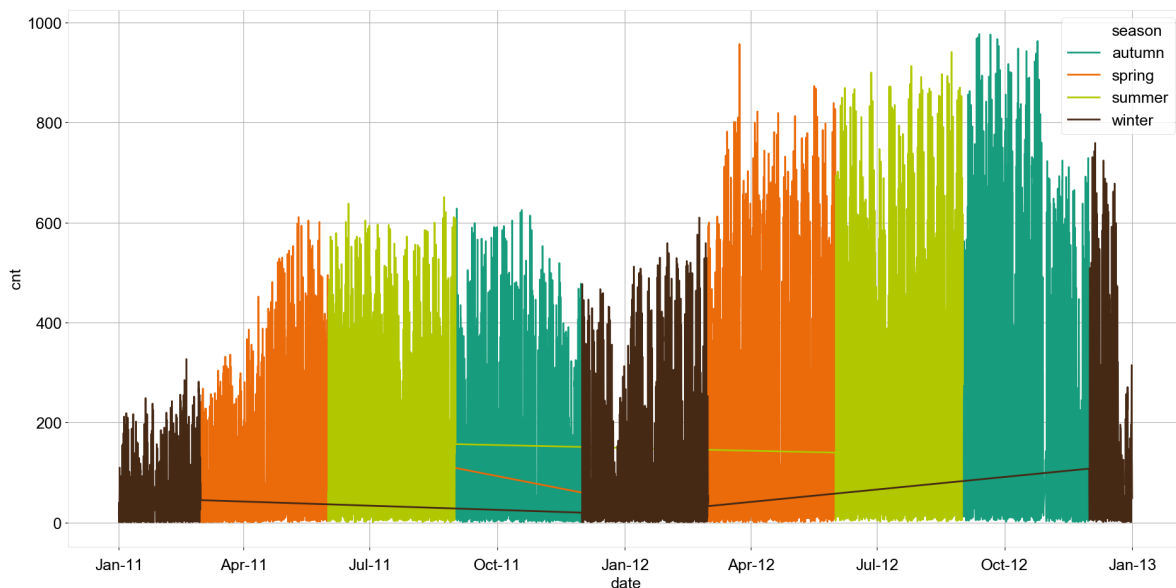# Bike Sharing Dataset Analysis

Data Processing

- Reformatting the columns names and types to be compatible with plotting
- Correcting the season's column

EDA Analysis

- The mean rate of renting bikes has increased from 2011 to 2012 from 144 to 235 bike respectively with a percentage increase of 63%.


- Over all seasons, winter represent the lowest mean rate of renting bike – as expected – with almost 50% reduction compared to other months. Generally, the registered bikes rental rate is always higher than casual bikes. However, the difference between increase of registered rental bikes is the highest with 400% compared to the casual.
- Summer represents the busiest season with a relative increase over Autumn and Spring.
- The influence of weather is substantial on the rental bikes as it reaches a maximum mean value during clear or with few clouds. Also, the misty dry weather has a higher rate than a light rain or snow. This can confirm that Autumn is better for renting than spring.
- What is surprising in Winter, is the variance. The variance in winter can vary significantly ranging between 25 to 150 bikes. This huge gap presents a difficulty in the future for predicting in the future and may lead to a reduction in accuracy.
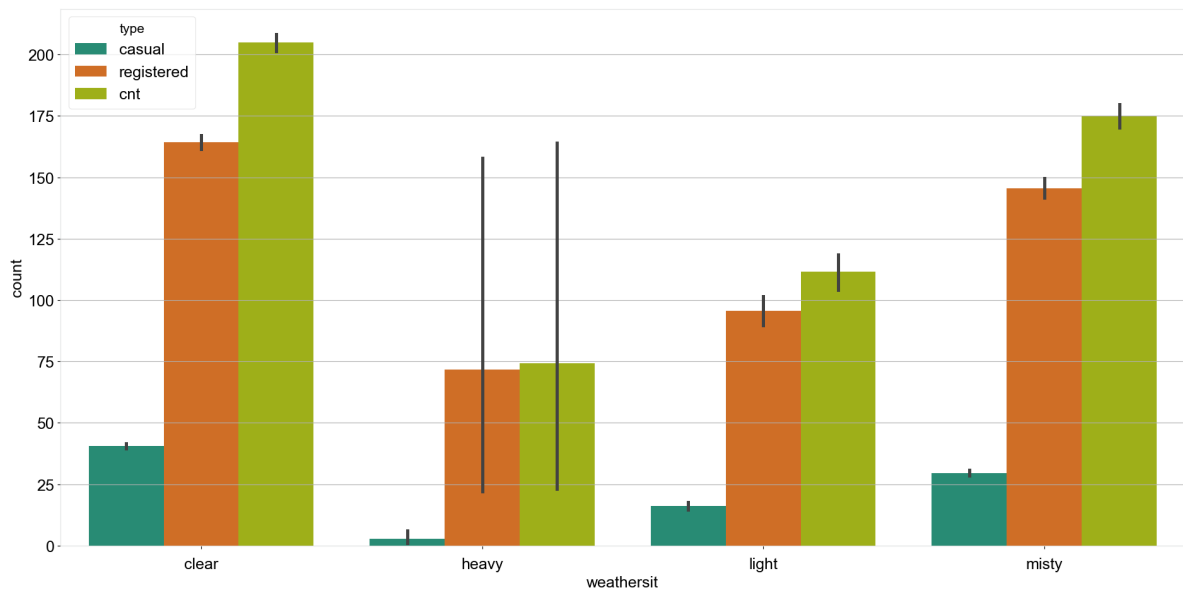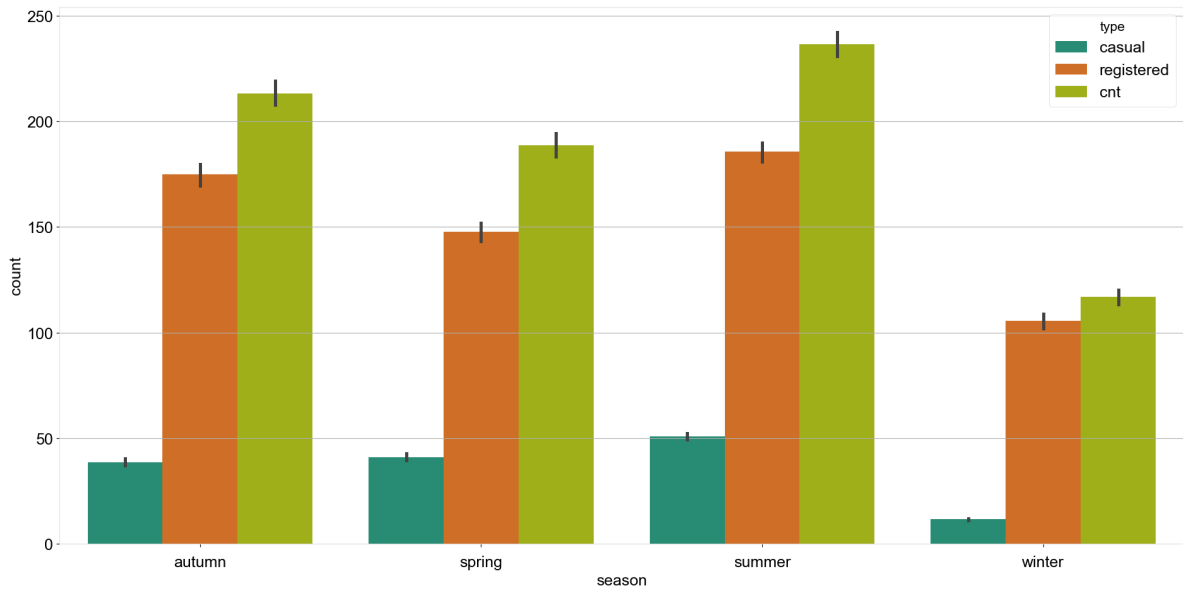
- Something was remarkable during analyzing the daily behavior for renting. The renting mean during weekends dropped for registered users and increased for casual riders. The rate is almost constant over all other working days.
- This can be explained in the light of individual behavior for consumers. Registered may tend to do other activities on weekends in contrast to casual.
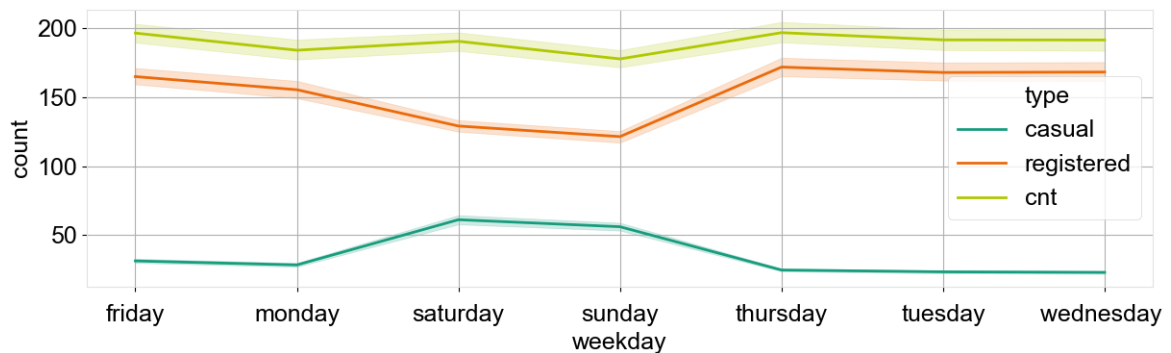
- The same trend occurs during holidays which can be confirmed from the graph below. The casual riders rate increases in holidays and decreases for registered riders. Overall, renting rate in holidays is lower than normal ones.
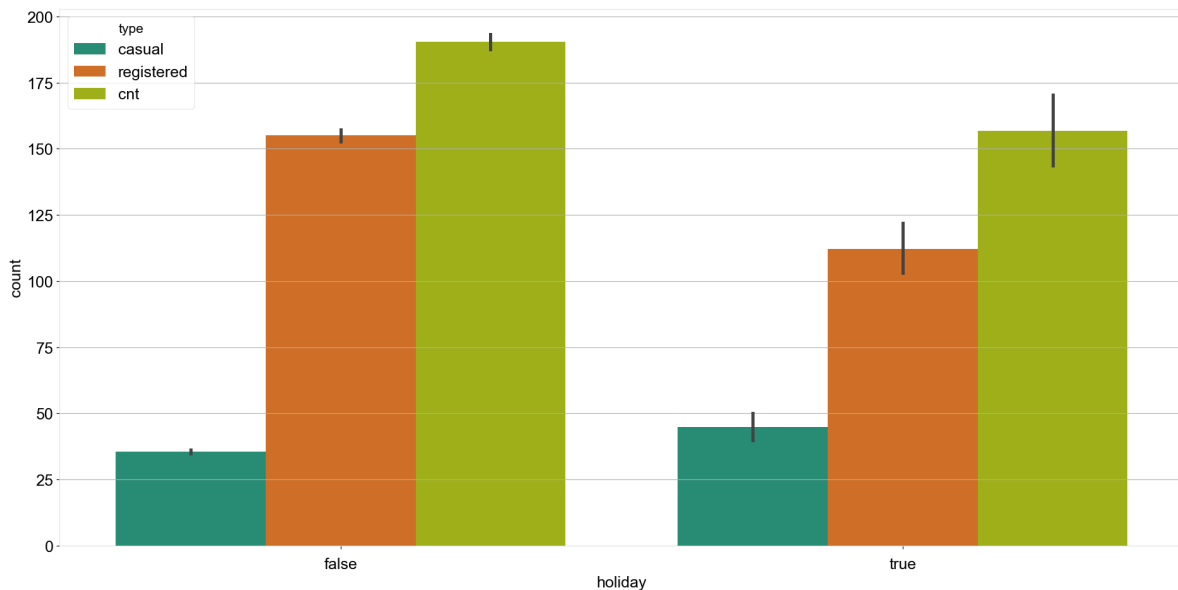
- Analyzing the hourly behavior for renting is important to indicate the busy times. By choosing an exemplary day for each season we can notice that there are two busy times between 06:00 - 09:00 and 17:00 – 20:00.
- Winter is different. It has only one peak which occurs where off-peak for other seasons happen at 15:00.
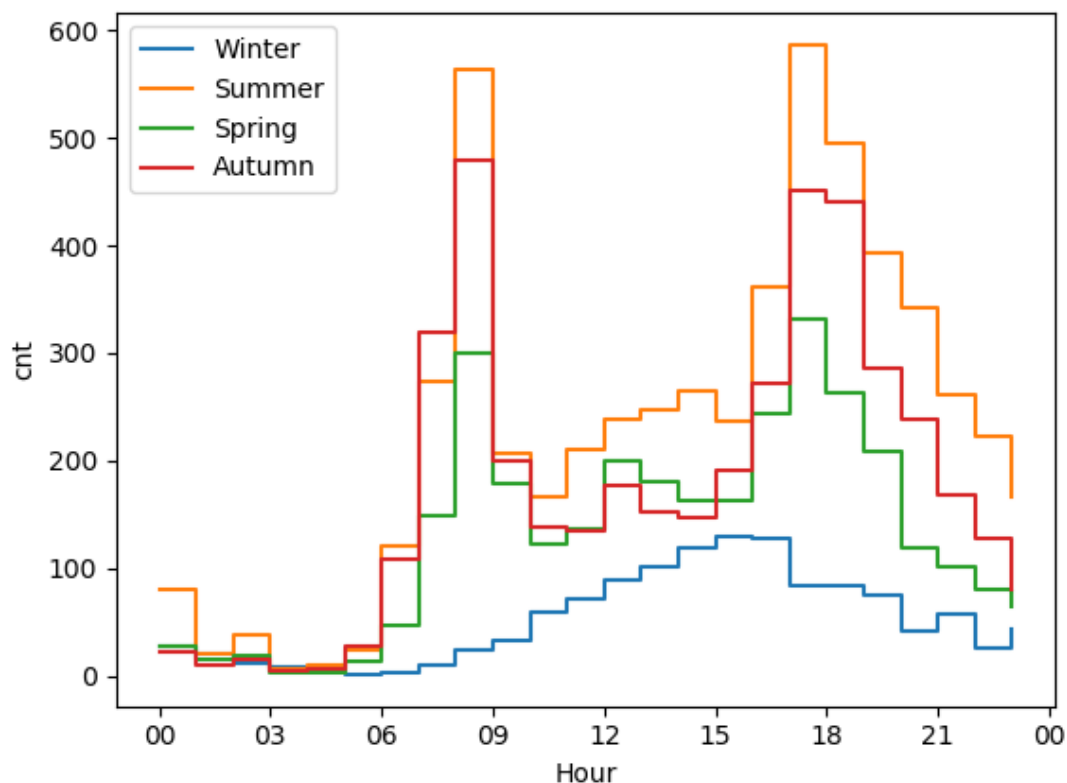
- Over all, there is two busy times for a whole year as mentioned above. With busy seasons in Summer and Autumn as indicated above.
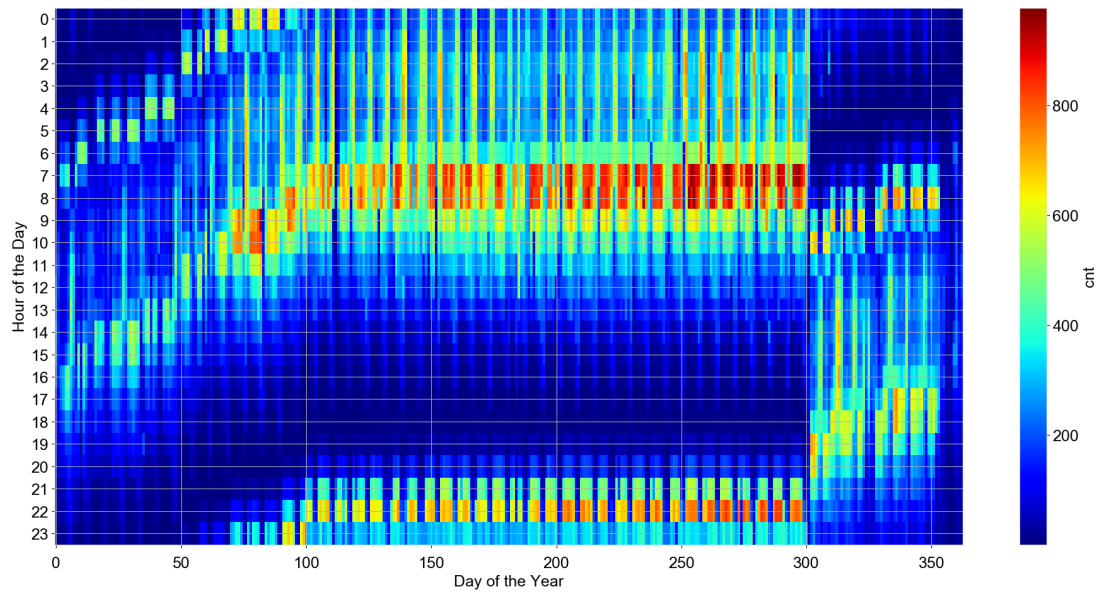

Prediction model

Features Selection

- By plotting the Pearson coefficient between the variables. It is clear that cnt has an inverse relation with humidity and direct relation with temperature.
- Also, to indicate it with coefficients, a LASSO coefficient graph is plotted. It confirms that the previous relationships with high coefficients.
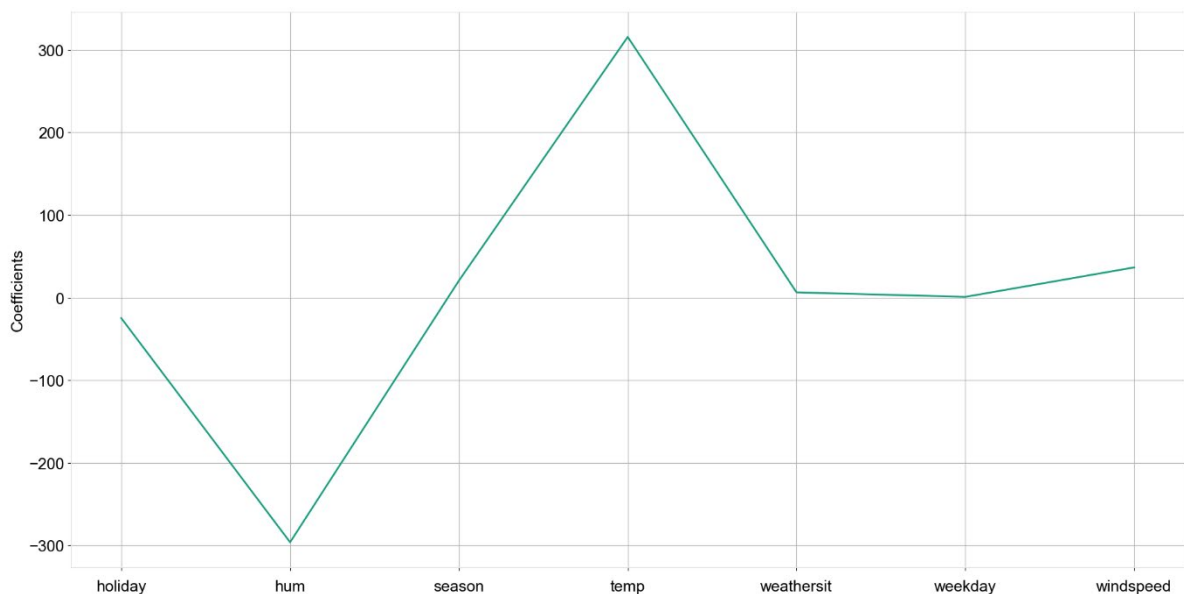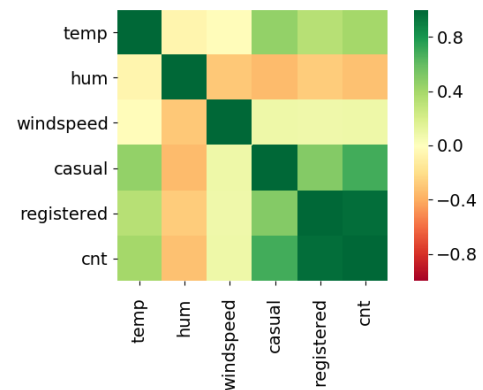
- Therefore, only temperature and humid are the most important features for the further prediction.
- Cnt is a continues variable, therefore, linear regression is an ideal easy model for prediction. However, its accuracy was not high enough. Thus, after increasing the polynomial value to 3, It increased by 6%.
- By building a polynomial regression model (n=3), an accuracy of 0.34 is achieved. The frequent and extreme variance results in a low accuracy. This can be tweaked with selecting only important features, cross-validation and Standardizing as has been done, however, it is still low.
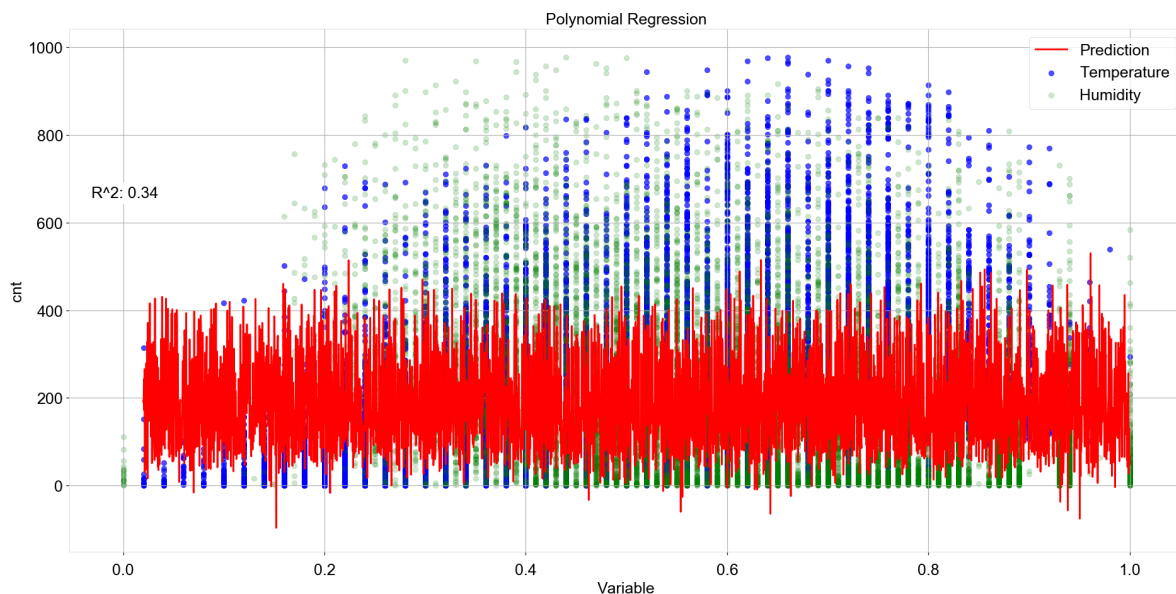
MAD values

Season	Holiday	Weekday	Weathersit	Temp	Hum	Windspeed
0.99	0.05	1.72	0.56	0.16	0.16	0.096
		Casual	Registered	Total		
		34	114	142		

- In case of working in a development team, the codes used be construed of patterns, direct for loops and regulation function expression. List comprehensions and lambda function can be confusing. Also, docstrings are crucial for others to understand what the utilization of this function is. Comments explain vague lines should also be included. This provide developers with insights for quick reacting to daily uncertainties.