**INVESTIGATING THE USAGE OF BLUE CARS FOR THE AUTOLIB NETWORK**

**Problem Statement**

The dataset and glossary I used for this project can be found here [<http://bit.ly/DSCoreAutolibDataset>] and here [[Link]](http://bit.ly/DSCoreAutolibDatasetGlossary) respectively. The provided dataset is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging).

The random variable under investigation was the sum of Blue Cars taken from various postal codes. The study focused on determining if the mean sum of Blue Cars taken were similar in the postal codes 75015 and 75017 respectively.

**Null Hypothesis**: Blue Cars taken from postal code 75015 is equal to the number of cars taken from postal code 75017.

**Alternative Hypothesis**: Blue Cars taken from postal code 75015 is not equal to the number of cars taken from postal code 75017.

It is important to investigate this hypothesis because this better pushes the expansion and profitization of the autolib network by checking the trends that dictate the consumption patterns of their client base.

**Data Description**

The datafile contains **16,085** rows and **13** columns.

Our data contains the following variables;

1. Postal code
2. date
3. n\_daily\_data\_points
4. dayOfWeek
5. day\_type
6. BlueCars\_taken\_sum
7. BlueCars\_returned\_sum
8. Utilib\_taken\_sum
9. Utilib\_returned\_sum
10. Utilib\_14\_taken\_sum
11. Utilib\_14\_returned\_sum
12. Slots\_freed\_sum
13. Slots\_taken\_sum

Variable definitions are as follows;

1. **Postal code**-postal code of the area (in Paris)
2. **Date**-date of the row aggregation
3. **n\_daily\_data\_points**-number of daily data poinst that were available
4. **DayOfWeek**-identifier of weekday (0: Monday -> 6: Sunday)
5. **Day\_type**-weekday or weekend
6. **BlueCars\_taken\_sum**-Number of bluecars taken that date in that area
7. **BlueCars\_returned\_sum**-Number of bluecars returned that date in that area
8. **Utilib\_taken\_sum**-Number of Utilib taken that date in that area
9. **Utilib\_returned\_sum**-Number of Utilib returned that date in that area
10. **Utilib\_14\_taken\_sum**-Number of Utilib 1.4 taken that date in that area
11. **Utilib\_14\_returned\_sum**-Number of Utilib 1.4 returned that date in that area
12. **Slots\_freed\_sum**-Number of recharging slots released that date in that area
13. **Slots\_taken\_sum**-Number of rechargign slots taken that date in that area

The categorical variables include the type of day (weekday/weekend). The rest of the columns contain numerical variables. The variable being tested is the sum of the blue cars taken in the postal codes 75015 and 75017 respectively.

**Hypothesis Testing Procedure**

The following procedure was taken in hypothesis testing;

1. Sampling
2. Formulating the null and alternative hypotheses
3. Deciding on a significance level (5%)
4. Calculating the z-score and the corresponding p-value
5. Drawing a conclusion based on the result of the test

Given the sample size will be greater than 30, hence a z-test will be ideal. However, if my sample size is less than 30, then I will resort to using a t-test. The alpha level I will use is 0.05 (confidence level of 95%) because it is the standard alpha level used

**Hypothesis Testing Results**

| **z-score** | 0.3326 |
| --- | --- |
| **p-value** | 0.739 |
| **Reject or accept** | Fail to reject the null hypothesis |

Our P-value 0.739 is greater than 0.05 significance level. Therefore the distribution is normal and we fail to reject the null hypothesis.

* population mean of **125.926951**
* a sample size of **2500**
* sample mean of **127.1792**
* sample std of **188.27865**
* calculated z-score of **0.3326**
* Andsignificance level of **0.05**

Our calculated p-value=**0.739**

**Summary and Conclusions**

The autolib analysis should aim to give insights that steer major decisions in the company, for instance maximizing profits and customer satisfaction and support.

From the hypothesis test, it was observed that the p-value was higher than the set significance level of 0.05, therefore there is no sufficient evidence to show that Blue Cars taken from postal code 75015 is equal to the number of cars taken from postal code 75017.

The full analysis can be found [here](https://colab.research.google.com/drive/1wzd-HtpINDaq6r-E39Rj9udbrjztxetM?usp=sharing).