Relation between daily average temperature and number of daily cyclists in Faria Lima's cycle path in the city of Sao Paulo in Brazil

Mascarenhas Alexandre[1]
[1]University of Tukuba, Tsukuba – Ibaraki, Japan
e-mail: mascarenhasav@gmail.com

***Abstract* – The objective of this paper it is to show how is the relation of the average temperature of the day and the number of cyclists riding on the Faria Lima's cycle path in Sao Paulo city in Brazil in this day.**

***Keywords* – Cyclists, Cycle paths**

I. INTRODUCTION

The big cities, such as city of Sao Paulo in Brazil have been constantly faced with problems related to the number of cars on the streets. This shows the needs in encourage other modes of transport. In the last ten years, the number of cyclist in Sao Paulo city has increased sharply. This increase has several reasons, such as public policies to encourage the use of alternatives ways of locomotion, for example, the bike.

II. ORGANIZATION OF THE PAPER

This paper will be organized like follows:

- How to obtain datasets: Will be described how to obtain the datasets used in the experiment;
- Pre-processing of the data: Usually a dataset has some problems in the data, such as data missing and outliers. Will be described how was made the treatment of this issues;
- Analaysis and presentation of the data: What results were obtained with this datas;
- Conclusion

III. EXPERIMENT

A. Obtaining the data

The data used in this experiment are all public and were obtained in the website of responsible institution. They are composed of two datasets. One containing the daily number of cyclists passing in the counter installed on the cycle path of Faria Lima Avenue in the city of Sao Paulo. Other containing, among other data, the daily average temperature of the city. Both datasets are files with extension .CSV that it is a tabular form of organise the data, with the columns separated by comma.

The way to get each of the dataset is as follows:

1) Dataset of the number of cyclists: The dataset of the number of cyclists in the Faria Lima's cycle path it is available on the page `https://data.eco-counter.com/public2/?id=100027495`. In this page it is possible select the desired period, since the start of counting on 18 January of 2016 to the current day. After select the period just click in button "export to csv" to download the .CSV file.

2) Dataset of the temperature: The dataset of temperature of the city of Sao Paulo it is available on the page `https://bdmep.inmet.gov.br/`. On the page it is necessary click in the "prosseguir", on the next page it is asked to enter an email address which one the dataset will be sent. After this, some settings must be made to set which variables will be in the file as well as the period from which you want the data. Once all the settings have been set, the file will be sent to the email.

3) Both Datasets: Other way to obtain the datasets used in this experiment is through the link `https://www.github.com/mascarenhasav/master/tree/main/courses/experimental-design-in-computer-science/report-1`, where the files of number of cyclists and temperature were named respectively as "faria-lima.csv" and "temperature-sp.csv".

B. Pre-processing

Before effectively starting the data analysis, it will be make a pre-processing of the datas. This is important, because allows to identify some issues with the dataset, such as, missing data, input error, incompatibility. For the purpose of this article, the period that will be considered in data analysis is from 1 January 2018 to 1 January 2020. This choice was made mainly to avoid abrupt behavior caused by the Covid 19. Once the data period is selected and there is no issues with then, it becomes possible to do some analysis.

C. Analysis and presentation

There are a number of possible approaches to do the relation of average daily temperature and the daily number of cyclists. Naturally, many factors can influence the numbers, such as the day of the week, if it is a weekend, a holiday, and if it is raining or not. As a first analysis, will be make a relation without any of this considerations, which would allow to see if the temperature is one of the major factor.

IV. OTHER INSTRUCTIONS

V. CONCLUSIONS

ACKNOWLEDGEMENTS

BIOGRAPHIES