Step 1: Reading and Understanding the Data

In [2]:

```python
# Supress Warnings

import warnings

warnings.filterwarnings('ignore')
```

In [3]:

```python
import numpy as np
import pandas as pd
```

In [4]:

```python
df = pd.read_csv("day.csv")
```

In [5]:

```python
df.head()
```

Out[5]:

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	reg
0	1	01-01-2018	1	0	1	0	6	0	2	14.110847	18.18125	80.5833	10.749882	331	
1	2	02-01-2018	1	0	1	0	0	0	2	14.902598	17.68695	69.6087	16.652113	131	
2	3	03-01-2018	1	0	1	0	1	1	1	8.050924	9.47025	43.7273	16.636703	120	
3	4	04-01-2018	1	0	1	0	2	1	1	8.200000	10.60610	59.0435	10.739832	108	
4	5	05-01-2018	1	0	1	0	3	1	1	9.305237	11.46350	43.6957	12.522300	82	

In [6]:

```python
df.shape
```

Out[6]:

```
(730, 16)
```

In [7]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 730 entries, 0 to 729
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   instant     730 non-null    int64
 1   dteday      730 non-null    object
 2   season      730 non-null    int64
 3   yr          730 non-null    int64
 4   mnth        730 non-null    int64
 5   holiday     730 non-null    int64
 6   weekday     730 non-null    int64
 7   workingday  730 non-null    int64
 8   weathersit  730 non-null    int64
```

```
 9   temp        730 non-null    float64
 10  atemp       730 non-null    float64
 11  hum         730 non-null    float64
 12  windspeed   730 non-null    float64
 13  casual      730 non-null    int64
 14  registered  730 non-null    int64
 15  cnt         730 non-null    int64
dtypes: float64(4), int64(11), object(1)
memory usage: 91.4+ KB
```

In [8]:

```
df.describe()
```

Out[8]:

	instant	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	
count	730.000000	730.000000	730.000000	730.000000	730.000000	730.000000	730.000000	730.000000	730.000000	730.000000	730.0
mean	365.500000	2.498630	0.500000	6.526027	0.028767	2.997260	0.683562	1.394521	20.319259	23.726322	62.7
std	210.877136	1.110184	0.500343	3.450215	0.167266	2.006161	0.465405	0.544807	7.506729	8.150308	14.2
min	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	2.424346	3.953480	0.0
25%	183.250000	2.000000	0.000000	4.000000	0.000000	1.000000	0.000000	1.000000	13.811885	16.889713	52.0
50%	365.500000	3.000000	0.500000	7.000000	0.000000	3.000000	1.000000	1.000000	20.465826	24.368225	62.6
75%	547.750000	3.000000	1.000000	10.000000	0.000000	5.000000	1.000000	2.000000	26.880615	30.445775	72.9
max	730.000000	4.000000	1.000000	12.000000	1.000000	6.000000	1.000000	3.000000	35.328347	42.044800	97.2

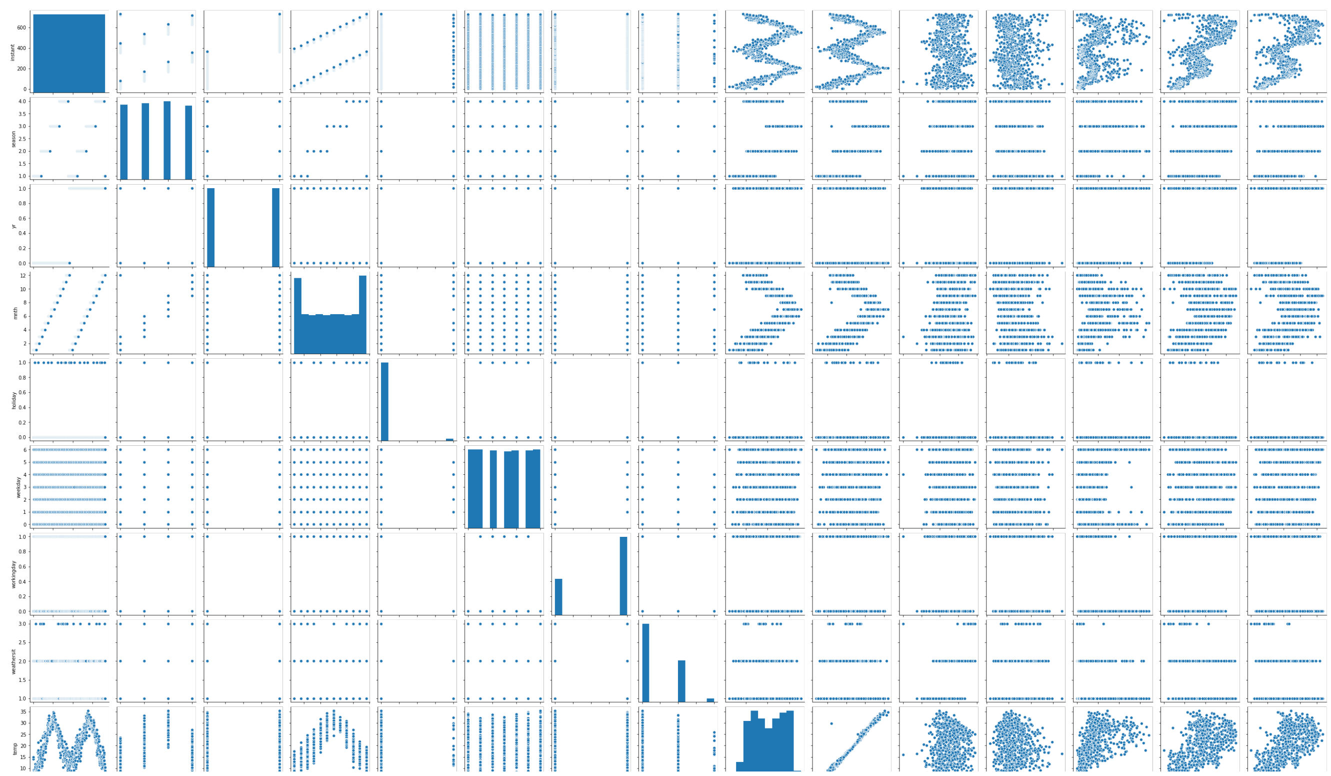
Step 2: Visualising the Data

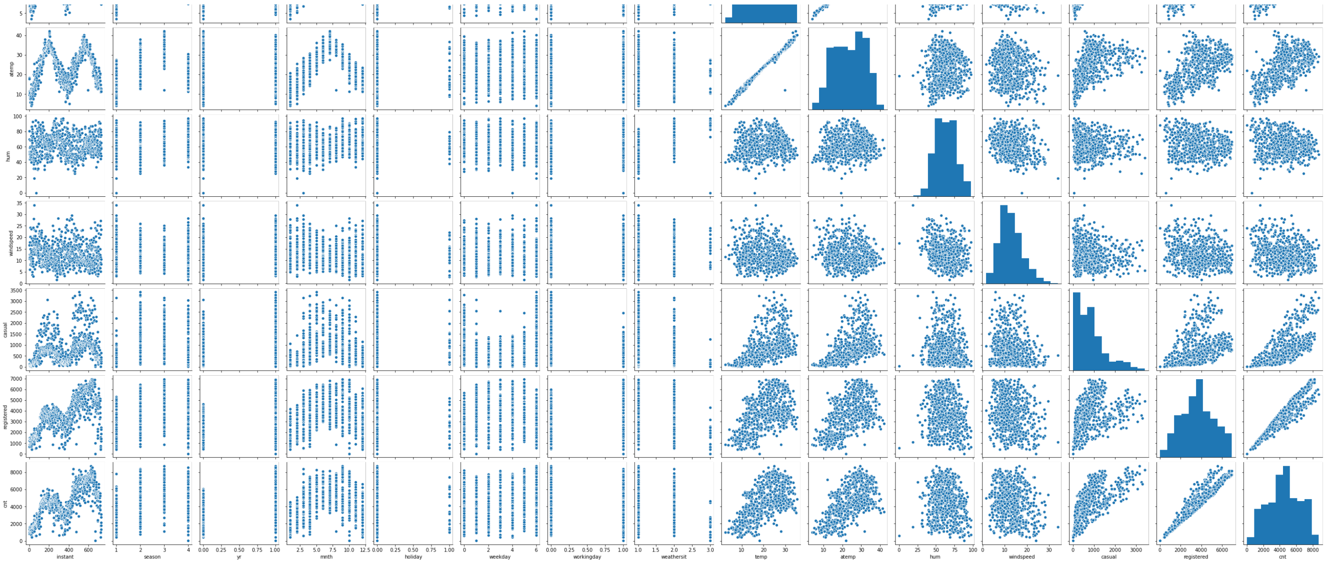
In [9]:

```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [10]:

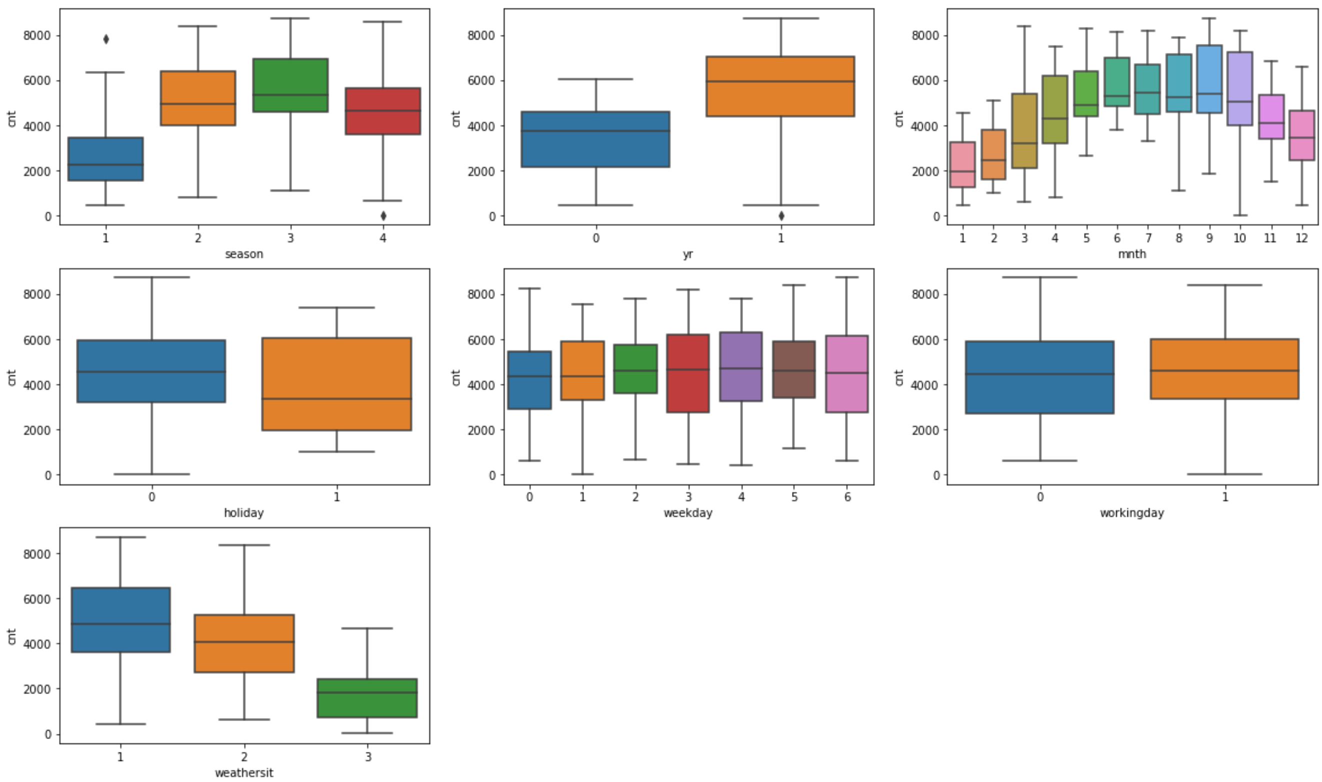
```
sns.pairplot(df)
plt.show()
```





In [11]:

```
plt.figure(figsize=(20, 12))
plt.subplot(3,3,1)
sns.boxplot(x = 'season', y = 'cnt', data = df)
plt.subplot(3,3,2)
sns.boxplot(x = 'yr', y = 'cnt', data = df)
plt.subplot(3,3,3)
sns.boxplot(x = 'mnth', y = 'cnt', data = df)
plt.subplot(3,3,4)
sns.boxplot(x = 'holiday', y = 'cnt', data = df)
plt.subplot(3,3,5)
sns.boxplot(x = 'weekday', y = 'cnt', data = df)
plt.subplot(3,3,6)
sns.boxplot(x = 'workingday', y = 'cnt', data = df)
plt.subplot(3,3,7)
sns.boxplot(x = 'weathersit', y = 'cnt', data = df)
plt.show()
```



Step 3: Data Preparation

In [12]:

```
s1 = pd.get_dummies(df['season'], drop_first = True)
s1
```

Out[12]:

	2	3	4
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
...	...	...	...
725	0	0	0
726	0	0	0
727	0	0	0
728	0	0	0
729	0	0	0

730 rows × 3 columns

In [13]:

```
s1= s1.rename(columns = {2:'summer',3:'fall',4:'winter'})
```

In [14]:

```
s1
```

Out[14]:

	summer	fall	winter
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
...	...	...	...
725	0	0	0
726	0	0	0
727	0	0	0
728	0	0	0
729	0	0	0

730 rows × 3 columns

In [15]:

```
s2 = pd.get_dummies(df['mnth'], drop_first = True)
s2
```

Out[15]:

	2	3	4	5	6	7	8	9	10	11	12
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0

	2	3	4	5	6	7	8	9	10	11	12
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
725	0	0	0	0	0	0	0	0	0	0	1
726	0	0	0	0	0	0	0	0	0	0	1
727	0	0	0	0	0	0	0	0	0	0	1
728	0	0	0	0	0	0	0	0	0	0	1
729	0	0	0	0	0	0	0	0	0	0	1

730 rows × 11 columns

In [16]:

```
s2= s2.rename(columns =
{2:'feb',3:'mar',4:'apr',5:'may',6:'jun',7:'jul',8:'aug',9:'sep',10:'oct',11:'nov',12:'dec'})
s2
```

Out[16]:

	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
725	0	0	0	0	0	0	0	0	0	0	1
726	0	0	0	0	0	0	0	0	0	0	1
727	0	0	0	0	0	0	0	0	0	0	1
728	0	0	0	0	0	0	0	0	0	0	1
729	0	0	0	0	0	0	0	0	0	0	1

730 rows × 11 columns

In [17]:

```
s3 = pd.get_dummies(df['weekday'], drop_first = True)
s3
```

Out[17]:

	1	2	3	4	5	6
0	0	0	0	0	0	1
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	0	0	1	0	0	0
...	...	...	...	...	...	...
725	0	0	0	1	0	0
726	0	0	0	0	1	0
727	0	0	0	0	0	1
728	0	0	0	0	0	0
729	1	0	0	0	0	0

730 rows × 6 columns

In [18]:

```
s3= s3.rename(columns = {1:'mon',2:'tue',3:'wed',4:'thu',5:'fri',6:'sat'})
s3
```

Out[18]:

	mon	tue	wed	thu	fri	sat
0	0	0	0	0	0	1
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	0	0	1	0	0	0
...	...	...	...	...	...	...
725	0	0	0	1	0	0
726	0	0	0	0	1	0
727	0	0	0	0	0	1
728	0	0	0	0	0	0
729	1	0	0	0	0	0

730 rows × 6 columns

In [19]:

```
s4 = pd.get_dummies(df['weathersit'], drop_first = True)
s4
```

Out[19]:

	2	3
0	1	0
1	1	0
2	0	0
3	0	0
4	0	0
...	...	...
725	1	0
726	1	0
727	1	0
728	0	0
729	1	0

730 rows × 2 columns

In [20]:

```
s4= s4.rename(columns = {2:'w2',3:'w3'})
s4
```

Out[20]:

	w2	w3
0	1	0
1	1	0
2	0	0

	w2	w3
3	0	0
4	0	0
...	...	...
725	1	0
726	1	0
727	1	0
728	0	0
729	1	0

730 rows × 2 columns

In [21]:

```
df = pd.concat([df, s1], axis = 1)
df = pd.concat([df, s2], axis = 1)
df = pd.concat([df, s3], axis = 1)
df = pd.concat([df, s4], axis = 1)
```

In [22]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 730 entries, 0 to 729
Data columns (total 38 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   instant     730 non-null    int64
 1   dteday      730 non-null    object
 2   season      730 non-null    int64
 3   yr          730 non-null    int64
 4   mnth        730 non-null    int64
 5   holiday     730 non-null    int64
 6   weekday     730 non-null    int64
 7   workingday  730 non-null    int64
 8   weathersit  730 non-null    int64
 9   temp        730 non-null    float64
 10  atemp       730 non-null    float64
 11  hum         730 non-null    float64
 12  windspeed   730 non-null    float64
 13  casual      730 non-null    int64
 14  registered  730 non-null    int64
 15  cnt         730 non-null    int64
 16  summer      730 non-null    uint8
 17  fall        730 non-null    uint8
 18  winter      730 non-null    uint8
 19  feb         730 non-null    uint8
 20  mar         730 non-null    uint8
 21  apr         730 non-null    uint8
 22  may         730 non-null    uint8
 23  jun         730 non-null    uint8
 24  jul         730 non-null    uint8
 25  aug         730 non-null    uint8
 26  sep         730 non-null    uint8
 27  oct         730 non-null    uint8
 28  nov         730 non-null    uint8
 29  dec         730 non-null    uint8
 30  mon         730 non-null    uint8
 31  tue         730 non-null    uint8
 32  wed         730 non-null    uint8
 33  thu         730 non-null    uint8
 34  fri         730 non-null    uint8
 35  sat         730 non-null    uint8
 36  w2          730 non-null    uint8
 37  w3          730 non-null    uint8
dtypes: float64(4), int64(11), object(1), uint8(22)
memory usage: 107.1+ KB
```

In [23]:

```
df.drop(['season'], axis = 1, inplace = True)
df.drop(['mnth'], axis = 1, inplace = True)
df.drop(['weekday'], axis = 1, inplace = True)
df.drop(['weathersit'], axis = 1, inplace = True)
df.drop(['instant'], axis = 1, inplace = True)
df.drop(['dteday'], axis = 1, inplace = True)
```

In [24]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 730 entries, 0 to 729
Data columns (total 32 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   yr          730 non-null    int64  
 1   holiday     730 non-null    int64  
 2   workingday  730 non-null    int64  
 3   temp        730 non-null    float64
 4   atemp       730 non-null    float64
 5   hum         730 non-null    float64
 6   windspeed   730 non-null    float64
 7   casual      730 non-null    int64  
 8   registered  730 non-null    int64  
 9   cnt         730 non-null    int64  
 10  summer      730 non-null    uint8  
 11  fall        730 non-null    uint8  
 12  winter      730 non-null    uint8  
 13  feb         730 non-null    uint8  
 14  mar         730 non-null    uint8  
 15  apr         730 non-null    uint8  
 16  may         730 non-null    uint8  
 17  jun         730 non-null    uint8  
 18  jul         730 non-null    uint8  
 19  aug         730 non-null    uint8  
 20  sep         730 non-null    uint8  
 21  oct         730 non-null    uint8  
 22  nov         730 non-null    uint8  
 23  dec         730 non-null    uint8  
 24  mon         730 non-null    uint8  
 25  tue         730 non-null    uint8  
 26  wed         730 non-null    uint8  
 27  thu         730 non-null    uint8  
 28  fri         730 non-null    uint8  
 29  sat         730 non-null    uint8  
 30  w2          730 non-null    uint8  
 31  w3          730 non-null    uint8  
dtypes: float64(4), int64(6), uint8(22)
memory usage: 72.8 KB
```

Step 4: Splitting the Data into Training and Testing Sets

In [25]:

```
from sklearn.model_selection import train_test_split

np.random.seed(0)
df_train, df_test = train_test_split(df, train_size = 0.7, test_size = 0.3, random_state = 100)
```

In [26]:

```
from sklearn.preprocessing import MinMaxScaler
```

In [27]:

```
scaler = MinMaxScaler()
```

In [28]:

In [28]:

```
df.head()
```

Out[28]:

	yr	holiday	workingday	temp	atemp	hum	windspeed	casual	registered	cnt	...	nov	dec	mon	tue	wed	thu	fri
0	0	0	0	14.110847	18.18125	80.5833	10.749882	331	654	985	...	0	0	0	0	0	0	0
1	0	0	0	14.902598	17.68695	69.6087	16.652113	131	670	801	...	0	0	0	0	0	0	0
2	0	0	1	8.050924	9.47025	43.7273	16.636703	120	1229	1349	...	0	0	1	0	0	0	0
3	0	0	1	8.200000	10.60610	59.0435	10.739832	108	1454	1562	...	0	0	0	1	0	0	0
4	0	0	1	9.305237	11.46350	43.6957	12.522300	82	1518	1600	...	0	0	0	0	1	0	0

5 rows × 32 columns

In [29]:

```
num_vars = ['temp', 'atemp', 'hum', 'windspeed', 'casual','registered','cnt']

df_train[num_vars] = scaler.fit_transform(df_train[num_vars])
```

In [30]:

```
df_train.head()
```

Out[30]:

	yr	holiday	workingday	temp	atemp	hum	windspeed	casual	registered	cnt	...	nov	dec	mon	tue	wed
653	1	0	1	0.509887	0.501133	0.575354	0.300794	0.280402	0.951776	0.864243	...	0	0	0	1	0
576	1	0	1	0.815169	0.766351	0.725633	0.264686	0.294422	0.899220	0.827658	...	0	0	0	1	0
426	1	0	0	0.442393	0.438975	0.640189	0.255342	0.290765	0.446145	0.465255	...	0	0	0	0	0
728	1	0	0	0.245101	0.200348	0.498067	0.663106	0.110332	0.203869	0.204096	...	0	1	0	0	0
482	1	0	0	0.395666	0.391735	0.504508	0.188475	0.340750	0.444701	0.482973	...	0	0	0	0	0

5 rows × 32 columns

In [31]:

```
df_train.describe()
```

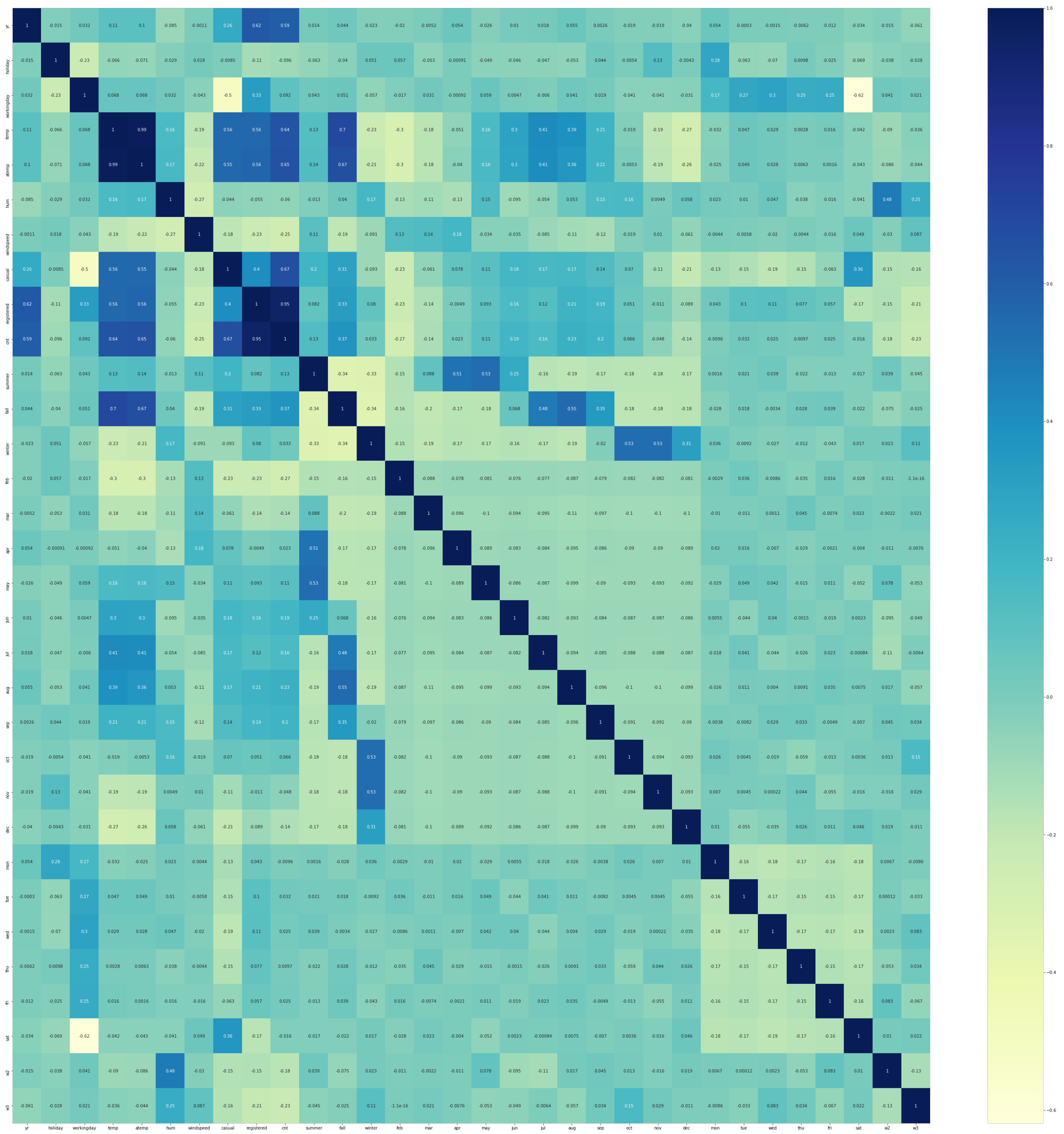
Out[31]:

	yr	holiday	workingday	temp	atemp	hum	windspeed	casual	registered	cnt	...	
count	510.000000	510.000000	510.000000	510.000000	510.000000	510.000000	510.000000	510.000000	510.000000	510.000000	...	5
mean	0.507843	0.025490	0.676471	0.537262	0.512989	0.650369	0.320768	0.254661	0.523944	0.513620	...	
std	0.500429	0.157763	0.468282	0.225844	0.212385	0.145882	0.169797	0.206011	0.228175	0.224593	...	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	
25%	0.000000	0.000000	0.000000	0.339853	0.332086	0.538643	0.199179	0.094179	0.353487	0.356420	...	
50%	1.000000	0.000000	1.000000	0.540519	0.526811	0.653714	0.296763	0.212740	0.525123	0.518638	...	
75%	1.000000	0.000000	1.000000	0.735215	0.688457	0.754830	0.414447	0.327415	0.696073	0.684710	...	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	

8 rows × 32 columns

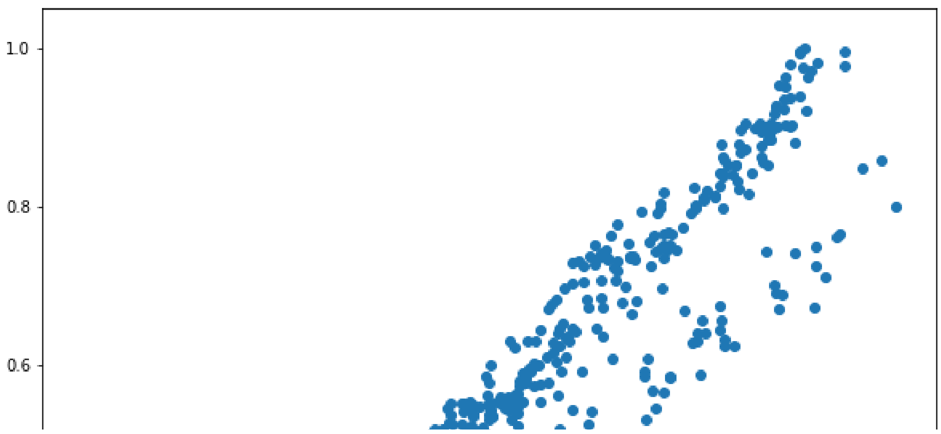
In [32]:

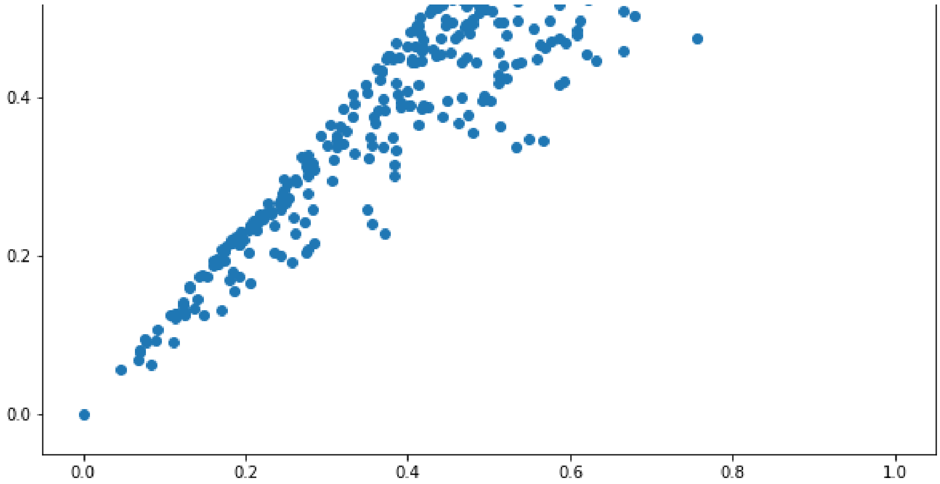
```
plt.figure(figsize = (50, 50))
sns.heatmap(df_train.corr(), annot = True, cmap="YlGnBu")
plt.show()
```



In [33]:

```
plt.figure(figsize=[10,10])
plt.scatter(df_train.cnt, df_train.registered)
plt.show()
```





In [34]:

```
y_train = df_train.pop('cnt')
X_train = df_train
```

Step 5: Building a linear model

In [35]:

```
df.columns
```

Out[35]:

```
Index(['yr', 'holiday', 'workingday', 'temp', 'atemp', 'hum', 'windspeed',
       'casual', 'registered', 'cnt', 'summer', 'fall', 'winter', 'feb', 'mar',
       'apr', 'may', 'jun', 'jul', 'aug', 'sep', 'oct', 'nov', 'dec', 'mon',
       'tue', 'wed', 'thu', 'fri', 'sat', 'w2', 'w3'],
      dtype='object')
```

In [36]:

```
import statsmodels.api as sm
X_train_lm = sm.add_constant(X_train)

lr_1 = sm.OLS(y_train, X_train_lm).fit()

lr_1.params
```

Out[36]:

```
const          4.232725e-16
yr            -2.133710e-16
holiday       -1.249001e-16
workingday     6.175616e-16
temp           2.442491e-15
atemp         -1.776357e-15
hum            6.938894e-17
windspeed     -2.775558e-16
casual         3.774735e-01
registered     7.968247e-01
summer         2.567391e-16
fall          -8.326673e-17
winter         1.249001e-16
feb           -1.942890e-16
mar           -1.769418e-16
apr           -5.377643e-16
may           -1.873501e-16
jun           -5.551115e-17
jul           -6.938894e-17
aug           -1.942890e-16
sep           -2.012279e-16
oct            1.006140e-16
nov            1.804112e-16
```

```
dec         -2.081668e-16
mon          1.908196e-16
tue          1.387779e-17
wed          6.591949e-17
thu          4.076600e-17
fri         -2.133710e-16
sat          2.498002e-16
w2           4.666406e-16
w3           1.526557e-16
dtype: float64
```

In [37]:

```
print(lr_1.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 3.731e+29
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:00   Log-Likelihood:                 16695.
No. Observations:                 510   AIC:                        -3.333e+04
Df Residuals:                     479   BIC:                        -3.320e+04
Df Model:                          30                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       4.233e-16   5.44e-16      0.779      0.437   -6.45e-16    1.49e-15
yr         -2.134e-16   2.42e-16     -0.880      0.379    -6.9e-16    2.63e-16
holiday    -1.249e-16   4.08e-16     -0.306      0.759   -9.26e-16    6.76e-16
workingday  6.176e-16   2.51e-16      2.465      0.014    1.25e-16    1.11e-15
temp        2.442e-15   2.46e-15      0.991      0.322    -2.4e-15    7.28e-15
atemp      -1.776e-15   2.37e-15     -0.749      0.454   -6.43e-15    2.88e-15
hum         6.939e-17    6.7e-16      0.104      0.918   -1.25e-15    1.39e-15
windspeed  -2.776e-16   4.75e-16     -0.585      0.559   -1.21e-15    6.55e-16
casual         0.3775    6.9e-16   5.47e+14      0.000       0.377       0.377
registered     0.7968    8.2e-16   9.71e+14      0.000       0.797       0.797
summer      2.567e-16    4.1e-16      0.627      0.531   -5.48e-16    1.06e-15
fall       -8.327e-17   5.16e-16     -0.161      0.872    -1.1e-15     9.3e-16
winter      1.249e-16    4.7e-16      0.266      0.791   -7.99e-16    1.05e-15
feb        -1.943e-16   3.54e-16     -0.549      0.584    -8.9e-16    5.02e-16
mar        -1.769e-16   3.86e-16     -0.458      0.647   -9.36e-16    5.82e-16
apr        -5.378e-16   5.77e-16     -0.933      0.351   -1.67e-15    5.95e-16
may        -1.874e-16   6.24e-16     -0.300      0.764   -1.41e-15    1.04e-15
jun        -5.551e-17   6.72e-16     -0.083      0.934   -1.38e-15    1.26e-15
jul        -6.939e-17   7.54e-16     -0.092      0.927   -1.55e-15    1.41e-15
aug        -1.943e-16   7.22e-16     -0.269      0.788   -1.61e-15    1.22e-15
sep        -2.012e-16   6.52e-16     -0.309      0.758   -1.48e-15    1.08e-15
oct         1.006e-16   5.94e-16      0.169      0.866   -1.07e-15    1.27e-15
nov         1.804e-16   5.63e-16      0.320      0.749   -9.26e-16    1.29e-15
dec        -2.082e-16   4.54e-16     -0.458      0.647    -1.1e-15    6.84e-16
mon         1.908e-16   1.71e-16      1.119      0.264   -1.44e-16    5.26e-16
tue         1.388e-17   1.99e-16      0.070      0.944   -3.77e-16    4.05e-16
wed         6.592e-17   1.91e-16      0.345      0.730    -3.1e-16    4.41e-16
thu         4.077e-17   1.94e-16      0.210      0.833    -3.4e-16    4.21e-16
fri        -2.134e-16   1.93e-16     -1.106      0.269   -5.93e-16    1.66e-16
sat         2.498e-16   2.52e-16      0.991      0.322   -2.46e-16    7.45e-16
w2          4.666e-16   1.85e-16      2.520      0.012    1.03e-16     8.3e-16
w3          1.527e-16   4.98e-16      0.307      0.759   -8.25e-16    1.13e-15
==============================================================================
Omnibus:                        3.888   Durbin-Watson:                   0.200
Prob(Omnibus):                  0.143   Jarque-Bera (JB):                3.422
Skew:                           0.124   Prob(JB):                        0.181
Kurtosis:                       2.684   Cond. No.                     2.25e+15
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.81e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

In [38]:

```
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

In [39]:

```
vif = pd.DataFrame()
vif['Features'] = X_train.columns
vif['VIF'] = [variance_inflation_factor(X_train.values, i) for i in range(X_train.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[39]:

	Features	VIF
25	wed	inf
2	workingday	inf
27	fri	inf
1	holiday	inf
24	tue	inf
23	mon	inf
26	thu	inf
3	temp	457.25
4	atemp	384.53
8	registered	48.36
5	hum	21.16
10	fall	15.55
11	winter	12.22
18	aug	11.12
7	casual	10.72
17	jul	9.64
9	summer	9.09
19	sep	7.59
16	jun	7.43
15	may	7.28
20	oct	6.77
0	yr	6.61
21	nov	6.03
14	apr	5.72
6	windspeed	4.93
22	dec	3.80
13	mar	3.17
29	w2	2.48
28	sat	2.06
12	feb	1.75
30	w3	1.58

In [40]:

```
X = X_train.drop('wed', 1)
```

In [41]:

```
X_train_lm = sm.add_constant(X)

lr_2 = sm.OLS(y_train, X_train_lm).fit()
```

In [42]:

```
print(lr_2.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 4.543e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:01   Log-Likelihood:                 17333.
No. Observations:                 510   AIC:                        -3.460e+04
Df Residuals:                     479   BIC:                        -3.447e+04
Df Model:                          30                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.678e-16   1.56e-16      2.361      0.019    6.16e-17    6.74e-16
yr         -2.463e-16   6.94e-17     -3.547      0.000   -3.83e-16    -1.1e-16
holiday     2.776e-16   1.51e-16      1.832      0.068   -2.01e-17    5.75e-16
workingday  5.412e-16   1.02e-16      5.298      0.000    3.41e-16    7.42e-16
temp        6.661e-16   7.06e-16      0.943      0.346   -7.21e-16    2.05e-15
atemp       -1.11e-15   6.79e-16     -1.634      0.103   -2.45e-15    2.25e-16
hum        -1.943e-16   1.92e-16     -1.012      0.312   -5.72e-16    1.83e-16
windspeed  -4.718e-16   1.36e-16     -3.469      0.001   -7.39e-16   -2.05e-16
casual         0.3775   1.98e-16   1.91e+15      0.000       0.377       0.377
registered     0.7968   2.35e-16   3.39e+15      0.000       0.797       0.797
summer     -7.633e-17   1.17e-16     -0.650      0.516   -3.07e-16    1.54e-16
fall        -2.22e-16   1.48e-16     -1.502      0.134   -5.13e-16    6.85e-17
winter      6.939e-17   1.35e-16      0.515      0.607   -1.95e-16    3.34e-16
feb         3.469e-17   1.01e-16      0.342      0.733   -1.65e-16    2.34e-16
mar         3.053e-16   1.11e-16      2.759      0.006    8.78e-17    5.23e-16
apr         2.498e-16   1.65e-16      1.512      0.131   -7.49e-17    5.74e-16
may         3.123e-17   1.79e-16      0.175      0.861    -3.2e-16    3.82e-16
jun        -1.665e-16   1.93e-16     -0.865      0.388   -5.45e-16    2.12e-16
jul        -2.567e-16   2.16e-16     -1.188      0.236   -6.81e-16    1.68e-16
aug         2.498e-16   2.07e-16      1.208      0.228   -1.57e-16    6.56e-16
sep        -2.776e-17   1.87e-16     -0.149      0.882   -3.95e-16    3.39e-16
oct        -3.123e-16    1.7e-16     -1.833      0.067   -6.47e-16    2.24e-17
nov        -2.637e-16   1.61e-16     -1.634      0.103   -5.81e-16    5.34e-17
dec         6.939e-17    1.3e-16      0.533      0.594   -1.86e-16    3.25e-16
mon        -2.776e-17   7.22e-17     -0.384      0.701    -1.7e-16    1.14e-16
tue        -1.527e-16   7.24e-17     -2.109      0.035   -2.95e-16   -1.04e-17
thu         2.012e-16    7.2e-17      2.793      0.005    5.97e-17    3.43e-16
fri         -2.22e-16   7.45e-17     -2.982      0.003   -3.68e-16   -7.57e-17
sat         1.527e-16   7.23e-17      2.112      0.035    1.07e-17    2.95e-16
w2          3.469e-17   5.31e-17      0.654      0.514   -6.96e-17    1.39e-16
w3         -1.388e-16   1.43e-16     -0.973      0.331   -4.19e-16    1.41e-16
==============================================================================
Omnibus:                       11.333   Durbin-Watson:                   1.897
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               10.976
Skew:                           0.320   Prob(JB):                      0.00414
Kurtosis:                       2.674   Cond. No.                         97.6
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [43]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[43]:

	Features	VIF
3	temp	457.25

	Features	VIF
4	atemp	384.53
8	registered	48.36
5	hum	21.16
2	workingday	17.88
10	fall	15.55
11	winter	12.22
18	aug	11.12
7	casual	10.72
17	jul	9.64
9	summer	9.09
19	sep	7.59
16	jun	7.43
15	may	7.28
20	oct	6.77
0	yr	6.61
21	nov	6.03
14	apr	5.72
6	windspeed	4.93
22	dec	3.80
13	mar	3.17
28	w2	2.48
23	mon	2.13
27	sat	2.06
26	fri	1.91
25	thu	1.87
24	tue	1.86
12	feb	1.75
29	w3	1.58
1	holiday	1.54

In [44]:

```
X = X.drop('temp', 1)
```

In [45]:

```
X_train_lm = sm.add_constant(X)

lr_3 = sm.OLS(y_train, X_train_lm).fit()
```

In [46]:

```
print(lr_3.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 9.447e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:02   Log-Likelihood:                 17510.
No. Observations:                 510   AIC:                        -3.496e+04
Df Residuals:                     480   BIC:                        -3.483e+04
Df Model:                          29
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
const         1.249e-16    1.1e-16      1.138      0.256    -9.08e-17    3.41e-16
yr           -1.11e-16     4.9e-17     -2.267      0.024    -2.07e-16   -1.48e-17
holiday       1.076e-16    1.06e-16     1.012      0.312    -1.01e-16    3.16e-16
workingday    2.498e-16    7.19e-17     3.474      0.001     1.09e-16    3.91e-16
atemp         4.718e-16    1.7e-16      2.782      0.006     1.39e-16    8.05e-16
hum          -5.135e-16    1.35e-16    -3.802      0.000    -7.79e-16   -2.48e-16
windspeed    -2.498e-16    9.3e-17     -2.685      0.007    -4.33e-16   -6.7e-17
casual           0.3775    1.38e-16    2.73e+15    0.000        0.377       0.377
registered       0.7968    1.66e-16    4.81e+15    0.000        0.797       0.797
summer        1.457e-16    8.28e-17     1.761      0.079    -1.69e-17    3.08e-16
fall          5.829e-16    1.04e-16     5.603      0.000     3.78e-16    7.87e-16
winter        6.939e-18    9.5e-17      0.073      0.942    -1.8e-16     1.94e-16
feb           3.469e-17    7.15e-17     0.485      0.628    -1.06e-16    1.75e-16
mar          -1.874e-16    7.75e-17    -2.417      0.016    -3.4e-16    -3.51e-17
apr          -2.22e-16     1.16e-16    -1.915      0.056    -4.5e-16     5.79e-18
may          -1.908e-16    1.23e-16    -1.546      0.123    -4.33e-16    5.18e-17
jun          -3.331e-16    1.31e-16    -2.544      0.011    -5.9e-16    -7.59e-17
jul          -4.163e-16    1.48e-16    -2.811      0.005    -7.07e-16   -1.25e-16
aug          -3.192e-16    1.41e-16    -2.263      0.024    -5.96e-16   -4.21e-17
sep          -2.637e-16    1.3e-16     -2.031      0.043    -5.19e-16   -8.64e-18
oct          -1.943e-16    1.2e-16     -1.624      0.105    -4.29e-16    4.07e-17
nov          -1.804e-16    1.14e-16    -1.587      0.113    -4.04e-16    4.3e-17
dec          -1.422e-16    9.17e-17    -1.550      0.122    -3.23e-16    3.8e-17
mon           3.469e-17    5.08e-17     0.682      0.495    -6.52e-17    1.35e-16
tue          -1.388e-16    5.11e-17    -2.718      0.007    -2.39e-16   -3.85e-17
thu          -9.021e-17    5.08e-17    -1.776      0.076    -1.9e-16     9.6e-18
fri                   0    5.25e-17         0      1.000    -1.03e-16    1.03e-16
sat          -1.388e-17    5.1e-17     -0.272      0.785    -1.14e-16    8.63e-17
w2            9.368e-17    3.74e-17     2.503      0.013     2.01e-17    1.67e-16
w3            4.163e-17    1e-16        0.415      0.678    -1.56e-16    2.39e-16
==============================================================================
Omnibus:                       11.050   Durbin-Watson:                   2.028
Prob(Omnibus):                  0.004   Jarque-Bera (JB):               10.730
Skew:                          -0.317   Prob(JB):                      0.00468
Kurtosis:                       2.680   Cond. No.                         46.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [47]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[47]:

	Features	VIF
7	registered	48.29
3	atemp	48.24
4	hum	21.07
2	workingday	17.83
9	fall	15.47
10	winter	12.22
6	casual	10.56
17	aug	10.40
16	jul	9.11
8	summer	9.08
18	sep	7.37
14	may	6.99
15	jun	6.91
19	oct	6.71

	Features	VIF
0	yr	6.61
20	nov	6.02
13	apr	5.67
5	windspeed	4.60
21	dec	3.80
12	mar	3.13
27	w2	2.48
22	mon	2.12
26	sat	2.06
25	fri	1.91
24	thu	1.87
23	tue	1.86
11	feb	1.74
28	w3	1.57
1	holiday	1.52

In [48]:

```
X = X.drop('winter', 1)
```

In [49]:

```
X_train_lm = sm.add_constant(X)

lr_4 = sm.OLS(y_train, X_train_lm).fit()
```

In [50]:

```
print(lr_4.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.668e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:02   Log-Likelihood:                 17448.
No. Observations:                 510   AIC:                        -3.484e+04
Df Residuals:                     481   BIC:                        -3.471e+04
Df Model:                          28
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       2.914e-16   1.24e-16      2.351      0.019    4.79e-17    5.35e-16
yr          1.587e-16   5.36e-17      2.959      0.003    5.33e-17    2.64e-16
holiday    -6.384e-16   1.19e-16     -5.345      0.000   -8.73e-16   -4.04e-16
workingday   3.07e-16      8e-17      3.839      0.000     1.5e-16    4.64e-16
atemp       -2.22e-16   1.91e-16     -1.160      0.247   -5.98e-16    1.54e-16
hum        -5.551e-17   1.52e-16     -0.364      0.716   -3.55e-16    2.44e-16
windspeed   3.261e-16   1.05e-16      3.107      0.002     1.2e-16    5.32e-16
casual         0.3775   1.56e-16   2.42e+15      0.000       0.377       0.377
registered     0.7968   1.77e-16    4.5e+15      0.000       0.797       0.797
summer     -1.457e-16   8.92e-17     -1.633      0.103   -3.21e-16    2.96e-17
fall       -8.327e-17   1.01e-16     -0.824      0.410   -2.82e-16    1.15e-16
feb        -2.359e-16   8.08e-17     -2.921      0.004   -3.95e-16   -7.72e-17
mar        -2.776e-16   8.74e-17     -3.176      0.002   -4.49e-16   -1.06e-16
apr         4.857e-17   1.29e-16      0.375      0.707   -2.06e-16    3.03e-16
may         1.775e-16   1.39e-16      1.280      0.201   -9.49e-17     4.5e-16
jun         1.665e-16   1.45e-16      1.145      0.253   -1.19e-16    4.52e-16
jul        -1.388e-16    1.6e-16     -0.866      0.387   -4.54e-16    1.76e-16
aug         8.327e-17   1.53e-16      0.546      0.586   -2.17e-16    3.83e-16
sep         1.249e-16   1.37e-16      0.913      0.361   -1.44e-16    3.94e-16
oct         2.012e-16   9.91e-17      2.032      0.043     6.6e-18    3.96e-16
nov        -5.551e-17   8.64e-17     -0.643      0.521   -2.25e-16    1.14e-16
dec         1.388e-17   8.16e-17      0.170      0.865   -1.46e-16    1.74e-16
```

```
mon        -2.082e-17   5.72e-17     -0.364      0.716    -1.33e-16    9.16e-17
tue        -1.388e-17   5.76e-17     -0.241      0.810    -1.27e-16    9.93e-17
thu        -7.286e-17   5.74e-17     -1.270      0.205    -1.86e-16    3.99e-17
fri        -2.082e-17   5.92e-17     -0.351      0.725    -1.37e-16    9.56e-17
sat         1.388e-17   5.75e-17      0.241      0.809    -9.91e-17    1.27e-16
w2         -9.368e-17   4.21e-17     -2.226      0.026    -1.76e-16    -1.1e-17
w3          6.939e-17   1.12e-16      0.619      0.536    -1.51e-16     2.9e-16
==============================================================================
Omnibus:                       67.492   Durbin-Watson:                   0.793
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              137.182
Skew:                           0.749   Prob(JB):                     1.63e-30
Kurtosis:                       5.053   Cond. No.                         43.1
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [51]:

```python
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[51]:

	Features	VIF
3	atemp	48.16
7	registered	43.18
4	hum	21.05
2	workingday	17.21
9	fall	11.45
6	casual	10.49
16	aug	9.54
15	jul	8.35
8	summer	8.28
13	may	6.91
14	jun	6.69
17	sep	6.41
0	yr	6.22
12	apr	5.54
5	windspeed	4.58
18	oct	3.61
11	mar	3.12
19	nov	2.71
26	w2	2.46
20	dec	2.34
21	mon	2.11
25	sat	2.05
24	fri	1.91
23	thu	1.87
22	tue	1.86
10	feb	1.74
27	w3	1.53
1	holiday	1.51

In [52]:

```
X = X.drop('atemp', 1)
```

In [53]:

```
X_train_lm = sm.add_constant(X)

lr_5 = sm.OLS(y_train, X_train_lm).fit()
```

In [54]:

```
print(lr_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.130e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:03   Log-Likelihood:                 16949.
No. Observations:                 510   AIC:                        -3.384e+04
Df Residuals:                     482   BIC:                        -3.372e+04
Df Model:                          27
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -7.633e-17   3.29e-16     -0.232      0.817   -7.22e-16     5.7e-16
yr         -9.346e-17    1.4e-16     -0.667      0.505   -3.69e-16    1.82e-16
holiday     3.018e-16   3.16e-16      0.956      0.339   -3.18e-16    9.22e-16
workingday -1.388e-16   2.12e-16     -0.655      0.513   -5.55e-16    2.78e-16
hum        -8.743e-16   3.85e-16     -2.273      0.023   -1.63e-15   -1.18e-16
windspeed  -1.665e-16   2.78e-16     -0.600      0.549   -7.12e-16    3.79e-16
casual         0.3775   4.01e-16   9.41e+14      0.000       0.377       0.377
registered     0.7968   4.56e-16   1.75e+15      0.000       0.797       0.797
summer     -9.021e-17   2.36e-16     -0.382      0.703   -5.54e-16    3.74e-16
fall       -2.082e-16   2.67e-16     -0.781      0.435   -7.32e-16    3.16e-16
feb         5.551e-17   2.12e-16      0.261      0.794   -3.62e-16    4.73e-16
mar        -1.006e-16   2.21e-16     -0.455      0.649   -5.35e-16    3.33e-16
apr         -4.25e-16   3.26e-16     -1.304      0.193   -1.07e-15    2.16e-16
may        -2.914e-16   3.35e-16     -0.870      0.385   -9.49e-16    3.67e-16
jun        -1.631e-16   3.33e-16     -0.490      0.625   -8.17e-16    4.91e-16
jul         2.706e-16   3.69e-16      0.733      0.464   -4.55e-16    9.96e-16
aug         1.388e-17   3.68e-16      0.038      0.970   -7.09e-16    7.36e-16
sep         1.422e-16   3.41e-16      0.417      0.677   -5.28e-16    8.13e-16
oct          1.11e-16   2.49e-16      0.446      0.656   -3.78e-16       6e-16
nov         3.886e-16   2.25e-16      1.727      0.085   -5.34e-17    8.31e-16
dec         2.246e-16   2.14e-16      1.048      0.295   -1.97e-16    6.46e-16
mon         4.857e-17   1.52e-16      0.320      0.749    -2.5e-16    3.47e-16
tue                 0   1.53e-16          0      1.000      -3e-16       3e-16
thu         9.714e-17   1.52e-16      0.638      0.523   -2.02e-16    3.96e-16
fri         4.163e-17   1.57e-16      0.265      0.791   -2.67e-16     3.5e-16
sat         2.637e-16   1.52e-16      1.730      0.084   -3.57e-17    5.63e-16
w2         -9.888e-17   1.11e-16     -0.888      0.375   -3.18e-16     1.2e-16
w3          1.943e-16   2.97e-16      0.654      0.514    -3.9e-16    7.78e-16
==============================================================================
Omnibus:                        5.455   Durbin-Watson:                   0.318
Prob(Omnibus):                  0.065   Jarque-Bera (JB):                5.333
Skew:                          -0.248   Prob(JB):                       0.0695
Kurtosis:                       3.071   Cond. No.                         36.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [55]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[55]:

	Features	VIF
6	registered	40.74
3	hum	17.23
2	workingday	17.17
8	fall	11.33
5	casual	9.86
7	summer	8.24
15	aug	7.87
14	jul	6.30
0	yr	6.03
12	may	5.73
16	sep	5.67
11	apr	5.00
13	jun	4.98
4	windspeed	4.55
17	oct	3.24
10	mar	2.83
18	nov	2.62
25	w2	2.45
19	dec	2.30
20	mon	2.11
24	sat	2.04
23	fri	1.90
22	thu	1.87
21	tue	1.86
9	feb	1.71
26	w3	1.53
1	holiday	1.49

In [56]:

```
X = X.drop('workingday', 1)
```

In [57]:

```
X_train_lm = sm.add_constant(X)

lr_6 = sm.OLS(y_train, X_train_lm).fit()
```

In [58]:

```
print(lr_6.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 2.563e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:03   Log-Likelihood:                 17148.
No. Observations:                 510   AIC:                        -3.424e+04
Df Residuals:                     483   BIC:                        -3.413e+04
Df Model:                          26
Covariance Type:            nonrobust
==============================================================================
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.886e-16   2.15e-16      1.809      0.071   -3.36e-17    8.11e-16
yr          3.548e-16      9e-17      3.942      0.000    1.78e-16    5.32e-16
holiday     -2.29e-16   1.92e-16     -1.192      0.234   -6.07e-16    1.49e-16
hum        -4.441e-16    2.6e-16     -1.706      0.089   -9.56e-16    6.74e-17
windspeed   6.106e-16   1.88e-16      3.252      0.001    2.42e-16     9.8e-16
casual         0.3775   2.17e-16   1.74e+15      0.000       0.377       0.377
registered     0.7968   2.56e-16   3.12e+15      0.000       0.797       0.797
summer     -4.094e-16    1.6e-16     -2.566      0.011   -7.23e-16   -9.59e-17
fall       -3.053e-16    1.8e-16     -1.695      0.091   -6.59e-16    4.86e-17
feb         -2.22e-16   1.44e-16     -1.547      0.122   -5.04e-16    5.99e-17
mar        -1.457e-16   1.49e-16     -0.975      0.330   -4.39e-16    1.48e-16
apr         7.633e-17   2.21e-16      0.346      0.729   -3.57e-16     5.1e-16
may         5.551e-17   2.26e-16      0.245      0.806   -3.89e-16       5e-16
jun         1.596e-16   2.25e-16      0.710      0.478   -2.82e-16    6.01e-16
jul          5.69e-16   2.49e-16      2.286      0.023    7.98e-17    1.06e-15
aug         4.372e-16   2.48e-16      1.762      0.079   -5.04e-17    9.25e-16
sep         4.163e-16    2.3e-16      1.812      0.071   -3.52e-17    8.68e-16
oct        -1.284e-16   1.67e-16     -0.766      0.444   -4.57e-16    2.01e-16
nov         1.804e-16    1.5e-16      1.200      0.231   -1.15e-16    4.76e-16
dec         1.576e-16   1.43e-16      1.105      0.270   -1.23e-16    4.38e-16
mon         2.637e-16   9.42e-17      2.800      0.005    7.86e-17    4.49e-16
tue         2.429e-16   9.67e-17      2.511      0.012    5.29e-17    4.33e-16
thu         1.214e-16    9.6e-17      1.265      0.207   -6.72e-17     3.1e-16
fri         4.996e-16   9.54e-17      5.238      0.000    3.12e-16    6.87e-16
sat        -2.498e-16   9.46e-17     -2.640      0.009   -4.36e-16   -6.38e-17
w2          1.769e-16   7.51e-17      2.356      0.019    2.94e-17    3.25e-16
w3           1.11e-16   1.98e-16      0.560      0.576   -2.78e-16    5.01e-16
==============================================================================
Omnibus:                        1.811   Durbin-Watson:                   0.640
Prob(Omnibus):                  0.404   Jarque-Bera (JB):                1.876
Skew:                          -0.141   Prob(JB):                        0.391
Kurtosis:                       2.904   Cond. No.                         33.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [59]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[59]:

	Features	VIF
5	registered	25.75
2	hum	15.74
7	fall	11.28
6	summer	8.23
14	aug	7.83
4	casual	6.62
13	jul	6.27
11	may	5.71
15	sep	5.60
0	yr	5.35
10	apr	5.00
12	jun	4.97
3	windspeed	4.46
16	oct	3.20
9	mar	2.82

	Features	VIF
17	nov	2.57
24	w2	2.45
18	dec	2.25
23	sat	1.79
19	mon	1.76
8	feb	1.71
20	tue	1.61
21	thu	1.60
22	fri	1.51
25	w3	1.50
1	holiday	1.24

In [60]:

```
X = X.drop('hum', 1)
```

In [61]:

```
X_train_lm = sm.add_constant(X)

lr_7 = sm.OLS(y_train, X_train_lm).fit()
```

In [62]:

```
print(lr_7.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.096e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:04   Log-Likelihood:                 16921.
No. Observations:                 510   AIC:                        -3.379e+04
Df Residuals:                     484   BIC:                        -3.368e+04
Df Model:                          25
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       5.551e-16   2.11e-16      2.635      0.009    1.41e-16    9.69e-16
yr          1.631e-16    1.4e-16      1.162      0.246   -1.13e-16    4.39e-16
holiday    -1.908e-16      3e-16     -0.637      0.525    -7.8e-16    3.98e-16
windspeed  -1.943e-16   2.82e-16     -0.690      0.490   -7.47e-16    3.59e-16
casual         0.3775   3.39e-16   1.11e+15      0.000       0.377       0.377
registered     0.7968   3.98e-16      2e+15      0.000       0.797       0.797
summer      1.388e-17   2.49e-16      0.056      0.955   -4.74e-16    5.02e-16
fall        5.967e-16   2.81e-16      2.124      0.034    4.48e-17    1.15e-15
feb         4.233e-16   2.24e-16      1.891      0.059   -1.65e-17    8.63e-16
mar         2.186e-16   2.33e-16      0.938      0.349   -2.39e-16    6.77e-16
apr        -6.939e-18   3.44e-16     -0.020      0.984   -6.83e-16    6.69e-16
may        -1.874e-16   3.51e-16     -0.533      0.594   -8.78e-16    5.03e-16
jun        -2.776e-17    3.5e-16     -0.079      0.937   -7.16e-16    6.61e-16
jul        -3.469e-16   3.87e-16     -0.895      0.371   -1.11e-15    4.14e-16
aug         -1.18e-15   3.85e-16     -3.062      0.002   -1.94e-15   -4.23e-16
sep        -6.939e-17   3.55e-16     -0.195      0.845   -7.67e-16    6.29e-16
oct        -4.857e-16   2.58e-16     -1.886      0.060   -9.92e-16    2.04e-17
nov        -4.857e-17   2.33e-16     -0.208      0.835   -5.07e-16     4.1e-16
dec          2.22e-16   2.21e-16      1.006      0.315   -2.11e-16    6.56e-16
mon         2.012e-16   1.47e-16      1.370      0.171   -8.74e-17     4.9e-16
tue         9.021e-17   1.51e-16      0.598      0.550   -2.06e-16    3.87e-16
thu          2.22e-16    1.5e-16      1.484      0.139    -7.2e-17    5.16e-16
fri         2.914e-16   1.49e-16      1.962      0.050   -3.96e-19    5.83e-16
sat         2.776e-17   1.47e-16      0.189      0.851   -2.62e-16    3.17e-16
w2         -5.898e-17   1.01e-16     -0.586      0.558   -2.57e-16    1.39e-16
w3           1.11e-16   2.92e-16      0.380      0.704   -4.63e-16    6.85e-16
==============================================================================
Omnibus:                        8.818   Durbin-Watson:                   0.301
```

```
Prob(Omnibus):                  0.012   Jarque-Bera (JB):                5.434
Skew:                           0.054   Prob(JB):                       0.0661
Kurtosis:                       2.506   Cond. No.                         30.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [63]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[63]:

	Features	VIF
4	registered	23.74
6	fall	11.28
5	summer	8.23
13	aug	7.71
3	casual	6.51
12	jul	6.17
10	may	5.56
14	sep	5.49
0	yr	5.20
9	apr	4.95
11	jun	4.90
2	windspeed	4.19
15	oct	3.04
8	mar	2.69
16	nov	2.41
17	dec	1.99
22	sat	1.77
18	mon	1.72
23	w2	1.67
19	tue	1.60
20	thu	1.59
7	feb	1.58
21	fri	1.50
24	w3	1.32
1	holiday	1.24

In [64]:

```
X = X.drop('summer', 1)
```

In [65]:

```
X_train_lm = sm.add_constant(X)

lr_8 = sm.OLS(y_train, X_train_lm).fit()
```

In [66]:

```
print(lr_8.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 8.462e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:04   Log-Likelihood:                 17431.
No. Observations:                 510   AIC:                        -3.481e+04
Df Residuals:                     485   BIC:                        -3.471e+04
Df Model:                          24
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       7.147e-16   7.73e-17      9.245      0.000    5.63e-16    8.67e-16
yr          1.197e-16   5.15e-17      2.326      0.020    1.86e-17    2.21e-16
holiday     1.249e-16    1.1e-16      1.135      0.257   -9.13e-17    3.41e-16
windspeed  -5.135e-16   1.03e-16     -4.966      0.000   -7.17e-16    -3.1e-16
casual         0.3775   1.24e-16   3.05e+15      0.000       0.377       0.377
registered     0.7968   1.46e-16   5.45e+15      0.000       0.797       0.797
fall        5.551e-17   8.99e-17      0.618      0.537   -1.21e-16    2.32e-16
feb        -4.163e-17   8.22e-17     -0.507      0.613   -2.03e-16     1.2e-16
mar        -4.857e-17      8e-17     -0.607      0.544   -2.06e-16    1.09e-16
apr        -2.776e-17   9.09e-17     -0.305      0.760   -2.06e-16    1.51e-16
may         2.828e-16   9.56e-17      2.957      0.003    9.49e-17    4.71e-16
jun         -1.11e-16   1.07e-16     -1.041      0.299   -3.21e-16    9.86e-17
jul        -1.318e-16   1.34e-16     -0.980      0.327   -3.96e-16    1.32e-16
aug        -1.249e-16   1.34e-16     -0.933      0.351   -3.88e-16    1.38e-16
sep         1.388e-16   1.26e-16      1.102      0.271   -1.09e-16    3.86e-16
oct         2.498e-16   9.45e-17      2.644      0.008    6.42e-17    4.35e-16
nov         6.939e-17   8.57e-17      0.810      0.418   -9.89e-17    2.38e-16
dec        -4.042e-16    8.1e-17     -4.989      0.000   -5.63e-16   -2.45e-16
mon        -1.527e-16   5.39e-17     -2.832      0.005   -2.59e-16   -4.67e-17
tue         2.776e-17   5.54e-17      0.501      0.616   -8.11e-17    1.37e-16
thu         2.706e-16   5.49e-17      4.927      0.000    1.63e-16    3.79e-16
fri        -3.747e-16   5.45e-17     -6.870      0.000   -4.82e-16   -2.68e-16
sat         1.665e-16   5.41e-17      3.081      0.002    6.03e-17    2.73e-16
w2         -1.271e-16   3.68e-17     -3.450      0.001   -1.99e-16   -5.47e-17
w3         -9.714e-17   1.07e-16     -0.907      0.365   -3.08e-16    1.13e-16
==============================================================================
Omnibus:                        4.051   Durbin-Watson:                   1.584
Prob(Omnibus):                  0.132   Jarque-Bera (JB):                3.827
Skew:                           0.195   Prob(JB):                        0.148
Kurtosis:                       3.167   Cond. No.                         27.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [67]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[67]:

	Features	VIF
4	registered	23.73
5	fall	8.55
12	aug	6.90
3	casual	6.46
11	jul	5.52
0	yr	5.19
13	sep	5.12

	Features	VIF
2	windspeed	4.1[illegible]
10	jun	3.37
9	may	3.04
14	oct	3.03
8	apr	2.55
15	nov	2.41
7	mar	2.35
16	dec	1.99
21	sat	1.77
17	mon	1.72
22	w2	1.66
18	tue	1.60
19	thu	1.59
6	feb	1.58
20	fri	1.50
23	w3	1.31
1	holiday	1.24

In [68]:

```
X = X.drop('fall', 1)
```

In [69]:

```
X_train_lm = sm.add_constant(X)

lr_9 = sm.OLS(y_train, X_train_lm).fit()
```

In [70]:

```
print(lr_9.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 2.210e+30
Date:                Thu, 03 Sep 2020   Prob (F-statistic):               0.00
Time:                        17:05:05   Log-Likelihood:                 17078.
No. Observations:                 510   AIC:                        -3.411e+04
Df Residuals:                     486   BIC:                        -3.401e+04
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -3.053e-16   1.54e-16     -1.977      0.049   -6.09e-16   -1.87e-18
yr          4.582e-16   1.03e-16      4.457      0.000    2.56e-16     6.6e-16
holiday     3.816e-17    2.2e-16      0.173      0.862   -3.94e-16     4.7e-16
windspeed  -1.874e-16   2.07e-16     -0.907      0.365   -5.93e-16    2.19e-16
casual         0.3775   2.48e-16   1.52e+15      0.000       0.377       0.377
registered     0.7968   2.91e-16   2.73e+15      0.000       0.797       0.797
feb        -4.857e-17   1.64e-16     -0.296      0.768   -3.71e-16    2.74e-16
mar        -7.286e-17    1.6e-16     -0.456      0.649   -3.87e-16    2.41e-16
apr         1.735e-16   1.82e-16      0.955      0.340   -1.83e-16     5.3e-16
may        -6.939e-17   1.91e-16     -0.363      0.717   -4.45e-16    3.06e-16
jun         3.123e-17      2e-16      0.156      0.876   -3.63e-16    4.25e-16
jul        -4.996e-16   1.95e-16     -2.565      0.011   -8.82e-16   -1.17e-16
aug        -2.984e-16   1.92e-16     -1.557      0.120   -6.75e-16    7.81e-17
sep        -3.331e-16   2.04e-16     -1.633      0.103   -7.34e-16    6.77e-17
oct        -1.943e-16   1.89e-16     -1.029      0.304   -5.65e-16    1.77e-16
nov         2.082e-17   1.71e-16      0.122      0.903   -3.16e-16    3.57e-16
dec         6.245e-17   1.62e-16      0.386      0.700   -2.56e-16    3.81e-16
mon         1.388e-16   1.08e-16      1.288      0.198   -7.29e-17     3.5e-16
tue         1.388e-17   1.11e-16      0.125      0.900   -2.04e-16    2.31e-16
```

```
tue          1.388e-17   1.11e-16      0.125      0.900   -2.04e-16    2.31e-16
thu          2.047e-16    1.1e-16      1.867      0.062   -1.07e-17     4.2e-16
fri          1.388e-17   1.09e-16      0.127      0.899      -2e-16    2.28e-16
sat          3.469e-16   1.08e-16      3.213      0.001    1.35e-16    5.59e-16
w2           2.724e-16   7.35e-17      3.705      0.000    1.28e-16    4.17e-16
w3           2.776e-16   2.14e-16      1.297      0.195   -1.43e-16    6.98e-16
==============================================================================
Omnibus:                       40.674   Durbin-Watson:                   0.789
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               14.352
Skew:                          -0.090   Prob(JB):                     0.000765
Kurtosis:                       2.198   Cond. No.                         24.7
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [71]:

```python
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[71]:

	Features	VIF
4	registered	23.66
3	casual	6.46
0	yr	5.18
2	windspeed	4.19
11	aug	3.49
12	sep	3.33
8	may	3.04
13	oct	3.03
9	jun	2.96
10	jul	2.84
7	apr	2.55
14	nov	2.41
6	mar	2.34
15	dec	1.99
20	sat	1.77
16	mon	1.72
21	w2	1.66
17	tue	1.59
5	feb	1.58
18	thu	1.58
19	fri	1.50
22	w3	1.31
1	holiday	1.23

In [72]:

```python
X = X.drop('casual', 1)
```

In [73]:

```python
X_train_lm = sm.add_constant(X)

lr_10 = sm.OLS(y_train, X_train_lm).fit()
```

In [74]:

```
print(lr_10.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.954
Model:                            OLS   Adj. R-squared:                  0.952
Method:                 Least Squares   F-statistic:                     461.8
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          9.47e-310
Time:                        17:05:05   Log-Likelihood:                 825.14
No. Observations:                 510   AIC:                            -1604.
Df Residuals:                     487   BIC:                            -1507.
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0381      0.011      3.620      0.000       0.017       0.059
yr             0.0351      0.007      5.064      0.000       0.021       0.049
holiday        0.0481      0.015      3.201      0.001       0.019       0.078
windspeed     -0.0616      0.014     -4.399      0.000      -0.089      -0.034
registered     0.8028      0.020     39.873      0.000       0.763       0.842
feb            0.0163      0.011      1.441      0.150      -0.006       0.039
mar            0.0679      0.011      6.394      0.000       0.047       0.089
apr            0.0999      0.012      8.542      0.000       0.077       0.123
may            0.1103      0.012      9.030      0.000       0.086       0.134
jun            0.1146      0.013      8.932      0.000       0.089       0.140
jul            0.1131      0.012      9.094      0.000       0.089       0.138
aug            0.1096      0.012      8.934      0.000       0.086       0.134
sep            0.1100      0.013      8.350      0.000       0.084       0.136
oct            0.0992      0.012      8.101      0.000       0.075       0.123
nov            0.0486      0.012      4.183      0.000       0.026       0.071
dec            0.0197      0.011      1.765      0.078      -0.002       0.042
mon           -0.0388      0.007     -5.358      0.000      -0.053      -0.025
tue           -0.0440      0.007     -5.965      0.000      -0.059      -0.030
thu           -0.0381      0.007     -5.171      0.000      -0.053      -0.024
fri           -0.0229      0.007     -3.066      0.002      -0.038      -0.008
sat            0.0632      0.007      9.174      0.000       0.050       0.077
w2            -0.0236      0.005     -4.761      0.000      -0.033      -0.014
w3            -0.0777      0.014     -5.414      0.000      -0.106      -0.050
==============================================================================
Omnibus:                       84.122   Durbin-Watson:                   1.901
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              163.743
Skew:                           0.930   Prob(JB):                     2.78e-36
Kurtosis:                       5.060   Cond. No.                         22.6
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [75]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[75]:

	Features	VIF
3	registered	23.49
0	yr	4.98
2	windspeed	4.11
10	aug	2.91
11	sep	2.85
12	oct	2.59

	Features	VIF
7	may	2.52
8	jun	2.47
9	jul	2.33
13	nov	2.28
6	apr	2.14
5	mar	2.07
14	dec	1.95
15	mon	1.64
20	w2	1.61
4	feb	1.55
17	thu	1.51
16	tue	1.50
18	fri	1.48
19	sat	1.45
21	w3	1.25
1	holiday	1.20

In [76]:

```
X = X.drop('feb', 1)
```

In [77]:

```
X_train_lm = sm.add_constant(X)

lr_11 = sm.OLS(y_train, X_train_lm).fit()
```

In [78]:

```
print(lr_11.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.954
Model:                            OLS   Adj. R-squared:                  0.952
Method:                 Least Squares   F-statistic:                     482.7
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          1.20e-310
Time:                        17:05:06   Log-Likelihood:                 824.06
No. Observations:                 510   AIC:                            -1604.
Df Residuals:                     488   BIC:                            -1511.
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0439      0.010      4.502      0.000       0.025       0.063
yr             0.0342      0.007      4.950      0.000       0.021       0.048
holiday        0.0491      0.015      3.267      0.001       0.020       0.079
windspeed     -0.0592      0.014     -4.254      0.000      -0.087      -0.032
registered     0.8060      0.020     40.240      0.000       0.767       0.845
mar            0.0604      0.009      6.512      0.000       0.042       0.079
apr            0.0921      0.010      8.875      0.000       0.072       0.113
may            0.1025      0.011      9.356      0.000       0.081       0.124
jun            0.1067      0.012      9.191      0.000       0.084       0.130
jul            0.1053      0.011      9.397      0.000       0.083       0.127
aug            0.1018      0.011      9.242      0.000       0.080       0.123
sep            0.1021      0.012      8.524      0.000       0.079       0.126
oct            0.0913      0.011      8.329      0.000       0.070       0.113
nov            0.0409      0.010      3.961      0.000       0.021       0.061
dec            0.0123      0.010      1.243      0.215      -0.007       0.032
mon           -0.0392      0.007     -5.417      0.000      -0.053      -0.025
tue           -0.0439      0.007     -5.944      0.000      -0.058      -0.029
thu           -0.0387      0.007     -5.250      0.000      -0.053      -0.024
fri           -0.0230      0.007     -3.087      0.002      -0.038      -0.008
```

```
sat            0.0630      0.007      9.135      0.000       0.049       0.077
w2            -0.0237      0.005     -4.767      0.000      -0.033      -0.014
w3            -0.0764      0.014     -5.328      0.000      -0.105      -0.048
==============================================================================
Omnibus:                       84.004   Durbin-Watson:                   1.911
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              163.457
Skew:                           0.929   Prob(JB):                     3.20e-36
Kurtosis:                       5.059   Cond. No.                         19.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [79]:

```python
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[79]:

	Features	VIF
3	registered	21.21
0	yr	4.83
2	windspeed	3.51
9	aug	2.45
10	sep	2.43
11	oct	2.17
6	may	2.12
7	jun	2.11
8	jul	2.00
12	nov	1.91
5	apr	1.77
4	mar	1.72
13	dec	1.68
14	mon	1.64
19	w2	1.58
16	thu	1.51
15	tue	1.50
17	fri	1.48
18	sat	1.44
20	w3	1.23
1	holiday	1.19

In [80]:

```python
X = X.drop('dec', 1)
```

In [81]:

```python
X_train_lm = sm.add_constant(X)

lr_12 = sm.OLS(y_train, X_train_lm).fit()
```

In [82]:

```python
print(lr_12.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.954
Model:                            OLS   Adj. R-squared:                  0.952
Method:                 Least Squares   F-statistic:                     506.2
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          1.13e-311
Time:                        17:05:06   Log-Likelihood:                 823.25
No. Observations:                 510   AIC:                            -1605.
Df Residuals:                     489   BIC:                            -1516.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0464      0.010      4.871      0.000       0.028       0.065
yr             0.0319      0.007      4.789      0.000       0.019       0.045
holiday        0.0497      0.015      3.303      0.001       0.020       0.079
windspeed     -0.0597      0.014     -4.287      0.000      -0.087      -0.032
registered     0.8140      0.019     42.891      0.000       0.777       0.851
mar            0.0556      0.008      6.593      0.000       0.039       0.072
apr            0.0868      0.009      9.165      0.000       0.068       0.105
may            0.0963      0.010      9.859      0.000       0.077       0.116
jun            0.1003      0.010      9.649      0.000       0.080       0.121
jul            0.0991      0.010      9.867      0.000       0.079       0.119
aug            0.0953      0.010      9.831      0.000       0.076       0.114
sep            0.0952      0.011      8.963      0.000       0.074       0.116
oct            0.0851      0.010      8.716      0.000       0.066       0.104
nov            0.0352      0.009      3.801      0.000       0.017       0.053
mon           -0.0395      0.007     -5.455      0.000      -0.054      -0.025
tue           -0.0447      0.007     -6.067      0.000      -0.059      -0.030
thu           -0.0390      0.007     -5.287      0.000      -0.053      -0.024
fri           -0.0235      0.007     -3.146      0.002      -0.038      -0.009
sat            0.0637      0.007      9.265      0.000       0.050       0.077
w2            -0.0231      0.005     -4.674      0.000      -0.033      -0.013
w3            -0.0740      0.014     -5.205      0.000      -0.102      -0.046
==============================================================================
Omnibus:                       84.195   Durbin-Watson:                   1.899
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              163.380
Skew:                           0.932   Prob(JB):                     3.33e-36
Kurtosis:                       5.052   Cond. No.                         17.2
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [83]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[83]:

	Features	VIF
3	registered	16.04
0	yr	4.33
2	windspeed	3.46
9	aug	1.90
10	sep	1.90
11	oct	1.73
6	may	1.70
7	jun	1.70
13	mon	1.64
8	jul	1.63
12	nov	1.56

	Features	VIF
18	w2	1.54
15	thu	1.51
14	tue	1.49
5	apr	1.48
16	fri	1.48
4	mar	1.45
17	sat	1.41
1	holiday	1.19
19	w3	1.18

In [84]:

```
X = X.drop('registered', 1)
```

In [85]:

```
X_train_lm = sm.add_constant(X)

lr_13 = sm.OLS(y_train, X_train_lm).fit()
```

In [86]:

```
print(lr_13.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                     91.74
Date:                Thu, 03 Sep 2020   Prob (F-statistic):         1.02e-147
Time:                        17:05:07   Log-Likelihood:                 425.28
No. Observations:                 510   AIC:                            -810.6
Df Residuals:                     490   BIC:                            -725.9
Df Model:                          19
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.2894      0.017     17.310      0.000       0.257       0.322
yr             0.2473      0.010     25.819      0.000       0.228       0.266
holiday       -0.0900      0.032     -2.811      0.005      -0.153      -0.027
windspeed     -0.2056      0.029     -6.988      0.000      -0.263      -0.148
mar            0.1102      0.018      6.066      0.000       0.074       0.146
apr            0.2024      0.020     10.224      0.000       0.164       0.241
may            0.2795      0.019     14.591      0.000       0.242       0.317
jun            0.3046      0.020     15.131      0.000       0.265       0.344
jul            0.2727      0.020     13.606      0.000       0.233       0.312
aug            0.3014      0.018     16.425      0.000       0.265       0.337
sep            0.3387      0.020     17.309      0.000       0.300       0.377
oct            0.2685      0.019     14.040      0.000       0.231       0.306
nov            0.1709      0.019      8.994      0.000       0.134       0.208
mon            0.0115      0.016      0.738      0.461      -0.019       0.042
tue            0.0168      0.016      1.069      0.286      -0.014       0.048
thu            0.0277      0.016      1.766      0.078      -0.003       0.059
fri            0.0342      0.016      2.137      0.033       0.003       0.066
sat            0.0324      0.015      2.173      0.030       0.003       0.062
w2            -0.0864      0.010     -8.378      0.000      -0.107      -0.066
w3            -0.2854      0.029     -9.810      0.000      -0.343      -0.228
==============================================================================
Omnibus:                       41.693   Durbin-Watson:                   1.938
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              102.361
Skew:                          -0.419   Prob(JB):                     5.92e-23
Kurtosis:                       5.029   Cond. No.                         9.94
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [87]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[87]:

	Features	VIF
2	windspeed	3.36
0	yr	1.91
17	w2	1.52
12	mon	1.50
16	sat	1.40
3	mar	1.35
13	tue	1.34
14	thu	1.33
15	fri	1.33
4	apr	1.31
8	aug	1.25
10	oct	1.25
11	nov	1.25
5	may	1.22
9	sep	1.20
7	jul	1.19
6	jun	1.18
1	holiday	1.16
18	w3	1.10

In [88]:

```
X = X.drop('mon', 1)
```

In [89]:

```
X_train_lm = sm.add_constant(X)

lr_14 = sm.OLS(y_train, X_train_lm).fit()
```

In [90]:

```
print(lr_14.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.780
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                     96.90
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          1.33e-148
Time:                        17:05:07   Log-Likelihood:                 425.00
No. Observations:                 510   AIC:                            -812.0
Df Residuals:                     491   BIC:                            -731.5
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.2928      0.016     18.234      0.000       0.261       0.324
```

```
yr              0.2476      0.010     25.893      0.000       0.229       0.266
holiday        -0.0837      0.031     -2.714      0.007      -0.144      -0.023
windspeed      -0.2057      0.029     -6.995      0.000      -0.264      -0.148
mar             0.1104      0.018      6.079      0.000       0.075       0.146
apr             0.2026      0.020     10.235      0.000       0.164       0.241
may             0.2791      0.019     14.583      0.000       0.242       0.317
jun             0.3046      0.020     15.135      0.000       0.265       0.344
jul             0.2727      0.020     13.613      0.000       0.233       0.312
aug             0.3013      0.018     16.430      0.000       0.265       0.337
sep             0.3384      0.020     17.307      0.000       0.300       0.377
oct             0.2686      0.019     14.051      0.000       0.231       0.306
nov             0.1704      0.019      8.979      0.000       0.133       0.208
tue             0.0133      0.015      0.887      0.376      -0.016       0.043
thu             0.0240      0.015      1.615      0.107      -0.005       0.053
fri             0.0305      0.015      2.008      0.045       0.001       0.060
sat             0.0288      0.014      2.045      0.041       0.001       0.057
w2             -0.0861      0.010     -8.362      0.000      -0.106      -0.066
w3             -0.2856      0.029     -9.823      0.000      -0.343      -0.228
==============================================================================
Omnibus:                       41.170   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              100.139
Skew:                          -0.416   Prob(JB):                     1.80e-22
Kurtosis:                       5.005   Cond. No.                         9.87
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [91]:

```python
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[91]:

	Features	VIF
2	windspeed	3.23
0	yr	1.88
16	w2	1.50
3	mar	1.34
4	apr	1.30
15	sat	1.30
12	tue	1.26
14	fri	1.25
13	thu	1.24
8	aug	1.24
10	oct	1.24
11	nov	1.24
5	may	1.22
9	sep	1.19
7	jul	1.18
6	jun	1.17
17	w3	1.10
1	holiday	1.06

In [92]:

```python
X = X.drop('tue', 1)
```

In [93]:

```
X_train_lm = sm.add_constant(X)

lr_15 = sm.OLS(y_train, X_train_lm).fit()
```

In [94]:

```
print(lr_15.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.780
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                     102.6
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          1.91e-149
Time:                        17:05:08   Log-Likelihood:                 424.59
No. Observations:                 510   AIC:                            -813.2
Df Residuals:                     492   BIC:                            -737.0
Df Model:                          17                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.2957      0.016     18.799      0.000       0.265       0.327
yr             0.2475      0.010     25.887      0.000       0.229       0.266
holiday       -0.0861      0.031     -2.802      0.005      -0.147      -0.026
windspeed     -0.2054      0.029     -6.988      0.000      -0.263      -0.148
mar            0.1106      0.018      6.092      0.000       0.075       0.146
apr            0.2029      0.020     10.256      0.000       0.164       0.242
may            0.2798      0.019     14.632      0.000       0.242       0.317
jun            0.3040      0.020     15.117      0.000       0.264       0.343
jul            0.2736      0.020     13.671      0.000       0.234       0.313
aug            0.3018      0.018     16.468      0.000       0.266       0.338
sep            0.3387      0.020     17.327      0.000       0.300       0.377
oct            0.2688      0.019     14.063      0.000       0.231       0.306
nov            0.1708      0.019      9.003      0.000       0.134       0.208
thu            0.0211      0.014      1.453      0.147      -0.007       0.050
fri            0.0274      0.015      1.854      0.064      -0.002       0.056
sat            0.0258      0.014      1.886      0.060      -0.001       0.053
w2            -0.0861      0.010     -8.364      0.000      -0.106      -0.066
w3            -0.2867      0.029     -9.871      0.000      -0.344      -0.230
==============================================================================
Omnibus:                       39.555   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               93.054
Skew:                          -0.409   Prob(JB):                     6.22e-21
Kurtosis:                       4.926   Cond. No.                         9.82
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [95]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[95]:

	Features	VIF
2	windspeed	3.13
0	yr	1.88
15	w2	1.50
3	mar	1.33
4	apr	1.30
14	sat	1.24

	Features	VIF
10	oct	1.23
11	nov	1.23
8	aug	1.22
12	thu	1.20
13	fri	1.20
5	may	1.20
9	sep	1.18
6	jun	1.16
7	jul	1.15
16	w3	1.10
1	holiday	1.06

In [96]:

```
X = X.drop('thu', 1)
```

In [97]:

```
X_train_lm = sm.add_constant(X)

lr_16 = sm.OLS(y_train, X_train_lm).fit()
```

In [98]:

```
print(lr_16.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.779
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                     108.6
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          5.11e-150
Time:                        17:05:08   Log-Likelihood:                 423.50
No. Observations:                 510   AIC:                            -813.0
Df Residuals:                     493   BIC:                            -741.0
Df Model:                          16
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.3000      0.015     19.396      0.000       0.270       0.330
yr             0.2473      0.010     25.839      0.000       0.228       0.266
holiday       -0.0868      0.031     -2.820      0.005      -0.147      -0.026
windspeed     -0.2055      0.029     -6.981      0.000      -0.263      -0.148
mar            0.1115      0.018      6.140      0.000       0.076       0.147
apr            0.2021      0.020     10.207      0.000       0.163       0.241
may            0.2792      0.019     14.588      0.000       0.242       0.317
jun            0.3037      0.020     15.087      0.000       0.264       0.343
jul            0.2728      0.020     13.622      0.000       0.233       0.312
aug            0.3021      0.018     16.467      0.000       0.266       0.338
sep            0.3394      0.020     17.346      0.000       0.301       0.378
oct            0.2672      0.019     13.986      0.000       0.230       0.305
nov            0.1714      0.019      9.028      0.000       0.134       0.209
fri            0.0236      0.015      1.621      0.106      -0.005       0.052
sat            0.0218      0.013      1.627      0.104      -0.005       0.048
w2            -0.0867      0.010     -8.413      0.000      -0.107      -0.066
w3            -0.2856      0.029     -9.824      0.000      -0.343      -0.228
==============================================================================
Omnibus:                       38.462   Durbin-Watson:                   1.940
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               88.756
Skew:                          -0.403   Prob(JB):                     5.33e-20
Kurtosis:                       4.878   Cond. No.                         9.76
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [99]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[99]:

	Features	VIF
2	windspeed	3.07
0	yr	1.87
14	w2	1.50
3	mar	1.32
4	apr	1.30
10	oct	1.23
11	nov	1.22
8	aug	1.21
5	may	1.20
13	sat	1.20
9	sep	1.17
12	fri	1.17
6	jun	1.16
7	jul	1.15
15	w3	1.10
1	holiday	1.06

In [100]:

```
X = X.drop('fri', 1)
```

In [101]:

```
X_train_lm = sm.add_constant(X)

lr_17 = sm.OLS(y_train, X_train_lm).fit()
```

In [102]:

```
print(lr_17.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.778
Model:                            OLS   Adj. R-squared:                  0.771
Method:                 Least Squares   F-statistic:                     115.3
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          1.71e-150
Time:                        17:05:09   Log-Likelihood:                 422.14
No. Observations:                 510   AIC:                            -812.3
Df Residuals:                     494   BIC:                            -744.5
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.3036      0.015     19.803      0.000       0.273       0.334
yr             0.2469      0.010     25.767      0.000       0.228       0.266
holiday       -0.0882      0.031     -2.862      0.004      -0.149      -0.028
windspeed     -0.2053      0.029     -6.962      0.000      -0.263      -0.147
mar            0.1111      0.018      6.110      0.000       0.075       0.147
apr            0.2018      0.020     10.175      0.000       0.163       0.241
```

```
may            0.2787      0.019     14.541      0.000       0.241       0.316
jun            0.3030      0.020     15.032      0.000       0.263       0.343
jul            0.2734      0.020     13.630      0.000       0.234       0.313
aug            0.3027      0.018     16.471      0.000       0.267       0.339
sep            0.3390      0.020     17.297      0.000       0.300       0.377
oct            0.2667      0.019     13.943      0.000       0.229       0.304
nov            0.1698      0.019      8.942      0.000       0.133       0.207
sat            0.0182      0.013      1.372      0.171      -0.008       0.044
w2            -0.0854      0.010     -8.299      0.000      -0.106      -0.065
w3            -0.2880      0.029     -9.904      0.000      -0.345      -0.231
==============================================================================
Omnibus:                       39.364   Durbin-Watson:                   1.924
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               95.554
Skew:                          -0.396   Prob(JB):                     1.78e-21
Kurtosis:                       4.967   Cond. No.                         9.70
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [103]:

```
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[103]:

	Features	VIF
2	windspeed	3.02
0	yr	1.87
13	w2	1.47
3	mar	1.32
4	apr	1.30
10	oct	1.23
11	nov	1.22
5	may	1.20
8	aug	1.20
12	sat	1.18
6	jun	1.16
9	sep	1.16
7	jul	1.14
14	w3	1.10
1	holiday	1.06

In [104]:

```
X = X.drop('sat', 1)
```

In [105]:

```
X_train_lm = sm.add_constant(X)

lr_18 = sm.OLS(y_train, X_train_lm).fit()
```

In [106]:

```
print(lr_18.summary())
```

```
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                    cnt   R-squared:                       0.777
Model:                            OLS   Adj. R-squared:                  0.771
Method:                 Least Squares   F-statistic:                     123.2
Date:                Thu, 03 Sep 2020   Prob (F-statistic):          3.85e-151
Time:                        17:05:09   Log-Likelihood:                 421.17
No. Observations:                 510   AIC:                            -812.3
Df Residuals:                     495   BIC:                            -748.8
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.3063      0.015     20.136      0.000       0.276       0.336
yr             0.2465      0.010     25.712      0.000       0.228       0.265
holiday       -0.0911      0.031     -2.963      0.003      -0.152      -0.031
windspeed     -0.2033      0.029     -6.899      0.000      -0.261      -0.145
mar            0.1110      0.018      6.100      0.000       0.075       0.147
apr            0.2012      0.020     10.140      0.000       0.162       0.240
may            0.2772      0.019     14.473      0.000       0.240       0.315
jun            0.3028      0.020     15.007      0.000       0.263       0.342
jul            0.2731      0.020     13.606      0.000       0.234       0.313
aug            0.3026      0.018     16.452      0.000       0.266       0.339
sep            0.3387      0.020     17.268      0.000       0.300       0.377
oct            0.2665      0.019     13.917      0.000       0.229       0.304
nov            0.1693      0.019      8.910      0.000       0.132       0.207
w2            -0.0852      0.010     -8.275      0.000      -0.105      -0.065
w3            -0.2874      0.029     -9.877      0.000      -0.345      -0.230
==============================================================================
Omnibus:                       37.139   Durbin-Watson:                   1.920
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               92.877
Skew:                          -0.358   Prob(JB):                     6.79e-21
Kurtosis:                       4.964   Cond. No.                         9.62
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [107]:

```python
vif = pd.DataFrame()
vif['Features'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['VIF'] = round(vif['VIF'], 2)
vif = vif.sort_values(by = "VIF", ascending = False)
vif
```

Out[107]:

	Features	VIF
2	windspeed	2.93
0	yr	1.87
12	w2	1.47
3	mar	1.32
4	apr	1.30
10	oct	1.23
11	nov	1.22
5	may	1.20
8	aug	1.20
6	jun	1.16
9	sep	1.16
7	jul	1.14
13	w3	1.10
1	holiday	1.05

Step 6: Residual Analysis of the train data

In [108]:

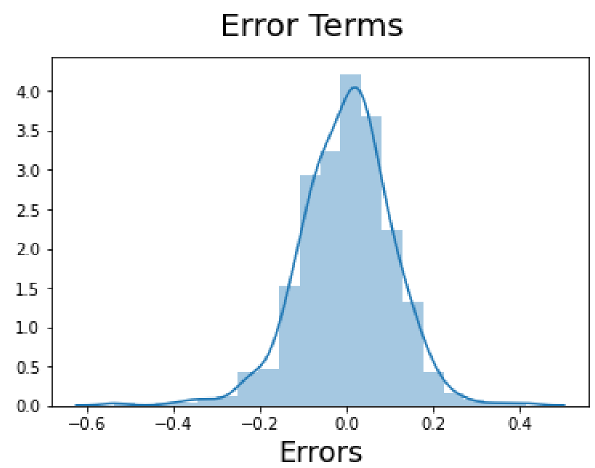
```
y_train_cnt = lr_18.predict(X_train_lm)
```

In [109]:

```
fig = plt.figure()
sns.distplot((y_train - y_train_cnt), bins = 20)
fig.suptitle('Error Terms', fontsize = 20)
plt.xlabel('Errors', fontsize = 18)
```

Out[109]:

```
Text(0.5, 0, 'Errors')
```



Step 7: Making Predictions Using the Final Model

In [110]:

```
num_vars = ['temp', 'atemp', 'hum', 'windspeed', 'casual','registered','cnt']

df_test[num_vars] = scaler.transform(df_test[num_vars])
```

In [111]:

```
df_test.describe()
```

Out[111]:

	yr	holiday	workingday	temp	atemp	hum	windspeed	casual	registered	cnt	...	
count	219.000000	219.000000	219.000000	219.000000	219.000000	219.000000	219.000000	219.000000	219.000000	219.000000	...	2
mean	0.479452	0.036530	0.698630	0.558941	0.532991	0.638508	0.313350	0.266372	0.527146	0.520592	...	
std	0.500722	0.188034	0.459904	0.233698	0.217888	0.148974	0.159947	0.217246	0.217921	0.218435	...	
min	0.000000	0.000000	0.000000	0.046591	0.025950	0.261915	-0.042808	0.002133	0.059486	0.048205	...	
25%	0.000000	0.000000	0.000000	0.354650	0.344751	0.527265	0.198517	0.101951	0.359154	0.377531	...	
50%	0.000000	0.000000	1.000000	0.558691	0.549198	0.627737	0.299459	0.223712	0.526567	0.524275	...	
75%	1.000000	0.000000	1.000000	0.759096	0.714132	0.743928	0.403048	0.362085	0.664742	0.672745	...	
max	1.000000	1.000000	1.000000	0.984424	0.980934	1.002146	0.807474	1.038708	0.985273	0.963300	...	

8 rows × 32 columns

In [112]:

```
y_test = df_test.pop('cnt')
X_test = df_test
```

In [113]:

```
X_test_m4 = sm.add_constant(X_test)
```

In [114]:

```
X_test_m4 = X_test_m4.drop(["wed", "temp", "winter","atemp", "workingday", "hum",
                            "summer", "fall", "casual","feb", "dec", "registered",
                            "mon", "tue", "thu","fri","sat"], axis = 1)
```

In [115]:

```
y_pred_m4 = lr_18.predict(X_test_m4)
```

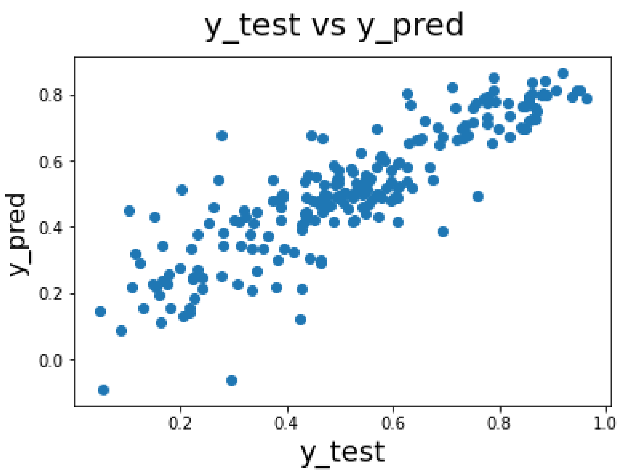
Step 8: Model Evaluation

In [116]:

```
fig = plt.figure()
plt.scatter(y_test, y_pred_m4)
fig.suptitle('y_test vs y_pred', fontsize = 20)
plt.xlabel('y_test', fontsize = 18)
plt.ylabel('y_pred', fontsize = 16)
```

Out[116]:

```
Text(0, 0.5, 'y_pred')
```



In [118]:

```
from sklearn.metrics import r2_score
r2_score(y_test, y_pred_m4)
```

Out[118]:

```
0.7600207568043166
```

Conclusion

cnt = (const × 0.3063)+ (yr × 0.2465)+ (holiday × -0.0911)+ (windspeed × -0.2033)+ (mar × 0.1110)+ (apr × 0.2012)+ (may × 0.2772)+ (jun × 0.3028)+ (jul × 0.2731)+ (aug × 0.3026)+ (sep × 0.3387)+ (oct × 0.2665)+ (nov × 0.1693)+ (w2 × -0.0852)+ (w3 × -0.2874)