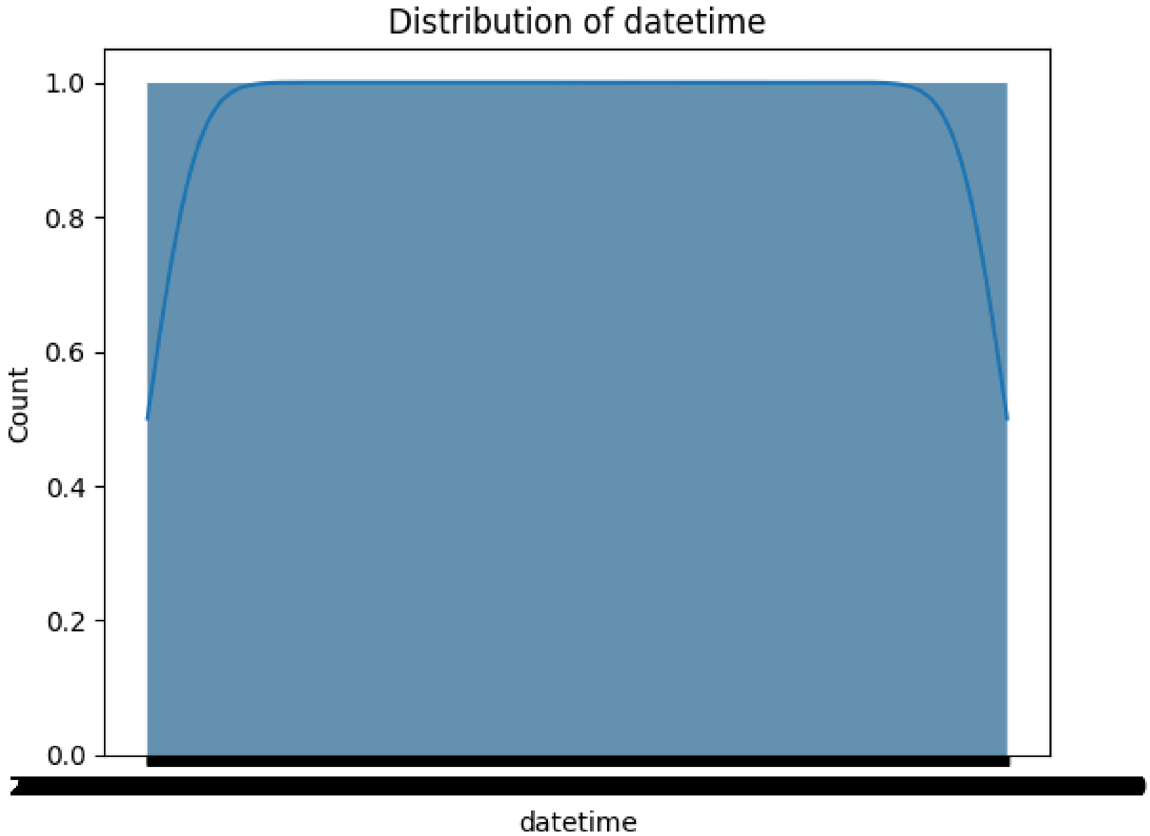
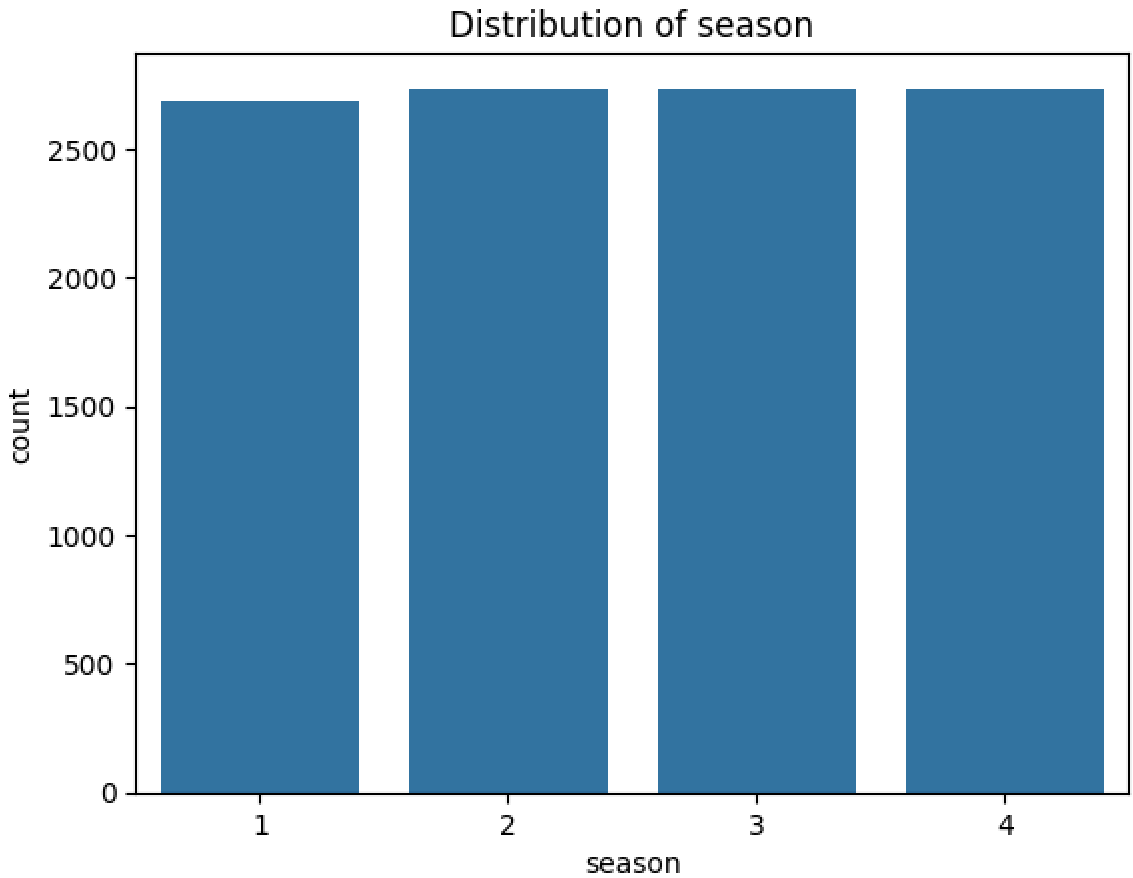
In [2]:
```
# Importing required libraries
import os
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import ttest_ind, f_oneway, chi2_contingency
from statsmodels.stats.weightstats import ztest

# Load the dataset
os.chdir("/Users/ayeshasiddiqha/Downloads")
data = pd.read_csv('yulu.csv')

# Initial Exploration
data.info()
print(data.describe())
print(data.isnull().sum())

# Convert categorical columns
categorical_columns = ['season', 'holiday', 'workingday', 'weather']
for col in categorical_columns:
    data[col] = data[col].astype('category')

# Univariate Analysis
for col in data.columns:
    if data[col].dtype == 'category':
        sns.countplot(x=col, data=data)
        plt.title(f"Distribution of {col}")
        plt.show()
    else:
        sns.histplot(data[col], kde=True)
        plt.title(f"Distribution of {col}")
        plt.show()

# Bivariate Analysis
# Exclude non-numeric columns like 'datetime'
numeric_data = data.select_dtypes(include=[np.number])
sns.heatmap(numeric_data.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()

# Scatter plots for continuous variables vs count
continuous_vars = ['temp', 'atemp', 'humidity', 'windspeed']
for var in continuous_vars:
    sns.scatterplot(x=data[var], y=data['count'])
    plt.title(f"{var} vs Count")
    plt.show()

# Boxplots for categorical variables vs count
for col in categorical_columns:
    sns.boxplot(x=col, y='count', data=data)
    plt.title(f"{col} vs Count")
    plt.show()

# Hypothesis Testing
alpha = 0.05

# 1. Working day effect on electric cycle rentals
```

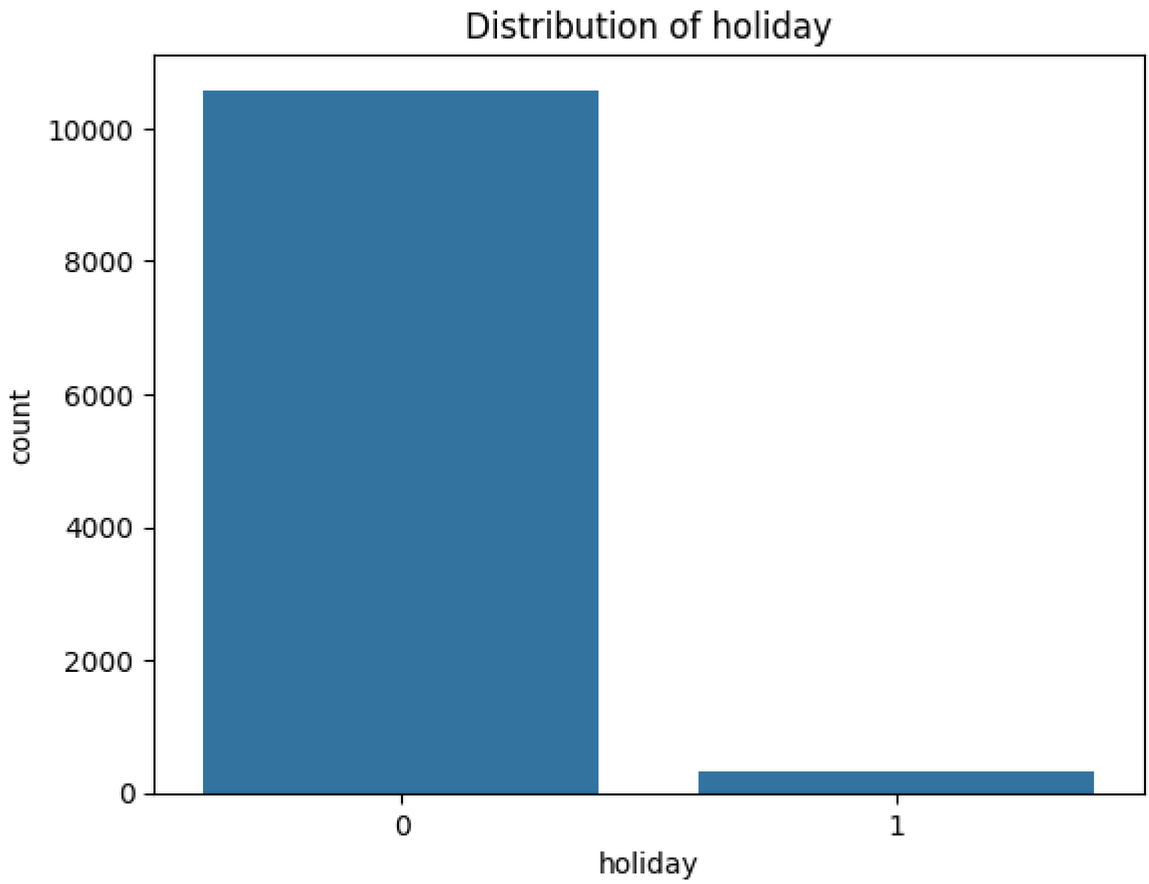
```
# H0: Working day has no effect on the number of cycles rented
# H1: Working day has an effect on the number of cycles rented
workingday_counts = [data[data['workingday'] == 1]['count'],
                     data[data['workingday'] == 0]['count']]
t_stat, p_val = ttest_ind(workingday_counts[0], workingday_counts[1], equ
print(f"T-test result: t_stat = {t_stat}, p_val = {p_val}")
if p_val < alpha:
    print("Reject the null hypothesis: Working day has an effect on the n
else:
    print("Fail to reject the null hypothesis: No effect of working day o

# 2. Cycles rented in different seasons
# H0: No difference in the number of cycles rented across seasons
# H1: Difference exists in the number of cycles rented across seasons
season_groups = [data[data['season'] == s]['count'] for s in data['season
f_stat, p_val = f_oneway(*season_groups)
print(f"ANOVA result (Seasons): f_stat = {f_stat}, p_val = {p_val}")
if p_val < alpha:
    print("Reject the null hypothesis: Cycles rented differ across season
else:
    print("Fail to reject the null hypothesis: No difference in rentals a

# 3. Cycles rented in different weathers
# H0: No difference in the number of cycles rented across weathers
# H1: Difference exists in the number of cycles rented across weathers
weather_groups = [data[data['weather'] == w]['count'] for w in data['weat
f_stat, p_val = f_oneway(*weather_groups)
print(f"ANOVA result (Weather): f_stat = {f_stat}, p_val = {p_val}")
if p_val < alpha:
    print("Reject the null hypothesis: Cycles rented differ across weathe
else:
    print("Fail to reject the null hypothesis: No difference in rentals a

# 4. Dependency of Weather on Season
# H0: Weather is independent of the season
# H1: Weather is dependent on the season
contingency_table = pd.crosstab(data['season'], data['weather'])
chi2, p_val, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-square result: chi2 = {chi2}, p_val = {p_val}")
if p_val < alpha:
    print("Reject the null hypothesis: Weather depends on the season.")
else:
    print("Fail to reject the null hypothesis: Weather is independent of 

# Final Insights and Recommendations
print("""
Insights:
1. Working day has a significant effect on the number of cycles rented.
2. The number of cycles rented significantly differs across seasons and w
3. Weather is dependent on the season, indicating strong seasonal pattern

Recommendations:
1. Increase availability of cycles during peak seasons and favorable weat
2. Focus marketing efforts on working days to maximize rentals.
3. Plan for adverse weather conditions with protective measures or altern
""")
```

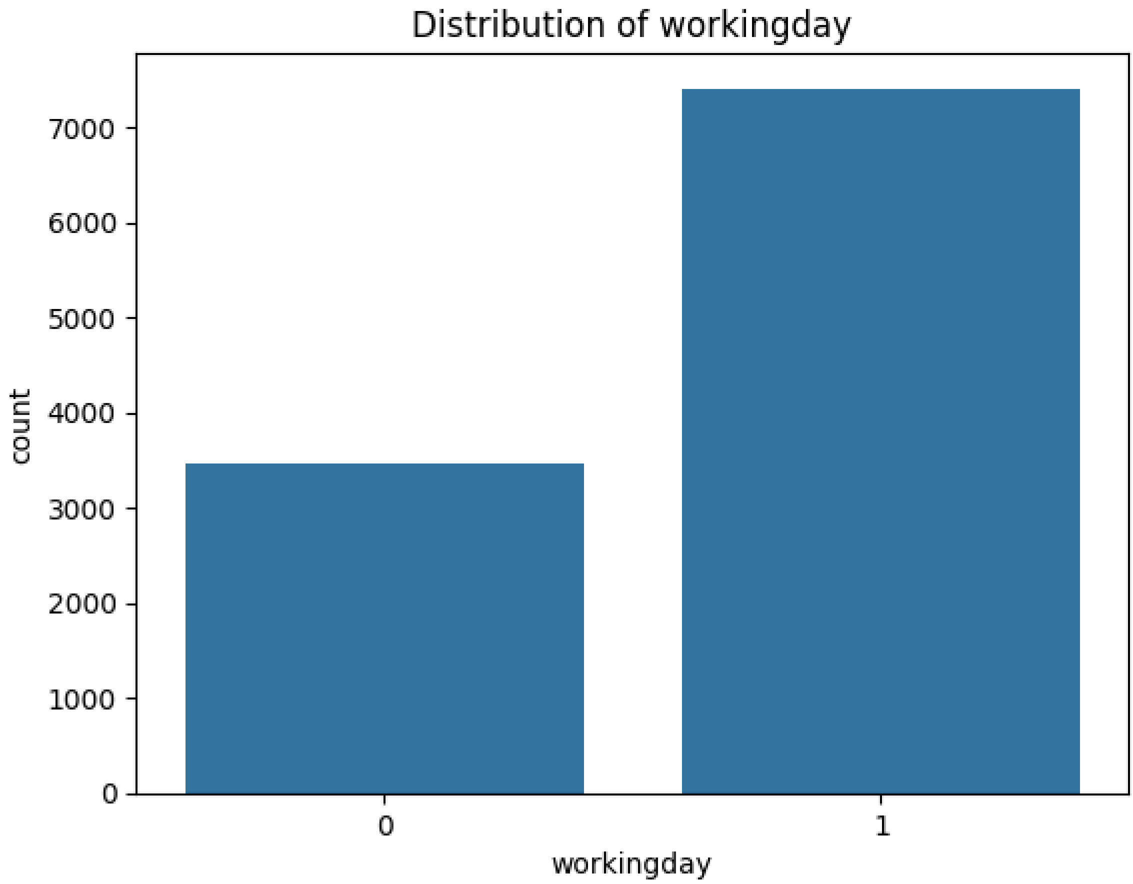
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10886 entries, 0 to 10885
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   datetime    10886 non-null  object 
 1   season      10886 non-null  int64  
 2   holiday     10886 non-null  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
             season       holiday    workingday       weather         temp 
\
count  10886.000000  10886.000000  10886.000000  10886.000000  10886.00000
mean       2.506614      0.028569      0.680875      1.418427     20.23086
std        1.116174      0.166599      0.466159      0.633839      7.79159
min        1.000000      0.000000      0.000000      1.000000      0.82000
25%        2.000000      0.000000      0.000000      1.000000     13.94000
50%        3.000000      0.000000      1.000000      1.000000     20.50000
75%        4.000000      0.000000      1.000000      2.000000     26.24000
max        4.000000      1.000000      1.000000      4.000000     41.00000

              atemp      humidity     windspeed        casual    registere
d  \
count  10886.000000  10886.000000  10886.000000  10886.000000  10886.00000
0   
mean      23.655084     61.886460     12.799395     36.021955    155.55217
7   
std        8.474601     19.245033      8.164537     49.960477    151.03903
3   
min        0.760000      0.000000      0.000000      0.000000      0.00000
0   
25%       16.665000     47.000000      7.001500      4.000000     36.00000
0   
50%       24.240000     62.000000     12.998000     17.000000    118.00000
0   
75%       31.060000     77.000000     16.997900     49.000000    222.00000
0   
max       45.455000    100.000000     56.996900    367.000000    886.00000
0   

              count  
count  10886.000000  
mean     191.574132  
std      181.144454  
min        1.000000  
25%       42.000000  
50%      145.000000  
75%      284.000000  
max      977.000000  
datetime      0
season        0
```

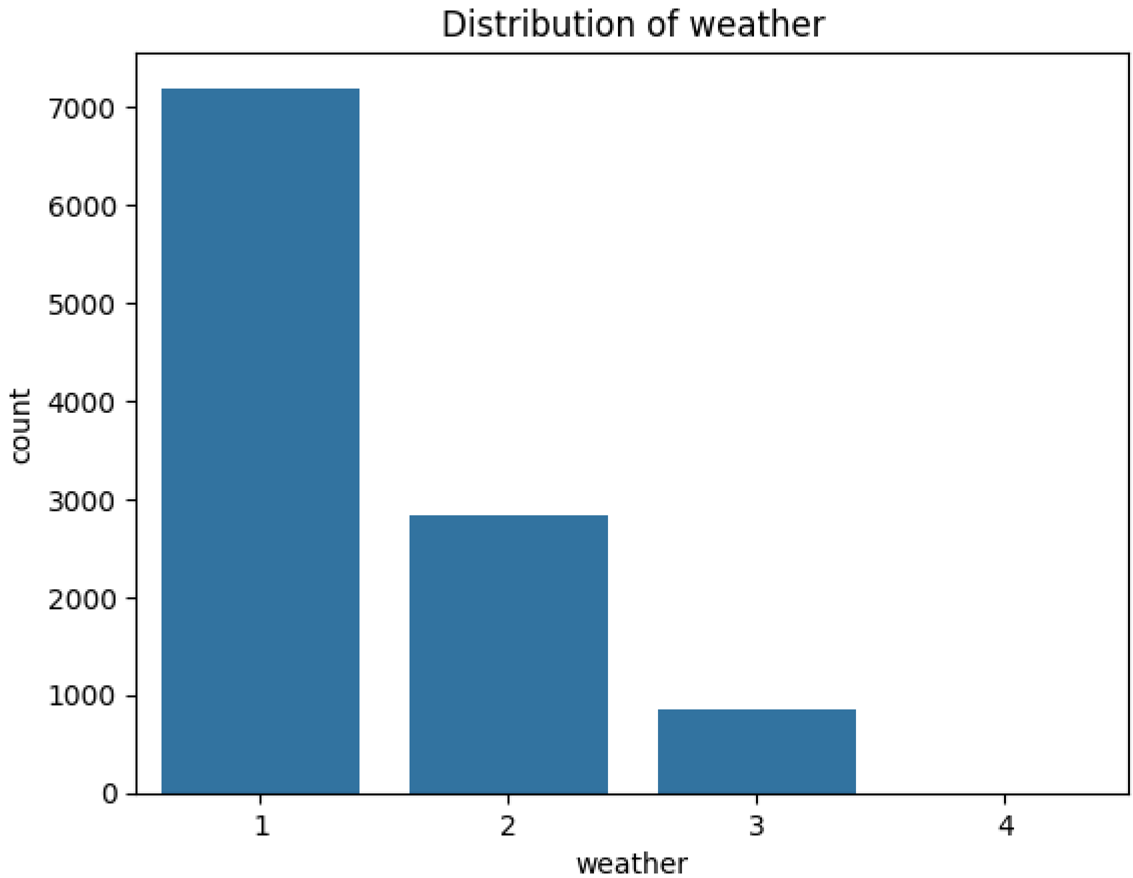
```
holiday       0
workingday    0
weather       0
temp          0
atemp         0
humidity      0
windspeed     0
casual        0
registered    0
count         0
dtype: int64
```

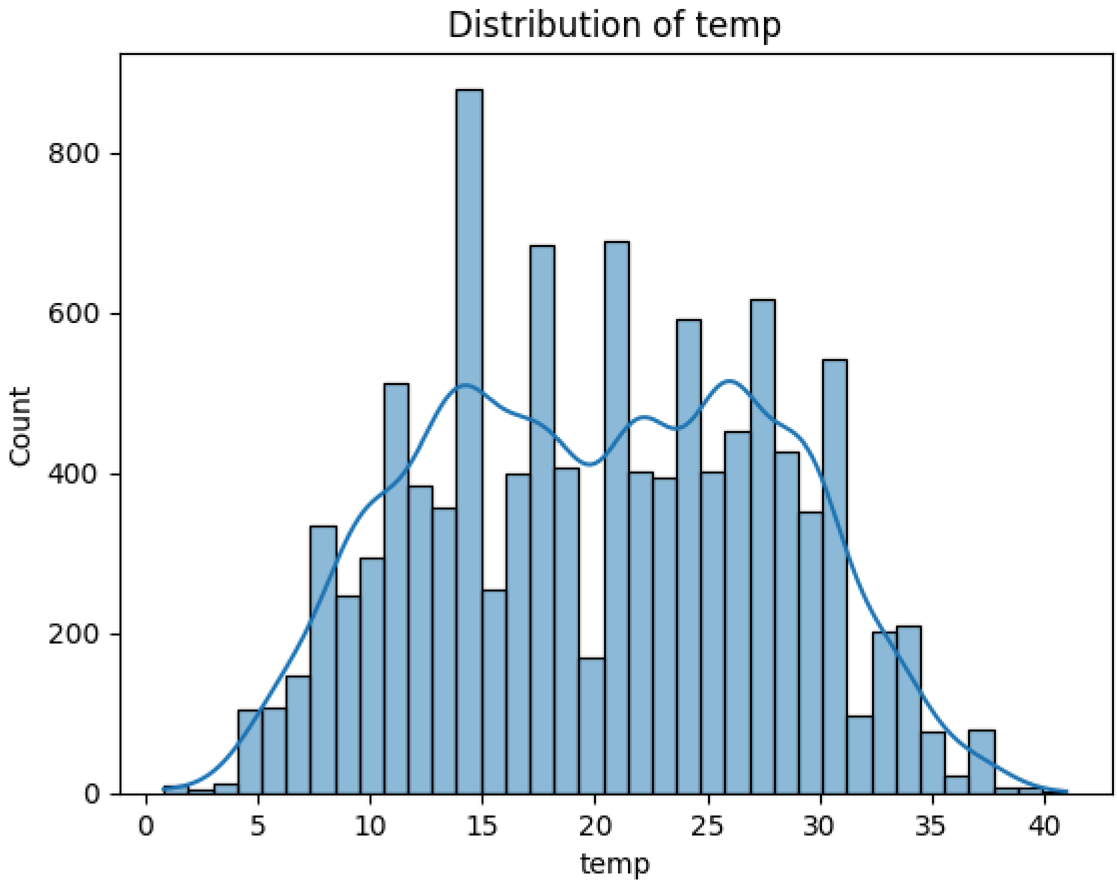


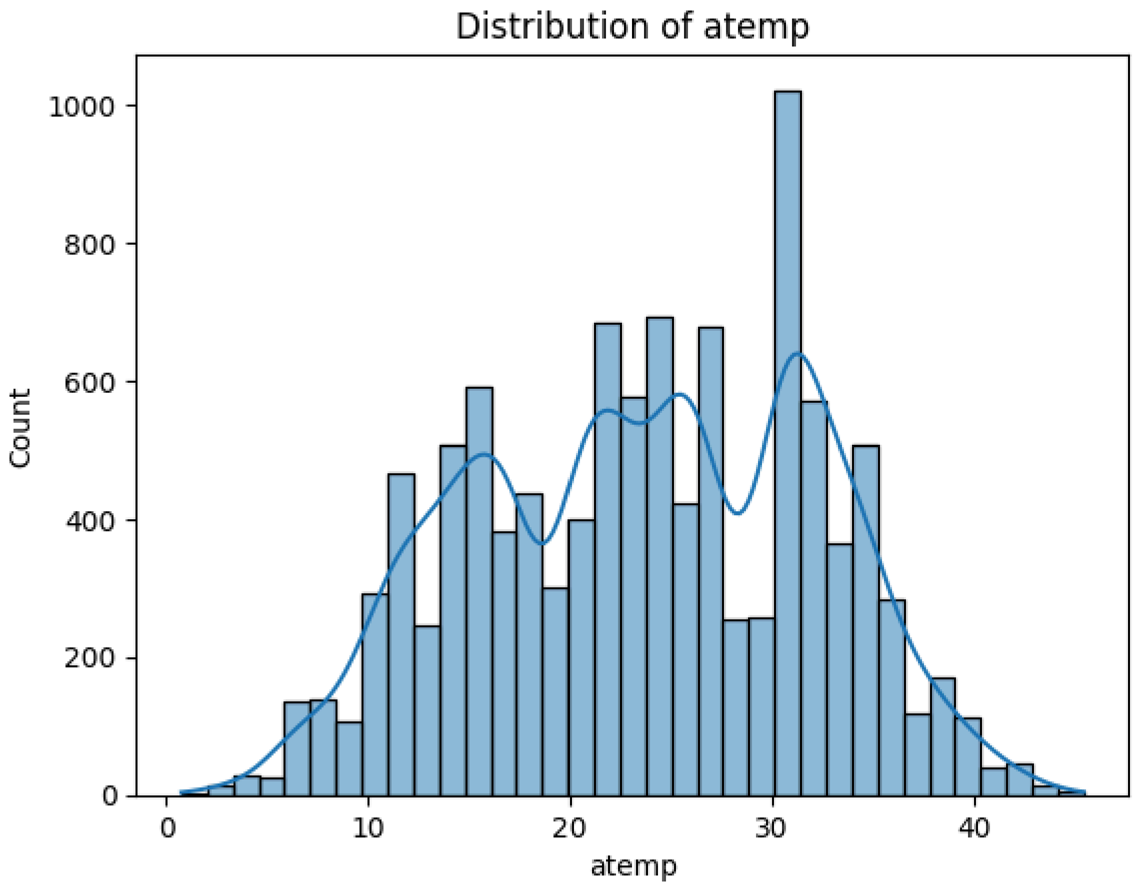

Distribution of season
count
2500
2000
1500
1000
500
0
1
2
3
4
season

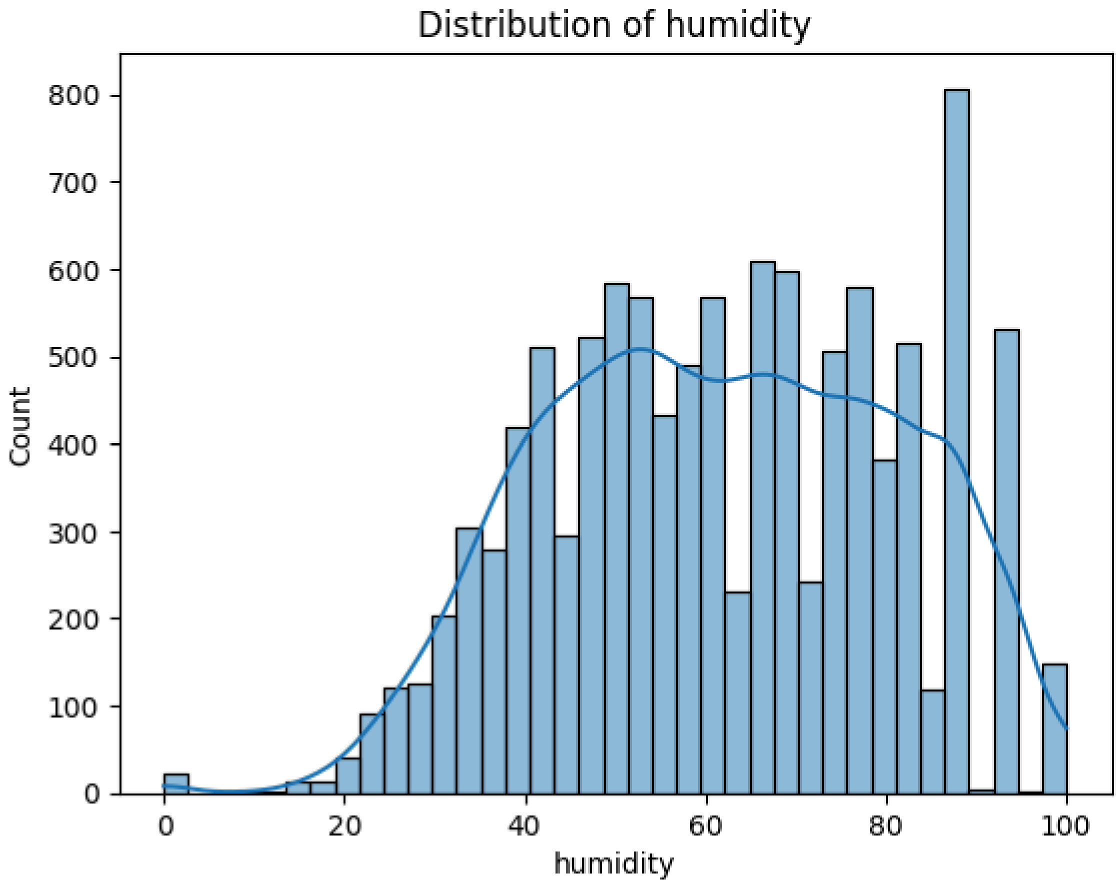

Distribution of holiday
count
10000
8000
6000
4000
2000
0
0
1
holiday

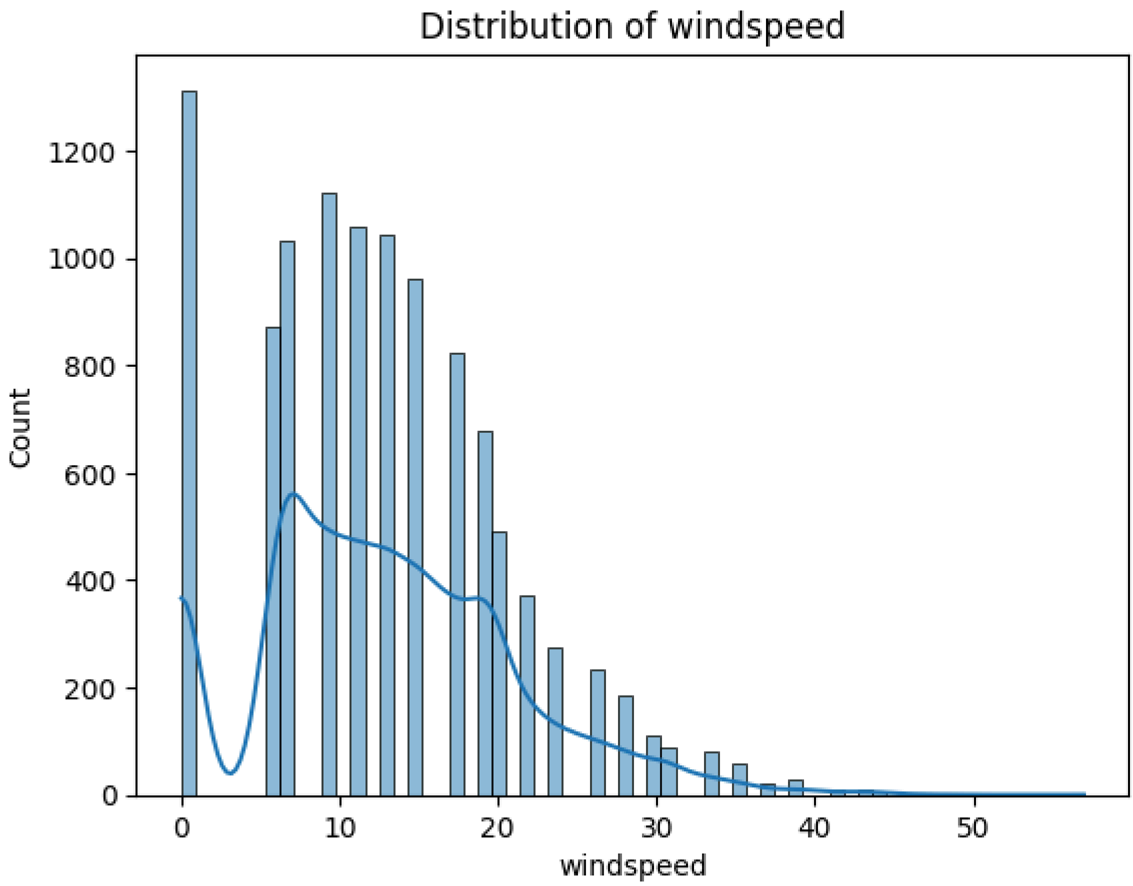

Distribution of workingday
count
0
1000
2000
3000
4000
5000
6000
7000
0
1
workingday

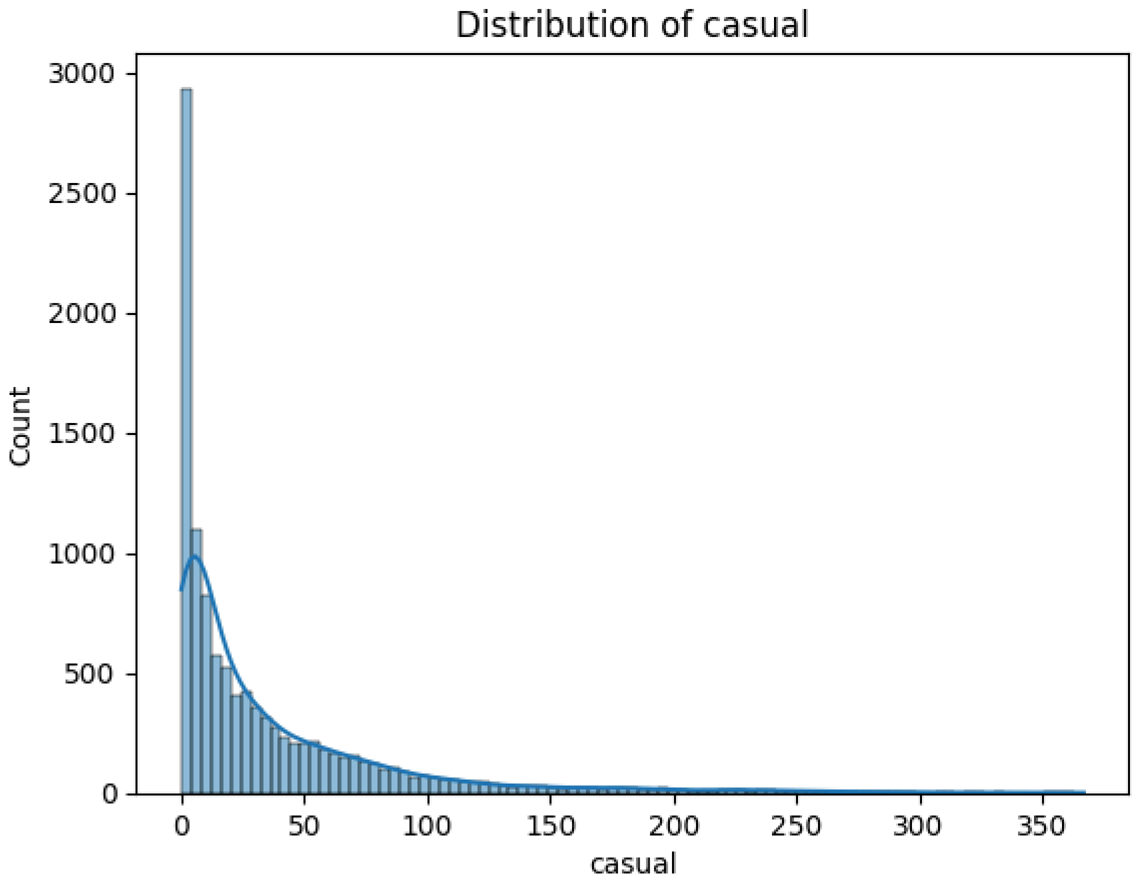

Distribution of weather
count
0
1000
2000
3000
4000
5000
6000
7000
1
2
3
4
weather

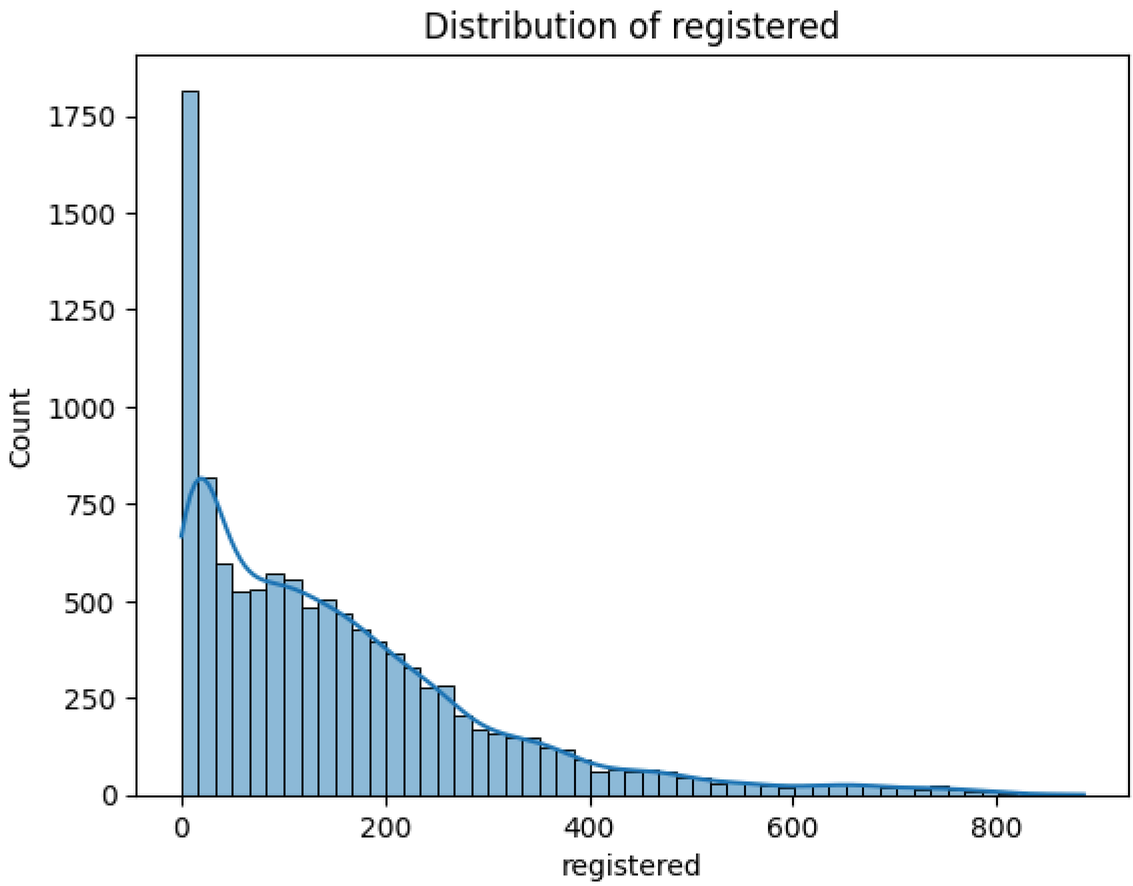

Distribution of temp
Count
0
200
400
600
800
0
5
10
15
20
25
30
35
40
temp

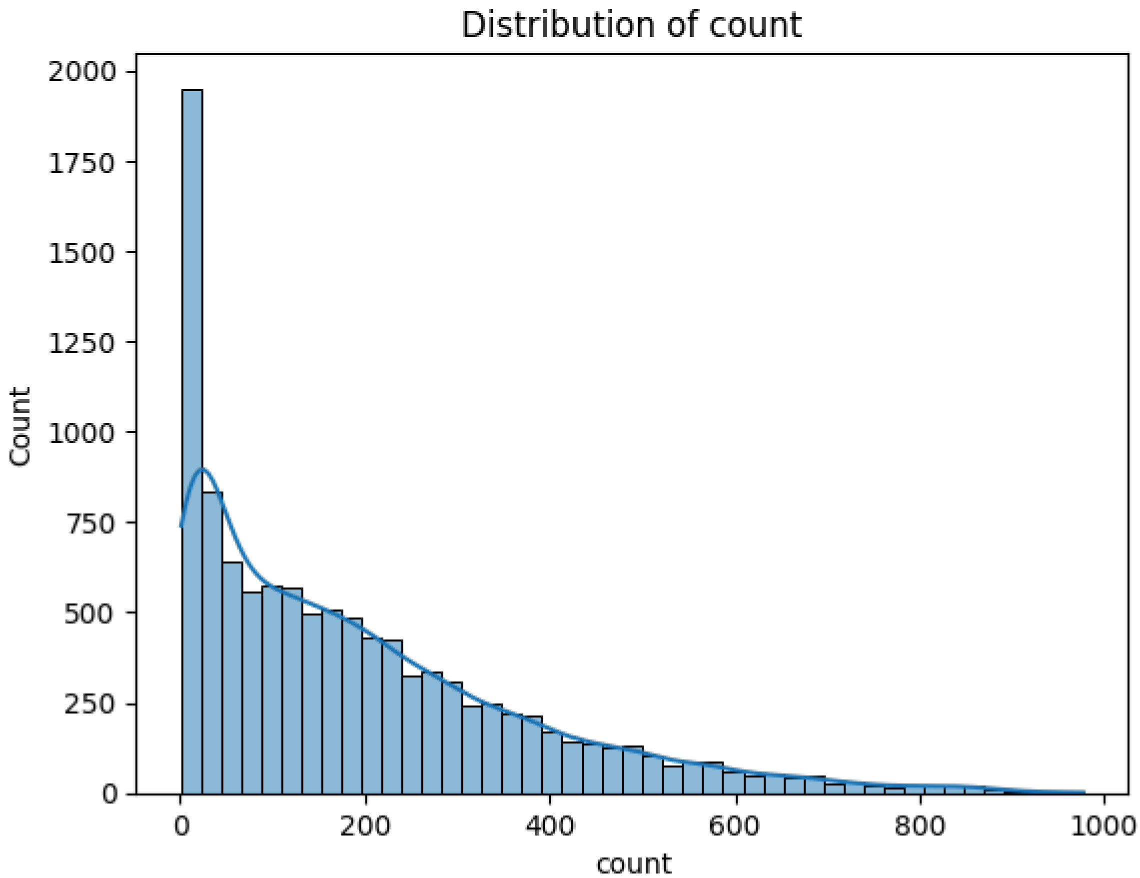

Distribution of atemp
Count
0
200
400
600
800
1000
0
10
20
30
40
atemp

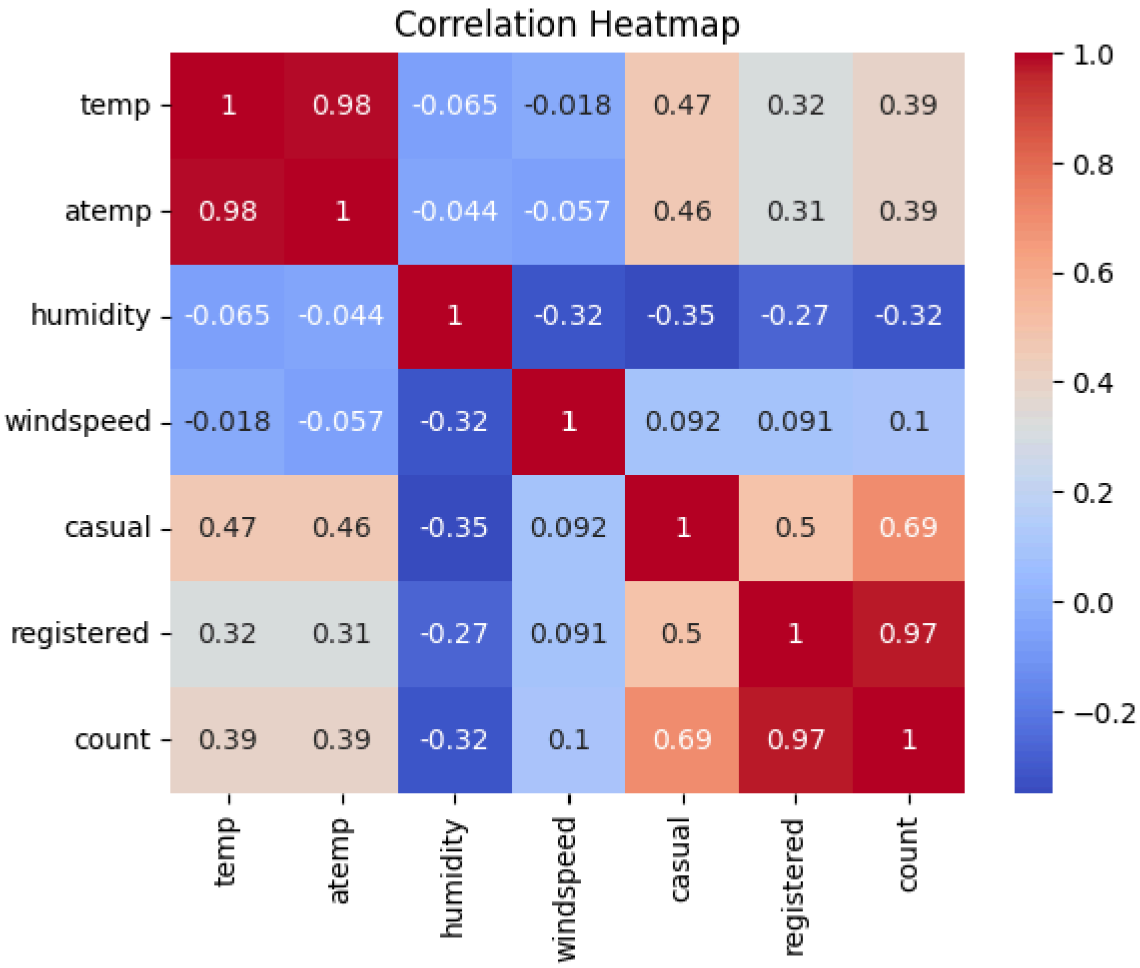

Distribution of humidity
Count
humidity
0
100
200
300
400
500
600
700
800
0
20
40
60
80
100

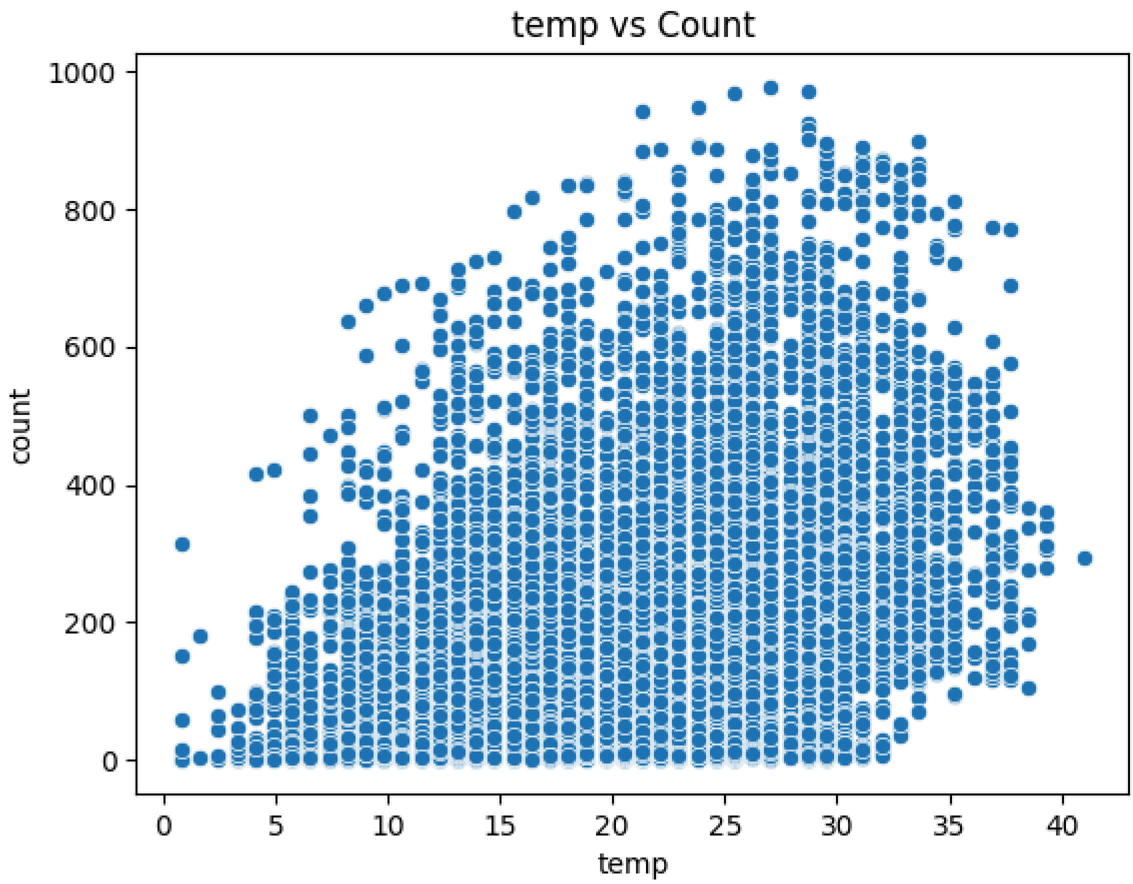

Distribution of windspeed
Count
windspeed
0
200
400
600
800
1000
1200
0
10
20
30
40
50

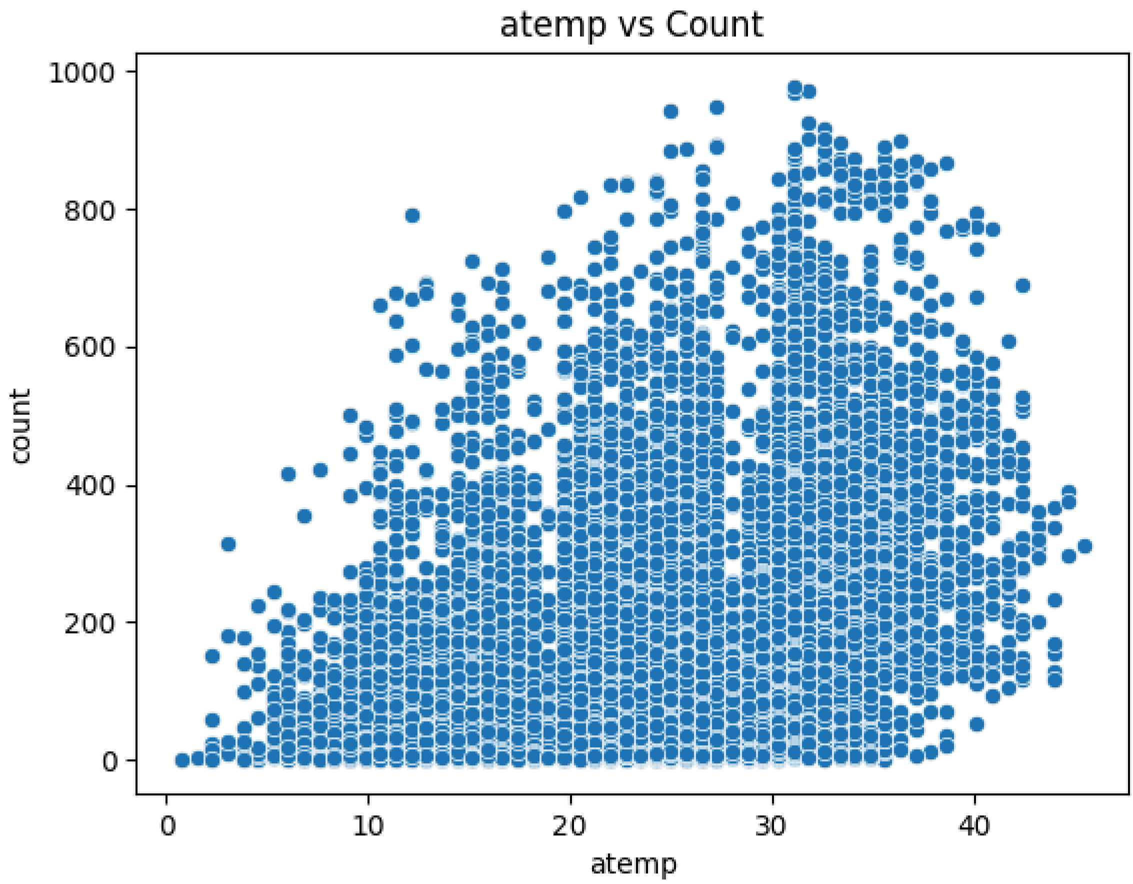

Distribution of casual
Count
3000
2500
2000
1500
1000
500
0
0
50
100
150
200
250
300
350
casual

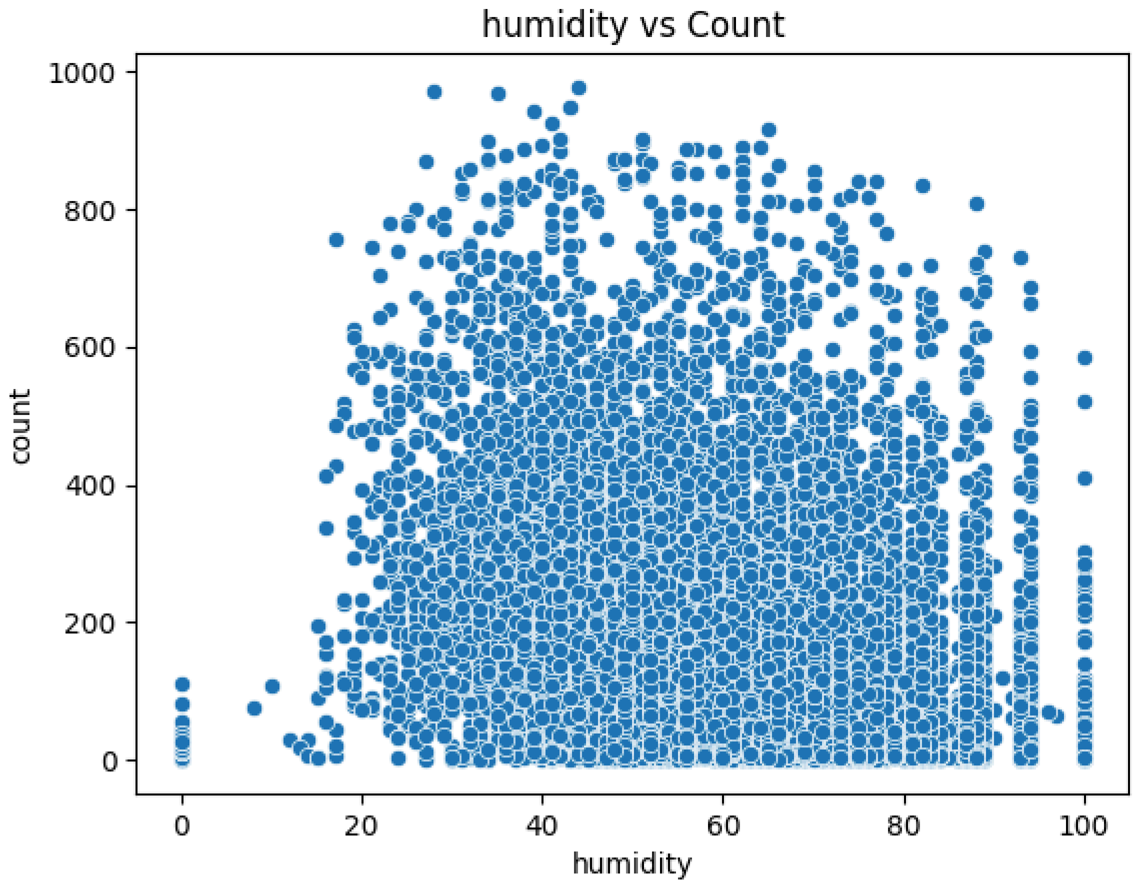

Distribution of registered
Count
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
registered

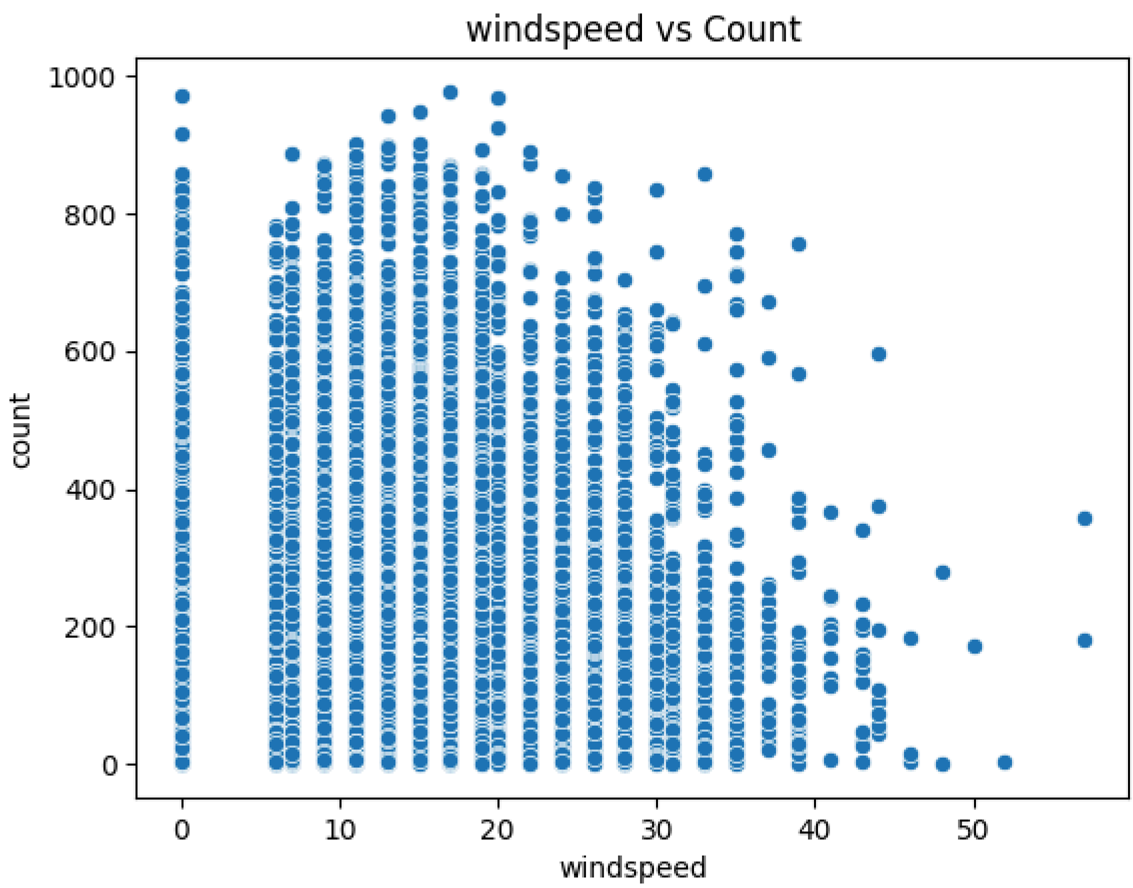

Distribution of count
Count
2000
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000
count

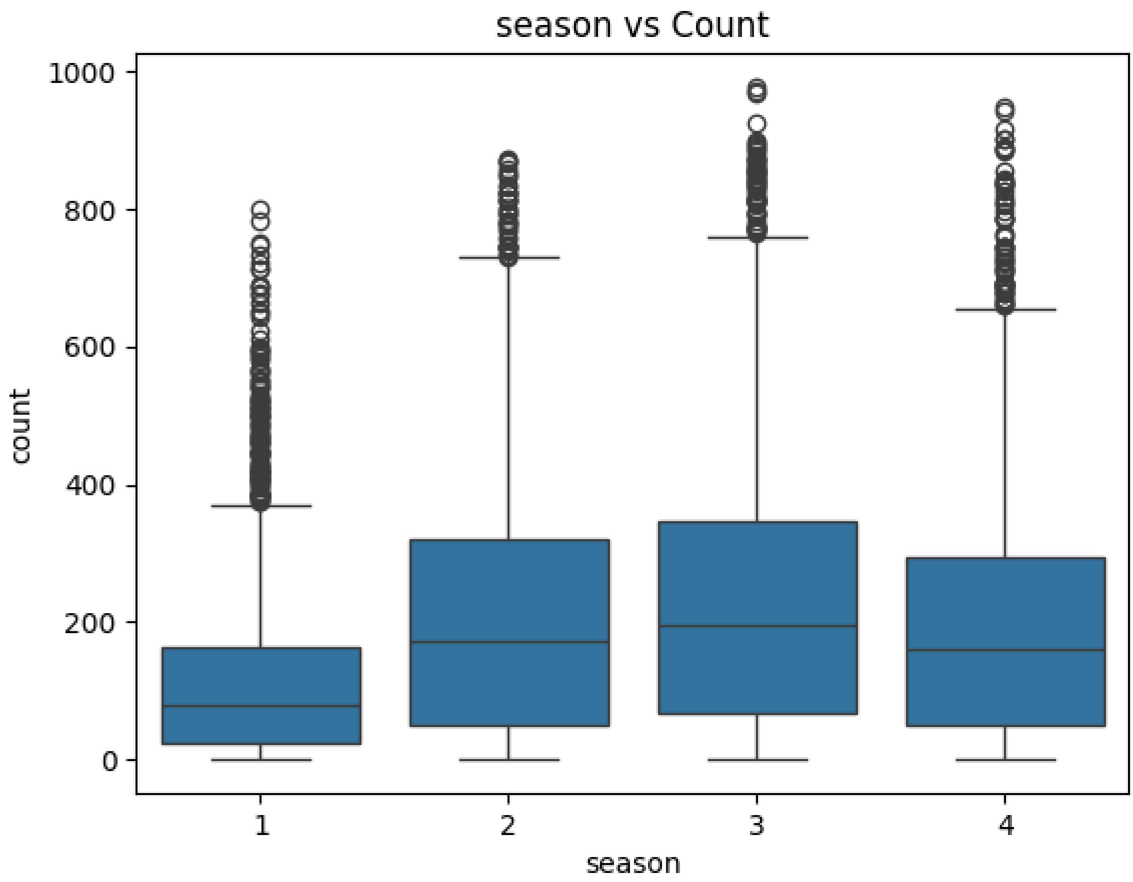

Correlation Heatmap
temp
atemp
humidity
windspeed
casual
registered
count
1 0.98 -0.065 -0.018 0.47 0.32 0.39
0.98 1 -0.044 -0.057 0.46 0.31 0.39
-0.065 -0.044 1 -0.32 -0.35 -0.27 -0.32
-0.018 -0.057 -0.32 1 0.092 0.091 0.1
0.47 0.46 -0.35 0.092 1 0.5 0.69
0.32 0.31 -0.27 0.091 0.5 1 0.97
0.39 0.39 -0.32 0.1 0.69 0.97 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

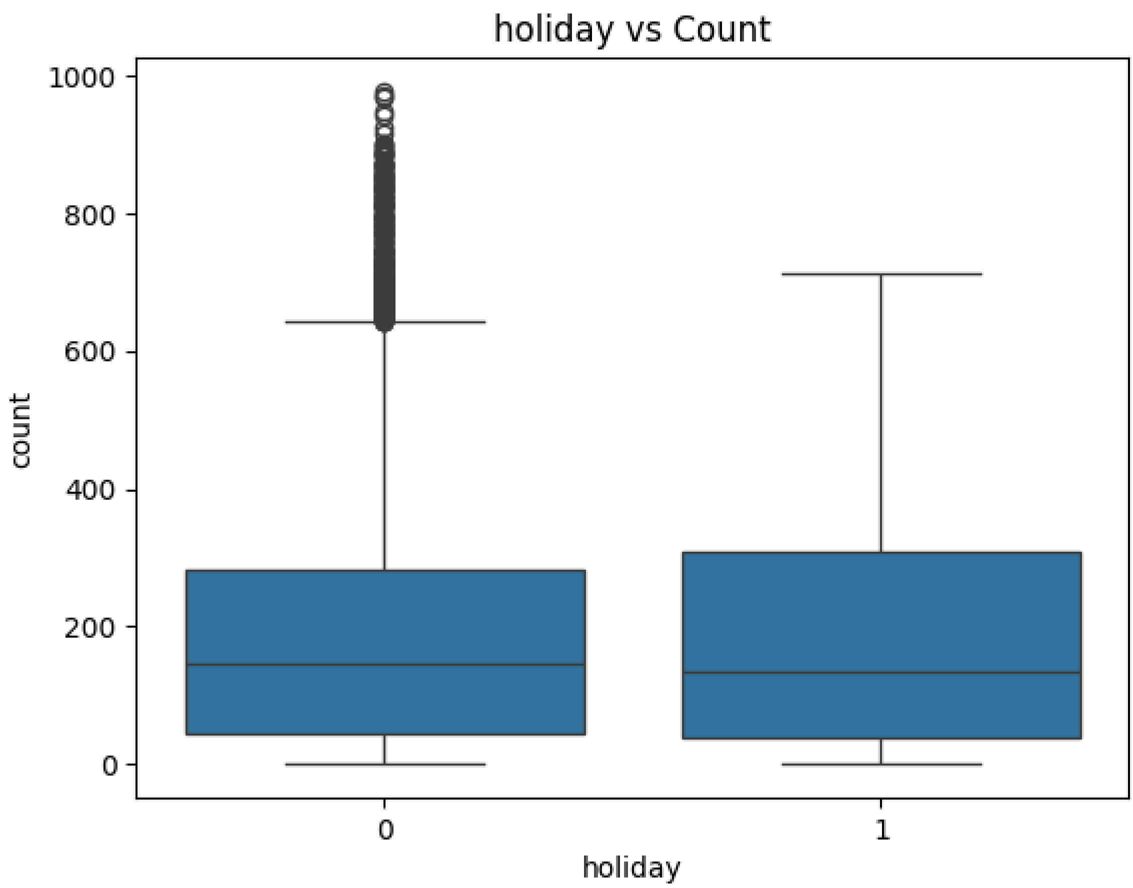

temp vs Count
count
temp
0
200
400
600
800
1000
0
5
10
15
20
25
30
35
40

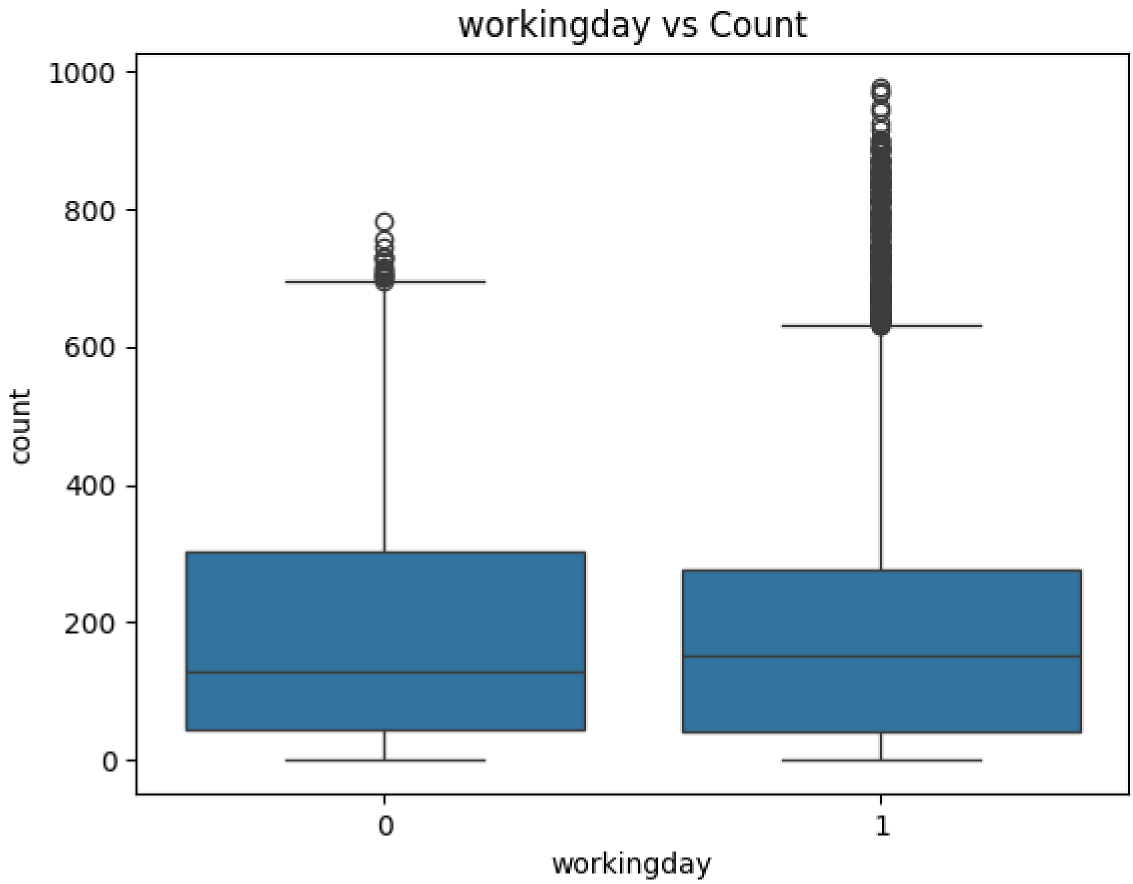

atemp vs Count
count
atemp
0
200
400
600
800
1000
0
10
20
30
40

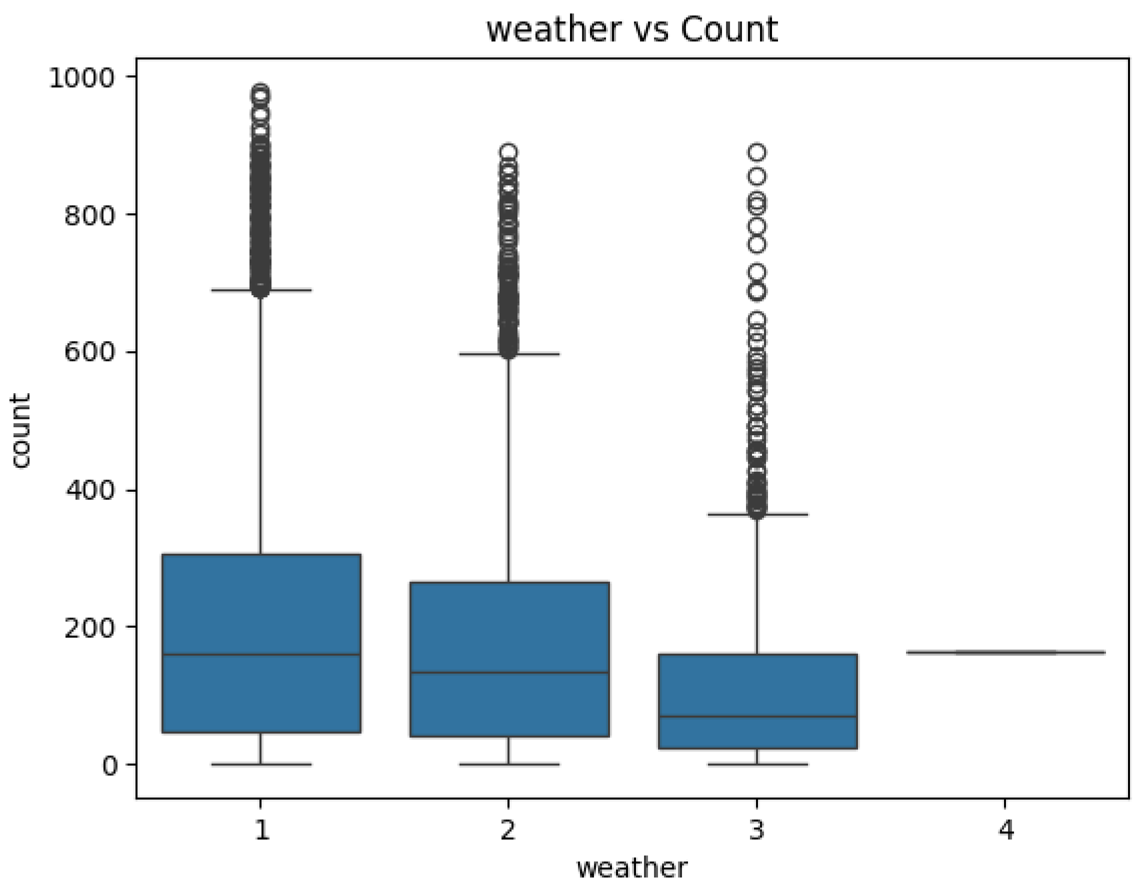

humidity vs Count
count
humidity
0
200
400
600
800
1000
0
20
40
60
80
100


windspeed vs Count
count
windspeed
0
200
400
600
800
1000
0
10
20
30
40
50


season vs Count
1000
800
600
400
200
0
count
1
2
3
4
season


holiday vs Count
1000
800
600
400
200
0
count
0
1
holiday


workingday vs Count
1000
800
600
400
200
0
count
0
1
workingday


weather vs Count
1000
800
600
400
200
0
count
1
2
3
4
weather

```
T-test result: t_stat = 1.2362580418223226, p_val = 0.21640312280695098
Fail to reject the null hypothesis: No effect of working day on cycle rent
als.
ANOVA result (Seasons): f_stat = 236.94671081032106, p_val = 6.16484338649
9654e-149
Reject the null hypothesis: Cycles rented differ across seasons.
ANOVA result (Weather): f_stat = 65.53024112793271, p_val = 5.482069475935
669e-42
Reject the null hypothesis: Cycles rented differ across weathers.
Chi-square result: chi2 = 49.158655596893624, p_val = 1.549925073686492e-0
7
Reject the null hypothesis: Weather depends on the season.

Insights:
1. Working day has a significant effect on the number of cycles rented.
2. The number of cycles rented significantly differs across seasons and we
ather conditions.
3. Weather is dependent on the season, indicating strong seasonal pattern
s.

Recommendations:
1. Increase availability of cycles during peak seasons and favorable weath
er conditions.
2. Focus marketing efforts on working days to maximize rentals.
3. Plan for adverse weather conditions with protective measures or alterna
te solutions.
```

In [ ]:

```

```