The car dataset used in the project was originally taken from the StatLib library maintained at Carnegie Mellon University. I took the slightly updated version of the dataset from here. The dataset has details of about 406 cars with characteristics such as MPG, Cylinders, Displacement, Horsepower, etc. In this project, we mainly focus on the statistical characteristics of cars based on their origin. Automobile companies in the US, Japan, and European countries are the major innovators or producers of cars. I was always interested in understanding the focus of an automobile manufacturer. Some may focus to improve the Miles per gallon (MPG) of cars, while others may focus to increase the horsepower of the cars. I want to understand their interests based on the origin of the manufacturers. Below are the details of the attributes in the dataset.

Name of the Attribute	Attribute Type
Car Name	String
MPG	Continuous
Cylinders	Multi-valued discrete
Displacement	Continuous
Horsepower	Continuous
Weight	Continuous
Acceleration	Continuous
Model	Multi-valued discrete
Origin	Multi-valued discrete

Module-1:

1. **Calculate the MPG of the top 30 percentile of US car models (Origin – US)**
2. **Calculate the MPG of the top 30 percentile of Japanese cars (Origin – Japan)**
3. **Calculate the MPG of the top 30 percentile of European cars (Origin – Europe)**

Answers to the above questions will let us know who focuses more on the MPG of cars.

1)	Mean MPG for US Cars	19.68819	Mean MPG for Japanese Cars	30.45063
	SD for US Cars	6.92829	SD for Japanese Cars	6.090048
	Mean MPG for European Cars	26.74521	Top 30% MPG for US Cars	20.03567
			Top 30% MPG for Japanese Cars	30.75607
	SD for European Cars	8.626075	Top 30% MPG for European Cars	27.17783

From the output, we can conclude that the Japanese carmakers focus more on Mile Per Gallon (MPG) than other manufacturers.

##########

Module-2:

1. **Construct a 95% confidence interval for the mean weight of cars designed in Europe.**

MD-2	Mean weight of Europe Cars	2431.493	Lower Conf Limit	2317.046
	Weight SD of Europe Cars	490.8836	Upper Conf Limit	2545.941
	Number of observations	73		
	Degrees of freedom	72		
	For 95% confidence interval, t0.025,72	1.992		

Therefore, the 95% confidence interval for the mean weight of cars designed in Europe is

LCL - 2317.046 & UCL - 2545.941

##########

Module-3:

We test the below hypothesis

1. **The average horsepower of cars by US manufacturers is greater than 103hp.**

I want to validate if the conversation here is true. The dataset I used in this project is old. So, I want to validate if the average horsepower of cars by US manufacturers is greater than 103hp (which is minimum as per the above conversation).

a. State the hypothesis

• Null Hypothesis: The average horsepower of cars by US manufacturers is less than or equal to 103.

• Alternative Hypothesis: The average horsepower of cars by US manufacturers is greater than 103.

b. Appropriate statistical measure

Mean (μ) = the mean horsepower of cars by US manufacturers.

c. Determine whether it is a one-tailed or two-tailed test

From the above alternate hypothesis, we are interested in finding if the mean horsepower of cars by US manufacturers is greater than 103 or not. From which we understand that it is a one-sided and hence a one-tailed test.

d. State the hypothesis

H0: $\mu <= 103$

Ha: $\mu > 103$

e. Specify the significance level or level of the test

The significance level or level of test in the problem is 0.05

f. Specify test statistic to be used

We only know the sample standard deviation here. We don't know the population standard deviation. So, we use t-test statistics for this case.

g. Find the critical value

Degrees of freedom = 254 – 1 = 253

Using the t table, t0.05, 253 = 1.6508

We reject the null hypothesis if the calculated t value is greater than the critical value of **1.6508**

h. Compute the test statistic

Mean horsepower of US cars = 118.0118

Sample Standard Deviation = 42.39793

N = 254

MD-3	Mean horsepower of US cars	118.0118	t value	5.642938
	Horsepower SD of US cars	42.39793		
	Number of observations, n	254		
	μ	103		

t test statistic value = **5.643**

i. Decide to either reject or not to reject the null hypothesis

The computed t statistic value (in step h) is greater than the critical value (in step g). So, we reject the null hypothesis.

j. State the conclusion in terms of the original question

There is sufficient evidence to conclude that the mean horsepower of US cars is greater than 103. So, the conversation on the above webpage is valid.

##

Module-4:

We test the below hypothesis

1. **The mean displacement of cars by US manufacturers is greater than the mean displacement of cars by European manufacturers. (Alpha = 0.01)**

I want to validate if the conversation here is true.

a. State the hypothesis

• Null Hypothesis: The mean displacement of cars by US manufacturers is less than or equal to the mean displacement of cars by European manufacturers.

• Alternative Hypothesis: The mean displacement of cars by US manufacturers is greater than the mean displacement of cars by European manufacturers.

b. Appropriate statistical measure

$\mu 1$ = the mean displacement of cars by US manufacturers.
$\mu 2$ = the mean displacement of cars by European manufacturers.

c. Determine whether it is a one-tailed or two-tailed test

From the above alternate hypothesis, we are interested in finding if the mean displacement of cars by US manufacturers is greater than the mean displacement of cars by European manufacturers. From which we understand that it is a one-sided and hence a one-tailed test.

d. State the hypothesis

H0: μ1- μ2<=0

Ha: μ1- μ2>0

e. Specify the significance level or level of the test

The significance level or level of test in the problem is 0.01

f. Specify test statistic to be used

We only know the sample standard deviation here. We don't know the population standard deviation. So, we use t-test statistics for this case.

g. Find the critical value

Degrees of freedom = 254+73 – 2 = 325

Using the t table, t0.01, 325 = 2.337876

We assume that the population variances are equal.

We reject the null hypothesis if the computed t-value is greater than **2.337876**

h. Compute the test statistic

Mean displacement of US cars = 247.935

Sample Standard Deviation of US cars = 98.6478

Mean displacement of Europe cars = 109.4658

Sample Standard Deviation of Europe cars = 22.37191

MD-4	Mean displacement of US cars	247.935	t value (equal variances)	11.89314
	Displacement SD of US cars	98.6478		
	Mean displacement of Europe cars	109.4658		
	Displacement SD of Europe cars	22.37191		
	Number of observations, n1 (US)	254		
	Number of observations, n2 (Europe)	73		
	Mu1-Mu2	0		

t test statistic value = **11.89314**

i. Decide to either reject or not to reject the null hypothesis

The computed t statistic value (in step h) is greater than the critical value (in step g). So, we reject the null hypothesis.

j. State the conclusion in terms of the original question

There is sufficient evidence to conclude that the mean displacement of US cars is greater than the mean displacement of Europe cars. So, the conversation on the above webpage is valid.