REPORT ON BIKE RENTAL SERVICES

Analysis

Chart, histogram

Description automatically generated

**Target Variable**

As we can see in the following figure, the target variable presents a skewed distribution. A way of tackling this regression problem could be by using a Poisson distribution.

Graphical user interface

Description automatically generated

Histograms of all the variables, no of hours, humidity, are higher than other factors

Diagram, box and whisker chart

Description automatically generated with medium confidence

Observing linear correlations, we see that there are several variables affecting the target, such as Year, Hour, Temp/Atemp and humidity. This will be interesting for building the baseline model.

A picture containing chart

Description automatically generated

**Year Variable**

We see that we have balanced entries in the edge of year. Also that the year distributions for the different targets are slightly different in the first (0-200) range, we could benefit from this for our predictions.

Chart, box and whisker chart

Description automatically generated

**Season**

In the following boxplot, we observe some outliers in the Season=1, which is Winter. We will look to those entries in special.

We see there are 79 rows in Winter that have more than 400 bikes rented. It could be because of Christmas Holidays or other celebrations; we can check that by using the holiday variable. Also we observe the subset distributions. After doing that check, we see it does not seems to exist any plausible reason for that high rent values and we decide to drop that observations in the training phase to avoid scaling effects.

**Algorithm**

The algorithm, we used is k-mean clustering which helps to form groups which have same type of characteristics in this algorithm there are 3 types of methods in which we are using only 2 methods to compare the cluster values.

The 3 methods are elbow method, silhouette method and gap statistic method

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

* In both the methods the k value is 2
* The range is usually (0,1) in elbow method and (-1,1) in silhouette method.
* The point k=2 is near to 0 which is the peak point and from there it keeps on decreasing.

IMPORTANT FACTORS EFFECTING BUSINESS

Text

Description automatically generated

* Bike rental prediction depends on whether condition, working day or a holiday and season
* As a biker rental service provider, I should make sure season time and based on weather condition the I should meet the customers requirement, to not to lose customers

**Reference**

<https://www.kaggle.com/code/lakshmi25npathi/bike-rental-count-prediction-using-python/data>