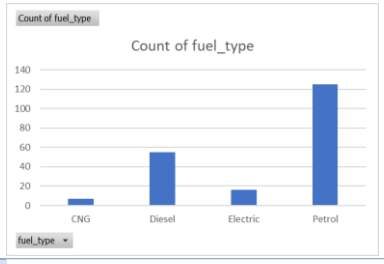
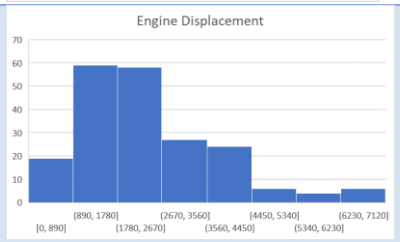
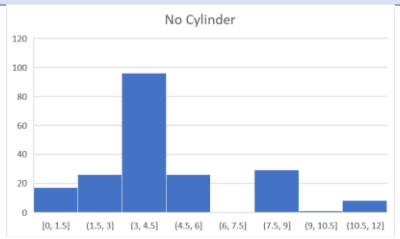
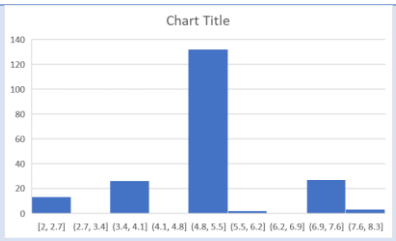
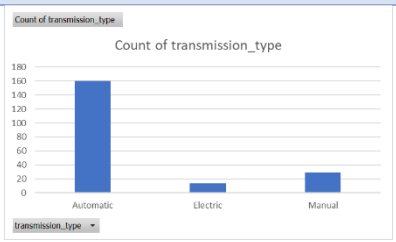
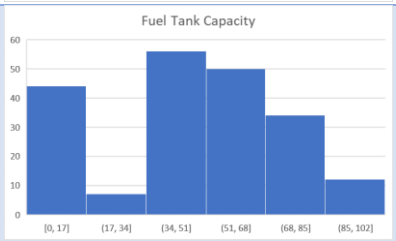
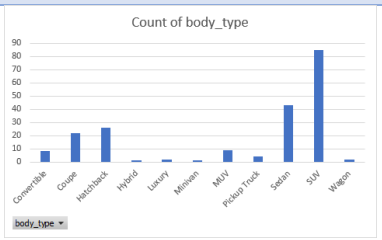
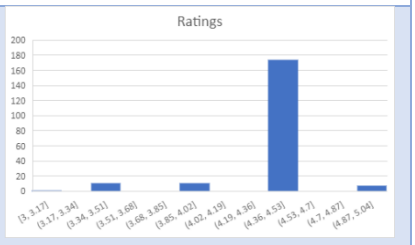
Report By: Brux D., Jens B., Malcolm H.

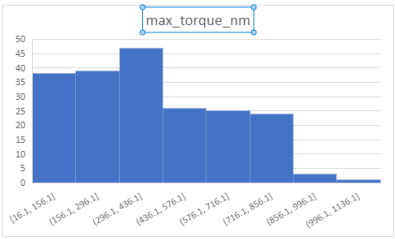
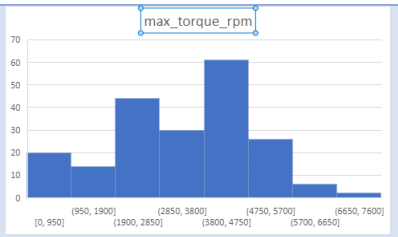
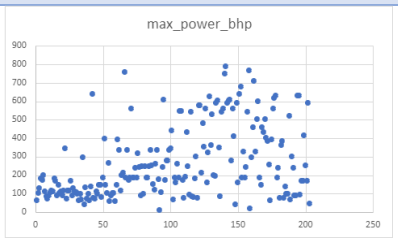
Introduction:

We are a group of data analysts that are providing an informational guide to aid consumers in finding the right, affordable car for them. To accomplish this, we have done data analysis to provide a navigable reference for consumers. Our data comes from a used car data set from 2022 that is updated quarterly. In this data set, there are 203 unique types of vehicles of which the features and technical performances of all of them are provided. New vehicles are marketed at exorbitant prices nowadays, this is increasing with the growing demand but struggling supply for computer chips like semiconductors. We are targeting those who are looking in the car market but don't want to break the bank on a new car. Our data and analysis will provide a deeper understanding of the used vehicle market and what vehicle types are ideal for specific purposes and necessities.

Descriptive Statistics

Variable Name	Variable Type	Discrete/ Continuous	Description	Variable Visuals and Mean/Standard Deviation
Car Name	Unique		The car name is a unique variable because there are more than 2 car names.	
Fuel Type	Categorical		The type of fuel the car runs on.	
Engine Displacement	Numerical	Discrete	Engine Displacement is the combined swept volume of the pistons inside the cylinders of the engine.	 Mean: 2305.92 StDev: 1493.84
Cylinders	Numerical	Discrete	The number of cylinders that the vehicle's engine has. This is important as more cylinders puts out more power.	 Mean: 4.71 StDev: 2.54

Seating Capacity	Numerical	Discrete	The number of seats each specific vehicle contains.	 Mean: 5 StDev: 1.18
Transmission	Categorical (Binary)		The transmission variable looks at whether the vehicle uses an automatic or manual transmission.	
Fuel Tank	Numeric	Discrete	The number of gallons the fuel tank can fit.	 Mean: 46.14 StDev: 28.90
Body Type	Categorical 3+		This variable indicates the body type of the vehicle, such as a hatchback, SUV, MUV, sedan, etc...	
Rating	Numerical	Continuous	The average of the customers ratings, out of a scale of five, for the car.	 Mean: 4.4 Standard Dev: 0.28
Starting Price	Numerical	Discrete	The cost of the cars starting base model without additions.	

				Mean: 115618.21 Standard Dev.: 166141.44
Ending Price	Numerical	Discrete	The cost of the cars model with all additions.	 Mean: 11120054.2 Standard Dev: 15517463.1
Max Torque NM	Numerical	Continuous	The more torque, the greater amount of power an engine can produce in terms of Newton-meters (NM).	 Mean: 402.8 Standard Dev: 237.12
Max Torque RPM	Numerical	Discrete	The more torque, the greater amount of power an engine can produce in terms of rotations per minute (RPM)	 Mean: 3346.51 Standard Dev: 1594.96
Max Power BHP	Numerical	Continuous	Refers to the maximum amount of power that an engine can produce at its peak performance. BHP stands for Brake Horsepower, which is a unit of power that measures the amount of power produced by an engine at the output shaft.	 Mean:266.58 Standard Dev: 193.47

Confidence Intervals

Starting Price	[92762.95, 138473.46]	Based on our data, this set shows that with 95% confidence the

		mean starting price will fall between 92762.95 and 138473.46. However, there is a 5% chance of error. This means we are confident in the prices of the cars we reviewed being in this level.
Rating	[3.884440569, 4.982554505]	Based on our data, this set shows that with 95% confidence the mean rating will fall between 3.88 and 4.98. However, there is a 5% chance of error. This means all of our analyzed vehicles have average to high levels of satisfaction.
Engine Displacement	[2100.415, 2511.427]	Based on our data, this set shows that with a 95% confidence, the mean engine displacement will fall between 2100.415 and 2511.427. However, there is a 5% chance of error. This means that there is a relatively low level of engine displacement.

Data Analysis Tests

3+ One Sample Hypothesis Tests:

Test 1:

Ho Hypothesis: The mean starting price for automatic transmission vehicles is greater than 800,000.
Ha Hypothesis: The mean starting price for automatic transmission vehicles is less than 800,000.
Decision: P < alpha? No, fail to reject Ho
P Value: 1 **Alpha:** 0.05
Conclusion: Based on the sample, we are 95% confident that the starting price is greater than 800,000, therefore we fail to reject Ho. However, there is a 5% chance of making a type II error.
Recommendation: With a competitive car market, it is difficult to attract more customers, a lower starting price could be implemented for vehicles in order to bring the mean starting price down

Test 2:

Ho Hypothesis: The mean ending price for automatic transmission vehicles is less than 12,000,000.
Ha Hypothesis: The mean ending price for automatic transmission vehicles is greater than 12,000,000.
Decision: P < alpha? No, fail to reject Ho
P-Value: 0.210 **Alpha:** 0.05
Conclusion: Based on the sample, we are 95% confident that the mean ending price is less than 12,000,000 for automatic vehicles, therefore we fail to reject Ho. However, there is a 5% chance that we make a type II error.
Recommendation: A comprehensive market analysis is necessary in order to evaluate the demand and potential profitability of automatic vehicles priced below 12,000,000. This analysis would consider factors such as customer preferences, market size, competition, and any specific trends or regulations affecting the automotive industry.

Test 3:

Ho Hypothesis: The mean max torque rpm for automatic transmission vehicles is less than 3000

Ha Hypothesis: The mean max torque rpm for automatic transmission vehicles is greater than 3000
Decision: P < alpha? No, fail to reject Ho
P-Value: 0.999 **Alpha:** 0.05
Conclusion: Based on the given sample, we are 95% confident that the mean max torque rpm is less than 3000 for automatic vehicles, therefore we fail to reject Ho. However, there is a 5% chance of making a type II error.
Recommendation: Continuously monitor the torque performance of competitor vehicles in the market to stay ahead of the competition. Identify opportunities for differentiation and capitalize on any weaknesses or limitations that other vehicles may have in terms of torque delivery at lower RPMs.

3+ Two Sample Tests:

Test 1:
Ho Hypothesis: The mean max power bhp for automatic transmission vehicles is greater than or equal to the mean max power bhp for manual transmission vehicles.

Ha Hypothesis: The mean max power bhp for automatic transmission vehicles is less than the mean max power bhp for manual transmission vehicles.

Outcomes: Based on this sample, we are 95% confident that the Max Power BHP of Manual transmission vehicles is greater than the mean max power bhp of automatic transmission vehicles, therefore we fail to reject Ho.

Possible Errors: There is a 5% chance of making a Type 1 error.

Recommendation: The car market is diverse and large with many consumers wanting different components to a vehicle. By concluding that Manual vehicles on average have a greater power output than automatic vehicles we can more accurately market options for consumers looking to go fast!

Test 2:
Ho Hypothesis: The mean engine displacement for Diesel fuel type vehicles is greater than or equal to the mean engine displacement for Petrol fuel type vehicles.

Ha Hypothesis: The mean engine displacement for Diesel fuel type vehicles is less than the mean engine displacement for Petrol fuel type vehicles.

Outcomes: Based on this sample, we are 95% confident that the mean Engine Displacement for Petrol fuel type vehicles is greater than the mean Engine Displacement for Diesel fuel type vehicles. Therefore, we reject Ho.

Possible Errors: There is a 5% chance of making a Type 1 error.

Recommendation: Engine displacement is an important specification because it provides an indication of the engine's power potential and performance characteristics. The size of the engine directly affects its ability to produce power and torque. Therefore, by investing in petrol vehicles, we can build our brand of dealing powerful, high performing vehicles.

Test 3:
Ho Hypothesis: The mean fuel tank capacity for 4-cylinder vehicles is greater than or equal to the mean fuel tank capacity for 6-cylinder vehicles.

Ha Hypothesis: The mean fuel tank capacity for 4-cylinder vehicles is less than the mean fuel tank capacity for 6-cylinder vehicles.

Outcomes: Based on this sample, we are 95% confident that the mean fuel tank capacity for 4-cylinder vehicles is greater than the mean fuel tank capacity for 6-cylinder vehicles.

Possible Errors: There is a 5% chance of making a Type 2 error.

Recommendation: All cylinders in a vehicle are injected with fuel while driving, therefore the more cylinders in a vehicle the more fuel is burned. Based on our data, even though 4-cylinder vehicles burn less fuel, they have on average a greater fuel tank capacity. As a result, we should avoid purchasing 6-cylinder vehicles without sufficient fuel tank capacity. These cars won't have positive customer feedback and will sit on the lot.

3+ ANOVA Tests:

Test 1:
Ho Hypothesis: There is a significant difference in starting prices among different fuel types.
Ha Hypothesis: There is no significant difference in prices among the body types.
Decision: P < alpha? Yes, reject Ho
P-Value: .019694
Alpha: 0.05
Conclusion: Based on the sample, we are 95% confident that there is a not a significant difference in starting prices among different fuel types, therefore we reject Ho. However, there is a 5% chance of making a type I error.
Recommendation:
Test 2:
Ho Hypothesis: There is significant difference in prices among the body types.
Ha Hypothesis: There is no significant difference in prices among the body types.
Decision: P < alpha? Yes, reject Ho
P-Value: 8E-5
Alpha: 0.05
Conclusion: Based on the sample, we are 95% confident that there is not significant difference in starting prices among different body types, therefore we reject Ho. However, there is a 5% chance of making a type I error.
Recommendation: By knowing this information, we are able to infer that we can have a larger variety of body types, opening ourselves up to a larger market while diversifying the inventory.
Test 3:
Ho Hypothesis: There is significant difference in max power among the cars different body types.
Ha Hypothesis: There is no significant difference in max power among the cars different body types.
Decision: P < alpha? Yes, reject Ho
P-Value: 6.36E-19
Alpha: 0.05
Conclusion: Based on the sample, we are 95% confident that there is not a significant difference in max power BHP among different body types, therefore we reject Ho. However, there is a 5% chance of making a type I error.
Recommendation: The max power a car can put out is a big factor in buying your car as it has a strong correlation with a car's acceleration and how it will drive depending on if it's a heavier or lighter car with higher or lower BHP. Our market is very open and has a lot of variation when it comes to looking for different BHP's as the cars different body types don't have a significant statistical difference in BHP.

2 Goodness of Fit Tests:

Test 1:
In this Test we determined whether or not the reviews count data fits a uniform distribution.
Ho Hypothesis: This data fits a uniform distribution.
Ha Hypothesis: This data does not fit a uniform distribution.
Decision: P < alpha? Yes, reject Ho
P-Value: 3.61E-172
Alpha: 0.05
Conclusion: Based on this sample, we are 95% confident that this data does not fit a uniform distribution. There is a 5% chance of making a type 1 error.
Test 1:
In this Test we determined whether the seating capacity data fits a uniform distribution.
Ho Hypothesis: This data fits a uniform distribution.
Ha Hypothesis: This data does not fit a uniform distribution.
Decision: P < alpha? Yes, reject Ho
P-Value: 3.01E-127
Alpha: 0.05
Conclusion: Based on this sample, we are 95% confident that this data does not fit a uniform distribution. There is a 5% chance of making a type 1 error.

1 Independence Test:

Ho Hypothesis: The vehicles body type and seating capacity are independent variables
Ha Hypothesis: The vehicles body type and seating capacity are not independent variables
Decision: P < alpha? No, fail to reject Ho
P-Value: 0.998
Alpha: 0.05
Conclusion: Based on the sample we do not have sufficient evidence to reject the Ho that the car body type and the car seating capacity are independent. There is a 5% chance of making a type II error.

References

Tyagi, M. (2022, September 3). *Cars 2022 dataset*. Kaggle. Retrieved April 13, 2023, from https://www.kaggle.com/datasets/tr1gg3rtrash/cars-2022-dataset

Tucker, S. (2021b, July 29). Is Now the Time to Buy, Sell, or Trade-In a Used Car? Retrieved from Kelley Blue Book website: https://www.kbb.com/car-advice/is-now-the-time-to-buy-sell-or-trade-in-a-used-car/

Carlier, M. (n.d.). *Topic: Automotive Industry Worldwide*. Statista. Retrieved April 13, 2023, from https://www.statista.com/topics/1487/automotive-industry/#topicOverview

Appendix A

Test	Variable Name
One Sample Test #1	Starting Price
One Sample Test #2	Ending Price
One Sample Test #3	Max Torque

Two Sample Test #1	Transmission Type, Max Power BHP
Two Sample Test #2	Engine Displacement, Fuel Type
Two Sample Test #3	Fuel Tank Capacity, # of Cylinders
ANOVA Test #1	Fuel Type, Starting Price
ANOVA Test #2	Body Type, Starting Price
ANOVA Test #3	Max Power, Body Type
Goodness of Fit Test#1	Reviews Count
Goodness of Fit Test #2	Seating Capacity
Independence	Body Type, Seating Capacity