**Analysis of AutoLib Electric Car Sharing Company**

1. **Problem Statement**

Autolib is an electric car-sharing company. As a data scientist, you have been tasked to research the claim that clients who take blue cars on the weekday will return them during the end of the weekday(Friday).  
The data used for the study were:

* [AutolibDataset](http://bit.ly/DSCoreAutolibDataset)
  + Contains 13 columns and 16085 rows:- after cleaning we will remain with 6 columns necessary for the study.
* [Glossary](http://bit.ly/DSCoreAutolibDatasetGlossary)
  + Contains description of column head meanings.

The following are our hypotheses:

1. Null hypothesis: The number of blue cars taken during the weekday is equal to the number of blue cars returned during the weekday.
2. Alternative hypothesis: The number of blue cars taken during the weekday is not equal to the number of blue cars returned during the weekend

This is important to understand the weekday business of hiring blue cars which is the most common car to be hired, if the hypothesis is true then we can make an assumption that the weekend business is not affected by the weekday business.

1. **Data Description**The data we will use for the research contains the following tables:
   1. Postal code- displays information about different locations
   2. Date- Day the car was taken or returned
   3. Day of the week- Indicates what day of the week it is. 0 is Monday and 6 is Sunday
   4. Day type- either a weekend or weekday
   5. Blue cars taken sum- number of blue cars taken on that particular date
   6. Blue cars returned sum- number of blue cars on that particular date

The Data cleaning procedures include:

* Checking for outliers-
  + Do a boxplot of the desired columns i.e blue cars taken and blue cars returned
* Check for null values
* Replacing the inconsistent data with the 75% quartile values.
* Performing univariate and multivariate analysis

1. **Hypothesis testing procedures**
   1. Since we have two sample data, we will use a t-test to carry out our hypothesis testing.  
      This is made possible by the scipy library to investigate the independent data.
   2. Our null and alternative hypotheses are interesting to us because they show the difference between blue cars taken versus returned during the weekday. This will in turn help us understand if clients like sharing a car during the weekday and returning them before the weekdays’ end.
   3. A T-test will be used but because we are working with the mean values of our sample data. The assumption necessary for the test statistic were met.
   4. The alpha leveled used is 5% (i.e 0.05)
2. **Hypothesis testing and results**

**Results:**

The test results found were as follows:

* Test statistics(t-stat) = -0.2966815419171609
* P\_value was found to be 0.7667123192859796

Since the p\_value was found to be higher than the alpha, the null hypothesis was accepted

A confidence interval of 62.145 was obtained.

1. **Discussion of Test Sensitivity**
   1. Since our null hypothesis was accepted, we can say that a type 1 error occurred. (1-α)
2. **Summaries and Conclusion**

The project process includes:-

* Importing relevant data
* Cleaning our data for better results
* Performing various univariate and multivariate analysis
* Formulating null and alternative hypothesis
* Proving our assumptions.

In conclusion, it was found out that clients who like sharing the blue car from Monday to Friday are likely to return the same car during the specified time gap.