Car Collection Data Report

Introduction

The Car Collection dataset provides detailed information on various car models, including their make, model, color, mileage, price, and cost. This report analyzes the dataset to derive insights for better decision-making in car purchases and understanding market trends. The dataset includes six different car models: Honda, Chevrolet, Nissan, Toyota, Dodge, and Ford.

This report targets car enthusiasts, industry professionals, analysts, and those interested in car market trends. It includes detailed dataset analysis, statistical evaluations, visualizations, and interpretation of results.

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

Questionnaire

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

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

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

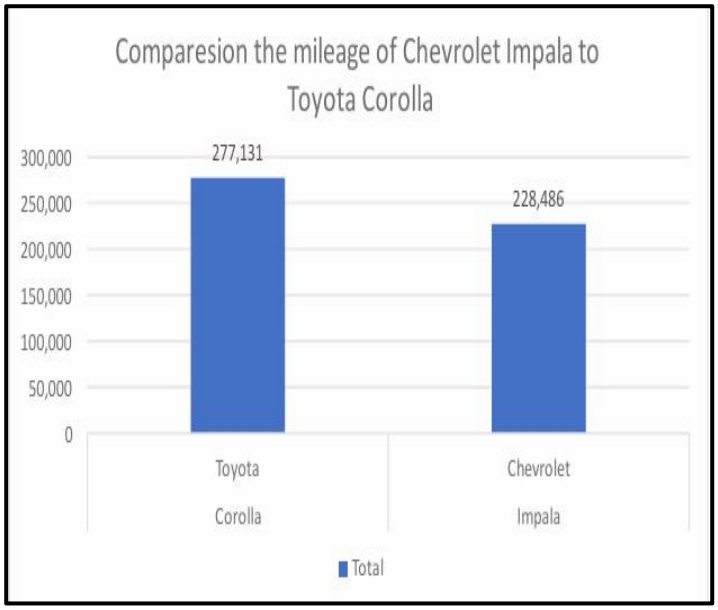
4. Compare all the cars which are of silver color to the green color in terms of Mileage.

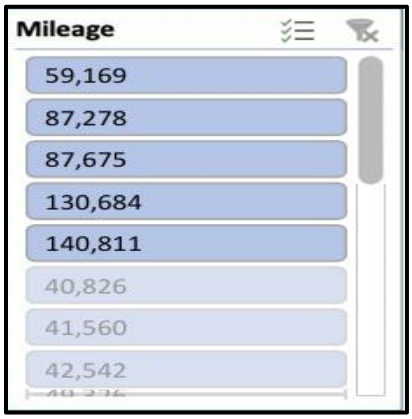
5. Find out all the cars, and their total cost which is more than $2000?

Analytics

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

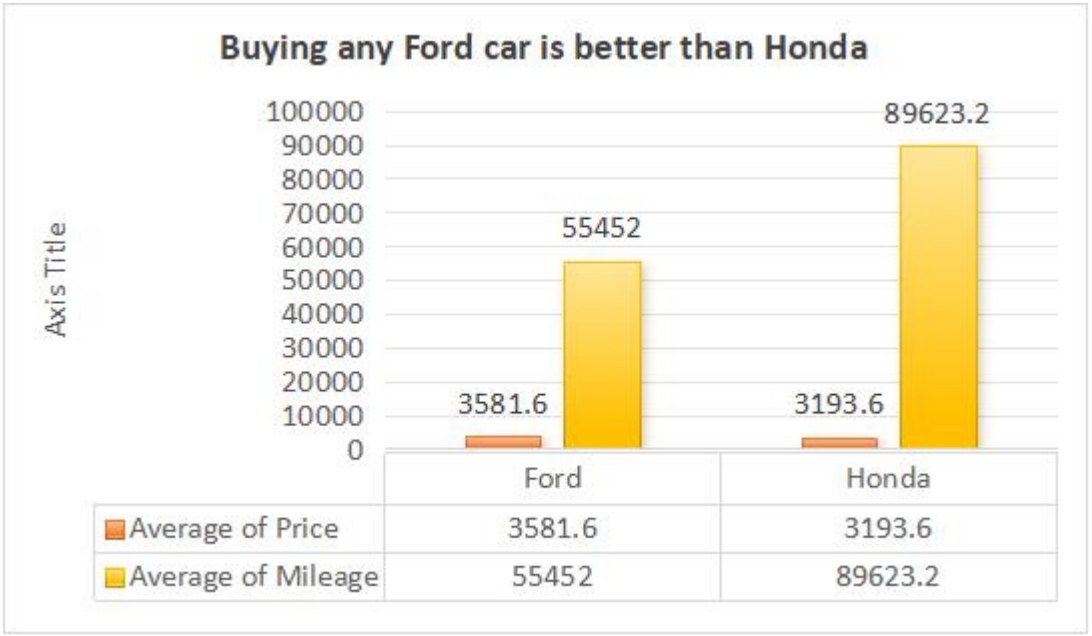
ANS: This analysis examines the fuel efficiency (mileage) of two popular car models, the Chevrolet Impala and the Toyota Corolla. To perform this comparison, the dataset was filtered to isolate relevant data, and a column chart was created to visualize the findings. Based on this analysis, it was concluded that the Chevrolet Impala, with a mileage of 114,243 miles, offers better fuel efficiency compared to the Toyota Corolla, which has a mileage of 92,377 miles. This insight helps potential buyers and industry analysts make informed decisions regarding these two car models' fuel performance.

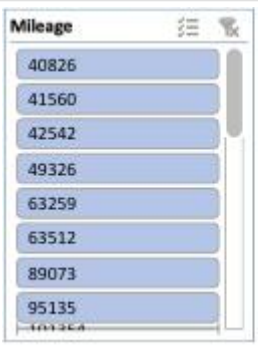


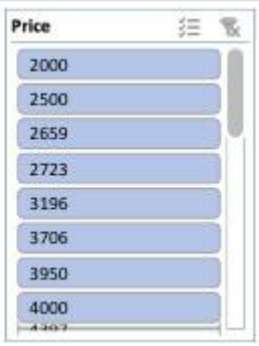


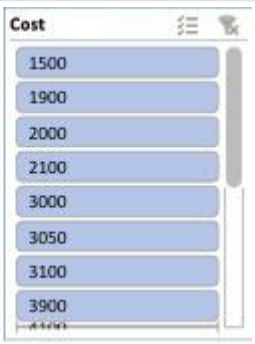
Q2. Justify, Buying of any Ford car is better than Honda.

ANS: This analysis compares Ford and Honda cars, focusing on price. Contrary to the initial statement, Honda cars have better average mileage (89,623.3) and a lower average price ($3,193.6) than Ford cars. Honda offers higher mileage at a lower cost. However, if low mileage and cost are prioritized, Ford is the better choice. Thus, the decision hinges on whether the buyer values higher mileage or lower cost.



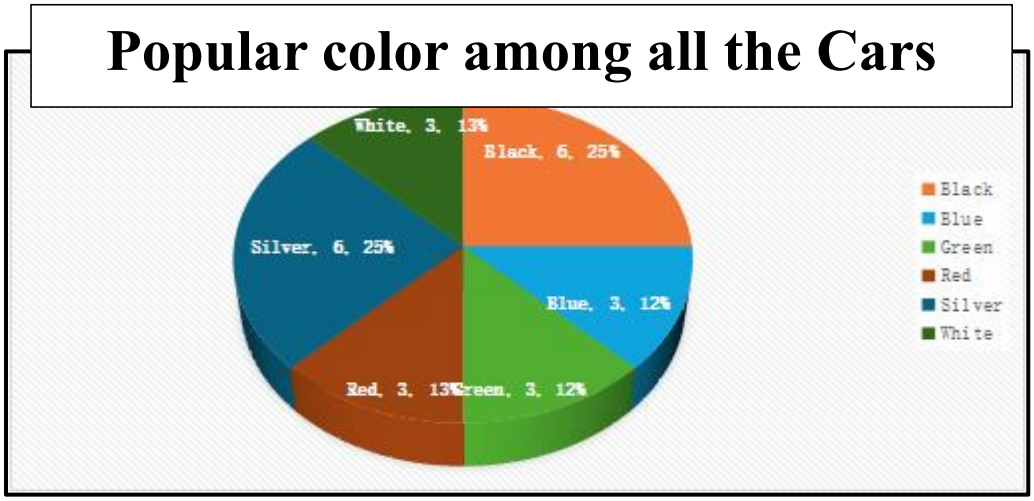






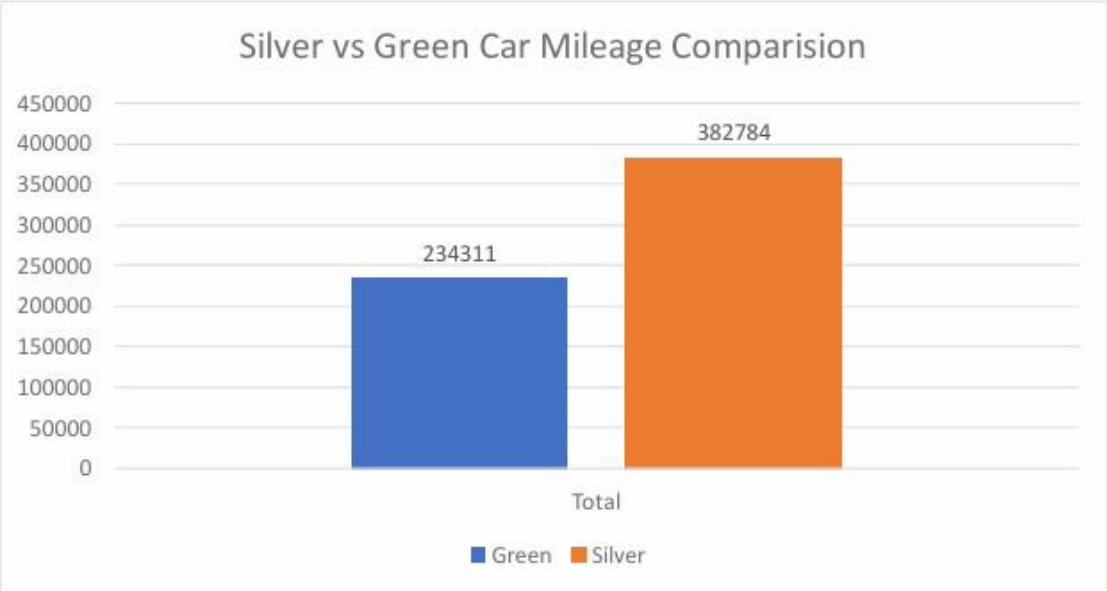
3. Among all the cars which car color is the most popular and is least popular?

ANS: This analysis identifies the most and least popular car colors in the dataset based on make counts. The results show that Black and White are the most popular colors, each accounting for 25% of the cars made by the company. Green and Blue are the least popular, each with 12% of the make. Additionally, Silver and Black are the top colors, each appearing six times, while Blue, Green, Red, and White are the least common, each appearing three times.



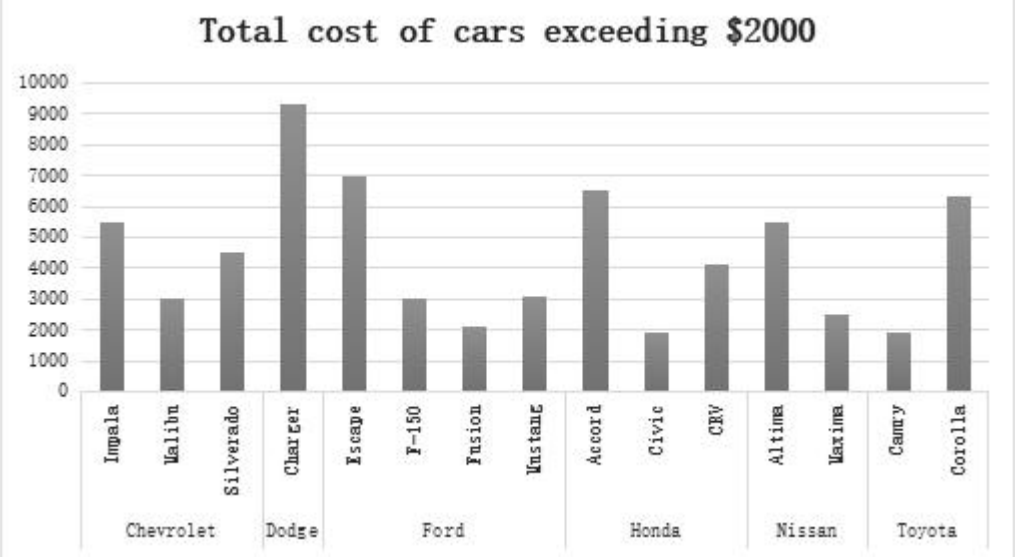
Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.

ANS: This analysis compares silver and green cars by mileage. There are five silver cars (Mustang, Impala, Corolla, Charger, Accord) with Accord having the highest mileage at 101,354. Among the two green cars (Silverado, Altima), Silverado has the highest mileage at 109,23



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

ANS: This analysis aims to find out the car's costs more than $2000. And by using bar graph and taking value as sum of cost it shows the desired result. The total grand cost of all cars exceeding $2000 is $66150.



Conclusion and Review

Mileage Comparison : The analysis revealed that the Chevrolet Impala offers better fuel efficiency compared to the Toyota Corolla.

Ford vs. Honda : Contrary to the initial assumption, Honda cars were found to have better average mileage and price compared to Ford cars, disproving the claim that Ford cars are superior.

Popular Car Colors : Black and White were identified as the most popular car colors, each comprising 25% of the production, while Green and Blue were the least popular, each at 12%.

Silver vs. Green Cars : Among silver cars, the Accord had the highest average mileage. The Silverado had the highest mileage among green cars.

Cost Analysis : The total cost of cars priced over $2000 was $66,150.

The analysis provided valuable insights into mileage, car color popularity, and costs. It highlighted discrepancies, particularly in the Ford vs. Honda comparison. The use of visualizations like column charts and bar graphs effectively presented findings. This report is valuable for car buyers, industry professionals, and researchers, but further exploration into other factors influencing car purchases is recommended.

Regression

Regression analysis examined the dataset, using mileage as the dependent variable and cost and price as independent variables, to determine the statistical relationship between mileage, cost, and price.
Regression Analysis Findings

The regression analysis shows a moderate positive relationship between the predictor and response variables, with a correlation coefficient of 0.40. The model explains 16% of the variance ($R^2 = 0.16$). Each unit increase in the predictor results in a 16.66 decrease in the response variable, with a marginally significant p-value of 0.056.

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

Correlation

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

Anova means the Analysis of variance

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

Anova: Single Factor

SUMMARY

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

ANOVA

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

Descriptive Statistics

The descriptive statistics provide insights into three variables: Mileage, Price, and Cost. The Mileage variable ranges from 34,853 to 140,811 miles, with an average of approximately 83,803 miles. Price ranges from $2,000 to $4,959, and Cost ranges from $1,500 to $4,500. The means and standard deviations offer insights into the central tendencies and variability within each variable. These statistics offer a comprehensive overview of the dataset, enhancing our understanding of the data distribution and characteristics.

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