Analysis of AutoLib Electric Car Sharing Company

Problem Statement

Autolib is an electric car-sharing company which maintains a fleet of all-electric Bolloré Bluecars for public use on a paid subscription basis, employing a citywide network of parking and charging stations. The task as the hired Data Scientist is to research the claim that there is a difference between the number of Bluecars taken on weekdays from those taken over weekends.

The data used for the study was **AutolibDataset** which consists of 16085 rows and 13 columns. Cleaning the data was done and unwanted columns were dropped in the process, leaving only 6 relevant columns to use for the analysis.

Hypotheses were created before the study to find the answer to the question and to anticipate the results to look for. Having a hypothesis is important as it helps to understand the weekday business of hiring Bluecars. If the hypothesis is true, then we can make an assumption that the weekend business is not affected by the weekday business.

The following were the hypotheses:

Null Hypothesis: The mean number of Bluecars taken over the weekend is the same as the mean number of Bluecars taken on weekdays

Alternative Hypothesis: The mean number of Bluecars taken over the weekend is greater than the mean number of Bluecars taken on weekdays.

Data Description

The data contains the following tables as described :

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 following data cleaning procedures were used:

1. Checking for appropriate data types
2. Standardizing the format of the column names
3. Renaming the column names
4. Checking for null values
5. Dropping duplicates
6. Checking for outliers and plotting boxplots of the desired columns
7. Replacing the inconsistent data with quartile values.
8. Performing univariate and bivariate analysis

Hypothesis testing procedures

1. The interest was comparing the means of two samples by testing them to reveal if the two means are statistically different from each other. To achieve this, the necessary libraries were imported.
2. Hypotheses were created and the logic behind these null and alternative hypotheses was to find the answer to the question and to anticipate the results to look for. Having hypotheses in this study was important as it helped to understand the daily business of hiring Bluecars.
3. To compare the means of two samples, a t-test was used to tell if the two means were statistically different from each other. This was the suitable test for our study.
4. The alpha leveled used is 5% or 0.05

Hypothesis testing and results

The test results found were as follows:

- Test statistics(t-stat) = -58.21448543848224
- P_value = 0.0
- Critical Value = 1.6458149541902078

Since the p_value was found to be lower than the alpha(0.05), the null hypothesis was rejected.

Discussion of Test Sensitivity

Since our null hypothesis was rejected, we can say that a type 2 error occurred. (1-β)

Summaries and Conclusion

The project process were as follows:

- Importing relevant data
- Cleaning data for better results
- Performing various univariate and bivariate analysis
- Formulating null and alternative hypothesis
- Proving the assumptions.

It was concluded that the mean number of Bluecars taken over the weekend was greater than the mean number of Bluecars taken on weekdays.