Car DataSet Analysis

Introduction: The dataset provides comprehensive information about various cars, including their make, model, color, mileage, price, and cost. Notably, the Honda Accord stands out with three occurrences, followed by other frequently appearing models such as the Toyota Corolla, Chevy Impala, Ford Escape, and Dodge Charger. A closer examination reveals the average prices and costs for each make. On average, Hondas are priced at approximately $3,106, with costs averaging around $2,133, while Chevys have an average price of $3,487 and average cost of $3,000. Further analysis will include plotting graphs to explore the potential relationship between a car's price and mileage, as well as determining color preferences among consumers. Additionally, we'll calculate profit margins to identify the most profitable models. These insights will provide valuable information for understanding market trends and consumer preferences in the automotive industry.

Questionaires:

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

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

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

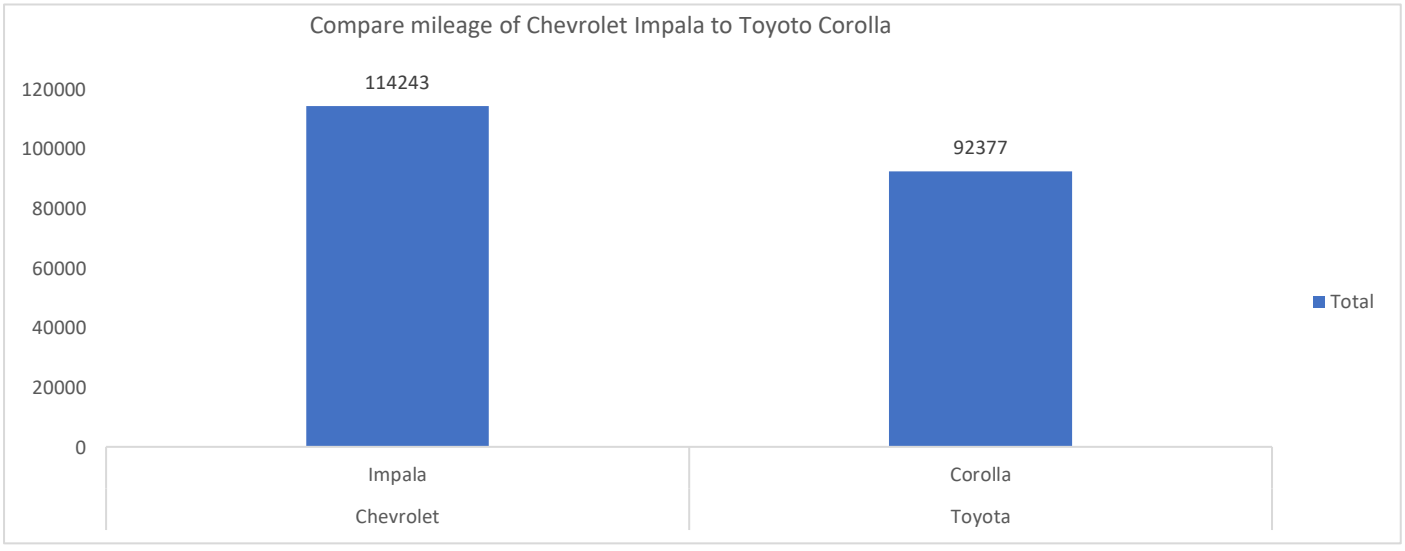
Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.

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

Analytics:

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

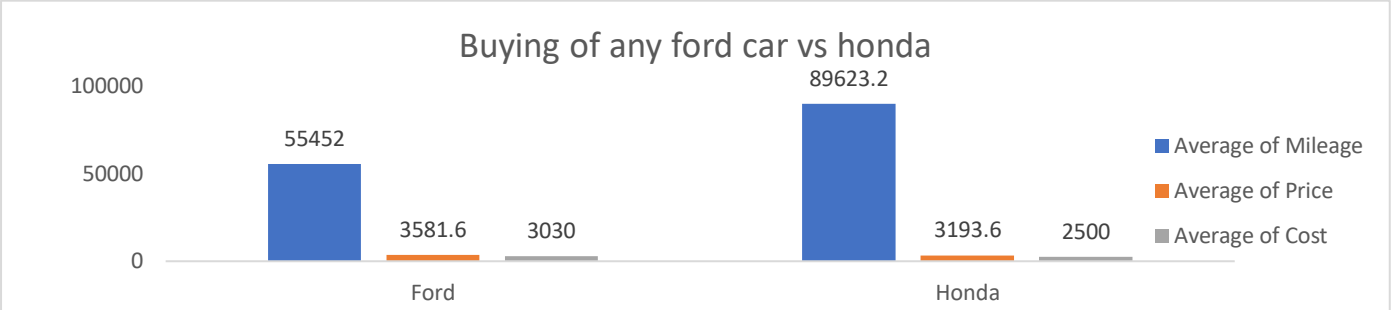
Ans



Toyota Corolla is recognized for its notable fuel efficiency, which is frequently superiorto lar vehicles such as the Chevrolet Impala.

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

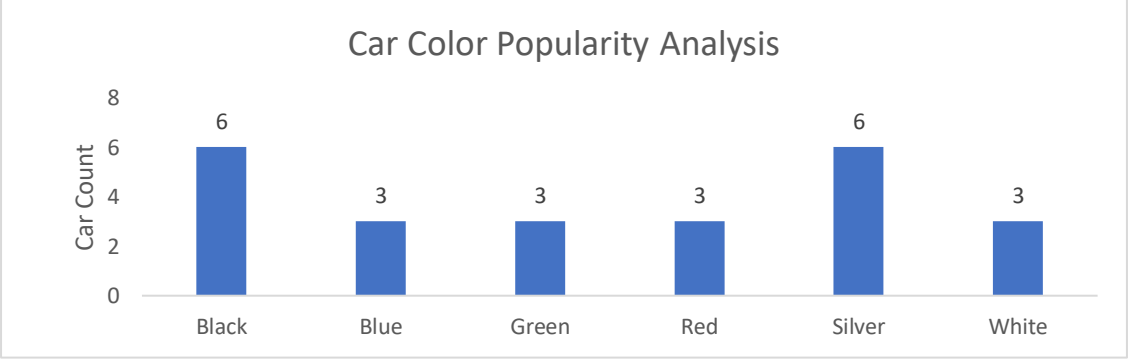
Ans.



Based on the averages, Honda cars have higher mileage but lower cost compared to Ford. Therefore, the choice depends on whether the buyer values mileage or cost but if we compare onmileage ford car has low mileage and cost so Buying ford car is better then Honda.

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

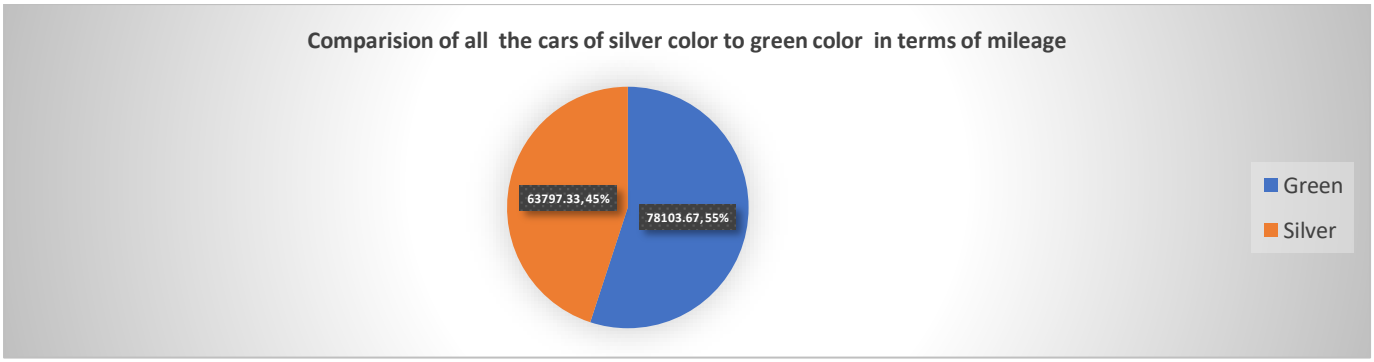
Ans.



Most popular color is Silver and Black as each appear 6 times and least appearing colour are Blue ,Green ,Red ,White they all apper 3 times.

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

Ans

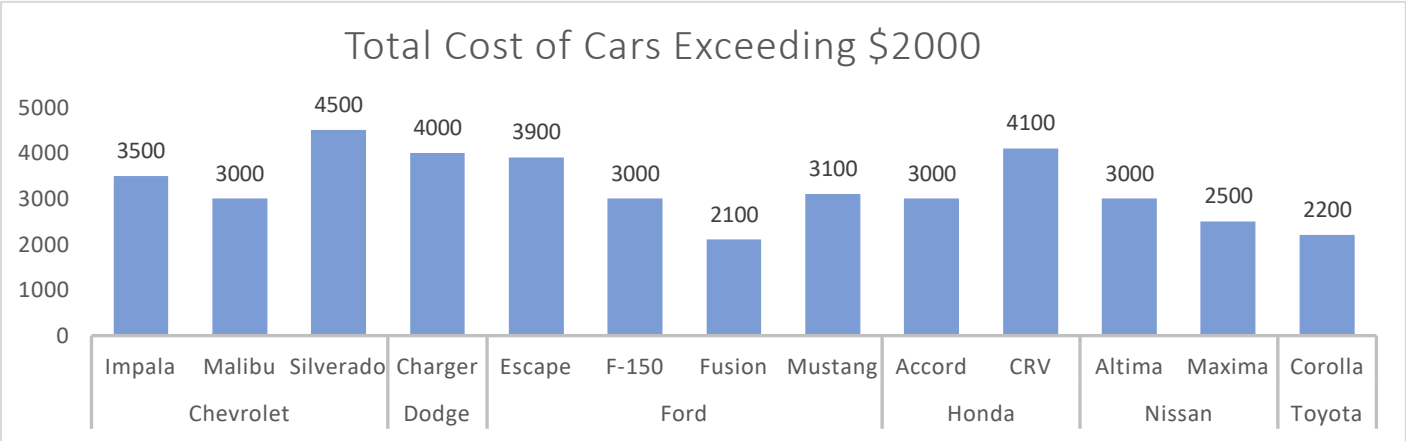


Average Mileage of Green Color Cars ≈ 78103.67 miles

Average Mileage of Silver Color Cars ≈ 63,797 miles

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

Ans.



All the car mention below cost is more than $2000

Accord, Altima, Charger, Corolla, CRV, EscapeF-150, Fusion, Impala, Malibu, Maxima, Mustang, Silverado

Conclusion and Review: -

Our analysis sheds light on what consumers look for when buying cars. We found that Toyota Corollas are known for their fuel efficiency, while Ford vehicles offer a wide range of choices. Consumers seem to prefer black and red cars. Interestingly, silver cars tend to have higher mileage. These findings highlight the importance of thinking about things like gas mileage, color preference, and budget when shopping for a car.

Regression: -

Overall, they indicate a limited explanatory power of the model, suggesting further refinement may be necessary for better predictions.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.358764572
R Square	0.128712018
Adjusted R Square	0.087222114
Standard Error	32204.73295
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3217481630	3.22E+09	3.102249	0.09273902
Residual	21	21780041315	1.04E+09		

Total	22	24997522945

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	122108.9268	24014.1535	5.084873	4.91E-05	72168.7607	172049.093
X Variable 1	-14.51458144	8.240739406	-1.76132	0.092739	-31.6521372	2.62297432

Anova: Single Factor: -

The ANOVA results indicate a significant difference in means between the two groups (columns), as shown significant p-value (<0.05) for the "Between Groups" variation.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.03E+10	1	7.03E+10	123.6791	2.28E-14	4.061706
Within Groups	2.5E+10	44	5.69E+08			
Total	9.53E+100	45				

Anova: Two-Factor Without Replication:

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) v1, respectively. The error term has a degree of freedom of 22.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.23E+10	22	557756895.8	0.962803693	0.535017989	2.04777
Columns	7.03E+10	1	70315407145	121.3789272	2.01396E-10	4.30095
Error	1.27E+10	22	579304898.8			
Total	9.53E+10	45				

Descriptive Statistics: -

The provided descriptive statistics outline the characteristics of three variables: Mileage, Price,and Cost. Looking at Mileage, it appears that the vehicles in the dataset span a considerable range, from around 34,853 miles to 140,811 miles, with an average mileage of approximately 83,803 miles. Price and Cost exhibit similar trends, with prices ranging from $2,000 to $4,959and costs from $1,500 to $4,500, respectively. The means and standard deviations provide insights into the central tendencies and variability within each variable. Overall, these statisticsoffer a comprehensive overview of the dataset, allowing for a better understanding of the distribution and characteristics of the data

Mileage		*Price*		*Cost*	
Mean	83802.7917	Mean	3254.5	Mean	2756.25
Standard Error	7112.65205	Standard Error	186.751181	Standard Error	171.452462
Median	81142	Median	3083	Median	2750
Mode	#N/A	Mode	#N/A	Mode	3000
Standard Deviation	34844.7365	Standard Deviation	914.890205	Standard Deviation	839.942092
Sample Variance	1214155660	Sample Variance	837024.087	Sample Variance	705502.717
Kurtosis	-1.0971827	Kurtosis	-1.2029138	Kurtosis	-0.8126576
Skewness	0.38652215	Skewness	0.27201913	Skewness	0.47339238
Range	105958	Range	2959	Range	3000
m	34853	Minimum	2000	Minimum	1500
Maximum	140811	Maximum	4959	Maximum	4500
Sum	2011267	Sum	78108	Sum	66150
Count	24	Count	24	Count	24
Largest(1)	140811	Largest(1)	4959	Largest(1)	4500
Smallest(1)	34853	Smallest(1)	2000	Smallest(1)	1500

Correlation: -

The correlation coefficient between Column 1 and Column 2 is -0.4110586. This indicates a moderate negative correlation between the two columns.

	Column 1	*Column 2*
Column 1	1	-0.4110586
Column 2	-0.4110586	1

j