# 1. Problem Statement

The research question of this study is as follows: What day is most popular for people to take bluecars? There is a general assumption that people usually take bluecars on the start of the weekend, technically which is Friday. Monday is expected to be generally a slow day for the electric car sharing company. We need to understand between the two days, which one is more popular for taking blue cars to test the theory.

The null hypothesis is the amount of bluecars taken on Friday is less or equal to that taken on Monday. In this case the claim and alternative hypothesis is that the amount of bluecars taken on Friday is more than that taken on Mondays as it is expected that usage is more during weekends.

The hypothesis is interesting as there is a general assumption that electric car sharing services usually have people taking the cars at the start of the weekend, in this case bluecars on Friday and returned on Monday. However, there is no sufficient evidence showing that Friday is the peak day for taking cars and Monday a slower day. The test will show if the claim is true or there is more to it.

# 2. Data Description

The random variable in this case is the number of bluecars taken. The data has records of the number of bluecars taken over different postal codes in a year (2018). Using the data, we can get the general behavior of people in terms of when they take cars most, least, and the most popular postal codes.

The source of the data is a secondary source (dataset) from the electric car sharing company. A sample was used from the dataset to better test the hypothesis. The sampling method used was stratified random sampling taking samples randomly from two strata, Monday and Friday. 10% of the strata was used as the sample size bringing a total of 464 values. The right proportions were used in the stratification.

# 3. Hypothesis Testing Procedure

The hypothesis will be tested using a sample of the whole population. This will be done by creating a stratum of Monday and Friday from the dataset and then doing a simple random sampling procedure from the strata to get a sample size.

The rationale for the null and alternate hypothesis is to challenge a common claim. The claim is that the amount of bluecars taken on Friday is more than that taken on Mondays as it is expected that usage is more during weekends.

The test statistic used is the z-test. The sample size was 464 meaning there are more than 30-items in the sample. With that, the t-test was eliminated. The other assumptions for a z-test are the following:

* Data points should be independent from each other. i.e, one data point isn’t related or doesn’t affect another data point.
* Data should be normally distributed. However, for large sample sizes (over 30) this doesn’t always matter.
* Data should be randomly selected from a population, where each item has an equal chance of being selected.
* Sample sizes should be equal if at all possible.

The assumptions were met, apart from the last one as it was not attainable.

Based on the alternate hypothesis, the test is one-tailed. For that, we will use an alpha level of 0.05. We use 0.05 as the alpha because the alpha level is the probability of making a Type I error, it seems to make sense that we make this area as tiny as possible.

# 4. Hypothesis Testing Results

* The results of the test lead us to fail to reject the null hypothesis. The z-statistic was 0.76026 and the p-value of 0.224 is greater than the alpha of 0.05.
* The point estimate of the parameter was 122 and the population mean was 116. The standard deviation was 170.

# 5. Discussion of Test Sensitivity

Sensitivity testing was not conducted as this is not a machine learning model.

# 6. Summary and Conclusions

The null and alternative hypothesis were as follows:

From that, we used stratified sampling to test the hypothesis. The alpha level used was 0.05 and the p-value was 0.224 meaning we fail to reject the null hypothesis.

From the results, we find that the sum of bluecars taken on Friday is less or equal to the sum of bluecars taken on Monday. That means that the claim people have about bluecars being taken more on Friday for weekend usage is false. This creates a basis of identifying more trends about the taking and returning of bluecars based on the day of the week. The findings and hypothesis testing were done using Python programming using Jupyter notebook and all the details can be found on my GitHub repository if the project. [[Here](https://github.com/Kibuye24/Joshua_Kibuye_Core_Week_4_IP)]