NAME – PREETI MONDAL

ROLL NO – 21419STC036

ENROLLMENT NO - 436872

BIKE Rental

Prediction

ACKNOWLEDGEMENT

First, I wish to express my sincere gratitude to my supervisor, Ass. Professor Jyoti Sing Kirar, for her insightful comments, helpful information, practical advice, and unceasing ideas that have helped me at all times in my Project. Her immense knowledge, profound experience, and professional expertise in Regression Analysis and Statistics have enabled me to complete this project successfully. Without her support and guidance, this project would not have been possible.

I would like to express my sincere gratitude to my groupmates AKSHAY KUMAR JHA, SRIRAM RAVIKISHOR and ARYAN SING for supporting throughout this Project.

We are ensuring that this project is finished by us.

PREETI MONDAL

M.Sc. Statistics &

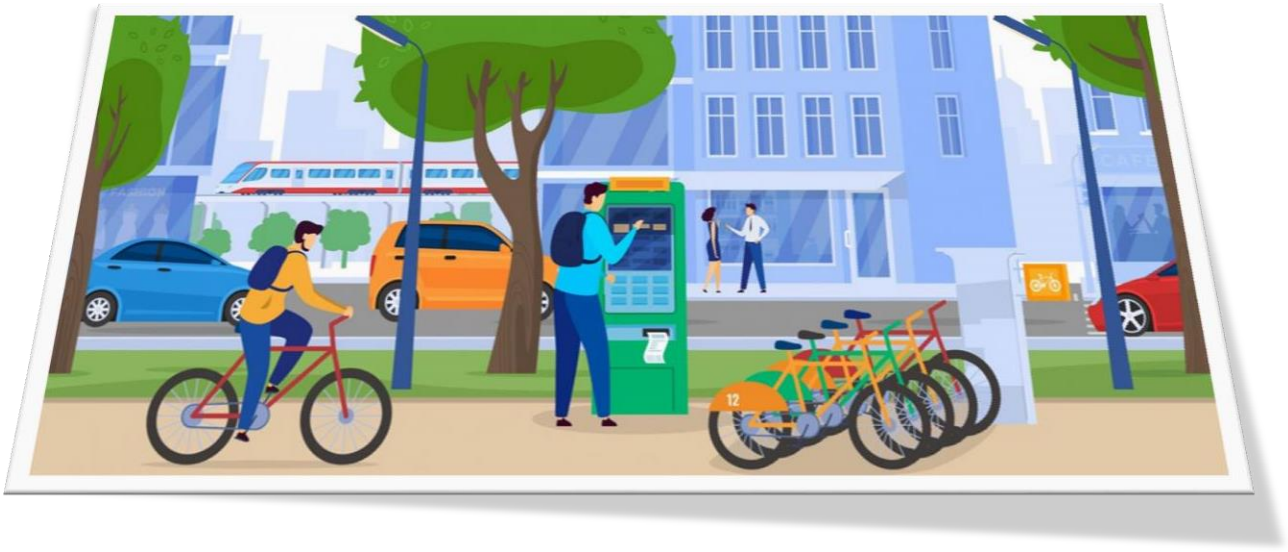
Computing

DST-CIMS, BHU

CONTENTS

Abstract – Sharing Bike has come a new concept now a days where one doesn't have to buy a bike to enjoy their daily ride. Currently It has introduced in so many urban countries for enhancement of mobility comfort. One can rent bikes on several basis like hourly, daily, monthly etc. It has a significant role to the rising issues related to the global warming, climate change, carbon emission and many more environmental anomalies. It is very important to make rental bikes available and accessible to the customers to the right time as it lessens the waiting time.

In this project, we choose to analyse a dataset containing the rental bike count and other climate related variables of Seoul city, The capital of South Korea. First, we have focused on the data pre-processing and data transformation, then we have used "Linear Regression" to predict the rent bike count required at each hour very efficiently.



Introduction

Bike Sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout acity. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis.

The first bike-share programs began in 1960s Europe, but the concept did not take off worldwideuntil the mid-2000s. In North America, they tend to be affiliated with municipal governments, though some programs, particularly in small college towns, centre on university campuses.

The typical bike-share has several defining characteristics and features, including station-based bikes and payment systems, membership, and pass fees, and per-hour usage fees. Programs are generally intuitive enough for novice users to understand. And, despite some variation, the differences are usually small enough to prevent confusion when a regular user of one city's bike-share uses another city's program for the first time.

With the onset of Industry 4.0, integration of Internet of Things (IoT) systems with bike-sharing ecosystem has eased the rental process to a significant extent. Real-time tracking of bikes, trafficdensity, and climate variables aids in gaining useful knowledge about trends, and patterns of renting process, thereby allowing an incisive prediction to meet future demand.

Considering the current ecosystem, bike-sharing can play a vital role in reducing the impact of carbon emissions and other greenhouse gases- major contributors in climate change. Sustainableand clean transport system, if successful, can provide a greener alternative to the traditional car-pool system, and help in reducing traffic congestion, too.

In addition to the environmental benefits, the sharing systems will impart healthier habits amongcommuting public, who in the hustle of tasking daily routine, often are unable to integrate optimum level of physical activity, which results in a barrage of ailments.

Exploratory Data Analysis

In this part of EDA each individual features are analysed by proper statistical methods.

So first let's look at the dataset attributes:-

Attribute Information:

- Date : year-month-day
- Rented Bike count - Count of bikes rented at each hour
- Hour - Hour of the day
- Temperature-Temperature in Celsius
- Humidity - %
- Wind speed - m/s
- Visibility - 10m
- Dew point temperature - Celsius
- Solar radiation - MJ/m2
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday/No holiday
- Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

Dataset Statistics

Number of variables	**14**
Number of observations	8760
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
DateTime	1

Numeric	10
Categorical	2
Boolean	1

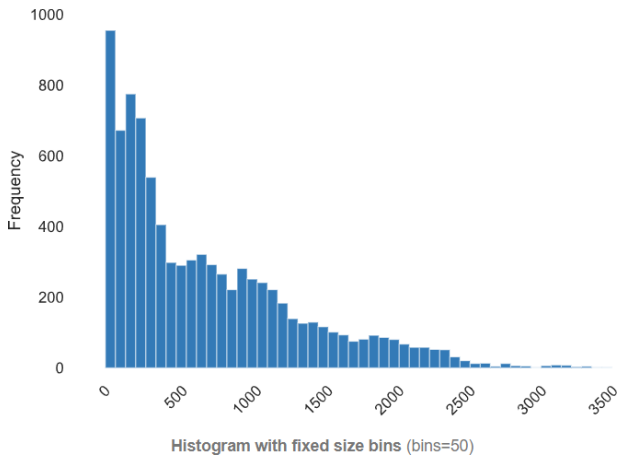
Variables

❖ Date

Distinct	**365**
Distinct (%)	4.2%
Missing	0
Missing (%)	0.0%
Minimum	2017-01-12 00:00:00
Maximum	2018-12-11 00:00:00

❖ Bike Count

Distinct	**2166**
Distinct (%)	24.7%
Missing	0
Mean	704.6020548
Minimum	0
Maximum	3556
Zeros	295
Zeros (%)	3.4%



Negative	0

Quantile statistics

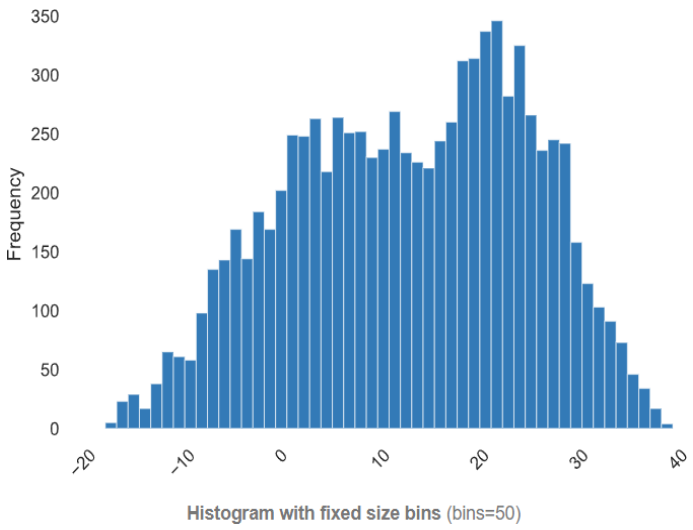
Minimum	**0**
5-th percentile	22
Q1	191
median	504.5
Q3	1065.25
95-th percentile	2043
Maximum	3556
Range	3556
Interquartile range (IQR)	874.25

Descriptive statistics

Standard deviation	**644.9974677**
Coefficient of variation (CV)	0.9154067368
Kurtosis	0.8533869902
Mean	704.6020548
Median Absolute Deviation (MAD)	373.5
Skewness	1.153428177
Sum	6172314
Variance	416021.7334
Monotonicity	Not monotonic

❖ Temperature

Distinct	**546**
Distinct (%)	6.2%
Missing	0
Mean	12.88292237
Minimum	-17.8
Maximum	39.4
Zeros	21
Zeros (%)	0.2%
Negative	1433
Negative (%)	16.4%



Quantile statistics

Minimum	**-17.8**
5-th percentile	-7.1
Q1	3.5
median	13.7
Q3	22.5
95-th percentile	30.7
Maximum	39.4
Range	57.2
Interquartile range (IQR)	19

Descriptive statistics

Standard deviation	**11.94482523**
Coefficient of variation (CV)	0.9271828924
Kurtosis	-0.837786292

Mean	12.88292237
Median Absolute Deviation (MAD)	9.4
Skewness	-0.1983255345
Sum	112854.4
Variance	142.6788498
Monotonicity	Not monotonic

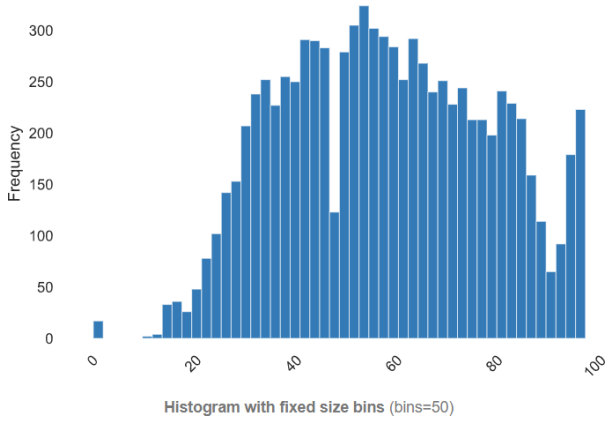
❖ Hour

Distinct	**24**
Distinct (%)	0.3%
Missing	0
Mean	11.5
Minimum	0
Maximum	23
Zeros	365
Zeros (%)	4.2%
Negative	0

❖ Humidity

Distinct	**90**
Distinct (%)	1.0%
Missing	0

Mean	58.22625571
Minimum	0
Maximum	98
Zeros	17
Zeros (%)	0.2%
Negative	0



Quantile statistics

Minimum	**0**
5-th percentile	27
Q1	42
median	57
Q3	74
95-th percentile	94
Maximum	98
Range	98
Interquartile range (IQR)	32

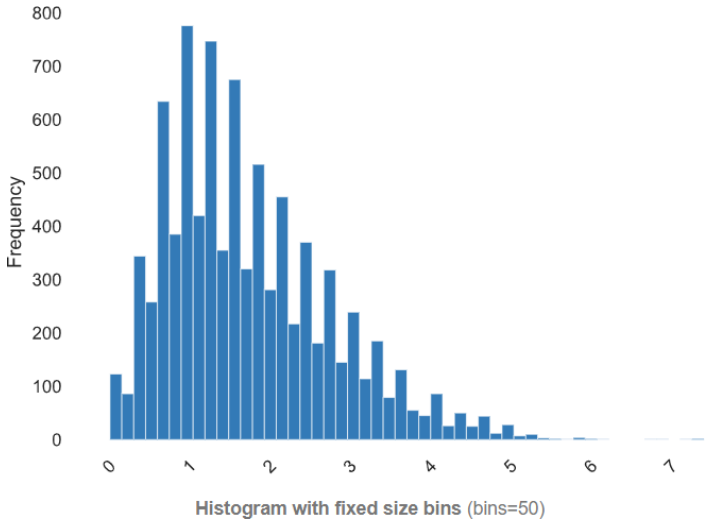
Descriptive statistics

Standard deviation	**20.3624133**
Coefficient of variation (CV)	0.3497118792
Kurtosis	-0.8035591857
Mean	58.22625571
Median Absolute Deviation (MAD)	16
Skewness	0.05957897258

Sum	510062
Variance	414.6278755
Monotonicity	Not monotonic

❖ Wind Speed

Distinct	**65**
Distinct (%)	0.7%
Missing	0
Mean	1.724908676
Minimum	0
Maximum	7.4
Zeros	74
Zeros (%)	0.8%
Negative	0



Quantile statistics

Minimum	**0**
5-th percentile	0.4
Q1	0.9
median	1.5
Q3	2.3
95-th percentile	3.7
Maximum	7.4
Range	7.4

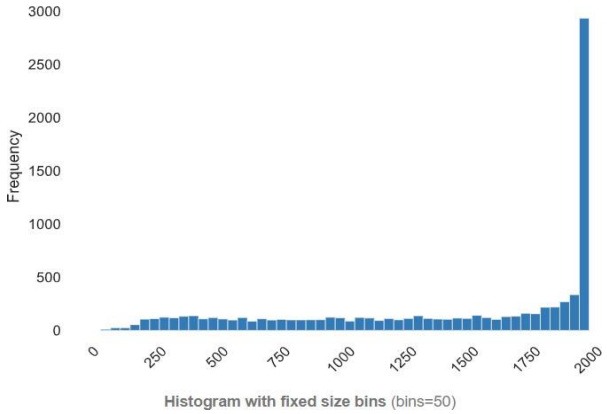
Interquartile range (IQR)	1.4

Descriptive statistics

Standard deviation	**1.036299993**
Coefficient of variation (CV)	0.6007854259
Kurtosis	0.7271794546
Mean	1.724908676
Median Absolute Deviation (MAD)	0.7
Skewness	0.890954798
Sum	15110.2
Variance	1.073917676
Monotonicity	Not monotonic

❖ Visibility

Distinct	**1789**
Distinct (%)	20.4%
Missing	0
Mean	1436.825799
Minimum	27
Maximum	2000
Zeros	0
Zeros (%)	0.0%
Negative	0



Quantile statistics

Minimum	**27**
5-th percentile	300
Q1	940
median	1698
Q3	2000
95-th percentile	2000
Maximum	2000
Range	1973
Interquartile range (IQR)	1060

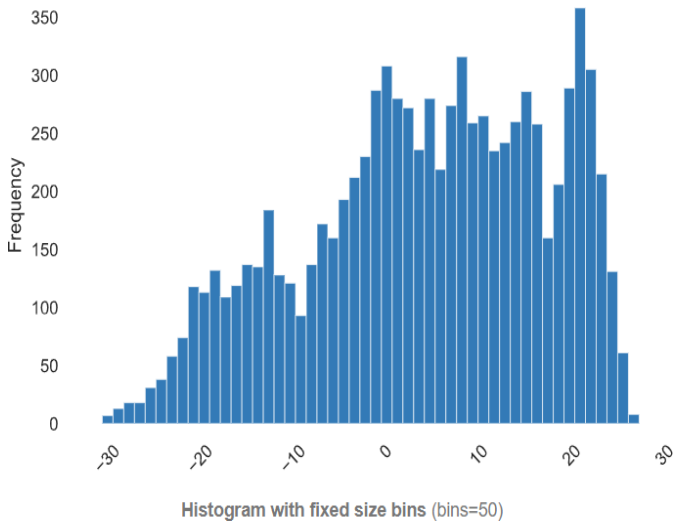
Descriptive statistics

Standard deviation	**608.298712**
Coefficient of variation (CV)	0.4233628825
Kurtosis	-0.961980131
Mean	1436.825799
Median Absolute Deviation (MAD)	302
Skewness	-0.701786449
Sum	12586594
Variance	370027.323
Monotonicity	Not monotonic

❖ Dew Point Temperature

Distinct	**556**
Distinct (%)	6.3%

Missing	0
Mean	4.073812785
Minimum	-30.6
Maximum	27.2
Zeros	60
Zeros (%)	0.7%
Negative	3138
Negative (%)	35.8%



Quantile statistics

Minimum	**-30.6**
5-th percentile	-19.505
Q1	-4.7
median	5.1
Q3	14.8
95-th percentile	22.405
Maximum	27.2
Range	57.8
Interquartile range (IQR)	19.5

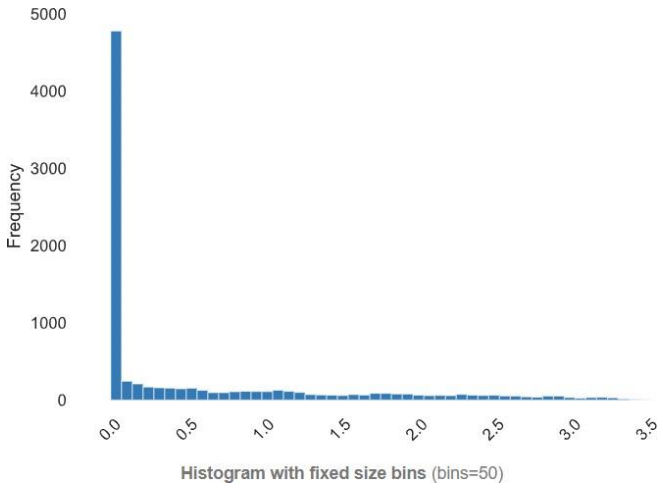
Descriptive statistics

Standard deviation	**13.06036934**
Coefficient of variation (CV)	3.20593263
Kurtosis	-0.7554295071
Mean	4.073812785

Median Absolute Deviation (MAD)	9.7
Skewness	-0.3672984397
Sum	35686.6
Variance	170.5732472
Monotonicity	Not monotonic

❖ Solar Radiation

Distinct	**345**
Distinct (%)	3.9%
Missing	0
Mean	0.5691107306
Minimum	0
Maximum	3.52
Zeros	4300
Zeros (%)	49.1%
Negative	0



Quantile statistics

Minimum	**0**
5-th percentile	0
Q1	0
median	0.01
Q3	0.93
95-th percentile	2.56

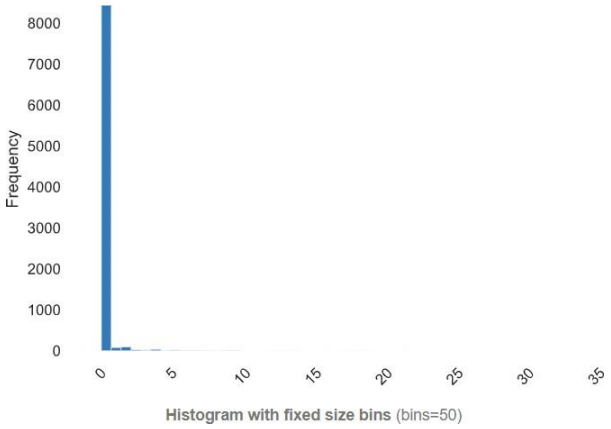
Maximum	3.52
Range	3.52
Interquartile range (IQR)	0.93

Descriptive statistics

Standard deviation	**0.8687462422**
Coefficient of variation (CV)	1.526497737
Kurtosis	1.126432996
Mean	0.5691107306
Median Absolute Deviation (MAD)	0.01
Skewness	1.504039717
Sum	4985.41
Variance	0.7547200334
Monotonicity	Not monotonic

❖ Rainfall

Distinct	**61**
Distinct (%)	0.7%
Missing	0
Mean	0.1486872146
Minimum	0
Maximum	35
Zeros	8232
Zeros (%)	94.0%
Negative	0



Quantile statistics

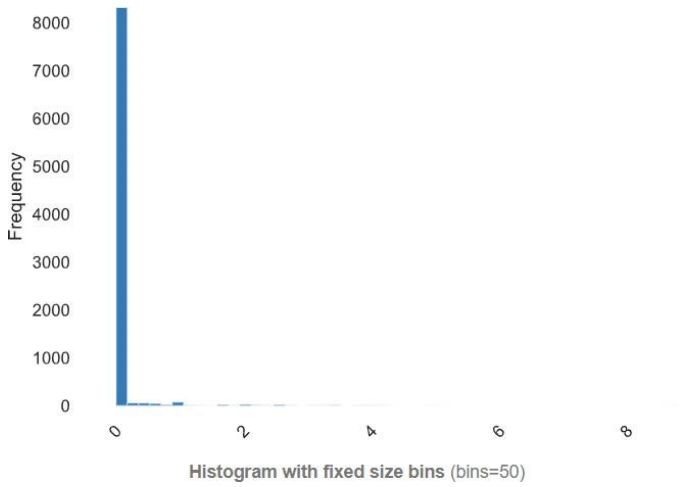
Minimum	**0**
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0.4
Maximum	35
Range	35
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	**1.128192969**
Coefficient of variation (CV)	7.58769321
Kurtosis	284.9910986
Mean	0.1486872146
Median Absolute Deviation (MAD)	0
Skewness	14.53323224
Sum	1302.5
Variance	1.272819375
Monotonicity	Not monotonic

❖ Snowfall

Distinct	**51**
Distinct (%)	0.6%
Missing	0
Mean	0.07506849315
Minimum	0
Maximum	8.8
Zeros	8317
Zeros (%)	94.9%
Negative	0



Quantile statistics

Minimum	**0**
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0.2
Maximum	8.8
Range	8.8
Interquartile range (IQR)	0

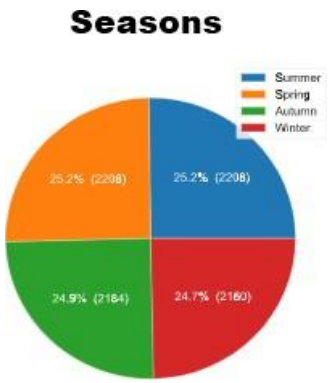
Descriptive statistics

Standard deviation	**0.4367461811**
Coefficient of variation (CV)	5.817969201
Kurtosis	93.80332357

Mean	0.07506849315
Median Absolute Deviation (MAD)	0
Skewness	8.440800781
Sum	657.6
Variance	0.1907472267
Monotonicity	Not monotonic

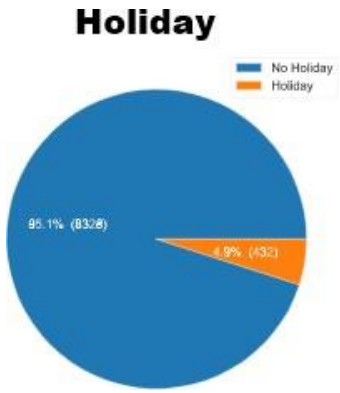
❖ Seasons

Distinct	**4**
Distinct (%)	< 0.1%
Missing	0
Max length	6
Median length	6
Mean length	6
Min length	6



❖ Holiday

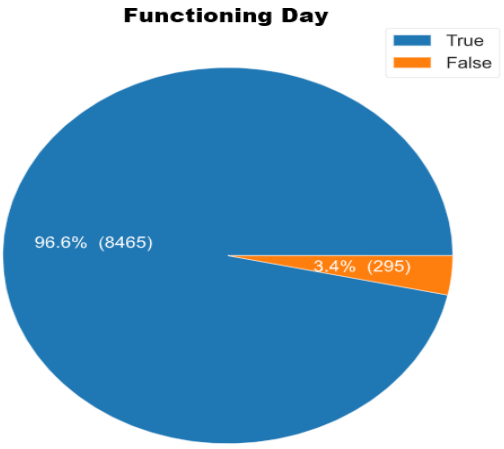
Distinct	**2**
Distinct (%)	< 0.1%
Missing	0
Max length	10
Median length	10



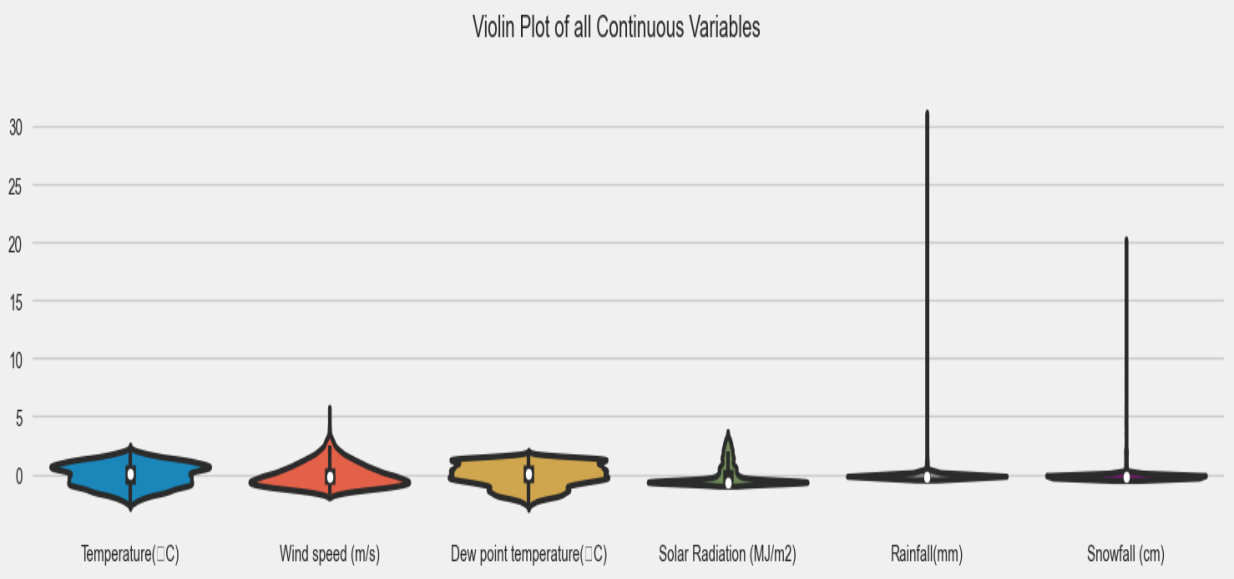
Mean length	9.852054795
Min length	7

❖Functional Day

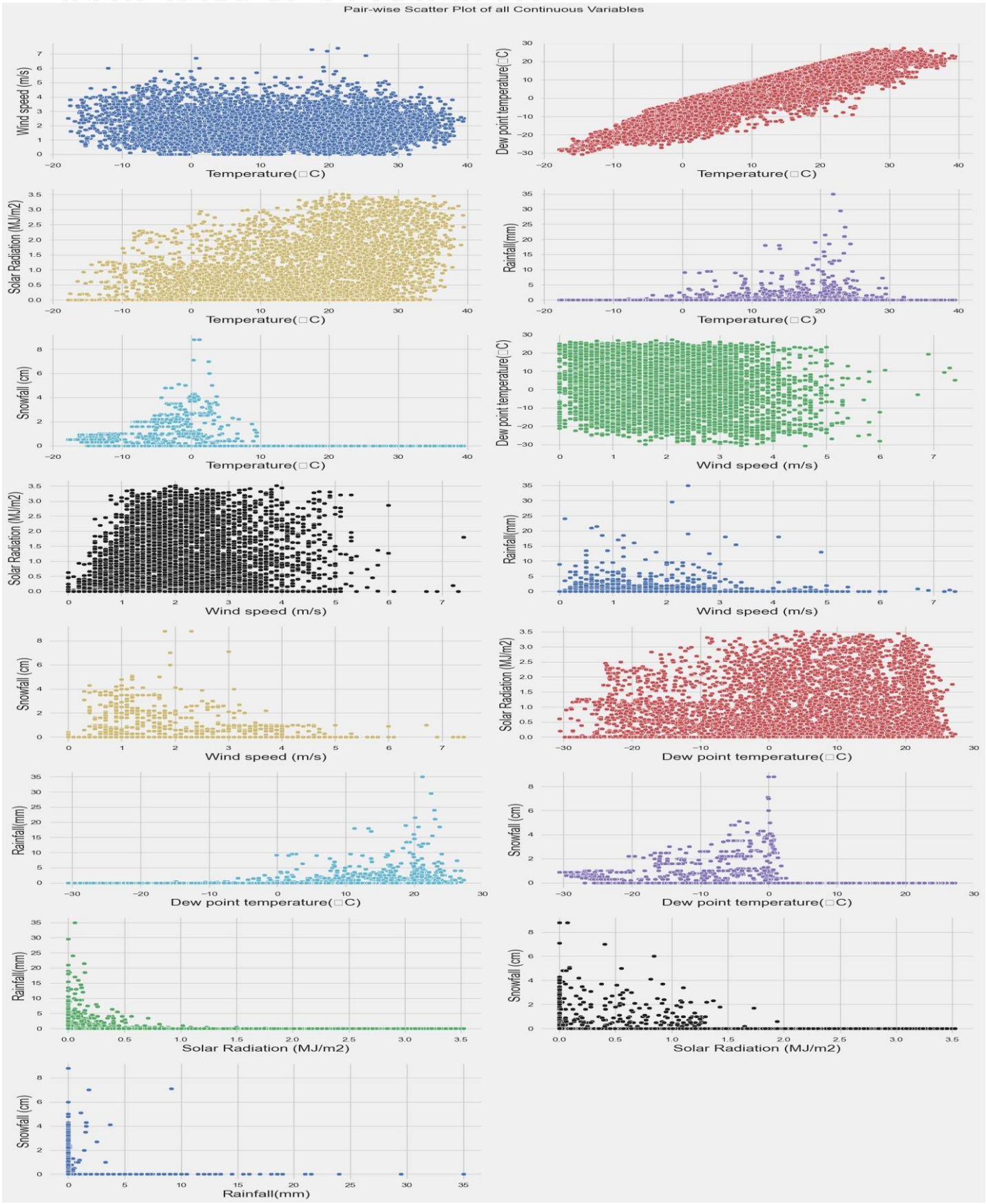
Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	8.7 KiB



VIOLIN PLOT



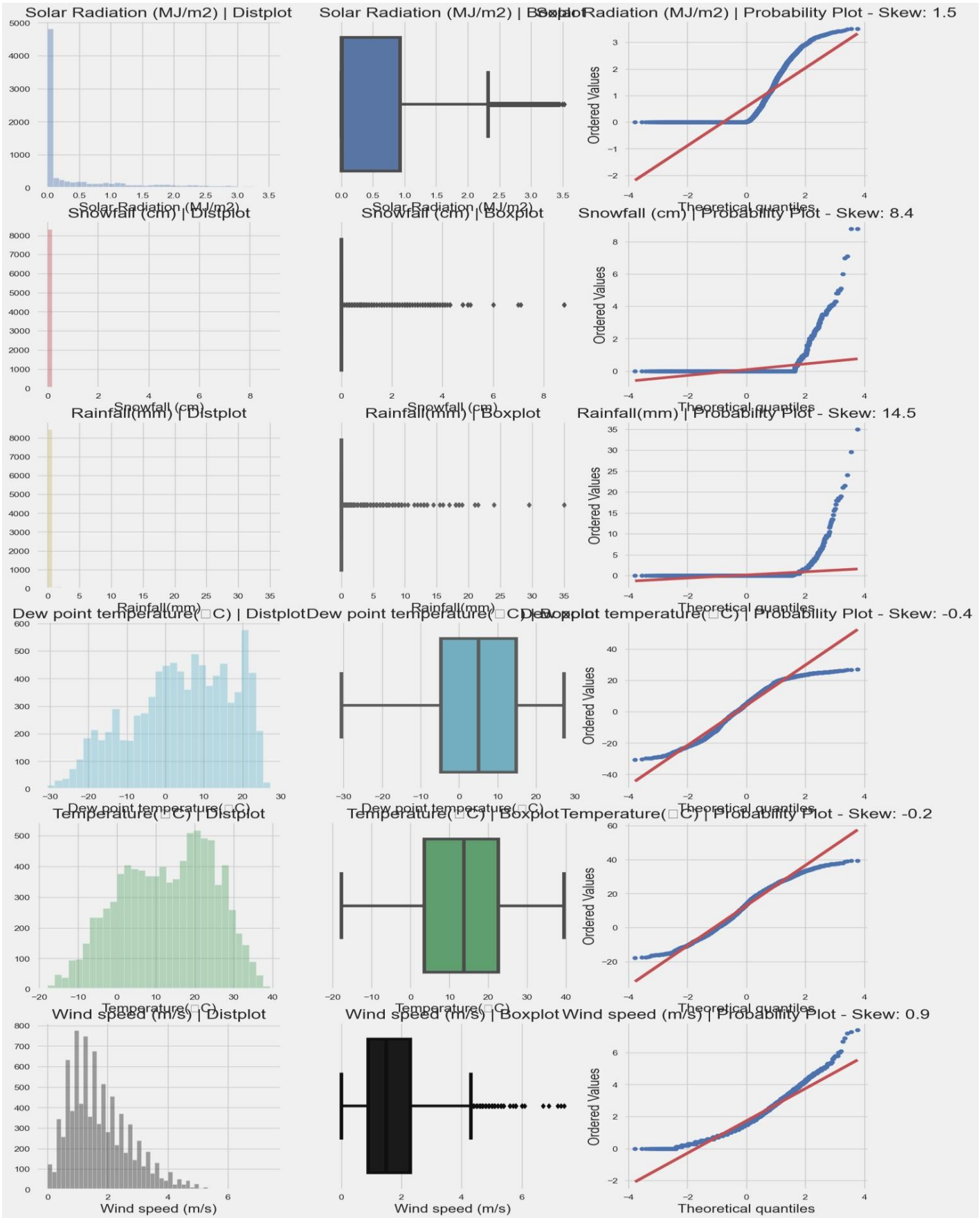
PAIR-WISE SCATTER PLOT



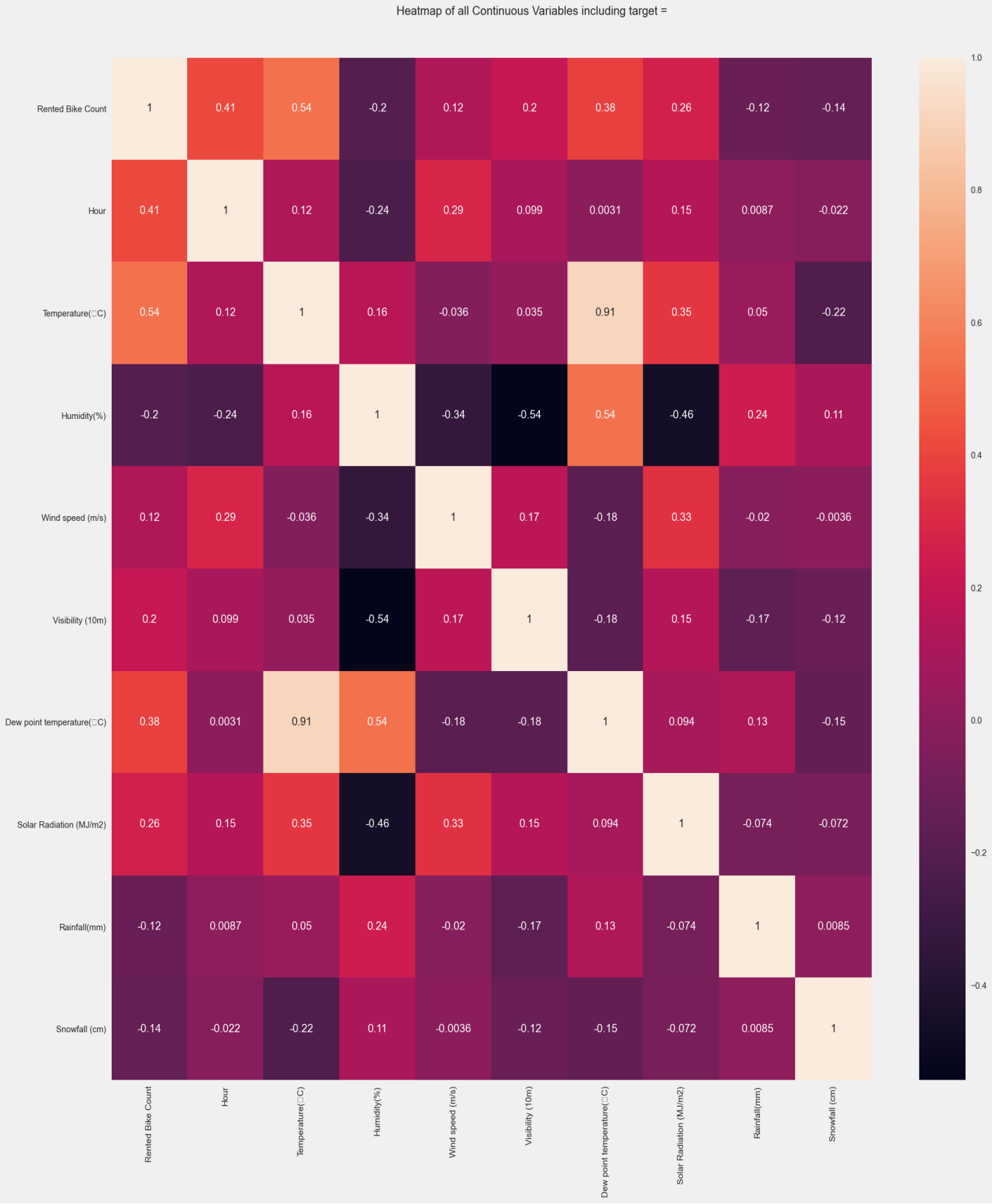
HISTOGRAM PLOT

BOX PLOT

QQ-PLOT



SCATTER PLOT



INSIGHTS OF OUR DATA VISUALIZATIONS

- **Rented Bike Count (Dependent Variable) and wind speed (Independent Variable) are Positive skewed and their skewness are 1.15 and 0.89.**
- **There are many outliers present in Rainfall and Snowfall.**
- **The data points of temperature and Dew point temperature are nearly symmetrical.**
- **Rental Bike Count (Dependent Variable) is highly correlated with Temperature (Independent Variable).**
- **Temperature (Independent Variable) is highly correlated with Dew point Temperature (Independent variable).**
- **Solar Radiation has 4300 (49.1%) zero values**
- **Rainfall has 8232 (94.0%) zero values**
- **Snowfall has 8317 (94.9%) zero values**
- **The categories in Seasons attribute (summer, Spring, Autumn, Winter) are equally distributed in our dataset.**

Data Preprocessing

▪ Data Cleaning

- **Missing Values**

  There are 14 attributes in our dataset and there is no any missing value in any of the attributes.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8760 entries, 0 to 8759
Data columns (total 14 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Date                        8760 non-null   object
 1   Rented Bike Count           8760 non-null   int64
 2   Hour                        8760 non-null   int64
 3   Temperature(C)              8760 non-null   float64
 4   Humidity(%)                 8760 non-null   int64
 5   Wind speed (m/s)            8760 non-null   float64
 6   Visibility (10m)            8760 non-null   int64
 7   Dew point temperature(C)    8760 non-null   float64
 8   Solar Radiation (MJ/m2)     8760 non-null   float64
 9   Rainfall(mm)                8760 non-null   float64
 10  Snowfall (cm)               8760 non-null   float64
 11  Seasons                     8760 non-null   object
 12  Holiday                     8760 non-null   object
 13  Functioning Day             8760 non-null   object
dtypes: float64(6), int64(4), object(4)
memory usage: 958.2+ KB
```

▪ Data Transformation

- **One Hot Encoding :** a **one-hot** is a group of bits among which the legal combinations of values are only those with a single high (1) bit and all the others low (0). One-hot encoding is often used for indicating the state of a state machine. When using binary or Gray code, a decoder is needed to determine the state. A one-hot state machine, however, does not need a decoder as the state machine is in the *n*th state if and only if the *n*th bit is high.
  We have considered four categorical attributes for one hot encoding and these variables are **Hour,Seasons,Holiday,Functioning Day.**

After One-Hot Encoding some variables increased in our dataset

```
Index(['Date', 'Rented Bike Count', 'Temperature(°C)', 'Humidity(%)',
       'Wind speed (m/s)', 'Visibility (10m)', 'Dew point temperature(°C)',
       'Solar Radiation (MJ/m2)', 'Rainfall(mm)', 'Snowfall (cm)',
       'Holiday_Holiday', 'Holiday_No Holiday', 'Seasons_Autumn',
       'Seasons_Spring', 'Seasons_Summer', 'Seasons_Winter',
       'Functioning Day_No', 'Functioning Day_Yes', 'Hour_0', 'Hour_1',
       'Hour_2', 'Hour_3', 'Hour_4', 'Hour_5', 'Hour_6', 'Hour_7', 'Hour_8',
       'Hour_9', 'Hour_10', 'Hour_11', 'Hour_12', 'Hour_13', 'Hour_14',
       'Hour_15', 'Hour_16', 'Hour_17', 'Hour_18', 'Hour_19', 'Hour_20',
       'Hour_21', 'Hour_22', 'Hour_23'],
      dtype='object')
```

- **Standardization :** The standard score of a sample x is calculated as
  z = (x - u) / s
  Standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data (e.g. Gaussian with 0 mean and unit variance).
  Whole dataset is Standardized by the class Standard Scaler.

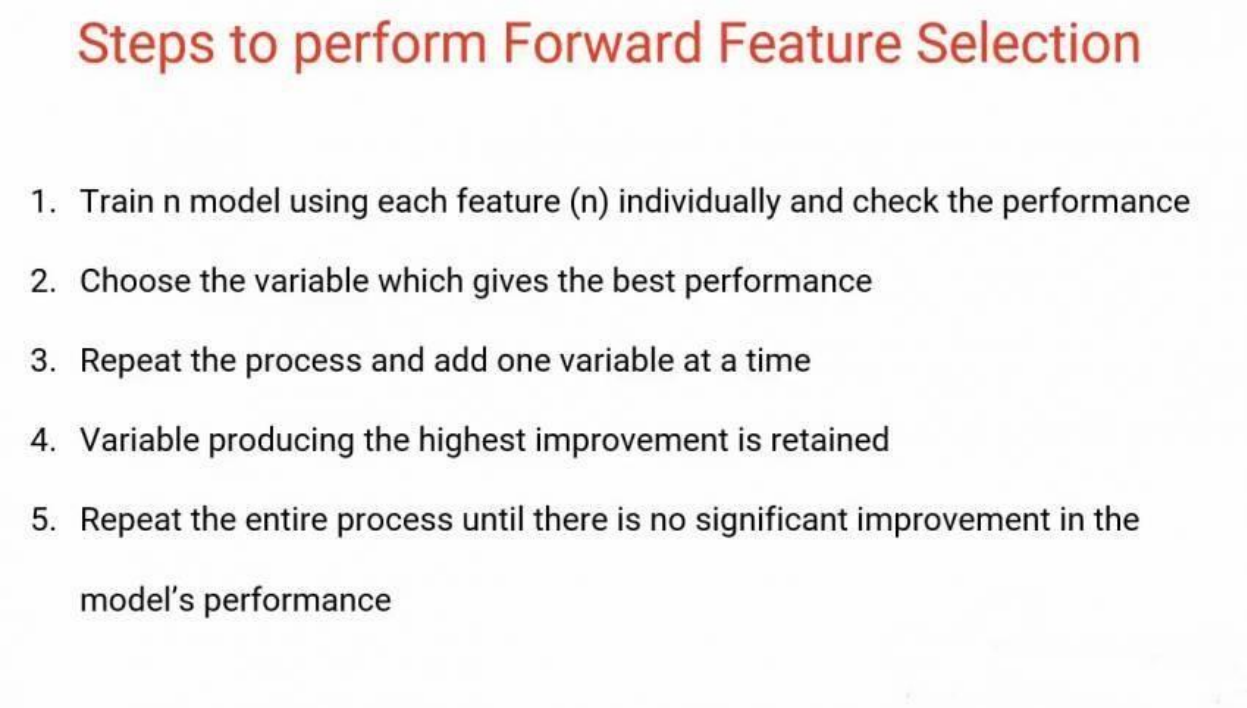
  After applying Standard Scaler in our dataset

```
[[-0.69865046 -1.51395724 -1.04248288 ... -0.20851441 -0.20851441
  -0.20851441]
 [-0.77617457 -1.53907415 -0.99336999 ... -0.20851441 -0.20851441
  -0.20851441]
 [-0.82423951 -1.58093567 -0.94425709 ... -0.20851441 -0.20851441
  -0.20851441]
 ...
 [-0.0164383  -0.86091752 -0.94425709 ...  4.79583152 -0.20851441
  -0.20851441]
 [ 0.01147038 -0.90277904 -0.8460313  ... -0.20851441  4.79583152
  -0.20851441]
 [-0.18699134 -0.91952365 -0.74780551 ... -0.20851441 -0.20851441
   4.79583152]]
```

Feature Selection

- **Forward Feature Selection**

Forward Feature Selection is to train n models using each feature individually and checking the performance. So if you have three independent variables, we will train three models using each of these three features individually .



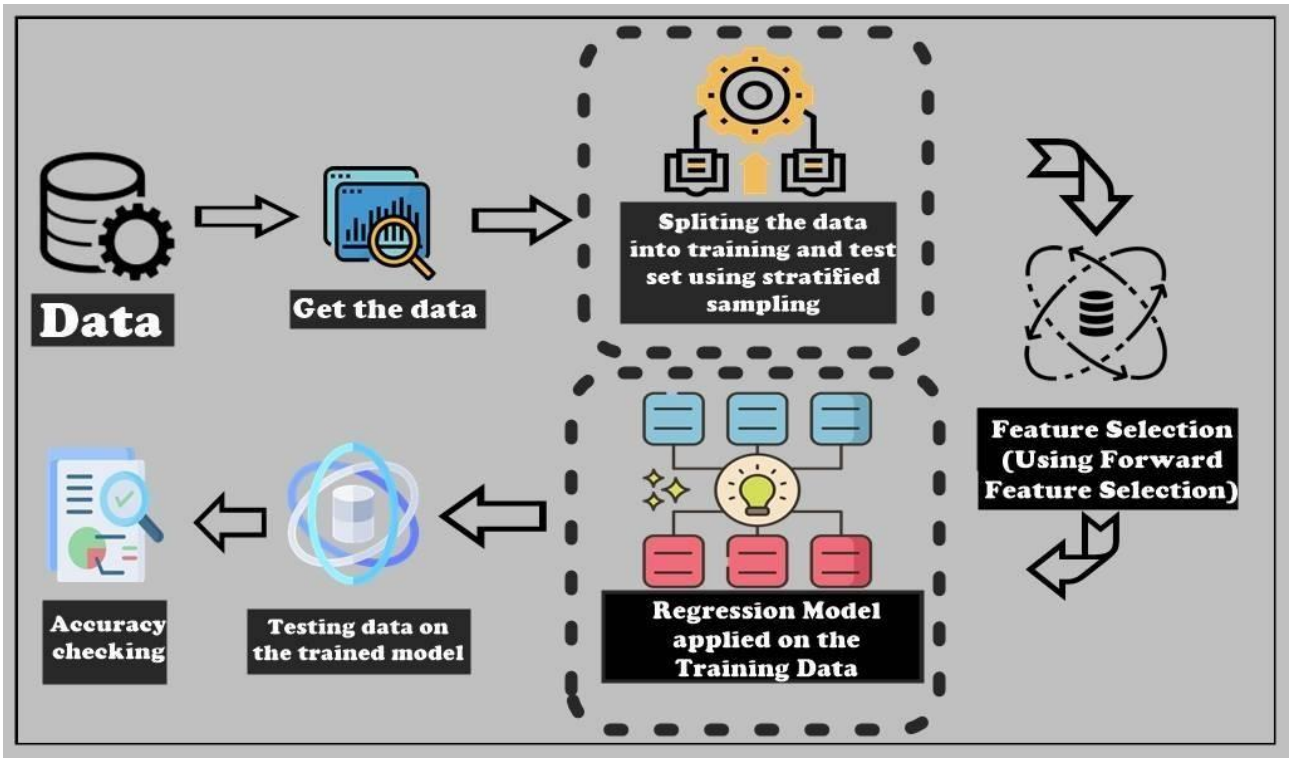
Each individual accuracy for different number of features in our model.

```
{1: 0.26730074786068014,
2: 0.470858021787809,
3: 0.5856553463150498,
4: 0.6311192862328596,
5: 0.6563970661427261,
6: 0.6799649834207429,
7: 0.6977112704795301,
8: 0.7152804174388814,
9: 0.7313535640856988,
10: 0.7424889291410854,
11: 0.7525072538157692,
12: 0.7595327258022639,
13: 0.7669392996384695,
14: 0.77287347113809,
15: 0.7783699334615449,
16: 0.784337887575459,
```

```
17: 0.7889450855082414,
18: 0.7940760200330776,
19: 0.7961989596579642,
20: 0.7978050281453273,
21: 0.7989428711131319,
22: 0.7998778832199286,
23: 0.8005563204861882,
24: 0.8012500787889908,
25: 0.8017588079547893,
26: 0.8027400452841543,
27: 0.8038855573969617,
28: 0.8064158435626755,
29: 0.8073601296528087,
30: 0.8077159976037313,
31: 0.807977216413485,
32: 0.8080482523538419,
33: 0.8081223477207902,
34: 0.8081382584674834,
35: 0.8081455442342659,
36: 0.808147444898313,
37: 0.8081474752575182,
38: 0.8081443434221877}
```

On the basis of forward feature selection we have selected all 38 features for our model.

Model Selection



▪ Multiple Linear Regression

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables. In essence, multiple regression is the extension of ordinary least-squares (OLS) regression because it involves more than one explanatory variable.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

where, for $i=n$ observations:

Y_i = Dependent variable

x_i =Explanatory variables

β_0 = Y-intercept (constant term)

β_p=Slope coefficients for each explanatory

variable

ϵ=the model's error term (also known as the residuals)

The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables.
- The independent variables are not too highly correlated with each other.
- Y_i observations are selected independently and randomly from the population.
- Residuals should be normally distributed with a mean of 0 and variance σ.

The coefficient of determination (R-squared) is a statistical metric that is used to measure how much of the variation in outcome can be explained by the variation in the independent variables. R^2 always increases as more predictors are added to the MLR model, even though the predictors may not be related to the outcome variable. R^2 by itself can't thus be used to identify which predictors should be included in a model and which should be excluded. So, we use Adjusted-R^2 to decide the test accuracy.

OLS Regression Results

Dep. Variable:	y	**R-squared (uncentered):**	0.809

Model:	OLS	**Adj. R-squared (uncentered):**	0.808
Method:	Least Squares	**F-statistic:**	770.6
Date:	Thu, 28 Apr 2022	**Prob (F-statistic):**	0.00
Time:	11:31:16	**Log-Likelihood:**	-3877.7
No. Observations:	6570	**AIC:**	7827.
Df Residuals:	6534	**BIC:**	8072.
Df Model:	36		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x1	6.025e+10	1.1e+11	0.546	0.585	-1.56e+11	2.76e+11
x2	6.033e+10	1.1e+11	0.546	0.585	-1.56e+11	2.77e+11
x3	6.02e+10	1.1e+11	0.546	0.585	-1.56e+11	2.76e+11
x4	6.028e+10	1.1e+11	0.546	0.585	-1.56e+11	2.77e+11
x5	0.3638	0.013	28.955	0.000	0.339	0.388
x6	-0.1194	0.009	-13.084	0.000	-0.137	-0.102
x7	-0.0172	0.006	-2.658	0.008	-0.030	-0.005
x8	0.0050	0.007	0.719	0.472	-0.009	0.019
x9	0.2349	0.016	14.651	0.000	0.203	0.266
x10	-0.2554	0.006	-42.001	0.000	-0.267	-0.244
x11	0.0046	0.006	0.777	0.437	-0.007	0.016
x12	-0.0455	0.005	-8.312	0.000	-0.056	-0.035

x13	0.4877	0.006	86.895	0.000	0.477	0.499
x14	-0.0202	0.007	-2.698	0.007	-0.035	-0.006
x15	-0.0585	0.008	-7.774	0.000	-0.073	-0.044
x16	-0.1068	0.008	-14.125	0.000	-0.122	-0.092
x17	-0.1428	0.008	-18.919	0.000	-0.158	-0.128
x18	-0.1970	0.008	-26.237	0.000	-0.212	-0.182
x19	-0.1803	0.007	-24.256	0.000	-0.195	-0.166
x20	-0.0961	0.007	-13.070	0.000	-0.110	-0.082
x21	-0.0181	0.008	-2.385	0.017	-0.033	-0.003
x22	0.0428	0.008	5.319	0.000	0.027	0.059
x23	-0.0583	0.009	-6.641	0.000	-0.076	-0.041
x24	-0.1383	0.009	-14.622	0.000	-0.157	-0.120
x25	-0.1389	0.010	-14.239	0.000	-0.158	-0.120
x26	-0.1134	0.010	-11.631	0.000	-0.133	-0.094
x27	-0.1249	0.010	-12.581	0.000	-0.144	-0.105
x28	-0.1216	0.010	-12.447	0.000	-0.141	-0.102
x29	-0.1004	0.010	-10.483	0.000	-0.119	-0.082
x30	-0.0755	0.009	-8.061	0.000	-0.094	-0.057

x31	-0.0107	0.009	-1.211	0.226	-0.028	0.007
x32	0.0927	0.008	11.535	0.000	0.077	0.108
x33	0.0616	0.008	8.105	0.000	0.047	0.076
x34	0.0661	0.008	8.808	0.000	0.051	0.081
x35	0.0752	0.007	10.103	0.000	0.061	0.090
x36	0.0497	0.007	6.684	0.000	0.035	0.064

Omnibus:	311.672	**Durbin-Watson:**	2.011
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	588.843
Skew:	-0.358	**Prob(JB):**	1.36e-128
Kurtosis:	4.280	**Cond. No.**	6.76e+13

From the above plot we can say that "Temperature" has the most linear relationship with the "Rented Bike Count" among others. Next, we have plotted the scatter plots of the "Rented Bike Count" with the other independent continuous variables

We have finished our Data preprocessing phase of our analysis.

Next, we have split the dataset into training and testing using Stratified Sampling technique based on "Season", 80% of the samples are used to train the proposed model and rest to validate that our model is doing well with the previously unseen data. Here we have used Stratification technique based on "Season" to avoid over or underfitting. There may be a situation that for some seasons our model is working very well, but for other it's giving bad results. Stratification will remove this issue.

Now, it's time to build some regression models based on our training data. In this project, we have used "Linear Regression" to get desired results.

- # Conclusion

  From the above results we can see that the Adjusted R^2 score for Multiple Linear Regression model is 0.808 which means our model is able to explain 80.8% of the variability in the dataset.

I. REFERENCES

II. V E, S., Park, J., & Cho, Y. (2020). Seoul bike trip duration prediction using data mining techniques. *IET Intelligent Transport Systems*, *14*(11), 1465–1474.

III. V E, S., & Cho, Y. (2020). A rule-based model for Seoul Bike sharing demand prediction using weather data. *European Journal of Remote Sensing*, *53*(sup1), 166–183.

IV. Liu, X., & Pelechrinis, K. (2021). Excess demand prediction for bike sharing systems. *PLOS ONE*, *16*(6), e0252894.

V. V E, S., & Cho, Y. (2020b). Season wise bike sharing demand analysis using random forest algorithm. *Computational Intelligence*. Published.

VI. Wang, Z. (2019). Regression Model for Bike-Sharing Service by Using Machine Learning. *Asian Journal of Social Science Studies*, *4*(4), 16.

VII. Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, *54*, 101882.

VIII. Goh, C. Y., Yan, C., & Jaillet, P. (2019). Estimating Primary Demand in Bike-sharing Systems. *SSRN Electronic Journal*. Published.

IX. Almannaa, M. H., Elhenawy, M., & Rakha, H. A. (2019). Dynamic linear models to predict bike availability in a bike sharing system. *International Journal of Sustainable Transportation*, *14*(3), 232–242.

X. Sachdeva, P., & Sarvanan, K. N. (2017). Prediction of Bike Sharing Demand. *Oriental Journal of Computer Science and Technology*, *10*(1), 219–226.

XI. Liu, X. N., Wang, J. J., & Zhang, T. F. (2014). A Method of Bike Sharing Demand Forecasting. *Applied Mechanics and Materials*, *587–589*, 1813–1816.