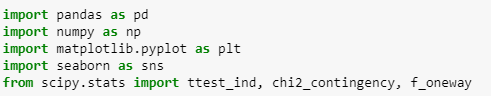
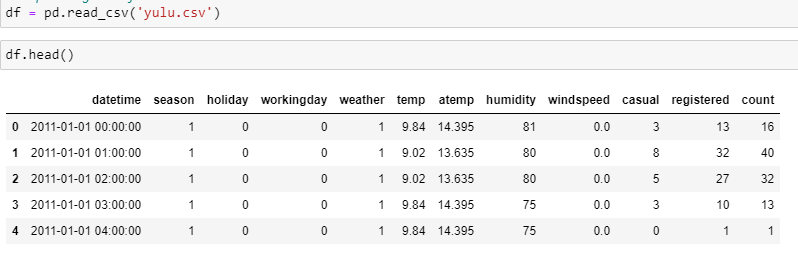
**Name:** Aloysius Lobo

**Batch:** Morning – DSML BeginnerFeb23

**Case-Study: Yulu – Hypothesis Testing**

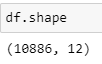
**# Importing all the necessary libraries**

****

**#Importing ‘Yulu’ csv file**

**# Performing basic EDA**

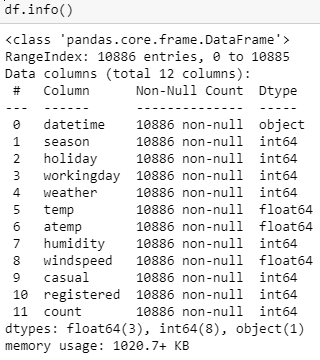
1. **Getting shape of dataset**

****

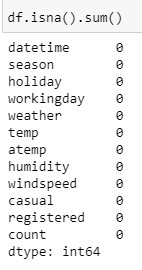
**Observations –**

There are total **10886** rows and **12** columns in the entire dataset.

1. **Getting column info on datatype and non-null value counts**

****

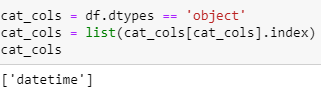
**# Finding missing/null values**

****

**Observations –**

There are no missing values found in the dataset.

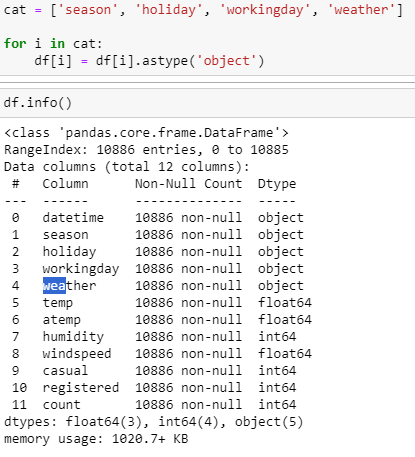
**# Columns containing object/categorical datatype**

****

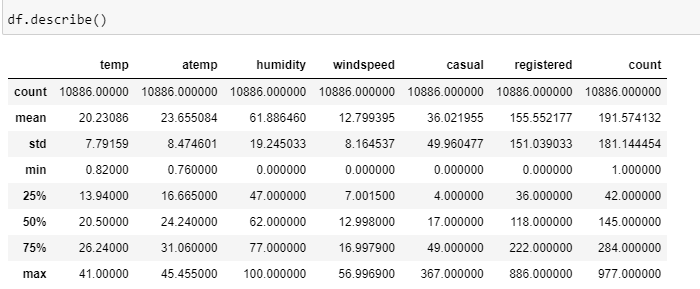
**Observations –**

Only ‘datetime’ column contains object datatype

**# Converting all categorical into ‘object’ datatype**

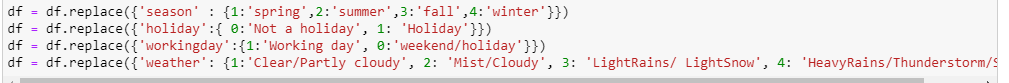


**# Getting statistical info on columns with numerical datatypes**

****

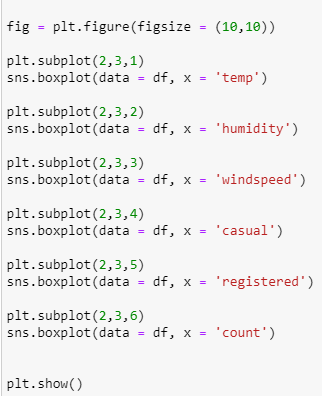
**Observations –**

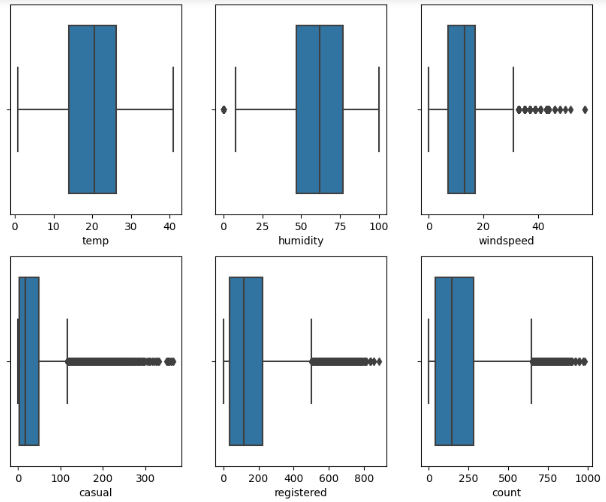
* Count of users renting the bicycle was found to be the maximum when the temperature in the city was 41 degrees Celcius while the count decreased when the temperature was 0.82 degrees Celcius.
* The number of registered customers renting the bicycle is 886 and is more than the non-registered customers which is 367.
* The average humidity is 62 when 50% customers rent the bike.
* The average temperature is 20.5 degrees when 50% customers rent the bike.

**# Replacing row-names with specific names in categorical columns**

|  |  |  |  |
| --- | --- | --- | --- |
| **Seasons** | **Weather** | **Holiday** | **Working day** |
| 1 – Spring | 1 – Clear/Partly cloudy | 0 – Not a holiday | 0 – Weekend/Holiday |
| 2 – Summer | 2 – Mist/Cloudy | 1 – Holiday | 1 – Working day |
| 3 – Fall | 3 – LightRains/ LightSnow | - | - |
| 4 - Winter | 4 – HeavyRains/ Thunderstorms/Snow | - | - |

**# Outlier detection**

****

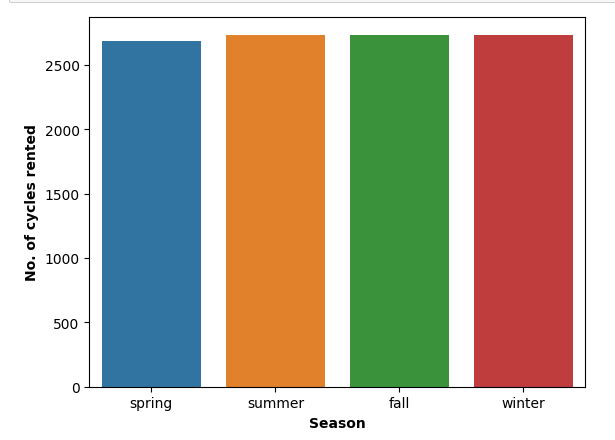


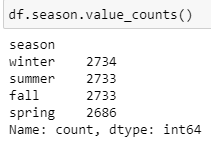
**Observations –**

* Maximum outliers are observed in casual, registered and count, while a few are observed in the windspeed data.
* Few customers do rent the bicycle even when the windspeed is over 30.

**# Univariate Analysis**

**Categorical variables**

1. **Seasons**

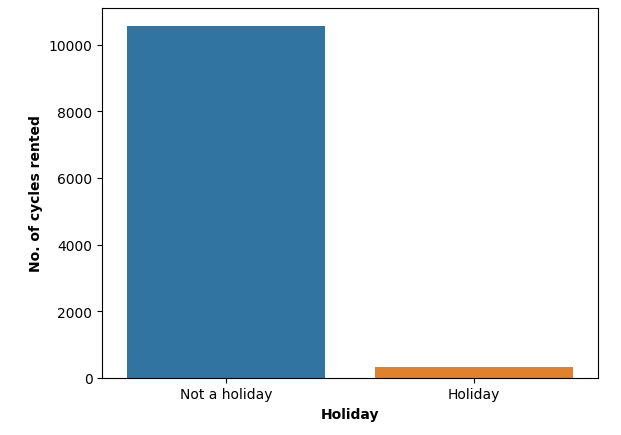
****

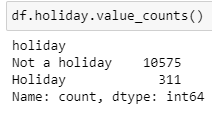
**Observations –**

Above analysis on seasons show that except during spring, the cycles rented were similar in all other seasons.

**Recommendations -**

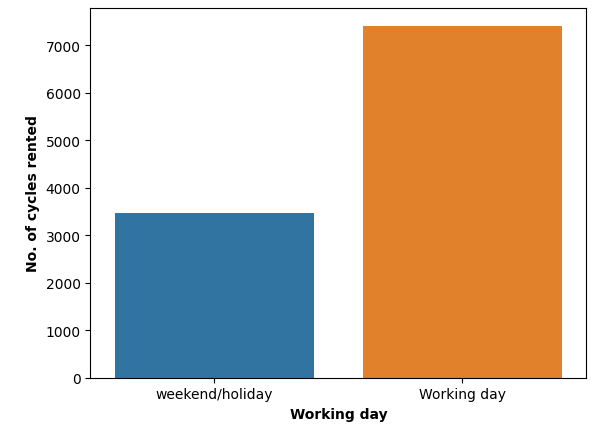
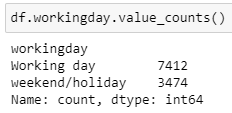
Seasons do affect the renting of Yulu bicycles.

1. **Holiday**

****

**Observations –**

Above analysis on holidays show that maximum cycles were rented when it’s not a holiday.

1. **Working days**

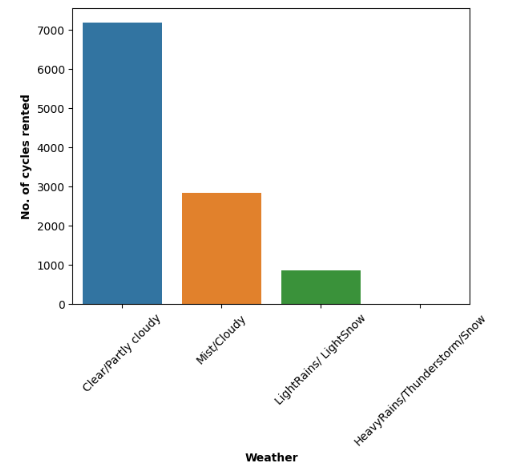
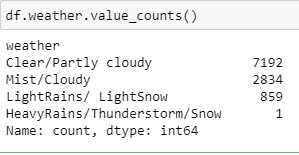
**Observations –**

Above analysis on working days show that maximum cycles were rented during working days and not on weekends/holidays.

**Recommendations -**

From above analysis on Holidays and Working days, it is clear that maximum count of bicylces were rented during the Working days as compared to on Weekends/ Holidays. Hence, Yulu must make provisions to increase the Yulu bikes over working days.

1. **Weather**

****

**Observations –**

* Above analysis show that maximum number of bicyles are rented only when the weather is clear or its partly cloudy.
* During very bad weather like thunderstorms or very heavy snowing, no one has rented Yulu bicyle.

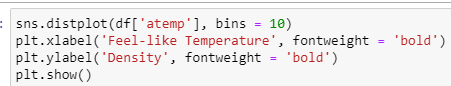
**Recommendations -**

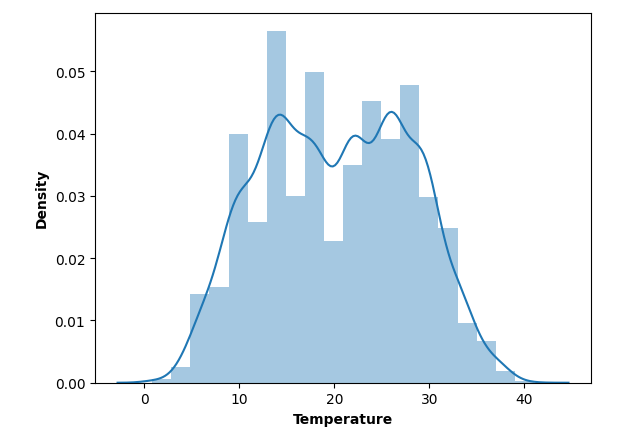
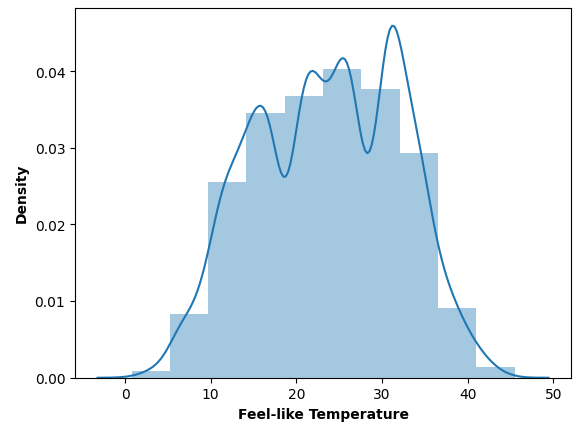
As the analysis show a drop in the bike rents during bad weathers like rains/ thunderstorms. It is thus suggested that Yulu can either drop the rent fares during such bad weather conditions or can introduce a bicycle with some protective covering for snow and rains so that users are comfortable to rent the bikes even in bad weather conditions.

**Numerical variables**

1. **Temp and Feel-like temperature**

****

****

****

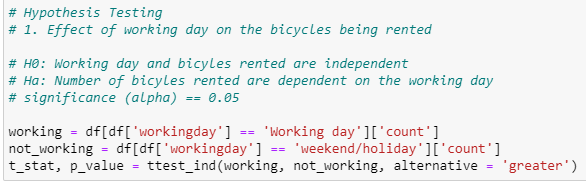
**Observations –**

* Mostly customers have rented bicycle in the temperature range of 15°C to 30°C.
* This is similar to the Feels-like temperature in the range of 15°C to 32°C.
* Some outliers are seen where customers have rented the bike in extreme temperature ranges.

**Recommendations –**

Yulu should increase the number of bikes for rent when the temperature is above 20 degrees to 30 degrees celcius, as the demand for bikes is more.

**# Hypothesis Testing**

1. **Working\_day**

****

**Observations –**

* As p\_value is greater than the significance value (alpha) i.e. (0.113 > 0.05), we conclude that there are no enough evidences to reject the null hypothesis. Hence, Count of bicycles being rented is independent if it is a working day or a holiday.
* Univariate analysis showed that bicycles are rented more on working days than on holidays/weekends but from the hypothesis testing its clear to confirm that with a 95% confidence they are independent of each other.

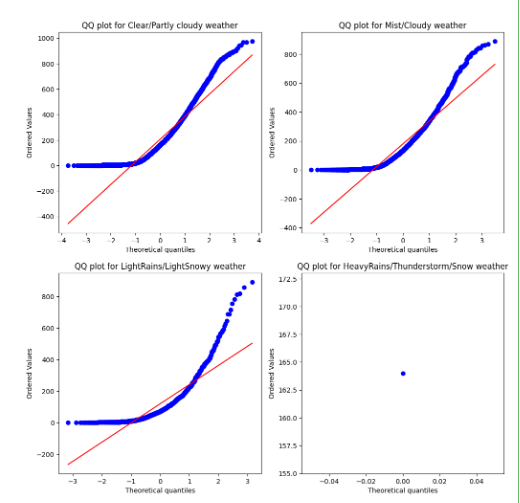
**Recommendations –**

As the count of bikes rented is independent if it’s a working day or a holiday, Yulu can make provisions to increase the bikes for rent any day of the week.

1. **Weather**

**# Testing for assumptions for using ANOVA**

1. **Testing for Normality using QQ Plot**



**Observations –**

It’s clear from the above analysis that **normality does not exist** for the count of bikes rented during various weather conditions.

1. **Testing for Normality using Shapiro-Wilk Test**

**H0: It’s a Gaussian/Normal distribution**

**Ha: It is not a Gaussian/Normal Distribution**

**significance(alpha): 0.05**

****

**Observations –**

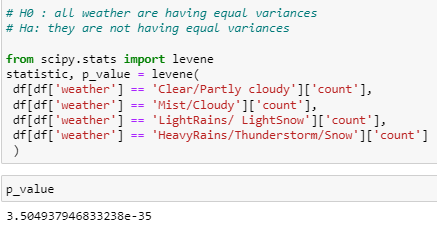
Since, p\_value < alpha(0.05), we can conclude that with enough evidence we reject the Null Hypothesis (H0) i.e. The above distribution is **not a normal distribution**.

**Transforming the data using boxcox transformation and checking if the transformed data follows normal distribution**

**Observations –**

* As p\_value for weather1 (0.0508) > alpha(0.05), we have enough evidence to fail to reject the Null Hypothesis. Hence, weather-1 i.e. Clear/Partly cloudy weather follows Normal Distribution.
* As p\_values for the remaining weather conditions are less than alpha (0.05), we can conclude that with 95% confidence level we reject the null hypothesis. Hence, **it does not follow Normal Distribution.**

1. **Testing for Equal Variance using Levene Test**

****

**Observations –**

As p\_value(3.5049e-35) < alpha(0.05) for a 95% confidence level, **all weather conditions have equal variance.**

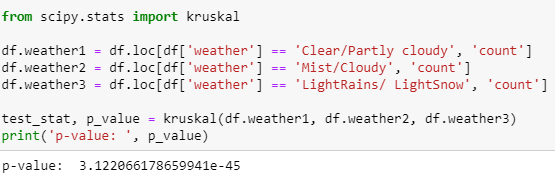
As the assumptions for ANOVA testing are **not satisfied**, we check the dependency of the count of bikes rented with the weather conditions using **Kruskal Wallis hypothesis test**.

**# Kruskal Wallis test**

**H0: Number of bikes rented is independent of weather**

**Ha: Number of bikes rented depends on weather**

**significance (alpha) : 0.05**

****

**Observations –**

As p\_value is very less than the significance value (alpha) i.e. (3.122e-45 < 0.05), we can conclude that with 95% confidence level, there are enough evidences to reject the null hypothesis. Hence, Count of bicycles being rented is dependent on the weather i.e. it is different for different weather.

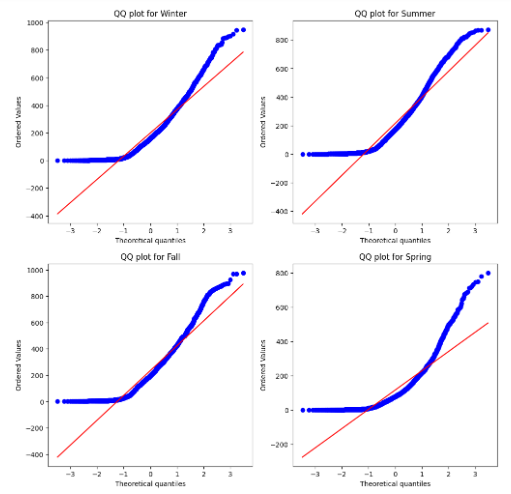
**Recommendations –**

The company should make provisions for bikes to be used in bad weather conditions. Bikes with provisions like protective overhead covering for rains/snow must be incorporated to increase the demands for bikes even during bad weathers.

1. **Seasons**

**# Testing for assumptions of ANOVA**

1. **Testing for Normality using QQ Plot**

****

**Observations -**

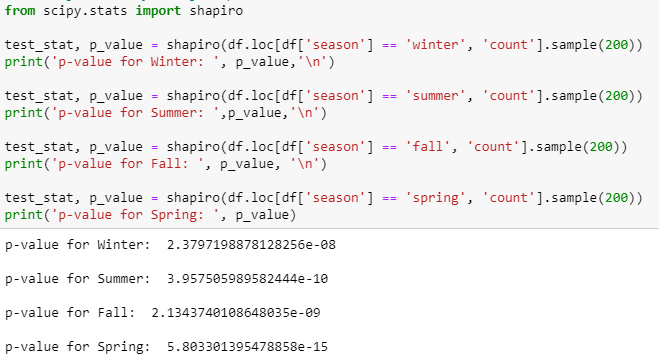
It’s clear from the above analysis that **normality do not exist** for the count of bikes rented during various seasons.

1. **Testing for Normality using Shapiro-Wilk Test**

**H0: It’s a Gaussian/Normal distribution**

**Ha: It is not a Gaussian/Normal Distribution**

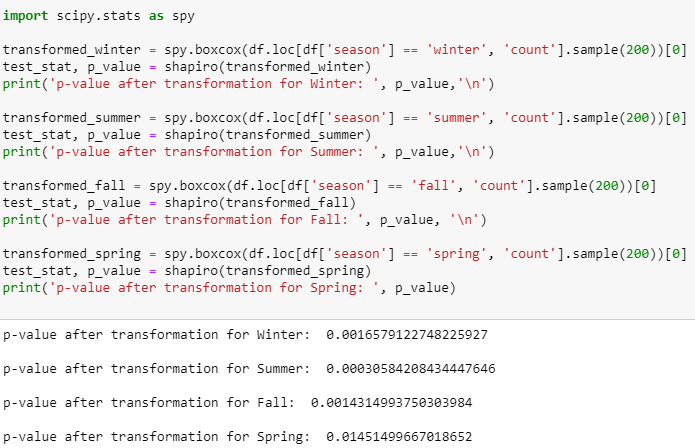
**significance(alpha): 0.05**



**Observations –**

Since, p\_value < alpha(0.05), we can conclude that with enough evidence we reject the Null Hypothesis (H0) i.e. The above distribution is **not a normal distribution**.

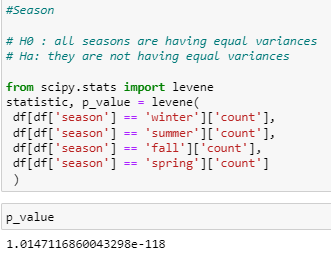
**Transforming the data using boxcox transformation and checking if the transformed data follows normal distribution**

****

**Observations –**

* As p\_values for the every season are less than alpha (0.05), we can conclude that with 95% confidence level we reject the null hypothesis. Hence, it **does not follow Normal Distribution.**

1. **Testing for Equal variance using Levene Test**

****

**Observations -**

As p\_value(1.015e-118) < alpha(0.05) for a 95% confidence level, **all seasons have equal variance.**

As the assumptions for **ANOVA testing are not satisfied**, we check the dependency of the count of bikes rented with the seasons using **Kruskal Wallis hypothesis test**.

**# Kruskal Wallis test**

**H0: Number of bikes rented is independent of weather**

**Ha: Number of bikes rented depends on weather**

**significance (alpha) : 0.05**

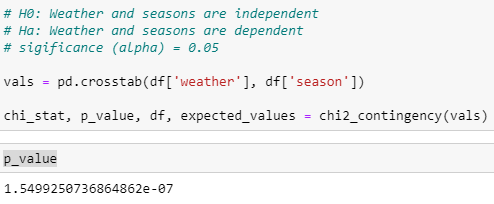
****

**Observations –**

As p\_value is very less than the significance value (alpha) i.e. (2.479e-151 < 0.05), we can conclude that with 95% confidence level, there are enough evidences to reject the null hypothesis. Hence, Count of bicycles being rented is dependent on the seasons i.e. it is different for different seasons.

**Recommendations –**

The company must increase the number of bikes for rent during specific seasons like summer abd fall when the demand for the bikes is high. Discount offers or some special offers can also be given during other seasons when the demand is less, in order to increase the sales.

1. **If weather and seasons are dependent using chi2\_contingency test**

**Observations –**

As p\_value is very less than the significance value (alpha) i.e. (1.549e-07 < 0.05), we conclude

that there are enough evidences to reject the null hypothesis. Hence, weather and seasons are dependent on each other.

**SUMMARY**

### **Insights**

* In **summer** and **fall** seasons more bikes are rented as compared to other seasons.
* Number of bikes rented is independent if it’s a working day or a holiday.
* Whenever there is **rain, thunderstorm, snow or fog**, there were less bikes were rented.
* Whenever the humidity is less than 20, number of bikes rented is very very low.
* Whenever the temperature is less than 10, number of bikes rented is less.
* Whenever the windspeed is greater than 35, number of bikes rented is less.

### **Recommendations**

* In **summer** and **fall** seasons the company should have more bikes in stock to be rented. Because the demand in these seasons is higher as compared to other seasons.
* In very low humid days, company should have less bikes in the stock to be rented.
* Whenever temperature is less than 10 or in very cold days, company should have less bikes.
* For bad weather conditions like rains/thunderstorms, the company should introduce bikes with protective overhead covering, to increase its demand in such weather conditions.