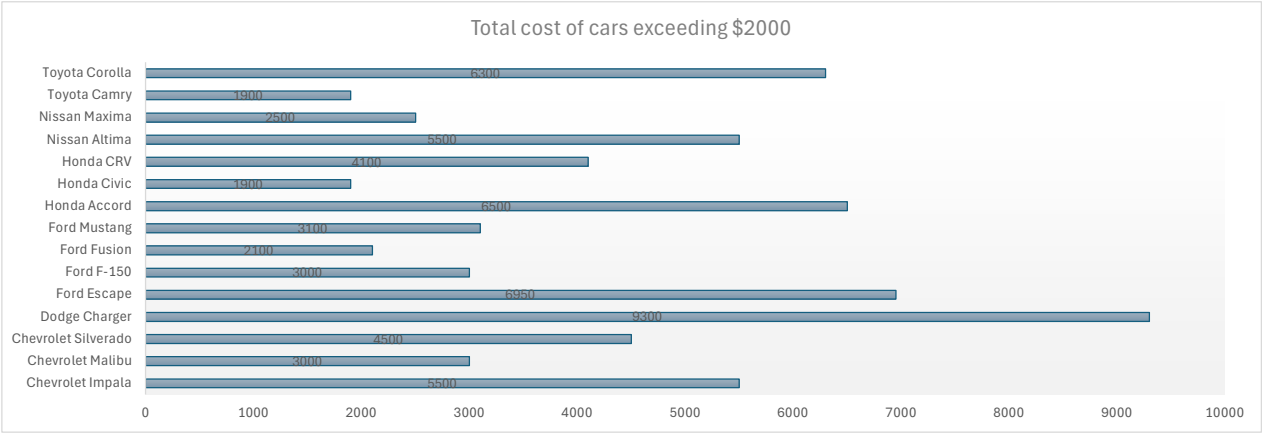
Dashboard of Car data



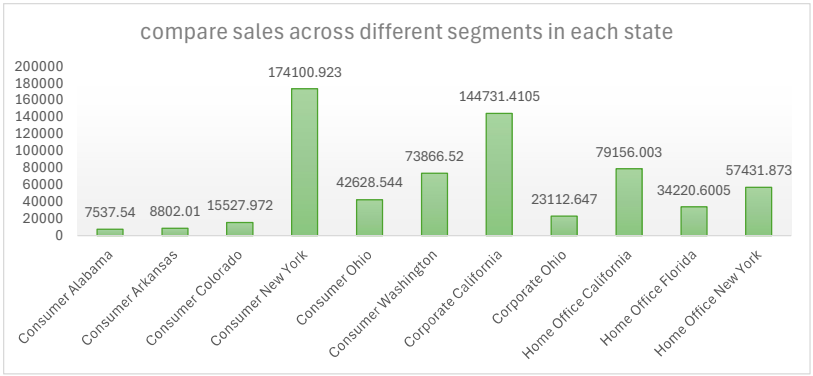


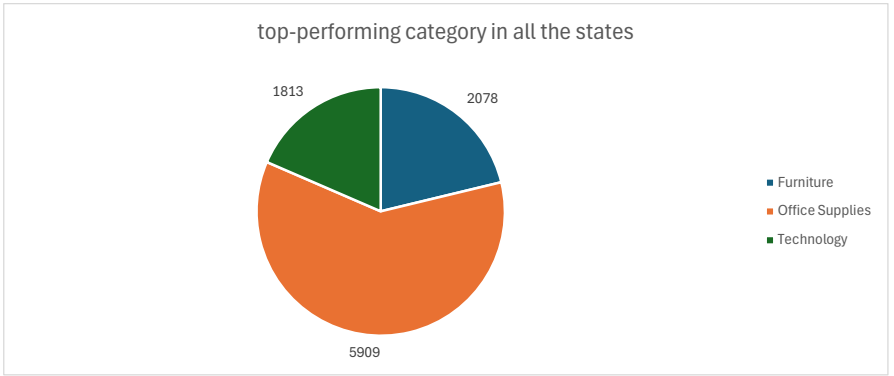


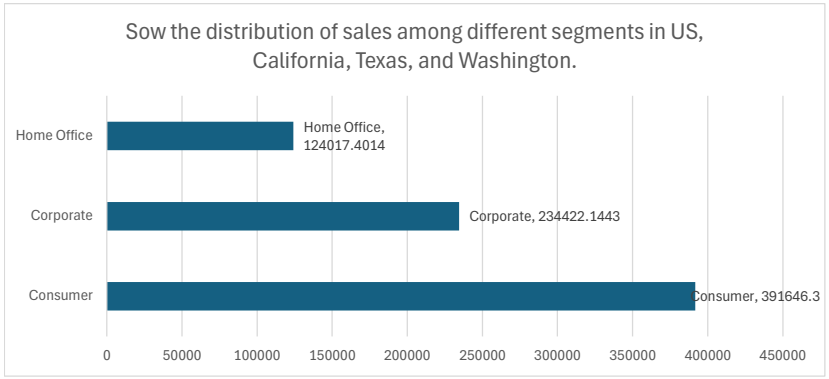


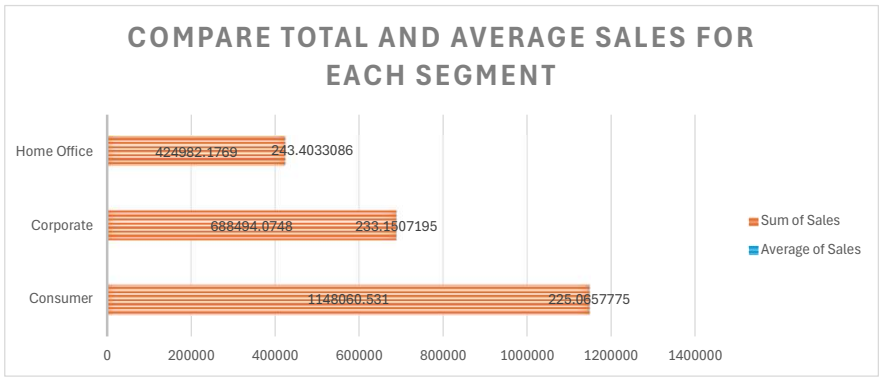


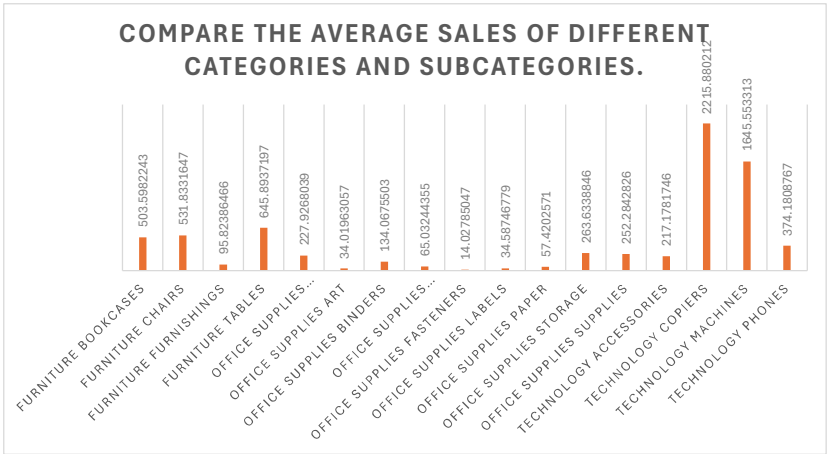
Dashboard of Order data



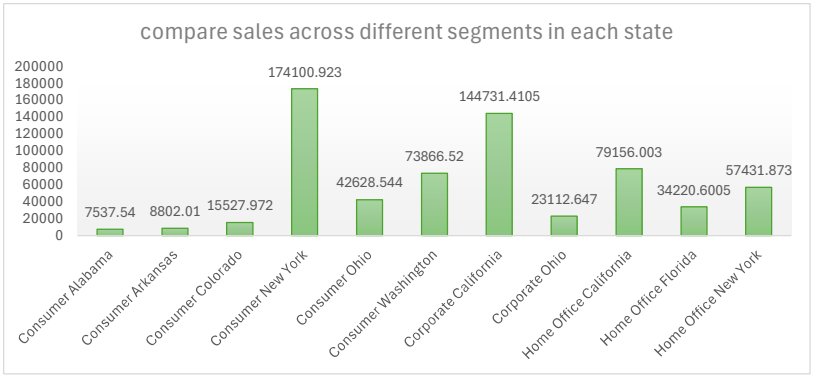


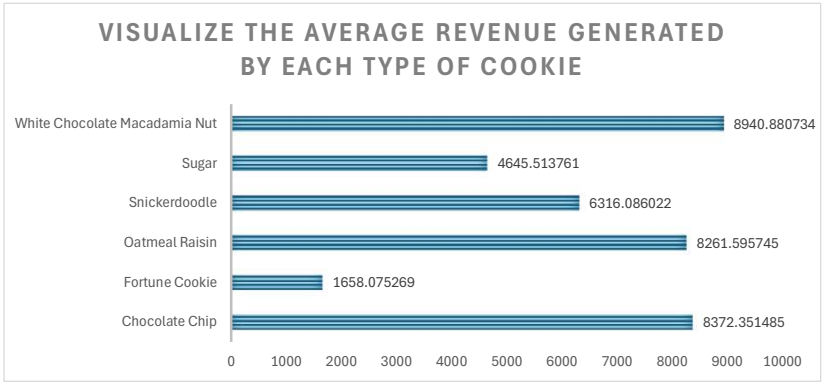


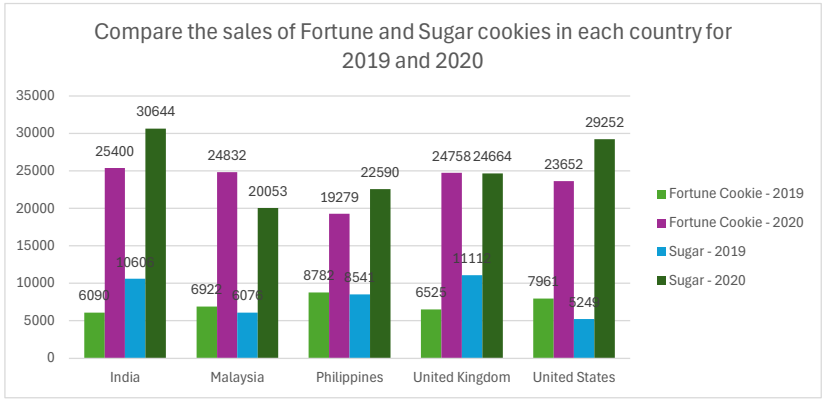


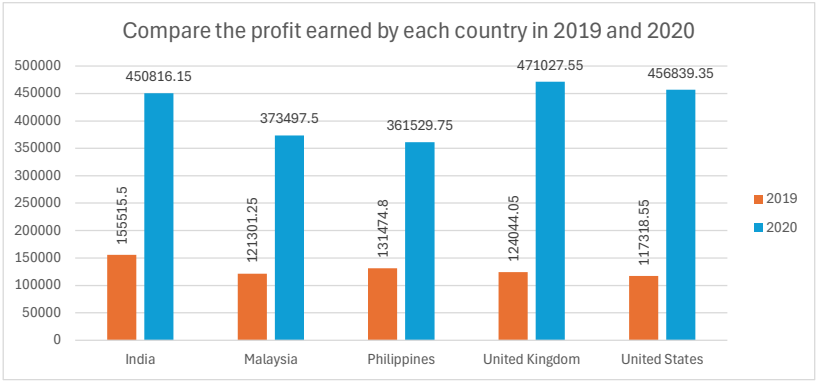


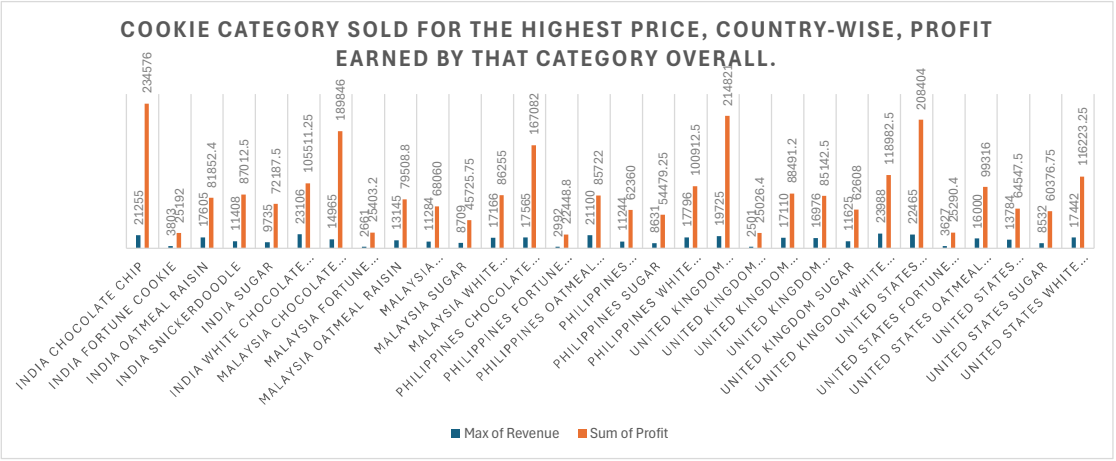
Dashboard of Cookie data



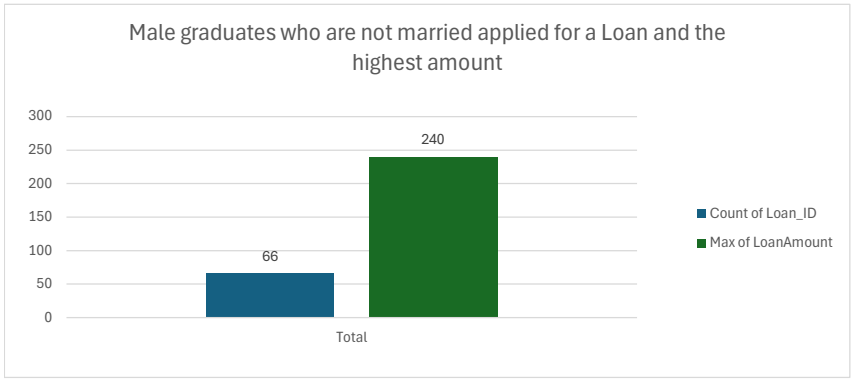


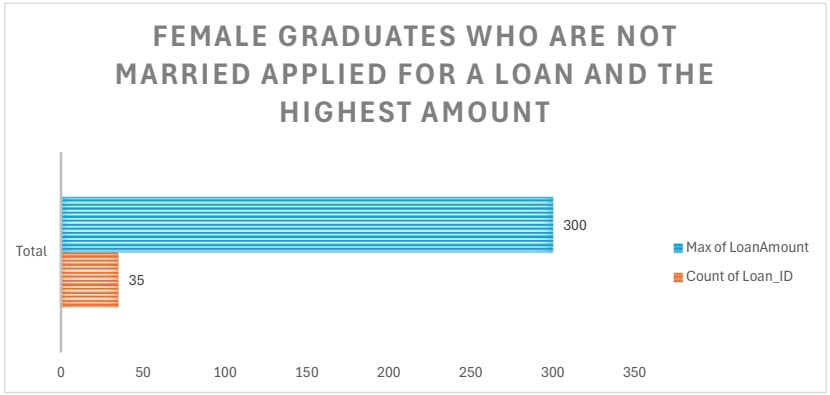


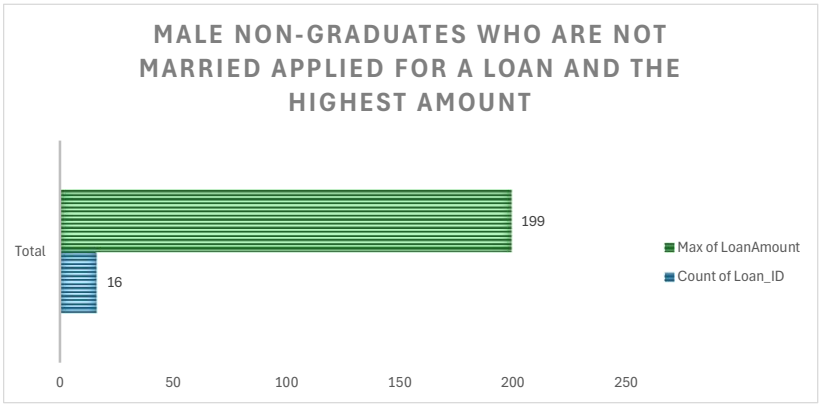


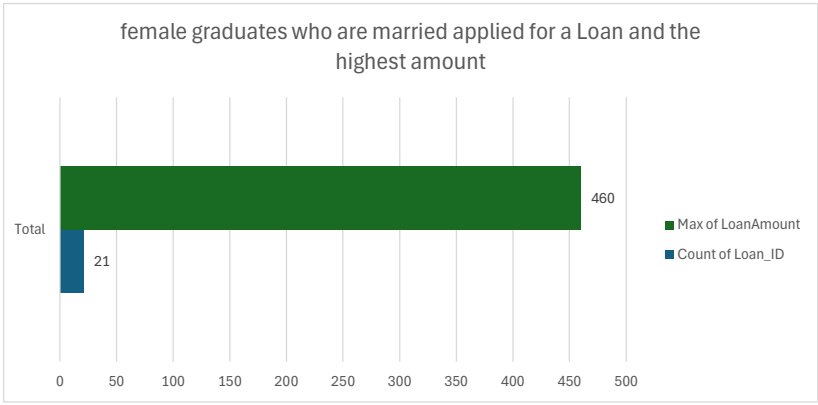


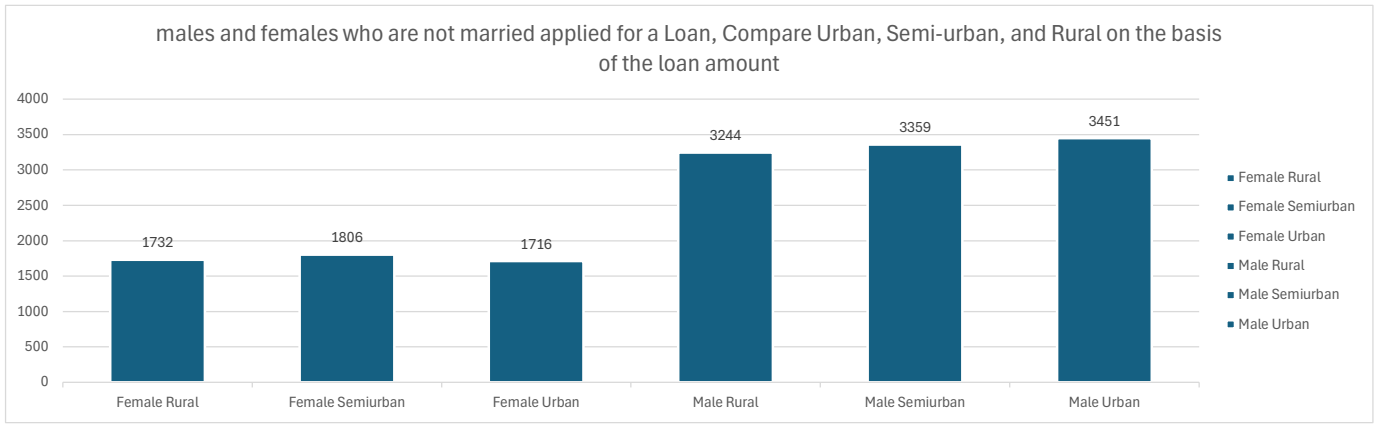
Dashboard of Loan data



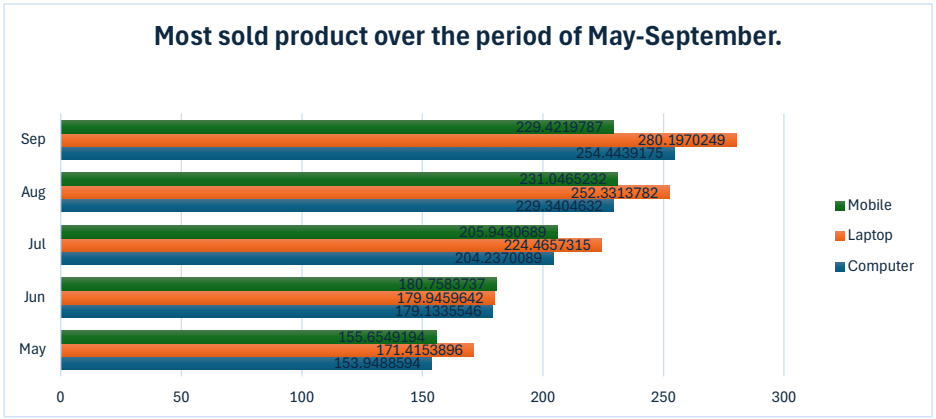


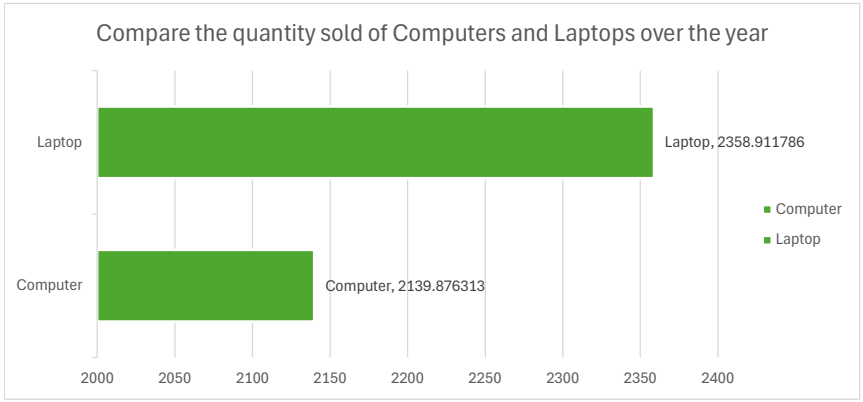


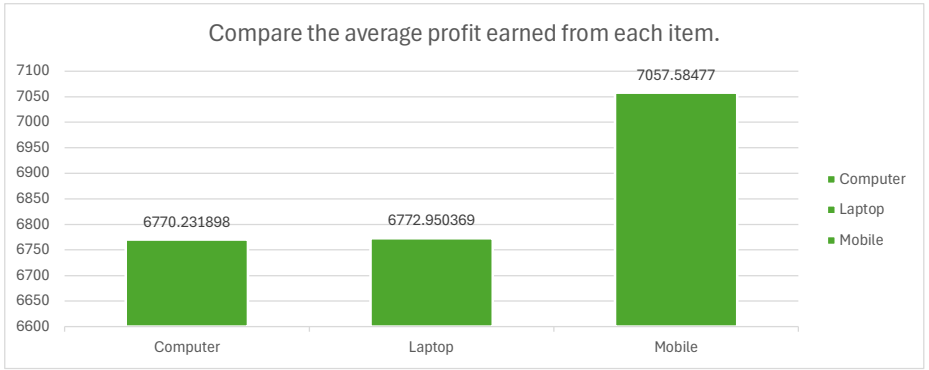


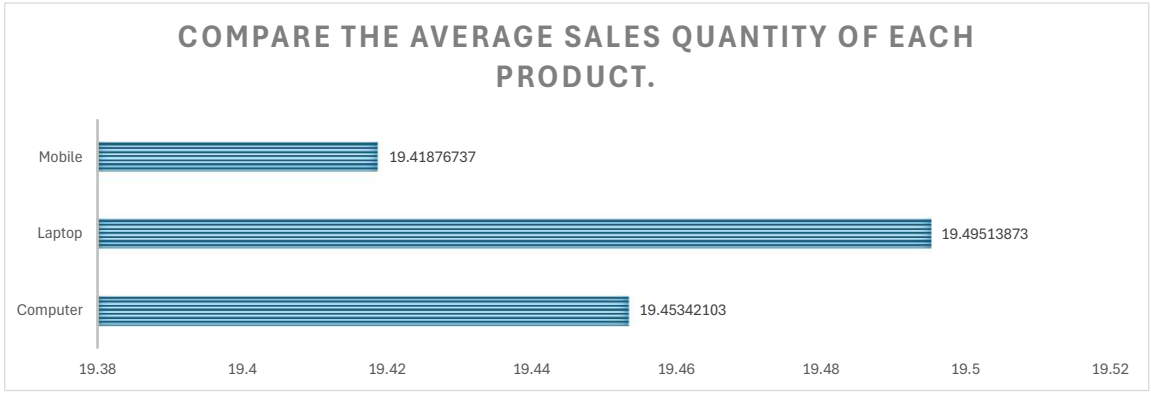


Dashboard of Shop sales data

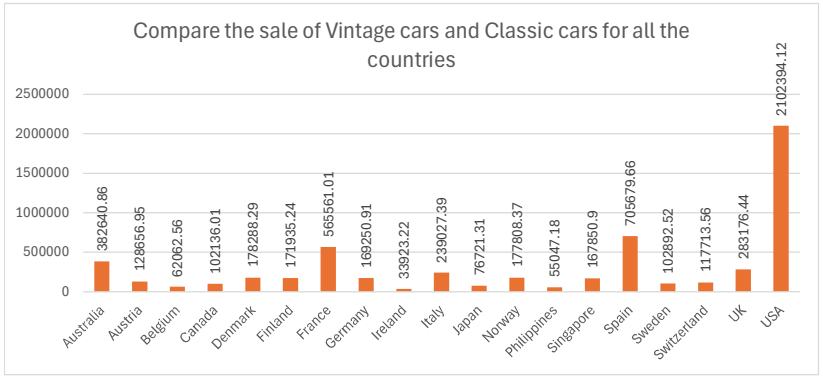


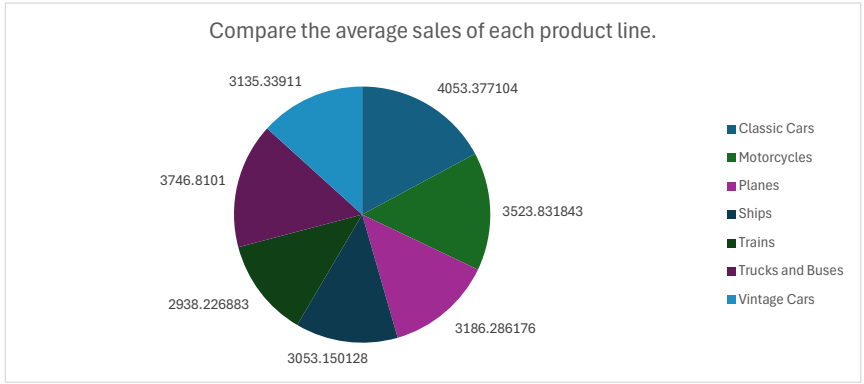


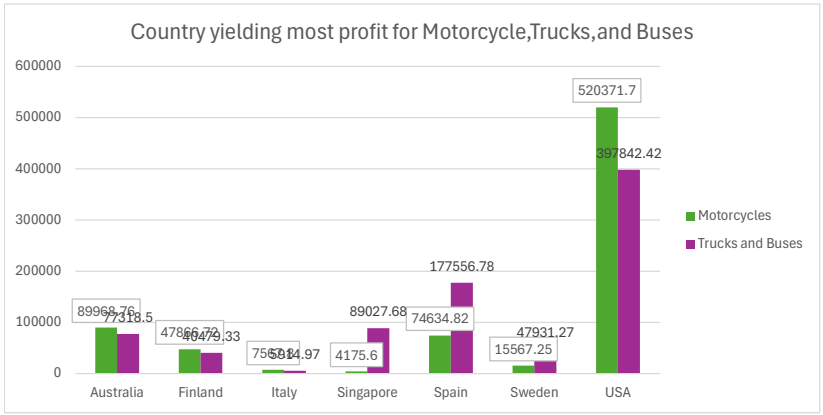


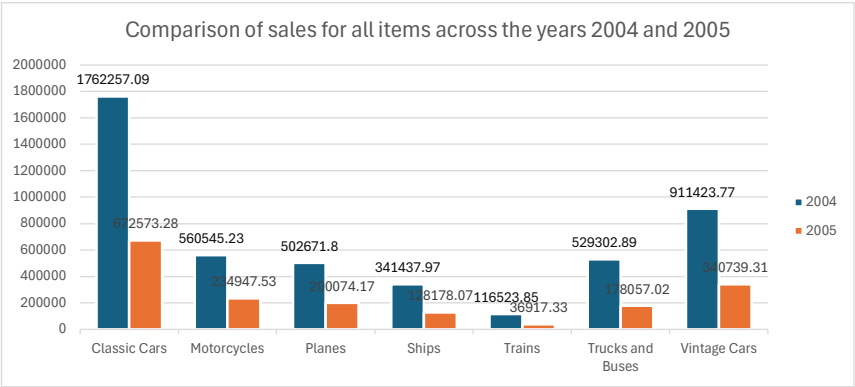


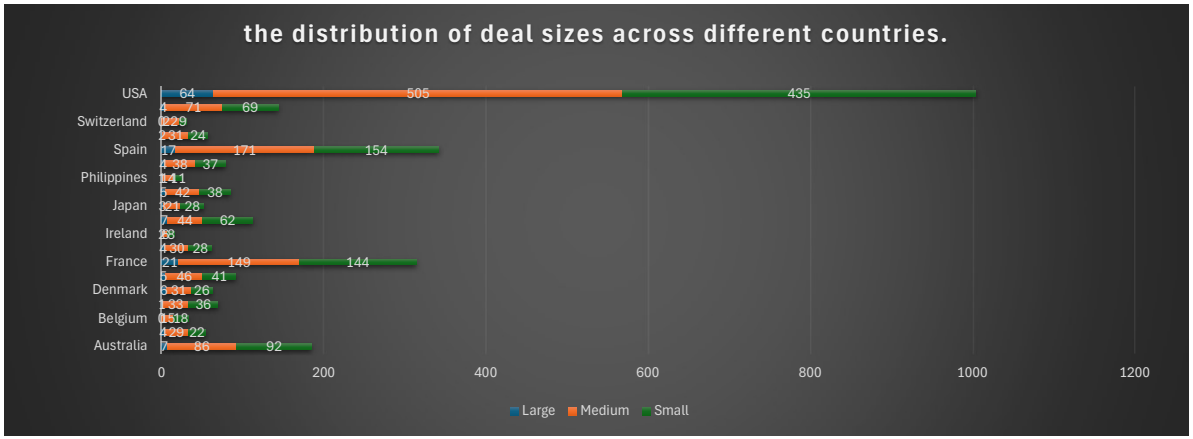
Dashboard of Sales Data Sample



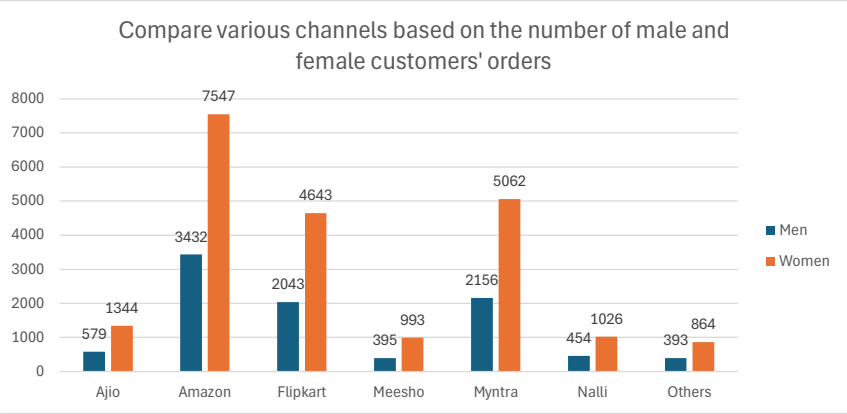


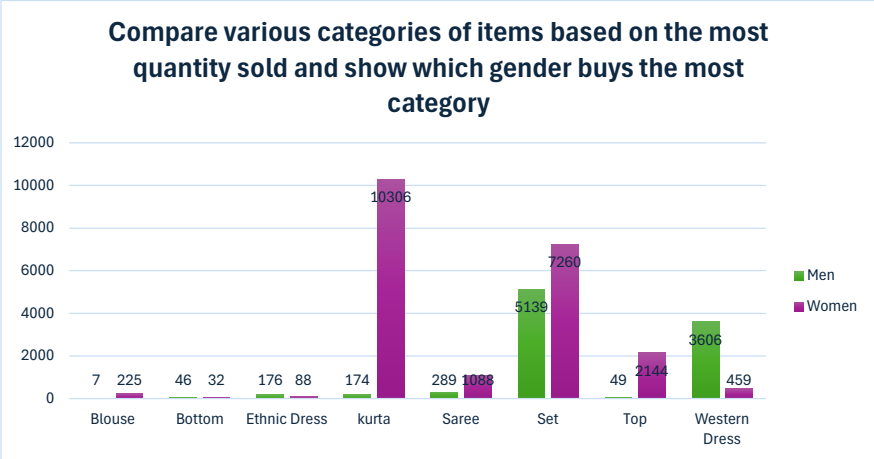






Dashboard of Store Data





Car Collection Data Report

Introduction

This report provides an in-depth analysis of a dataset containing various attributes of different car models, such as make, model, color, mileage, price, and cost. The goal is to derive insights to assist in decision-making regarding car purchases and understanding market trends. The dataset includes six cars: Honda, Chevrolet, Nissan, Toyota, Dodge, and Ford. This report targets car enthusiasts, automotive industry professionals, analysts, and anyone interested in car market trends. It includes statistical analyses, visualizations, and interpretations.

Questionnaire

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

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

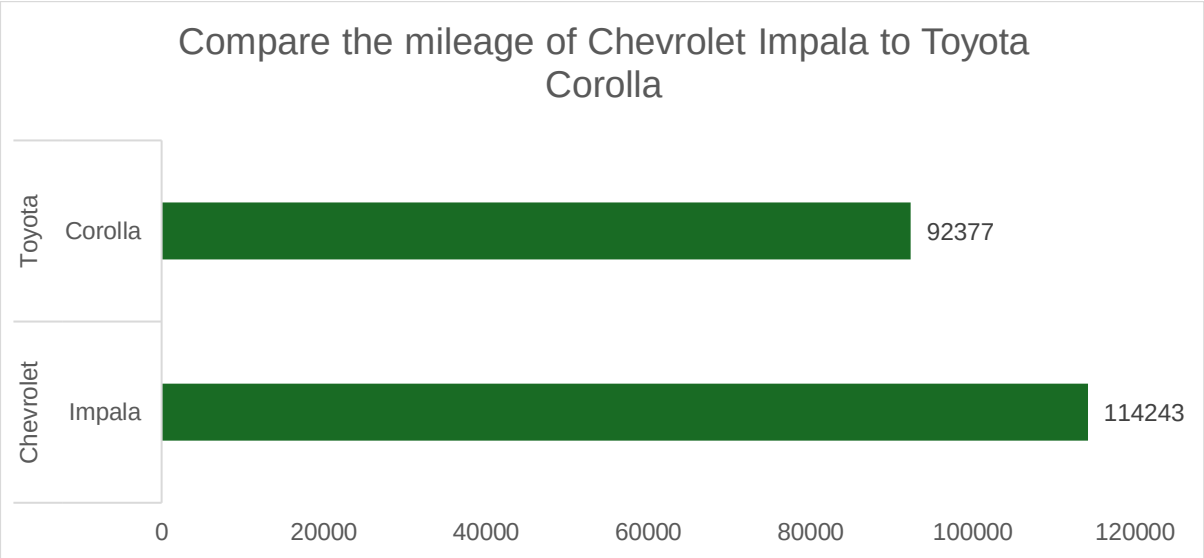
3. Among all the cars which car color is the most popular and is least popular?

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

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

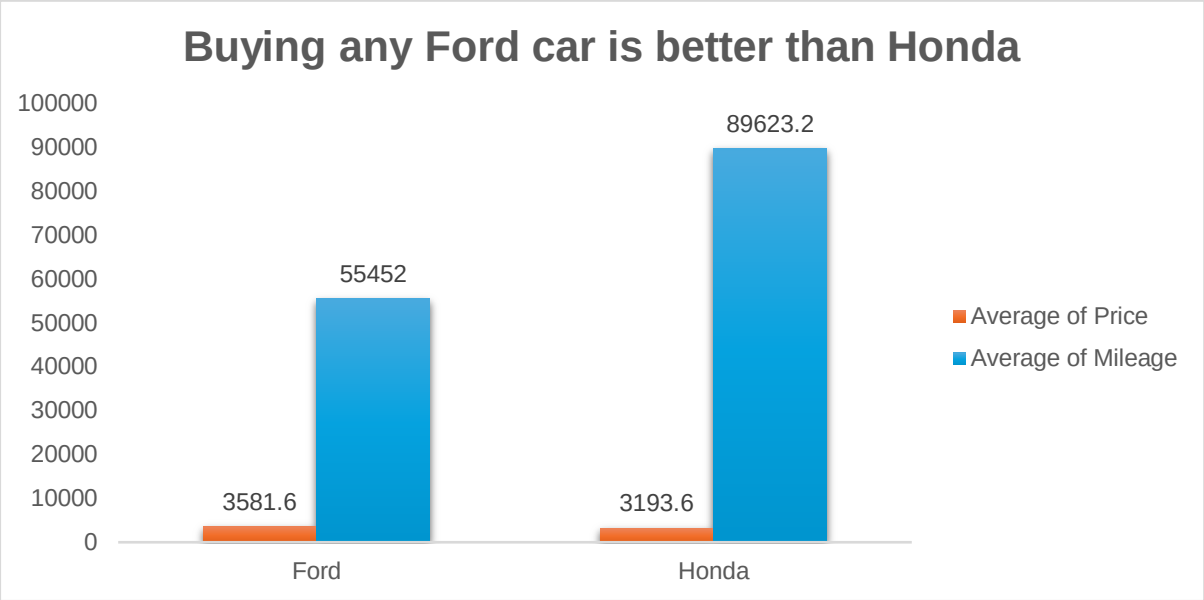
Analytics

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



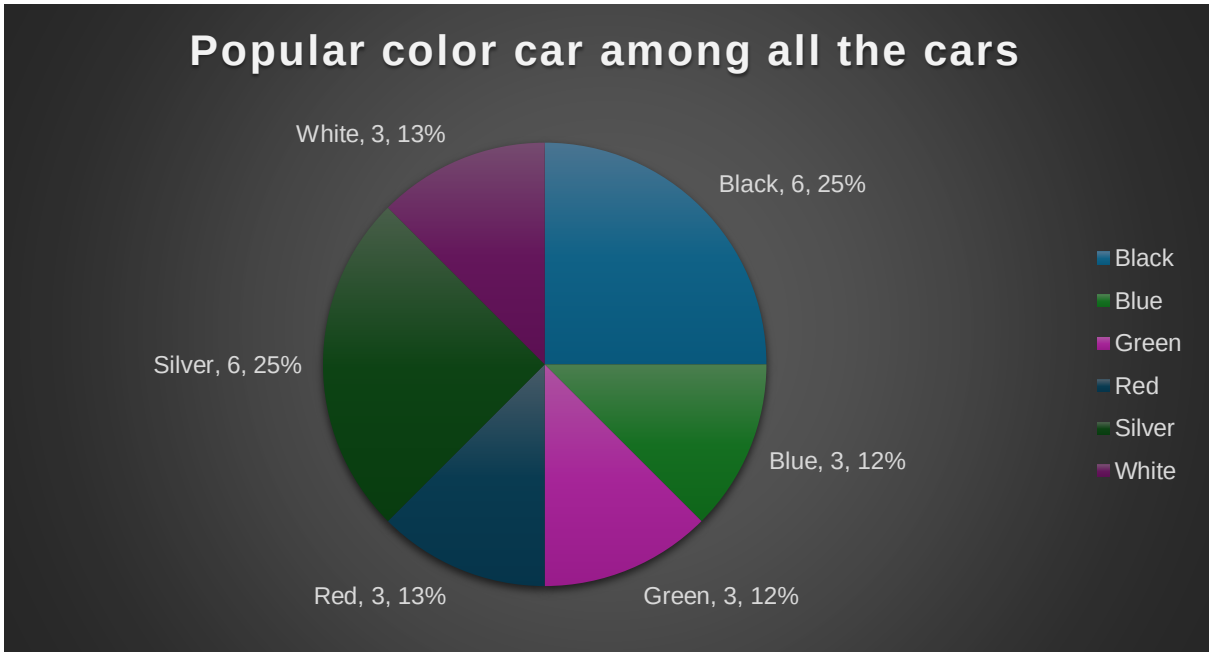
This study evaluates the fuel efficiency (mileage) of two widely recognized car models: the Chevrolet Impala and the Toyota Corolla. To achieve this, the dataset was filtered to select pertinent data, and a column chart was created for visualization. The findings from the analysis revealed that the Chevrolet Impala, with a mileage of 114,243, outperforms the Toyota Corolla, which has a mileage of 92,377.

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



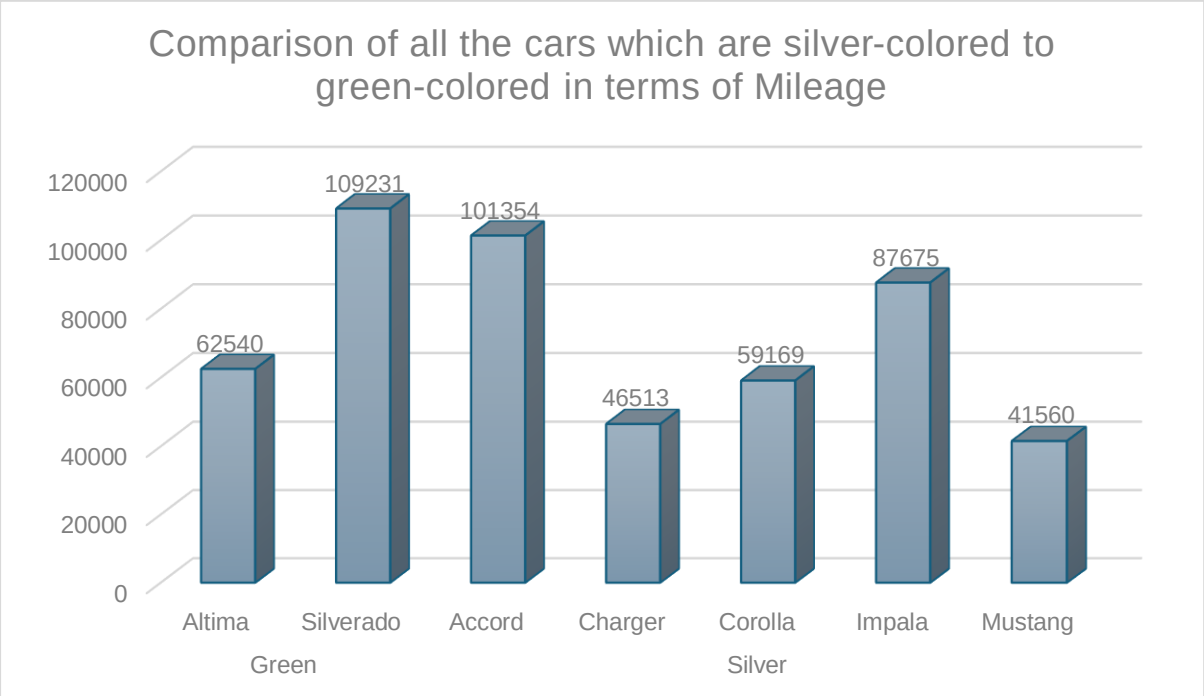
This analysis seeks to justify purchasing a Ford car over a Honda by comparing their respective attributes, with a particular focus on price. However, after analyzing the dataset, the findings do not support this statement. Instead, Honda cars were found to have better average mileage (89,623.3) and a lower average price (3,193.6) compared to Ford cars.

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



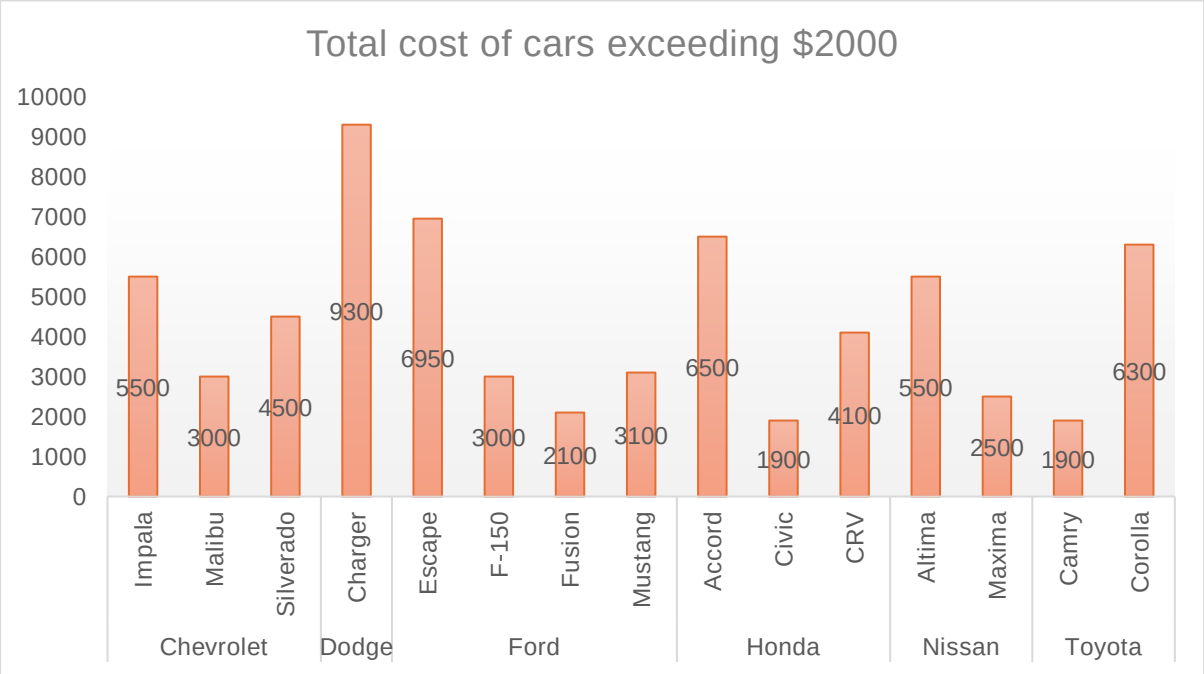
This analysis aims to identify the most and least popular car colors among all the cars in the dataset based on the count of each make. The results indicate that Black and White are the most popular car colors, each accounting for 25% of the total makes. In contrast, Green and Blue are the least popular, each representing 12% of the makes.

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



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

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



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

Conclusion and Review

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

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

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

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

Cars Costing More Than $2000: The analysis determined that the total cost of cars exceeding $2,000 amounted to $66,150.

The analysis provided valuable insights into various aspects of the dataset, including mileage comparisons, car color popularity, and cost considerations. However, there were discrepancies between the initial assumptions and the findings, particularly in the comparison between Ford and Honda cars. The analysis was thorough and utilized appropriate visualizations, such as column charts and bar graphs, to present the findings effectively. Overall, the report offers valuable information for car buyers, industry professionals, and researchers interested in understanding trends within the car market. Nonetheless, it's important to note the limitations of the analysis, such as the dataset's completeness and the need for further exploration into other factors influencing car purchasing decisions.

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 Regression Analysis table sheds light on how the dependent variable (Mileage) relates to the independent variables (Cost and Price) in the Car Collection Dataset. The analysis reveals a strong positive correlation (Multiple R = 0.962639) between these variables, indicating that Cost and Price together explain about 92.67% of the variation in Mileage (R Square = 0.926673). Both Cost and Price significantly influence Mileage, as shown by their coefficients and low p-values. Notably, Price has a greater impact on Mileage compared to Cost. The ANOVA table further validates the regression model's overall significance with a very low p-value (1.22E-12) for the F-statistic. In summary, the regression analysis suggests that Cost and Price significantly affect Mileage in the car collection dataset.

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				

The Single Factor ANOVA table examines three groups: Mileage, Cost, and Price. Each group consists of 24 observations. For the Mileage group, the total sum is 2,011,267, with an average of 83,802.79 and a variance of 1.21E+09. The Cost group has a total sum of 66,150, an average of 2,756.25, and a variance of 705,502.7. The Price group has a total sum of 78,108, an average of 3,254.5, and a variance of 837,024.1.

The ANOVA analysis evaluates the differences in means among these groups. The Between Groups Sum of Squares (SS) is 1.04E+11 with 2 degrees of freedom (df), resulting in a Mean Squares (MS) of 5.22E+10. This indicates substantial variation among the group means. The F-statistic is 128.8822, with a p-value of 5E-24, which is significantly lower than the

conventional significance level of 0.05. This suggests a highly significant difference in means among the groups. The Within Groups SS is 2.8E+10 with 69 df, leading to 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				

The Two Factor ANOVA table examines two factors: Rows and Columns, each with levels of Mileage, Cost, and Price. The summary section provides details for each row and column combination, including the count, sum, average, and variance. For instance, Row 1 (Mileage) has 3 observations, a total sum of 70,512, an average of 23,504, and a variance of 1.2E+09. Row 2 (Cost) also has 3 observations, a total sum of 99,635, an average of 33,211.67, and a variance of 2.88E+09.

In the ANOVA section, the analysis evaluates the sources of variation in the data. The Sum of Squares for Rows (SS) is 8.95E+09 with 23 degrees of freedom (df), resulting in a Mean Squares (MS) of 3.89E+08. The Sum of Squares for Columns (SS) is 1.04E+11 with 2 df, yielding an MS of 5.22E+10. Both Rows and Columns have p-values greater than 0.05, indicating that neither Rows nor Columns significantly affect the observed variances. The Error Sum of Squares (SS) is 1.9E+10 with 46 df, and the Total Sum of Squares (SS) is 1.32E+11.

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

The descriptive statistics offer key insights into the variables Mileage, Cost, and Price. For Mileage, the mean value is 83,802.79 with a standard error of 7,112.652. The median is 81,142, and the standard deviation is 34,844.74, showing a moderate data spread around the mean. The range is 105,958, with a minimum of 34,853 and a maximum of 140,811. The skewness is slightly positive at 0.386522, indicating a right-skewed distribution, and the kurtosis is -1.09718, suggesting a relatively flat distribution.

Regarding Cost, the mean is 2,756.25 with a standard error of 171.4525. The median is 2,750, and the standard deviation is 839.9421. The range is 3,000, with a minimum of 1,500 and a maximum of 4,500. The skewness is slightly positive at 0.473392, indicating a right-skewed distribution, and the kurtosis is -0.81266, indicating a relatively flat distribution.

For Price, the mean is 3,254.5 with a standard error of 186.7512. The median is 3,083, and the standard deviation is 914.8902. The range is 2,959, with a minimum of 2,000 and a maximum of 4,959. The skewness is slightly positive at 0.272019, indicating a right-skewed distribution, and the kurtosis is -1.20291, suggesting a relatively flat distribution. Overall, these statistics provide a detailed look at the central tendency, variability, and distribution shape for each variable.

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 report offers a deep dive into a comprehensive dataset capturing sales transactions within the automotive industry. It includes various attributes such as Order ID, Order Date, Ship Date, Customer Details, Product Information, and Sales Figures. The primary objective is to extract actionable insights to inform decision-making processes and drive business growth within the automotive sector.

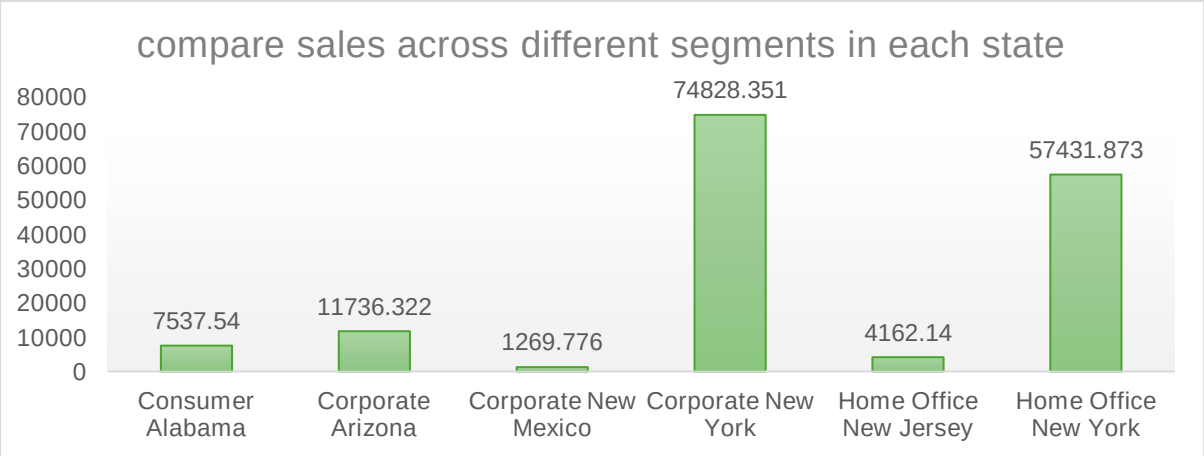
By analyzing sales data across different US states, segments, categories, and sub-categories, this report aims to pinpoint key trends, identify top-performing segments, and highlight areas of potential growth. The insights derived from this analysis will be invaluable for automotive industry stakeholders, including sales managers, marketers, and executives, who are keen on optimizing sales strategies, enhancing customer satisfaction, and maximizing revenue.

Questionnaire

1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?
2. Find out top performing category in all the states?
3. Which segment has the most sales in the US, California, Texas, and Washington?
4. Compare total and average sales for all different segments?
5. Compare the average sales of different categories and subcategory of all the states.

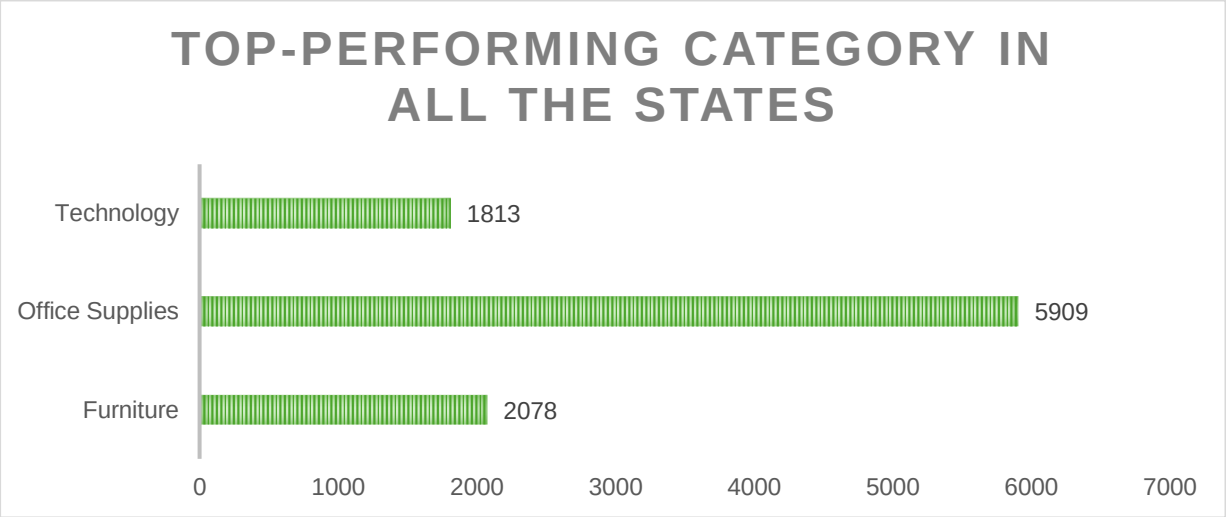
Analytics

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



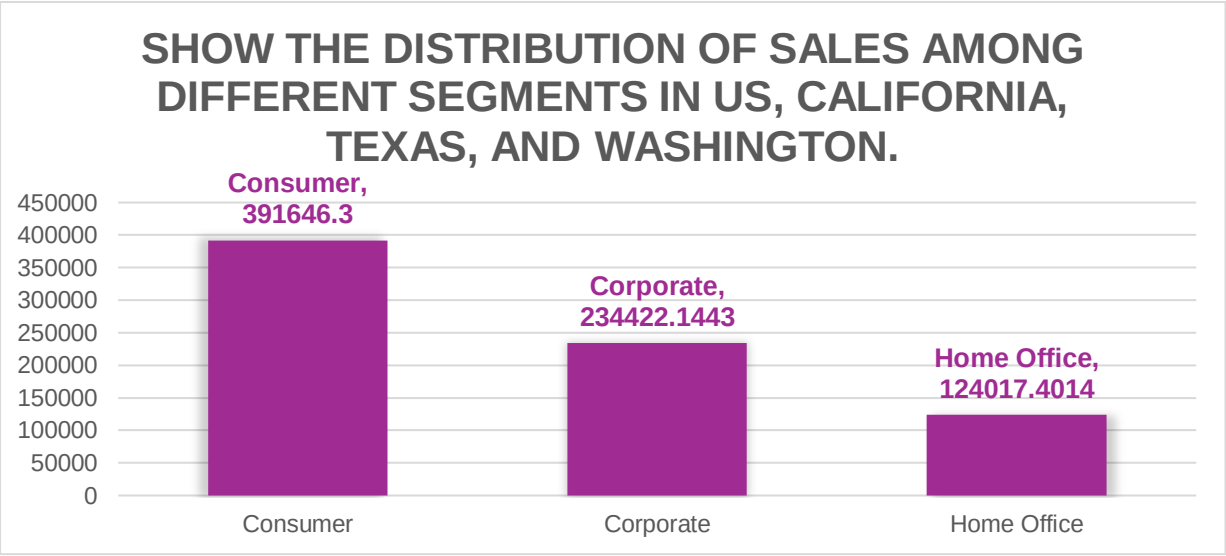
After conducting a comprehensive comparison of all states in terms of segment and sales, California emerged as the state with the highest number of sales, totaling $222,419.05. Additionally, the Consumer segment performed notably well across all states, accumulating a total sales figure of $1,148,060.531.

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



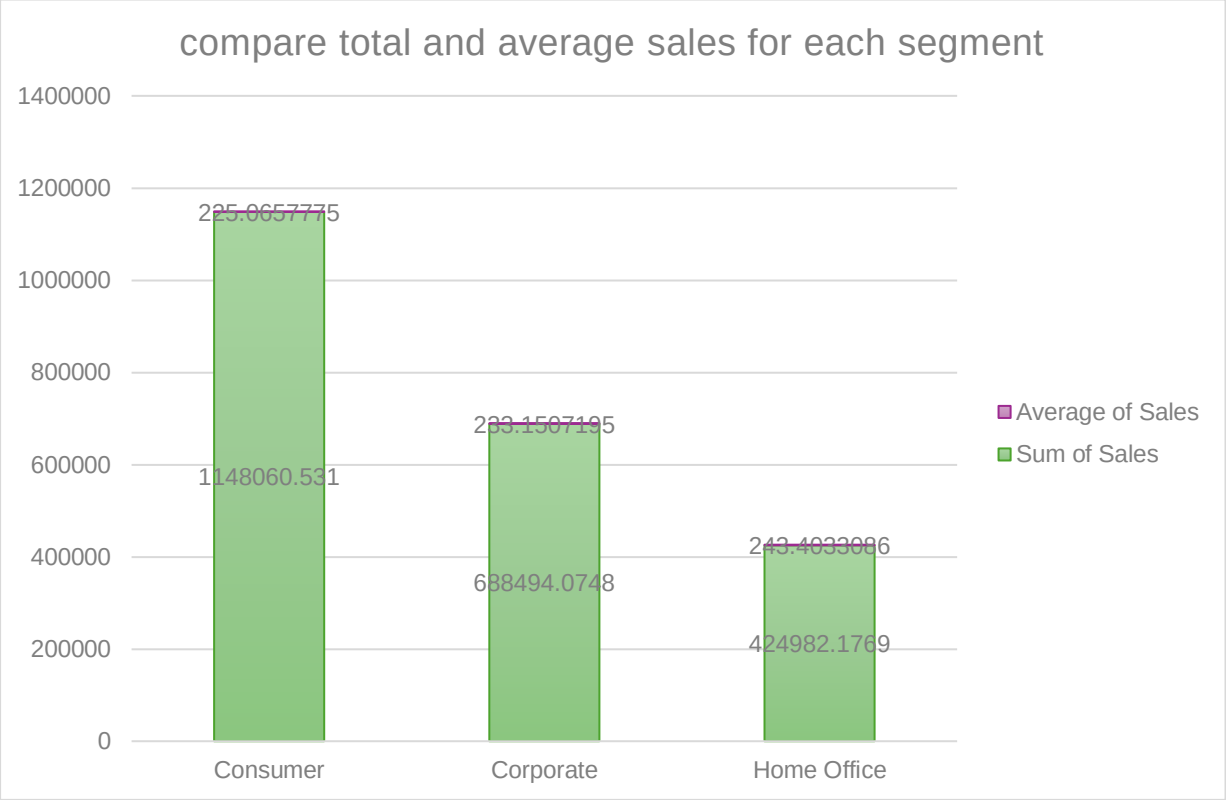
Office Supplies emerges as the top-performing category across all states, boasting a total count of sales reaching 5,909. Following closely behind, Furniture records 2,078 sales, while Technology comes in third with 1,813 sales.

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



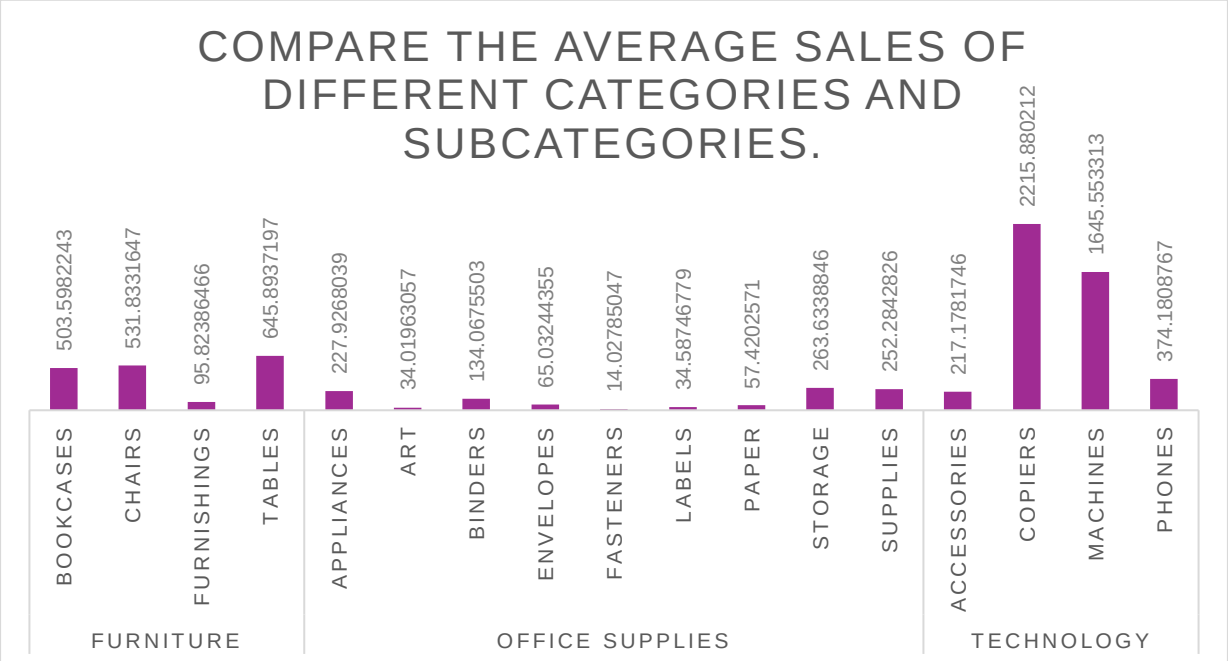
Filtering the states for the total sales count and showing the percentage of distribution through Bar Graph. The consumer segment has the most sales in US, California, Texas, and Washington.

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



The data clearly indicates that the Consumer segment boasts a significantly higher average sales figure of $1,148,060.531, while the Home Office segment registers a total sales amount of $243.40.

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



The analysis provides the average sales figures for three categories, each comprising multiple subcategories: Furniture, Office Supplies, and Technology.

Conclusion and Review

The analysis of sales data within the automotive industry unveils significant insights. California emerges as the leading state in terms of sales volume, with the Consumer segment displaying robust performance across all states. Moreover, Office Supplies emerges as the top-performing category, followed by Furniture and Technology, underscoring consumer preferences.

Consistently, the Consumer segment commands sales dominance across the US, especially in California, Texas, and Washington. Furthermore, the analysis accentuates the higher average sales of the Consumer segment relative to the Home Office segment.

Overall, these insights offer valuable guidance for optimizing sales strategies, enhancing customer engagement, and fostering business success within the automotive industry.

Regression

SUMMARY OUTPUT

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

ANOVA

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

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

In this regression analysis of the Order dataset, there appears to be little to no association between Order ID and Sales. This is evidenced by the very low multiple R and R-squared values (0.000434 and 1.88E-07, respectively). The coefficient for Order ID is not statistically significant, with a p-value of 0.965747, indicating that Order ID is not a predictor of Sales. The ANOVA test further supports this lack of significance, with 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

In the Sales dataset, the mean sales amount is 230.1162, with a standard error of 6.320053. The median sales value is 54.384, while the mode is 12.96. The standard deviation is 625.3021, indicating substantial variability in sales amounts. The data exhibits highly positive skewness, with a skewness value of 13.05363, and high kurtosis at 307.3056, suggesting a heavy-tailed distribution. The range of sales values spans from 0.444 to 22638.48, with a total sum of 2252607 across 9789 observations.

Cookie Data Report

Introduction

In our cookie dataset, we have detailed information on six types of cookies: Chocolate Chip, Fortune Cookie, Sugar, Oatmeal Raisin, Snickerdoodle, and White Chocolate Macadamia Nut. This dataset encompasses sales volumes, costs, revenue, and profits for these cookies across various countries and dates. Beyond simply analyzing cookies, this report delves into consumer preferences, pricing dynamics, and regional popularity trends. By exploring these insights, businesses can gain valuable understanding of market preferences and opportunities within the cookie industry. Get ready to uncover intriguing insights that could have significant implications for businesses 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?

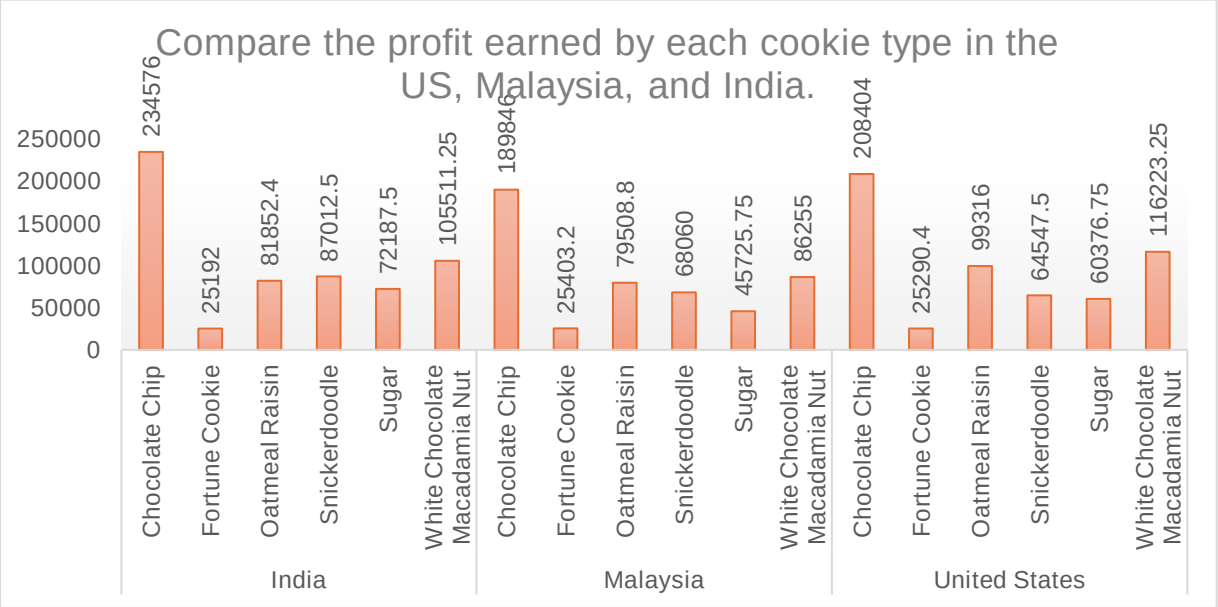
3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

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?

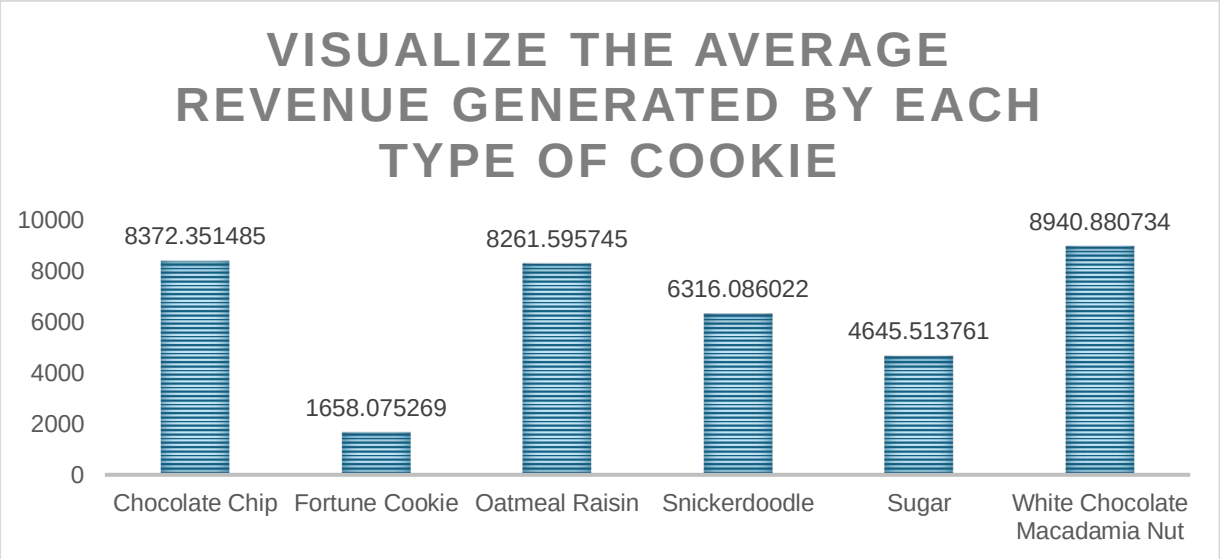
Analytics

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



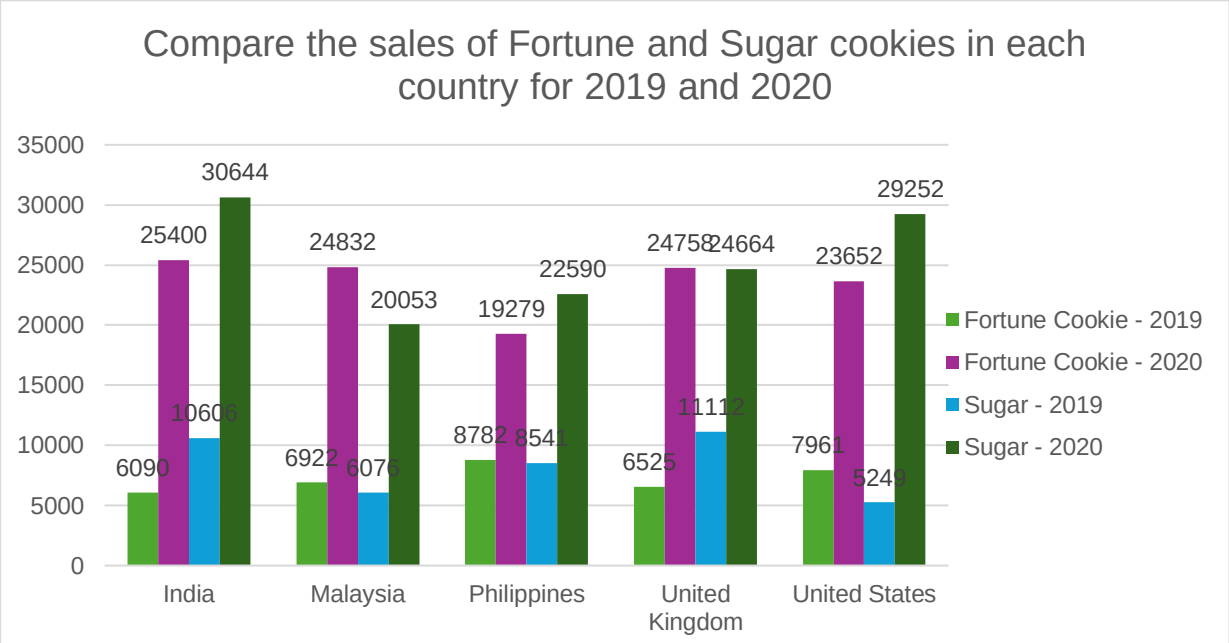
This analysis examines the profits generated by all cookie types in three different countries: the United States, Malaysia, and India. The highest profit for Chocolate Chip cookies is observed in India, followed by Malaysia and the United States.

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



This analysis aims to present the average revenue generated by each cookie type. It is evident that White Chocolate Macadamia Nut generates the highest average revenue at $8,940.88, followed by Chocolate Chip.

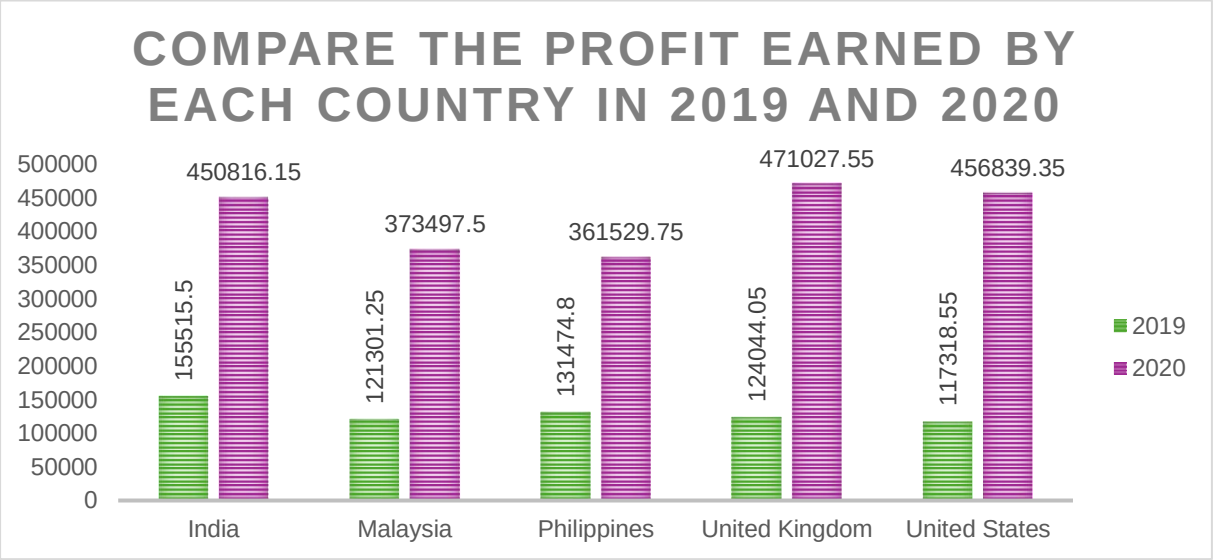
3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?



This analysis seeks to compare the sales of Fortune and Sugar cookies across different countries for the years 2019 and 2020. In 2020, India exhibits significant sales of Sugar cookies, totaling 30,644 units. Conversely, the United Kingdom led in Sugar cookie sales for 2019, followed by India.

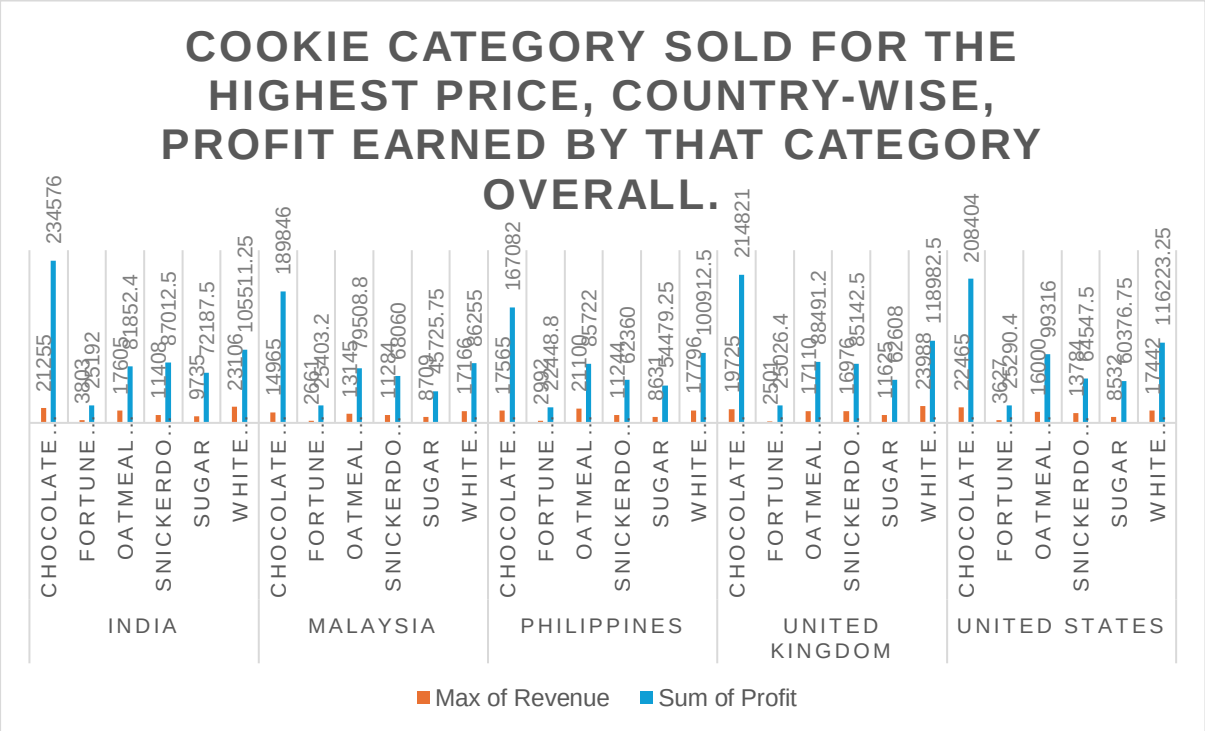
For Fortune cookies, India records the highest sales of 25,400 units, followed by Malaysia. On the other hand, the Philippines leads in Fortune cookie sales, with 8,782 units sold, followed by the United States.

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



This analysis aims to compare the profits earned by countries in the financial years 2019 and 2020. According to the graph, the United Kingdom demonstrates the highest profit earned in 2020, amounting to $471,027.55 in sales, followed closely by the United States with $456,839.35. Conversely, the highest profit in 2019 was recorded by India, totaling $155,515.5 in sales, followed by the Philippines with $131,474.8.

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



This analysis identifies the cookie category sold for the highest price in each country. Chocolate Chip cookies yield the highest revenue, and Sugar cookies generate the most profit, particularly in India followed by the United Kingdom.

Conclusion and Review

The analysis provided insights into the profit earned by different cookie types in the US, Malaysia, and India. India emerged with the highest profit for chocolate chip cookies, followed by Malaysia and the United States. White chocolate macadamia nut cookies generated the highest average revenue, closely followed by chocolate chip cookies.

In terms of sales, India showed significant sales of sugar cookies in 2020, while the United Kingdom had the highest sales of sugar cookies in 2019. For fortune cookies, India and Malaysia exhibited higher sales in both years, with the Philippines and the United States also contributing notable sales.

Regarding profit comparison by country for 2019 and 2020, the United Kingdom recorded the highest profit in 2020, followed by the United States. In 2019, India had the highest profit, followed by the Philippines.

Chocolate chip cookies were sold for the highest price in terms of revenue, while sugar cookies generated the highest profit overall.

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

Regression

SUMMARY OUTPUT

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

ANOVA

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

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.3E-11	7.3E-13	-18.0657	4.09E-60	-1.5E-11	-1.2E-11	-1.5E-11	-1.2E-11

X Variable 1	6.56E-17	8.42E-16	0.077892	0.937936	-1.6E-15	1.72E-15	-1.6E-15	1.72E-15
X Variable 2	1	8.38E-16	1.19E+15	0	1	1	1	1
X Variable 3	-1	1.72E-15	-5.8E+14	0	-1	-1	-1	-1

In the regression analysis of the Cookie dataset, the results indicate a perfect linear relationship between the independent and dependent variables. The multiple R value is 1, suggesting a perfect correlation. Both the R-squared and adjusted R-squared values are also 1, indicating that the independent variables explain all the variability in the dependent variable. The standard error is exceptionally small (9.16E-12), indicating precise estimates.

The ANOVA results confirm that the regression model is highly significant ($p < 0.05$), with an F-statistic of 1.9E+31. Despite the model's overall significance, the coefficients for the independent variables (X Variable 1, X Variable 2, X Variable 3) are all very close to 0, suggesting no meaningful effect on the dependent variable. The p-values associated with these coefficients are all greater than 0.05, further supporting the lack of statistical significance.

In summary, while the regression model itself is highly significant due to the perfect fit, the independent variables do not have a significant effect on the dependent variable, as indicated by their coefficients and associated p-values.

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 single-factor ANOVA analysis compares the variance between two groups: Cost and Profit. The Cost group consists of 700 observations, with a total sum of 1,926,955 and an average of 2,752.79. The Profit group also comprises 700 observations, with a total sum of 2,763,364 and an average of 3,947.66.

The ANOVA results indicate a significant difference between the means of the Cost and Profit groups ($F = 90.92153$, $p < 0.05$). This suggests that there is a statistically significant variation in the average values of Cost and Profit. The p-value (6.36E-21) is much smaller than the significance level ($\alpha = 0.05$), providing strong evidence against the null hypothesis. Therefore, we reject the null hypothesis and conclude that there is a significant difference in the mean values of Cost and Profit.

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 two-factor ANOVA without replication assesses the effects of two categorical independent variables, Revenue and Cost, on the dependent variable, Profit. The table provides a summary of the data for Revenue, Cost, and Profit, indicating the count, sum, average, and variance for each factor level.

The ANOVA results reveal significant main effects for both Revenue ($F = 14.75112$, $p < 0.05$) and Cost ($F = 1484.458$, $p < 0.05$), as well as a significant interaction effect between Revenue and Cost ($MS = 28507277$, $p < 0.05$). The p-values for all factors are less than the significance level ($\alpha = 0.05$), indicating strong evidence against the null hypothesis. Therefore, we reject the null hypothesis and conclude that both Revenue and Cost have a significant impact on Profit, and there is also a significant interaction effect between Revenue and Cost.

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 descriptive statistics offer valuable insights into the distribution and characteristics of the variables Unit Sold, Revenue, Cost, and Profit.

For Unit Sold, the mean value is 1608.32 units, with a standard error of 32.79 units. The median value of 1542.5 units provides a measure of central tendency, while the mode of 727 units represents the most frequently occurring value. The standard deviation, skewness, and kurtosis values provide information about the dispersion, symmetry, and shape of the distribution, respectively.

Similarly, for Revenue, Cost, and Profit, the descriptive statistics provide measures of central tendency, variability, and distributional characteristics. These statistics offer valuable insights into the distribution and variability of the variables, which in turn aid in better understanding their underlying characteristics and inform further analysis.

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 correlation matrix provides a detailed view of the relationships between Unit Sold, Revenue, Cost, and Profit. A correlation coefficient near 1 indicates a strong positive relationship, while a value near -1 indicates a strong negative relationship. For Unit Sold and Revenue, the correlation coefficient is approximately 0.796, indicating a moderately strong positive correlation. Similarly, Unit Sold and Profit have a correlation coefficient of about 0.829, showing a moderately strong positive relationship. Revenue and Cost are highly correlated with a coefficient of around 0.992, indicating a strong positive relationship. Additionally, Revenue and Profit have a correlation coefficient of approximately 0.995, signifying a very strong positive relationship. Cost and Profit also exhibit a strong positive correlation with a coefficient of approximately 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

The loan dataset contains detailed information on loan applicants, including factors like gender, marital status, education level, income, loan amount, and property location. This dataset provides valuable insights into the dynamics of loan requests.

In this examination, our objective is to investigate the traits of loan applicants and identify trends within the data. By utilizing pivot tables and visualizations, we aim to answer specific questions regarding the demographics of loan applicants, their educational backgrounds, and the amounts they borrow.

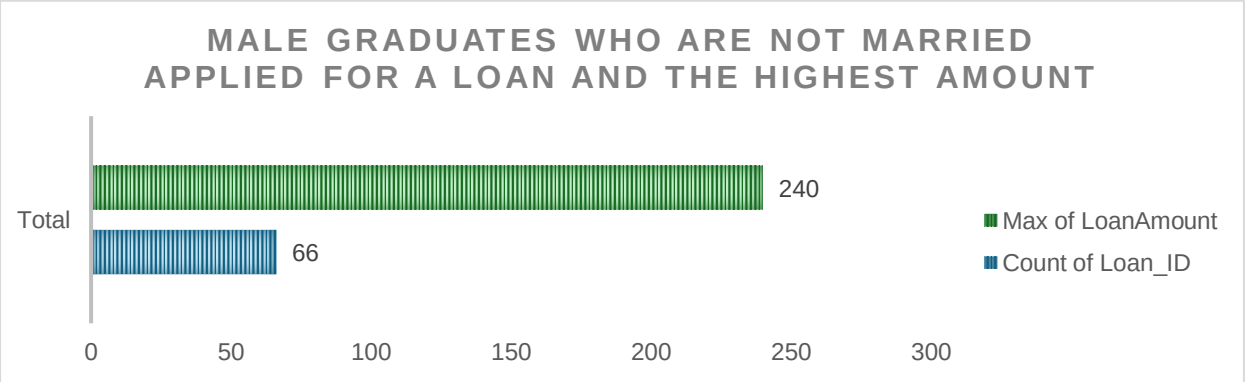
Comprehending the intricacies of loan requests is essential for financial institutions to make well-informed choices, streamline lending procedures, and customize services to suit the varied needs of clients. Through this analysis, we aspire to uncover practical insights that can inform strategic decision-making and improve the effectiveness of loan management systems.

Questionnaire

1. How many male graduates who are not married applied for Loan? What was the highest amount?
2. How many female graduates who are not married applied for Loan? What was the highest amount?
3. How many male non-graduates who are not married applied for Loan? What was the highest amount?
4. How many female graduates who are married applied for Loan? What was the highest amount?
5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rural based on amount.

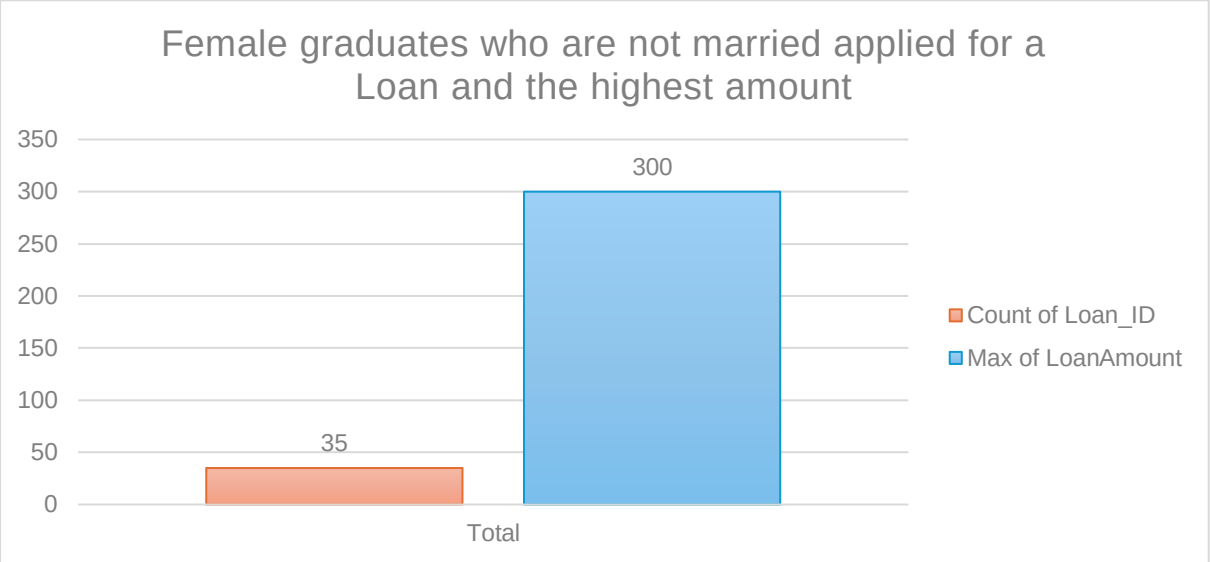
Analytics

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



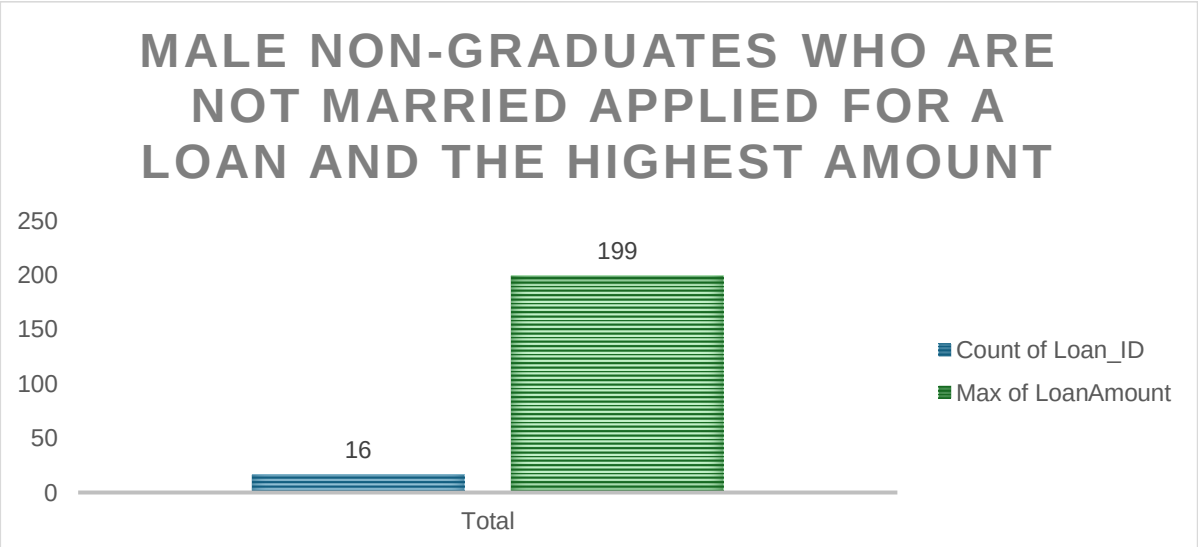
The analysis reveals that among the loan applicants, the highest number consists of unmarried male graduates, and they have applied for the highest loan amount. Specifically, out of the total 66 loan applications analyzed, the maximum loan amount requested is 240.

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



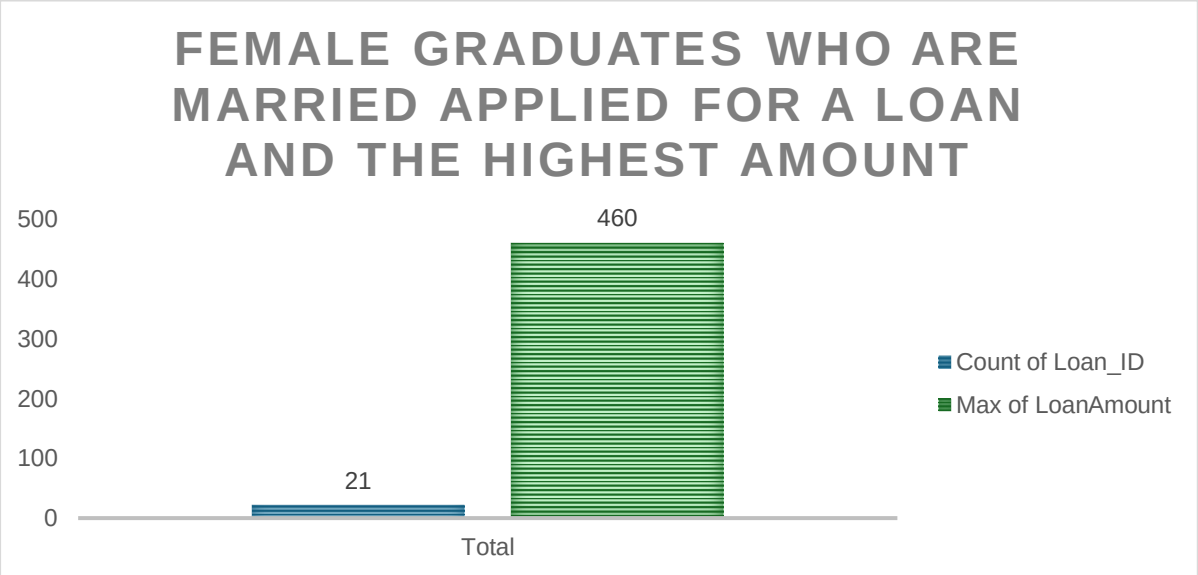
According to the analysis, the highest number of loan applicants among females are unmarried graduates, and they have applied for the highest loan amount. Specifically, out of the total 35 loan applications examined, the maximum loan amount requested is 300.

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



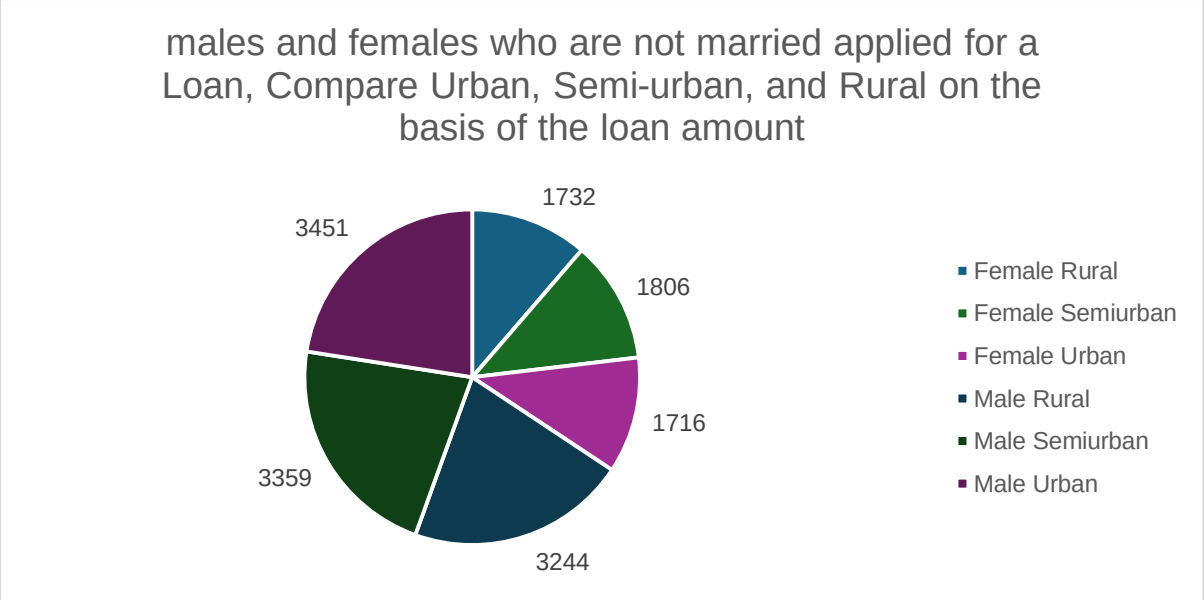
The analysis indicates that the highest number of loan applicants among males are non-graduates who are not married, and they have applied for the highest loan amount. Specifically, out of the total 16 loan applications reviewed, the maximum loan amount requested is 199.

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



The analysis indicates that among the loan applicants, the highest number comprises unmarried female graduates, and they have applied for the highest loan amount. Specifically, out of the total 21 loan applications analyzed, the maximum loan amount requested is 460.

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



This analysis seeks to contrast the number of loan applicants who are unmarried, categorized by rural, semi-urban, and urban areas, across both genders. While the count of female applicants is lower, it is significantly higher among male applicants. Specifically, the count of female loan applicants in rural areas is 1732, in semi-urban areas is 1806, and in urban areas is 1716. In contrast, the count of male loan applicants in rural areas is 3244, in semi-urban areas is 3359, and in urban areas is 3451.

Conclusion and Review

The analysis underscores pronounced gender gaps in loan applications. Unmarried male graduates emerged as the primary applicants, closely followed by unmarried female graduates. While both unmarried male non-graduates and married female graduates also sought loans, their numbers were comparatively smaller. Importantly, males outnumbered females significantly across rural, semi-urban, and urban regions.

This analysis effectively delineates gender-based patterns in loan requests, offering valuable insights into borrower demographics. It suggests further exploration into factors influencing loan decisions, alongside visual enhancements to enhance data presentation. Ultimately, the report establishes a groundwork for comprehending loan dynamics, with prospects for deeper insights.

Regression

SUMMARY OUTPUT

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

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

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

The regression analysis of the loan dataset reveals several key findings. The multiple R coefficient is approximately 0.531, indicating a moderate positive relationship between the predictors and the loan amount. The R-squared value of around 0.282 suggests that approximately 28.2% of the variability in the loan amount can be explained by the independent variables.

The coefficients for the predictors are as follows: the coefficient for Applicant Income is approximately 0.096, and for Co-applicant Income, it's about 0.0068. These coefficients indicate the impact of each predictor on the loan amount.

The ANOVA table shows a significant F-value of 37.32 ($p < 0.05$), confirming the statistical significance of the regression model. This implies that the model as a whole explains a significant amount of the variance in the loan amount.

In summary, this analysis provides insights into how applicant and co-applicant incomes influence the loan amount, thereby aiding in a better understanding of the loan approval process.

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				

In the ANOVA table for the single-factor analysis of the loan dataset, the data is segmented into two groups based on the factors Loan Amount and Loan Amount Term. The total sum of squares (SS) is approximately 8392703, with 2267909 within-group SS and 6124794 between-group SS. This results in a mean square (MS) of 3937.343 within groups and 6124794 between groups.

The F-value of 1555.565 and the associated p-value of approximately 8.4E-166 indicate that there is a significant difference between the means of the two groups. This means that the factor being considered (Loan Amount vs. Loan Amount Term) has a substantial impact on the loan dataset. The high F-value and very low p-value suggest strong evidence against the null hypothesis, supporting the conclusion that the difference in means is not due to random chance. Thus, the factor chosen significantly affects the loan dataset.

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				

In the two-factor ANOVA without replication, the loan dataset is analyzed based on Loan Amount and Loan Amount Term. The total sum of squares (SS) is approximately 8392703, with 1264619 SS for Loan Amount Term (rows), 6124794 SS for Loan Amount (columns), and 1003290 SS for error.

The mean square (MS) for Loan Amount Term is 4391.038, and for Loan Amount is 6124794. The F-value for Loan Amount Term is 1.260472, and for Loan Amount is 1758.156, both with associated 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 correlation matrix for the loan dataset variables shows the relationships between Applicant Income, Co-Applicant Income, and Loan Amount. There is a weak negative correlation of approximately -0.084 between Applicant Income and Co-Applicant Income. Applicant Income and Loan Amount exhibit a moderate positive correlation of approximately 0.446. Similarly, Co-Applicant Income and Loan Amount show a weak positive correlation of approximately 0.230. These correlation coefficients provide insights into the direction and strength of the relationships between the variables, which are crucial for understanding their potential impacts on loan outcomes.

Shop Sales Data Report

Introduction

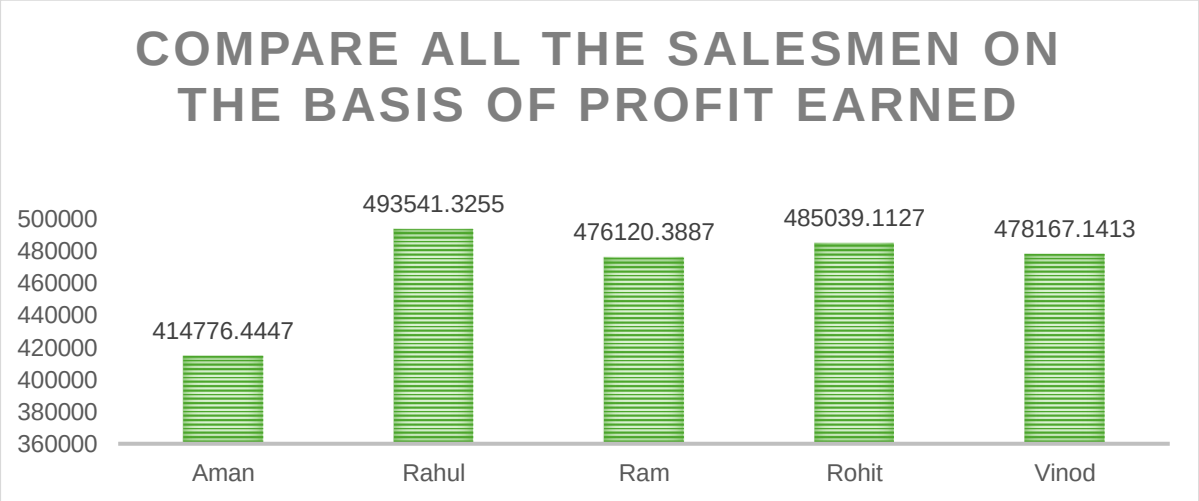
This report explores an extensive sales dataset, focusing on the examination of sales performance and product trends among sales representatives. The dataset includes various attributes such as sales personnel details, product specifications, sales volumes, and generated profits. The primary aim of this analysis is to unveil valuable insights that can guide the formulation of sales strategies and improve overall business performance. By scrutinizing sales data across a defined timeframe and comparing product performance, the report endeavors to identify top-performing sales representatives, evaluate product popularity, and grasp sales trends. The insights gained from this analysis will be highly beneficial for sales managers, marketing experts, and executives aiming to refine sales strategies, optimize revenue, and foster business expansion. Through this examination, our goal is to present actionable insights that can facilitate decision-making and contribute to the overall success of the business.

Questionaries

1. Compare all the salesmen based on profit earn.
2. Find out most sold product over the period of May-September.
3. Find out which of the two product sold the most over the year Computer or Laptop?
4. Which item yield most average profit?
5. Find out average sales of all the products and compare them.

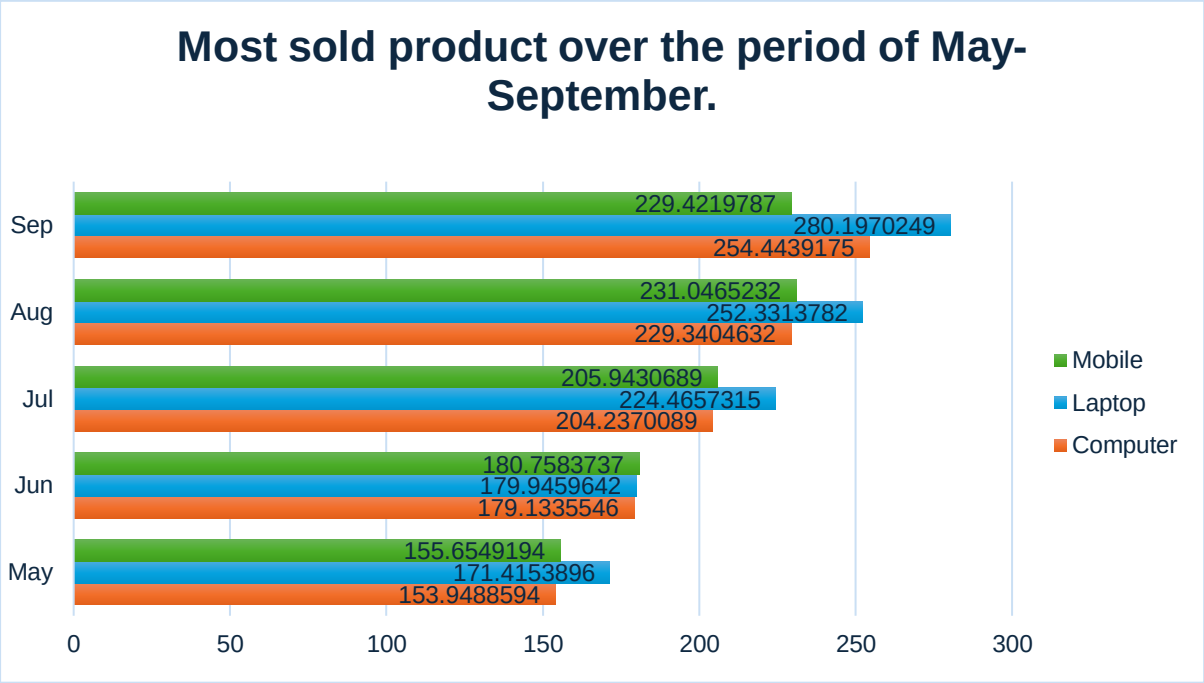
Analytics

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



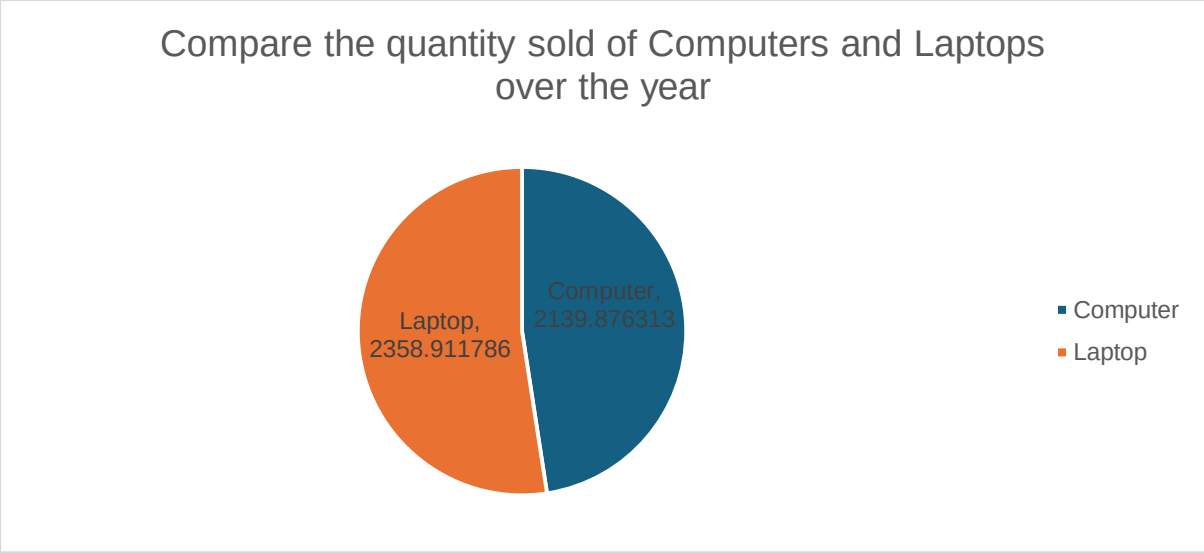
The comparison of all the salesmen on the basis of profit earned and the bar graph shows that the rahul has the highest profit earned with value 493541.3255, compared to all the salesmen.

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



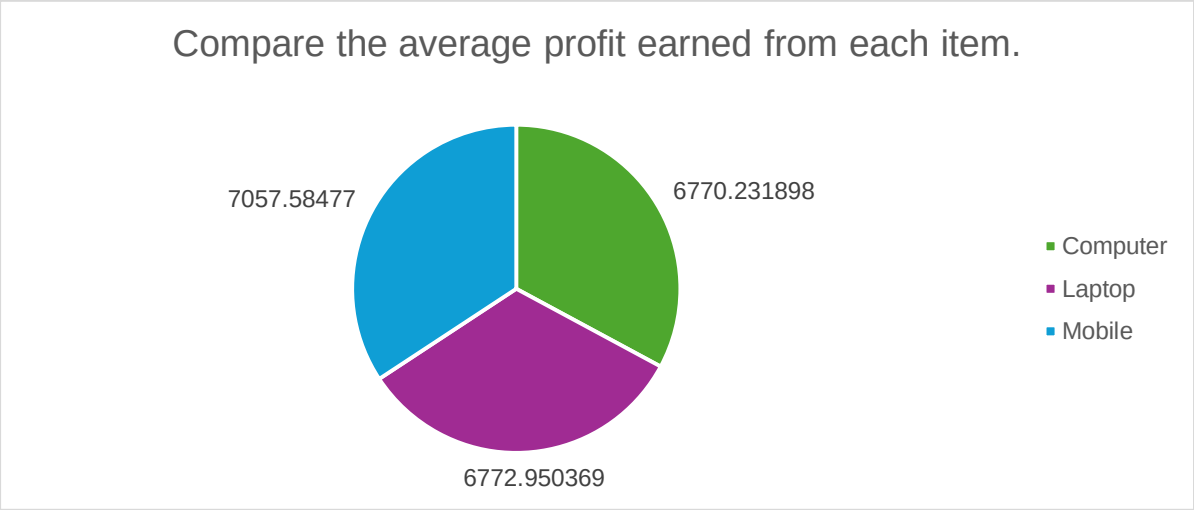
To determine the highest-selling product between May and September, we must analyze the sales data within this timeframe. By summing up the quantities sold for each product across all transactions during these months, we find that the top-selling product is the Laptop, with the highest sales occurring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

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



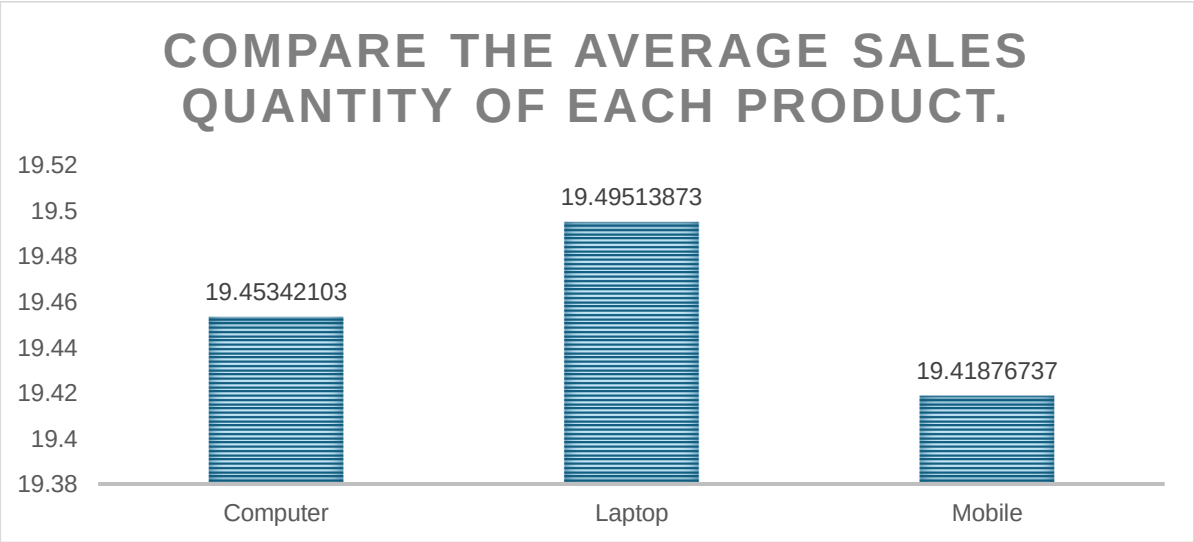
The two products that sold the most over the year were computers and laptops. Computers had a total sold quantity of 2139.876313 units, while laptops had a higher sold quantity of 2358.911786 units.

4 . Which item yield most average profit?



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

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



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

This report examines a detailed sales dataset containing various attributes like ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and SALES, with the goal of deriving insights to steer sales strategies and bolster business efficacy. It targets sales managers, marketers, and executives aiming to refine sales operations and amplify revenue generation. Key analyses involve juxtaposing sales figures of Vintage cars and Classic cars, calculating average sales, pinpointing top-selling items, assessing country-specific profits for particular product lines, comparing sales trends across different years, and evaluating countries based on transaction size. By conducting these analyses, the report seeks to furnish actionable insights to propel sales expansion and enhance overall business outcomes.

Questionnaire

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

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

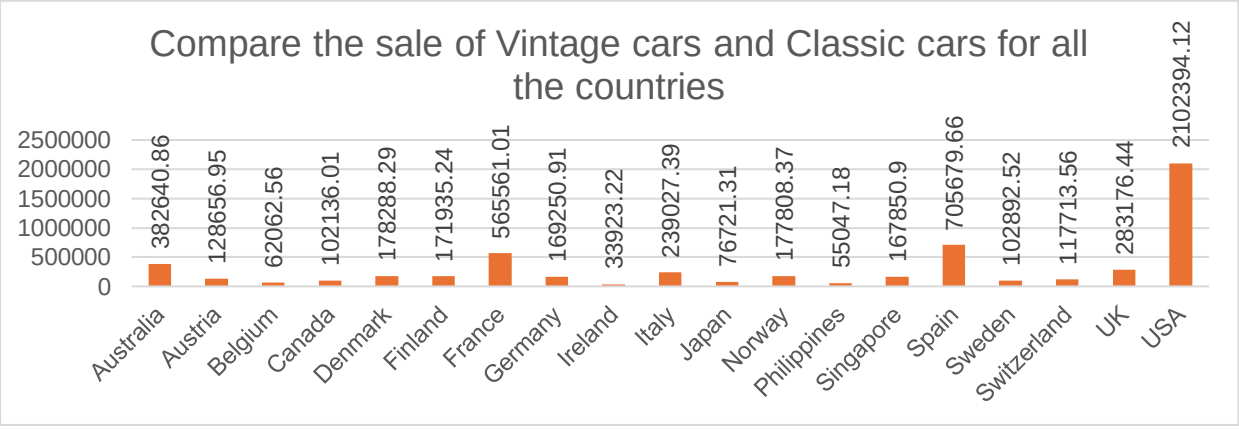
3. Assessment of the country yielding the most profit for Motorcycles, Trucks, and Buses.

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

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

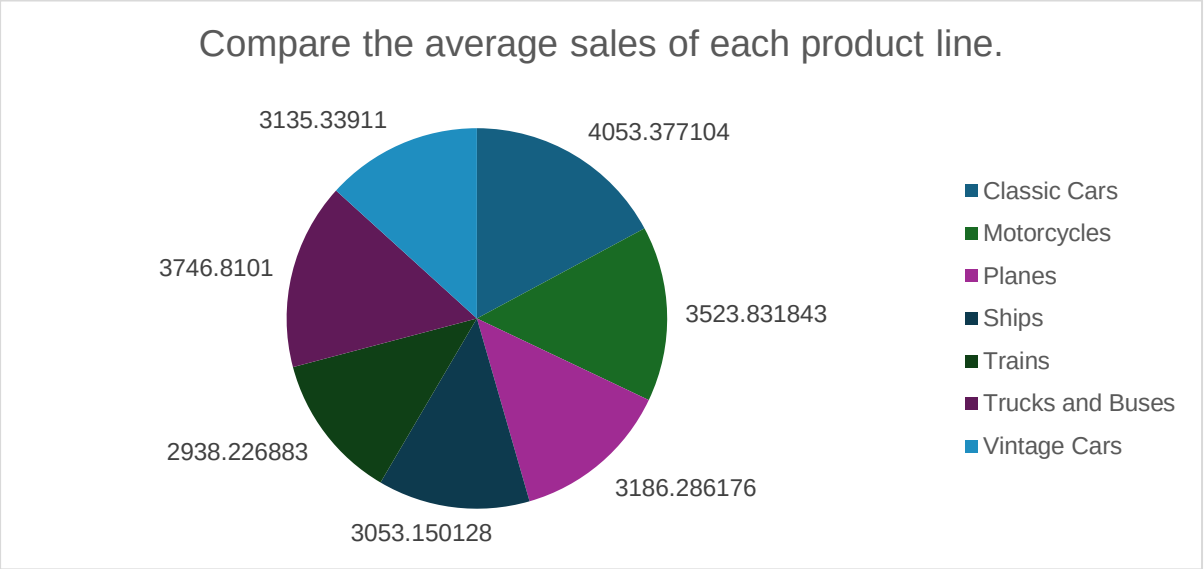
Analytics

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



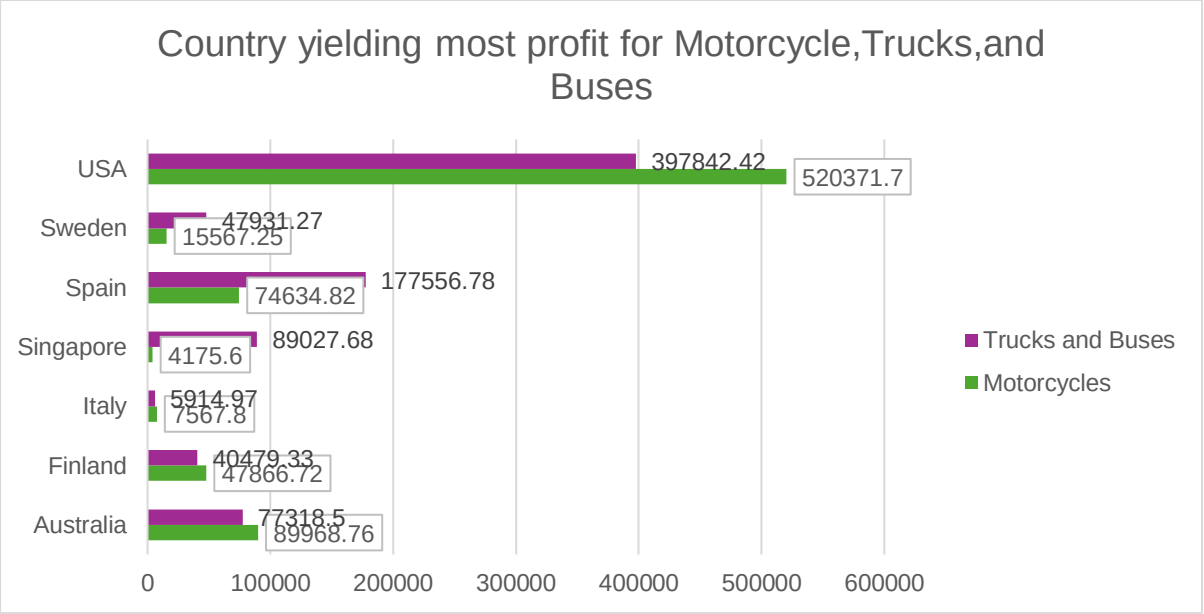
This analysis Compare the sale of Vintage cars and Classic cars for all the countries. Where USA(2102394.02) has the highest sales followed by Spain, France, and Australia. This is represented by using line graph.

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



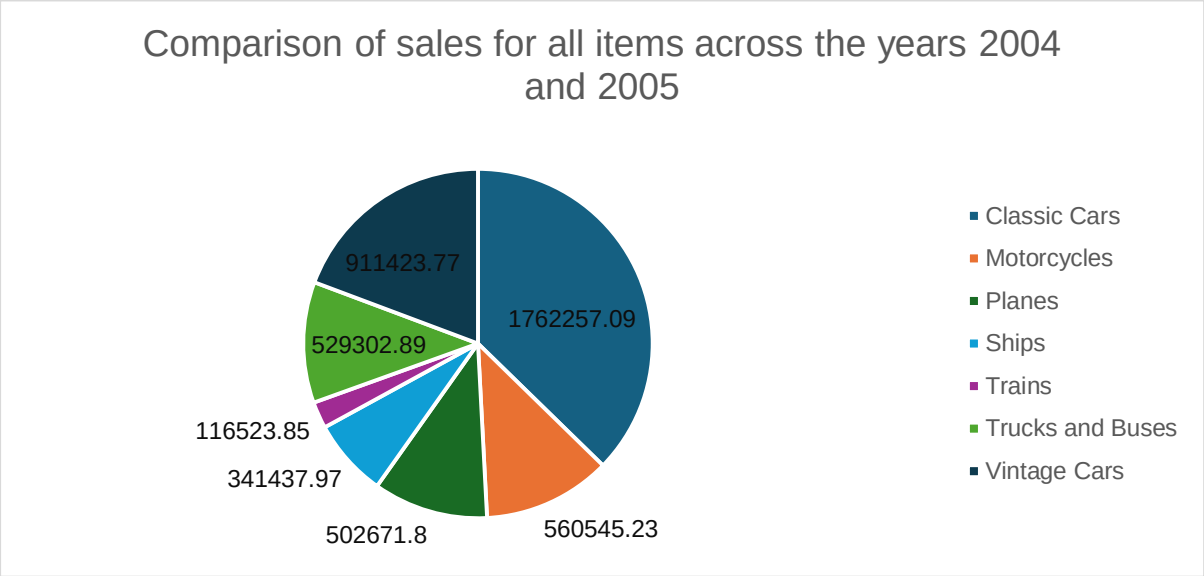
This analysis seeks to present the average sales figures for all products and pinpoint the highest-selling product. The graphical representation highlights that Classic Cars lead the sales, boasting an average of 4053.377104 units sold, followed by Trucks and Buses, and Motorcycles.

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



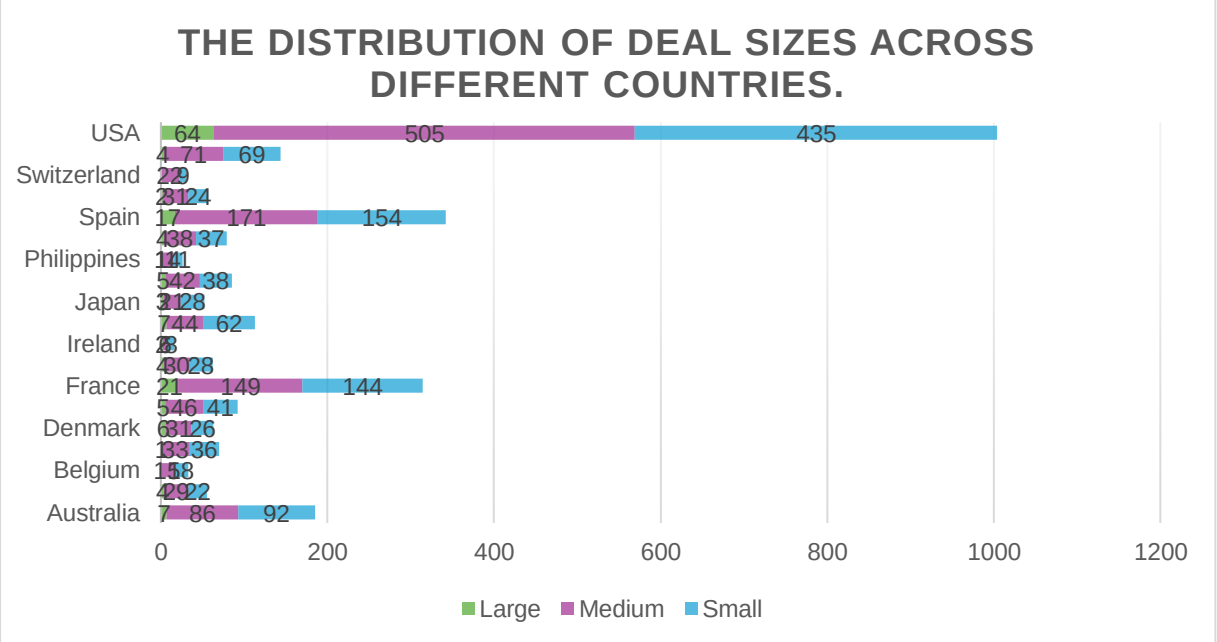
The objective of this analysis is to determine the country generating the highest profit for Motorcycles, Trucks, and Buses. The bar chart illustrates that the USA leads with the highest sales, totaling $397,842.42 for Trucks and Buses, and $520,371.70 for Motorcycles, followed by France and Spain in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

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



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.

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



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

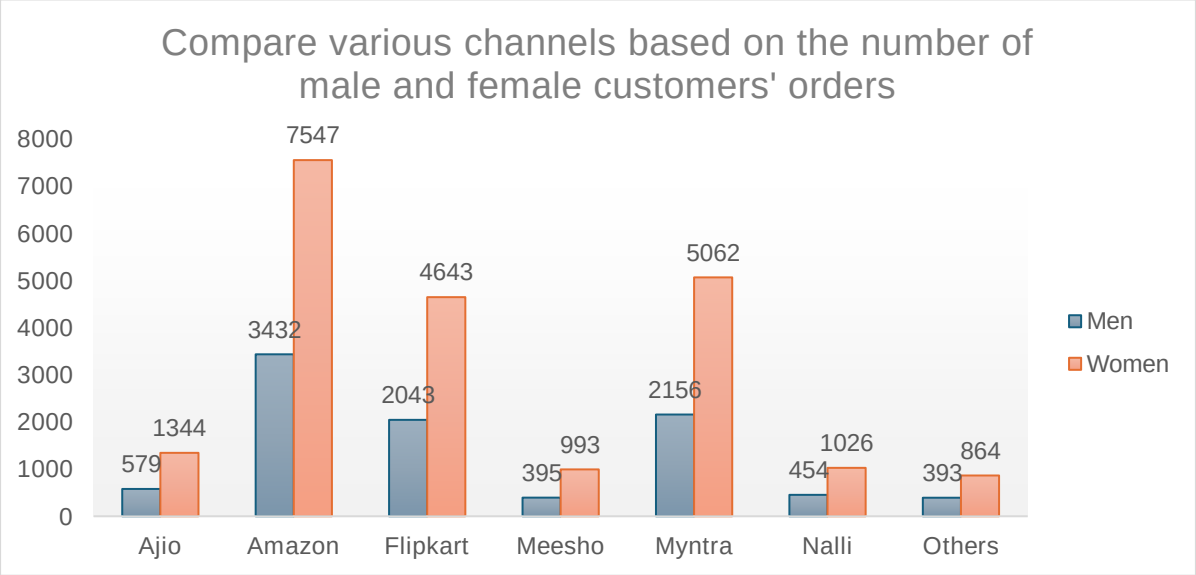
This dataset contains sales information from a retail outlet, including diverse details like customer demographics (Gender, Age Group), transaction specifics (Order ID, Status), product details (Category, SKU), and shipping data. Our examination focuses on understanding customer actions and product patterns, aiming to reveal trends, preferences, and associations present in the dataset. Leveraging these insights, companies can enhance marketing approaches, optimize inventory control, and boost overall customer contentment.

Questionnaire

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

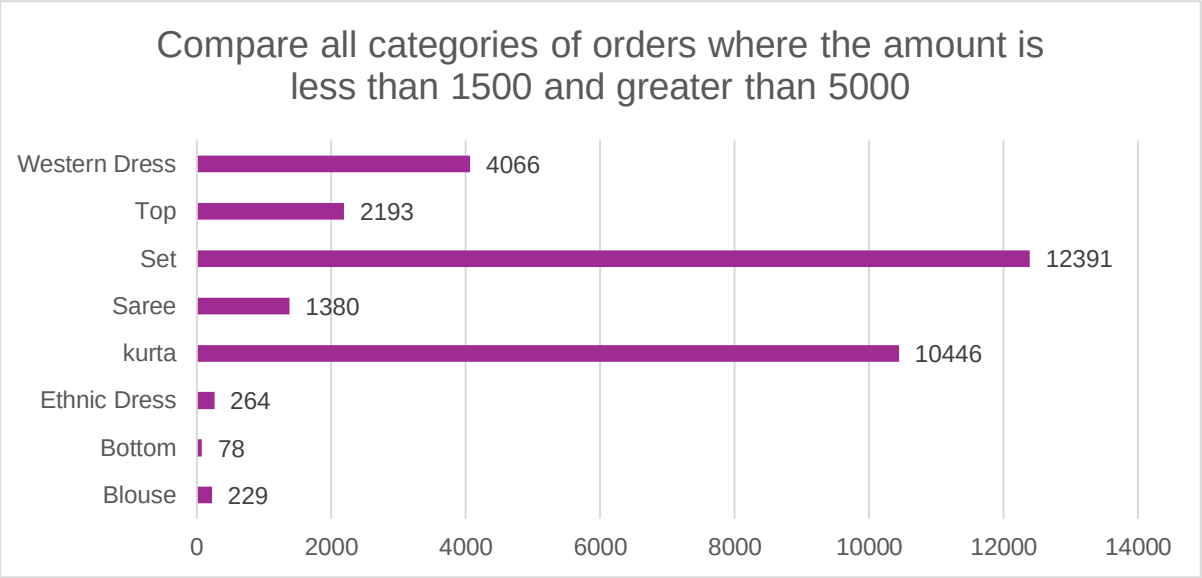
Analytics

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



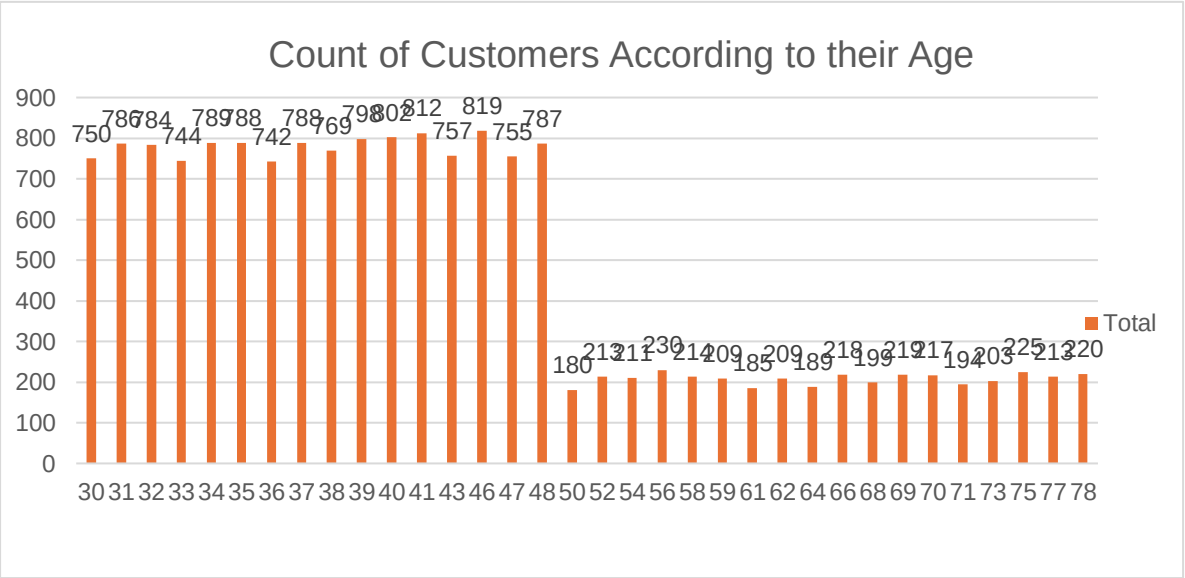
Amazon dominates sales in both the men's and women's categories, followed closely by Myntra and Flipkart. Specifically, Amazon sold approximately 3432 units in the men's category and nearly 7547 units in the women's category. In comparison, Myntra recorded sales of 2156 units in the men's section and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

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



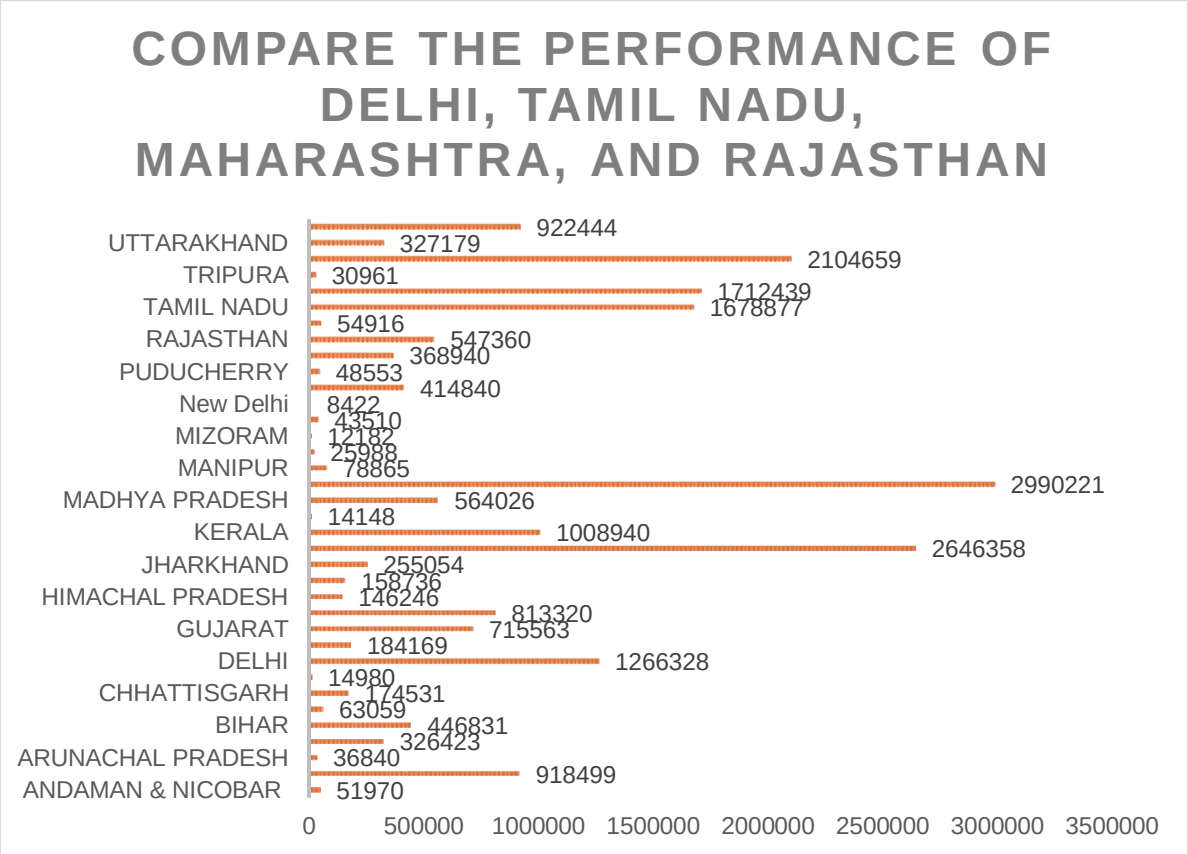
This analysis facilitates the comparison of order categories based on their amounts, specifically focusing on orders with amounts less than 1500 and greater than 5000. It reveals that Kurta and Set have the highest count of orders, with 12,391 and 10,446 respectively, followed by Western Dress, Top,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

3. How many Customers are there whose age is 30 and above and state is Delhi.



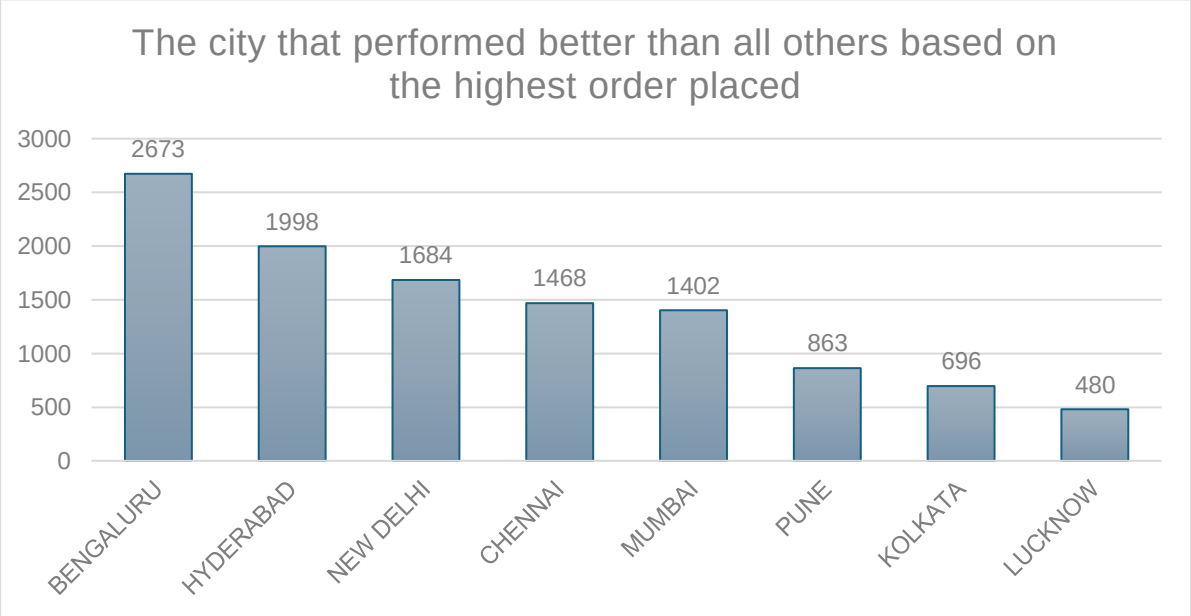
This analysis facilitates the comparison of count of Customers based on age specifically focused on the age and the state is Delhi. The highest value of count is in 46 age group and the lowest value of count is in 50 age group, showing that most customers belong to 46 age group.

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



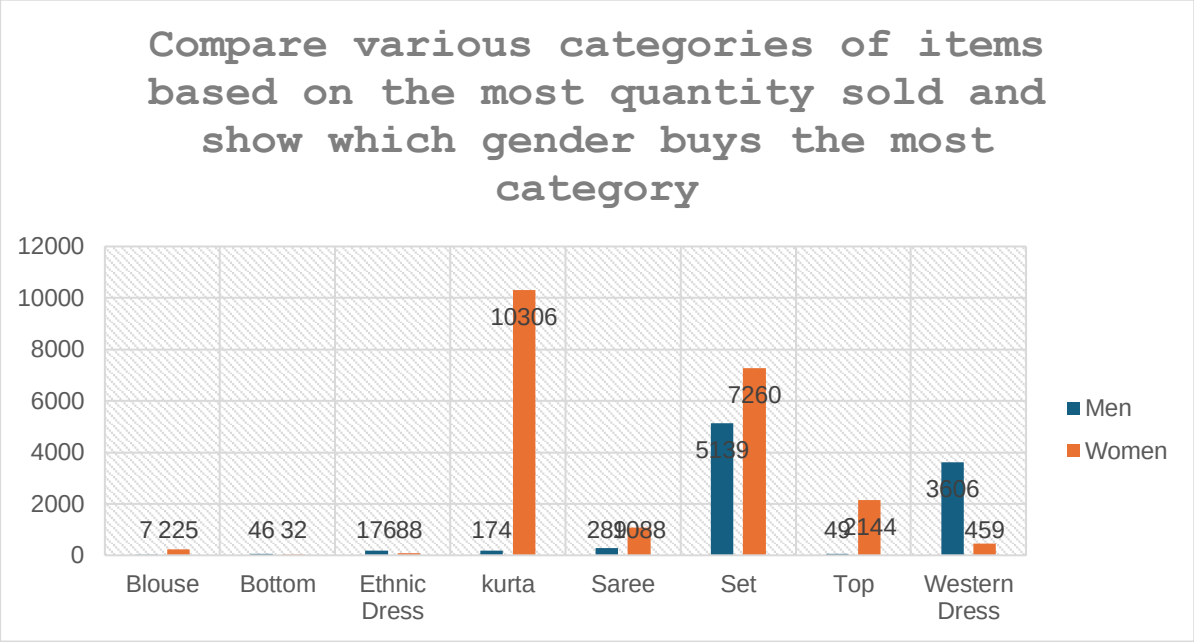
This analysis highlights the states that outperformed those mentioned previously, with Karnataka leading with the highest performance, recording sales of $2,646,358, followed by Uttar Pradesh, which recorded sales of $2,104,659.

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



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.

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



This analysis compares various categories of items based on the 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.

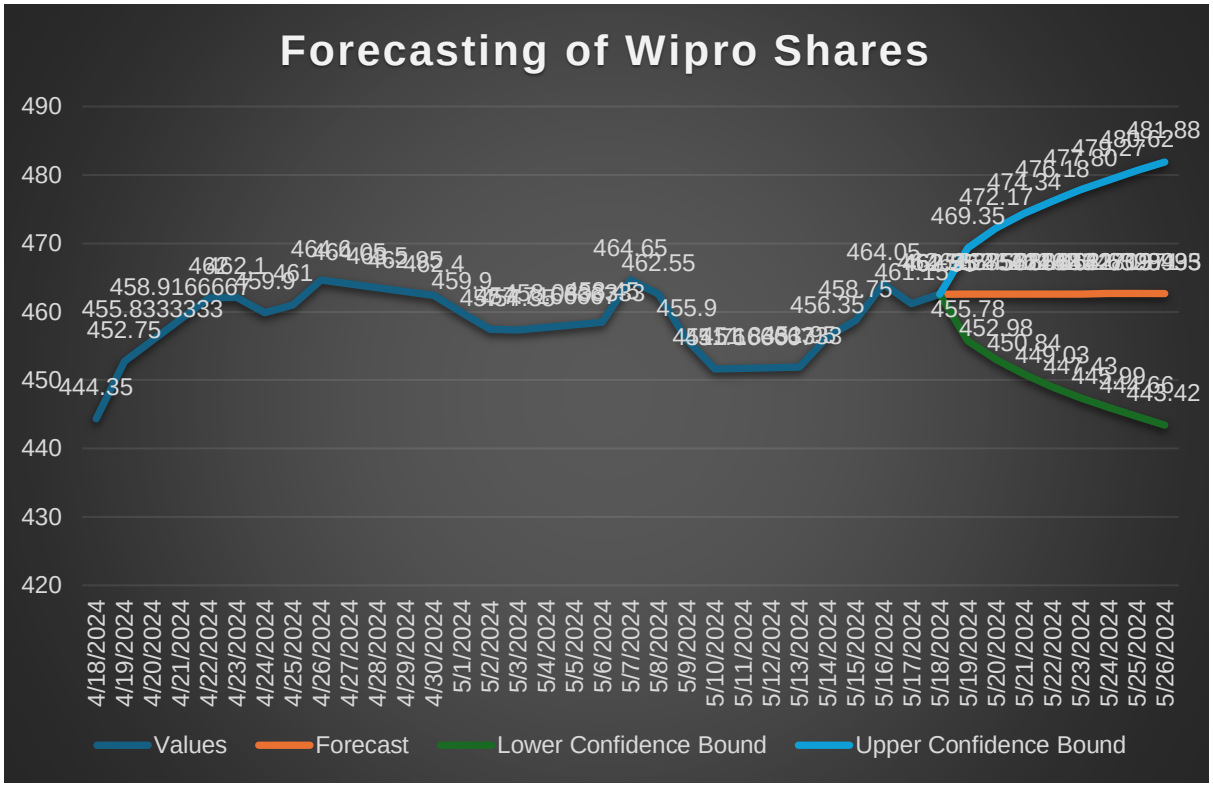
Forecasting of Wipro Shares

The shares dataset is showing the share price of Wipro shares from 18 April 2024 to 18 May 2024.It also helps in understanding the lower bound and upper bound.

Timeline	Values	Forecast	Lower Bound	Confidence	Upper Confidence Bound
18-04-2024	444.35				
19-04-2024	452.75				
20-04-2024	455.8333				
21-04-2024	458.9167				
22-04-2024	462				
23-04-2024	462.1				
24-04-2024	459.9				
25-04-2024	461				
26-04-2024	464.6				
27-04-2024	464.05				
28-04-2024	463.5				
29-04-2024	462.95				
30-04-2024	462.4				
01-05-2024	459.9				
02-05-2024	457.4				
03-05-2024	457.35				
04-05-2024	457.7167				
05-05-2024	458.0833				
06-05-2024	458.45				
07-05-2024	464.65				
08-05-2024	462.55				
09-05-2024	455.9				
10-05-2024	451.6				

11-05-2024	451.7167			
12-05-2024	451.8333			
13-05-2024	451.95			
14-05-2024	456.35			
15-05-2024	458.75			
16-05-2024	464.05			
17-05-2024	461.15			
18-05-2024	462.55	462.55	462.55	462.55
19-05-2024		462.56286	455.78	469.35
20-05-2024		462.57571	452.98	472.17
21-05-2024		462.58857	450.84	474.34
22-05-2024		462.60142	449.03	476.18
23-05-2024		462.61428	447.43	477.80
24-05-2024		462.62714	445.99	479.27
25-05-2024		462.63999	444.66	480.62
26-05-2024		462.65285	443.42	481.88

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.