# Bike Sharing System

# Problem Statement

- What is Boombike?

A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short term basis for a price or free. Allows people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it. This bike can then be returned to another dock belonging to the same system.

- Business Goal:
  Model the demand for shared bikes with the available independent variables. Used by the management to understand how exactly the demands vary with different features. They can accordingly manipulate the business strategy to meet the demand levels and meet the customer's expectations. Further, the model will be a good way for management to understand the demand dynamics of a new market.

Requirements :

- **Requirements:**

Which variables are significant to predict the demand of share bikes.

How strongly these variables predict the demand.

- **What we need to do?**

Create a linear model that describes the effect of different features on demand of shared bikes.

Model should be easy to understand and used by the management

# Steps:

Data Visualization

Data Preparation

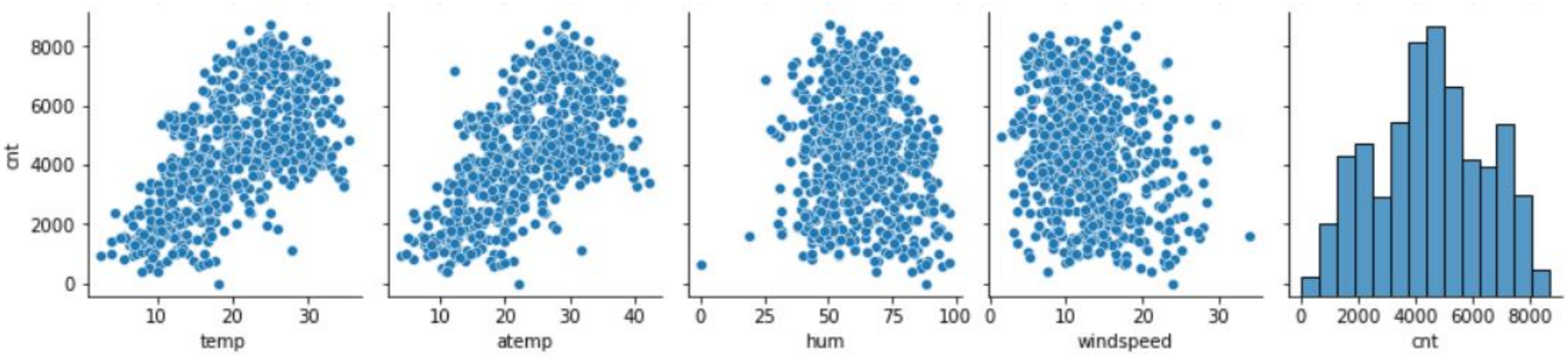
Data modelling and evaluation

# Data Visualization

Sample data :

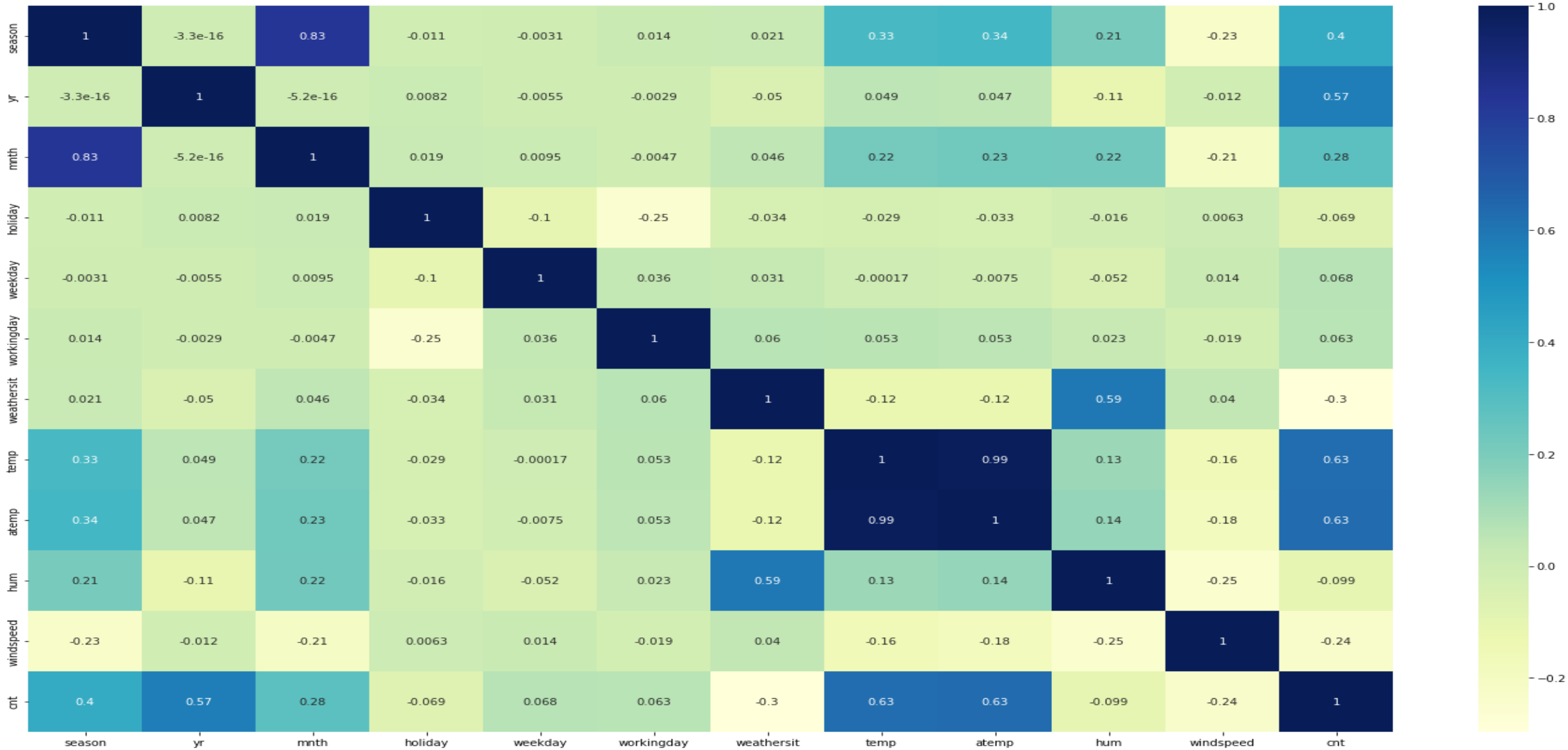
| | instant | dteday | season | yr | mnth | holiday | weekday | workingday | weathersit | temp | atemp | hum | windspeed | casual | registered | cnt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 01-01-2018 | 1 | 0 | 1 | 0 | 6 | 0 | 2 | 14.110847 | 18.18125 | 80.5833 | 10.749882 | 331 | 654 | 985 |
| **1** | 2 | 02-01-2018 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 14.902598 | 17.68695 | 69.6087 | 16.652113 | 131 | 670 | 801 |
| **2** | 3 | 03-01-2018 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 8.050924 | 9.47025 | 43.7273 | 16.636703 | 120 | 1229 | 1349 |
| **3** | 4 | 04-01-2018 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 8.200000 | 10.60610 | 59.0435 | 10.739832 | 108 | 1454 | 1562 |
| **4** | 5 | 05-01-2018 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 9.305237 | 11.46350 | 43.6957 | 12.522300 | 82 | 1518 | 1600 |

We can observe that below columns are not required for our analysis : instant, dteday, casual, registered



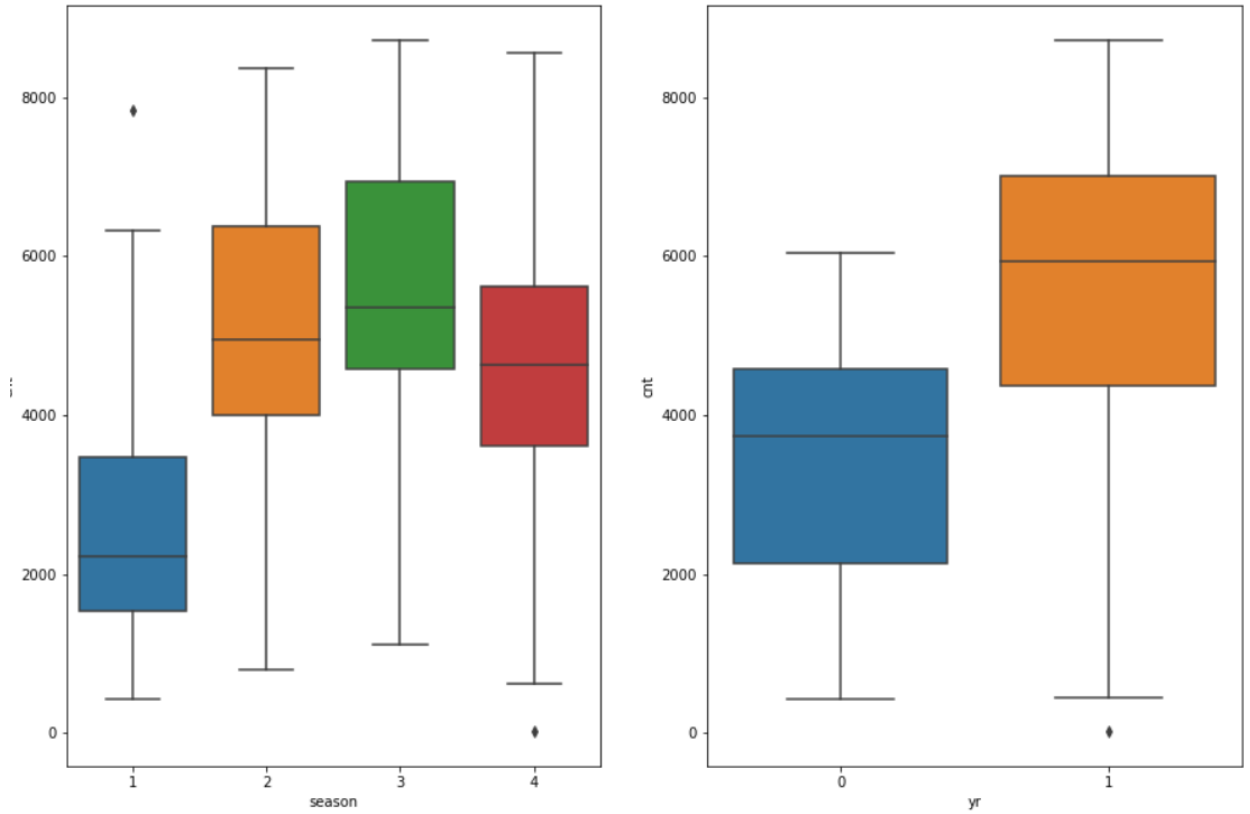

- Numerical Data analysis
- We can observe a linear relationship between temp, atemp and cnt:

# Correlation analysis:

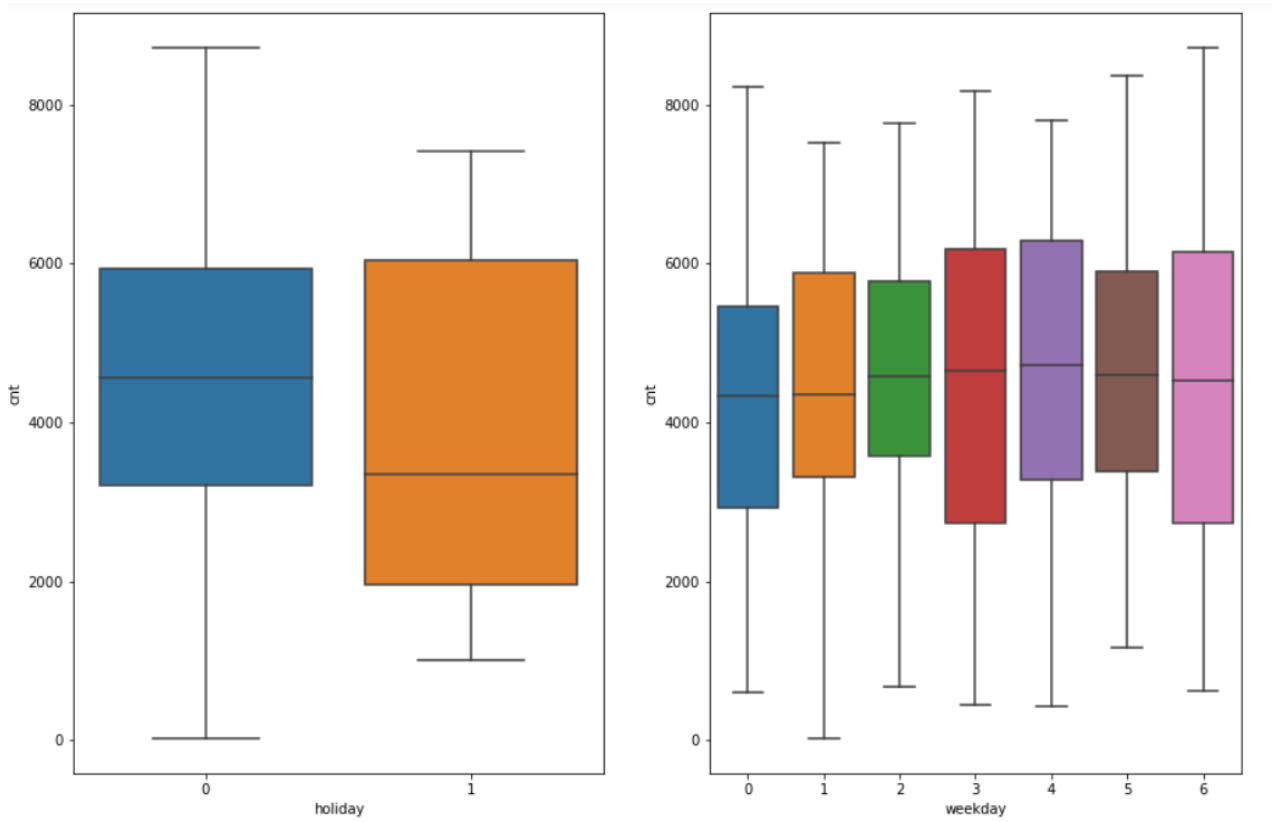


# Categorical data analysis

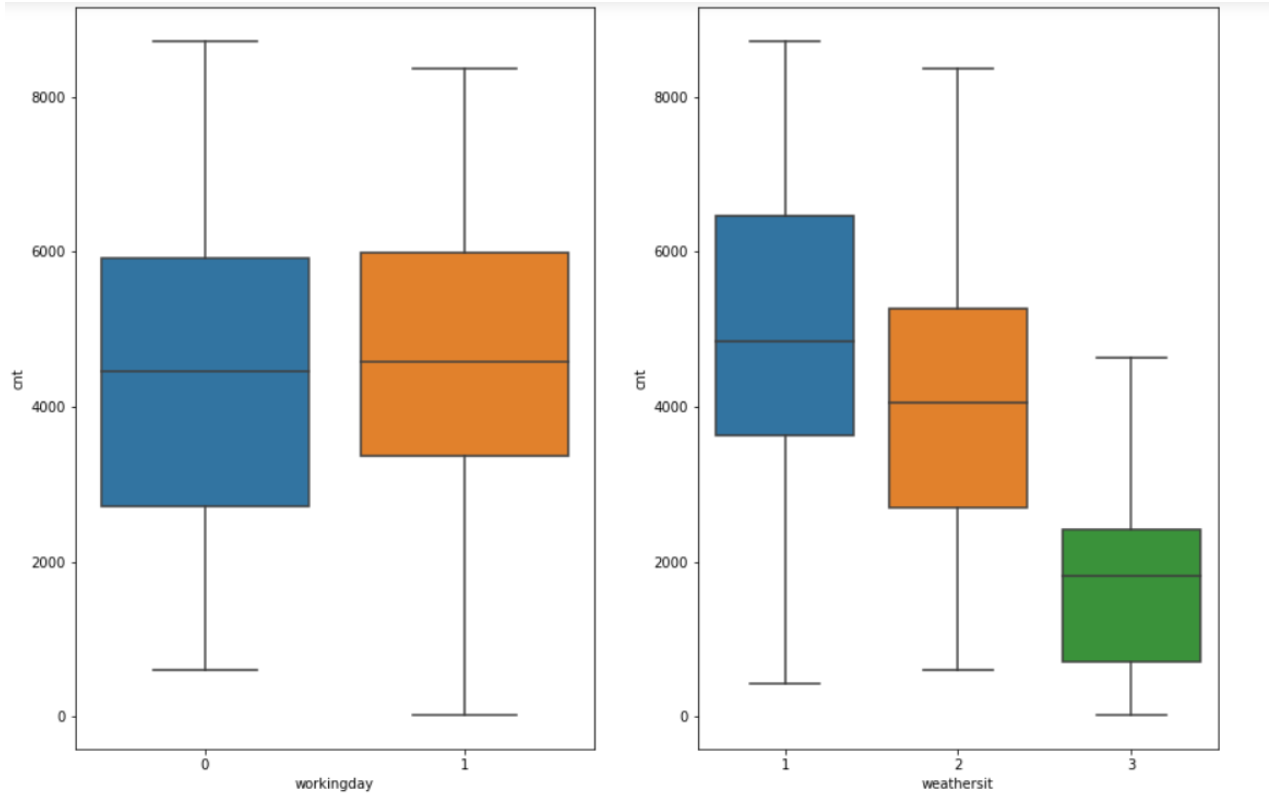
- Observations:
- Season -> Demand of bike increases in 1 2 and 3 , while decrease in 4th season
- Yr(year) -> Demand increased in 2019 as compared to 2018



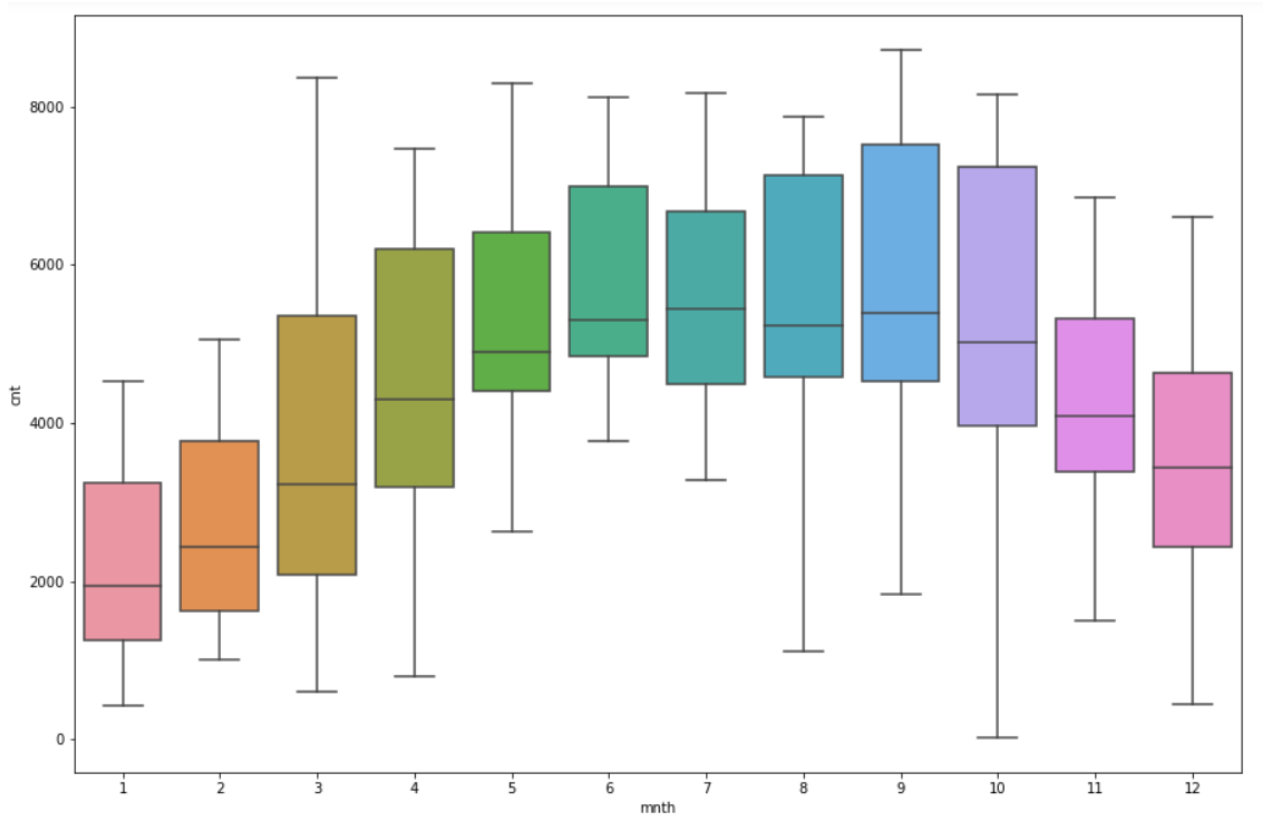
- Observations:
- Holiday -> Median of bike rented on holiday is more as compared to not a holiday
- Weekday -> All most same/constant number of bookings on weekdays. Very close trends of bike booking



- Observations:
- Workingday -> Median of rented bike booking is same on working and non working day
- Weathersit -> Demand of bike booking decreasing  from clear to mist to light snow weather



- Observation :
- An increase in booking of rented bike till 7th month and the it's start falling.



# Data Preparation

- Create one-hot encoding variables for all the categorical features. As ML can deal with only numerical data, we can change all categorical to numerical using one-hot encoding.

Example:

| Salary |
|---|
| Minimum |
| Medium |
| Maximum |

**One-Hot Encoding**

| Min_Salary | Medium_Salary | Max_Salary |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |

- Splitting data

  Split data into train and test set of 70% and 30% ratio respectively

Data Normalization (Scaling)

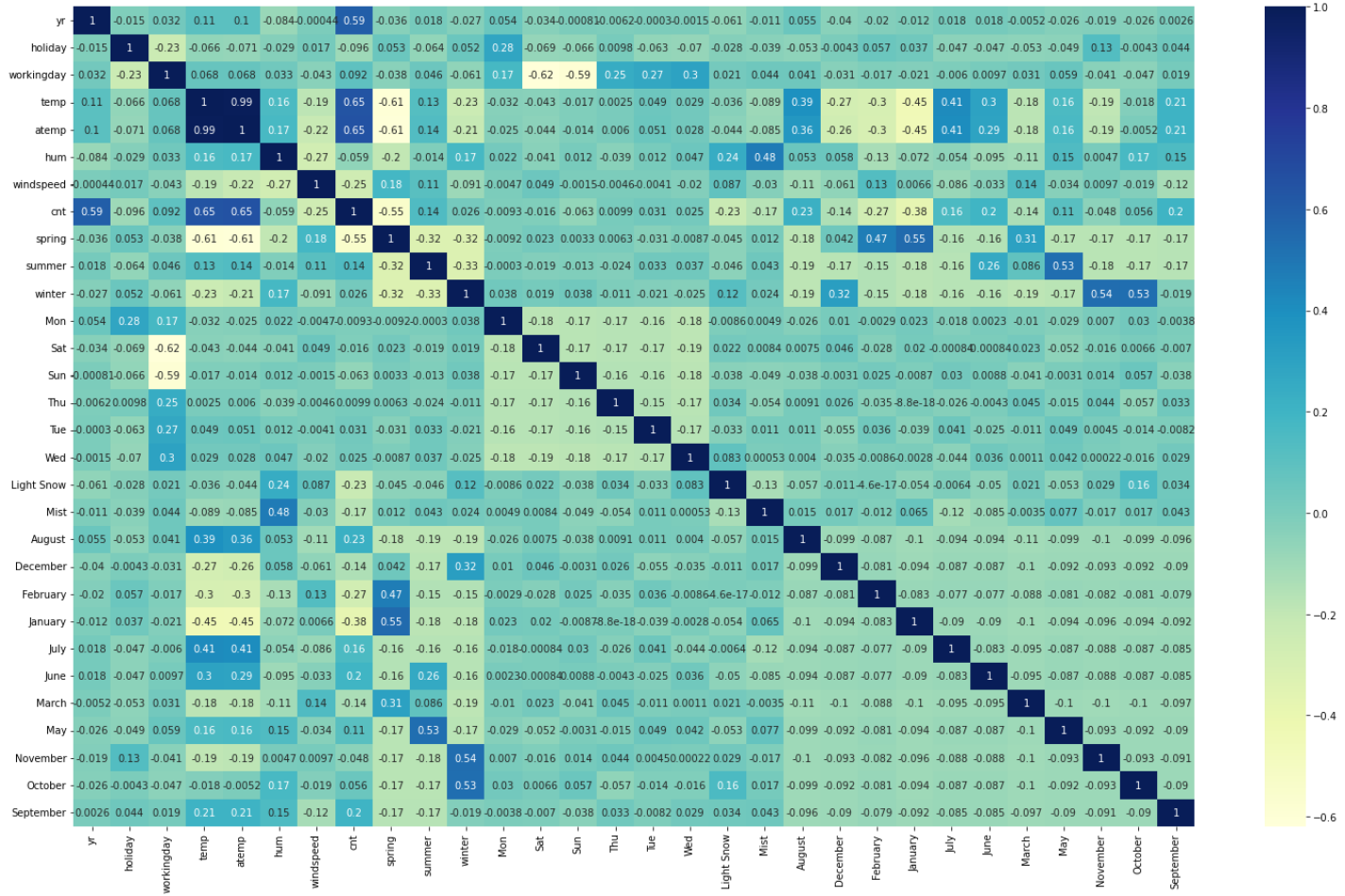
Use MinMaxScaler to normalize the data. So that all numerical falls in range of 0 and 1

Divide Data

Divide data into dependent and independent variables.

High positive correlation between cnt, atemp, temp, yr and season .

Negative correlation with windspeed, hum, weathersit and holiday



# Data Modeling

- Create Linear Regression model using mixed approach (RFE & VIF/p-value).
- Check the various assumptions.
- Check the Adjusted R-Square for both train & Test data.
- Report the final model.

# Final Model
# VIF<5 and p-value <0.05

| | features | VIF |
|---|---|---|
| 0 | const | 51.11 |
| 4 | hum | 1.88 |
| 2 | workingday | 1.65 |
| 8 | Sat | 1.64 |
| 3 | temp | 1.60 |
| 10 | Mist | 1.56 |
| 11 | July | 1.43 |
| 6 | summer | 1.33 |
| 7 | winter | 1.29 |
| 9 | Light Snow | 1.24 |
| 12 | September | 1.19 |
| 5 | windspeed | 1.18 |
| 1 | yr | 1.03 |

| | | | |
|---|---|---|---|
| **Dep. Variable:** | cnt | **R-squared:** | 0.843 |
| **Model:** | OLS | **Adj. R-squared:** | 0.839 |
| **Method:** | Least Squares | **F-statistic:** | 222.7 |
| **Date:** | Wed, 09 Mar 2022 | **Prob (F-statistic):** | 4.14e-191 |
| **Time:** | 16:11:39 | **Log-Likelihood:** | 511.32 |
| **No. Observations:** | 510 | **AIC:** | -996.6 |
| **Df Residuals:** | 497 | **BIC:** | -941.6 |
| **Df Model:** | 12 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 0.1712 | 0.028 | 6.014 | 0.000 | 0.115 | 0.227 |
| **yr** | 0.2286 | 0.008 | 28.267 | 0.000 | 0.213 | 0.244 |
| **workingday** | 0.0524 | 0.011 | 4.791 | 0.000 | 0.031 | 0.074 |
| **temp** | 0.5960 | 0.022 | 26.667 | 0.000 | 0.552 | 0.640 |
| **hum** | -0.1709 | 0.037 | -4.558 | 0.000 | -0.245 | -0.097 |
| **windspeed** | -0.1888 | 0.026 | -7.393 | 0.000 | -0.239 | -0.139 |
| **summer** | 0.0827 | 0.011 | 7.770 | 0.000 | 0.062 | 0.104 |
| **winter** | 0.1355 | 0.010 | 12.930 | 0.000 | 0.115 | 0.156 |
| **Sat** | 0.0625 | 0.014 | 4.429 | 0.000 | 0.035 | 0.090 |
| **Light Snow** | -0.2391 | 0.026 | -9.100 | 0.000 | -0.291 | -0.188 |
| **Mist** | -0.0536 | 0.010 | -5.129 | 0.000 | -0.074 | -0.033 |
| **July** | -0.0439 | 0.018 | -2.450 | 0.015 | -0.079 | -0.009 |
| **September** | 0.0928 | 0.016 | 5.816 | 0.000 | 0.061 | 0.124 |

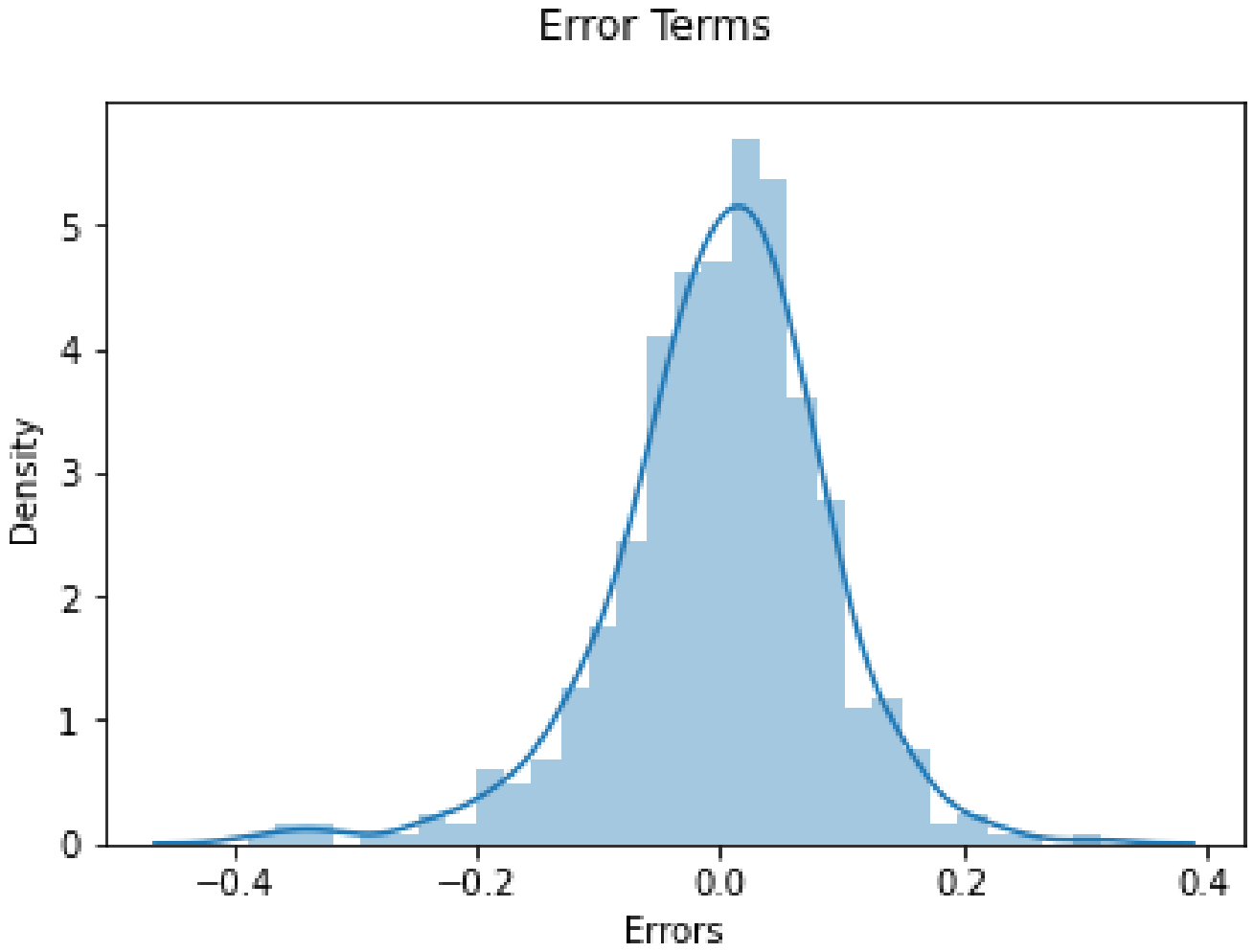
# Final Equation:

- cnt = 0.171244 + (0.228581 x **yr**) + (0.052441 x **workingday**) + (0.595988 x **temp**) - (0.170916 x **hum**) - (0.188772 x **windspeed**) + (0.082670 x **summer**) + (0.135504 x **winter**) + (0.062458 x **sat**) - (0.239137 x **Light Snow**) - (0.053623 x **Mist**) - (0.043937 x **July**) + (0.092832 x **September**)

# Coefficient Interpretation:

- **yr :** (0.228581) A unit increase in yr , increase the bike hire by 0.228581
- workingday : (0.052441) A unit increase in workingday , increase the bike hire by 0.052441
- temp : (0.595988) A unit increase in temp , increase the bike hire by 0.595988
- hum : (-0.170916) A unit increase in hum , decrease the bike hire by 0.170916
- windspeed : (-0.188772) A unit increase in windspeed , decrease the bike hire by 0.188772
- summer : (0.082670) A unit increase in summer , increase the bike hire by 0.082670
- winter : (0.135504) A unit increase in winter , increase the bike hire by 0.135504
- sat : (0.062458) A unit increase in sat, increase the bike hire by 0.062458
- Light Snow : (-0.239137) A unit increase in Light Snow , decrease the bike hire by 0.239137
- Mist : (-0.053623) A unit increase in Mist , decrease the bike hire by 0.053623
- July : (-0.043937) A unit increase in July , decrease the bike hire by 0.043937
- September : (0.092832) A unit increase in September , increase the bike hire by 0.092832

Residual analysis: Error terms normally distributed and mean to zero



# Final Results:

- Train R^2 = 0.839
- Train Adjusted R^2 = 0.843
- Test R^2 = 0.806950460291592 8
- Test Adjusted R^2 = 0.777426842105263
- We can say that model is good

# Final Report: Top 3 predictor variables that influences the bike booking

Temperature (**0.595988**) -: A coefficient value of '**0.595988**' indicated that a unit increase in temp variable increases the bike hire numbers by **0.595988** units

weathersit (**Light Snow: (-0.239137) Mist : (-0.053623) =0.29276):-** A coefficient value of '-**0.29276**' indicated that, w.r.t Weathersit1, a unit increase in Weathersit3 variable decreases the bike hire numbers by **0.29276** units

**Year (yr)** - A coefficient value of '**0.228581**' indicated that a unit increase in yr variable increases the bike hire numbers by **0.228581** units.

- Thank you