Bike Sharing Demand and Prediction in Washington DC

1. Project Introduction

1.1 Introduction

The bike sharing system is one of the most popular way for short-time bicycle rental to promote the clean and sustainable transport system. The OBIS defined the scheme as the self-service, one-way-capable public services and ECF said that it is ideal for point-to-point trips (Chelma, Meunier and Wolfler, 2011). There are many popular platforms nowadays, such as LimeBike, Mobike. People can rent the bike from any location and return it to their destination through these systems. The bike sharing system clearly records travel time, departure location, arrival location and time. Therefore, it can be used to study the mobility in the cities. In this project, it is required to combine historical usage patterns with weather related data to analyse the influencing factors of bike rent and make prediction about the rental needs in Washington, DC as Washington is one of the most developed cities in U.S. and around the world.

1.2 Context

The real time bike availability can be easily accessed through the platform. According to Fishman (2013), the urban development encourages replacing car journeys with bicycles which also will reduce the land consumption. There are over 600 cities in 49 countries that have bike sharing system which is experienced the fastest growth of the public bike system transport mode nowadays(DeMiao, 2009)). The bike sharing system should meet the demand of fluctuating demand of bike at each station. Lin and Yang (2011) did the strategic research for rental capacity and locations.

1.3 Aims

This project is aimed to analyse the influencing factors of the bike sharing demand and would make the predictions of the future use as to study the effectiveness and impact on urban mobility management. It would reflect the current state of the public bike system in Washington and gather valuable evidence for the assessment of the system. The methodology includes the advanced regression and classification methods.

In [1]:
```python
#It would import the packages that would be used first.
# Import data package
import pandas as pd
import numpy as np

#Import plot package
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

#Import time series
from datetime import datetime

import sys
#Import ignore warning package
```

```
import warnings
warnings.filterwarnings("ignore")
```

2. Data Prepration

2.1 Import the Data

Data source: The data is divided into two parts, the train.csv and test.csv. The training part consists of data from the first 19 days of each month, and the test set is data from the 20th day of each month to the end of the month. The data provides hourly rental data spanning two years, including weather, season and other related information which is obtained from Kaggle from Capital Bikeshare.

In [2]:
```
# Import the dataset
train = pd.read_csv('C:/Users/36111/Desktop/DSSS-Bike/bike-sharing-demand/train.csv'
test = pd.read_csv('C:/Users/36111/Desktop/DSSS-Bike/bike-sharing-demand/test.csv')
Bike = train.append(test, sort=False)
Bike.head()
```

Out[2]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	r
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3.0	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8.0	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5.0	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3.0	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0.0	

2.2 Look Into Data

In [3]:
```
Bike.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 17379 entries, 0 to 6492
Data columns (total 12 columns):
datetime      17379 non-null object
season        17379 non-null int64
holiday       17379 non-null int64
workingday    17379 non-null int64
weather       17379 non-null int64
temp          17379 non-null float64
atemp         17379 non-null float64
humidity      17379 non-null int64
windspeed     17379 non-null float64
casual        10886 non-null float64
registered    10886 non-null float64
count         10886 non-null float64
dtypes: float64(6), int64(5), object(1)
memory usage: 1.7+ MB
```

2.3 Variable Description

Types of variables:

- Categorical: Season, Holiday, Working day, Weather
- Timeseries: Datetime
- Numerical: Temp, aTemp, Humidity, Windspeed, Casual, Registed, Count

- Datetime – year, month, day and time
- Season – 1:Spring, 2:Summer, 3:Autumn, 4:Winter
- Holiday – whether it is the holiday, 0 represents not the holiday while 1 represents the holiday
- Workingday – 0 represents the weekend while 1 represents the working day
- Weather– 1: : Clear, Few clouds, Partly cloudy, Partly cloudy 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp - temperature in Celsius
- atemp - "feels like" temperature in Celsius
- humidity - relative humidity
- windspeed - wind speed
- casual - number of non-registered user rentals initiated
- registered - number of registered user rentals initiated
- count - number of total rentals

In [4]:
```
Bike.describe()
```

Out[4]:

	season	holiday	workingday	weather	temp	atemp	hun
count	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.000000	17379.0(
mean	2.501640	0.028770	0.682721	1.425283	20.376474	23.788755	62.7:
std	1.106918	0.167165	0.465431	0.639357	7.894801	8.592511	19.2!
min	1.000000	0.000000	0.000000	1.000000	0.820000	0.000000	0.0(
25%	2.000000	0.000000	0.000000	1.000000	13.940000	16.665000	48.0(
50%	3.000000	0.000000	1.000000	1.000000	20.500000	24.240000	63.0(
75%	3.000000	0.000000	1.000000	2.000000	27.060000	31.060000	78.0(
max	4.000000	1.000000	1.000000	4.000000	41.000000	50.000000	100.0(

It would check the data type and whether there are some NaN values. To inspect into the data, there are not some NaN values in the dataset so it just needs to deal with the outliers.

In [5]:
```
Bike.isnull().values.any()
```

Out[5]:
```
True
```

In [6]:
```
Bike.isnull().sum()
```

Out[6]:
```
datetime        0
season          0
```

```
holiday          0
workingday       0
weather          0
temp             0
atemp            0
humidity         0
windspeed        0
casual        6493
registered    6493
count         6493
dtype: int64
```

The columns ‘season’, ‘holiday’, ‘working day’ and ‘weather’ should be of ‘categorical’ data type. It would transform the dataset so that it can get started up. It divides the datetime to the date, year, month, week, day and hour for further analysis.

In [7]:
```
train['date']=train['datetime'].apply(lambda c: c.split()[0])
train['week']=train['date'].apply(lambda c : datetime.strptime(c,'%Y-%m-%d').isoweek
train['Year']=train['datetime'].apply(lambda c : c.split()[0].split('-')[0]).astype(
train['Month']=train['datetime'].apply(lambda c : c.split()[0].split('-')[1]).astype
train['Day']=train['datetime'].apply(lambda c : c.split()[0].split('-')[2]).astype('
train['Hour']=train['datetime'].apply(lambda c : c.split()[1].split(':')[0]).astype(
train.head()
```

Out[7]:

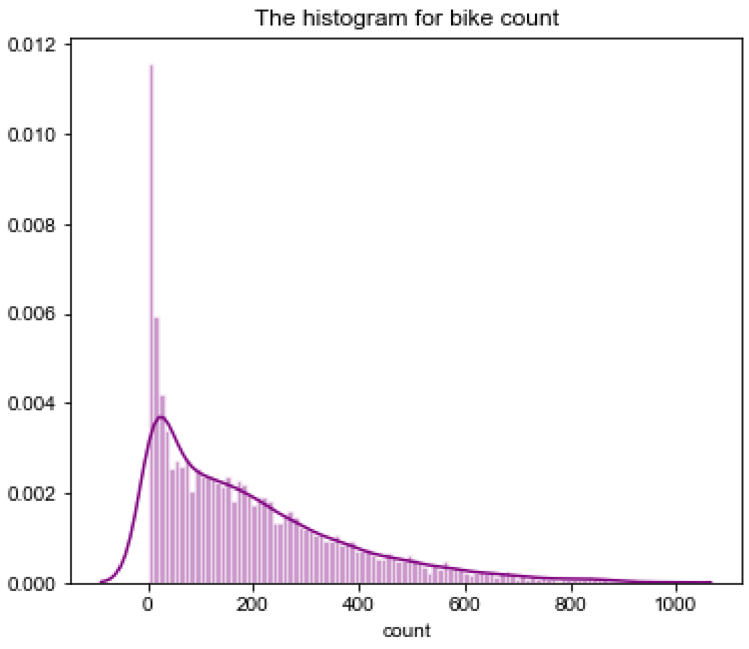
	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	r
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0	

3. Data Pattern Analysis

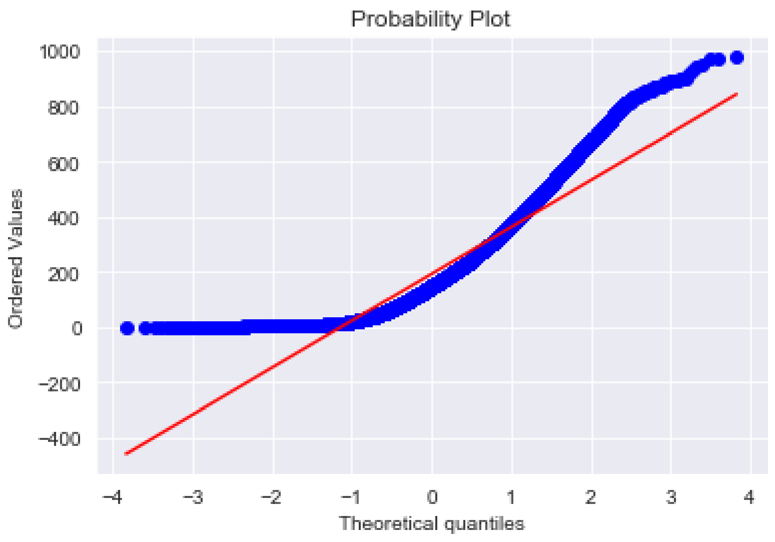
3.1 Exploratory Data Analysis

In [8]:
```
# Frist, it will plot the histogram for count
fig = plt.figure()
ax = fig.add_subplot(1,1,1)
fig.set_size_inches(6,5)
sns.set_style('darkgrid')
sns.distplot(train['count'], bins = 100, color = 'purple')
ax.set(xlabel='count', title='The histogram for bike count')
```

Out[8]:
```
[Text(0.5, 0, 'count'), Text(0.5, 1.0, 'The histogram for bike count')]
```

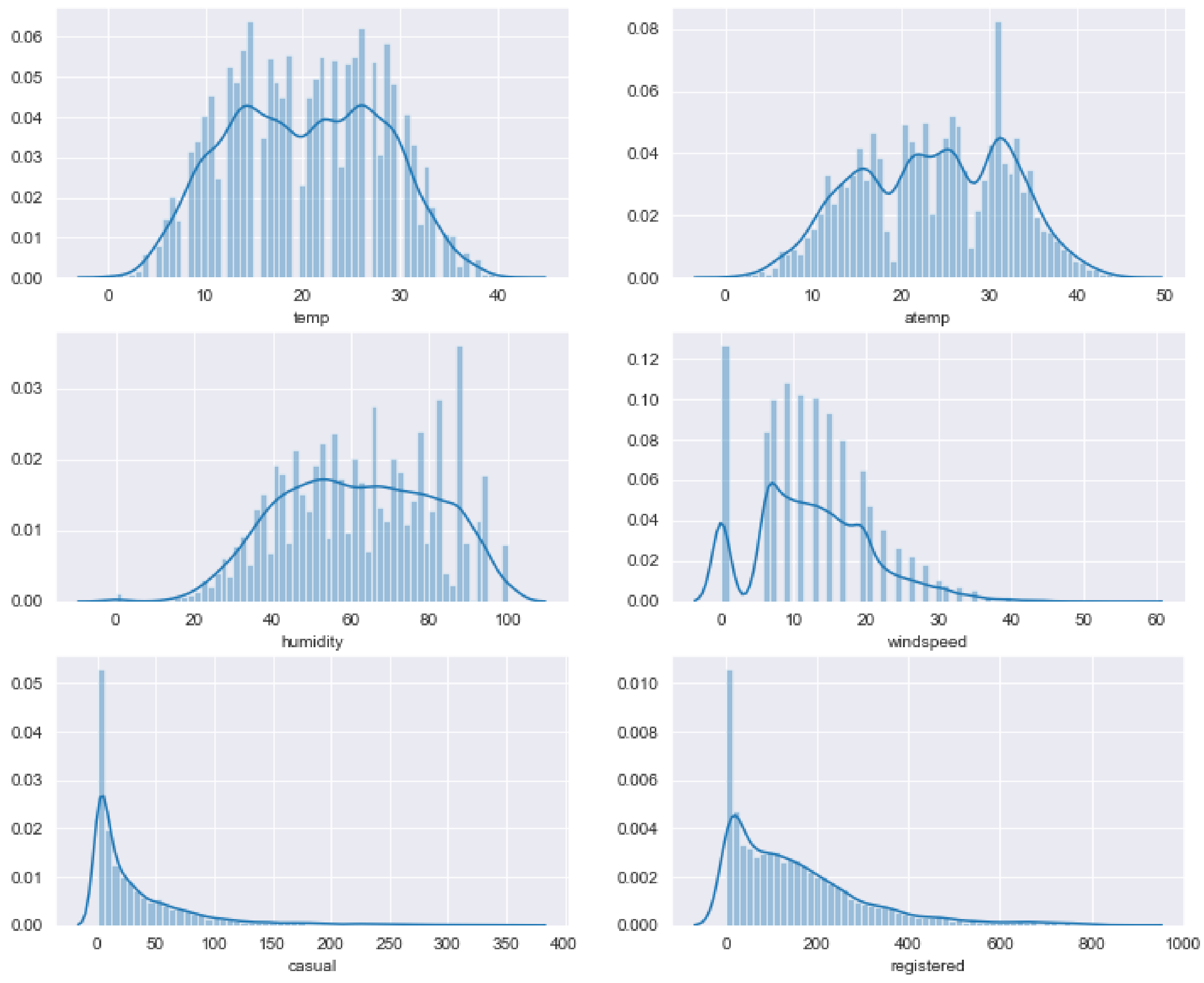


In [9]:
```python
# The Q-Q plot will also be did
from scipy import stats
plt = stats.probplot(train['count'], plot = sns.mpl.pyplot)
```



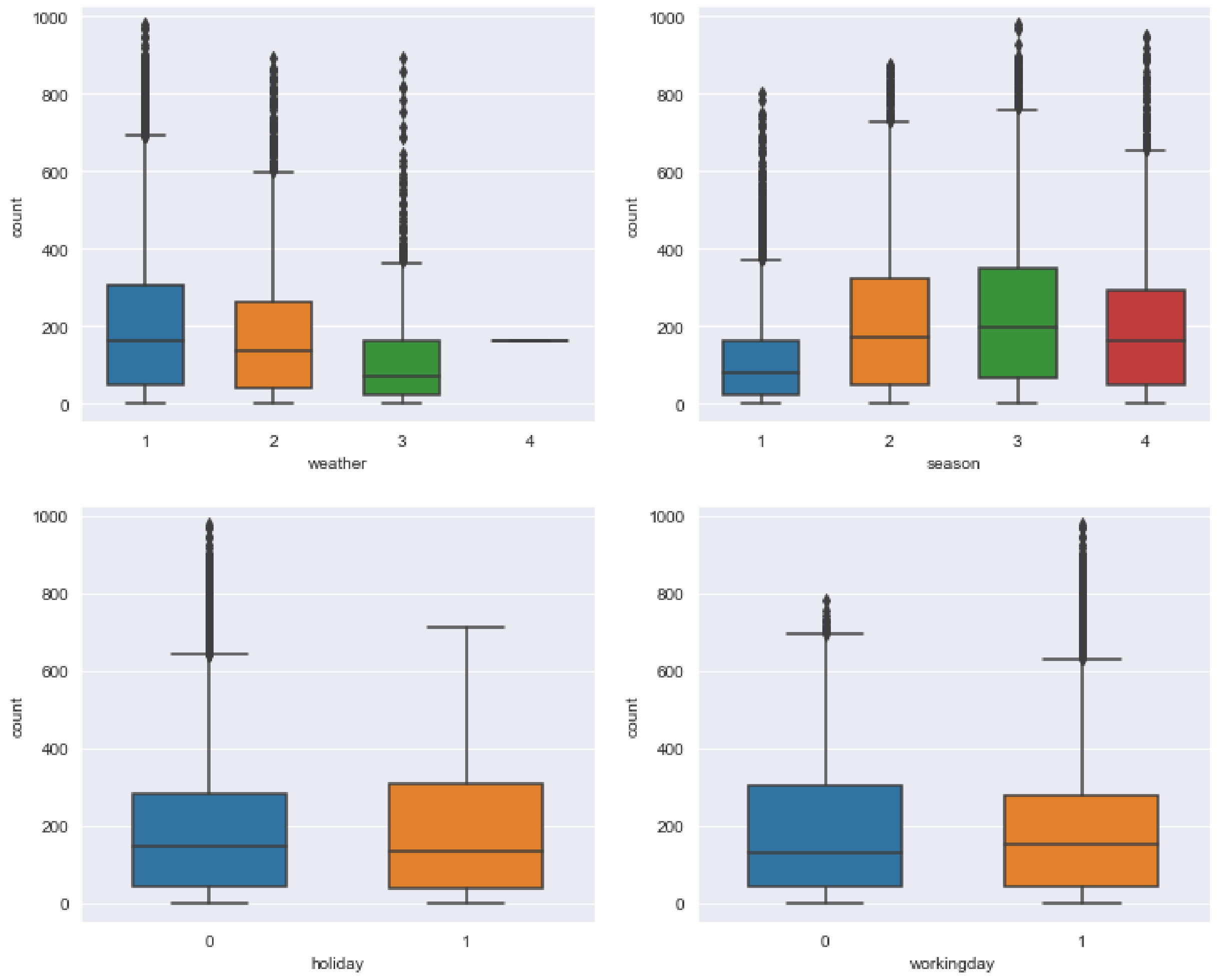
In [10]:
```python
import matplotlib.pyplot as plt
fig, axes = plt.subplots(3,2)
fig.set_size_inches(12,10)

sns.distplot(train['temp'], bins = 60, ax=axes[0,0])
sns.distplot(train['atemp'], bins = 60, ax=axes[0,1])
sns.distplot(train['humidity'], bins = 60, ax=axes[1,0])
sns.distplot(train['windspeed'], bins = 60, ax=axes[1,1])
sns.distplot(train['casual'], bins = 60, ax=axes[2,0])
sns.distplot(train['registered'], bins = 60, ax=axes[2,1])
```

Out[10]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d2af19940>
```



In [11]:
```
# Boxplot
fig, axes = plt.subplots(2,2)
fig.set_size_inches(12,10)

sns.boxplot(x='weather', y='count', data=train, orient='v',width=0.6, ax=axes[0,0])
sns.boxplot(x='season', y='count', data=train, orient='v',width=0.6, ax=axes[0,1])
sns.boxplot(x='holiday', y='count', data=train, orient='v',width=0.6, ax=axes[1,0])
sns.boxplot(x='workingday', y='count', data=train, orient='v',width=0.6, ax=axes[1,1
```

Out[11]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d2b321978>
```



The histogram plot and q-q plot for bike counts are made. The data is desired to normally distributed as most of machine learning algorithms require the data to be normally distributed. The count number distribution looks good but not ideally following normal distribution. Then, it has made the distribution and boxplot of other aspects. The spring season has relatively lower count. Most of the outliers are mainly contributed from ‘working day’ than ‘non-working day’. From the plots, it can be found that there are some extreme outliers between windspeed 0-10 that need to be replaced. In this project, it would use the random forest to fill these outliers. It would train the model to predict the windspeed and using it to replace the outliers data of original windspeed data. After that, it checked the windspeed data distribution again and found that it is normally distributed.

In [12]:
```
# replace zero value in windspeed
# understnd the co of windspeed
corrws = train.corr()
corrws['windspeed'].sort_values(ascending =False)
```

Out[12]:
```
windspeed     1.000000
Hour          0.146631
count         0.101369
casual        0.092276
registered    0.091052
Day           0.036157
workingday    0.013373
holiday       0.008409
weather       0.007261
Year         -0.015221
temp         -0.017852
week         -0.024804
atemp        -0.057473
season       -0.147121
Month        -0.150192
```

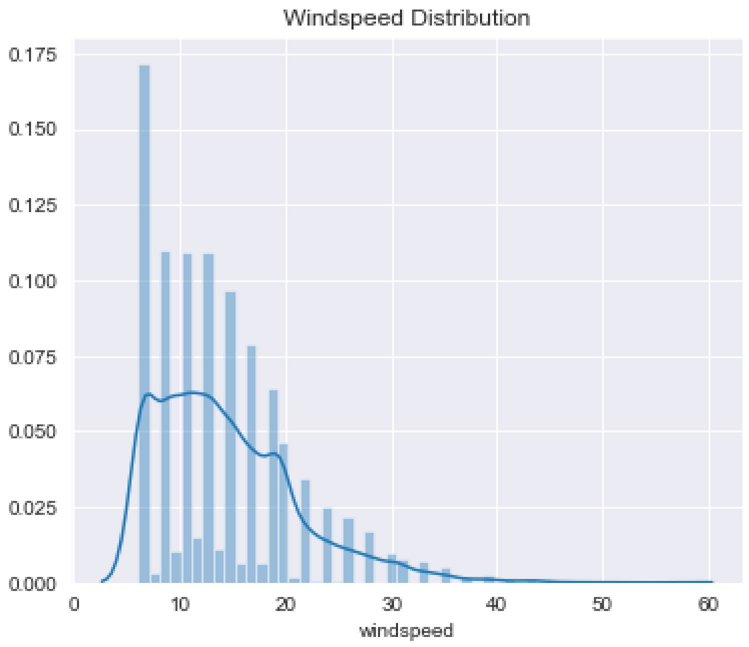
```
humidity     -0.318607
Name: windspeed, dtype: float64
```

In [13]:
```
from sklearn.ensemble import RandomForestRegressor

# Divide the data to two parts, the windspeed = 0 or != 0
wind0 = train[train['windspeed']==0]
wind = train[train['windspeed']!=0]

# Construct model
model = RandomForestRegressor(n_estimators=1000, random_state=42)

# Select characteristic value
wind_x = wind[['Hour','count','season','Month','humidity']]

# Select lable value
wind_y = wind['windspeed']

# Select prediction characteristic value
windpre_x = wind0[['Hour','count','season','Month','humidity']]

# It will take the data that windspeed not equal to zero as the training set and fit
model.fit(wind_x,wind_y)

# Predict windspeed through the pre-trained model
wind0Values = model.predict(X = windpre_x)

# Fulfill the value into the wind0 dataset
wind0.loc[:,'windspeed'] = wind0Values
```

In [14]:
```
data = wind.append(wind0)
data.reset_index(inplace=True)
data.drop('index',inplace=True,axis=1)
```

In [15]:
```
# windspeed distribution

fig = plt.figure()
ax = fig.add_subplot(1,1,1)

fig.set_size_inches(6,5)
sns.distplot(data['windspeed'])

ax.set(xlabel='windspeed',title='Windspeed Distribution')
```

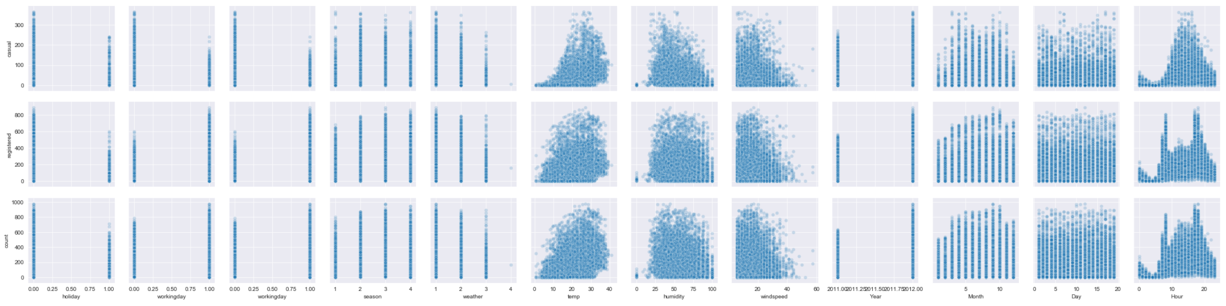
Out[15]:
```
[Text(0.5, 0, 'windspeed'), Text(0.5, 1.0, 'Windspeed Distribution')]
```



3.2 Overall Data pattern

In [16]:
```
#Plot the correlation plot
sns.pairplot(data, x_vars=['holiday','workingday','workingday','season','weather','t
             y_vars=['casual','registered','count'], plot_kws={'alpha': 0.2})
```

Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x24d57700400>
```



From the overall perspective, it can be observed that:

1. The total number of bike rental in non-holiday is higher than the total number of holiday bike rental.
2. Members would borrow more bikes on working days, and non-members would borrow more on weekends.
3. There are least bicycle rental in the first quarter of the year.
4. The weather has a significant effect on the amount of bike borrowed.
5. The temperature and humidity have a greater impact on non-members and less on members
6. The number of hours has a significant impact on the rental situation. Members show two peaks of the distribution, and non-members are normally distributed.

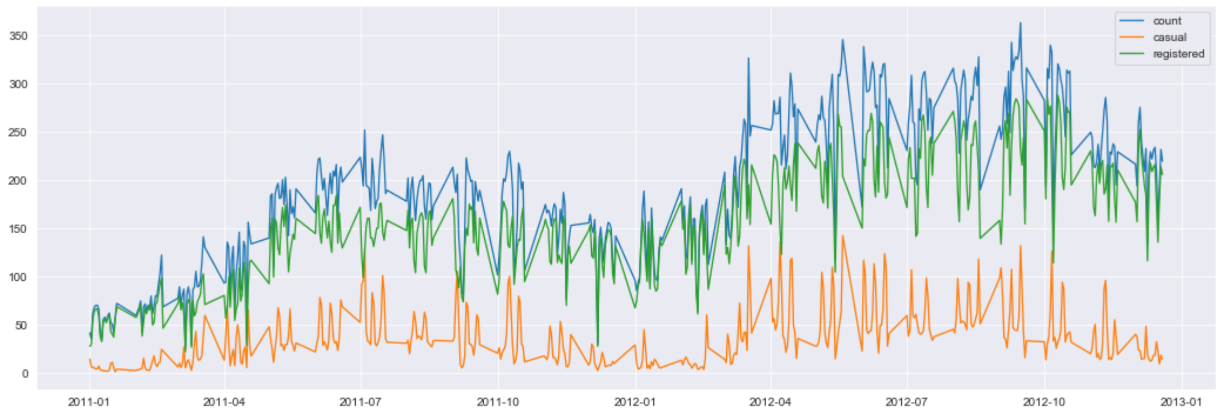
3.3 Item-by-item Anlysis

Then, it would analyse the pattern from specific perspective.

In [17]:
```
date = data.groupby(['date'], as_index=False).agg({'count': 'mean','casual': 'mean',
                                                  'registered' : 'mean'})
# transfer date to datestamp
date['date']=pd.to_datetime(date['date'])
```

```
#plot
fig = plt.figure(figsize=(18,6))
ax= fig.add_subplot(1,1,1)
plt.plot(date['date'],date['count'], linewidth=1.3, label='count')
plt.plot(date['date'],date['casual'], linewidth=1.3, label='casual')
plt.plot(date['date'],date['registered'], linewidth=1.3, label='registered')
plt.legend()
```

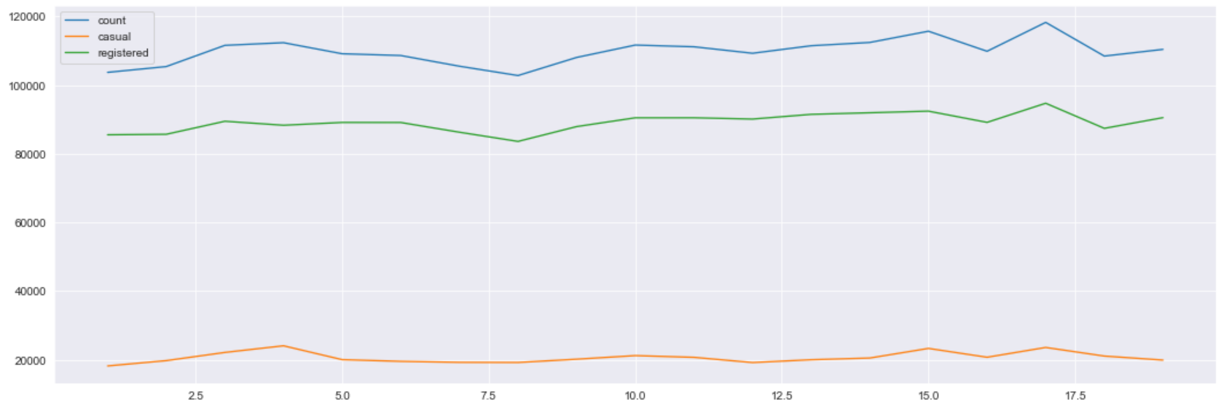
Out[17]:
```
<matplotlib.legend.Legend at 0x24d5872a518>
```



As to the time series analysis, it would plot the trend of total count change from year, month, week and day. As to the year, it can be found that more people prefer to use in the spring, summer and fall season and there has less people use in the winter maybe owing to the lower temperature. The bike count number is increasing from 2011 to 2012.

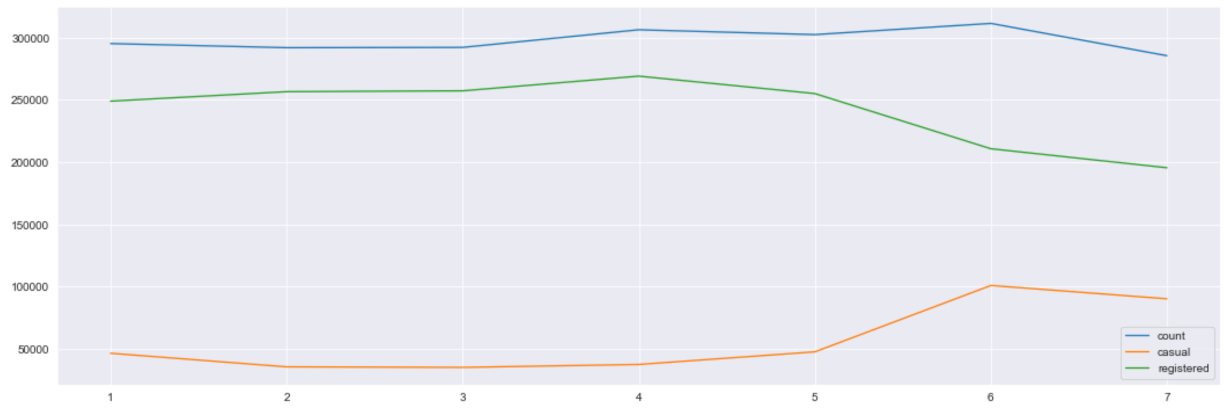
In [18]:
```
# For Month
date = data.groupby(['Day'], as_index=False).agg({'count': 'sum','casual': 'sum',
                                                  'registered' : 'sum'})

#plot
fig = plt.figure(figsize=(18,6))
ax= fig.add_subplot(1,1,1)
plt.plot(date['Day'],date['count'], linewidth=1.3, label='count')
plt.plot(date['Day'],date['casual'], linewidth=1.3, label='casual')
plt.plot(date['Day'],date['registered'], linewidth=1.3, label='registered')
plt.legend()
```

Out[18]:
```
<matplotlib.legend.Legend at 0x24d58e63e80>
```



In [19]:
```
#For Week
date = data.groupby(['week'], as_index=False).agg({'count': 'sum','casual': 'sum',
                                                   'registered' : 'sum'})

#plot
fig = plt.figure(figsize=(18,6))
ax= fig.add_subplot(1,1,1)
plt.plot(date['week'],date['count'], linewidth=1.3, label='count')
plt.plot(date['week'],date['casual'], linewidth=1.3, label='casual')
plt.plot(date['week'],date['registered'], linewidth=1.3, label='registered')
plt.legend()
```

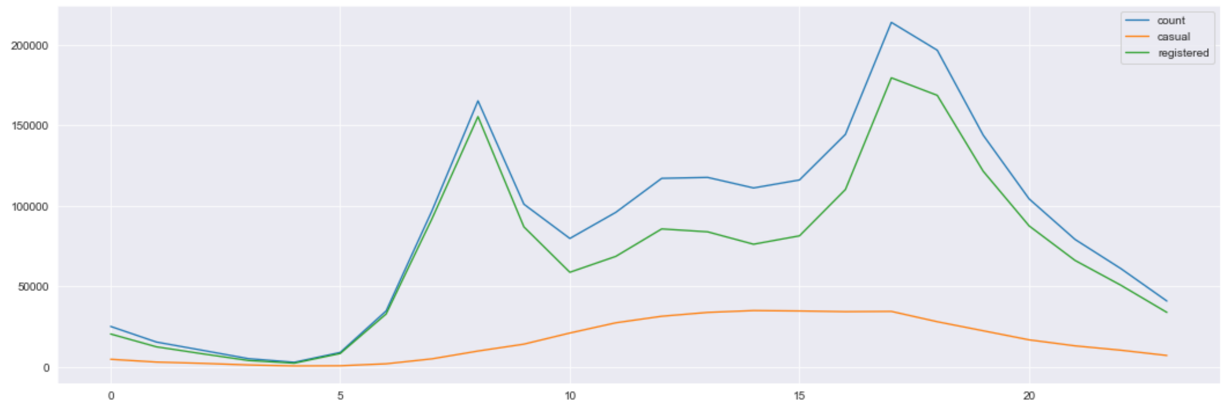
Out[19]:
```
<matplotlib.legend.Legend at 0x24d58c364e0>
```



As to the month and week, there are not too much change over the time.

In [20]:
```
# For day
date = data.groupby(['Hour'], as_index=False).agg({'count': 'sum','casual': 'sum',
                                                  'registered' : 'sum'})

#plot
fig = plt.figure(figsize=(18,6))
ax= fig.add_subplot(1,1,1)
plt.plot(date['Hour'],date['count'], linewidth=1.3, label='count')
plt.plot(date['Hour'],date['casual'], linewidth=1.3, label='casual')
plt.plot(date['Hour'],date['registered'], linewidth=1.3, label='registered')
plt.legend()
```

Out[20]:
```
<matplotlib.legend.Legend at 0x24d594098d0>
```



There are two obvious peaks over the daily changes at around 7 am - 8 am and 5 pm - 6 pm and this may be attributed to regular school and office commuters. However, this pattern is not observed on weekend that more people tend to use the bike sharing system between 10 am and 4 pm.

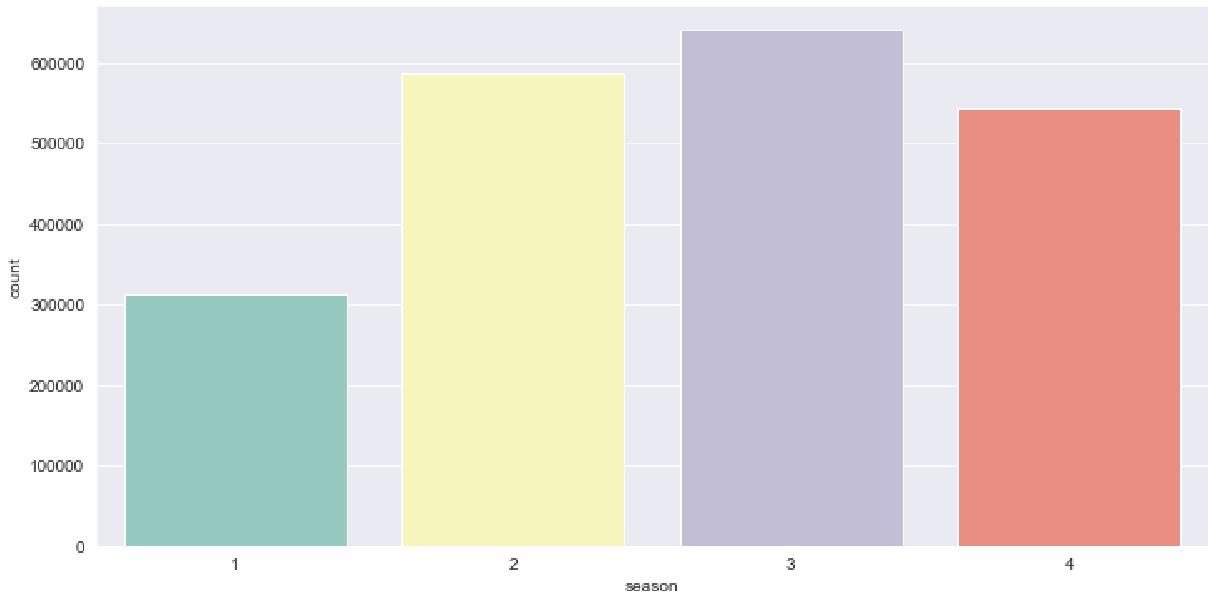
Then, it will analyse the influencing factors.

- Season: There are fewest bike rentals in spring while the most in summer.
- Weather: The weather condition impacts a lot on the count and it has the decreasing trend.
- Humidity: The count of bike rental has a clear decreasing trend while the humidity is increasing.
- Windspeed: The bike rental number is first increasing and then decreasing with the windspeed rising.

- Temperature: The total bike rental is clearly rising with the temperature.
- Workingday and holiday: The working day and holiday does not influence a lot to the bike rental number.

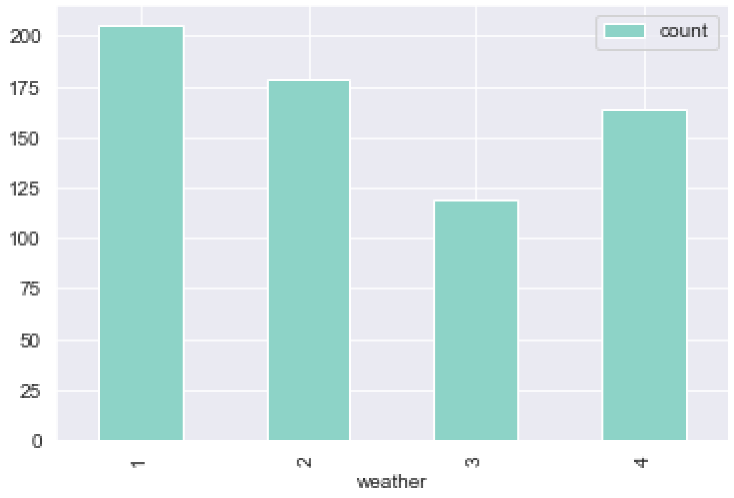
In [21]:
```
# For season
season = data.groupby(['season'], as_index=False).agg({'count':'sum','casual':'sum',

#plot
seasonx = season['season']
seasony = season['count']

fig = plt.figure()
ax=fig.add_subplot(1,1,1)
fig.set_size_inches(12,6)

sns.barplot(seasonx,seasony, palette='Set3')
```

Out[21]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d59442fd0>
```



In [22]:
```
# For weather
weather = data.groupby(['weather'], as_index=True).agg({'count':'mean'})
#plot
weather.plot(kind='bar',colormap='Set3')
```

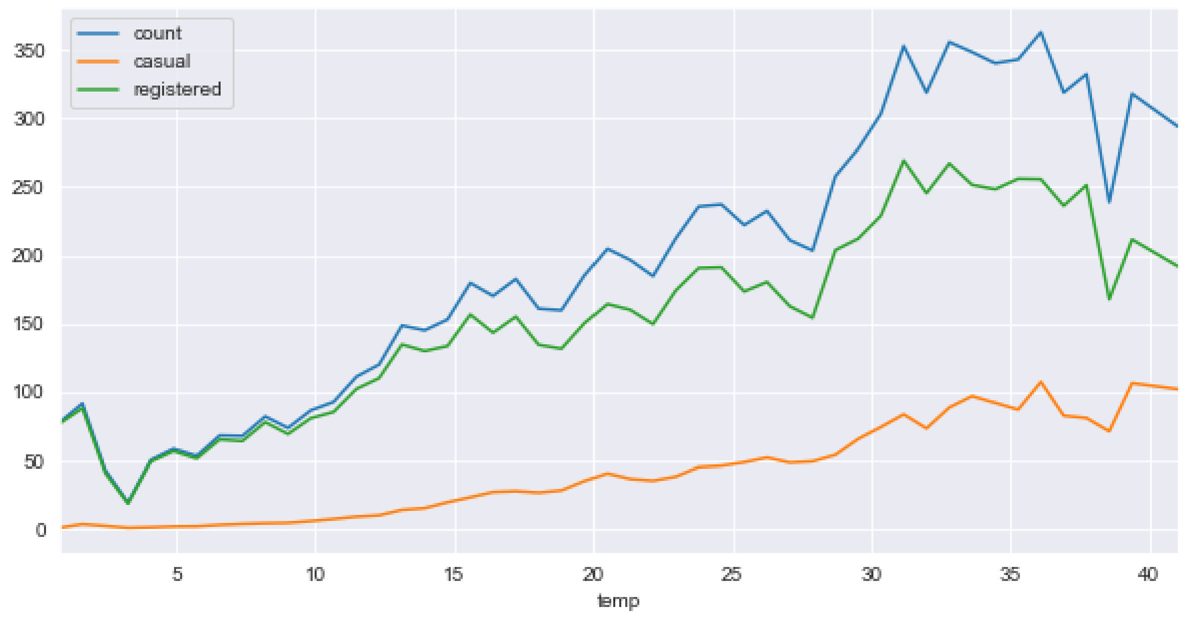
Out[22]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d5903d828>
```



In [23]:
```
# For temperature
```

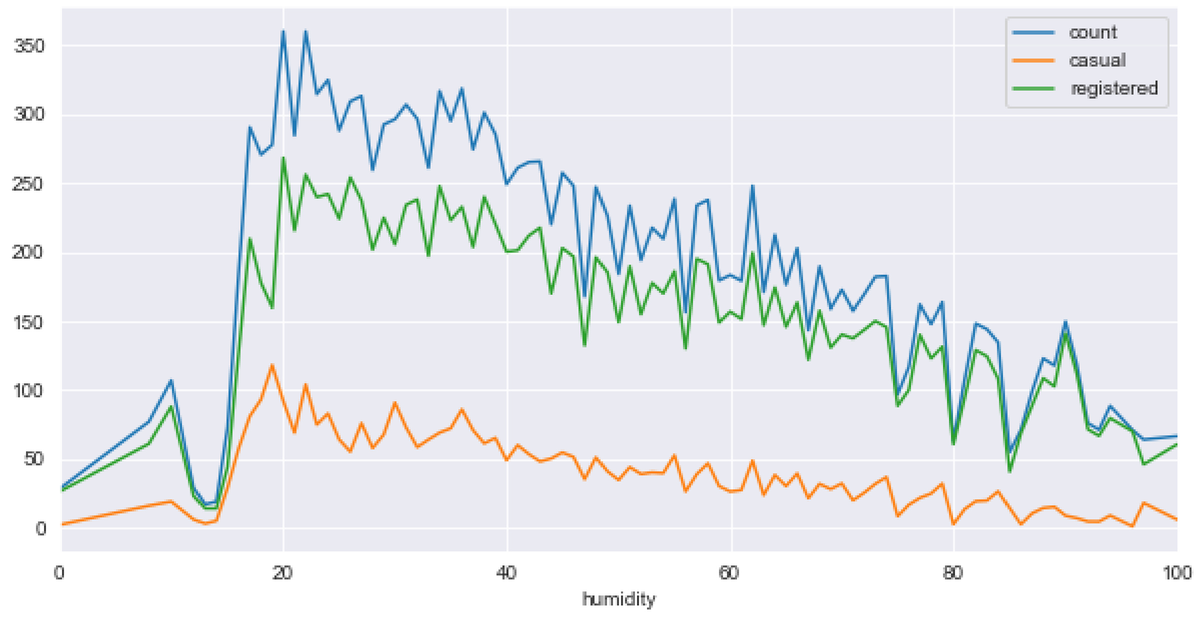
```
temp = data.groupby(['temp'], as_index=True).agg({'count':'mean','casual':'mean','re
#plot
temp.plot(figsize=(10,5))
```

Out[23]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d5907dba8>
```



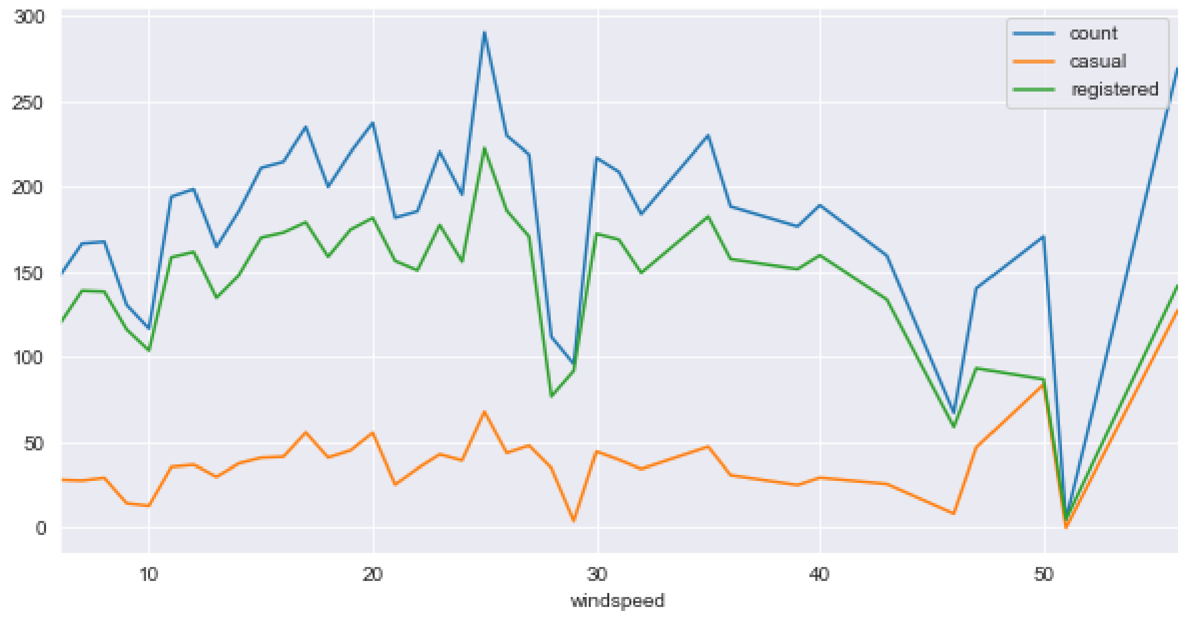
In [24]:
```
# For humidity
humi = data.groupby(['humidity'], as_index=True).agg({'count':'mean','casual':'mean'
#plot
humi.plot(figsize=(10,5))
```

Out[24]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d59108358>
```



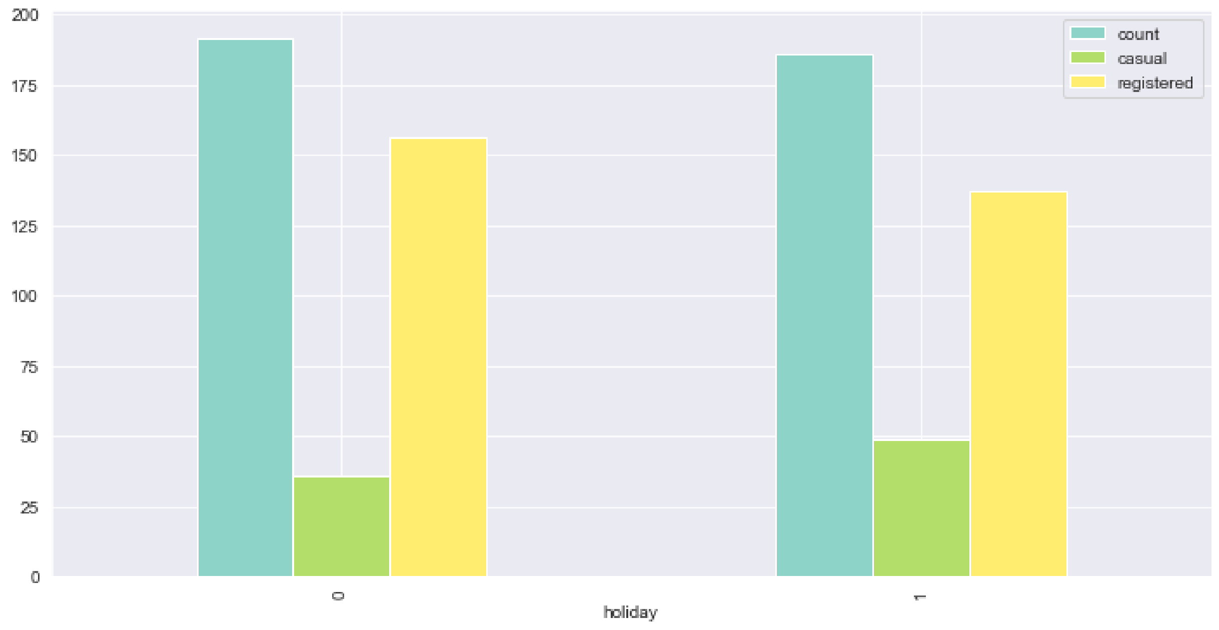
In [25]:
```
# For temperature
data['windspeed'] = data['windspeed'].astype(int)
winds = data.groupby(['windspeed'], as_index=True).agg({'count':'mean','casual':'mea
#plot
winds.plot(figsize=(10,5))
```

Out[25]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d5918f2b0>
```



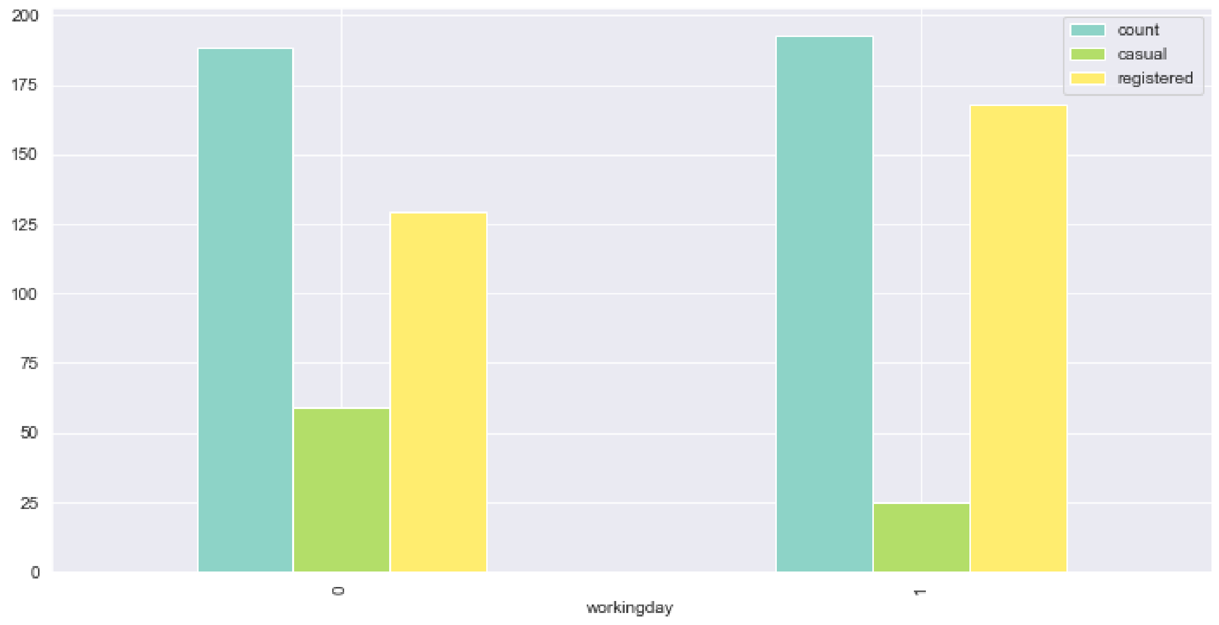
In [26]:
```
# For Holiday
holiday = data.groupby(['holiday'], as_index=True).agg({'count':'mean','casual':'mea
#plot
holiday.plot(kind='bar', figsize=(12,6),colormap='Set3')
```

Out[26]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d59733be0>
```



In [27]:
```
# For workday
workday = data.groupby(['workingday'], as_index=True).agg({'count':'mean','casual':'
#plot
workday.plot(kind='bar', figsize=(12,6),colormap='Set3')
```

Out[27]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d597822b0>
```



4. Data preprocessing

4.1 Data transformation

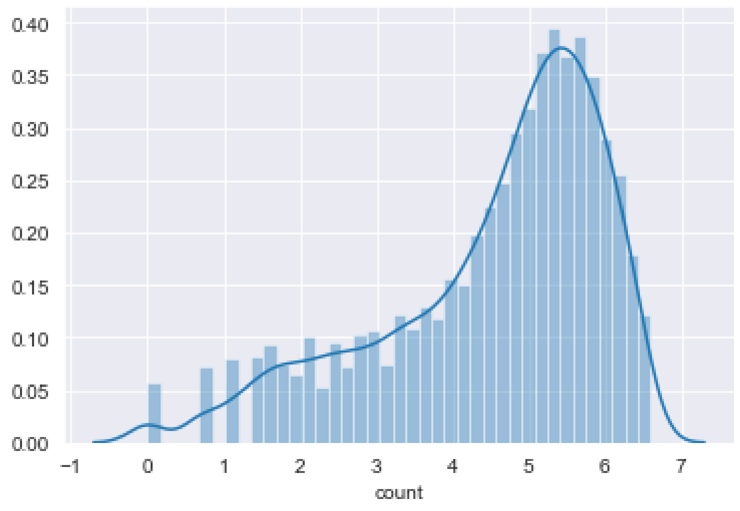
In this project, the value is set to the ‘count’ number. Due to the uneven distribution of the values, it would remove the values other than three times the variance, and then the logarithmic transformation processing is performed.

In [28]:
```
# Remove values other than 3 times the variance
data_std = data[np.abs(data['count'] -
                       data['count'].mean())<=(3*data['count'].std())]
data_std.shape
```

Out[28]:
```
(10739, 18)
```

In [29]:
```
# Log transformation
ylabels = data_std['count']
ylabels_log = np.log(ylabels)
sns.distplot(ylabels_log)
```

Out[29]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d59813eb8>
```



In [30]:
```
# transfer the datetime data of test dataset
test['date']=test['datetime'].apply(lambda c: c.split()[0])
```

```
test['week']=test['date'].apply(lambda c : datetime.strptime(c,'%Y-%m-%d').isoweekda
test['Year']=test['datetime'].apply(lambda c : c.split()[0].split('-')[0]).astype('i
test['Month']=test['datetime'].apply(lambda c : c.split()[0].split('-')[1]).astype('
test['Day']=test['datetime'].apply(lambda c : c.split()[0].split('-')[2]).astype('in
test['Hour']=test['datetime'].apply(lambda c : c.split()[1].split(':')[0]).astype('i
test.head()
```

Out[30]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	date	w
0	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027	2011-01-20	
1	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000	2011-01-20	
2	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000	2011-01-20	
3	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014	2011-01-20	
4	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014	2011-01-20	

In [31]:

```
# make log transformation and unit the dataset
data_std['count'] = np.log(data_std['count'])
Full_Bike = data_std.append(test, ignore_index = True)
print('The full dataset:', Full_Bike.shape)
```

```
The full dataset: (17232, 18)
```

In [32]:

```
Full_Bike
```

Out[32]:

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	holiday	humidity	reg
0	1	5	1	2011	12.880	0.0	0.000000	2011-01-01	2011-01-01 05:00:00	0	75	
1	1	10	1	2011	19.695	12.0	3.583519	2011-01-01	2011-01-01 10:00:00	0	76	
2	1	11	1	2011	16.665	26.0	4.025352	2011-01-01	2011-01-01 11:00:00	0	81	
3	1	12	1	2011	21.210	29.0	4.430817	2011-01-01	2011-01-01 12:00:00	0	77	
4	1	13	1	2011	22.725	47.0	4.543295	2011-01-01	2011-01-01 13:00:00	0	72	
5	1	14	1	2011	22.725	35.0	4.663439	2011-01-01	2011-01-01 14:00:00	0	72	

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	holiday	humidity	reg
6	1	15	1	2011	21.970	40.0	4.700480	2011-01-01	2011-01-01 15:00:00	0	77	
7	1	16	1	2011	21.210	41.0	4.532599	2011-01-01	2011-01-01 16:00:00	0	82	
8	1	17	1	2011	21.970	15.0	4.204693	2011-01-01	2011-01-01 17:00:00	0	82	
9	1	18	1	2011	21.210	9.0	3.555348	2011-01-01	2011-01-01 18:00:00	0	88	
10	1	19	1	2011	21.210	6.0	3.610918	2011-01-01	2011-01-01 19:00:00	0	88	
11	1	20	1	2011	20.455	11.0	3.583519	2011-01-01	2011-01-01 20:00:00	0	87	
12	1	21	1	2011	20.455	3.0	3.526361	2011-01-01	2011-01-01 21:00:00	0	87	
13	1	22	1	2011	20.455	11.0	3.332205	2011-01-01	2011-01-01 22:00:00	0	94	
14	1	23	1	2011	22.725	15.0	3.663562	2011-01-01	2011-01-01 23:00:00	0	88	
15	2	0	1	2011	22.725	4.0	2.833213	2011-01-02	2011-01-02 00:00:00	0	88	
16	2	1	1	2011	21.970	1.0	2.833213	2011-01-02	2011-01-02 01:00:00	0	94	
17	2	2	1	2011	21.210	1.0	2.197225	2011-01-02	2011-01-02 02:00:00	0	100	
18	2	3	1	2011	22.725	2.0	1.791759	2011-01-02	2011-01-02 03:00:00	0	94	
19	2	4	1	2011	22.725	2.0	1.098612	2011-01-02	2011-01-02 04:00:00	0	94	
20	2	6	1	2011	21.210	0.0	0.693147	2011-01-02	2011-01-02 06:00:00	0	77	
21	2	7	1	2011	20.455	0.0	0.000000	2011-01-02	2011-01-02 07:00:00	0	76	

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	holiday	humidity	reg
22	2	8	1	2011	20.455	0.0	2.079442	2011-01-02	2011-01-02 08:00:00	0	71	
23	2	9	1	2011	19.695	1.0	2.995732	2011-01-02	2011-01-02 09:00:00	0	76	
24	2	10	1	2011	17.425	7.0	3.970292	2011-01-02	2011-01-02 10:00:00	0	81	
25	2	11	1	2011	16.665	16.0	4.248495	2011-01-02	2011-01-02 11:00:00	0	71	
26	2	12	1	2011	16.665	20.0	4.532599	2011-01-02	2011-01-02 12:00:00	0	66	
27	2	13	1	2011	17.425	11.0	4.317488	2011-01-02	2011-01-02 13:00:00	0	66	
28	2	14	1	2011	17.425	4.0	4.077537	2011-01-02	2011-01-02 14:00:00	0	76	
29	2	15	1	2011	16.665	19.0	4.304065	2011-01-02	2011-01-02 15:00:00	0	81	
...	...	...	...	...	...	...	...	...	...	...	...	
17202	30	18	12	2012	10.605	NaN	NaN	2012-12-30	2012-12-30 18:00:00	0	44	
17203	30	19	12	2012	18.180	NaN	NaN	2012-12-30	2012-12-30 19:00:00	0	61	
17204	30	20	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 20:00:00	0	47	
17205	30	21	12	2012	10.605	NaN	NaN	2012-12-30	2012-12-30 21:00:00	0	51	
17206	30	22	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 22:00:00	0	55	
17207	30	23	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 23:00:00	0	51	
17208	31	0	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 00:00:00	0	55	
17209	31	1	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 01:00:00	0	55	

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	holiday	humidity	reg
17210	31	2	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 02:00:00	0	59	
17211	31	3	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 03:00:00	0	59	
17212	31	4	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 04:00:00	0	69	
17213	31	5	12	2012	7.575	NaN	NaN	2012-12-31	2012-12-31 05:00:00	0	64	
17214	31	6	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 06:00:00	0	64	
17215	31	7	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 07:00:00	0	64	
17216	31	8	12	2012	7.575	NaN	NaN	2012-12-31	2012-12-31 08:00:00	0	69	
17217	31	9	12	2012	10.605	NaN	NaN	2012-12-31	2012-12-31 09:00:00	0	64	
17218	31	10	12	2012	10.605	NaN	NaN	2012-12-31	2012-12-31 10:00:00	0	69	
17219	31	11	12	2012	11.365	NaN	NaN	2012-12-31	2012-12-31 11:00:00	0	60	
17220	31	12	12	2012	11.365	NaN	NaN	2012-12-31	2012-12-31 12:00:00	0	56	
17221	31	13	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 13:00:00	0	44	
17222	31	14	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 14:00:00	0	45	
17223	31	15	12	2012	14.395	NaN	NaN	2012-12-31	2012-12-31 15:00:00	0	45	
17224	31	16	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 16:00:00	0	48	
17225	31	17	12	2012	14.395	NaN	NaN	2012-12-31	2012-12-31 17:00:00	0	48	

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	holiday	humidity
17226	31	18	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 18:00:00	0	48
17227	31	19	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 19:00:00	0	60
17228	31	20	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 20:00:00	0	60
17229	31	21	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 21:00:00	0	60
17230	31	22	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 22:00:00	0	56
17231	31	23	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 23:00:00	0	65

17232 rows × 18 columns

In [33]:
```
#check the shape of training data and testing data
row_train = data_std.shape[0]
row_test = test.shape[0]
print('The training dataset has:',row_train,
     '; The testing dataset has:', row_test)
```

```
The training dataset has: 10739 ; The testing dataset has: 6493
```

4.2 Set up feature value

In [34]:
```
# Check the correlation
#First, drop the 'casual' and 'registered' column
Full_BikeDf = Full_Bike.drop(['casual','registered'],axis=1)
corrDf = Full_BikeDf.corr()
corrDf
```

Out[34]:

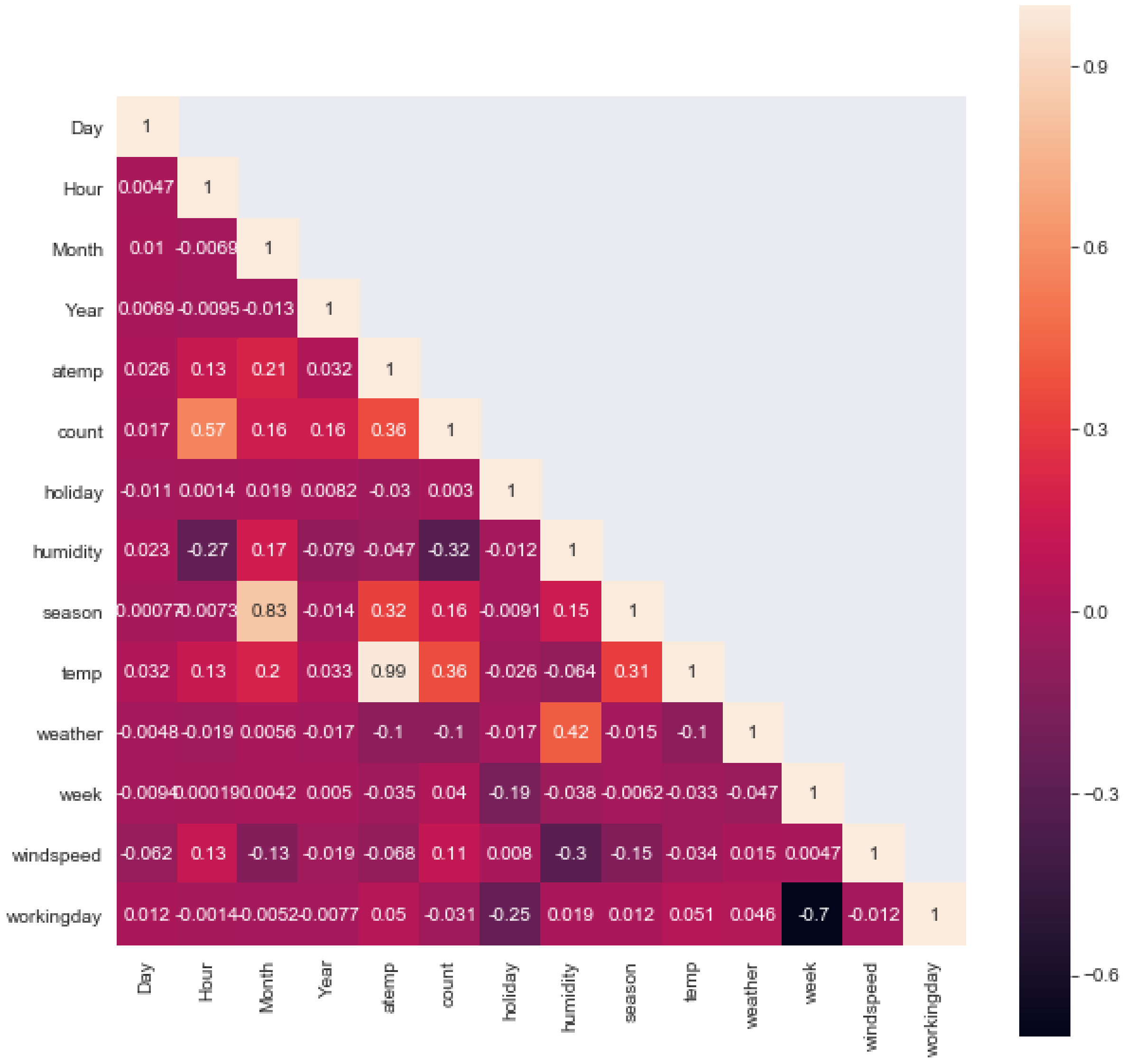
	Day	Hour	Month	Year	atemp	count	holiday	humidity
Day	1.000000	0.004739	0.010042	0.006944	0.026175	0.017273	-0.011003	0.023387
Hour	0.004739	1.000000	-0.006866	-0.009517	0.128743	0.566538	0.001449	-0.273240
Month	0.010042	-0.006866	1.000000	-0.013427	0.207730	0.160221	0.018968	0.165208
Year	0.006944	-0.009517	-0.013427	1.000000	0.031976	0.158753	0.008188	-0.078624
atemp	0.026175	0.128743	0.207730	0.031976	1.000000	0.359495	-0.029816	-0.046832
count	0.017273	0.566538	0.160221	0.158753	0.359495	1.000000	0.003002	-0.322558
holiday	-0.011003	0.001449	0.018968	0.008188	-0.029816	0.003002	1.000000	-0.011614
humidity	0.023387	-0.273240	0.165208	-0.078624	-0.046832	-0.322558	-0.011614	1.000000
season	-0.000771	-0.007309	0.829884	-0.014035	0.319993	0.156657	-0.009057	0.151517
temp	0.032067	0.132439	0.201269	0.033499	0.987888	0.363572	-0.026139	-0.064472

	Day	Hour	Month	Year	atemp	count	holiday	humidity	
weather	-0.004833	-0.018574	0.005555	-0.016977	-0.104424	-0.102830	-0.017491	0.417826	-
week	-0.009427	0.000192	0.004167	0.005041	-0.035210	0.039716	-0.190192	-0.037979	-
windspeed	-0.061655	0.127101	-0.132771	-0.018780	-0.067567	0.110735	0.007975	-0.302039	-
workingday	0.011760	-0.001450	-0.005163	-0.007658	0.050190	-0.031389	-0.252103	0.019332	

In [35]:
```python
# Compare the correlation score
corrDf['count'].sort_values(ascending=False)
```

Out[35]:
```
count         1.000000
Hour          0.566538
temp          0.363572
atemp         0.359495
Month         0.160221
Year          0.158753
season        0.156657
windspeed     0.110735
week          0.039716
Day           0.017273
holiday       0.003002
workingday   -0.031389
weather      -0.102830
humidity     -0.322558
Name: count, dtype: float64
```

In [36]:
```python
#plot the heatmap of correlation
corrDf
mask = np.array(corrDf)
mask[np.tril_indices_from(mask)]=False
fig = plt.figure(figsize=(10,10))
ax=sns.heatmap(corrDf,mask=mask,annot=True,square=True)
```



It can be seen that the influence of the characteristic value on the rental number is: time> temperature> humidity> year> month> season> weather> wind speed> day > whether it is a working day> whether it is a holiday.

It would transform strings in the predictors into binary values using one-hot encoding. These features are characterized using one-hot (get_dummies) transformation.

In [37]:
```
# one-hot transformation
Full_Bike_ = pd.get_dummies(data=Full_Bike, columns=['season','holiday','workingday'
train_ = Full_Bike_[pd.notnull(Full_Bike_['count'])].sort_values(by=['datetime'])
test_ = Full_Bike_[~pd.notnull(Full_Bike_['count'])].sort_values(by=['datetime'])
```

According to the data pattern analysis, it would decide to take the temp, humidity, windspeed, weather, season, year, month, week and hour as 8 features.

In [38]:
```
# Set up the feature value
from sklearn.preprocessing import StandardScaler
cols = ['temp','atemp','humidity','windspeed','Month','week','Hour','Day']
features = Full_Bike[cols]

scaler = StandardScaler().fit(features.values)
Full_Bike[cols] = scaler.transform(features.values)
```

In [39]:
```
#Check the column name of dataset
Full_Bike_
```

Out[39]:

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	humidity	...	season_3
0	1	5	1	2011	12.880	0.0	0.000000	2011-01-01	2011-01-01 05:00:00	75	...	0
1	1	10	1	2011	19.695	12.0	3.583519	2011-01-01	2011-01-01 10:00:00	76	...	0
2	1	11	1	2011	16.665	26.0	4.025352	2011-01-01	2011-01-01 11:00:00	81	...	0
3	1	12	1	2011	21.210	29.0	4.430817	2011-01-01	2011-01-01 12:00:00	77	...	0
4	1	13	1	2011	22.725	47.0	4.543295	2011-01-01	2011-01-01 13:00:00	72	...	0
5	1	14	1	2011	22.725	35.0	4.663439	2011-01-01	2011-01-01 14:00:00	72	...	0
6	1	15	1	2011	21.970	40.0	4.700480	2011-01-01	2011-01-01 15:00:00	77	...	0
7	1	16	1	2011	21.210	41.0	4.532599	2011-01-01	2011-01-01 16:00:00	82	...	0
8	1	17	1	2011	21.970	15.0	4.204693	2011-01-01	2011-01-01 17:00:00	82	...	0
9	1	18	1	2011	21.210	9.0	3.555348	2011-01-01	2011-01-01 18:00:00	88	...	0
10	1	19	1	2011	21.210	6.0	3.610918	2011-01-01	2011-01-01 19:00:00	88	...	0
11	1	20	1	2011	20.455	11.0	3.583519	2011-01-01	2011-01-01 20:00:00	87	...	0
12	1	21	1	2011	20.455	3.0	3.526361	2011-01-01	2011-01-01 21:00:00	87	...	0
13	1	22	1	2011	20.455	11.0	3.332205	2011-01-01	2011-01-01 22:00:00	94	...	0
14	1	23	1	2011	22.725	15.0	3.663562	2011-01-01	2011-01-01 23:00:00	88	...	0
15	2	0	1	2011	22.725	4.0	2.833213	2011-01-02	2011-01-02 00:00:00	88	...	0

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	humidity	...	season_3
16	2	1	1	2011	21.970	1.0	2.833213	2011-01-02	2011-01-02 01:00:00	94	...	C
17	2	2	1	2011	21.210	1.0	2.197225	2011-01-02	2011-01-02 02:00:00	100	...	C
18	2	3	1	2011	22.725	2.0	1.791759	2011-01-02	2011-01-02 03:00:00	94	...	C
19	2	4	1	2011	22.725	2.0	1.098612	2011-01-02	2011-01-02 04:00:00	94	...	C
20	2	6	1	2011	21.210	0.0	0.693147	2011-01-02	2011-01-02 06:00:00	77	...	C
21	2	7	1	2011	20.455	0.0	0.000000	2011-01-02	2011-01-02 07:00:00	76	...	C
22	2	8	1	2011	20.455	0.0	2.079442	2011-01-02	2011-01-02 08:00:00	71	...	C
23	2	9	1	2011	19.695	1.0	2.995732	2011-01-02	2011-01-02 09:00:00	76	...	C
24	2	10	1	2011	17.425	7.0	3.970292	2011-01-02	2011-01-02 10:00:00	81	...	C
25	2	11	1	2011	16.665	16.0	4.248495	2011-01-02	2011-01-02 11:00:00	71	...	C
26	2	12	1	2011	16.665	20.0	4.532599	2011-01-02	2011-01-02 12:00:00	66	...	C
27	2	13	1	2011	17.425	11.0	4.317488	2011-01-02	2011-01-02 13:00:00	66	...	C
28	2	14	1	2011	17.425	4.0	4.077537	2011-01-02	2011-01-02 14:00:00	76	...	C
29	2	15	1	2011	16.665	19.0	4.304065	2011-01-02	2011-01-02 15:00:00	81	...	C
...	...	...	...	...	...	...	...	...	...	...	...	..
17202	30	18	12	2012	10.605	NaN	NaN	2012-12-30	2012-12-30 18:00:00	44	...	C
17203	30	19	12	2012	18.180	NaN	NaN	2012-12-30	2012-12-30 19:00:00	61	...	C

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	humidity	...	season_3
17204	30	20	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 20:00:00	47	...	C
17205	30	21	12	2012	10.605	NaN	NaN	2012-12-30	2012-12-30 21:00:00	51	...	C
17206	30	22	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 22:00:00	55	...	C
17207	30	23	12	2012	9.850	NaN	NaN	2012-12-30	2012-12-30 23:00:00	51	...	C
17208	31	0	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 00:00:00	55	...	C
17209	31	1	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 01:00:00	55	...	C
17210	31	2	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 02:00:00	59	...	C
17211	31	3	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 03:00:00	59	...	C
17212	31	4	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 04:00:00	69	...	C
17213	31	5	12	2012	7.575	NaN	NaN	2012-12-31	2012-12-31 05:00:00	64	...	C
17214	31	6	12	2012	8.335	NaN	NaN	2012-12-31	2012-12-31 06:00:00	64	...	C
17215	31	7	12	2012	9.090	NaN	NaN	2012-12-31	2012-12-31 07:00:00	64	...	C
17216	31	8	12	2012	7.575	NaN	NaN	2012-12-31	2012-12-31 08:00:00	69	...	C
17217	31	9	12	2012	10.605	NaN	NaN	2012-12-31	2012-12-31 09:00:00	64	...	C
17218	31	10	12	2012	10.605	NaN	NaN	2012-12-31	2012-12-31 10:00:00	69	...	C
17219	31	11	12	2012	11.365	NaN	NaN	2012-12-31	2012-12-31 11:00:00	60	...	C

	Day	Hour	Month	Year	atemp	casual	count	date	datetime	humidity	...	season_3
17220	31	12	12	2012	11.365	NaN	NaN	2012-12-31	2012-12-31 12:00:00	56	...	0
17221	31	13	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 13:00:00	44	...	0
17222	31	14	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 14:00:00	45	...	0
17223	31	15	12	2012	14.395	NaN	NaN	2012-12-31	2012-12-31 15:00:00	45	...	0
17224	31	16	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 16:00:00	48	...	0
17225	31	17	12	2012	14.395	NaN	NaN	2012-12-31	2012-12-31 17:00:00	48	...	0
17226	31	18	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 18:00:00	48	...	0
17227	31	19	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 19:00:00	60	...	0
17228	31	20	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 20:00:00	60	...	0
17229	31	21	12	2012	12.880	NaN	NaN	2012-12-31	2012-12-31 21:00:00	60	...	0
17230	31	22	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 22:00:00	56	...	0
17231	31	23	12	2012	13.635	NaN	NaN	2012-12-31	2012-12-31 23:00:00	65	...	0

17232 rows × 26 columns

In [40]:
```python
# Set up the column name
cols = ['temp','atemp','humidity','windspeed','Month','week','Hour','Day','season_1'
        'workingday_0','weather_1','weather_2','weather_3']
```

5. Prediction Model Construction

5.1 Set Training Set and Testing Set

The training data and testing data should be splitted fairly under using train_test_split function in Python that the prediction model can have higher accuracy.

In [41]:
```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

In [42]:
```
# Split the training dataset and testing dataset
X = train_[cols]
y = train_['count']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_stat
```

5.2 Regression Model

First, it uses the advanced regression techniques for the prediction. Regression analysis is a method of predictive modelling technology used in forecasting, time series modelling and finding causal relationships between variables. In this part, it adopts linear regression, Lasso regression and random forest regression methods and calculate the model prediction accuracy using score function and comparing rmlse、rmse and r-squared score.

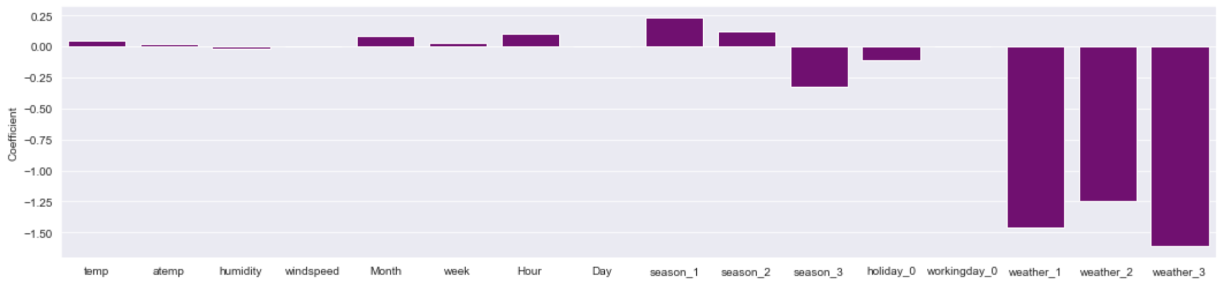
Linear Regression

In [43]:
```
lm = LinearRegression()
lm.fit(X_train, y_train)
print(lm.intercept_)
```

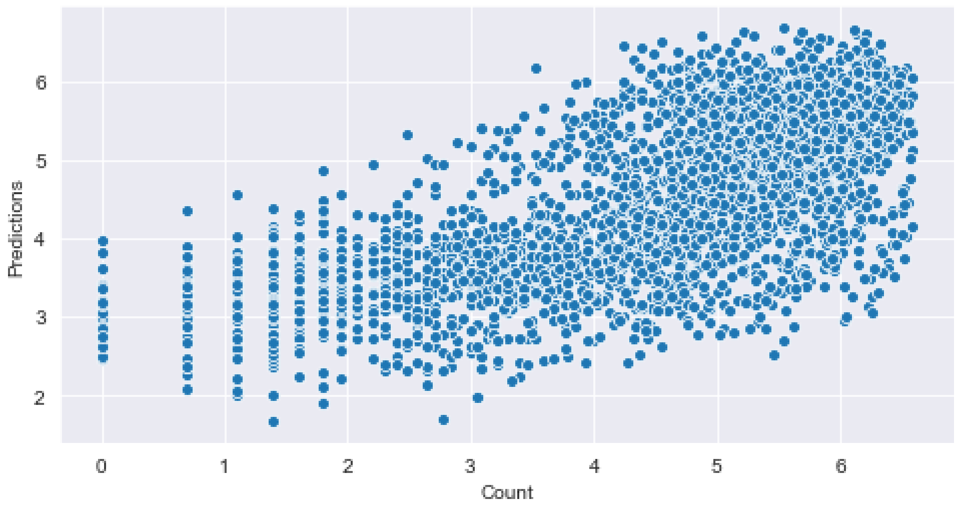
```
3.7798197990027917
```

In [44]:
```
plt.figure(figsize = (18,4))
coeff = pd.DataFrame(lm.coef_, index = X.columns, columns=['Coefficient'])
sns.barplot(x=coeff.index,y = 'Coefficient', data = coeff, color ='purple')
```

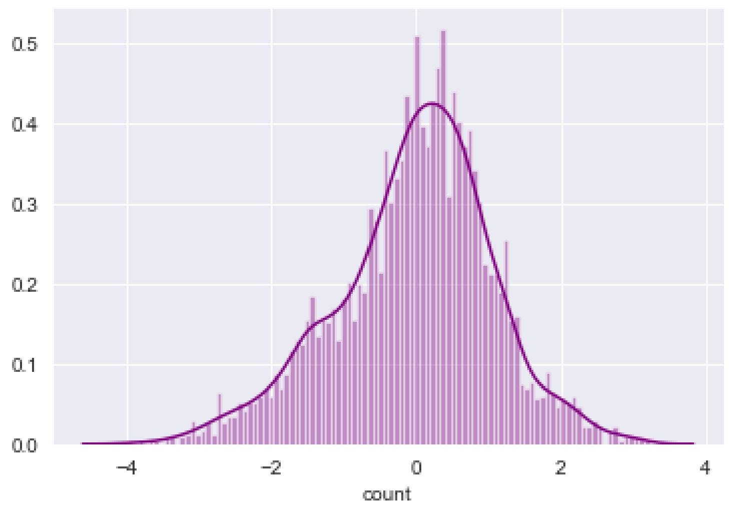
Out[44]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d598fb940>
```



In [45]:
```
plt.figure(figsize=(8,4))
pred = lm.predict(X_test)
sns.scatterplot(x=y_test, y=pred)
plt.xlabel('Count')
plt.ylabel('Predictions')
plt.show()
```



In [46]:
```python
sns.distplot((y_test-pred), bins=100, color='purple')
plt.show()
```



First, it uses the advanced regression techniques for the prediction. Regression analysis is a method of predictive modelling technology used in forecasting, time series modelling and finding causal relationships between variables. In this part, it adopts linear regression, Lasso regression and random forest regression methods and calculate the model prediction accuracy using score function.

Lasso Regression

In [47]:
```python
#Import Lasso regression model
from sklearn.linear_model import Lasso

temp_msle = {}
for i in np.logspace(-10,-1,20):
    lasso = Lasso(alpha = i, normalize = True, tol = 0.1)
    #fit lasso model
    lasso.fit(X_train, y_train)
    #make prediction
    pred = lasso.predict(X_test)
    #calculate msle
    msle= np.sqrt(metrics.mean_squared_log_error(np.exp(y_test), np.exp(pred)))
    temp_msle[i]=msle
```

The main idea of LASSO is to construct a first-order punishment function to obtain a refined model, and to finally select the coefficients of some variables as 0 for feature screening.

Random Forest

In [48]:
```python
#Import random forest regressor model
from sklearn.ensemble import RandomForestRegressor
#Fit the model
RF = RandomForestRegressor(n_estimators=1000, random_state=42)
RF.fit(X_train, y_train)
```

Out[48]:
```
RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                      max_depth=None, max_features='auto', max_leaf_nodes=None,
                      max_samples=None, min_impurity_decrease=0.0,
                      min_impurity_split=None, min_samples_leaf=1,
                      min_samples_split=2, min_weight_fraction_leaf=0.0,
                      n_estimators=1000, n_jobs=None, oob_score=False,
                      random_state=42, verbose=0, warm_start=False)
```

In [49]:
```python
#calculate the score of the random forest model
RF.score(X_test, y_test)
```

Out[49]:
```
0.9193167871884363
```

In [50]:
```
#make predictions
predictions = RF.predict(X_test)
pred = pd.Series(predictions, index = y_test.index)
```

The random forest regressor prediction has the highest score and prediction accuracy. The random forest regression is a neighbourhood-based ideal model. It has the highest accuracy score in this case. However, it may have disadvantages in making accurate predictions for the time ranges that are outside the training data.

5.3 Comparison of regression methods

In [51]:
```
from sklearn.svm import SVR
from sklearn.model_selection import KFold
from sklearn.metrics import mean_squared_log_error, mean_squared_error, r2_score,mea
```

In [52]:
```
models = [RandomForestRegressor(),SVR(),LinearRegression(), Lasso()]
modelna = ['RandomForestRegressor','SVR','LinearRegression','Lasso']
rmsle=[]
d={}
rmse=[]
e={}
r2score=[]
f={}
for model in range(len(models)):
    clf=models[model]
    clf.fit(X_train,y_train)
    pred=clf.predict(X_test)
    rmsle.append(np.sqrt(mean_squared_log_error(y_test,pred)))
    rmse.append(np.sqrt(mean_squared_error(y_test,pred)))
    r2score.append(r2_score(y_test,pred))

d={'Modelling Alogo':modelna,'RMSLE':rmsle}
d
```

Out[52]:
```
{'Modelling Alogo': ['RandomForestRegressor',
  'SVR',
  'LinearRegression',
  'Lasso'],
 'RMSLE': [0.12877524610259763,
  0.26578884779700307,
  0.2749054126476759,
  0.2889874724674801]}
```

In [53]:
```
e={'Modelling':modelna,'RMSE':rmse}
e
```

Out[53]:
```
{'Modelling': ['RandomForestRegressor', 'SVR', 'LinearRegression', 'Lasso'],
 'RMSE': [0.4222163993351084,
  0.9840728969914119,
  1.0813945053274945,
  1.1205801243907545]}
```

In [54]:
```
f={'Modelling':modelna,'R-squared score':r2score}
f
```

Out[54]:
```
{'Modelling': ['RandomForestRegressor', 'SVR', 'LinearRegression', 'Lasso'],
 'R-squared score': [0.9182325163441553,
  0.5558138847801415,
  0.46361236918372906,
  0.4240347686795062]}
```

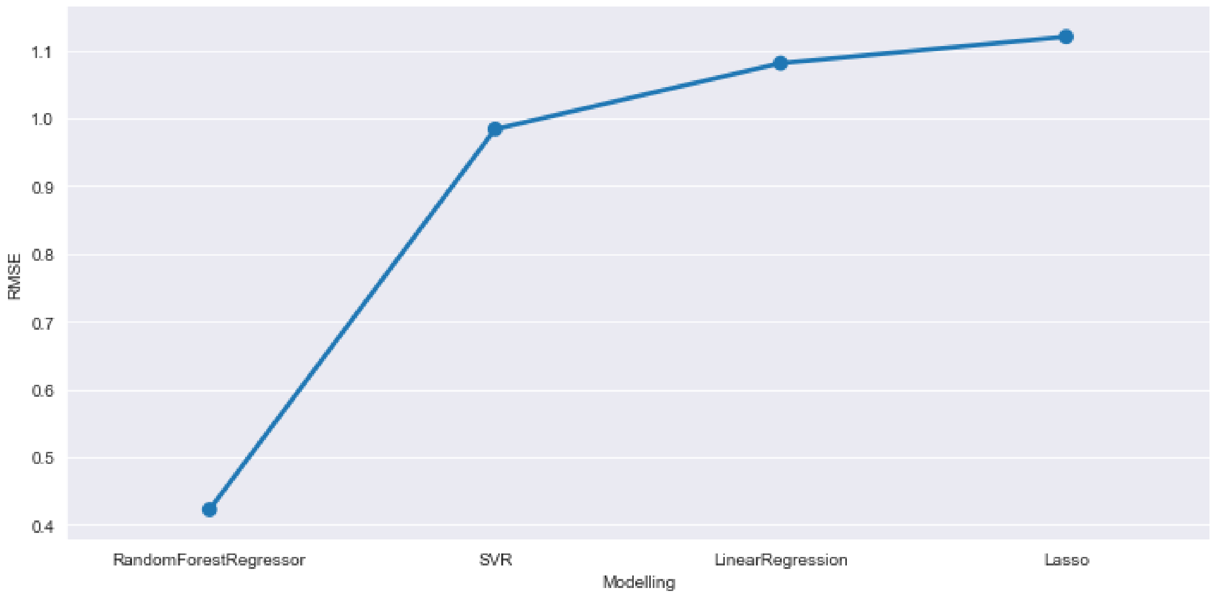
In [55]:
```
rmse_frame=pd.DataFrame(e)
rmse_frame
```

Out[55]:

	Modelling	RMSE
0	RandomForestRegressor	0.422216
1	SVR	0.984073
2	LinearRegression	1.081395
3	Lasso	1.120580

In [56]:
```
sns.factorplot(x='Modelling', y='RMSE',data=rmse_frame,kind='point',size=5,aspect=2)
```

Out[56]:
```
<seaborn.axisgrid.FacetGrid at 0x24d5a0c4860>
```



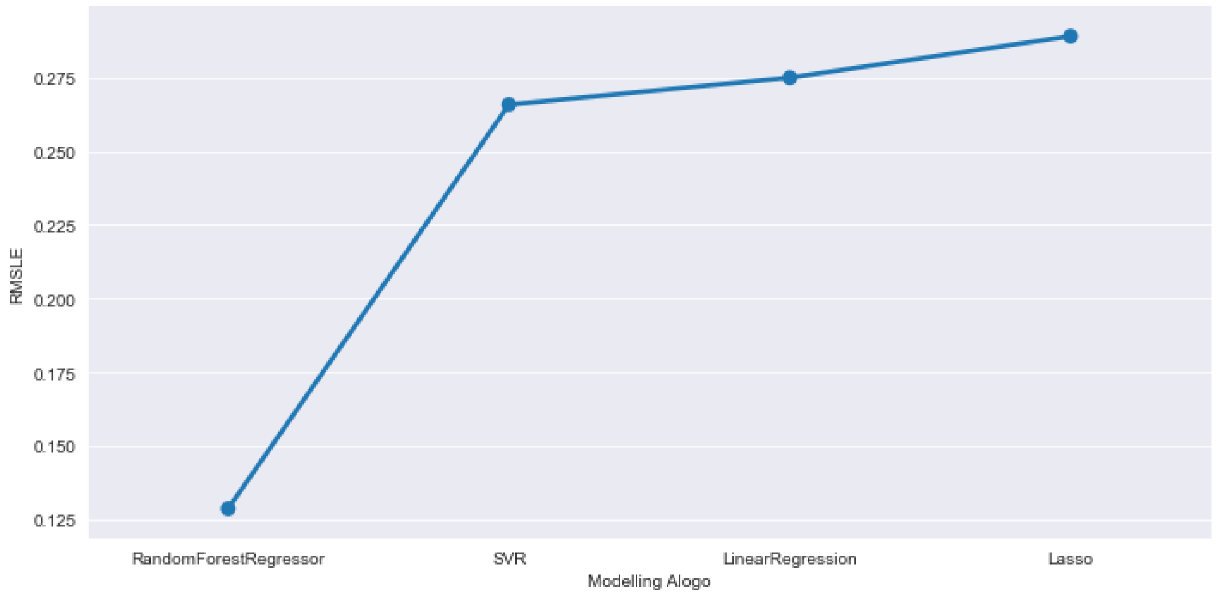
In [57]:
```
rmsle_frame=pd.DataFrame(d)
rmsle_frame
```

Out[57]:

	Modelling Alogo	RMSLE
0	RandomForestRegressor	0.128775
1	SVR	0.265789
2	LinearRegression	0.274905
3	Lasso	0.288987

In [58]:
```
sns.factorplot(x='Modelling Alogo', y='RMSLE',data=rmsle_frame,kind='point',size=5,a
```

Out[58]:
```
<seaborn.axisgrid.FacetGrid at 0x24d5a09d710>
```



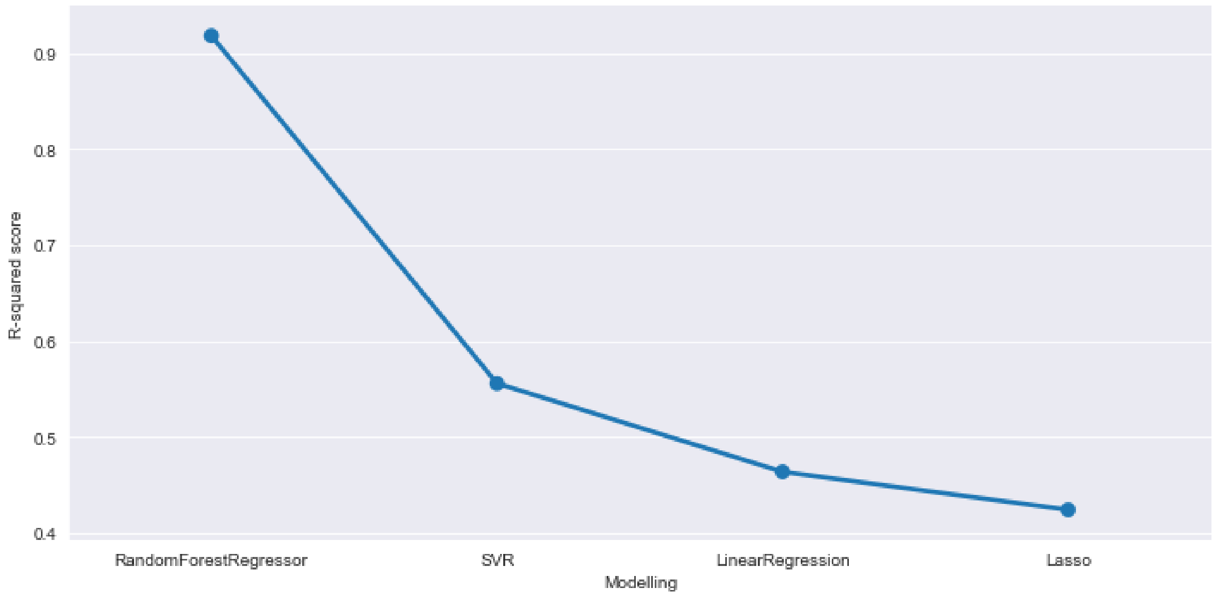
In [59]:
```
r2score_frame=pd.DataFrame(f)
r2score_frame
```

Out[59]:

	Modelling	R-squared score
0	RandomForestRegressor	0.918233
1	SVR	0.555814
2	LinearRegression	0.463612
3	Lasso	0.424035

In [60]:
```
sns.factorplot(x='Modelling', y='R-squared score',data=r2score_frame,kind='point',si
```

Out[60]:
```
<seaborn.axisgrid.FacetGrid at 0x24d5a11a550>
```



It has made the table of the Root Mean Squared Error(RMSE),Root Mean Squared Logarithmic Error(RMSLE) and R-squared results of each regression model. It can be found that the random forest regressor has the lowest score of RMSLE and RMSE among the four regression models. For RMSLE, it takes the log of the predictions and actual values and both RMSE and RMSLE are usually used in evaluation of supervised learning.Theproblem with RMSLE is that it penalizes the models that actually have an unbiased estimate. In addition, it has compared the r-squared value that a higher value represents higher correlation, thus means higher accuracy of the model prediction result. Among four models, random forest has the highest score higher than 0.9 while the score of linear regression and lasso is lower than 0.5 that means a bad model for prediction.

5.4 Advanced Classification Model

In addition, it also makes use of advanced classification techniques for predicting. KNN, Decision Tree, Random Forest and SVM are run in this part. In this case, as the log transformed type 'float64' cannot be fitted into the classification model, the original training data is used.KNN is to choose the weight neighbours and the SVM is to find a linear division in a high-dimension space. The decision tree is to find out the classification. The random forest also creates many decision trees for each subset and can handle numerical and categorical data.

In [61]:
```
# check data type
Full_Bike_.describe(include=[np.number])
```

Out[61]:

	Day	Hour	Month	Year	atemp	casual	
count	17232.000000	17232.000000	17232.000000	17232.000000	17232.000000	10739.000000	10739.0
mean	15.726149	11.507950	6.527855	2011.498317	23.723688	35.220039	4.5
std	8.800706	6.924261	3.446353	0.500012	8.587677	49.546882	1.4
min	1.000000	0.000000	1.000000	2011.000000	0.000000	0.000000	0.0
25%	8.000000	6.000000	4.000000	2011.000000	16.665000	4.000000	3.7
50%	16.000000	11.500000	7.000000	2011.000000	24.240000	16.000000	4.9
75%	23.000000	18.000000	10.000000	2012.000000	31.060000	47.000000	5.6
max	31.000000	23.000000	12.000000	2012.000000	50.000000	367.000000	6.5

8 rows × 24 columns

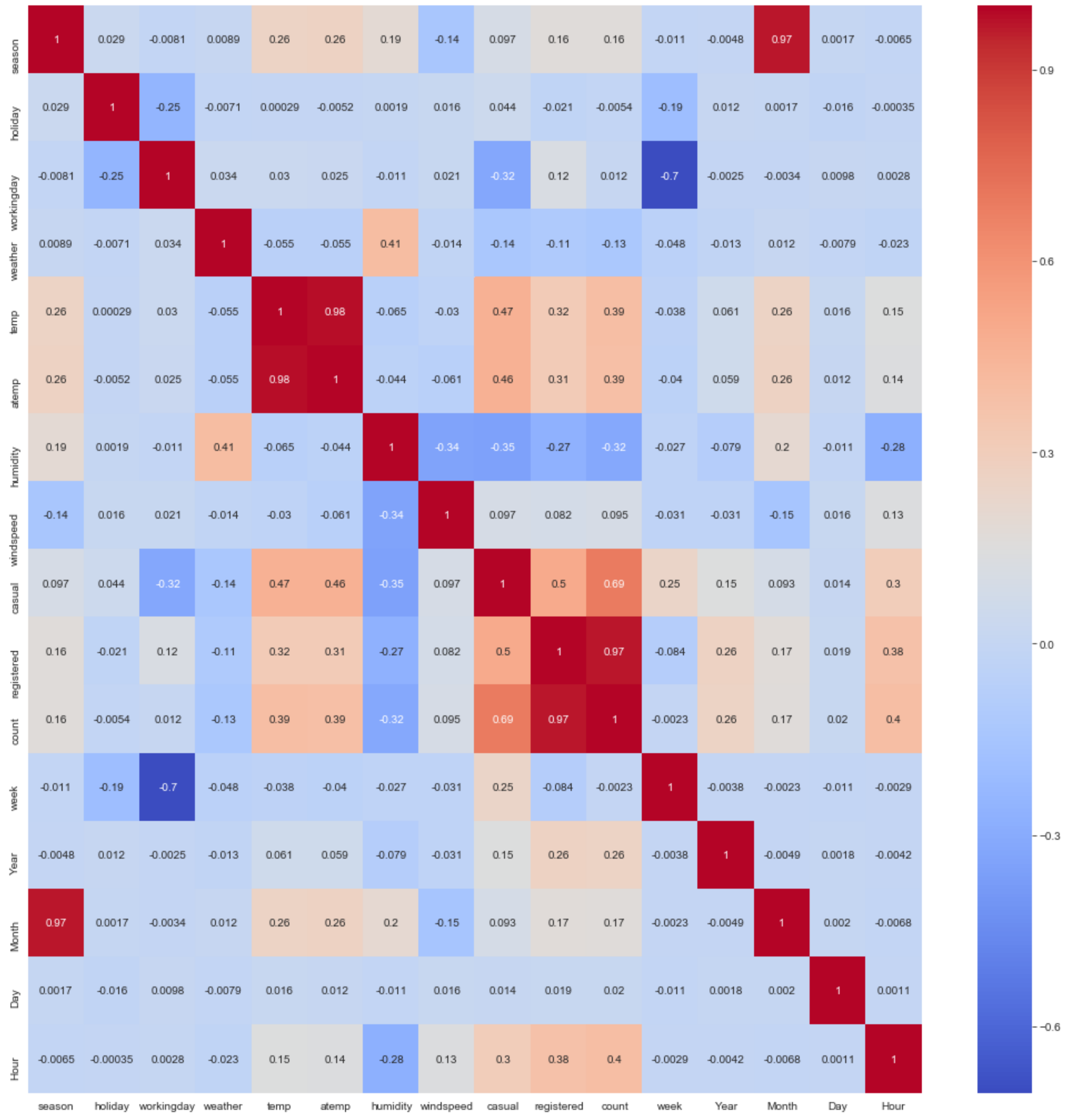
In [62]:
```
# check the type of the data, 'O' means Object type
Full_Bike_.describe(include=['O'])
```

Out[62]:

	date	datetime
count	17232	17232
unique	731	17232
top	2012-04-20	2011-01-21 10:00:00
freq	24	1

In [63]:
```
# Plot the heatmap of the data and check the correlation
plt.figure(figsize = (18,18))
sns.heatmap(data.corr(),cmap='coolwarm', annot = True)
```

Out[63]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x24d5a17d5c0>
```



The heatmap could define that the humidity, temp and atemp have greater impact on the count and the coefficient of temp and atemp are close so it would use temp. However, the workingday, weekend, and holiday has little impact on the count.

In [64]:
```
#select fatures that would be used in the classification algorithm
select_features = data[['temp','casual','registered','Hour','humidity']]
select_features.head()
```

Out[64]:

	temp	casual	registered	Hour	humidity
0	9.84	0	1	5	75
1	15.58	12	24	10	76
2	14.76	26	30	11	81
3	17.22	29	55	12	77
4	18.86	47	47	13	72

In [65]:
```
scaler = StandardScaler() # initialise a scaler object to run on a dataframe
scaler.fit(select_features)
scaled_features = scaler.transform(select_features)
```

In [66]:
```
from sklearn.feature_extraction import DictVectorizer
data_dict = data.to_dict('record')
print(data_dict[1])
```

```
{'datetime': '2011-01-01 10:00:00', 'season': 1, 'holiday': 0, 'workingday': 0, 'wea
ther': 1, 'temp': 15.58, 'atemp': 19.695, 'humidity': 76, 'windspeed': 16, 'casual':
12, 'registered': 24, 'count': 36, 'date': '2011-01-01', 'week': 6, 'Year': 2011, 'M
onth': 1, 'Day': 1, 'Hour': 10}
```

In [67]:
```
# create the DicVectorizer object
vec = DictVectorizer()
data_mat=vec.fit_transform(data_dict)
print(vec.feature_names_[0:5])
```

```
['Day', 'Hour', 'Month', 'Year', 'atemp']
```

In [68]:
```
# split the training data and testing data for classification algorithm
from sklearn.model_selection import train_test_split
train_X, test_X, train_y, test_y = train_test_split(select_features, data['count'],
```

In [69]:
```
#Check out the training data shape and testing data shape
print(train_X.shape)
print(test_X.shape)
```

```
(7620, 5)
(3266, 5)
```

In [70]:
```
#Import KNN model
from sklearn.neighbors import KNeighborsClassifier
```

In [71]:
```
#calculate the running time of KNN model
import time
start_time = time.time()
#In KNN, it set the neighbors number as 20
knn = KNeighborsClassifier(n_neighbors=20)
```

In [72]:
```
knn.fit(train_X, train_y)
end_time = time.time()

print(end_time - start_time)
```

```
0.024935245513916016
```

In [73]:
```
# Make predictioins
pred = knn.predict(test_X)
```

In [74]:
```
from sklearn.metrics import mean_squared_error
from math import sqrt

rms = sqrt(mean_squared_error(test_y, pred))
rms
```

Out[74]:
```
12.359361929796224
```

In [75]:
```
from sklearn import metrics

print('Table 1.Classification Report KNN:\n', metrics.classification_report(test_y,p
```

```
Table 1.Classification Report KNN:
               precision    recall  f1-score   support

           1       0.26      0.26      0.26        34
           2       0.25      0.27      0.26        44
           3       0.20      0.28      0.23        47
```

```
          4       0.13      0.28      0.18        36
          5       0.17      0.29      0.21        52
          6       0.25      0.16      0.20        56
          7       0.17      0.15      0.16        41
          8       0.10      0.08      0.09        25
          9       0.05      0.04      0.05        23
         10       0.26      0.25      0.25        28
         11       0.15      0.08      0.11        36
         12       0.11      0.09      0.10        22
         13       0.00      0.00      0.00        20
         14       0.10      0.15      0.12        13
         15       0.06      0.07      0.06        15
         16       0.21      0.19      0.20        27
         17       0.25      0.05      0.08        20
         18       0.00      0.00      0.00         9
         19       0.00      0.00      0.00         8
         20       0.15      0.22      0.18        18
         21       0.12      0.12      0.12        16
         22       0.12      0.20      0.15         5
         23       0.33      0.05      0.08        22
         24       0.12      0.18      0.14        11
         25       0.07      0.08      0.08        12
         26       0.05      0.08      0.06        13
         27       0.00      0.00      0.00        11
         28       0.14      0.21      0.17        14
         29       0.12      0.12      0.12         8
         30       0.00      0.00      0.00        13
         31       0.00      0.00      0.00         9
         32       0.00      0.00      0.00        12
         33       0.12      0.20      0.15        10
         34       0.00      0.00      0.00        11
         35       0.33      0.13      0.19        15
         36       0.08      0.20      0.11         5
         37       0.50      0.07      0.12        14
         38       0.00      0.00      0.00         8
         39       0.17      0.13      0.15        15
         40       0.06      0.17      0.09         6
         41       0.13      0.20      0.16        10
         42       0.00      0.00      0.00         5
         43       0.00      0.00      0.00        10
         44       0.00      0.00      0.00        11
         45       0.10      0.08      0.09        12
         46       0.00      0.00      0.00         7
         47       0.00      0.00      0.00         8
         48       0.17      0.09      0.12        11
         49       0.12      0.11      0.12         9
         50       0.09      0.17      0.12         6
         51       0.10      0.12      0.11         8
         52       0.09      0.07      0.08        14
         53       0.00      0.00      0.00        12
         54       0.11      0.17      0.13         6
         55       0.00      0.00      0.00         8
         56       0.10      0.17      0.12         6
         57       0.00      0.00      0.00        14
         58       0.00      0.00      0.00         6
         59       0.00      0.00      0.00         8
         60       0.00      0.00      0.00         5
         61       0.00      0.00      0.00         4
         62       0.00      0.00      0.00         9
         63       0.00      0.00      0.00         5
         64       0.22      0.13      0.17        15
         65       0.00      0.00      0.00         8
         66       0.00      0.00      0.00         7
         67       0.00      0.00      0.00         2
         68       0.00      0.00      0.00         7
         69       0.12      0.10      0.11        10
         70       0.00      0.00      0.00         7
         71       0.38      0.33      0.35         9
         72       0.00      0.00      0.00        10
```

```
          73       0.25      0.11      0.15         9
          74       0.00      0.00      0.00         6
          75       0.10      0.11      0.11         9
          76       0.00      0.00      0.00         5
          77       0.00      0.00      0.00         5
          78       0.00      0.00      0.00        14
          79       0.00      0.00      0.00         9
          80       0.00      0.00      0.00         4
          81       0.50      0.14      0.22         7
          82       0.00      0.00      0.00         6
          83       0.14      0.22      0.17         9
          84       0.05      0.12      0.07         8
          85       0.00      0.00      0.00         3
          86       0.04      0.25      0.07         4
          87       0.00      0.00      0.00        10
          88       0.50      0.09      0.15        11
          89       0.00      0.00      0.00         9
          90       0.00      0.00      0.00        10
          91       0.00      0.00      0.00         6
          92       0.25      0.10      0.14        10
          93       0.00      0.00      0.00         9
          94       0.00      0.00      0.00        10
          95       0.00      0.00      0.00        10
          96       0.00      0.00      0.00         8
          97       0.00      0.00      0.00         5
          98       0.00      0.00      0.00         6
          99       0.07      0.25      0.11         4
         100       0.25      0.50      0.33         2
         101       0.00      0.00      0.00         4
         102       0.11      0.14      0.12         7
         103       0.00      0.00      0.00         8
         104       0.00      0.00      0.00         6
         105       0.00      0.00      0.00         9
         106       0.40      0.17      0.24        12
         107       0.00      0.00      0.00         8
         108       0.11      0.09      0.10        11
         109       0.05      0.14      0.08         7
         110       0.00      0.00      0.00         7
         111       0.00      0.00      0.00         4
         112       0.00      0.00      0.00         8
         113       0.00      0.00      0.00         9
         114       0.00      0.00      0.00        11
         115       0.00      0.00      0.00         5
         116       0.20      0.25      0.22         4
         117       0.00      0.00      0.00         5
         118       0.07      0.09      0.08        11
         119       0.00      0.00      0.00         8
         120       0.00      0.00      0.00         9
         121       0.00      0.00      0.00         5
         122       0.00      0.00      0.00         8
         123       0.14      0.50      0.22         4
         124       0.20      0.33      0.25        12
         125       0.00      0.00      0.00         7
         126       0.00      0.00      0.00        10
         127       0.00      0.00      0.00         7
         128       0.00      0.00      0.00         6
         129       0.00      0.00      0.00         5
         130       0.00      0.00      0.00         8
         131       0.00      0.00      0.00         3
         132       0.00      0.00      0.00         7
         133       0.00      0.00      0.00         7
         134       0.00      0.00      0.00         9
         135       0.00      0.00      0.00         8
         136       0.00      0.00      0.00         9
         137       0.00      0.00      0.00         6
         138       0.00      0.00      0.00         8
         139       0.00      0.00      0.00         9
         140       0.00      0.00      0.00        10
         141       0.00      0.00      0.00         9
```

```
142        0.25      0.33      0.29         3
143        0.00      0.00      0.00         5
144        0.00      0.00      0.00         5
145        0.00      0.00      0.00         5
146        0.00      0.00      0.00         6
147        0.00      0.00      0.00         9
148        0.14      0.12      0.13         8
149        0.00      0.00      0.00         4
150        0.00      0.00      0.00         8
151        0.10      0.20      0.13         5
152        0.00      0.00      0.00        10
153        0.10      0.18      0.13        11
154        0.12      0.18      0.14        11
155        0.00      0.00      0.00         9
156        0.00      0.00      0.00         6
157        0.00      0.00      0.00         8
158        0.00      0.00      0.00         5
159        0.33      0.12      0.18         8
160        0.00      0.00      0.00         6
161        0.00      0.00      0.00         6
162        0.00      0.00      0.00         5
163        0.00      0.00      0.00         6
164        0.00      0.00      0.00         5
165        0.00      0.00      0.00         9
166        0.00      0.00      0.00         5
167        0.06      0.11      0.08         9
168        0.00      0.00      0.00         7
169        0.00      0.00      0.00         6
170        0.17      0.12      0.14         8
171        0.00      0.00      0.00         9
172        0.07      0.08      0.08        12
173        0.00      0.00      0.00         6
174        0.00      0.00      0.00         5
175        0.00      0.00      0.00         9
176        0.09      0.25      0.13         4
177        0.00      0.00      0.00         8
178        0.00      0.00      0.00        10
179        0.00      0.00      0.00        10
180        0.11      0.11      0.11         9
181        0.25      0.25      0.25        12
182        0.07      0.17      0.10         6
183        0.00      0.00      0.00         7
184        0.20      0.14      0.17         7
185        0.00      0.00      0.00         8
186        0.00      0.00      0.00         6
187        0.00      0.00      0.00         6
188        0.00      0.00      0.00         4
189        0.00      0.00      0.00         6
190        0.00      0.00      0.00        10
191        0.00      0.00      0.00         6
192        0.00      0.00      0.00         6
193        0.00      0.00      0.00         5
194        0.00      0.00      0.00         3
195        0.00      0.00      0.00        13
196        0.17      0.12      0.14         8
197        0.00      0.00      0.00         4
198        0.00      0.00      0.00         4
199        0.00      0.00      0.00         2
200        0.00      0.00      0.00         5
201        0.50      0.10      0.17        10
202        0.00      0.00      0.00         5
203        0.18      0.25      0.21         8
204        0.09      0.20      0.13         5
205        0.05      0.12      0.07         8
206        0.11      0.12      0.12         8
207        0.25      0.60      0.35         5
208        0.00      0.00      0.00         4
209        0.00      0.00      0.00         1
210        0.00      0.00      0.00         6
```

```
         211       0.20      0.12      0.15         8
         212       0.00      0.00      0.00         3
         213       0.00      0.00      0.00         9
         214       0.00      0.00      0.00        11
         215       0.00      0.00      0.00         6
         216       0.11      0.17      0.13         6
         217       0.00      0.00      0.00        10
         218       0.00      0.00      0.00         9
         219       0.00      0.00      0.00        11
         220       0.00      0.00      0.00        10
         221       0.00      0.00      0.00         3
         222       0.17      0.10      0.12        10
         223       0.20      0.20      0.20         5
         224       0.00      0.00      0.00        10
         225       0.00      0.00      0.00         7
         226       0.00      0.00      0.00         5
         227       0.00      0.00      0.00         6
         228       0.09      0.14      0.11         7
         229       0.00      0.00      0.00         4
         230       0.00      0.00      0.00         9
         231       0.00      0.00      0.00         3
         232       0.00      0.00      0.00         6
         233       0.00      0.00      0.00         9
         234       0.00      0.00      0.00         2
         235       0.00      0.00      0.00         6
         236       0.00      0.00      0.00         4
         237       0.10      0.20      0.13         5
         238       0.00      0.00      0.00         5
         239       0.00      0.00      0.00         6
         240       0.00      0.00      0.00         2
         241       0.33      0.17      0.22         6
         242       0.00      0.00      0.00         1
         243       0.00      0.00      0.00         7
         244       0.25      0.50      0.33         4
         245       0.00      0.00      0.00         5
         246       0.00      0.00      0.00         3
         247       0.50      0.20      0.29         5
         248       0.00      0.00      0.00         9
         249       0.00      0.00      0.00         5
         250       0.00      0.00      0.00         4
         251       0.00      0.00      0.00         3
         252       0.00      0.00      0.00         3
         253       0.00      0.00      0.00         4
         254       0.00      0.00      0.00         4
         255       0.00      0.00      0.00         3
         256       0.08      0.14      0.11         7
         257       0.00      0.00      0.00         4
         258       0.00      0.00      0.00         9
         259       0.00      0.00      0.00         5
         260       0.00      0.00      0.00         8
         261       0.00      0.00      0.00         0
         262       0.00      0.00      0.00         4
         263       0.00      0.00      0.00         5
         264       0.00      0.00      0.00         7
         265       0.00      0.00      0.00         2
         266       0.00      0.00      0.00         5
         267       0.00      0.00      0.00         5
         268       0.00      0.00      0.00         8
         269       0.00      0.00      0.00         5
         270       0.00      0.00      0.00         3
         271       0.00      0.00      0.00         3
         272       0.14      0.12      0.13         8
         273       0.00      0.00      0.00         5
         274       0.00      0.00      0.00         6
         275       0.00      0.00      0.00         2
         276       0.00      0.00      0.00         5
         277       0.00      0.00      0.00         6
         278       0.00      0.00      0.00         6
         279       0.00      0.00      0.00         2
         280       0.08      0.17      0.11         6
         281       0.00      0.00      0.00         8
         282       0.00      0.00      0.00         6
         283       0.00      0.00      0.00         5
         284       0.17      1.00      0.29         1
         285       0.00      0.00      0.00         3
         286       0.07      0.25      0.11         4
         287       0.00      0.00      0.00         4
         288       0.00      0.00      0.00         5
         289       0.00      0.00      0.00         4
         290       0.00      0.00      0.00         3
         291       0.00      0.00      0.00         6
         292       0.25      0.12      0.17         8
         293       0.00      0.00      0.00         2
         294       0.00      0.00      0.00         8
         295       0.00      0.00      0.00         0
         296       0.00      0.00      0.00         3
         297       0.00      0.00      0.00         6
         298       0.00      0.00      0.00         2
         299       0.00      0.00      0.00         4
         300       0.00      0.00      0.00         5
         301       0.00      0.00      0.00         1
         302       0.00      0.00      0.00         4
         303       0.00      0.00      0.00         4
         304       0.00      0.00      0.00         6
         305       0.00      0.00      0.00         2
         306       0.00      0.00      0.00         2
         307       0.00      0.00      0.00         1
         308       0.25      0.50      0.33         4
         309       0.00      0.00      0.00         0
         310       0.00      0.00      0.00         2
         311       0.00      0.00      0.00         2
         312       0.00      0.00      0.00         5
         313       0.17      0.25      0.20         4
         314       0.00      0.00      0.00         4
         315       0.00      0.00      0.00         4
         316       0.00      0.00      0.00         3
         317       0.00      0.00      0.00         3
         318       0.00      0.00      0.00         2
         319       0.33      0.50      0.40         4
         320       0.00      0.00      0.00         3
         321       0.00      0.00      0.00         1
         322       0.00      0.00      0.00         3
         323       0.00      0.00      0.00         4
         324       0.00      0.00      0.00         1
         325       0.00      0.00      0.00         4
         326       0.00      0.00      0.00         2
         327       0.00      0.00      0.00         3
         328       0.00      0.00      0.00         8
         329       0.00      0.00      0.00         2
         330       0.00      0.00      0.00         2
         331       0.00      0.00      0.00         3
         332       0.00      0.00      0.00         6
         334       0.00      0.00      0.00         7
         335       0.00      0.00      0.00         6
         336       0.00      0.00      0.00         2
         337       0.00      0.00      0.00         3
         338       0.00      0.00      0.00         3
         339       0.00      0.00      0.00         1
         340       0.00      0.00      0.00         3
         341       0.00      0.00      0.00         3
         342       0.00      0.00      0.00         5
         343       0.17      0.14      0.15         7
         344       0.00      0.00      0.00         1
         345       0.00      0.00      0.00         1
         346       0.00      0.00      0.00         0
         347       0.00      0.00      0.00         6
         348       0.00      0.00      0.00         3
         349       0.00      0.00      0.00         4
         350       0.00      0.00      0.00         6
         351       0.00      0.00      0.00         4
         352       0.00      0.00      0.00         2
         353       0.00      0.00      0.00         4
         354       0.00      0.00      0.00         3
         355       0.00      0.00      0.00         2
         356       0.00      0.00      0.00         7
         357       0.00      0.00      0.00         4
         358       0.00      0.00      0.00         3
         359       0.00      0.00      0.00         4
         360       0.00      0.00      0.00         2
         361       0.00      0.00      0.00         3
         362       1.00      0.25      0.40         4
         363       0.00      0.00      0.00         6
         364       0.00      0.00      0.00         1
         365       0.00      0.00      0.00         6
         366       0.00      0.00      0.00         1
         367       0.00      0.00      0.00         4
         369       0.00      0.00      0.00         2
         370       0.00      0.00      0.00         4
         371       0.00      0.00      0.00         2
         372       0.00      0.00      0.00         6
         373       0.00      0.00      0.00         2
         374       0.00      0.00      0.00         7
         375       0.00      0.00      0.00         3
         376       0.00      0.00      0.00         1
         377       0.00      0.00      0.00         4
         378       0.00      0.00      0.00         1
         379       0.00      0.00      0.00         2
         380       0.00      0.00      0.00         1
         381       0.33      0.33      0.33         3
         382       0.00      0.00      0.00         2
         383       0.00      0.00      0.00         1
         384       0.00      0.00      0.00         2
         385       0.11      0.20      0.14         5
         386       0.00      0.00      0.00         1
         387       0.00      0.00      0.00         4
         388       0.00      0.00      0.00         1
         389       0.00      0.00      0.00         4
         390       0.00      0.00      0.00         4
         392       0.00      0.00      0.00         1
         393       0.00      0.00      0.00         3
         394       0.00      0.00      0.00         1
         395       0.00      0.00      0.00         2
         396       0.00      0.00      0.00         5
         397       0.00      0.00      0.00         2
         398       0.00      0.00      0.00         4
         399       0.00      0.00      0.00         0
         400       0.00      0.00      0.00         2
         401       0.00      0.00      0.00         3
         402       0.00      0.00      0.00         1
         403       0.00      0.00      0.00         2
         404       0.22      0.50      0.31         4
         405       0.00      0.00      0.00         4
         406       0.00      0.00      0.00         1
         407       0.00      0.00      0.00         0
         408       0.00      0.00      0.00         2
         409       0.00      0.00      0.00         1
         410       0.00      0.00      0.00         1
         411       0.00      0.00      0.00         1
         412       0.00      0.00      0.00         0
         413       0.00      0.00      0.00         4
         414       0.00      0.00      0.00         1
         415       0.00      0.00      0.00         1
         416       0.00      0.00      0.00         2
         417       0.00      0.00      0.00         1
         418       0.00      0.00      0.00         0
         419       0.00      0.00      0.00         5
         420       0.00      0.00      0.00         3
         421       0.00      0.00      0.00         3
         422       0.00      0.00      0.00         1
         423       0.00      0.00      0.00         2
         425       0.00      0.00      0.00         3
         426       0.00      0.00      0.00         1
         427       0.00      0.00      0.00         0
         428       0.00      0.00      0.00         5
         429       0.00      0.00      0.00         1
         430       0.50      0.25      0.33         4
         431       0.00      0.00      0.00         1
         432       0.00      0.00      0.00         2
         433       0.00      0.00      0.00         2
         434       0.00      0.00      0.00         1
         435       0.00      0.00      0.00         3
         436       0.00      0.00      0.00         2
         437       0.00      0.00      0.00         0
         439       0.00      0.00      0.00         1
         441       0.00      0.00      0.00         1
         442       0.00      0.00      0.00         1
         443       0.00      0.00      0.00         2
         444       0.00      0.00      0.00         1
         445       0.00      0.00      0.00         3
         446       0.50      0.25      0.33         4
         447       0.00      0.00      0.00         4
         448       0.00      0.00      0.00         1
         449       0.00      0.00      0.00         1
         450       0.00      0.00      0.00         1
         451       0.33      0.33      0.33         3
         452       0.00      0.00      0.00         3
         453       0.00      0.00      0.00         4
         454       0.00      0.00      0.00         3
         455       0.00      0.00      0.00         4
         456       0.00      0.00      0.00         4
         457       0.00      0.00      0.00         2
         458       0.00      0.00      0.00         1
         459       0.00      0.00      0.00         2
         460       0.00      0.00      0.00         2
         461       0.00      0.00      0.00         2
         462       0.00      0.00      0.00         2
         463       0.00      0.00      0.00         4
         464       0.00      0.00      0.00         2
         465       0.00      0.00      0.00         0
         466       0.00      0.00      0.00         6
         467       0.00      0.00      0.00         4
         468       0.00      0.00      0.00         1
         469       0.00      0.00      0.00         2
         470       0.00      0.00      0.00         3
         471       0.00      0.00      0.00         1
         472       0.00      0.00      0.00         1
         473       0.00      0.00      0.00         2
         474       0.00      0.00      0.00         1
         476       0.00      0.00      0.00         1
         477       0.00      0.00      0.00         1
         478       0.00      0.00      0.00         2
         479       0.00      0.00      0.00         1
         480       0.00      0.00      0.00         2
         481       0.00      0.00      0.00         3
         482       0.00      0.00      0.00         1
         483       0.00      0.00      0.00         4
         484       0.00      0.00      0.00         1
         485       0.00      0.00      0.00         0
         486       0.00      0.00      0.00         2
         487       0.00      0.00      0.00         1
         488       0.00      0.00      0.00         3
         489       0.00      0.00      0.00         1
         490       0.00      0.00      0.00         3
         491       0.00      0.00      0.00         2
         492       0.00      0.00      0.00         1
         493       0.00      0.00      0.00         1
         494       0.00      0.00      0.00         1
         495       0.00      0.00      0.00         3
         496       0.00      0.00      0.00         2
         497       0.00      0.00      0.00         1
         498       0.00      0.00      0.00         2
         499       0.12      1.00      0.22         1
         500       0.00      0.00      0.00         2
         501       0.00      0.00      0.00         1
         502       0.00      0.00      0.00         0
         503       0.00      0.00      0.00         1
         505       0.00      0.00      0.00         2
         506       0.00      0.00      0.00         1
         508       0.00      0.00      0.00         0
         509       0.00      0.00      0.00         2
         511       0.00      0.00      0.00         1
         512       0.00      0.00      0.00         3
         513       0.25      0.25      0.25         4
         514       0.00      0.00      0.00         2
         516       0.00      0.00      0.00         2
         517       0.00      0.00      0.00         2
         518       0.00      0.00      0.00         1
         520       0.00      0.00      0.00         4
         521       0.00      0.00      0.00         2
         522       0.00      0.00      0.00         1
         523       0.00      0.00      0.00         2
         524       0.00      0.00      0.00         0
         525       0.00      0.00      0.00         3
         526       0.00      0.00      0.00         1
         527       0.00      0.00      0.00         1
         529       0.00      0.00      0.00         0
         530       0.00      0.00      0.00         1
         531       0.00      0.00      0.00         1
         533       0.00      0.00      0.00         1
         536       0.00      0.00      0.00         2
         537       0.00      0.00      0.00         2
         538       0.00      0.00      0.00         1
         539       0.00      0.00      0.00         3
         541       0.00      0.00      0.00         2
         543       0.00      0.00      0.00         1
         544       0.00      0.00      0.00         1
         545       0.00      0.00      0.00         2
         546       0.00      0.00      0.00         2
         547       0.00      0.00      0.00         3
         549       0.00      0.00      0.00         1
         550       0.00      0.00      0.00         2
         552       0.00      0.00      0.00         1
         553       0.00      0.00      0.00         1
         554       0.00      0.00      0.00         0
         555       0.00      0.00      0.00         2
         556       0.20      1.00      0.33         1
         557       0.00      0.00      0.00         0
         558       0.00      0.00      0.00         3
         559       0.00      0.00      0.00         1
         560       0.00      0.00      0.00         0
         561       0.00      0.00      0.00         0
         562       0.00      0.00      0.00         1
         563       0.00      0.00      0.00         0
         564       0.00      0.00      0.00         3
         565       0.00      0.00      0.00         1
         566       0.00      0.00      0.00         3
         568       0.00      0.00      0.00         3
         569       0.00      0.00      0.00         3
         570       0.00      0.00      0.00         1
         571       0.00      0.00      0.00         3
         572       0.00      0.00      0.00         2
         573       0.00      0.00      0.00         1
         575       0.00      0.00      0.00         1
         576       0.00      0.00      0.00         1
         577       0.00      0.00      0.00         1
```

```
578      0.00      0.00      0.00        1
579      0.00      0.00      0.00        5
581      0.00      0.00      0.00        1
582      0.00      0.00      0.00        0
584      0.00      0.00      0.00        1
585      0.00      0.00      0.00        4
586      0.00      0.00      0.00        3
588      0.00      0.00      0.00        1
589      0.00      0.00      0.00        1
590      0.00      0.00      0.00        1
591      0.00      0.00      0.00        1
592      0.00      0.00      0.00        1
593      0.00      0.00      0.00        1
594      0.00      0.00      0.00        1
595      0.00      0.00      0.00        1
596      0.00      0.00      0.00        1
598      0.00      0.00      0.00        1
600      0.00      0.00      0.00        0
601      0.00      0.00      0.00        1
602      0.00      0.00      0.00        2
603      0.00      0.00      0.00        1
604      0.00      0.00      0.00        1
607      0.00      0.00      0.00        0
608      0.00      0.00      0.00        1
610      0.00      0.00      0.00        2
611      0.00      0.00      0.00        0
613      0.00      0.00      0.00        1
615      0.00      0.00      0.00        3
616      0.00      0.00      0.00        1
617      0.00      0.00      0.00        3
618      0.00      0.00      0.00        1
619      0.00      0.00      0.00        1
620      0.00      0.00      0.00        1
622      0.00      0.00      0.00        0
626      0.00      0.00      0.00        1
627      0.00      0.00      0.00        0
631      0.00      0.00      0.00        2
632      0.00      0.00      0.00        1
633      0.00      0.00      0.00        1
634      0.00      0.00      0.00        1
635      0.00      0.00      0.00        0
638      0.00      0.00      0.00        0
639      0.00      0.00      0.00        1
640      0.00      0.00      0.00        0
641      0.00      0.00      0.00        1
642      0.00      0.00      0.00        0
643      0.00      0.00      0.00        0
644      0.00      0.00      0.00        0
646      0.00      0.00      0.00        0
647      0.00      0.00      0.00        2
648      0.00      0.00      0.00        1
649      0.00      0.00      0.00        2
651      0.00      0.00      0.00        1
652      0.00      0.00      0.00        1
653      0.00      0.00      0.00        1
654      0.00      0.00      0.00        2
655      0.00      0.00      0.00        1
656      0.00      0.00      0.00        1
659      0.00      0.00      0.00        2
660      0.00      0.00      0.00        1
662      0.00      0.00      0.00        1
668      0.00      0.00      0.00        2
669      0.00      0.00      0.00        1
671      0.00      0.00      0.00        2
673      0.00      0.00      0.00        1
676      0.00      0.00      0.00        2
678      0.00      0.00      0.00        3
679      0.00      0.00      0.00        1
681      0.00      0.00      0.00        2
```

```
682      0.00      0.00      0.00          1
683      0.00      0.00      0.00          1
684      0.00      0.00      0.00          1
685      0.00      0.00      0.00          1
686      1.00      1.00      1.00          1
687      0.00      0.00      0.00          1
689      0.00      0.00      0.00          1
690      0.00      0.00      0.00          1
692      0.00      0.00      0.00          0
693      0.00      0.00      0.00          1
694      0.00      0.00      0.00          1
696      0.00      0.00      0.00          1
704      0.00      0.00      0.00          1
710      0.00      0.00      0.00          1
711      0.00      0.00      0.00          0
713      0.00      0.00      0.00          2
719      0.00      0.00      0.00          0
721      0.00      0.00      0.00          1
723      0.00      0.00      0.00          1
724      0.00      0.00      0.00          1
725      0.00      0.00      0.00          1
729      0.00      0.00      0.00          1
730      0.00      0.00      0.00          1
731      0.00      0.00      0.00          2
734      0.00      0.00      0.00          1
737      0.00      0.00      0.00          0
743      0.00      0.00      0.00          3
744      0.00      0.00      0.00          0
745      0.00      0.00      0.00          1
749      0.00      0.00      0.00          1
757      0.00      0.00      0.00          1
759      0.00      0.00      0.00          2
770      0.00      0.00      0.00          0
771      0.00      0.00      0.00          1
772      0.00      0.00      0.00          0
776      0.00      0.00      0.00          1
777      0.00      0.00      0.00          1
782      0.00      0.00      0.00          1
783      0.00      0.00      0.00          1
784      0.00      0.00      0.00          1
788      0.00      0.00      0.00          1
791      0.00      0.00      0.00          1
795      0.00      0.00      0.00          2
798      0.00      0.00      0.00          1
806      0.00      0.00      0.00          1
809      0.00      0.00      0.00          1
811      0.00      0.00      0.00          1
812      0.00      0.00      0.00          0
814      0.00      0.00      0.00          0
817      0.00      0.00      0.00          1
818      0.00      0.00      0.00          1
823      0.00      0.00      0.00          1
827      0.00      0.00      0.00          1
831      0.00      0.00      0.00          1
832      0.00      0.00      0.00          1
834      0.00      0.00      0.00          1
835      0.00      0.00      0.00          1
839      0.00      0.00      0.00          0
842      0.00      0.00      0.00          1
846      0.00      0.00      0.00          1
849      0.00      0.00      0.00          1
850      0.00      0.00      0.00          1
851      0.00      0.00      0.00          1
852      0.00      0.00      0.00          0
858      0.00      0.00      0.00          0
863      0.00      0.00      0.00          1
868      0.00      0.00      0.00          1
869      0.00      0.00      0.00          1
872      0.00      0.00      0.00          0
```

```
         884       0.00      0.00      0.00         0
         888       0.00      0.00      0.00         1
         977       0.00      0.00      0.00         1

    accuracy                           0.07      3266
   macro avg       0.03      0.04      0.03      3266
weighted avg       0.07      0.07      0.06      3266
```

In [76]:
```
from sklearn.tree import DecisionTreeClassifier
Decision_tree = DecisionTreeClassifier()
Decision_tree.fit(train_X, train_y)
test_pred_decision_tree = Decision_tree.predict(test_X)
```

In [77]:
```
print(metrics.classification_report(test_y, test_pred_decision_tree))
```

```
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        34
           2       0.98      0.98      0.98        44
           3       0.98      0.98      0.98        47
           4       0.97      1.00      0.99        36
           5       0.98      1.00      0.99        52
           6       1.00      0.98      0.99        56
           7       0.98      0.98      0.98        41
           8       0.96      1.00      0.98        25
           9       1.00      1.00      1.00        23
          10       1.00      0.96      0.98        28
          11       0.97      1.00      0.99        36
          12       1.00      1.00      1.00        22
          13       0.90      0.90      0.90        20
          14       0.71      0.77      0.74        13
          15       0.72      0.87      0.79        15
          16       1.00      0.70      0.83        27
          17       0.66      0.95      0.78        20
          18       0.86      0.67      0.75         9
          19       0.75      0.75      0.75         8
          20       1.00      0.83      0.91        18
          21       1.00      0.81      0.90        16
          22       0.44      0.80      0.57         5
          23       0.88      0.68      0.77        22
          24       0.50      0.64      0.56        11
          25       0.54      0.58      0.56        12
          26       0.70      0.54      0.61        13
          27       0.67      0.73      0.70        11
          28       0.77      0.71      0.74        14
          29       0.50      0.62      0.56         8
          30       0.33      0.23      0.27        13
          31       0.27      0.44      0.33         9
          32       0.57      0.33      0.42        12
          33       0.60      0.60      0.60        10
          34       0.58      0.64      0.61        11
          35       0.27      0.27      0.27        15
          36       0.40      0.80      0.53         5
          37       0.67      0.43      0.52        14
          38       0.40      0.50      0.44         8
          39       0.33      0.20      0.25        15
          40       0.11      0.17      0.13         6
          41       0.40      0.60      0.48        10
          42       0.14      0.20      0.17         5
          43       0.25      0.20      0.22        10
          44       0.30      0.27      0.29        11
          45       0.20      0.08      0.12        12
          46       0.21      0.57      0.31         7
          47       0.25      0.38      0.30         8
          48       0.17      0.09      0.12        11
          49       1.00      0.33      0.50         9
          50       0.14      0.17      0.15         6
```

```
 51       0.50       0.50       0.50           8
 52       0.73       0.57       0.64          14
 53       0.38       0.42       0.40          12
 54       0.25       0.33       0.29           6
 55       0.29       0.25       0.27           8
 56       0.20       0.50       0.29           6
 57       0.62       0.36       0.45          14
 58       0.17       0.17       0.17           6
 59       0.17       0.12       0.14           8
 60       0.12       0.20       0.15           5
 61       0.25       0.25       0.25           4
 62       0.50       0.44       0.47           9
 63       0.25       0.20       0.22           5
 64       0.42       0.53       0.47          15
 65       0.67       0.25       0.36           8
 66       0.43       0.43       0.43           7
 67       0.00       0.00       0.00           2
 68       0.40       0.29       0.33           7
 69       0.50       0.40       0.44          10
 70       0.17       0.14       0.15           7
 71       0.42       0.56       0.48           9
 72       0.33       0.30       0.32          10
 73       0.45       0.56       0.50           9
 74       0.00       0.00       0.00           6
 75       0.44       0.44       0.44           9
 76       0.00       0.00       0.00           5
 77       0.11       0.20       0.14           5
 78       0.29       0.14       0.19          14
 79       0.29       0.22       0.25           9
 80       0.25       0.25       0.25           4
 81       0.40       0.29       0.33           7
 82       0.20       0.17       0.18           6
 83       0.67       0.22       0.33           9
 84       0.40       0.75       0.52           8
 85       0.00       0.00       0.00           3
 86       0.20       0.50       0.29           4
 87       0.18       0.20       0.19          10
 88       0.25       0.09       0.13          11
 89       0.42       0.56       0.48           9
 90       0.08       0.10       0.09          10
 91       0.00       0.00       0.00           6
 92       0.29       0.20       0.24          10
 93       0.12       0.11       0.12           9
 94       0.50       0.20       0.29          10
 95       0.13       0.20       0.16          10
 96       0.22       0.25       0.24           8
 97       0.33       0.20       0.25           5
 98       0.12       0.17       0.14           6
 99       0.25       0.25       0.25           4
100       0.00       0.00       0.00           2
101       0.50       0.25       0.33           4
102       0.25       0.29       0.27           7
103       0.00       0.00       0.00           8
104       0.14       0.17       0.15           6
105       0.33       0.33       0.33           9
106       0.27       0.33       0.30          12
107       0.30       0.38       0.33           8
108       0.56       0.45       0.50          11
109       0.00       0.00       0.00           7
110       0.14       0.14       0.14           7
111       0.00       0.00       0.00           4
112       0.40       0.25       0.31           8
113       0.27       0.33       0.30           9
114       0.00       0.00       0.00          11
115       0.00       0.00       0.00           5
116       0.08       0.25       0.12           4
117       0.00       0.00       0.00           5
118       0.14       0.18       0.16          11
119       0.14       0.12       0.13           8
```

```
120       0.40       0.22       0.29        9
121       0.12       0.20       0.15        5
122       0.25       0.25       0.25        8
123       0.25       0.50       0.33        4
124       0.50       0.42       0.45       12
125       0.00       0.00       0.00        7
126       0.33       0.40       0.36       10
127       0.22       0.29       0.25        7
128       0.29       0.33       0.31        6
129       0.40       0.40       0.40        5
130       0.00       0.00       0.00        8
131       0.20       0.67       0.31        3
132       0.00       0.00       0.00        7
133       0.57       0.57       0.57        7
134       0.33       0.33       0.33        9
135       0.67       0.25       0.36        8
136       0.25       0.22       0.24        9
137       0.12       0.17       0.14        6
138       0.00       0.00       0.00        8
139       0.09       0.11       0.10        9
140       0.29       0.20       0.24       10
141       0.07       0.11       0.09        9
142       0.00       0.00       0.00        3
143       0.20       0.20       0.20        5
144       0.12       0.20       0.15        5
145       0.00       0.00       0.00        5
146       0.12       0.17       0.14        6
147       0.12       0.11       0.12        9
148       0.29       0.25       0.27        8
149       0.00       0.00       0.00        4
150       0.12       0.12       0.12        8
151       0.00       0.00       0.00        5
152       0.33       0.20       0.25       10
153       0.25       0.36       0.30       11
154       0.14       0.09       0.11       11
155       0.00       0.00       0.00        9
156       0.25       0.17       0.20        6
157       0.38       0.38       0.38        8
158       0.00       0.00       0.00        5
159       0.33       0.12       0.18        8
160       0.33       0.17       0.22        6
161       0.10       0.17       0.12        6
162       0.07       0.20       0.11        5
163       0.25       0.17       0.20        6
164       0.17       0.20       0.18        5
165       0.20       0.11       0.14        9
166       0.00       0.00       0.00        5
167       0.00       0.00       0.00        9
168       0.00       0.00       0.00        7
169       0.33       0.17       0.22        6
170       0.00       0.00       0.00        8
171       0.14       0.11       0.12        9
172       0.37       0.58       0.45       12
173       0.00       0.00       0.00        6
174       0.25       0.20       0.22        5
175       0.23       0.33       0.27        9
176       0.00       0.00       0.00        4
177       0.00       0.00       0.00        8
178       0.29       0.20       0.24       10
179       0.50       0.40       0.44       10
180       0.20       0.22       0.21        9
181       0.27       0.25       0.26       12
182       0.08       0.17       0.11        6
183       0.00       0.00       0.00        7
184       0.00       0.00       0.00        7
185       0.20       0.25       0.22        8
186       0.00       0.00       0.00        6
187       0.00       0.00       0.00        6
188       0.10       0.25       0.14        4
```

```
        189       0.20      0.33      0.25         6
        190       0.00      0.00      0.00        10
        191       0.20      0.33      0.25         6
        192       0.00      0.00      0.00         6
        193       0.10      0.20      0.13         5
        194       0.00      0.00      0.00         3
        195       0.00      0.00      0.00        13
        196       0.20      0.12      0.15         8
        197       0.33      0.25      0.29         4
        198       0.17      0.25      0.20         4
        199       0.25      0.50      0.33         2
        200       0.00      0.00      0.00         5
        201       0.00      0.00      0.00        10
        202       0.00      0.00      0.00         5
        203       0.22      0.25      0.24         8
        204       0.12      0.20      0.15         5
        205       0.17      0.25      0.20         8
        206       0.00      0.00      0.00         8
        207       0.10      0.20      0.13         5
        208       0.00      0.00      0.00         4
        209       0.00      0.00      0.00         1
        210       0.25      0.17      0.20         6
        211       0.14      0.12      0.13         8
        212       0.15      0.67      0.25         3
        213       0.50      0.22      0.31         9
        214       0.14      0.09      0.11        11
        215       0.00      0.00      0.00         6
        216       0.00      0.00      0.00         6
        217       0.20      0.10      0.13        10
        218       0.20      0.22      0.21         9
        219       0.14      0.09      0.11        11
        220       0.25      0.10      0.14        10
        221       0.00      0.00      0.00         3
        222       0.08      0.10      0.09        10
        223       0.00      0.00      0.00         5
        224       0.45      0.50      0.48        10
        225       0.33      0.43      0.38         7
        226       0.00      0.00      0.00         5
        227       0.00      0.00      0.00         6
        228       0.20      0.14      0.17         7
        229       0.00      0.00      0.00         4
        230       0.20      0.11      0.14         9
        231       0.20      0.33      0.25         3
        232       0.12      0.17      0.14         6
        233       0.18      0.22      0.20         9
        234       0.33      0.50      0.40         2
        235       0.00      0.00      0.00         6
        236       0.00      0.00      0.00         4
        237       0.40      0.40      0.40         5
        238       0.00      0.00      0.00         5
        239       0.14      0.17      0.15         6
        240       0.00      0.00      0.00         2
        241       0.40      0.33      0.36         6
        242       0.00      0.00      0.00         1
        243       0.00      0.00      0.00         7
        244       0.14      0.50      0.22         4
        245       0.00      0.00      0.00         5
        246       0.00      0.00      0.00         3
        247       0.00      0.00      0.00         5
        248       0.40      0.22      0.29         9
        249       0.00      0.00      0.00         5
        250       0.00      0.00      0.00         4
        251       0.00      0.00      0.00         3
        252       0.00      0.00      0.00         3
        253       0.14      0.25      0.18         4
        254       0.25      0.25      0.25         4
        255       0.00      0.00      0.00         3
        256       0.00      0.00      0.00         7
        257       0.25      0.25      0.25         4
```

```
         258       0.33      0.11      0.17         9
         259       0.14      0.20      0.17         5
         260       0.00      0.00      0.00         8
         261       0.00      0.00      0.00         0
         262       0.00      0.00      0.00         4
         263       0.00      0.00      0.00         5
         264       0.00      0.00      0.00         7
         265       0.00      0.00      0.00         2
         266       0.33      0.20      0.25         5
         267       0.09      0.20      0.13         5
         268       0.00      0.00      0.00         8
         269       0.00      0.00      0.00         5
         270       0.00      0.00      0.00         3
         271       0.00      0.00      0.00         3
         272       0.14      0.12      0.13         8
         273       0.00      0.00      0.00         5
         274       0.00      0.00      0.00         6
         275       0.00      0.00      0.00         2
         276       0.00      0.00      0.00         5
         277       0.00      0.00      0.00         6
         278       0.00      0.00      0.00         6
         279       0.00      0.00      0.00         2
         280       0.00      0.00      0.00         6
         281       0.00      0.00      0.00         8
         282       0.29      0.33      0.31         6
         283       0.17      0.20      0.18         5
         284       0.00      0.00      0.00         1
         285       0.00      0.00      0.00         3
         286       0.00      0.00      0.00         4
         287       0.00      0.00      0.00         4
         288       0.00      0.00      0.00         5
         289       0.00      0.00      0.00         4
         290       0.33      0.33      0.33         3
         291       0.25      0.17      0.20         6
         292       0.00      0.00      0.00         8
         293       0.00      0.00      0.00         2
         294       0.00      0.00      0.00         8
         295       0.00      0.00      0.00         0
         296       0.00      0.00      0.00         3
         297       0.25      0.17      0.20         6
         298       0.00      0.00      0.00         2
         299       0.00      0.00      0.00         4
         300       0.12      0.20      0.15         5
         301       0.00      0.00      0.00         1
         302       0.00      0.00      0.00         4
         303       0.00      0.00      0.00         4
         304       0.00      0.00      0.00         6
         305       0.00      0.00      0.00         2
         306       0.00      0.00      0.00         2
         307       0.00      0.00      0.00         1
         308       0.00      0.00      0.00         4
         309       0.00      0.00      0.00         0
         310       0.00      0.00      0.00         2
         311       0.00      0.00      0.00         2
         312       0.00      0.00      0.00         5
         313       0.33      0.25      0.29         4
         314       0.00      0.00      0.00         4
         315       0.00      0.00      0.00         4
         316       0.00      0.00      0.00         3
         317       0.00      0.00      0.00         3
         318       0.00      0.00      0.00         2
         319       0.17      0.25      0.20         4
         320       0.00      0.00      0.00         3
         321       0.00      0.00      0.00         1
         322       0.20      0.33      0.25         3
         323       0.00      0.00      0.00         4
         324       0.00      0.00      0.00         1
         325       0.00      0.00      0.00         4
         326       0.00      0.00      0.00         2
         327       0.00      0.00      0.00         3
         328       0.00      0.00      0.00         8
         329       0.00      0.00      0.00         2
         330       0.20      0.50      0.29         2
         331       0.00      0.00      0.00         3
         332       0.00      0.00      0.00         6
         333       0.00      0.00      0.00         0
         334       1.00      0.14      0.25         7
         335       0.00      0.00      0.00         6
         336       0.00      0.00      0.00         2
         337       0.00      0.00      0.00         3
         338       0.00      0.00      0.00         3
         339       0.00      0.00      0.00         1
         340       0.00      0.00      0.00         3
         341       0.00      0.00      0.00         3
         342       0.00      0.00      0.00         5
         343       0.00      0.00      0.00         7
         344       0.00      0.00      0.00         1
         345       0.00      0.00      0.00         1
         346       0.00      0.00      0.00         0
         347       0.00      0.00      0.00         6
         348       0.00      0.00      0.00         3
         349       0.00      0.00      0.00         4
         350       0.00      0.00      0.00         6
         351       0.00      0.00      0.00         4
         352       0.00      0.00      0.00         2
         353       0.00      0.00      0.00         4
         354       0.00      0.00      0.00         3
         355       0.00      0.00      0.00         2
         356       0.00      0.00      0.00         7
         357       0.00      0.00      0.00         4
         358       0.00      0.00      0.00         3
         359       0.00      0.00      0.00         4
         360       0.00      0.00      0.00         2
         361       0.00      0.00      0.00         3
         362       0.00      0.00      0.00         4
         363       0.00      0.00      0.00         6
         364       0.00      0.00      0.00         1
         365       0.00      0.00      0.00         6
         366       0.00      0.00      0.00         1
         367       0.00      0.00      0.00         4
         369       0.00      0.00      0.00         2
         370       0.00      0.00      0.00         4
         371       0.11      0.50      0.18         2
         372       0.00      0.00      0.00         6
         373       0.00      0.00      0.00         2
         374       0.00      0.00      0.00         7
         375       0.00      0.00      0.00         3
         376       0.00      0.00      0.00         1
         377       0.00      0.00      0.00         4
         378       0.00      0.00      0.00         1
         379       0.00      0.00      0.00         2
         380       0.00      0.00      0.00         1
         381       0.00      0.00      0.00         3
         382       0.17      0.50      0.25         2
         383       0.00      0.00      0.00         1
         384       0.00      0.00      0.00         2
         385       0.33      0.20      0.25         5
         386       0.00      0.00      0.00         1
         387       0.00      0.00      0.00         4
         388       0.00      0.00      0.00         1
         389       0.00      0.00      0.00         4
         390       0.00      0.00      0.00         4
         392       0.00      0.00      0.00         1
         393       0.00      0.00      0.00         3
         394       0.00      0.00      0.00         1
         395       0.00      0.00      0.00         2
         396       0.00      0.00      0.00         5
         397       0.00      0.00      0.00         2
         398       0.50      0.25      0.33         4
         399       0.00      0.00      0.00         0
         400       0.00      0.00      0.00         2
         401       0.00      0.00      0.00         3
         402       0.00      0.00      0.00         1
         403       0.00      0.00      0.00         2
         404       0.50      0.50      0.50         4
         405       0.00      0.00      0.00         4
         406       0.00      0.00      0.00         1
         407       0.00      0.00      0.00         0
         408       0.00      0.00      0.00         2
         409       0.00      0.00      0.00         1
         410       0.00      0.00      0.00         1
         411       0.00      0.00      0.00         1
         413       0.00      0.00      0.00         4
         414       0.00      0.00      0.00         1
         415       0.00      0.00      0.00         1
         416       0.00      0.00      0.00         2
         417       0.00      0.00      0.00         1
         419       0.00      0.00      0.00         5
         420       0.00      0.00      0.00         3
         421       0.00      0.00      0.00         3
         422       0.00      0.00      0.00         1
         423       0.00      0.00      0.00         2
         425       0.00      0.00      0.00         3
         426       0.00      0.00      0.00         1
         427       0.00      0.00      0.00         0
         428       0.00      0.00      0.00         5
         429       0.00      0.00      0.00         1
         430       0.00      0.00      0.00         4
         431       0.00      0.00      0.00         1
         432       0.00      0.00      0.00         2
         433       0.00      0.00      0.00         2
         434       0.00      0.00      0.00         1
         435       0.00      0.00      0.00         3
         436       0.00      0.00      0.00         2
         437       0.00      0.00      0.00         0
         438       0.00      0.00      0.00         0
         439       0.00      0.00      0.00         1
         440       0.00      0.00      0.00         0
         441       0.00      0.00      0.00         1
         442       0.00      0.00      0.00         1
         443       0.00      0.00      0.00         2
         444       0.00      0.00      0.00         1
         445       0.25      0.33      0.29         3
         446       0.50      0.25      0.33         4
         447       0.00      0.00      0.00         4
         448       0.00      0.00      0.00         1
         449       0.00      0.00      0.00         1
         450       0.00      0.00      0.00         1
         451       0.00      0.00      0.00         3
         452       0.00      0.00      0.00         3
         453       0.00      0.00      0.00         4
         454       0.00      0.00      0.00         3
         455       0.50      0.25      0.33         4
         456       0.00      0.00      0.00         4
         457       0.00      0.00      0.00         2
         458       0.00      0.00      0.00         1
         459       0.00      0.00      0.00         2
         460       0.00      0.00      0.00         2
         461       0.00      0.00      0.00         2
         462       0.00      0.00      0.00         2
         463       0.00      0.00      0.00         4
         464       0.00      0.00      0.00         2
         465       0.00      0.00      0.00         0
         466       0.00      0.00      0.00         6
         467       0.00      0.00      0.00         4
         468       0.00      0.00      0.00         1
         469       0.00      0.00      0.00         2
         470       0.00      0.00      0.00         3
         471       0.00      0.00      0.00         1
         472       0.00      0.00      0.00         1
         473       0.00      0.00      0.00         2
         474       0.00      0.00      0.00         1
         475       0.00      0.00      0.00         0
         476       0.00      0.00      0.00         1
         477       0.00      0.00      0.00         1
         478       0.00      0.00      0.00         2
         479       0.00      0.00      0.00         1
         480       0.00      0.00      0.00         2
         481       0.00      0.00      0.00         3
         482       0.50      1.00      0.67         1
         483       0.00      0.00      0.00         4
         484       0.00      0.00      0.00         1
         486       0.00      0.00      0.00         2
         487       0.00      0.00      0.00         1
         488       0.00      0.00      0.00         3
         489       0.00      0.00      0.00         1
         490       0.00      0.00      0.00         3
         491       0.00      0.00      0.00         2
         492       0.00      0.00      0.00         1
         493       0.00      0.00      0.00         1
         494       0.00      0.00      0.00         1
         495       0.00      0.00      0.00         3
         496       0.00      0.00      0.00         2
         497       0.00      0.00      0.00         1
         498       0.00      0.00      0.00         2
         499       0.00      0.00      0.00         1
         500       0.00      0.00      0.00         2
         501       0.00      0.00      0.00         1
         502       0.00      0.00      0.00         0
         503       0.00      0.00      0.00         1
         504       0.00      0.00      0.00         0
         505       0.00      0.00      0.00         2
         506       0.00      0.00      0.00         1
         507       0.00      0.00      0.00         0
         508       0.00      0.00      0.00         0
         509       0.00      0.00      0.00         2
         511       0.00      0.00      0.00         1
         512       0.00      0.00      0.00         3
         513       0.00      0.00      0.00         4
         514       0.00      0.00      0.00         2
         515       0.00      0.00      0.00         0
         516       0.00      0.00      0.00         2
         517       0.00      0.00      0.00         2
         518       0.00      0.00      0.00         1
         520       0.00      0.00      0.00         4
         521       0.00      0.00      0.00         2
         522       0.00      0.00      0.00         1
         523       0.00      0.00      0.00         2
         524       0.00      0.00      0.00         0
         525       0.00      0.00      0.00         3
         526       0.00      0.00      0.00         1
         527       0.00      0.00      0.00         1
         528       0.00      0.00      0.00         0
         530       0.00      0.00      0.00         1
         531       0.00      0.00      0.00         1
         532       0.00      0.00      0.00         0
         533       0.00      0.00      0.00         1
         534       0.00      0.00      0.00         0
         536       0.00      0.00      0.00         2
         537       0.00      0.00      0.00         2
         538       1.00      1.00      1.00         1
         539       0.00      0.00      0.00         3
         540       0.00      0.00      0.00         0
         541       0.33      0.50      0.40         2
         542       0.00      0.00      0.00         0
         543       0.00      0.00      0.00         1
         544       0.00      0.00      0.00         1
         545       0.00      0.00      0.00         2
         546       0.00      0.00      0.00         2
         547       0.00      0.00      0.00         3
         549       0.00      0.00      0.00         1
         550       0.00      0.00      0.00         2
         551       0.00      0.00      0.00         0
         552       0.00      0.00      0.00         1
         553       0.00      0.00      0.00         1
         554       0.00      0.00      0.00         0
         555       0.00      0.00      0.00         2
         556       0.50      1.00      0.67         1
         557       0.00      0.00      0.00         0
         558       0.00      0.00      0.00         3
         559       0.00      0.00      0.00         1
         561       0.00      0.00      0.00         0
         562       0.00      0.00      0.00         1
         563       0.00      0.00      0.00         0
         564       0.00      0.00      0.00         3
         565       0.00      0.00      0.00         1
         566       0.00      0.00      0.00         3
         567       0.00      0.00      0.00         0
         568       0.00      0.00      0.00         3
         569       0.00      0.00      0.00         3
         570       0.00      0.00      0.00         1
         571       0.00      0.00      0.00         3
         572       0.00      0.00      0.00         2
         573       0.00      0.00      0.00         1
         575       0.00      0.00      0.00         1
         576       0.00      0.00      0.00         1
         577       0.00      0.00      0.00         1
         578       0.00      0.00      0.00         1
         579       1.00      0.20      0.33         5
         580       0.00      0.00      0.00         0
         581       0.00      0.00      0.00         1
         582       0.00      0.00      0.00         0
         584       0.00      0.00      0.00         1
         585       0.00      0.00      0.00         4
         586       0.00      0.00      0.00         3
         588       0.00      0.00      0.00         1
         589       0.00      0.00      0.00         1
         590       0.00      0.00      0.00         1
         591       0.00      0.00      0.00         1
         592       0.00      0.00      0.00         1
         593       0.00      0.00      0.00         1
         594       0.00      0.00      0.00         1
         595       0.00      0.00      0.00         1
         596       0.00      0.00      0.00         1
         598       0.00      0.00      0.00         1
         601       0.00      0.00      0.00         1
         602       0.00      0.00      0.00         2
         603       0.00      0.00      0.00         1
         604       0.00      0.00      0.00         1
         605       0.00      0.00      0.00         0
         606       0.00      0.00      0.00         0
         607       0.00      0.00      0.00         0
         608       0.00      0.00      0.00         1
         610       0.00      0.00      0.00         2
         611       0.00      0.00      0.00         0
         613       0.00      0.00      0.00         1
         614       0.00      0.00      0.00         0
         615       0.00      0.00      0.00         3
         616       0.00      0.00      0.00         1
         617       0.00      0.00      0.00         3
         618       0.00      0.00      0.00         1
         619       0.00      0.00      0.00         1
         620       0.00      0.00      0.00         1
         622       0.00      0.00      0.00         0
         626       0.00      0.00      0.00         1
         628       0.00      0.00      0.00         0
         629       0.00      0.00      0.00         0
         631       0.00      0.00      0.00         2
         632       0.00      0.00      0.00         1
         633       1.00      1.00      1.00         1
         634       0.00      0.00      0.00         1
         635       0.00      0.00      0.00         0
         638       0.00      0.00      0.00         0
         639       0.00      0.00      0.00         1
         641       0.00      0.00      0.00         1
         644       0.00      0.00      0.00         0
         646       0.00      0.00      0.00         0
         647       0.00      0.00      0.00         2
         648       0.00      0.00      0.00         1
         649       0.00      0.00      0.00         2
         651       0.00      0.00      0.00         1
         652       0.00      0.00      0.00         1
         653       0.00      0.00      0.00         1
         654       0.00      0.00      0.00         2
         655       0.00      0.00      0.00         1
         656       0.00      0.00      0.00         1
         657       0.00      0.00      0.00         0
         659       0.00      0.00      0.00         2
         660       0.00      0.00      0.00         1
         662       0.00      0.00      0.00         1
         667       0.00      0.00      0.00         0
         668       0.00      0.00      0.00         2
         669       1.00      1.00      1.00         1
         671       0.00      0.00      0.00         2
         672       0.00      0.00      0.00         0
         673       0.00      0.00      0.00         1
         676       0.00      0.00      0.00         2
         677       0.00      0.00      0.00         0
         678       0.00      0.00      0.00         3
         679       0.00      0.00      0.00         1
         681       0.00      0.00      0.00         2
         682       0.00      0.00      0.00         1
         683       0.00      0.00      0.00         1
         684       0.00      0.00      0.00         1
         685       0.00      0.00      0.00         1
         686       1.00      1.00      1.00         1
         687       0.00      0.00      0.00         1
         688       0.00      0.00      0.00         0
         689       0.00      0.00      0.00         1
         690       0.00      0.00      0.00         1
         691       0.00      0.00      0.00         0
         692       0.00      0.00      0.00         0
         693       0.00      0.00      0.00         1
         694       0.00      0.00      0.00         1
         696       0.00      0.00      0.00         1
         698       0.00      0.00      0.00         0
         700       0.00      0.00      0.00         0
         701       0.00      0.00      0.00         0
         704       0.33      1.00      0.50         1
         706       0.00      0.00      0.00         0
         710       0.00      0.00      0.00         1
         711       0.00      0.00      0.00         0
         713       0.00      0.00      0.00         2
         719       0.00      0.00      0.00         0
         721       0.00      0.00      0.00         1
         722       0.00      0.00      0.00         0
         723       0.00      0.00      0.00         1
         724       0.00      0.00      0.00         1
         725       0.00      0.00      0.00         1
         729       0.00      0.00      0.00         1
         730       0.00      0.00      0.00         1
         731       0.00      0.00      0.00         2
         734       0.00      0.00      0.00         1
         738       0.00      0.00      0.00         0
         739       0.00      0.00      0.00         0
         743       0.00      0.00      0.00         3
         744       0.00      0.00      0.00         0
         745       0.00      0.00      0.00         1
         746       0.00      0.00      0.00         0
         749       0.00      0.00      0.00         1
         757       0.00      0.00      0.00         1
         759       0.00      0.00      0.00         2
         766       0.00      0.00      0.00         0
         769       0.00      0.00      0.00         0
         770       0.00      0.00      0.00         0
         771       0.00      0.00      0.00         1
         774       0.00      0.00      0.00         0
         776       0.00      0.00      0.00         1
         777       0.00      0.00      0.00         1
         779       0.00      0.00      0.00         0
         782       0.00      0.00      0.00         1
         783       0.00      0.00      0.00         1
         784       0.00      0.00      0.00         1
         785       0.00      0.00      0.00         0
         788       0.00      0.00      0.00         1
         791       0.00      0.00      0.00         1
         795       0.00      0.00      0.00         2
         798       0.00      0.00      0.00         1
         806       0.00      0.00      0.00         1
         809       0.00      0.00      0.00         1
         811       0.00      0.00      0.00         1
         814       0.00      0.00      0.00         0
         817       0.00      0.00      0.00         1
         818       0.00      0.00      0.00         1
         823       1.00      1.00      1.00         1
         825       0.00      0.00      0.00         0
         827       0.00      0.00      0.00         1
         831       0.00      0.00      0.00         1
         832       0.00      0.00      0.00         1
         834       0.00      0.00      0.00         1
         835       0.00      0.00      0.00         1
         837       0.00      0.00      0.00         0
         839       0.00      0.00      0.00         0
         842       0.00      0.00      0.00         1
         843       0.00      0.00      0.00         0
         846       0.00      0.00      0.00         1
         849       0.00      0.00      0.00         1
         850       0.00      0.00      0.00         1
         851       0.00      0.00      0.00         1
         854       0.00      0.00      0.00         0
         863       0.00      0.00      0.00         1
         867       0.00      0.00      0.00         0
         868       0.00      0.00      0.00         1
         869       0.00      0.00      0.00         1
         884       0.00      0.00      0.00         0
         888       0.00      0.00      0.00         1
         917       0.00      0.00      0.00         0
         943       0.00      0.00      0.00         0
         948       0.00      0.00      0.00         0
         977       0.00      0.00      0.00         1

    accuracy                           0.31      3266
   macro avg       0.12      0.12      0.11      3266
weighted avg       0.33      0.31      0.31      3266
```

In [78]:
```python
from sklearn.ensemble import RandomForestClassifier
```

In [79]:
```python
start_time = time.time()
RFC = RandomForestClassifier()
RFC.fit(train_X, train_y)
```

```
end_time = time.time()
print(end_time-start_time)
```

```
14.846244096755981
```

In [80]:
```
pred = RFC.predict(test_X)
```

In [81]:
```
RFC.score(test_X, test_y)
```

Out[81]:
```
0.22841396203306796
```

In [82]:
```
print('Table 3.Classification Random Forest:\n',metrics.classification_report(test_y
```

```
Table 3.Classification Random Forest:
               precision    recall  f1-score   support

           1       0.97      0.94      0.96        34
           2       0.93      0.95      0.94        44
           3       0.98      0.98      0.98        47
           4       0.85      0.97      0.91        36
           5       0.96      0.94      0.95        52
           6       0.96      0.95      0.95        56
           7       0.95      0.88      0.91        41
           8       0.85      0.88      0.86        25
           9       0.83      0.83      0.83        23
          10       0.78      0.89      0.83        28
          11       0.84      0.75      0.79        36
          12       0.65      0.77      0.71        22
          13       0.80      0.60      0.69        20
          14       0.40      0.62      0.48        13
          15       0.52      0.73      0.61        15
          16       0.56      0.37      0.44        27
          17       0.38      0.40      0.39        20
          18       0.56      0.56      0.56         9
          19       0.50      0.38      0.43         8
          20       0.57      0.67      0.62        18
          21       0.50      0.56      0.53        16
          22       0.33      0.40      0.36         5
          23       0.62      0.45      0.53        22
          24       0.33      0.45      0.38        11
          25       0.25      0.25      0.25        12
          26       0.25      0.15      0.19        13
          27       0.50      0.27      0.35        11
          28       0.39      0.50      0.44        14
          29       0.27      0.38      0.32         8
          30       0.12      0.08      0.10        13
          31       0.12      0.22      0.16         9
          32       0.23      0.25      0.24        12
          33       0.29      0.40      0.33        10
          34       0.08      0.09      0.09        11
          35       0.14      0.07      0.09        15
          36       0.16      0.60      0.25         5
          37       0.14      0.07      0.10        14
          38       0.29      0.25      0.27         8
          39       0.40      0.13      0.20        15
          40       0.00      0.00      0.00         6
          41       0.23      0.30      0.26        10
          42       0.09      0.20      0.13         5
          43       0.27      0.40      0.32        10
          44       0.00      0.00      0.00        11
          45       0.25      0.08      0.12        12
          46       0.30      0.43      0.35         7
          47       0.31      0.50      0.38         8
          48       0.22      0.18      0.20        11
          49       0.00      0.00      0.00         9
          50       0.09      0.17      0.12         6
          51       0.17      0.25      0.20         8
          52       0.27      0.29      0.28        14
          53       0.60      0.25      0.35        12
          54       0.11      0.17      0.13         6
          55       0.00      0.00      0.00         8
          56       0.08      0.17      0.11         6
          57       0.25      0.14      0.18        14
          58       0.33      0.33      0.33         6
          59       0.20      0.25      0.22         8
          60       0.00      0.00      0.00         5
          61       0.00      0.00      0.00         4
          62       0.22      0.22      0.22         9
          63       0.11      0.20      0.14         5
          64       0.44      0.47      0.45        15
          65       0.50      0.12      0.20         8
          66       0.25      0.14      0.18         7
          67       0.00      0.00      0.00         2
          68       0.25      0.14      0.18         7
          69       0.38      0.30      0.33        10
          70       0.00      0.00      0.00         7
          71       0.08      0.11      0.10         9
          72       0.29      0.40      0.33        10
          73       0.00      0.00      0.00         9
          74       0.38      0.50      0.43         6
          75       0.29      0.44      0.35         9
          76       0.25      0.20      0.22         5
          77       0.00      0.00      0.00         5
          78       0.22      0.14      0.17        14
          79       0.20      0.11      0.14         9
          80       0.00      0.00      0.00         4
          81       0.50      0.14      0.22         7
          82       0.17      0.17      0.17         6
          83       0.50      0.33      0.40         9
          84       0.15      0.25      0.19         8
          85       0.00      0.00      0.00         3
          86       0.10      0.50      0.16         4
          87       0.12      0.20      0.15        10
          88       0.00      0.00      0.00        11
          89       0.22      0.22      0.22         9
          90       0.15      0.20      0.17        10
          91       0.00      0.00      0.00         6
          92       0.00      0.00      0.00        10
          93       0.20      0.11      0.14         9
          94       0.00      0.00      0.00        10
          95       0.29      0.40      0.33        10
          96       0.20      0.12      0.15         8
          97       0.25      0.20      0.22         5
          98       0.00      0.00      0.00         6
          99       0.33      0.50      0.40         4
         100       0.00      0.00      0.00         2
         101       0.25      0.25      0.25         4
         102       0.25      0.29      0.27         7
         103       0.00      0.00      0.00         8
         104       0.33      0.17      0.22         6
         105       0.12      0.11      0.12         9
         106       0.36      0.33      0.35        12
         107       0.08      0.12      0.10         8
         108       0.25      0.18      0.21        11
         109       0.40      0.29      0.33         7
         110       0.00      0.00      0.00         7
         111       0.00      0.00      0.00         4
         112       0.29      0.25      0.27         8
         113       0.18      0.22      0.20         9
         114       0.11      0.09      0.10        11
         115       0.00      0.00      0.00         5
         116       0.08      0.25      0.12         4
         117       0.00      0.00      0.00         5
         118       0.25      0.36      0.30        11
         119       0.08      0.12      0.10         8
         120       0.00      0.00      0.00         9
         121       0.00      0.00      0.00         5
```

```
122        0.25       0.25       0.25           8
123        0.09       0.25       0.13           4
124        0.19       0.25       0.21          12
125        0.00       0.00       0.00           7
126        0.20       0.10       0.13          10
127        0.00       0.00       0.00           7
128        0.00       0.00       0.00           6
129        0.00       0.00       0.00           5
130        0.00       0.00       0.00           8
131        0.00       0.00       0.00           3
132        0.00       0.00       0.00           7
133        0.00       0.00       0.00           7
134        0.00       0.00       0.00           9
135        0.00       0.00       0.00           8
136        0.11       0.11       0.11           9
137        0.12       0.17       0.14           6
138        0.11       0.12       0.12           8
139        0.14       0.11       0.12           9
140        0.12       0.10       0.11          10
141        0.00       0.00       0.00           9
142        0.00       0.00       0.00           3
143        0.00       0.00       0.00           5
144        0.14       0.20       0.17           5
145        0.00       0.00       0.00           5
146        0.17       0.17       0.17           6
147        0.11       0.11       0.11           9
148        0.14       0.12       0.13           8
149        0.00       0.00       0.00           4
150        0.00       0.00       0.00           8
151        0.00       0.00       0.00           5
152        0.00       0.00       0.00          10
153        0.00       0.00       0.00          11
154        0.09       0.09       0.09          11
155        0.14       0.11       0.12           9
156        0.00       0.00       0.00           6
157        0.00       0.00       0.00           8
158        0.00       0.00       0.00           5
159        0.00       0.00       0.00           8
160        0.00       0.00       0.00           6
161        0.00       0.00       0.00           6
162        0.09       0.20       0.13           5
163        0.00       0.00       0.00           6
164        0.08       0.20       0.12           5
165        0.08       0.11       0.09           9
166        0.00       0.00       0.00           5
167        0.00       0.00       0.00           9
168        0.00       0.00       0.00           7
169        0.00       0.00       0.00           6
170        0.00       0.00       0.00           8
171        0.00       0.00       0.00           9
172        0.09       0.08       0.09          12
173        0.12       0.17       0.14           6
174        0.12       0.20       0.15           5
175        0.67       0.22       0.33           9
176        0.00       0.00       0.00           4
177        0.10       0.12       0.11           8
178        0.00       0.00       0.00          10
179        0.14       0.10       0.12          10
180        0.30       0.33       0.32           9
181        0.44       0.33       0.38          12
182        0.00       0.00       0.00           6
183        0.00       0.00       0.00           7
184        0.00       0.00       0.00           7
185        0.00       0.00       0.00           8
186        0.00       0.00       0.00           6
187        0.00       0.00       0.00           6
188        0.10       0.25       0.14           4
189        0.12       0.17       0.14           6
190        0.00       0.00       0.00          10
```

```
        191       0.29      0.33      0.31         6
        192       0.00      0.00      0.00         6
        193       0.00      0.00      0.00         5
        194       0.00      0.00      0.00         3
        195       0.00      0.00      0.00        13
        196       0.33      0.12      0.18         8
        197       0.25      0.25      0.25         4
        198       0.00      0.00      0.00         4
        199       0.00      0.00      0.00         2
        200       0.00      0.00      0.00         5
        201       0.00      0.00      0.00        10
        202       0.12      0.20      0.15         5
        203       0.15      0.25      0.19         8
        204       0.12      0.20      0.15         5
        205       0.00      0.00      0.00         8
        206       0.00      0.00      0.00         8
        207       0.08      0.20      0.12         5
        208       0.00      0.00      0.00         4
        209       0.00      0.00      0.00         1
        210       0.17      0.17      0.17         6
        211       0.00      0.00      0.00         8
        212       0.09      0.33      0.14         3
        213       0.00      0.00      0.00         9
        214       0.17      0.09      0.12        11
        215       0.00      0.00      0.00         6
        216       0.00      0.00      0.00         6
        217       0.22      0.20      0.21        10
        218       0.00      0.00      0.00         9
        219       0.00      0.00      0.00        11
        220       0.14      0.10      0.12        10
        221       0.00      0.00      0.00         3
        222       0.00      0.00      0.00        10
        223       0.00      0.00      0.00         5
        224       0.40      0.40      0.40        10
        225       0.00      0.00      0.00         7
        226       0.00      0.00      0.00         5
        227       0.00      0.00      0.00         6
        228       0.10      0.14      0.12         7
        229       0.00      0.00      0.00         4
        230       0.00      0.00      0.00         9
        231       0.00      0.00      0.00         3
        232       0.00      0.00      0.00         6
        233       0.10      0.11      0.11         9
        234       0.00      0.00      0.00         2
        235       0.00      0.00      0.00         6
        236       0.00      0.00      0.00         4
        237       0.00      0.00      0.00         5
        238       0.00      0.00      0.00         5
        239       0.00      0.00      0.00         6
        240       0.00      0.00      0.00         2
        241       0.25      0.17      0.20         6
        242       0.00      0.00      0.00         1
        243       0.33      0.14      0.20         7
        244       0.00      0.00      0.00         4
        245       0.00      0.00      0.00         5
        246       0.00      0.00      0.00         3
        247       0.17      0.20      0.18         5
        248       0.40      0.22      0.29         9
        249       0.00      0.00      0.00         5
        250       0.00      0.00      0.00         4
        251       0.00      0.00      0.00         3
        252       0.00      0.00      0.00         3
        253       0.00      0.00      0.00         4
        254       0.00      0.00      0.00         4
        255       0.00      0.00      0.00         3
        256       0.00      0.00      0.00         7
        257       0.00      0.00      0.00         4
        258       0.14      0.11      0.12         9
        259       0.00      0.00      0.00         5
```

```
         260       0.00      0.00      0.00         8
         261       0.00      0.00      0.00         0
         262       0.00      0.00      0.00         4
         263       0.00      0.00      0.00         5
         264       0.14      0.14      0.14         7
         265       0.08      0.50      0.13         2
         266       0.00      0.00      0.00         5
         267       0.00      0.00      0.00         5
         268       0.00      0.00      0.00         8
         269       0.00      0.00      0.00         5
         270       0.00      0.00      0.00         3
         271       0.00      0.00      0.00         3
         272       0.00      0.00      0.00         8
         273       0.50      0.20      0.29         5
         274       0.00      0.00      0.00         6
         275       0.00      0.00      0.00         2
         276       0.00      0.00      0.00         5
         277       0.00      0.00      0.00         6
         278       0.00      0.00      0.00         6
         279       0.00      0.00      0.00         2
         280       0.00      0.00      0.00         6
         281       0.25      0.25      0.25         8
         282       0.00      0.00      0.00         6
         283       0.00      0.00      0.00         5
         284       0.00      0.00      0.00         1
         285       0.00      0.00      0.00         3
         286       0.10      0.25      0.14         4
         287       0.00      0.00      0.00         4
         288       0.00      0.00      0.00         5
         289       0.00      0.00      0.00         4
         290       0.00      0.00      0.00         3
         291       0.33      0.17      0.22         6
         292       0.00      0.00      0.00         8
         293       0.00      0.00      0.00         2
         294       0.00      0.00      0.00         8
         295       0.00      0.00      0.00         0
         296       0.00      0.00      0.00         3
         297       0.00      0.00      0.00         6
         298       0.00      0.00      0.00         2
         299       0.14      0.25      0.18         4
         300       0.00      0.00      0.00         5
         301       0.00      0.00      0.00         1
         302       0.00      0.00      0.00         4
         303       0.00      0.00      0.00         4
         304       0.00      0.00      0.00         6
         305       0.00      0.00      0.00         2
         306       0.00      0.00      0.00         2
         307       0.00      0.00      0.00         1
         308       0.17      0.25      0.20         4
         309       0.00      0.00      0.00         0
         310       0.33      0.50      0.40         2
         311       0.00      0.00      0.00         2
         312       0.00      0.00      0.00         5
         313       0.33      0.25      0.29         4
         314       0.00      0.00      0.00         4
         315       0.00      0.00      0.00         4
         316       1.00      0.33      0.50         3
         317       0.00      0.00      0.00         3
         318       0.00      0.00      0.00         2
         319       0.50      0.25      0.33         4
         320       0.00      0.00      0.00         3
         321       0.00      0.00      0.00         1
         322       0.00      0.00      0.00         3
         323       0.00      0.00      0.00         4
         324       0.00      0.00      0.00         1
         325       0.00      0.00      0.00         4
         326       0.00      0.00      0.00         2
         327       0.00      0.00      0.00         3
         328       0.11      0.12      0.12         8
         329       0.00      0.00      0.00         2
         330       0.17      0.50      0.25         2
         331       0.00      0.00      0.00         3
         332       0.00      0.00      0.00         6
         333       0.00      0.00      0.00         0
         334       1.00      0.14      0.25         7
         335       0.00      0.00      0.00         6
         336       0.00      0.00      0.00         2
         337       0.00      0.00      0.00         3
         338       0.00      0.00      0.00         3
         339       0.00      0.00      0.00         1
         340       0.00      0.00      0.00         3
         341       0.00      0.00      0.00         3
         342       0.00      0.00      0.00         5
         343       0.00      0.00      0.00         7
         344       0.00      0.00      0.00         1
         345       0.00      0.00      0.00         1
         346       0.00      0.00      0.00         0
         347       0.00      0.00      0.00         6
         348       0.00      0.00      0.00         3
         349       0.00      0.00      0.00         4
         350       0.00      0.00      0.00         6
         351       0.00      0.00      0.00         4
         352       0.00      0.00      0.00         2
         353       0.00      0.00      0.00         4
         354       0.00      0.00      0.00         3
         355       0.00      0.00      0.00         2
         356       0.00      0.00      0.00         7
         357       0.00      0.00      0.00         4
         358       0.00      0.00      0.00         3
         359       0.00      0.00      0.00         4
         360       0.00      0.00      0.00         2
         361       0.00      0.00      0.00         3
         362       0.00      0.00      0.00         4
         363       0.00      0.00      0.00         6
         364       0.00      0.00      0.00         1
         365       0.00      0.00      0.00         6
         366       0.00      0.00      0.00         1
         367       0.00      0.00      0.00         4
         368       0.00      0.00      0.00         0
         369       0.00      0.00      0.00         2
         370       0.00      0.00      0.00         4
         371       0.14      0.50      0.22         2
         372       0.00      0.00      0.00         6
         373       0.00      0.00      0.00         2
         374       0.00      0.00      0.00         7
         375       0.00      0.00      0.00         3
         376       0.00      0.00      0.00         1
         377       0.00      0.00      0.00         4
         378       0.00      0.00      0.00         1
         379       0.00      0.00      0.00         2
         380       0.00      0.00      0.00         1
         381       0.00      0.00      0.00         3
         382       0.12      0.50      0.20         2
         383       0.00      0.00      0.00         1
         384       0.00      0.00      0.00         2
         385       0.50      0.20      0.29         5
         386       0.00      0.00      0.00         1
         387       0.00      0.00      0.00         4
         388       0.00      0.00      0.00         1
         389       0.00      0.00      0.00         4
         390       0.00      0.00      0.00         4
         391       0.00      0.00      0.00         0
         392       0.00      0.00      0.00         1
         393       0.00      0.00      0.00         3
         394       0.00      0.00      0.00         1
         395       0.00      0.00      0.00         2
         396       0.00      0.00      0.00         5
         397       0.00      0.00      0.00         2
         398       0.33      0.25      0.29         4
         399       0.00      0.00      0.00         0
         400       0.00      0.00      0.00         2
         401       0.00      0.00      0.00         3
         402       0.00      0.00      0.00         1
         403       0.00      0.00      0.00         2
         404       0.50      0.25      0.33         4
         405       0.00      0.00      0.00         4
         406       0.00      0.00      0.00         1
         407       0.00      0.00      0.00         0
         408       0.00      0.00      0.00         2
         409       0.00      0.00      0.00         1
         410       0.00      0.00      0.00         1
         411       0.00      0.00      0.00         1
         413       0.00      0.00      0.00         4
         414       0.00      0.00      0.00         1
         415       0.00      0.00      0.00         1
         416       0.00      0.00      0.00         2
         417       0.00      0.00      0.00         1
         418       0.00      0.00      0.00         0
         419       0.00      0.00      0.00         5
         420       0.17      0.33      0.22         3
         421       0.00      0.00      0.00         3
         422       0.00      0.00      0.00         1
         423       0.00      0.00      0.00         2
         425       0.00      0.00      0.00         3
         426       0.00      0.00      0.00         1
         427       0.00      0.00      0.00         0
         428       0.00      0.00      0.00         5
         429       0.00      0.00      0.00         1
         430       0.00      0.00      0.00         4
         431       0.00      0.00      0.00         1
         432       0.00      0.00      0.00         2
         433       0.00      0.00      0.00         2
         434       0.00      0.00      0.00         1
         435       0.00      0.00      0.00         3
         436       0.00      0.00      0.00         2
         437       0.00      0.00      0.00         0
         438       0.00      0.00      0.00         0
         439       0.00      0.00      0.00         1
         440       0.00      0.00      0.00         0
         441       0.00      0.00      0.00         1
         442       0.00      0.00      0.00         1
         443       0.00      0.00      0.00         2
         444       0.00      0.00      0.00         1
         445       0.00      0.00      0.00         3
         446       0.00      0.00      0.00         4
         447       0.00      0.00      0.00         4
         448       0.00      0.00      0.00         1
         449       0.00      0.00      0.00         1
         450       0.00      0.00      0.00         1
         451       0.00      0.00      0.00         3
         452       0.00      0.00      0.00         3
         453       0.00      0.00      0.00         4
         454       0.00      0.00      0.00         3
         455       0.00      0.00      0.00         4
         456       0.00      0.00      0.00         4
         457       0.00      0.00      0.00         2
         458       0.00      0.00      0.00         1
         459       0.00      0.00      0.00         2
         460       0.00      0.00      0.00         2
         461       0.00      0.00      0.00         2
         462       0.00      0.00      0.00         2
         463       0.00      0.00      0.00         4
         464       0.00      0.00      0.00         2
         465       0.00      0.00      0.00         0
         466       0.00      0.00      0.00         6
         467       0.00      0.00      0.00         4
         468       0.00      0.00      0.00         1
```

```
469       0.00       0.00       0.00          2
470       0.00       0.00       0.00          3
471       0.00       0.00       0.00          1
472       0.00       0.00       0.00          1
473       0.00       0.00       0.00          2
474       0.00       0.00       0.00          1
475       0.00       0.00       0.00          0
476       0.00       0.00       0.00          1
477       0.00       0.00       0.00          1
478       0.00       0.00       0.00          2
479       0.00       0.00       0.00          1
480       0.00       0.00       0.00          2
481       0.00       0.00       0.00          3
482       0.00       0.00       0.00          1
483       0.00       0.00       0.00          4
484       0.00       0.00       0.00          1
485       0.00       0.00       0.00          0
486       0.00       0.00       0.00          2
487       0.00       0.00       0.00          1
488       0.00       0.00       0.00          3
489       0.00       0.00       0.00          1
490       0.00       0.00       0.00          3
491       0.00       0.00       0.00          2
492       0.00       0.00       0.00          1
493       0.00       0.00       0.00          1
494       0.00       0.00       0.00          1
495       0.00       0.00       0.00          3
496       0.00       0.00       0.00          2
497       0.00       0.00       0.00          1
498       0.00       0.00       0.00          2
499       0.00       0.00       0.00          1
500       0.00       0.00       0.00          2
501       0.00       0.00       0.00          1
502       0.00       0.00       0.00          0
503       0.00       0.00       0.00          1
504       0.00       0.00       0.00          0
505       0.00       0.00       0.00          2
506       0.00       0.00       0.00          1
507       0.00       0.00       0.00          0
508       0.00       0.00       0.00          0
509       0.00       0.00       0.00          2
511       0.00       0.00       0.00          1
512       0.00       0.00       0.00          3
513       0.00       0.00       0.00          4
514       0.00       0.00       0.00          2
516       0.00       0.00       0.00          2
517       0.00       0.00       0.00          2
518       0.00       0.00       0.00          1
520       0.00       0.00       0.00          4
521       0.00       0.00       0.00          2
522       0.00       0.00       0.00          1
523       0.00       0.00       0.00          2
524       0.00       0.00       0.00          0
525       0.00       0.00       0.00          3
526       0.00       0.00       0.00          1
527       0.00       0.00       0.00          1
529       0.00       0.00       0.00          0
530       0.00       0.00       0.00          1
531       0.00       0.00       0.00          1
532       0.00       0.00       0.00          0
533       0.00       0.00       0.00          1
534       0.00       0.00       0.00          0
536       0.00       0.00       0.00          2
537       0.00       0.00       0.00          2
538       0.00       0.00       0.00          1
539       0.00       0.00       0.00          3
540       0.00       0.00       0.00          0
541       0.00       0.00       0.00          2
543       0.00       0.00       0.00          1
```

```
         544       0.00      0.00      0.00         1
         545       0.00      0.00      0.00         2
         546       0.00      0.00      0.00         2
         547       0.00      0.00      0.00         3
         549       0.00      0.00      0.00         1
         550       0.00      0.00      0.00         2
         551       0.00      0.00      0.00         0
         552       0.00      0.00      0.00         1
         553       1.00      1.00      1.00         1
         554       0.00      0.00      0.00         0
         555       0.50      0.50      0.50         2
         556       1.00      1.00      1.00         1
         557       0.00      0.00      0.00         0
         558       0.00      0.00      0.00         3
         559       0.00      0.00      0.00         1
         560       0.00      0.00      0.00         0
         561       0.00      0.00      0.00         0
         562       0.00      0.00      0.00         1
         563       0.00      0.00      0.00         0
         564       0.00      0.00      0.00         3
         565       0.00      0.00      0.00         1
         566       0.00      0.00      0.00         3
         567       0.00      0.00      0.00         0
         568       0.00      0.00      0.00         3
         569       0.00      0.00      0.00         3
         570       0.00      0.00      0.00         1
         571       0.00      0.00      0.00         3
         572       0.00      0.00      0.00         2
         573       0.00      0.00      0.00         1
         575       0.00      0.00      0.00         1
         576       0.00      0.00      0.00         1
         577       0.00      0.00      0.00         1
         578       0.00      0.00      0.00         1
         579       0.00      0.00      0.00         5
         580       0.00      0.00      0.00         0
         581       0.00      0.00      0.00         1
         582       0.00      0.00      0.00         0
         584       0.00      0.00      0.00         1
         585       0.00      0.00      0.00         4
         586       0.00      0.00      0.00         3
         588       0.00      0.00      0.00         1
         589       0.00      0.00      0.00         1
         590       0.00      0.00      0.00         1
         591       0.00      0.00      0.00         1
         592       0.00      0.00      0.00         1
         593       0.00      0.00      0.00         1
         594       0.00      0.00      0.00         1
         595       0.00      0.00      0.00         1
         596       0.00      0.00      0.00         1
         597       0.00      0.00      0.00         0
         598       0.00      0.00      0.00         1
         600       0.00      0.00      0.00         0
         601       0.00      0.00      0.00         1
         602       0.00      0.00      0.00         2
         603       0.00      0.00      0.00         1
         604       0.00      0.00      0.00         1
         607       0.00      0.00      0.00         0
         608       0.00      0.00      0.00         1
         610       0.00      0.00      0.00         2
         611       0.00      0.00      0.00         0
         613       0.00      0.00      0.00         1
         614       0.00      0.00      0.00         0
         615       0.00      0.00      0.00         3
         616       0.00      0.00      0.00         1
         617       0.00      0.00      0.00         3
         618       0.00      0.00      0.00         1
         619       0.00      0.00      0.00         1
         620       0.00      0.00      0.00         1
         623       0.00      0.00      0.00         0
         626       0.00      0.00      0.00         1
         627       0.00      0.00      0.00         0
         628       0.00      0.00      0.00         0
         631       0.00      0.00      0.00         2
         632       0.00      0.00      0.00         1
         633       1.00      1.00      1.00         1
         634       0.00      0.00      0.00         1
         635       0.00      0.00      0.00         0
         636       0.00      0.00      0.00         0
         638       0.00      0.00      0.00         0
         639       0.00      0.00      0.00         1
         640       0.00      0.00      0.00         0
         641       0.00      0.00      0.00         1
         642       0.00      0.00      0.00         0
         643       0.00      0.00      0.00         0
         646       0.00      0.00      0.00         0
         647       0.00      0.00      0.00         2
         648       0.00      0.00      0.00         1
         649       0.00      0.00      0.00         2
         650       0.00      0.00      0.00         0
         651       0.00      0.00      0.00         1
         652       0.00      0.00      0.00         1
         653       0.00      0.00      0.00         1
         654       0.00      0.00      0.00         2
         655       0.00      0.00      0.00         1
         656       0.00      0.00      0.00         1
         658       0.00      0.00      0.00         0
         659       0.00      0.00      0.00         2
         660       0.00      0.00      0.00         1
         662       0.00      0.00      0.00         1
         665       0.00      0.00      0.00         0
         667       0.00      0.00      0.00         0
         668       0.33      0.50      0.40         2
         669       0.00      0.00      0.00         1
         671       0.00      0.00      0.00         2
         673       0.00      0.00      0.00         1
         676       0.00      0.00      0.00         2
         678       0.00      0.00      0.00         3
         679       0.00      0.00      0.00         1
         681       0.00      0.00      0.00         2
         682       0.00      0.00      0.00         1
         683       0.00      0.00      0.00         1
         684       0.00      0.00      0.00         1
         685       0.00      0.00      0.00         1
         686       0.00      0.00      0.00         1
         687       0.00      0.00      0.00         1
         688       0.00      0.00      0.00         0
         689       0.00      0.00      0.00         1
         690       0.00      0.00      0.00         1
         692       0.00      0.00      0.00         0
         693       0.00      0.00      0.00         1
         694       0.00      0.00      0.00         1
         696       0.00      0.00      0.00         1
         698       0.00      0.00      0.00         0
         701       0.00      0.00      0.00         0
         702       0.00      0.00      0.00         0
         704       0.50      1.00      0.67         1
         708       0.00      0.00      0.00         0
         710       0.00      0.00      0.00         1
         712       0.00      0.00      0.00         0
         713       0.00      0.00      0.00         2
         715       0.00      0.00      0.00         0
         721       0.00      0.00      0.00         1
         723       0.00      0.00      0.00         1
         724       0.00      0.00      0.00         1
         725       0.00      0.00      0.00         1
         729       0.00      0.00      0.00         1
         730       0.00      0.00      0.00         1
         731       0.00      0.00      0.00         2
         734       0.00      0.00      0.00         1
         738       0.00      0.00      0.00         0
         741       0.00      0.00      0.00         0
         743       1.00      0.33      0.50         3
         744       0.00      0.00      0.00         0
         745       0.00      0.00      0.00         1
         747       0.00      0.00      0.00         0
         748       0.00      0.00      0.00         0
         749       0.00      0.00      0.00         1
         750       0.00      0.00      0.00         0
         757       0.00      0.00      0.00         1
         759       0.00      0.00      0.00         2
         761       0.00      0.00      0.00         0
         766       0.00      0.00      0.00         0
         767       0.00      0.00      0.00         0
         771       0.00      0.00      0.00         1
         776       0.00      0.00      0.00         1
         777       0.00      0.00      0.00         1
         782       0.00      0.00      0.00         1
         783       0.00      0.00      0.00         1
         784       0.00      0.00      0.00         1
         788       0.00      0.00      0.00         1
         790       0.00      0.00      0.00         0
         791       0.00      0.00      0.00         1
         794       0.00      0.00      0.00         0
         795       0.00      0.00      0.00         2
         798       0.00      0.00      0.00         1
         800       0.00      0.00      0.00         0
         806       0.00      0.00      0.00         1
         809       0.00      0.00      0.00         1
         811       0.00      0.00      0.00         1
         812       0.00      0.00      0.00         0
         814       0.00      0.00      0.00         0
         817       0.00      0.00      0.00         1
         818       0.00      0.00      0.00         1
         822       0.00      0.00      0.00         0
         823       0.00      0.00      0.00         1
         827       0.00      0.00      0.00         1
         831       0.00      0.00      0.00         1
         832       0.00      0.00      0.00         1
         834       0.00      0.00      0.00         1
         835       0.00      0.00      0.00         1
         837       0.00      0.00      0.00         0
         838       0.00      0.00      0.00         0
         839       0.00      0.00      0.00         0
         842       0.00      0.00      0.00         1
         846       0.00      0.00      0.00         1
         848       0.00      0.00      0.00         0
         849       0.00      0.00      0.00         1
         850       0.00      0.00      0.00         1
         851       0.00      0.00      0.00         1
         857       0.00      0.00      0.00         0
         858       0.00      0.00      0.00         0
         862       0.00      0.00      0.00         0
         863       0.00      0.00      0.00         1
         868       0.00      0.00      0.00         1
         869       0.00      0.00      0.00         1
         873       0.00      0.00      0.00         0
         888       0.00      0.00      0.00         1
         890       0.00      0.00      0.00         0
         894       0.00      0.00      0.00         0
         948       0.00      0.00      0.00         0
         977       0.00      0.00      0.00         1

    accuracy                           0.23      3266
   macro avg       0.07      0.07      0.07      3266
weighted avg       0.24      0.23      0.23      3266
```

For classification model, it uses accuracy score, recall score, precision score and f1 score for evaluating the model prediction results. The accuracy rate is the proportion of correctly classified samples to the total number of samples. The precision rate refers to the proportion of the samples that are predicted to be positive by the model that are actually positive to the samples that are predicted to be positive. Recall score refers to the proportion of actually positive samples to actually positive samples in the samples that are actually positive. F1 score is the harmonic mean of precision and recall score. Precision embodies the model's ability to distinguish negative samples. The higher the Precision score, the stronger the model's ability to distinguish negative samples. Recall embodies the model's ability to recognize positive samples. The higher the Recall score, the stronger the model's ability to recognize positive samples. The F1 score is a combination of the two. The model is more robust with higher F1 score. In this case, the classification method is not appropriate for the prediction as the score of the model is not good.

From the three classification models above, the decision tree classifier model has the highest accuracy and precision score that the model can predict whereas the knn method has the lowest score that the model performs worst.Sometimes, the result of precision score and recall score may be contradict. The F1 score can provide the balance of these two scores.When both the accuracy and the recall rate are high, the F1 value will also be high. The F1 value reaches the best value at 1 with perfect precision and recall score,and the worst is 0. The decision tree also has the highest F1 score among three models thus this model performs well among these three models, however, the score is quite low for prediction.This may be owing to the class-imbalance and not well classified data.It could use the cost function to learn the weight of each class. It can set the reciprocal of the weight of each category to the number of samples, and then uses oversampling for tuning.

6. Comparison About Prediction Model

It has made the regression prediction and classification prediction separately. The use of regression or classification technique depends on the continuous output y or discrete y. In this case, the prediction result is the bike sharing demand in the numerical shape so it is the consistent value that regression is more applicable. According to the result, the classification result is not good as the data in this type is not appropriate for the classification prediction while the regression technique has higher accuracy. SVM could minimize the structural risk. The decision tree is a simulation of human decision-making process, and the model is trained through experience, that is the loss of minimize data. So in this case, it would use random forest regressor for forecasting the bike sharing demand.

Random forest is also considered a very convenient and easy-to-use algorithm because its default hyperparameters usually produce good prediction results. In addition, in order to improve the accuracy of random forest regressor,it could select less and more correlated features，but not so strong correlation to train the prediction model.Random forest can deal with the variables with strong correlation to a certain extent, but it is recommended to eliminate the variables with strong correlation to reduce the effect.

7. Conclusion

During the 20th century, the government mainly focuses on the wealth and property development, however, it switches to the more sustainable development as the urgent necessity. To conclude, people has increasing demand for the public bike as it provides great convenience for the daily life. The total number is influenced a lot by the weather condition and datetime. It also tried to make the prediction of bike sharing demand using various techniques.The random forest regressor would be used of highest accuray so that can help both companies and government for the public bike management, thus to promote the development of the sustainable public transport system. In addition, further related data can be obtained to help the analysis, such as the public attitude towards the use of the public bike system, the distribution of the bike within the city and even across the country, the money people spent on and the budget that the companies need to invest (Fricker and Gast, 2012). There are some suggestions for the public bike system improvement. First of all, the government plays an important role in management that they need to combine the actual development situation when formulating relevant management policies (Contrado, Morency and Rousseau,2012). The companies can invest money from large city to some other city in the long term. For example, during the peak hours of the work, the tidal effect is particularly serious. A large number of bicycles flock to bus or subway station and this may be occupying the public areas causing some problems. In view of this situation, big data can be used for intelligent scheduling to intervene in tidal effects.

Reference List

Chemla.D, Meunier. F and Wolfler. C. R.(2011). Bike hiring system: solving the rebalancing provlem in the static case.Discrete Optimiztion. Contrado.C, Morency.C and Rousseau.L.M.(2012). Balancing a dynamic public bike-sharing system.[Online].Avaiable from: https://www.cirrrelt.ca/DocumentsTravail/CIRRELT-2012-09.pdf [Accessed:26/04/2020] DeMiao.P.(2009).Bike-sharing:history,impacts,models of provision, and future. J Public Transp 12(4):41-56. Fricker.C and Gast.N (2012). Incentives and regulations in bike-sharing systems with statioins of finite capacity.[Online].Avaiable from: https://arxiv.org/pdf/1201.1178v1.pdf.[Accessed: 28/04/2020] Lin.J.R and Yang.T.H.(2009).On the one-commodity pickup-and-delivery travelling salesman problem with stochastic demands. Math Prog Ser A 119:169-194. Raviv.T, Tzur.M. and Forma.I.(2012). Static repositioning in a bike-sharing system: models and solution approaches.Unabridged version, Industrial Engineering department, Tel Aviv Univeristy, Israel.