**AUTOLIB DATASET DATA REPORT**

1. **Problem Statement**

Autolib is an electric car sharing company. I would like to examine the blue car claim shown in the dataset above. The main purpose of this study is to identify interesting periods in the dataset by sampling. Then run the hypothesis test according to the claim. The main question is whether the total of blue cars in the area of ​​zip code 75015 is different from that of zip code 75017.

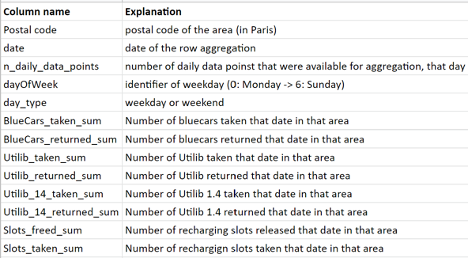
1. **Data Description**

Autolib was an electric car sharing company in existence between the years 2011 and 2018. They mainly dealt in three types of cars: blue cars, Utilib cars and Utilib 1.4 cars. Blue car was the most popular type and they were accessible in a number of cities and postal codes across France. The renters could pick at one station and return them to a station closer to their destination, our dataset provides information about the company’s activities between January 2018 and June 2018.

My null hypothesis is that the number of blue cars taken from postal code 75015 is same to the postal code 75017 during the weekdays. My alternative hypothesis is that the number of Blue cars taken from postal code 75015 is not same as that one in postal code 75017 during the weekdays. I then decided to use the Z-score to test my hypothesis and chose a representative from each postal region for the analysis. I also tested the hypothesis at a 95% confidence level.

1. **Hypothesis Testing Procedure**

The dataset used in this study consists of electric vehicle usage records in France from January to June 2018, taken from opendataparis.com. Given the null hypothesis and alternative hypothesis above, an important variable is the number of blue cars picked up. This exact attribute is accessible in a separate zip code dataset for each day of the week from January 1, 2018 to June 19, 2018. Below is a list of all the attributes of the record and their descriptions.



1. **Hypothesis Testing Results**

I calculated both z statistic and the p-value to execute the z-test. The two sampled z-test computed the p-value as 5.6794344568043504e-94which is less than 0.05 significantly. This means that the average number of blue cars returned are not the same as the number of blue cars taken and so we reject the null hypothesis. The estimated z statistic is -20.48 when the crucial value expected is -1.94. The data points are 20.48 standard deviations away from the mean, more reason to reject the null hypothesis.

1. **Discussion of Test Sensitivity**

Increasing the sample size and effect size can increase the power of your experiment. The study becomes more meaningful if all other variables remain constant and the effect and sample size are large.

A link to my GitHub can be found below [GitHub](https://github.com/stogaja/Python-ML-Projects/tree/main/Autolib)

1. **Summary and Conclusions**

The null hypothesis was rejected as a result of the development of null and alternative hypotheses, the implementation of sampling procedures, and the success of hypothesis testing. We can conclude that the number of blue cars received from the zip code of 75015 is not the same as the zip code of 75017 on weekdays. Poor test performance is an issue that needs to be considered in more detail. Large fluctuations in the target population may have affected test statistics and increased the likelihood of type I errors.