REPORT ON HYPOTHESIS TESTING

Introduction

We have been tasked to understand electric car usage by solving a research question. We will work as a Data Scientist for the Autolib electric car-sharing service company to investigate a claim about the blue cars from the provided Autolib dataset.

Problem Statement

- The data to be used has been provided through the following link; "http://bit.ly/DSCoreAutolibDataset". A glossary for the same dataset has also been provided separately. This study will be looking to investigate if there is no difference in the number of Blue cars and Utilib cars picked over the weekend

- Our null and alternative hypotheses are as below.

  Null hypothesis: The Average number of Bluecars taken in postal code 75001 over weekdays is not different from average number of BlueCars taken in postal code 75002

  Alternative hypothesis: The Average number of Bluecars taken in postal code 75001 over weekdays is different from average number of Bluecars taken in postal code 75002

- We are interested in this hypothesis because we would like to establish if there is consumer preference for Blue cars based on location.

Data Description

- The data provided for this study is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging).

Hypothesis Testing Procedure

- We use a Z Test since our sample size is > 30

. We additionally assume that our data is normally distributed and the variables are independent of each other

- We are interested in this test as we'd like to understand consumer preference for Blue cars based on location. .

- Our test statistic will be the average number of cars picked during weekdays.

- We will use an Alpha level of 0.25 since this will be a two tailed test.

Hypothesis Testing Results

- From our test, the p-value result was 5.628557826570973e-33 which was lower than our alpha value of 0.025. We then rejected the null hypothesis.

Discussion of Test Sensitivity

- A test sensitivity was not carried out for this study

Summary and Conclusions

- Since the calculated pvalue was less than our level of significance, we rejected the null and concluded that There was a statistically significant difference in the average number of blue cars picked in postal codes 75001 and 75002 over the weekdays.