Date of Submission : May 24, 2022

Bike Sharing Demand

Advanced Data Analytics
xxxxxxxxxxxX Project Report Xxxxxxxxxxxx

Kajal - S20190020215,

Regression

Abstract: Bike sharing systems are new generation of traditional bike rentals where whole process from membership, rental and return back has become automatic. Through these systems, user is able to easily rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousands bicycles. Today, there exists great interest in these systems due to their important role in traffic, environmental and health issues. Opposed to other transport services such as bus or subway, the duration of travel, departure and arrival position is explicitly recorded in these systems. This feature turns bike sharing system into a virtual sensor network that can be used for sensing mobility in the city. Hence, it is expected that most of important events in the city could be detected via monitoring these data.

1. Introduction

Bike-sharing rental process is highly correlated to the environmental and seasonal settings.
For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors.
The core data set is related to the two-year historical log corresponding to years 2011 and 2012 from Capital Bikeshare system, Washington D.C., USA . We aggregated the data on two hourly and daily basis and then the corresponding weather and seasonal information.

1.1. Dataset characteristics

Both hour.csv and day.csv have the following fields, except hr which is not available in day.csv

-instant: record index
- dteday : date
- season : season (1:springer, 2:summer, 3:fall, 4:winter)
- yr : year (0: 2011, 1:2012)
- mnth : month (1 to 12)
- hr : hour (0 to 23)
- holiday : weather day is holiday or not (extracted from http://dchr.dc.gov/page/holiday-schedule)
- weekday : day of the week
- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
+ weathersit :
- 1: Clear, Few clouds, Partly cloudy, Partly cloudy
- 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
- 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp : Normalized temperature in Celsius. The values are divided to 41 (max)
- atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
- hum: Normalized humidity. The values are divided to 100 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)
- casual: count of casual users
- registered: count of registered users
- cnt: count of total rental bikes including both casual and registered

2. Key concepts for project implementation:

2.1. Coefficient of Determination

The coefficient of determination (R^2 or r-squared) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, the coefficient of determination tells one how well the data fits the model (the goodness of fit).

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} \tag{1}$$

SSregression – The sum of squares due to regression (explained sum of squares)
SStotal – The total sum of squares

The coefficient of determination can take any values between 0 to 1.

2.2. Correlation Analysis

In statistics, the word correlation is used to denote some form of association between two variables.The correlation may be positive, negative or zero.
Correlation coefficient(r) is used to measure the degree of correlation. A correlation coefficient is a number between -1 and 1 that tells you the strength and direction of a relationship between variables.

we must choose the most relevant and non-redundant features from the original feature set to reduce the number of features.Here we use correlation analysis.

2.3. Mean Squared Error(MSE)

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. As model error increases, its value increases.

$$MSE = \frac{\sum(y_i - y_i^{hat})^2}{n} \tag{2}$$

Where:
yi is the ith observed value.
yi is the corresponding predicted value.
n = the number of observations.

2.4. Models used in this project

2.4.1 Linear Regression

Ordinary least squares Linear Regression.
LinearRegression fits a linear model with coefficients w = (w1, . . . , wp) to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation.

2.4.2 DecisionTreeRegressor

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

2.4.3 RandomForestRegressor

A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the max samples parameter if bootstrap=True (default), otherwise the whole dataset is used to build each tree.

2.4.4 AdaBoostRegressor

An AdaBoost regressor is a meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction. As such, subsequent regressors focus more on difficult cases.

3. Methodology

3.0.1 Import necessary libraries

3.0.2 Load and understand the dataset

3.0.3 Anomaly Detection

* Done in case of day dataset

3.0.4 Missing value Analysis

1.Finding missing values and either removing them or replacing them with appropriate statistic

3.0.5 Exploratory Data Analysis

1.Plot various graphs,barcharts ,pie charts etc on the data
2.Understand how the data is distributed with respect to different attributes

3.0.6 Outlier Detection and data preprocessing

1.Outliers are found using IQR and removed
2.Having outliers effects the model prediction significantly
3.Data preprocessing also include converting the data to the form that the model can be built on that .

3.0.7 Correlation Analysis

1.Correlation analysis involves the study of linear associtíon between various columns and if found a strong positive linear correlation one of the column need to be removed .

3.0.8 Model Building

1. Modelling the training dataset using models like
* Linear Regression Model
* Decision Tree Regressor Model
* Random Forest Model
* Ada Boost Regressor

2. Model performance on test dataset like
* Linear Regression Prediction
* Decision Tree Regressor Prediction
* Random Forest Prediction
* Ada Boost Regressor

3. Model Evaluation Metrics
* R-squared score
* Mean square error

4.Choosing best model for predicting bike rental count

3.0.9 Model Deployment

1.A pickle file is generated for the model that is performing the best 2.Using github and streamlit a web app has been built and two models have been deployed one for daily prediction and one for hourly prediction

4. Experimental results for day dataset

4.1. Anomaly Detection

4.1.1

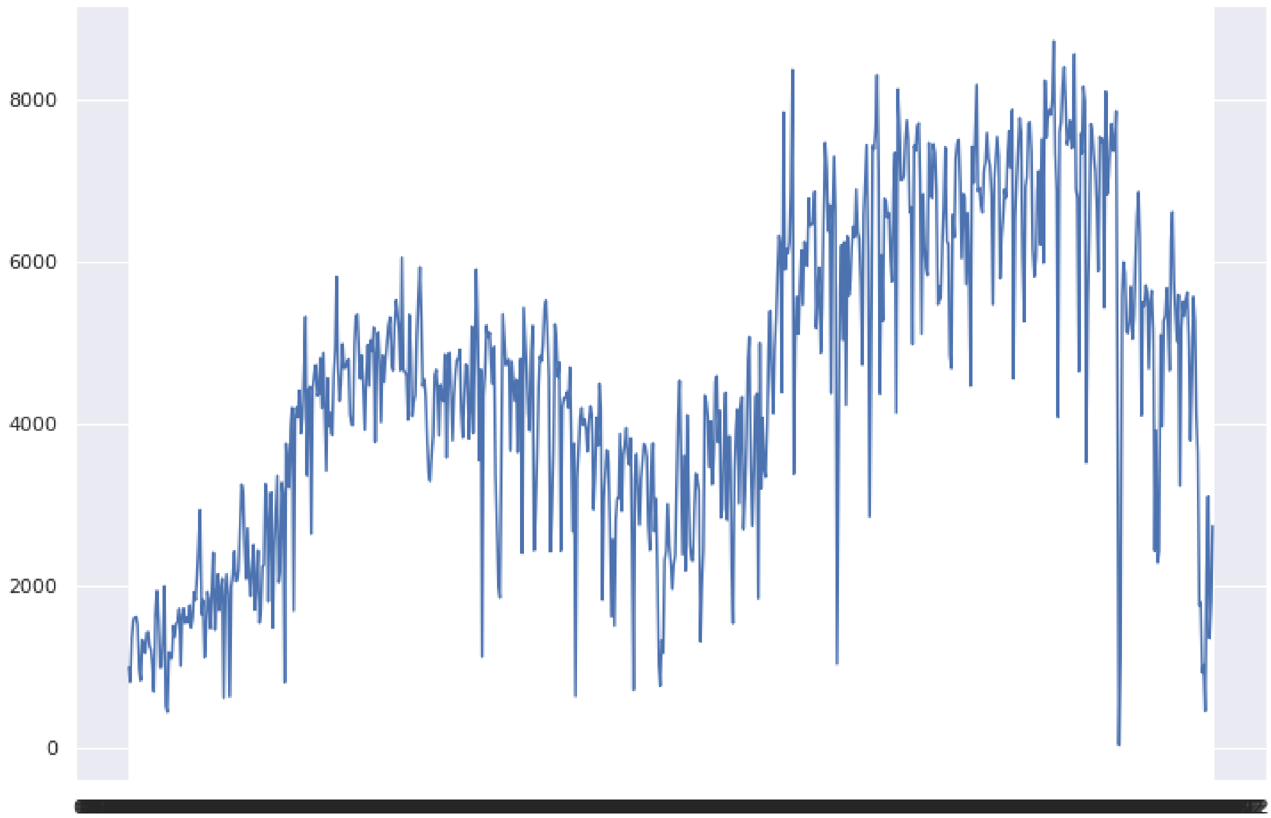

Figure 1: Date vs count

4.1.2

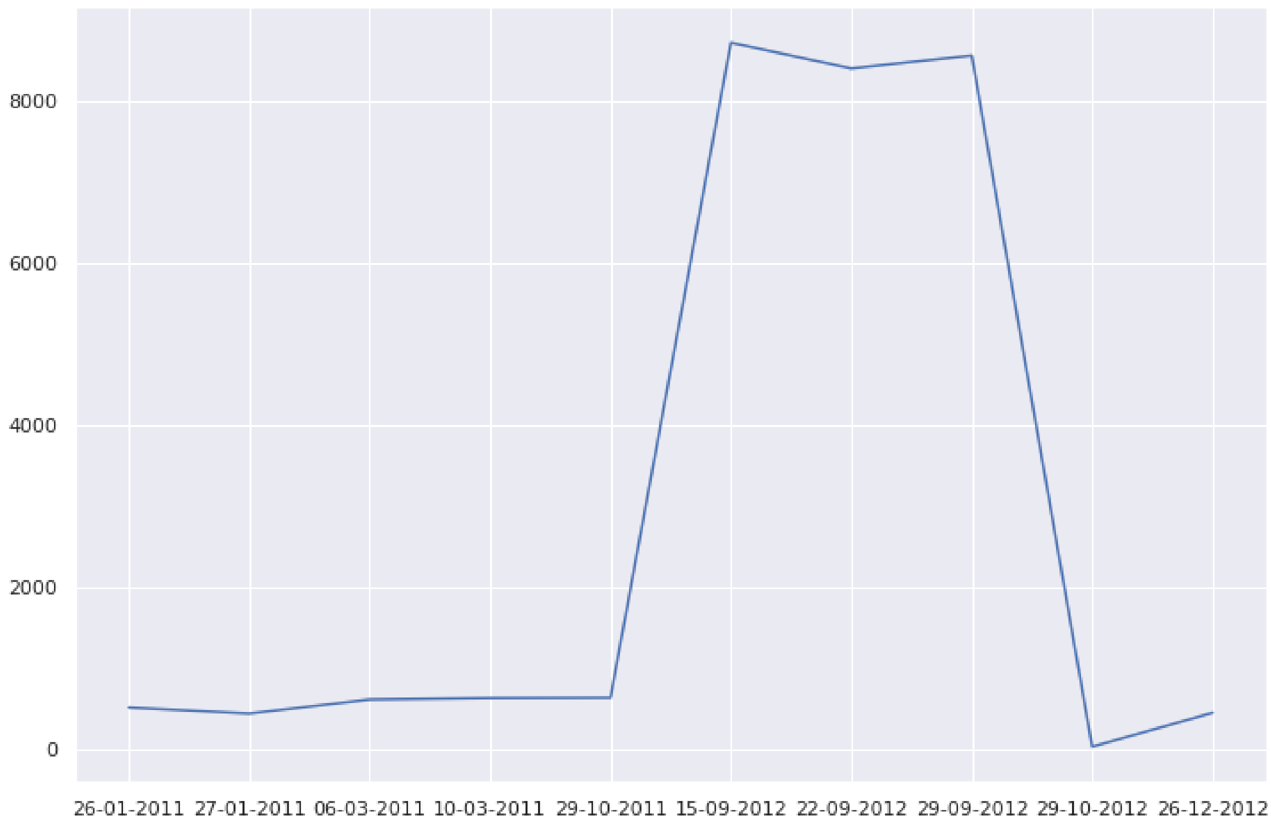

Figure 2: Anomaly dates filtered vs the count

Few Reasons why few of the above anamolies might have happened are analyzed:

1.Reasons for having high bike rental demand on 29-09-2012 might be analyzed from google search result like this

On September 29, 2013, a group of motorcyclists were participating in an annual rally, titled "Hollywood Stuntz," organized by a man named Jamie Lao.[4] Rallies organized by Lao in the past involved performing motorcycle stunts and an unauthorized ride through Times Square, Manhattan. During the 2012 ride a year prior, "well over a thousand motorcycles, dirt bikes, quads, four-wheel vehicles" rode through Times Square, according to New York Police Commissioner Raymond Kelly. The group did not have permits to do so.[5][6] Driving through the section of Broadway within Times Square has been illegal since it was pedestrianized in 2009. Kelly reported that the altercation with Lien was not the only problem involving the group on September 29, as over 200 other people had complained to the police about the reckless driving of the bikers on Manhattan's streets that day.

2.Reasons for having bike rental demand on 2012-10-30 might be:

query like "2012-10-30 washington d.c." in Google returns related results to Hurricane Sandy.

4.2. Summary statistics and missing value analysis

4.2.1 Summary Statistics of the day dataset

	season	yr	mnth	holiday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
count	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000
mean	2.496580	0.500684	6.519836	0.028728	0.683995	1.395349	0.495385	0.474354	0.627894	0.190486	848.176471	3656.172367	4504.348837
std	1.110807	0.500342	3.451913	0.167155	0.465233	0.544894	0.183051	0.162961	0.142429	0.077498	686.622488	1560.256377	1937.211452
min	1.000000	0.000000	1.000000	0.000000	0.000000	1.000000	0.059130	0.079070	0.000000	0.022392	2.000000	20.000000	22.000000
25%	2.000000	0.000000	4.000000	0.000000	0.000000	1.000000	0.337083	0.337842	0.520000	0.134950	315.500000	2497.000000	3152.000000
50%	3.000000	1.000000	7.000000	0.000000	1.000000	1.000000	0.498333	0.486733	0.626667	0.180975	713.000000	3662.000000	4548.000000
75%	3.000000	1.000000	10.000000	0.000000	1.000000	2.000000	0.655417	0.608602	0.730209	0.233214	1096.000000	4776.500000	5956.000000
max	4.000000	1.000000	12.000000	1.000000	1.000000	3.000000	0.861667	0.840896	0.972500	0.507463	3410.000000	6946.000000	8714.000000

Figure 3: Summary Statistics

4.2.2 Missing values

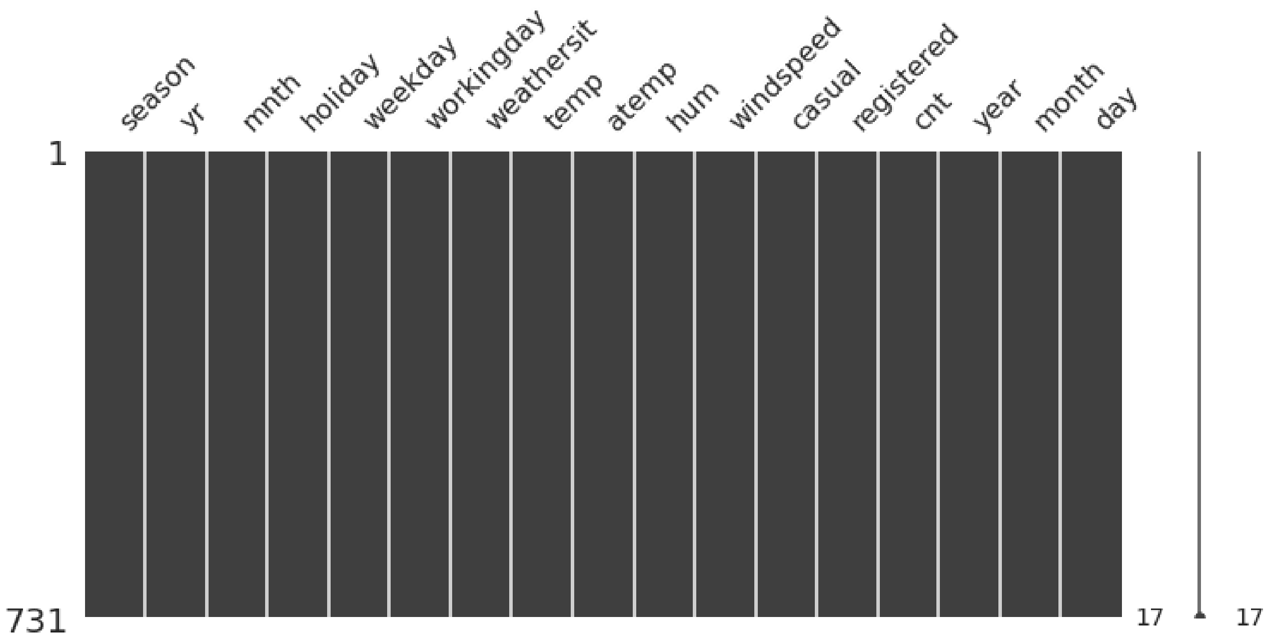

Figure 4: Matrix plot

Visualisation to understand the presence and distribution of missing data within a pandas dataframe.
The matrix plot is a great tool if we are working with depth-related data or time-series data. It provides a colour fill for each column. When data is present, the plot is shaded in grey (or your colour of choice), and when it is absent the plot is displayed in white.

4.3. Exploratory Data Analytics

4.3.1 Barplots of various variables with count

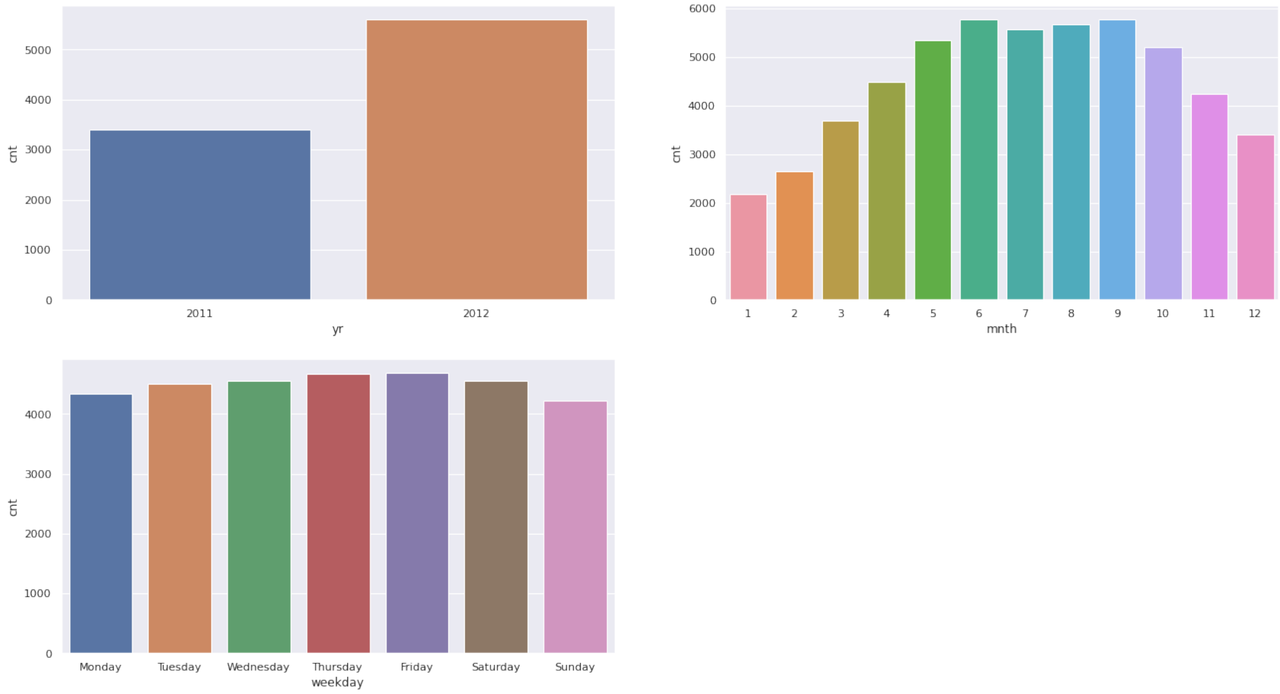

Figure 5: Barplots for count on year ,month and weekday

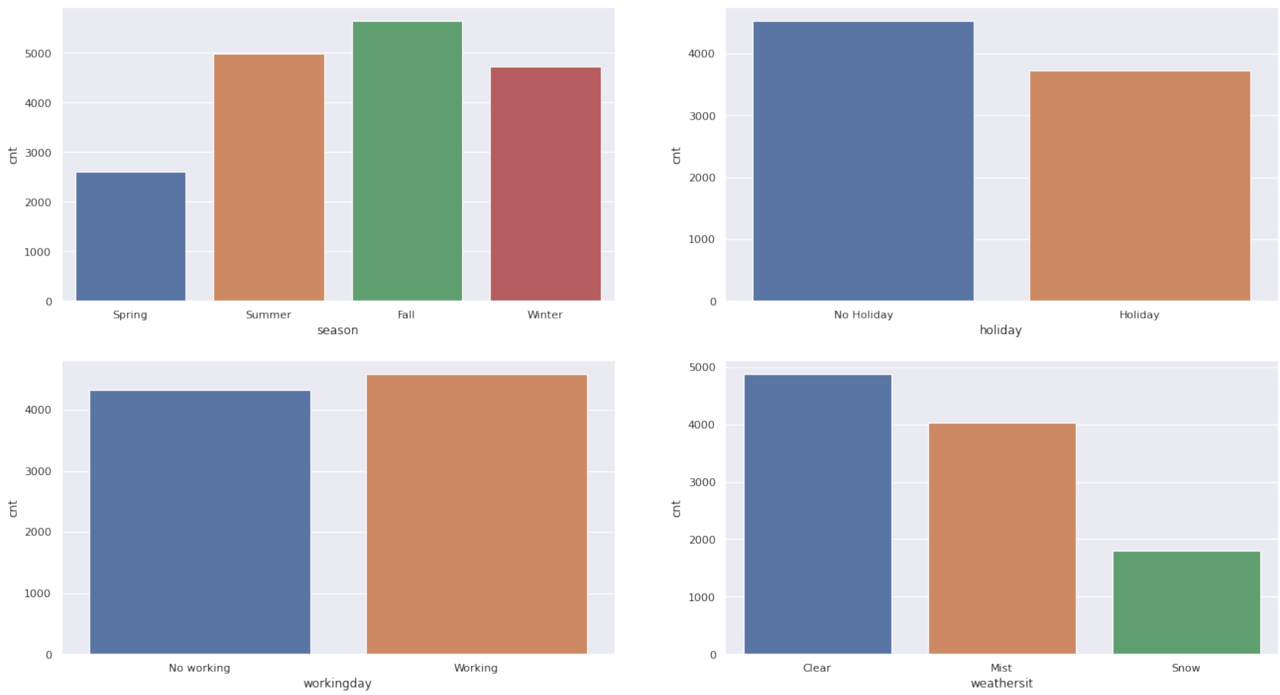

Figure 6: Barplots for count on season ,holiday ,workingday and weathersit

4.3.2 Pie Charts of various variables with count

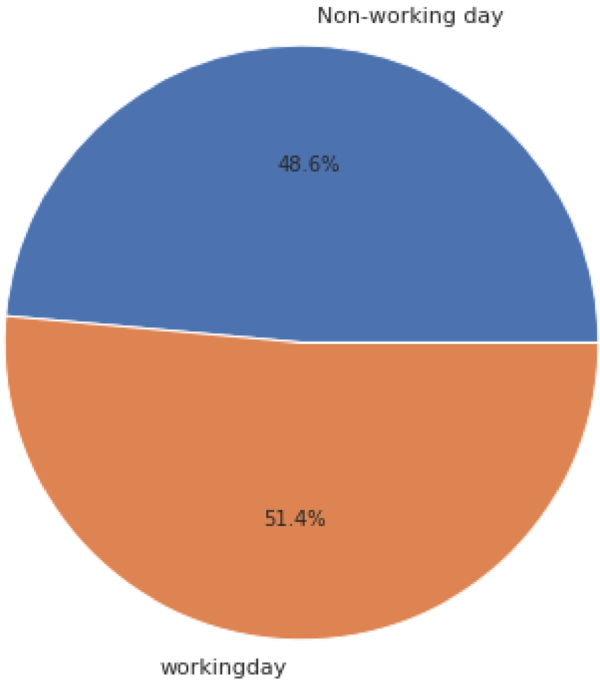

Figure 7: Proportion of counts on working days vs non working days

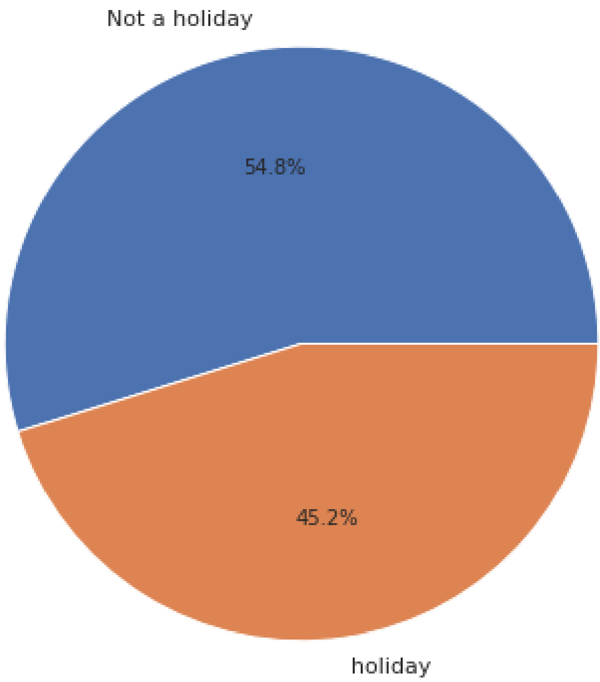

Figure 8: Proportion of counts on working days vs non working days

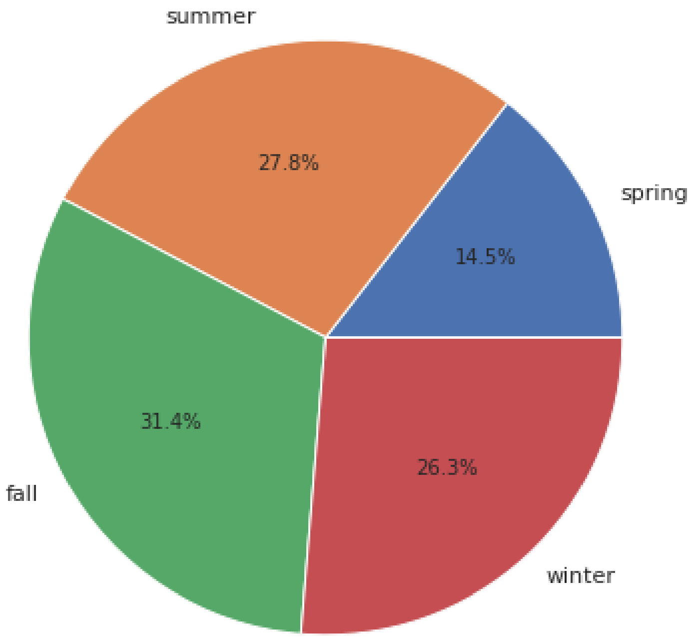

Figure 9: Proportion of counts on summer vs spring vs winter vs fall

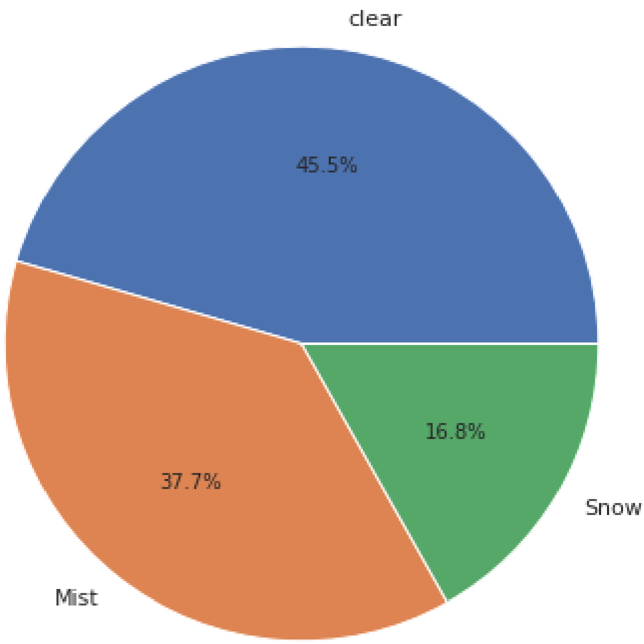

Figure 10: Proportion of counts on Mist vs Snow vs Clear

4.3.3 Boxplots for outliers detection

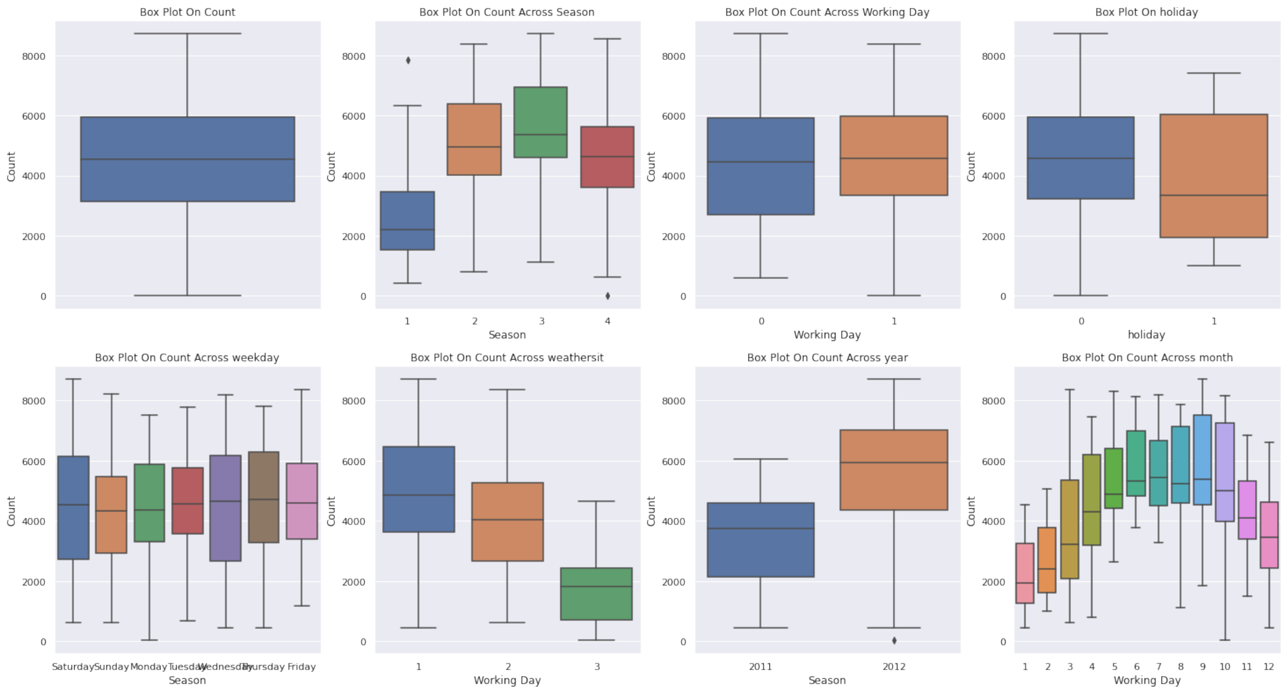

Figure 11: Boxplots

We don't have much outliers in this case

4.3.4 Correlation Analysis

	season	yr	mnth	holiday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt	year	month	day
season	1.000000	-0.001844	0.831440	-0.010537	0.012485	0.019211	0.334315	0.342876	0.205445	-0.229046	0.210399	0.411623	0.406100	-0.001844	0.831440	-0.000861
yr	-0.001844	1.000000	-0.001792	0.007954	-0.002013	-0.048727	0.047604	0.046106	-0.110651	-0.011817	0.248546	0.594248	0.566710	1.000000	-0.001792	0.002061
mnth	0.831440	-0.001792	1.000000	0.019191	-0.005901	0.043528	0.220205	0.227459	0.222204	-0.207502	0.123006	0.293488	0.279977	-0.001792	1.000000	0.009157
holiday	-0.010537	0.007954	0.019191	1.000000	-0.253023	-0.034627	-0.028556	-0.032507	-0.015937	0.006292	0.054274	-0.108745	-0.068348	0.007954	0.019191	-0.010710
workingday	0.012485	-0.002013	-0.005901	-0.253023	1.000000	0.061200	0.052660	0.052182	0.024327	-0.018796	-0.518044	0.303907	0.061156	-0.002013	-0.005901	0.009907
weathersit	0.019211	-0.048727	0.043528	-0.034627	0.061200	1.000000	-0.120602	-0.121583	0.591045	0.039511	-0.247353	-0.260388	-0.297391	-0.048727	0.043528	-0.023824
temp	0.334315	0.047604	0.220205	-0.028556	0.052660	-0.120602	1.000000	0.991702	0.126963	-0.157944	0.543285	0.540012	0.627494	0.047604	0.220205	0.024253
atemp	0.342876	0.046106	0.227459	-0.032507	0.052182	-0.121583	0.991702	1.000000	0.139988	-0.183643	0.543864	0.544192	0.631066	0.046106	0.227459	0.018286
hum	0.205445	-0.110651	0.222204	-0.015937	0.024327	0.591045	0.126963	0.139988	1.000000	-0.248489	-0.077008	-0.091089	-0.100659	-0.110651	0.222204	0.042753
windspeed	-0.229046	-0.011817	-0.207502	0.006292	-0.018796	0.039511	-0.157944	-0.183643	-0.248489	1.000000	-0.167613	-0.217449	-0.234545	-0.011817	-0.207502	0.021588
casual	0.210399	0.248546	0.123006	0.054274	-0.518044	-0.247353	0.543285	0.543864	-0.077008	-0.167613	1.000000	0.395282	0.672804	0.248546	0.123006	-0.013365
registered	0.411623	0.594248	0.293488	-0.108745	0.303907	-0.260388	0.540012	0.544192	-0.091089	-0.217449	0.395282	1.000000	0.945517	0.594248	0.293488	-0.024171
cnt	0.406100	0.566710	0.279977	-0.068348	0.061156	-0.297391	0.627494	0.631066	-0.100659	-0.234545	0.672804	0.945517	1.000000	0.566710	0.279977	-0.024204
year	-0.001844	1.000000	-0.001792	0.007954	-0.002013	-0.048727	0.047604	0.046106	-0.110651	-0.011817	0.248546	0.594248	0.566710	1.000000	-0.001792	0.002061
month	0.831440	-0.001792	1.000000	0.019191	-0.005901	0.043528	0.220205	0.227459	0.222204	-0.207502	0.123006	0.293488	0.279977	-0.001792	1.000000	0.009157
day	-0.000861	0.002061	0.009157	-0.010710	0.009907	-0.023824	0.024253	0.018286	0.042753	0.021588	-0.013365	-0.024171	-0.024204	0.002061	0.009157	1.000000

Figure 12: Correlation matrix

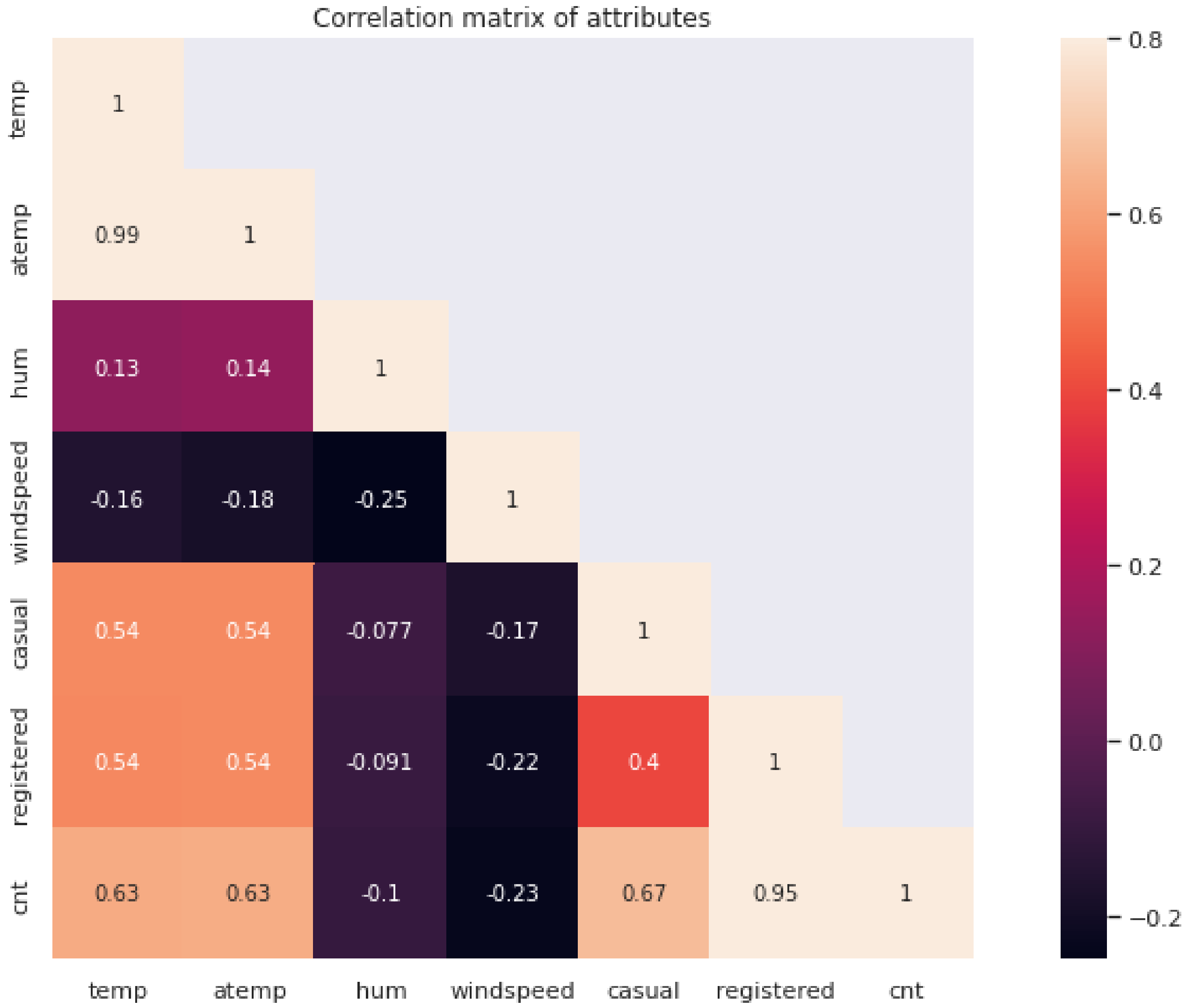

Figure 13

4.4. Model building

4.4.1 Models built

Registered Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	526682.5021514217	0.7269987856902445	working days
DecisionTreeRegressor	530713.3288037037	0.7768219525151165	working days
RandomForestRegressor	390773.1989712338	0.8176904220786967	working days
AdaBoostRegressor	383203.2035494365	0.8184584802180302	working days

Registered Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	522103.7914015562	0.5752410756446696	Non working days
DecisionTreeRegressor	642302.231948495	0.48588217139634626	Non working days
RandomForestRegressor	311973.24900839065	0.7258216884136692	Non working days
AdaBoostRegressor	299575.12113376	0.719147220517912	Non working days

Casual Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	42942.58157980533	0.5933670414985597	working days
DecisionTreeRegressor	57563.74079981994	0.5447970448884745	working days
RandomForestRegressor	43918.39289579461	0.6276223635699258	working days
AdaBoostRegressor	41621.95893782781	0.6292062784675707	working days

Casual Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	245165.23756501125	0.5634671086080872	Non working days
DecisionTreeRegressor	191122.1190290945	0.7238177993533758	Non working days
RandomForestRegressor	146910.12462449734	0.7755580184431531	Non working days
AdaBoostRegressor	41621.95893782781	0.6292062784675707	Non working days
All Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	640249.1401510865	0.7718907251327742	working days
DecisionTreeRegressor	620736.561061285	0.8090308997622665	working days
RandomForestRegressor	452803.61124522984	0.858826384478959	working days
AdaBoostRegressor	469047.1839846259	0.8499280918731321	working days
All Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	1083602.5635630684	0.6661323871775375	Non working days
DecisionTreeRegressor	927796.1974305003	0.7157184504443861	Non working days
RandomForestRegressor	653843.5491918524	0.7964308956686346	Non working days
AdaBoostRegressor	525744.3981968537	0.8282466521747489	Non working days
All Bike Rental			
Model	MSE	r2score	Built for
Linear Regression	799383.1653506244	0.7268478591679416	All days
DecisionTreeRegressor	617441.7654921587	0.8055195185581923	All days
RandomForestRegressor	617441.7654921587	0.8055195185581923	All days
AdaBoostRegressor	463245.88189044665	0.8489198551714614	All days

In each case the model that have the highest r2score is taken as the best model .
That model can be taken and put into production .

5. Experimental results for hour dataset

5.1. Summary statistics and missing value analysis

5.1.1 Summary Statistics of the hour dataset

	season	yr	mnth	hr	holiday	weekday	workingday	weathersit
count	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000
mean	2.501640	0.502561	6.537775	11.546752	0.028770	3.003683	0.682721	1.425283
std	1.106918	0.500008	3.438776	6.914405	0.167165	2.005771	0.465431	0.639357
min	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	4.000000	6.000000	0.000000	1.000000	0.000000	1.000000
50%	3.000000	1.000000	7.000000	12.000000	0.000000	3.000000	1.000000	1.000000
75%	3.000000	1.000000	10.000000	18.000000	0.000000	5.000000	1.000000	2.000000
max	4.000000	1.000000	12.000000	23.000000	1.000000	6.000000	1.000000	4.000000

	temp	atemp	hum	windspeed	casual	registered	cnt
count	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000
mean	0.496987	0.475775	0.627229	0.190098	35.676218	153.786869	189.463088
std	0.192556	0.171850	0.192930	0.122340	49.305030	151.357286	181.387599
min	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	0.340000	0.333300	0.480000	0.104500	4.000000	34.000000	40.000000
50%	0.500000	0.484800	0.630000	0.194000	17.000000	115.000000	142.000000
75%	0.660000	0.621200	0.780000	0.253700	48.000000	220.000000	281.000000
max	1.000000	1.000000	1.000000	0.850700	367.000000	886.000000	977.000000

Figure 14: Summary Statistics

5.1.2 Missing values

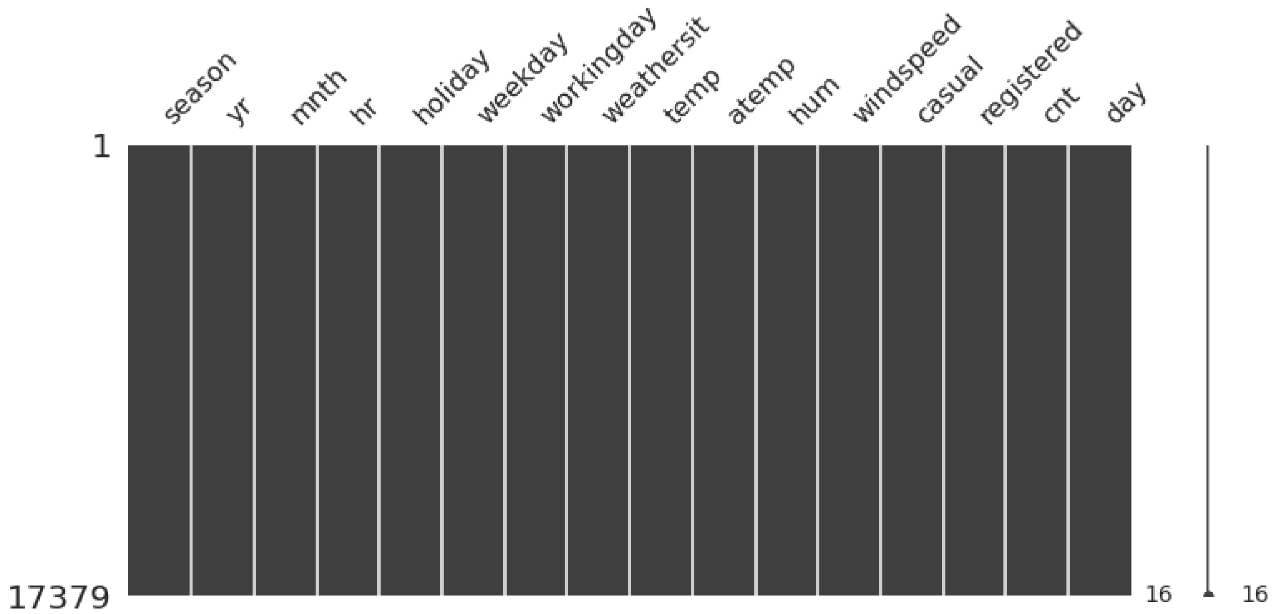

Figure 15: Matrix plot

Visualisation to understand the presence and distribution of missing data within a pandas dataframe.
The matrix plot is a great tool if we are working with depth-related data or time-series data. It provides a colour fill for each column. When data is present, the plot is shaded in grey (or your colour of choice), and when it is absent the plot is displayed in white.

5.2. Exploratory Data Analytics

5.2.1 Barplots of various variables with count

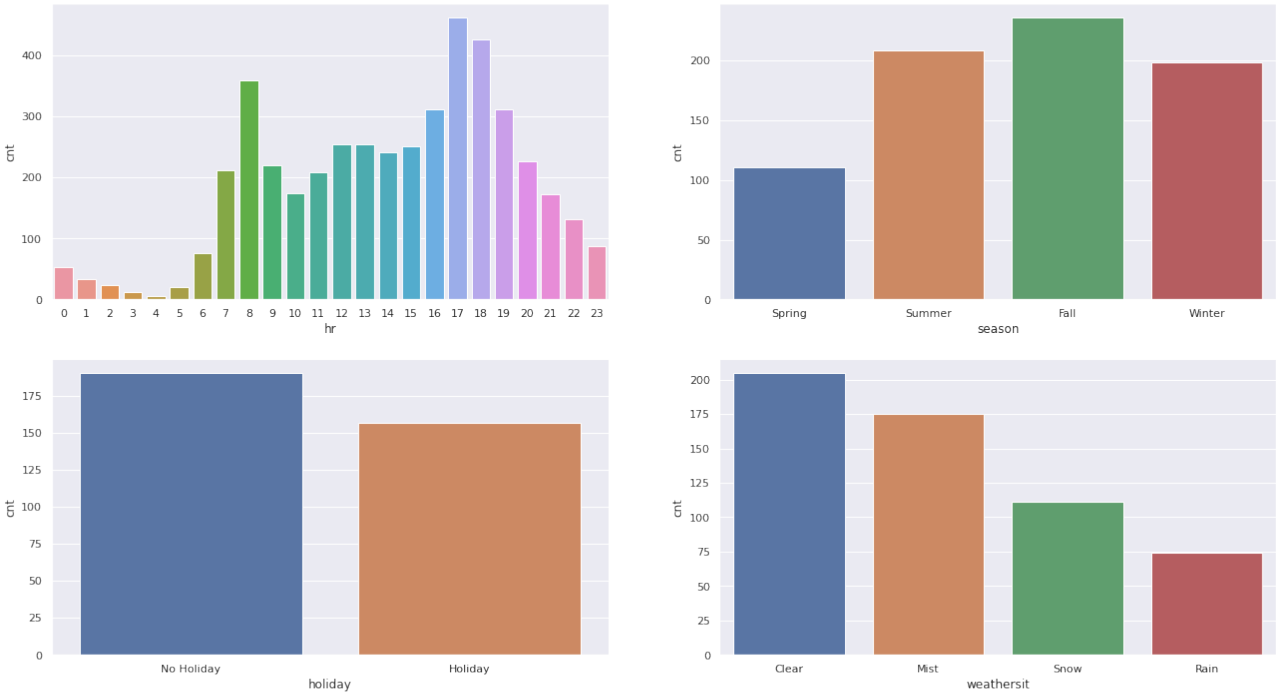

Figure 16: Barplots for count on hour ,season ,holiday and weathersit

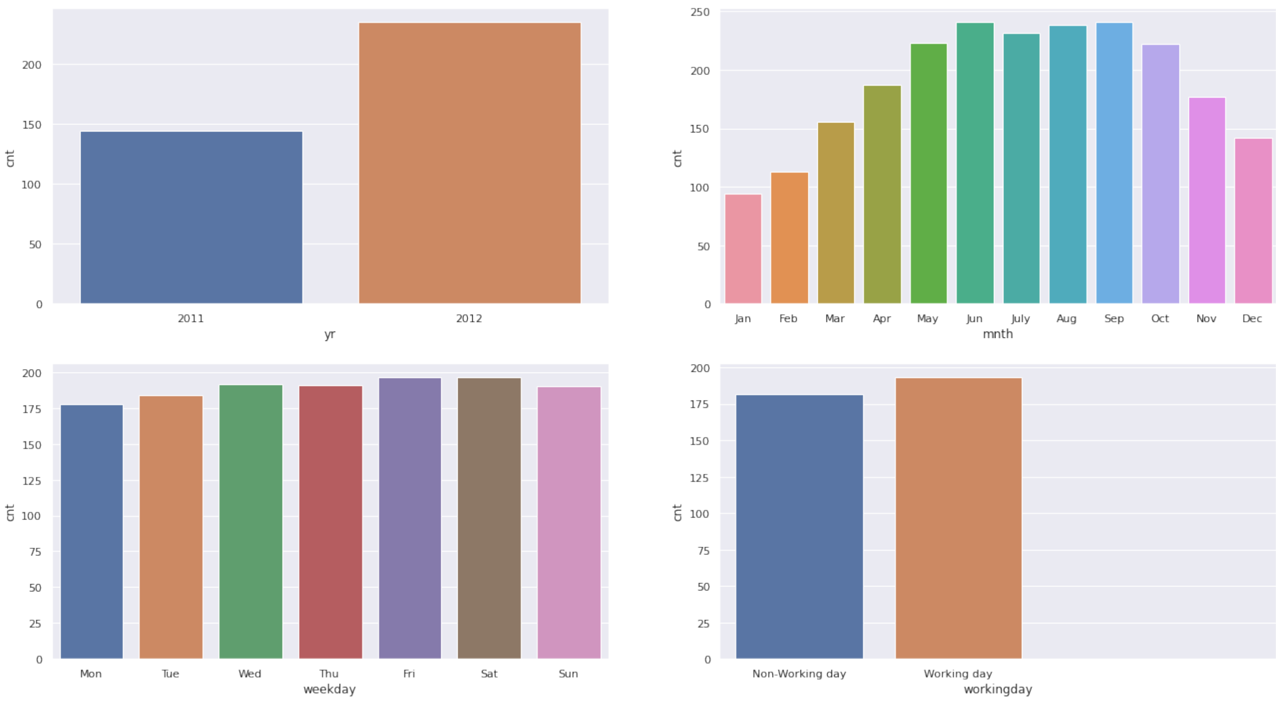

Figure 17: Barplots for count on year,month ,weekday and workingday

5.2.2 Pie Charts of various variables with count

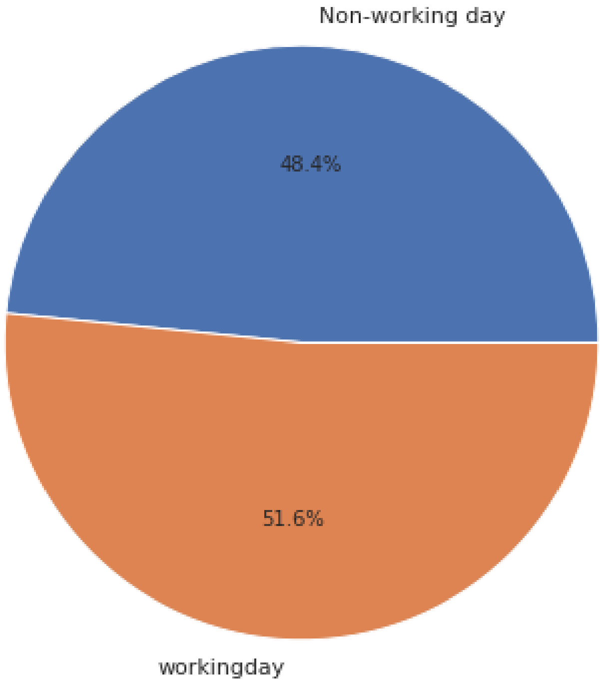

Figure 18: Proportion of counts on working days vs non working days

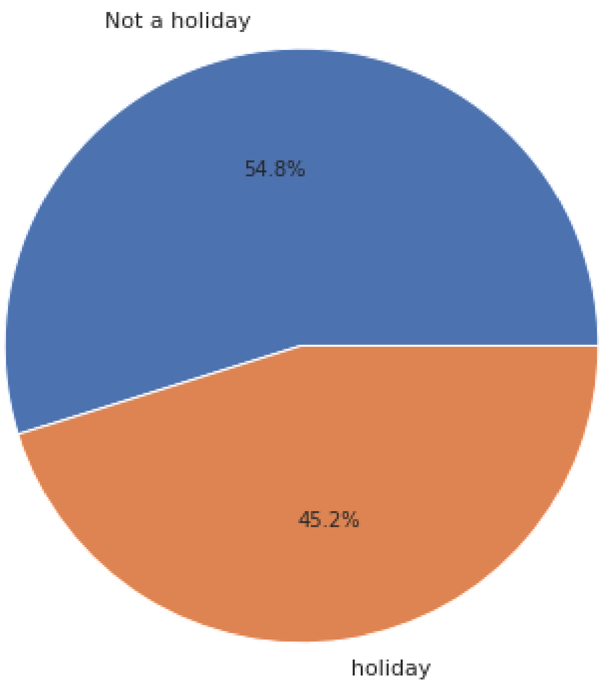

Figure 19: Proportion of counts on holiday vs not a holiday

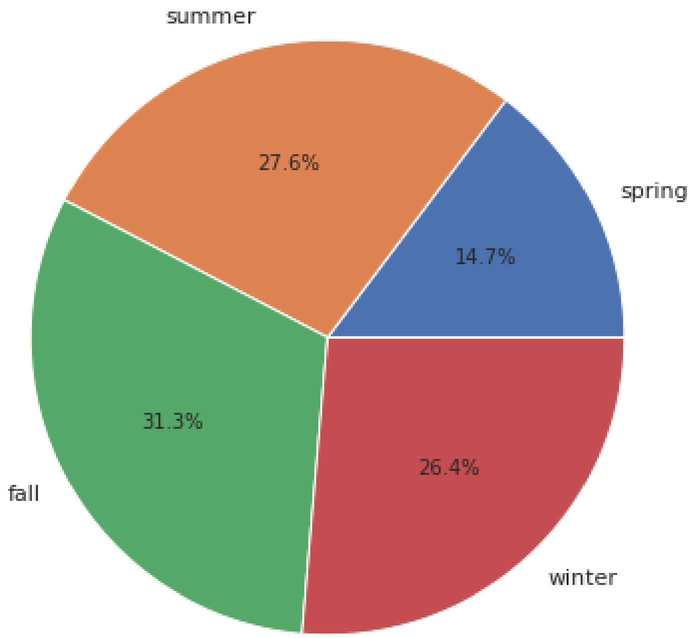

Figure 20: Proportion of counts on summer vs spring vs winter vs fall

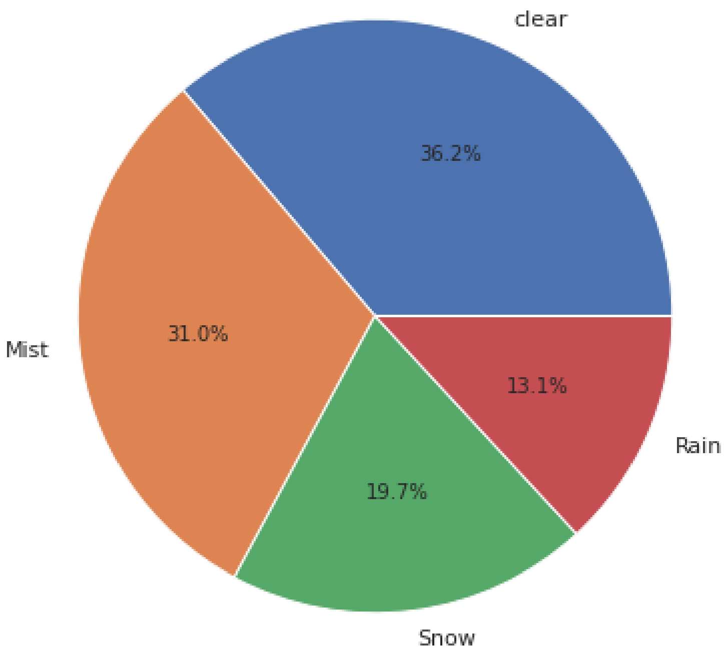

Figure 21: Proportion of counts on Mist vs Snow vs Clear vs Rain

5.2.3 Boxplots for outliers detection

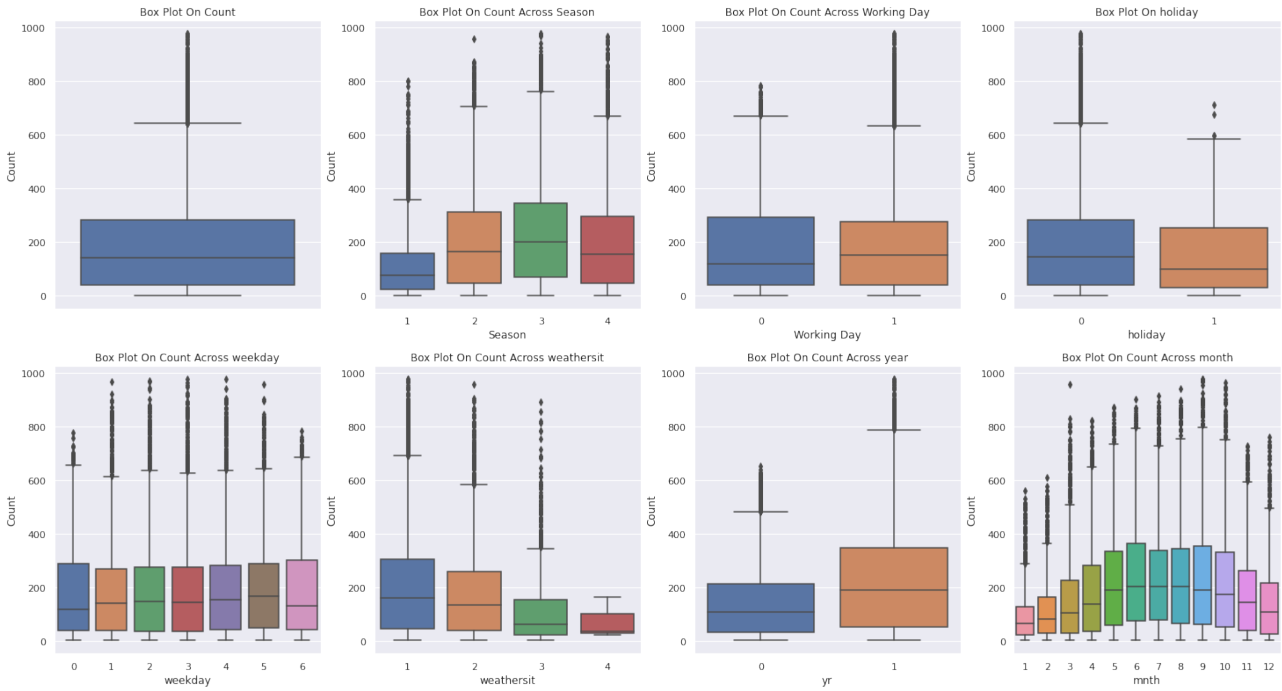

Figure 22: Boxplots

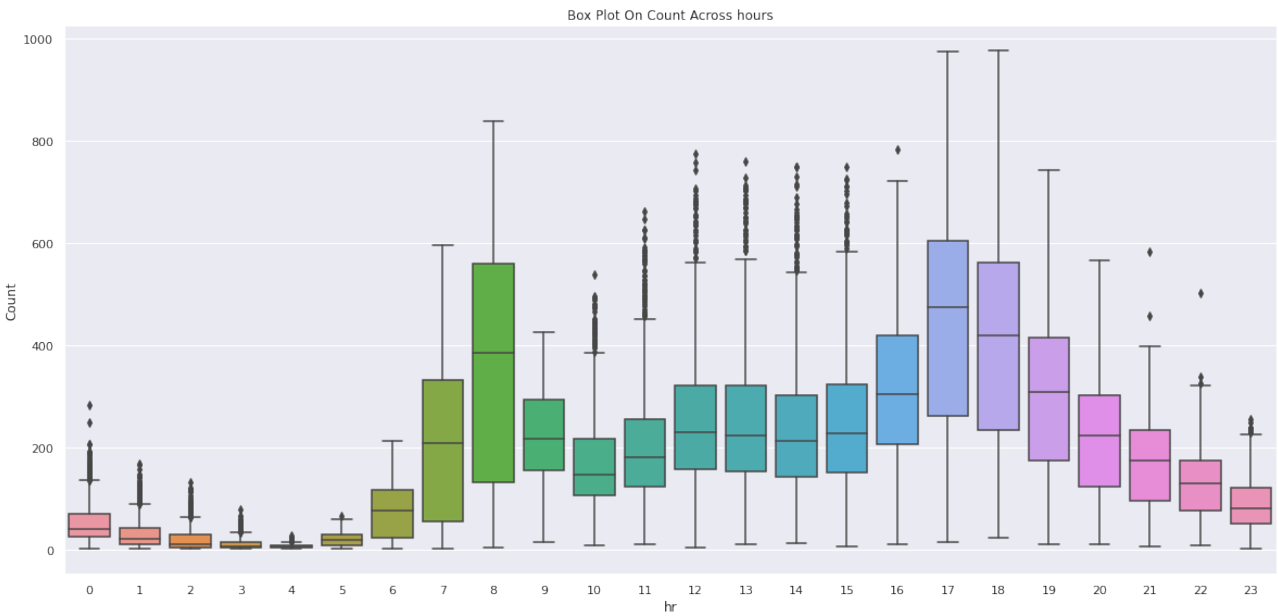

Figure 23: Boxplots

We have much outliers in this case . We need to remove them from the dataset.

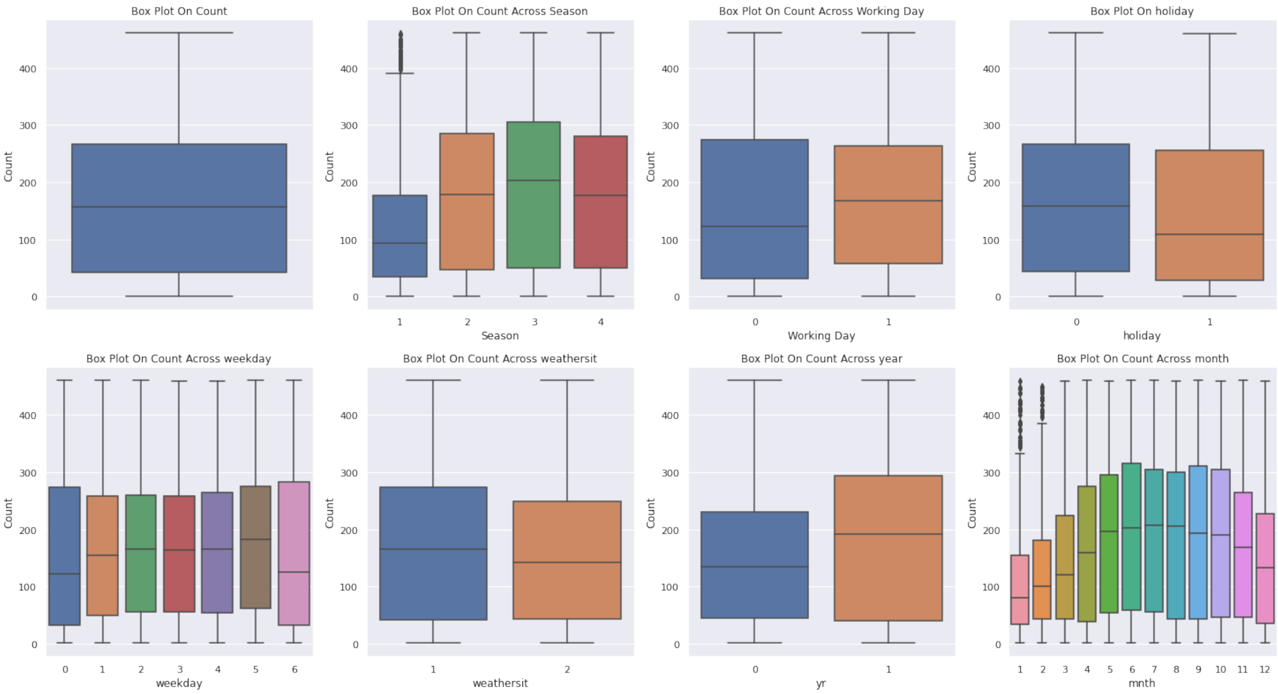

Figure 24: Boxplots

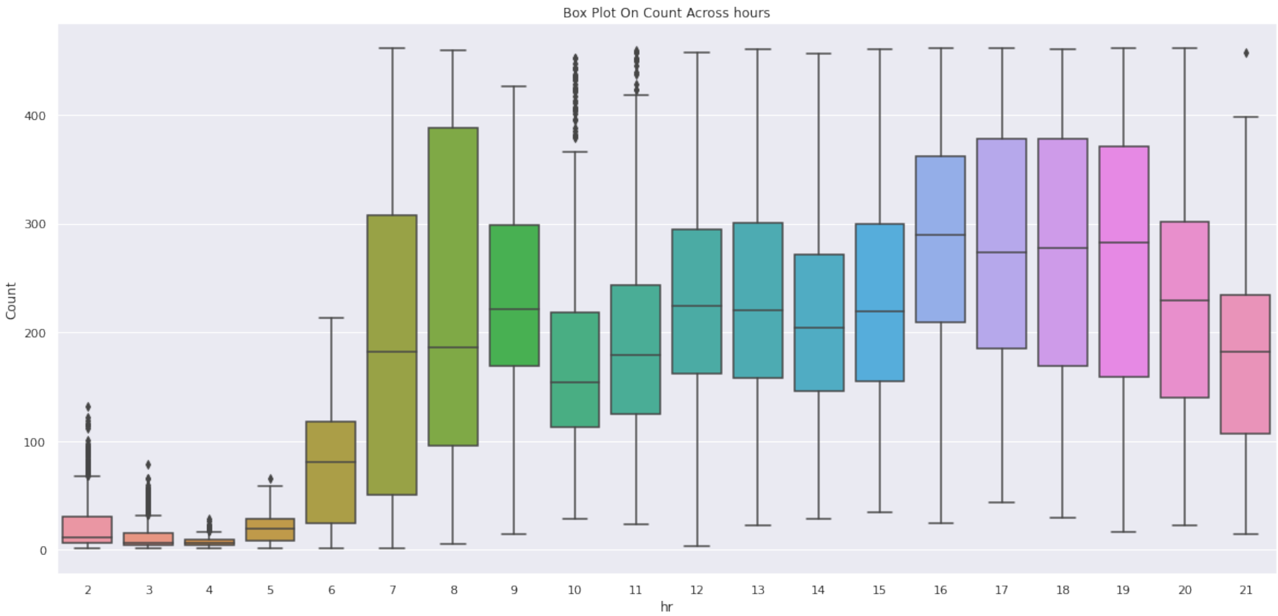

Figure 25: Boxplots

5.2.4 Correlation Analysis

	season	yr	mnth	hr	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt	day
season	1.000000	-0.019953	0.824975	-0.035685	-0.007487	0.005286	0.010787	0.003481	0.321950	0.331320	0.173502	-0.171743	0.137555	0.177126	0.185314	-0.005674
yr	-0.019953	1.000000	-0.015167	-0.056766	0.013395	-0.003657	0.022333	0.020754	-0.013187	-0.015346	-0.037004	-0.015417	0.022554	0.153447	0.131144	0.003613
mnth	0.824975	-0.015167	1.000000	-0.028515	0.023775	0.022637	-0.015504	0.023085	0.204715	0.212655	0.178673	-0.148260	0.082349	0.128628	0.129237	0.020666
hr	-0.035685	-0.056766	-0.028515	1.000000	0.000728	-0.002893	0.002292	-0.041659	0.136669	0.131775	-0.383440	0.179846	0.407492	0.546476	0.566598	-0.000016
holiday	-0.007487	0.013395	0.023775	0.000728	1.000000	-0.100729	-0.263006	-0.009186	-0.029783	-0.033943	-0.011469	-0.000988	0.053222	-0.048053	-0.022726	-0.003801
weekday	0.005286	-0.003657	0.022637	-0.002893	-0.100729	1.000000	0.052347	0.000025	-0.000932	-0.007997	-0.035562	0.021084	0.000874	0.026619	0.021826	0.002906
workingday	0.010787	0.022333	-0.015504	0.002292	-0.263006	0.052347	1.000000	0.043079	0.072582	0.072308	-0.016755	0.002811	-0.269182	0.153943	0.042779	0.001779
weathersit	0.003481	0.020754	0.023085	-0.041659	-0.009186	0.000025	0.043079	1.000000	-0.072076	-0.068170	0.308556	-0.045027	-0.080434	-0.025890	-0.045443	0.004735
temp	0.321950	-0.013187	0.204715	0.136669	-0.029783	-0.000932	0.072582	-0.072076	1.000000	0.988382	-0.022946	-0.035396	0.484589	0.298642	0.389328	0.037453
atemp	0.331320	-0.015346	0.212655	0.131775	-0.033943	-0.007997	0.072308	-0.068170	0.988382	1.000000	0.000255	-0.076143	0.477499	0.297604	0.386330	0.032951
hum	0.173502	-0.037004	0.178673	-0.383440	-0.011469	-0.035562	-0.016755	0.308556	-0.022946	0.000255	1.000000	-0.340159	-0.311593	-0.259411	-0.304915	0.023104
windspeed	-0.171743	-0.015417	-0.148260	0.179846	-0.000988	0.021084	0.002811	-0.045027	-0.035396	-0.076143	-0.340159	1.000000	0.082333	0.099764	0.105854	-0.000484
casual	0.137555	0.022554	0.082349	0.407492	0.053222	0.000874	-0.269182	-0.080434	0.484589	0.477499	-0.311593	0.082333	1.000000	0.510112	0.717430	-0.004389
registered	0.177126	0.153447	0.128628	0.546476	-0.048053	0.026619	0.153943	-0.025890	0.298642	0.297604	-0.259411	0.099764	0.510112	1.000000	0.965147	-0.012014
cnt	0.185314	0.131144	0.129237	0.566598	-0.022726	0.021826	0.042779	-0.045443	0.389328	0.386330	-0.304915	0.105854	0.717430	0.965147	1.000000	-0.011066
day	-0.005674	0.003613	0.020666	-0.000016	-0.003801	0.002906	0.001779	0.004735	0.037453	0.032951	0.023104	-0.000484	-0.004389	-0.012014	-0.011066	1.000000

Figure 26: Correlation matrix

5.2.5 plot of counts with respect to registered and casual on hourly basis

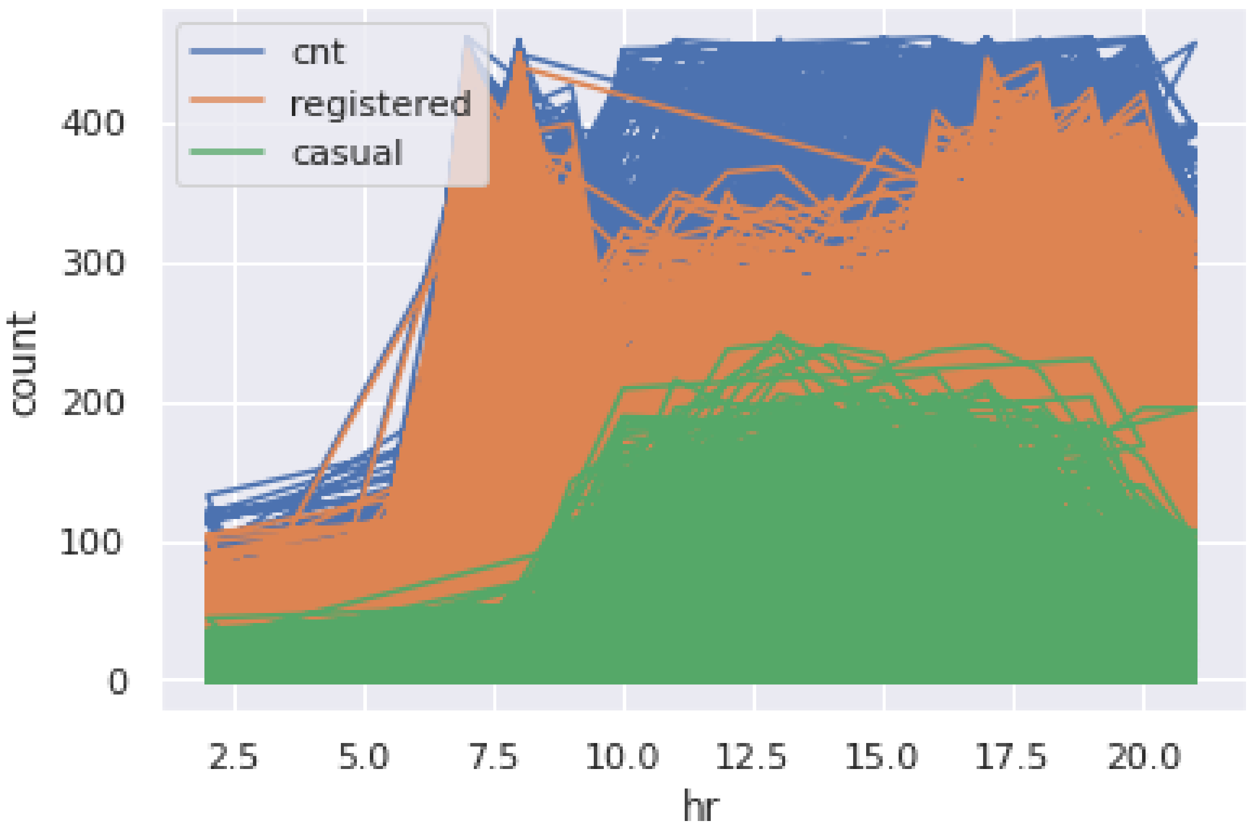

Figure 27

5.3. Model Building

5.3.1 Models built

Registered Bike Rental			
Model	MSE	r2score	Built for
DecisionTreeRegressor	1194.2214827563914	0.8901380510218632	working days
RandomForestRegressor	975.2334462516674	0.906314828991459	working days
AdaBoostRegressor	788.4156222355674	0.9225317899549949	working days

Registered Bike Rental			
Model	MSE	r2score	Built for
DecisionTreeRegressor	1181.8076493883377	0.8453814318008556	Non working days
RandomForestRegressor	876.532385627407	0.8765904308255217	Non working days
AdaBoostRegressor	709.9459906809523	0.8945840065748162	Non working days

Casual Bike Rental			
Model	MSE	r2score	Built for
DecisionTreeRegressor	173.25309804848297	0.5114963628560475	working days
RandomForestRegressor	164.2645480195017	0.7027711748205366	working days
AdaBoostRegressor	133.7554396642034	0.7764826321584307	working days

Casual Bike Rental			
Model	MSE	r2score	Built for
DecisionTreeRegressor	469.5691557373242	0.8170644311320279	Non working days
RandomForestRegressor	378.4419075512794	0.8409912990338057	Non working days
AdaBoostRegressor	311.34926821836183	0.8653122415936776	Non working days

All Bike Rental			
Model	MSE	r2score	Built for
DecisionTreeRegressor	1931.0749590598139	0.8773054178787941	All days
RandomForestRegressor	1749.0787279978333	0.8778275195853532	All days
AdaBoostRegressor	1589.0667073941697	0.8812372001789832	All days

In each case the model that have the highest r2score is taken as the best model .
That model can be taken and put into production .

6. Link for github repo of our project

Bike sharing demand prediction

7. Link for streamlit web app of our project where we deployed it

Bike sharing demand prediction Streamlit web app

8. Conclusion

Thus we have performed EDA to understand various things about. We have built various models for each case and finally taken two general models are taken and deployed them on streamlit .We can use this to predict on custom inputs.