**PETER KIRIMI SAMPLING PLAN**

**THE SAMPLING PLAN**

**1.Problem Statement**

We will work as a Data Scientist for the Autolib electric car-sharing service company to investigate a claim about the blue cars from the provided Autolib dataset.

In an effort to understand the company’s electric car usage by solving research questions,we are needed to solve a claim at hand that the Blue Cars are taken mostly during the weekends,compared to the other days of the week(weekdays).

In this case,the variables used in this test were namely:

1.Blue\_cars\_taken\_sum -  **Number of bluecars taken that date in that area**

**For this research,all areas that had any information on bluecars taken were used.**

2.Days\_ of\_the\_week -identifier of weekday (0: Monday -> 6: Sunday

Null Hypothesis(H0):mean=125.9

Alternate Hypothesis(Ha) :mean is not equals to(!==)125.9

Where mean stands for the mean of the bluecars taken

This research is to help the company make important financial decisions based on information acquired from the research so as to increase profit.

1. **Data Description**

To solve the research question, a proper study and consideration of the dataset has to be made as this will help the researcher in the understanding and analysis of the provisional data and sample.

For this stage, the variables under consideration are the (Blue\_cars\_taken\_sum,Days\_ of\_the\_week) respectively.

A random sample of 100 was taken from the dataset for the hypothesis testing,with a mean of (128.610000) and a standard deviation of(157.324659) respectively.

The minimum and maximum values were(2.000000 & 840.000000) respectively.

A normality test was done for the sample obtained from the dataset.

I opted for the shapiro wilk test for this test.In the analysis it is seen that the test gives a value of(0.440) which is greater than the alpha value,hence the distribution looks like a Gaussian distribution.This tells that the distribution is normal.

**3.Hypothesis Testing Procedure**

The sampling method used in this was random sampling as it is easier to implement and every element has an equal probability of being chosen.

After confirming that our selected sample data has no null and duplicated values and normal distribution since that’s the requirement for using the t\_test,we now performed the hypothesis testing assuming the conditions for inference have been met.

In this section I opted for the t-test.This is because the population sample is 100,that is more than 30.It is known that the t-test is used for samples that are greater than 30.

Our claim from the analysis states that the Bluecars are used mostly over the weekend compared to the other days of the week.So our expectations before the hypothesis testing would be that the claim is true.

In an effort to prove the claim,we chose for our Null hypothesis to be the mean of the population(125.926951).And a Null hypothesis of (125.926951) to disagree with the claim .

The formula to performing the T-test Is

t = (sample mean-Null hypothesis/mean of the population)

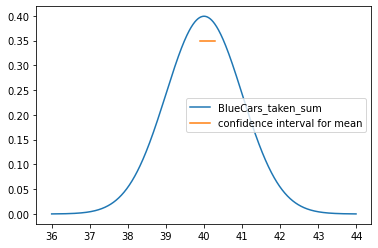
Standard Deviation of sample / the square root of (n)

4.**Hypothesis Testing Results**

The T\_test gave a value of (0.861)

So in this case we fail to reject the null hypothesis as the P value of(0.8054640791511911) is above our alpha/Significance level(0.05).

The points of estimation are (39.89206439763399 , 40.284057194542)



1. **Discussion of Test Sensitivity**

The practical significance of the finding in the event that your null hypothesis is rejected and the power of the test (1-alpha) for the given alpha level and sample size, and the effect of changing the sample size.