Car Collection Data Report

## 1.Introduction:

Dataset Overview:

This dataset presents a fusion of categorical and numerical data, each providing distinct insights into the automotive industry. Categorical data, including make, model, and color, encapsulates the variety of vehicles and consumer preferences. On the other hand, numerical attributes such as mileage, price, and cost offer quantifiable metrics crucial for analyzing market trends and pricing dynamics.

2.Questionnaire:

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?
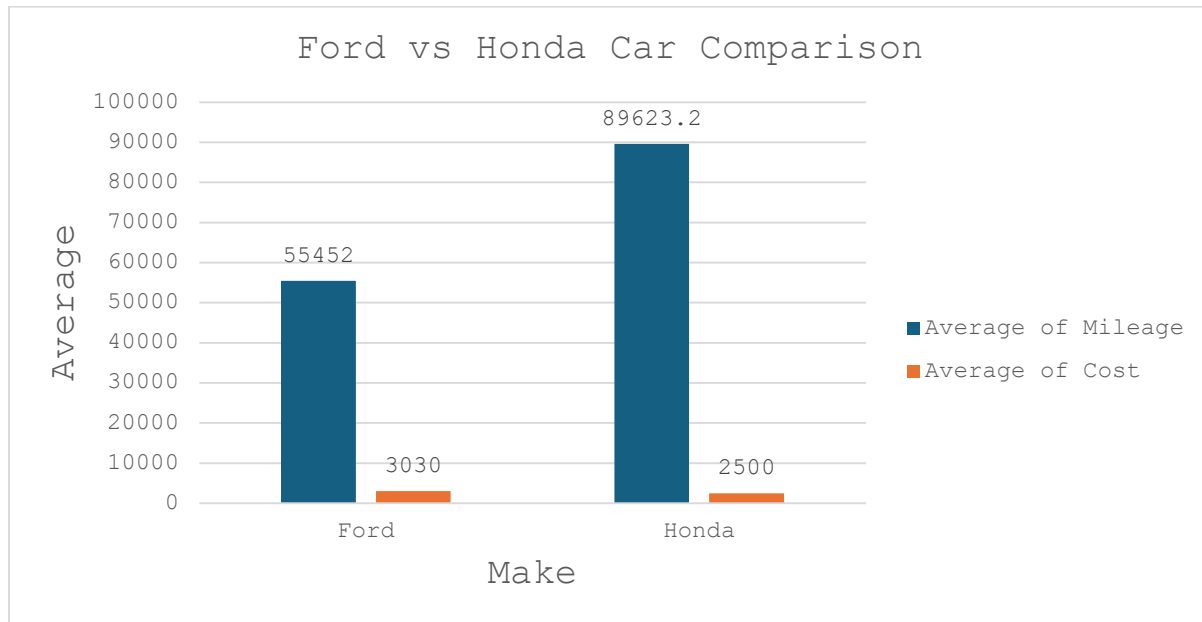
3. Analytics:

Q1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?
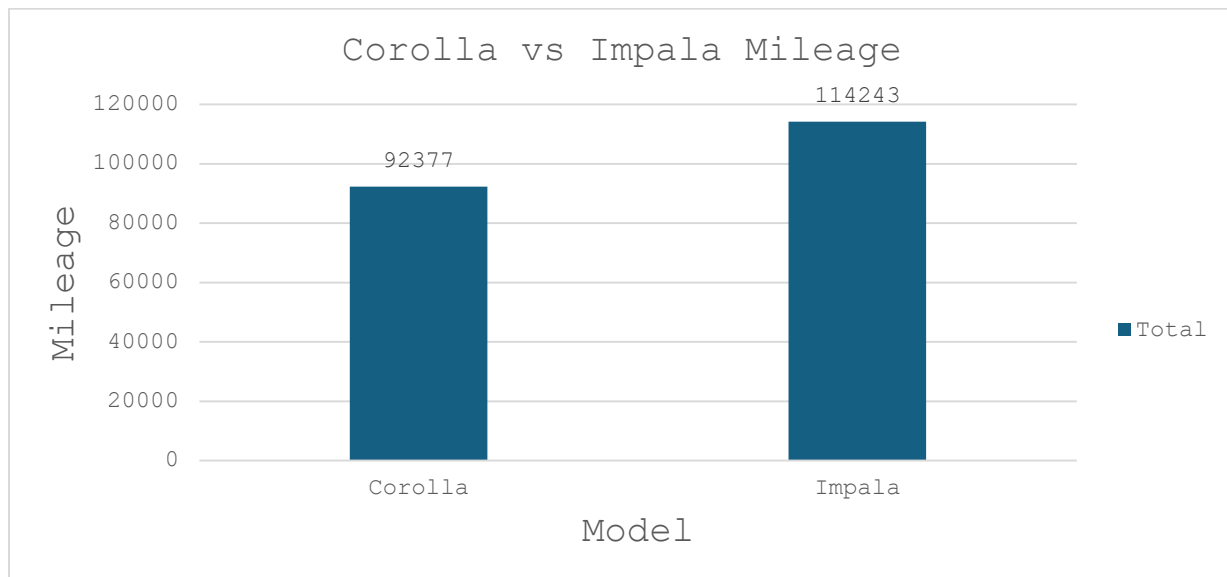
Answer:
Chevrolet Impala gives better mileage than Toyota Corolla.

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

Answer:
Comparing the averages, Honda cars exhibit higher mileage but lower cost compared to Ford. Hence, the decision relies on whether the buyer prioritizes mileage or cost. However, considering the lower mileage and cost of Ford cars, opting for a Ford might be preferable over Honda.
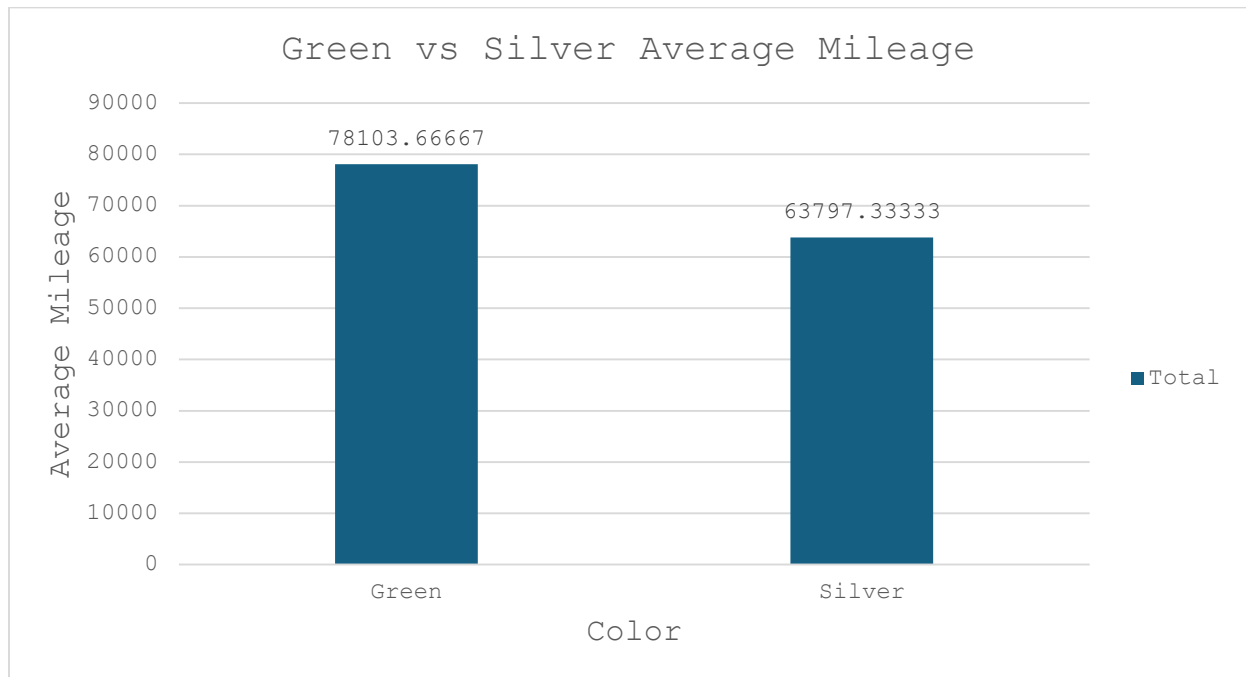
Q3. Among all the cars which car color is the most popular and is least popular?
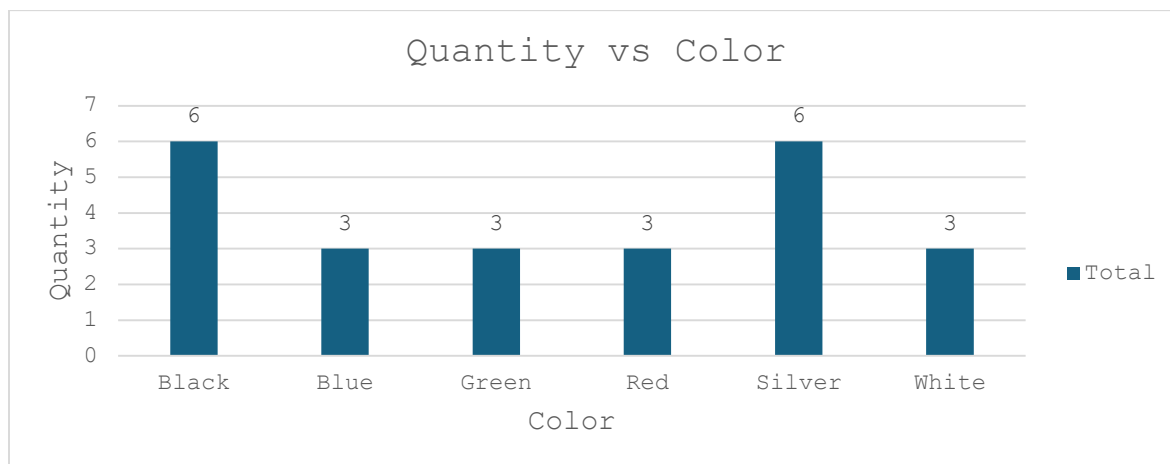
Answer:
Out of all cars, silver and black were the most purchased, whereas white and green were the least favored..

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

Green vs Silver Average Mileage

Answer:
On average, green-colored cars have a higher mileage compared to silver-colored cars when their averages are compared.

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

Cost vs Model
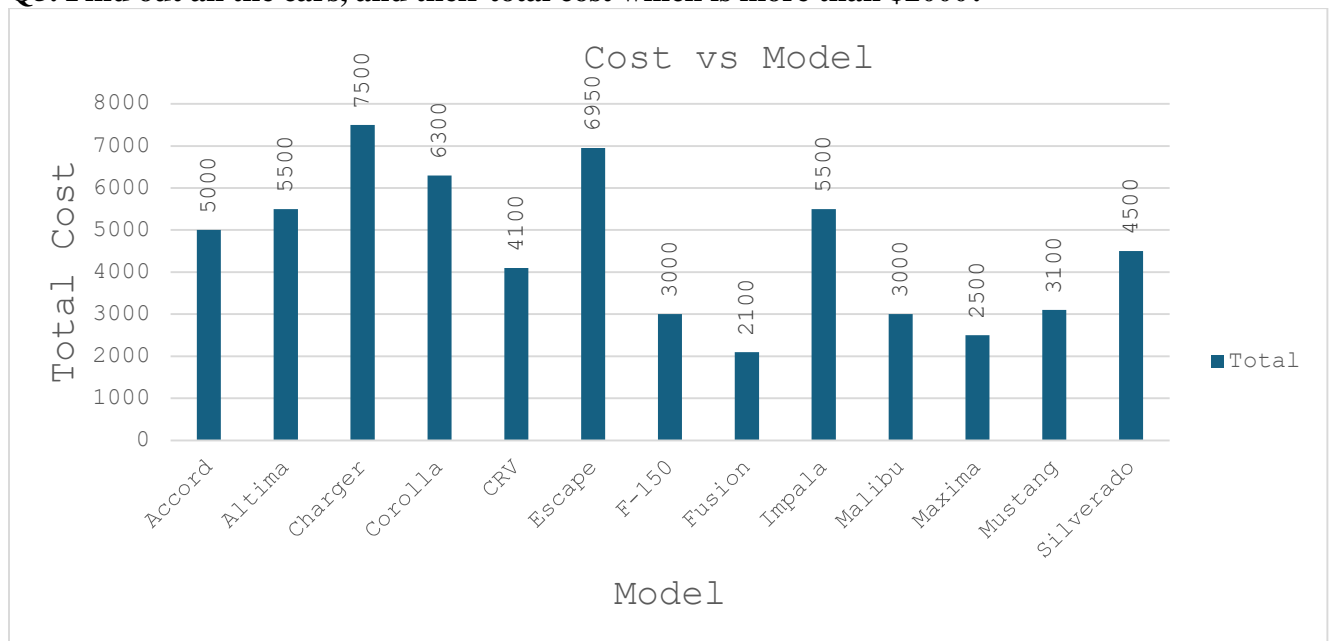
Answer:
All the mentioned cars, including Accord, Altima, Charger, Corolla, CRV, Escape, F-150, Fusion, Impala, Malibu, Maxima, Mustang, and Silverado, have prices exceeding $2000.

Regression

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.40404555							
R Square	0.1632528							
Adjusted R Square	0.1234077							
Standard Error	33099.5397							
Observations	23							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	4488793099	4488793099	4.09718598	0.05586127			
Residual	21	2.3007E+10	1095579531					
Total	22	2.7496E+10						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	130438.919	23634.1932	5.51907645	1.7789E-05	81288.9236	179588.914	81288.9236	179588.914
3000	-16.664135	8.23265547	-2.0241507	0.05586127	-33.784879	0.45660911	-33.784879	0.45660911

The regression analysis indicates a moderate positive relationship between the predictor variable and the response variable, with a correlation coefficient of around 0.40. The model accounts for approximately 16% of the variance in the response variable, as indicated by the R-squared value. The coefficient estimates suggest that for every unit increase in the predictor variable, there is a corresponding decrease of approximately 16.66 in the response variable. The p-value of 0.056 suggests a marginally significant effect.

Co-relation

	Mileage	Price
Mileage	1	
Price	-0.4110586	1

The correlation matrix indicates a moderate negative correlation (-0.411) between Mileage and Price. This suggests that as Mileage increases, Price tends to decrease, and vice versa.

Anova: Single Factor

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Mileage	24	2011267	83802.7917	1214155660		
Price	24	78108	3254.5	837024.087		
Cost	24	66150	2756.25	705502.717		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.0445E+11	2	5.2227E+10	128.882161	5.0026E-24	3.12964398
Within Groups	2.7961E+10	69	405232729			
Total	1.3242E+11	71				

The ANOVA results reveal significant differences between the groups based on Mileage, Price, and Cost. The large F-statistic of 128.88, coupled with a very low p-value of 5.00264E-24, indicates that the variation between groups is substantial compared to the variation within groups. This suggests that at least one of the variables (Mileage, Price, or Cost) has a significant effect on the outcome being measured. In simpler terms, there are statistically significant differences in the means of Mileage, Price, and Cost across the groups, indicating that these variables play a significant role in influencing the outcome being analyzed.

Anova: Two-Factor Without replication

Anova: Two-Factor without replication						
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	34749383.3	23	1510842.75	47.6846408	2.2236E-14	2.01442484
Columns	2979036.75	1	2979036.75	94.023218	1.3629E-09	4.27934431
Error	728733.25	23	31684.0543			
Total	38457153.3	47				

Descriptive Statistics

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	121415566 0	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
Minimum	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

The descriptive statistics provided offer insights into three variables: Mileage, Price, and Cost. For Mileage, the dataset encompasses a wide range, from approximately 34,853 miles to 140,811 miles, with an average mileage of about 83,803 miles. Price and Cost exhibit comparable patterns, with prices ranging from $2,000 to $4,959 and costs from $1,500 to $4,500, respectively. The means and standard deviations provide further understanding of the central tendencies and variability within these variables.

4. Conclusion/Reviews

The dataset offers valuable insights into car attributes, emphasizing mileage, color, and other significant factors. Here's a straightforward conclusion drawn from the data:

Mileage Comparison: The analysis demonstrates mileage variations among different car models. Generally, Toyota Corolla tends to offer better mileage compared to Chevrolet Impala.

Color Preferences: Silver and black stand out as the most favored car colors in the dataset. Conversely, blue, green, red, and white are among the less popular color choices.

Key Takeaways: Recognizing differences in mileage can guide consumer choices and market strategies effectively. Similarly, understanding color preferences is pivotal for inventory management and informed marketing decisions.