# **Problem Statement**

I am going to be looking at the autolib dataset. This is a dataset that contains information about an electric-car lending company. The data entails details on the picking and dropping times of the vehicles as well as other details.

I will focus more on the returning of the blue-cars on Sunday and Monday.

**Null**: More blue-cars are returned on Sunday than Monday.

**Claim**: Less blue-cars are returned on Sunday than Monday.

The reason for my choosing of the hypothesis is to see when most people return the cars after the weekend. Whether they are returned on Sunday or on Monday more.

# **Data Description**

The data that I have was provided by Moringa. You can have a view of it through the like at the bottom of the page. Both the data itself and the description are found in the links.

The data we have contains fields that we are going to use to analyse our hypothesis and determine whether to accept or reject them.

The fields in the dataset include:

-Postal code

-Date

-Data points

-Day of the Week

-Day type

-Blue-cars returned

-Blue-cars taken

-Utilib-14 returned

-Utilib-14 taken

-Utilib returned

-Utilib taken

-Free slots

-Taken slots

The data can help me validate my hypothesis as it contains useful data for my analysis.

# **Hypothesis Testing Procedure**

### **Null and alternative Hypothesis**

The null and alternative hypothesis are as follow; (**H0** is the null hypothesis while **H1** is the alternative hypothesis)

**H0:** More blue-cars are returned on Sunday than Monday

**H1:** Less blue-cars are returned on Sunday than Monday.

### **Significance Level**

I have decided to go with a 0.05 significance level in my hypothesis testing. This generally means that there is a 5% chance that I will accept my alternative hypothesis. Otherwise on the contrary, there is a 95% chance my null hypothesis is actually true.

### **Calculating the Test Statistic and Corresponding P-value**

I have decided to use the z-statistic. This is because the data I have exceeded the value of 30.

The resulting p-value will determine whether I will reject or accept my null value.

### **Drawing a Conclusion**

I will then draw a conclusion after analysis of my p-value.

# **Hypothesis Testing Results**

First I used stratified sampling for my data because I was going to be analyzing data based on groups and this method would help maintain the proportions. I then did the z-test on the sample data in order to obtain a z-score for interpretation into a p-value. If my p-value became less than my significance level, then I would have to reject my null hypothesis. If the p-value was higher than the significance level, then I would have to accept my null hypothesis.

After calculations, I found the following;

* The z-score is: 0.25301631811195047
* The p-value is: 0.4001278016768607

The point estimators for the parameter were as follows;

* The mean is: 128.19212462137602
* The standard deviation is: 180.59071025551745

The confidence interval, after employing a 95% confidence value is:

* 128.19 +/- 5.891163381097424

# **Discussion of Test Sensitivity**

I used the data on the collection days of my interest and the sum for the blue-cars returned on those days. This was able to give me a picture on whether to accept or reject the null hypothesis that I had.

# **Summary and Conclusions**

I got a p-value that was more than my significance level and thus that means that I have to accept the null hypothesis. On confirmation of my hypothesis,(whether or not it may be true, I went on ahead to plot a bar graph for the two days to have a visual view and indeed the hypothesis testing was a success.

I can thus conclude that more blue-cars are returned on Sunday

# **Links**

[**Colab notebook**](https://colab.research.google.com/drive/1KbAtdPJf1vLV-nVBq7PYOjmEoFvihmvg#scrollTo=GnAVUF0mIJDy)

[**Dataset information**](http://bit.ly/DSCoreAutolibDatasetGlossary)

[**Dataset**](http://bit.ly/DSCoreAutolibDataset)