

mxycbsyvg

January 23, 2025

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as py
```

```
[3]: data = pd.read_csv(r"C:\Users\91703\Downloads\bike dataset.csv")
```

```
[4]: data.head()
```

```
[4]:
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
0	1	01-01-2011	1	0	1	0	6	0	
1	2	02-01-2011	1	0	1	0	0	0	
2	3	03-01-2011	1	0	1	0	1	1	
3	4	04-01-2011	1	0	1	0	2	1	
4	5	05-01-2011	1	0	1	0	3	1	

	weathersit	temp	atemp	hum	windspeed	casual	registered	\
0	2	0.344167	0.363625	0.805833	0.160446	331	654	
1	2	0.363478	0.353739	0.696087	0.248539	131	670	
2	1	0.196364	0.189405	0.437273	0.248309	120	1229	
3	1	0.200000	0.212122	0.590435	0.160296	108	1454	
4	1	0.226957	0.229270	0.436957	0.186900	82	1518	

	cnt
0	985
1	801
2	1349
3	1562
4	1600

```
[5]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
#   Column      Non-Null Count  Dtype
---  -
0   instant     731 non-null   int64
1   dteday      731 non-null   object
```

```

2  season      731 non-null   int64
3  yr          731 non-null   int64
4  mnth        731 non-null   int64
5  holiday     731 non-null   int64
6  weekday     731 non-null   int64
7  workingday  731 non-null   int64
8  weathersit   731 non-null   int64
9  temp        731 non-null   float64
10 atemp       731 non-null   float64
11 hum         731 non-null   float64
12 windspeed   731 non-null   float64
13 casual      731 non-null   int64
14 registered  731 non-null   int64
15 cnt         731 non-null   int64
dtypes: float64(4), int64(11), object(1)
memory usage: 91.5+ KB

```

```
[7]: data.describe()
```

```

[7]:
count    instant      season      yr      mnth      holiday      weekday  \
count    731.000000    731.000000    731.000000    731.000000    731.000000    731.000000
mean      366.000000      2.496580      0.500684      6.519836      0.028728      2.997264
std       211.165812      1.110807      0.500342      3.451913      0.167155      2.004787
min         1.000000      1.000000      0.000000      1.000000      0.000000      0.000000
25%       183.500000      2.000000      0.000000      4.000000      0.000000      1.000000
50%       366.000000      3.000000      1.000000      7.000000      0.000000      3.000000
75%       548.500000      3.000000      1.000000     10.000000      0.000000      5.000000
max       731.000000      4.000000      1.000000     12.000000      1.000000      6.000000

count    workingday  weathersit      temp      atemp      hum      windspeed  \
count    731.000000    731.000000    731.000000    731.000000    731.000000    731.000000
mean       0.683995      1.395349      0.495385      0.474354      0.627894      0.190486
std       0.465233      0.544894      0.183051      0.162961      0.142429      0.077498
min       0.000000      1.000000      0.059130      0.079070      0.000000      0.022392
25%       0.000000      1.000000      0.337083      0.337842      0.520000      0.134950
50%       1.000000      1.000000      0.498333      0.486733      0.626667      0.180975
75%       1.000000      2.000000      0.655417      0.608602      0.730209      0.233214
max       1.000000      3.000000      0.861667      0.840896      0.972500      0.507463

count    casual    registered      cnt
count    731.000000    731.000000    731.000000
mean     848.176471    3656.172367    4504.348837
std     686.622488    1560.256377    1937.211452
min       2.000000      20.000000     22.000000
25%     315.500000    2497.000000    3152.000000
50%     713.000000    3662.000000    4548.000000
75%    1096.000000    4776.500000    5956.000000

```

```
max      3410.000000  6946.000000  8714.000000
```

```
[8]: data.isnull()
```

```
[8]:      instant  dteday  season    yr  mnth  holiday  weekday  workingday  \
0      False   False   False  False False  False   False    False      False
1      False   False   False  False False  False   False    False      False
2      False   False   False  False False  False   False    False      False
3      False   False   False  False False  False   False    False      False
4      False   False   False  False False  False   False    False      False
..      ...      ...      ...      ...  ...      ...      ...      ...
726    False   False   False  False False  False   False    False      False
727    False   False   False  False False  False   False    False      False
728    False   False   False  False False  False   False    False      False
729    False   False   False  False False  False   False    False      False
730    False   False   False  False False  False   False    False      False
```

```
      weathersit  temp  atemp  hum  windspeed  casual  registered  cnt
0      False  False  False  False    False    False    False    False
1      False  False  False  False    False    False    False    False
2      False  False  False  False    False    False    False    False
3      False  False  False  False    False    False    False    False
4      False  False  False  False    False    False    False    False
..      ...      ...      ...      ...      ...      ...      ...
726    False  False  False  False    False    False    False    False
727    False  False  False  False    False    False    False    False
728    False  False  False  False    False    False    False    False
729    False  False  False  False    False    False    False    False
730    False  False  False  False    False    False    False    False
```

```
[731 rows x 16 columns]
```

```
[9]: data.isnull().sum()
```

```
[9]: instant      0
      dteday      0
      season      0
      yr          0
      mnth        0
      holiday     0
      weekday     0
      workingday  0
      weathersit   0
      temp        0
      atemp       0
      hum         0
      windspeed   0
```

```
casual      0
registered  0
cnt         0
dtype: int64
```

```
[10]: #What is the average temperature (temp) in the dataset?**
average_temp = data['temp'].mean()
print("Average Temperature:", average_temp)
```

Average Temperature: 0.495384788508892

```
[18]: # What is the distribution of season in the dataset?
season_distribution = data['season'].value_counts()
print("Season Distribution:")
print(season_distribution)
```

Season Distribution:

```
season
3      188
2      184
1      181
4      178
```

Name: count, dtype: int64

```
[14]: #What is the total count of bike users (cnt) on weekdays?
weekday_cnt = data[data['weekday'] != 0]['cnt'].sum()
print("Total Weekday Count:", weekday_cnt)
```

Total Weekday Count: 2848652

```
[15]: #What is the average wind speed (windspeed) in the dataset?
average_windspeed = data['windspeed'].mean()
print("Average Wind Speed:", average_windspeed)
```

Average Wind Speed: 0.190486211627907

```
[16]: #What is the total count of registered bike users (registered) in the dataset?
total_registered = data['registered'].sum()
print("Total Registered Count:", total_registered)
```

Total Registered Count: 2672662

```
[17]: #What is the distribution of weekday in the dataset?
weekday_distribution = data['weekday'].value_counts()
print("Weekday Distribution:")
print(weekday_distribution)
```

Weekday Distribution:

```
weekday
6      105
0      105
1      105
2      104
3      104
4      104
5      104
Name: count, dtype: int64
```

```
[22]: #What is the average yearly bike count (cnt) in the dataset?
data.groupby('yr')['cnt'].mean()
```

```
[22]: yr
0      3405.761644
1      5599.934426
Name: cnt, dtype: float64
```

```
[23]: #What is the total count of bike users (cnt) on holidays?
holiday_cnt = data[data['holiday'] == 1]['cnt'].sum()
print("Total Holiday Count:", holiday_cnt)
```

Total Holiday Count: 78435

```
[24]: #What is the distribution of weathersit (weather situation) in the dataset?
weathersit_distribution = data['weathersit'].value_counts()
print("Weather Situation Distribution:")
print(weathersit_distribution)
```

```
Weather Situation Distribution:
weathersit
1      463
2      247
3       21
Name: count, dtype: int64
```

```
[25]: #What is the average temperature (temp) when bike count (cnt) is highest?
max_cnt_temp = data.loc[data['cnt'].idxmax()]['temp']
print("Average Temp at Max Bike Count:", max_cnt_temp)
```

Average Temp at Max Bike Count: 0.608333