yulu-buss-case-study

April 19, 2024

Yulu, India's pioneering micro-mobility service provider, has embarked on a mission to revolutionize daily commutes by offering unique, sustainable transportation solutions. However, recent revenue setbacks have prompted Yulu to seek the expertise of a consulting company to delve into the factors influencing the demand for their shared electric cycles, specifically in the Indian market.

[8]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[9]:
```
df= pd.read_csv('F:\\buss_cass\\data\\bike_sharing.txt')
```

[10]:
```
df.head()
```

[10]:

	datetime	season	holiday	workingday	weather	temp	atemp	\
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	

	humidity	windspeed	casual	registered	count
0	81	0.0	3	13	16
1	80	0.0	8	32	40
2	80	0.0	5	27	32
3	75	0.0	3	10	13
4	75	0.0	0	1	1

[11]:
```
df.shape
```

[11]:
```
(10886, 12)
```

[12]:
```
df.isnull().sum()
```

[12]:
```
datetime      0
season        0
holiday       0
```

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

[13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10886 entries, 0 to 10885
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   datetime    10886 non-null  object
 1   season      10886 non-null  int64
 2   holiday     10886 non-null  int64
 3   workingday  10886 non-null  int64
 4   weather     10886 non-null  int64
 5   temp        10886 non-null  float64
 6   atemp       10886 non-null  float64
 7   humidity    10886 non-null  int64
 8   windspeed   10886 non-null  float64
 9   casual      10886 non-null  int64
 10  registered  10886 non-null  int64
 11  count       10886 non-null  int64
dtypes: float64(3), int64(8), object(1)
memory usage: 1020.7+ KB
```

Datatype of following attributes needs to change to proper data type

datetime - to datetime, season - to categorical, holiday - to categorical, workingday - to categorical, weather - to categorical

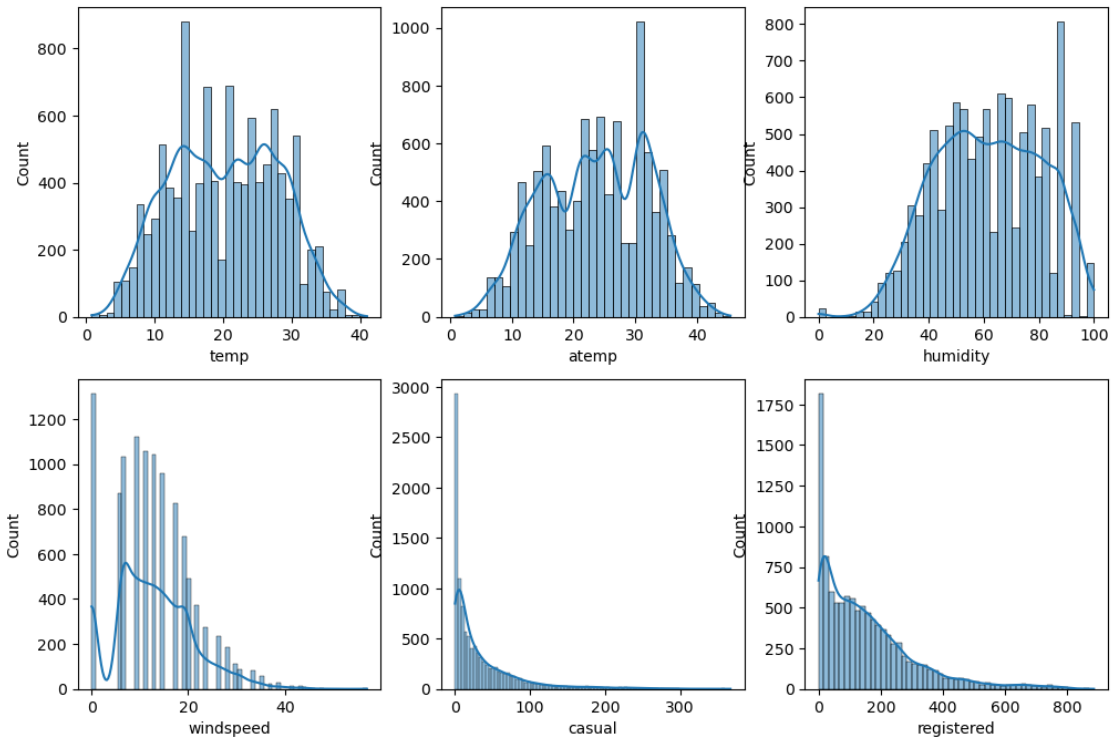
[14]:
```
df['datetime'] = pd.to_datetime(df['datetime'])
cat_cols= ['season', 'holiday', 'workingday', 'weather']
for col in cat_cols:
    df[col] = df[col].astype('object')
```

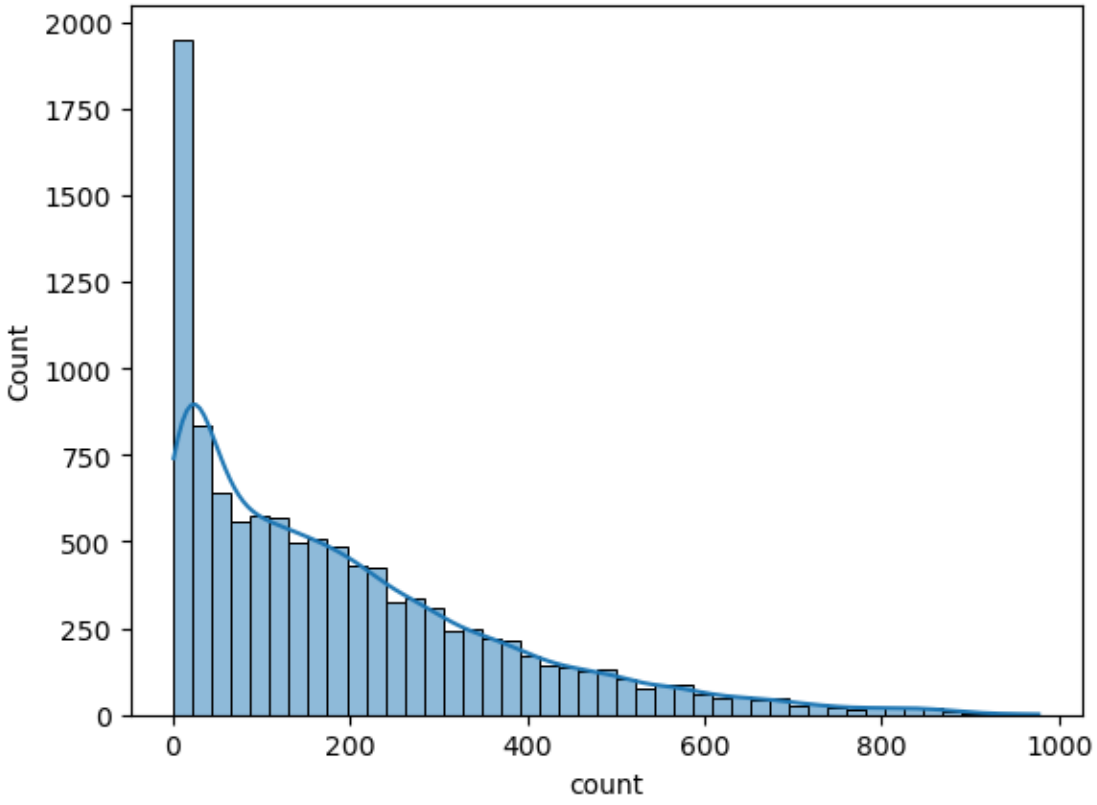
[71]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10886 entries, 0 to 10885
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype
```

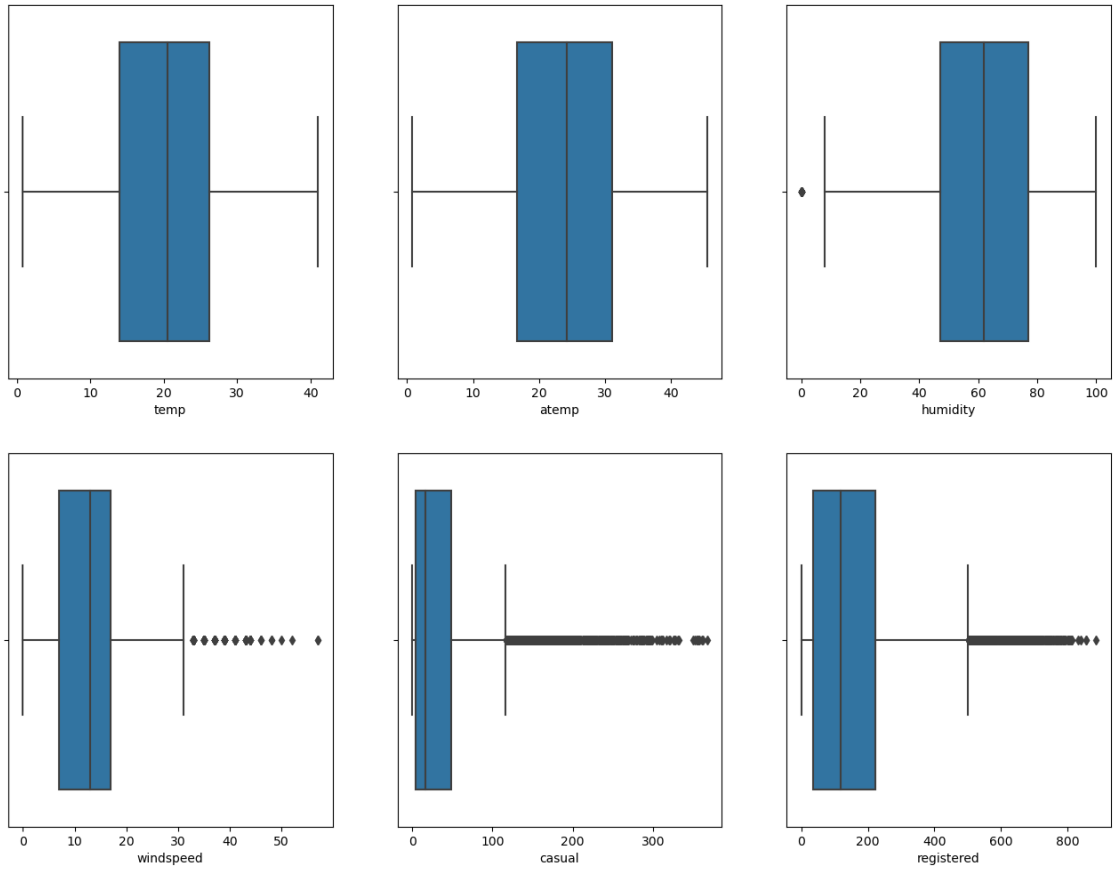
```
---  ------      --------------  -----
 0   datetime    10886 non-null  datetime64[ns]
 1   season      10886 non-null  object
 2   holiday     10886 non-null  object
 3   workingday  10886 non-null  object
 4   weather     10886 non-null  object
 5   temp        10886 non-null  float64
 6   atemp       10886 non-null  float64
 7   humidity    10886 non-null  int64
 8   windspeed   10886 non-null  float64
 9   casual      10886 non-null  int64
 10  registered  10886 non-null  int64
 11  count       10886 non-null  int64
dtypes: datetime64[ns](1), float64(3), int64(4), object(4)
memory usage: 1020.7+ KB
```

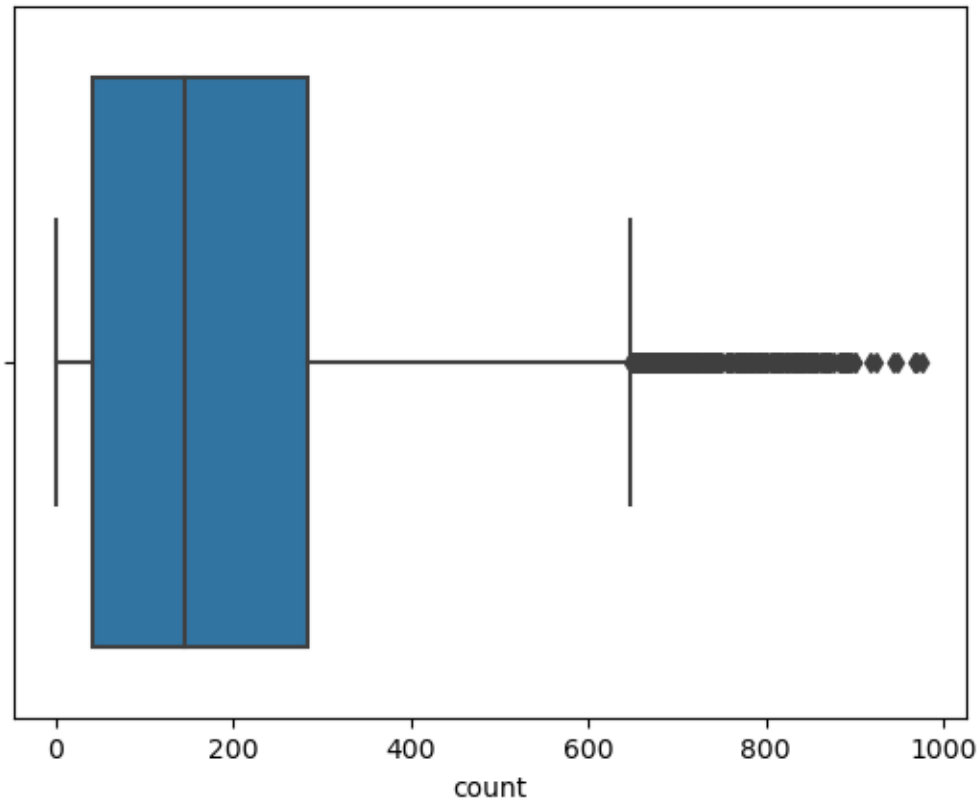
[78]:
```
num_cols = ['temp', 'atemp', 'humidity', 'windspeed', 'casual',
'registered','count']
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(12, 8))
index = 0
for row in range(2):
    for col in range(3):
        sns.histplot(df[num_cols[index]], ax=axis[row, col], kde=True)
        index += 1
plt.show()
sns.histplot(df[num_cols[-1]], kde=True)
plt.show()
```


Count
temp
atemp
humidity
windspeed
casual
registered


Count
count

[81]:
```
# plotting box plots to detect outliers in the data
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(16, 12))
index = 0
for row in range(2):
    for col in range(3):
        sns.boxplot(x=df[num_cols[index]], ax=axis[row, col])
        index += 1
plt.show()
sns.boxplot(x=df[num_cols[-1]])
plt.show()
```

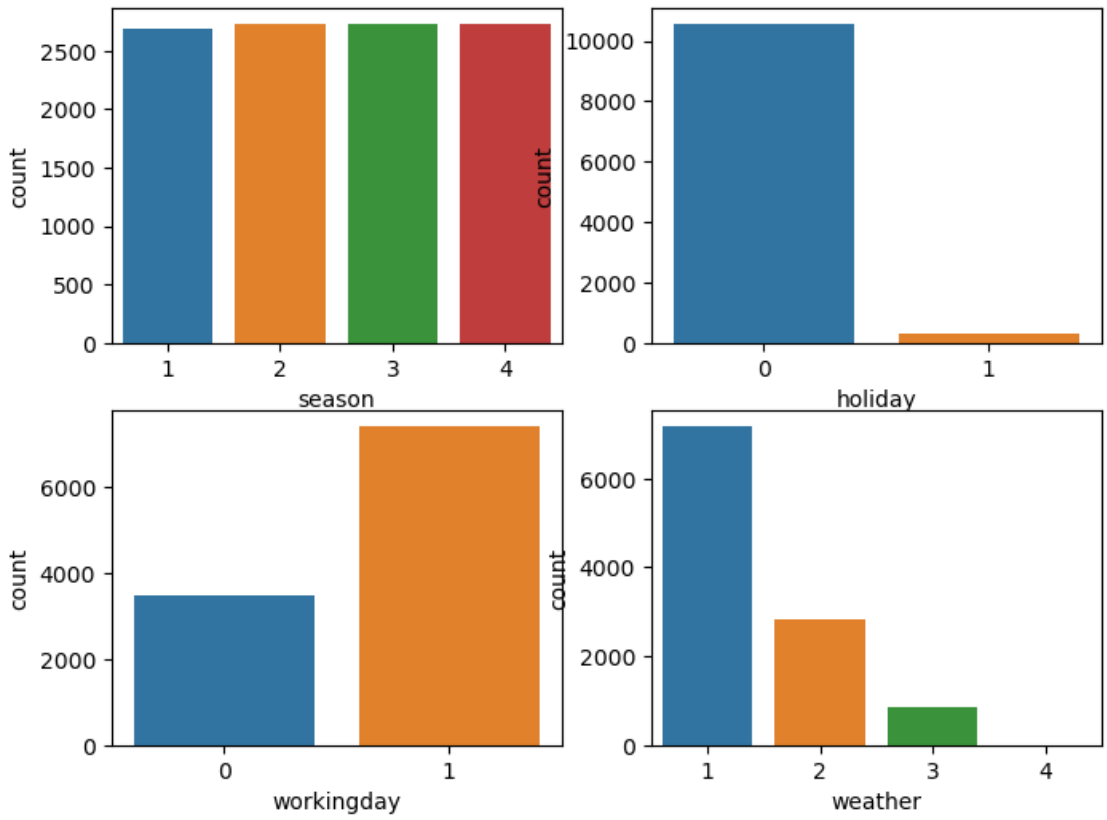




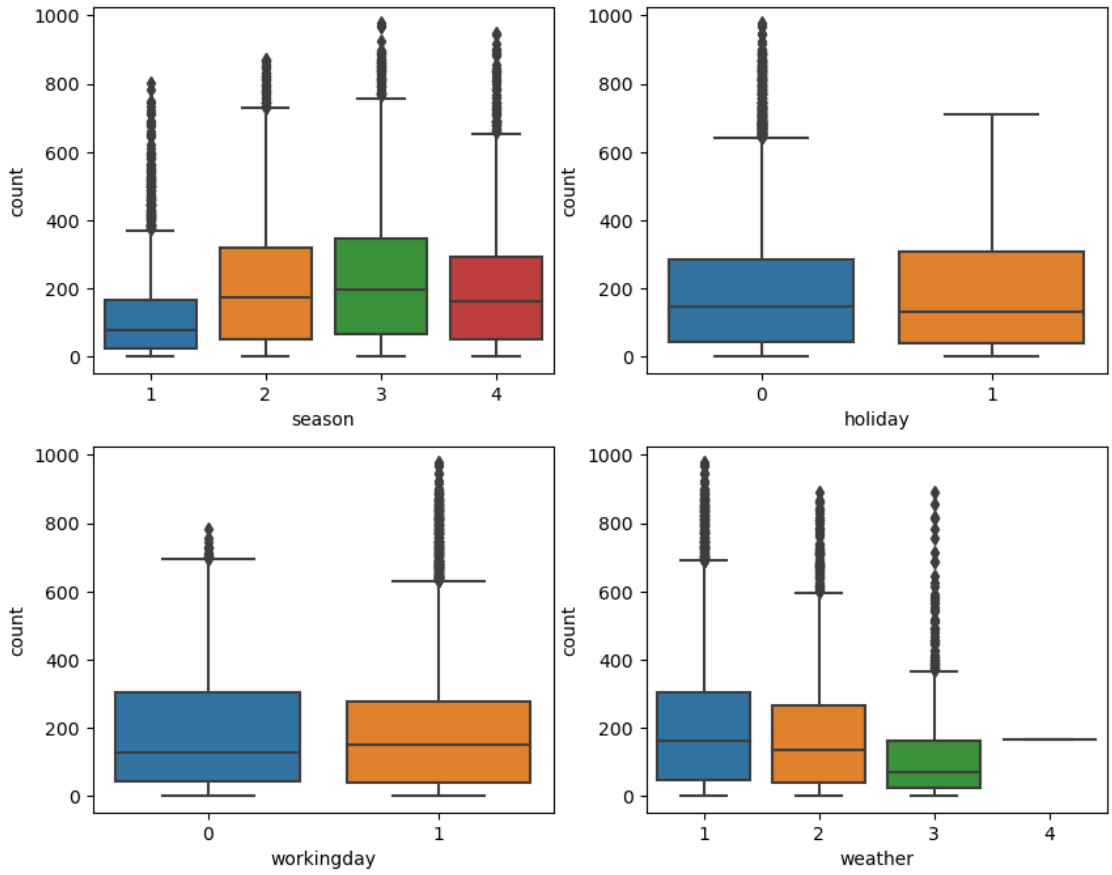
humidity, casual, registered and count have outliers in the data.

[ ]:

[82]:
```
# countplot of each categorical column
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(8, 6))
index = 0
for row in range(2):
    for col in range(2):
        sns.countplot(data=df, x=cat_cols[index], ax=axis[row, col])
        index += 1
plt.show()
```



[83]:
```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(10, 8))
index = 0
for row in range(2):
    for col in range(2):
        sns.boxplot(data=df, x=cat_cols[index], y='count', ax=axis[row,col])
        index += 1
plt.show()
```



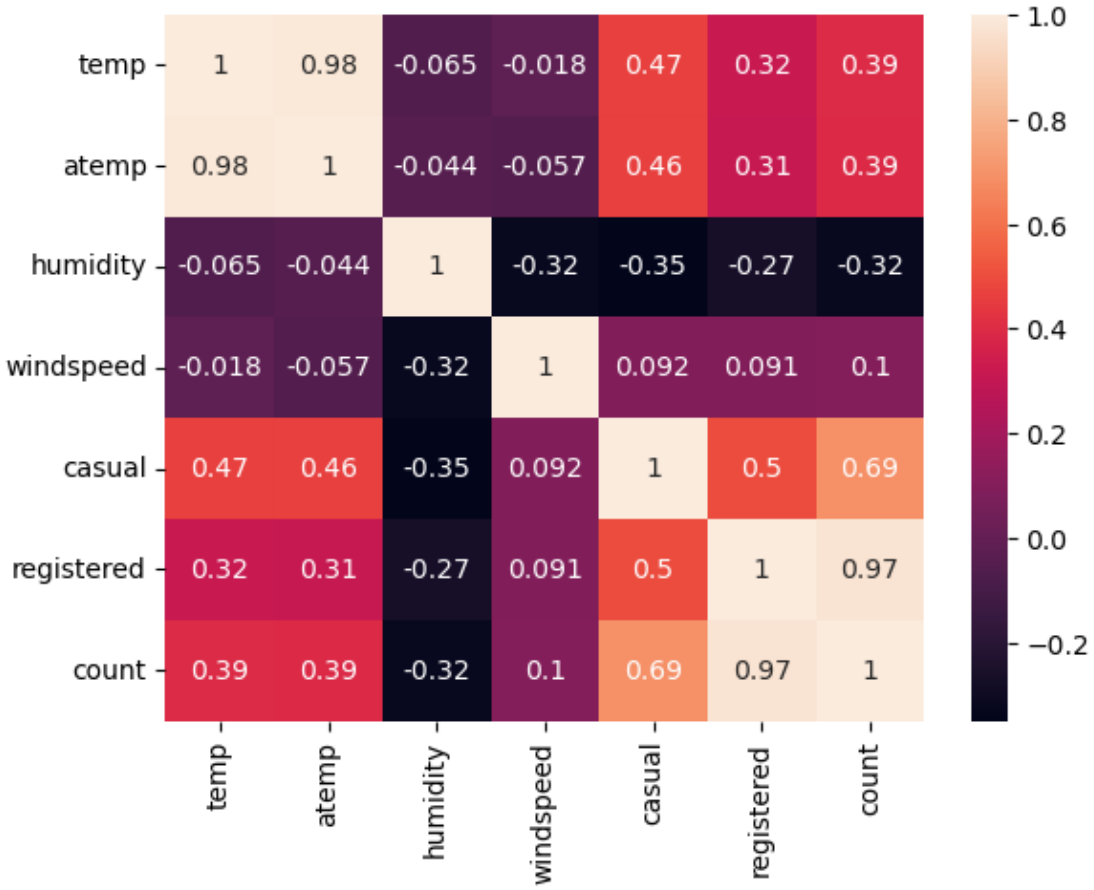
Corelation

[76]:
```
df_group =␣
↪df[['temp','atemp','humidity','windspeed','casual','registered','count']]
df_group.head()
```

[76]:

	temp	atemp	humidity	windspeed	casual	registered	count
0	9.84	14.395	81	0.0	3	13	16
1	9.02	13.635	80	0.0	8	32	40
2	9.02	13.635	80	0.0	5	27	32
3	9.84	14.395	75	0.0	3	10	13
4	9.84	14.395	75	0.0	0	1	1

[77]:
```
df_group.corr()['count']
sns.heatmap(df_group.corr(), annot=True)
plt.show()
```



[ ]:

Hypothesis Testing Two Sample T-test: Checking there any significant difference between the no. of bike rides on Weekdays and Weekends? H0 : There is no siginficant difference between the no of rides on weekdays and weekends H1 : There is a siginficant difference between the no of rides on weekdays and weekends Significance level (alpha): 0.05

[22]:
```
df['workingday'].value_counts()
```

[22]:
```
workingday
1    7412
0    3474
Name: count, dtype: int64
```

[23]:
```
weekdays = df[df['workingday']==0]['count'].values
weekends = df[df['workingday']==1]['count'].values
```

[26]:
```
from scipy.stats import ttest_ind
p = ttest_ind(weekdays, weekends, alternative = "two-sided")
```

```
p
```

[26]:
```
TtestResult(statistic=-1.2096277376026694, pvalue=0.22644804226361348,
df=10884.0)
```

p >= aplha p value is high Therefore we cannot reject null We don't have the sufficient evidence to say that working day has effect on the number of cycles being rented.

-ANNOVA to check if No. of cycles rented is similar or different in different weather & Season

H0: Number of cycles rented is similar in different weather and season. H1: Number of cycles rented is not similar in different weather and season. Significance level (alpha): 0.05

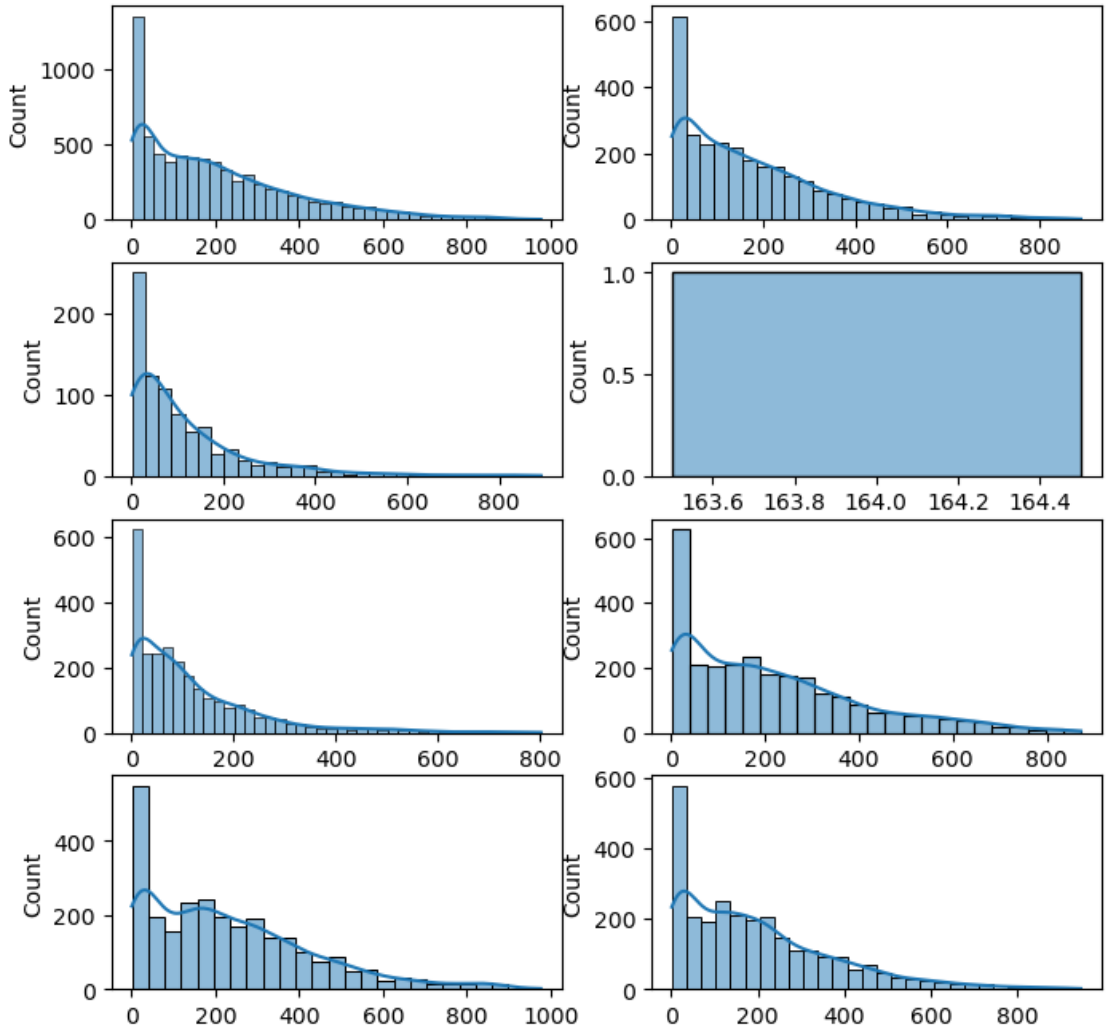
[45]:
```
group = df[['weather','season']].value_counts()
group
```

[45]:
```
weather  season
1        3         1930
         2         1801
         1         1759
         4         1702
2        4          807
         1          715
         2          708
         3          604
3        4          225
         2          224
         1          211
         3          199
4        1            1
Name: count, dtype: int64
```

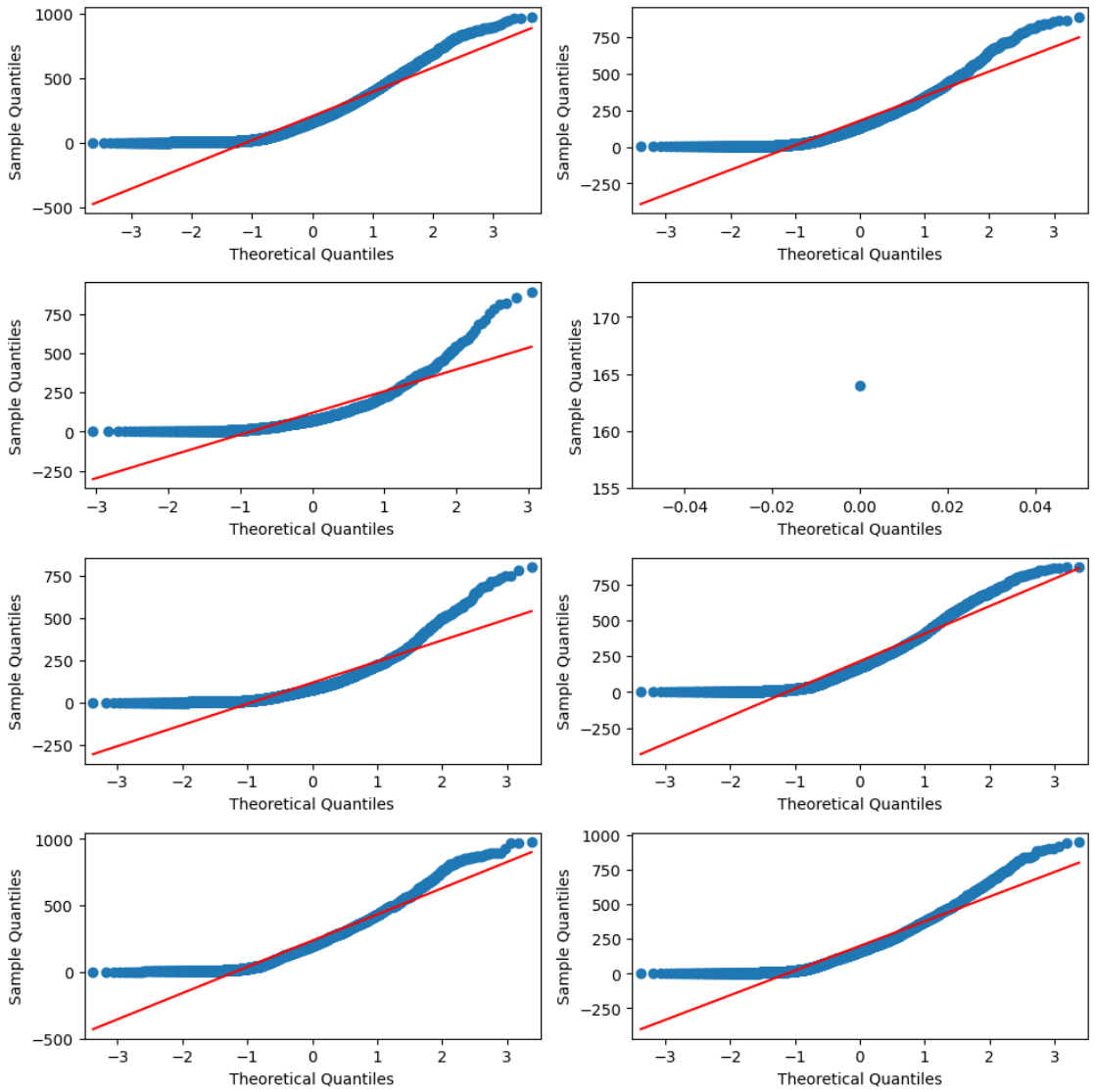
[56]:
```
gp1 = df[df['weather']==1]['count'].values
gp2 = df[df['weather']==2]['count'].values
gp3 = df[df['weather']==3]['count'].values
gp4 = df[df['weather']==4]['count'].values

gp5 = df[df['season']==1]['count'].values
gp6 = df[df['season']==2]['count'].values
gp7 = df[df['season']==3]['count'].values
gp8 = df[df['season']==4]['count'].values
groups=[gp1,gp2,gp3,gp4,gp5,gp6,gp7,gp8]
```

[57]:
```
fig, axis = plt.subplots(nrows=4, ncols=2, figsize=(8, 8))

index = 0
for row in range(4):
    for col in range(2):
```

```
        sns.histplot(groups[index], ax=axis[row, col], kde=True)
        index += 1

plt.show()
```



[65]:
```
fig, axs = plt.subplots(4, 2, figsize=(10, 10))

index = 0
for row in range(4):
    for col in range(2):
        qqplot(groups[index], line="s", ax=axs[row, col])
        index += 1

plt.tight_layout()
plt.show()
```



As per above graphs, all groups are not following Gaussian distribution Data is Independent ### Levene's Test H0: Variances is similar in different weather and season. H1: Variances is not similar in different weather and season. Significance level (alpha): 0.05

[68]:
```
from scipy.stats import levene
levene(gp1,gp2,gp3,gp4,gp5,gp6,gp7,gp8)
```

[68]:
```
LeveneResult(statistic=102.5026306304148, pvalue=3.463531888897594e-148)
```

p_value: 3.463531888897594e-148 Reject the Null hypothesis. Variances are not equal As per QQ plot and Levene's Test, We cannot ANOVA Test.

ANOVA fail, use Kruskal

[69]:
```
from scipy.stats import kruskal
kruskal(gp1,gp2,gp3,gp4,gp5,gp6,gp7,gp8)
```

[69]:
```
KruskalResult(statistic=904.7105757287106, pvalue=4.614440933900297e-191)
```

As p value is low we can reject H0, therefore Number of cycles rented is not similar in different weather and season

[ ]:
```

```

Chi-square test to check if Weather is dependent on the season

H0: Weather is independent of the season H1: Weather is not independent of the season Significance level (alpha): 0.05

[66]:
```
data_table = pd.crosstab(df['season'], df['weather'])
data_table
```

[66]:

weather	1	2	3	4
season				
1	1759	715	211	1
2	1801	708	224	0
3	1930	604	199	0
4	1702	807	225	0

[42]:
```
from scipy.stats import chi2_contingency
chi2_contingency(data_table)
```

[42]:
```
Chi2ContingencyResult(statistic=49.158655596893624,
pvalue=1.549925073686492e-07, dof=9, expected_freq=array([[1.77454639e+03,
6.99258130e+02, 2.11948742e+02, 2.46738931e-01],
       [1.80559765e+03, 7.11493845e+02, 2.15657450e+02, 2.51056403e-01],
       [1.80559765e+03, 7.11493845e+02, 2.15657450e+02, 2.51056403e-01],
       [1.80625831e+03, 7.11754180e+02, 2.15736359e+02, 2.51148264e-01]]))
```

p < alpha p value is low we reject the null, therefore Weather is not independent of the season

Recommendations

• In summer and fall seasons the company should have more bikes in stock to be rented. • Because the demand in these seasons is higher as compared to other seasons. • With a significance level of 0.05, workingday has no effect on the number of bikes being rented. • In very low humid days, company should have less bikes in the stock to be rented. • Whenever temperature is less than 10 or in very cold days, company should have less bikes. • Whenever the windspeed is greater than 35 or in thunderstorms, company should have less bikes in stock to be rented.

[ ]:
```

```