Yulu Business Case Study

Yulu has recently suffered considerable dips in its revenues. They have contracted a consulting company to understand the factors on which the demand for these shared electric cycles depends. Specifically, they want *to understand the factors affecting the demand for these shared electric cycles in the Indian market*.

Problem Statement

The company wants to know:

- Which variables are significant in predicting the demand for shared electric cycles in the Indian market?
- How well those variables describe the electric cycle demands

Non-Graphical Analysis (Analyzing the Basic Metrics)

- Total Customer Base: There are a total of 10886 customer details.
- Datetime: The data given is between the time frame: 2011-01-01 to 2012-12-19
    - Total number of days: 718 days
- Season: season (1: spring, 2: summer, 3: fall, 4: winter)
    - winter: 2734
    - summer: 2733
    - fall: 2733
    - spring: 2686
- Holiday:The Unique Values in holiday are:
    - Number of Holidays: 311
    - Not a holiday: 10575
- Workingday : if day is neither weekend nor holiday
    - Working day: 7412
    - Not a working day: 3474
- weather: The Unique Values in weather are:
    - 1: Clear: 7192
    - 2: Mist: 2834
    - 3: Light Rain: 859
    - 4: Heavy Rain : 1
- temp: The count of unique values in temp are: 49
- atemp: There are 60 unique values in the atemp column.
- humidity: There are 89 unique values in the humidity column.
- windspeed: There are 28 unique values in the windspeed column.
- casual: There are 309 unique values in the casual column.

- registered:There are 731 unique values in the registered column.
- count: The count of unique values in count are 822

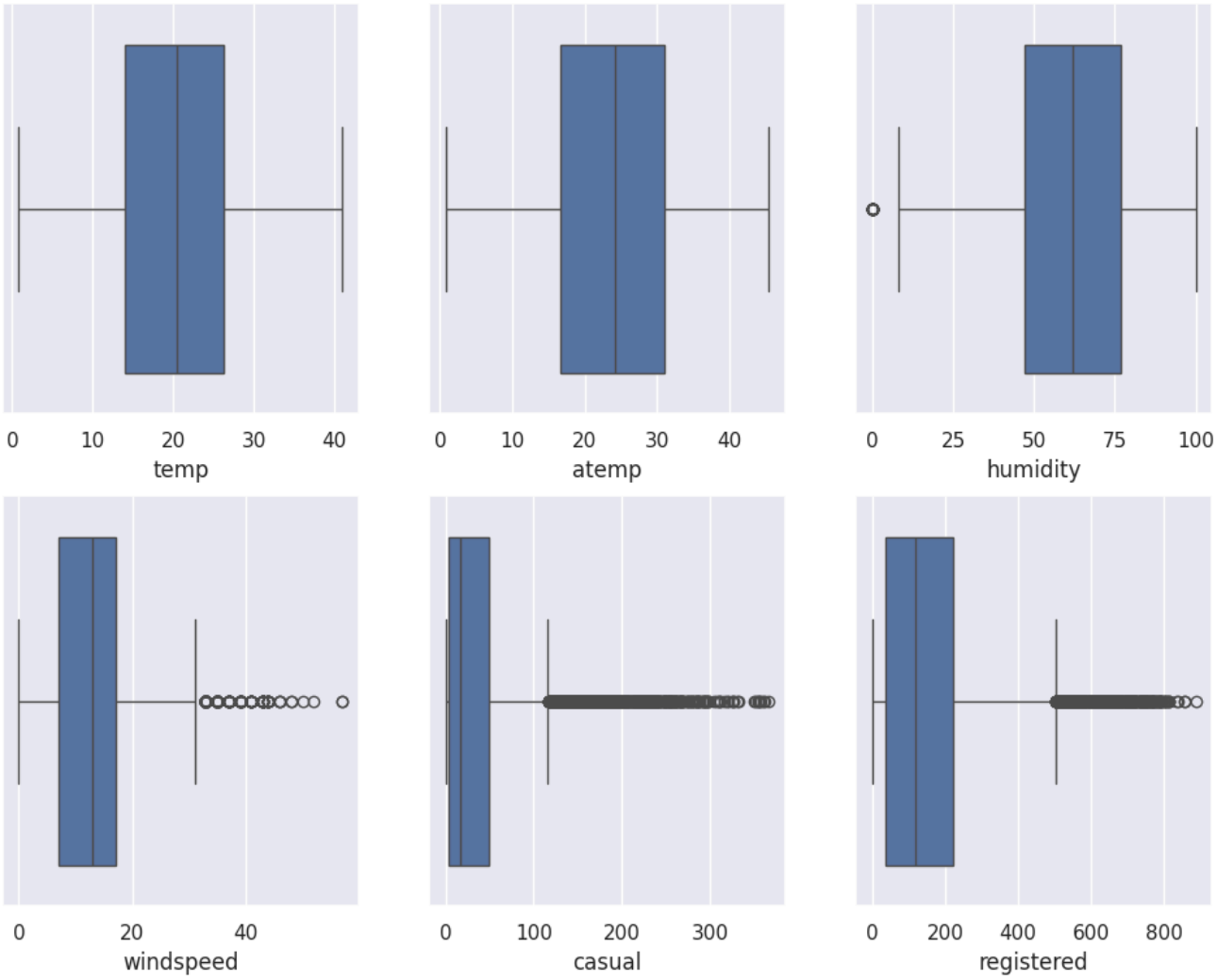
Finding Duplicates

- There are no duplicate values in the dataset.

Checking for Missing values

- There are no missing values in the dataset.

Outliers Detection



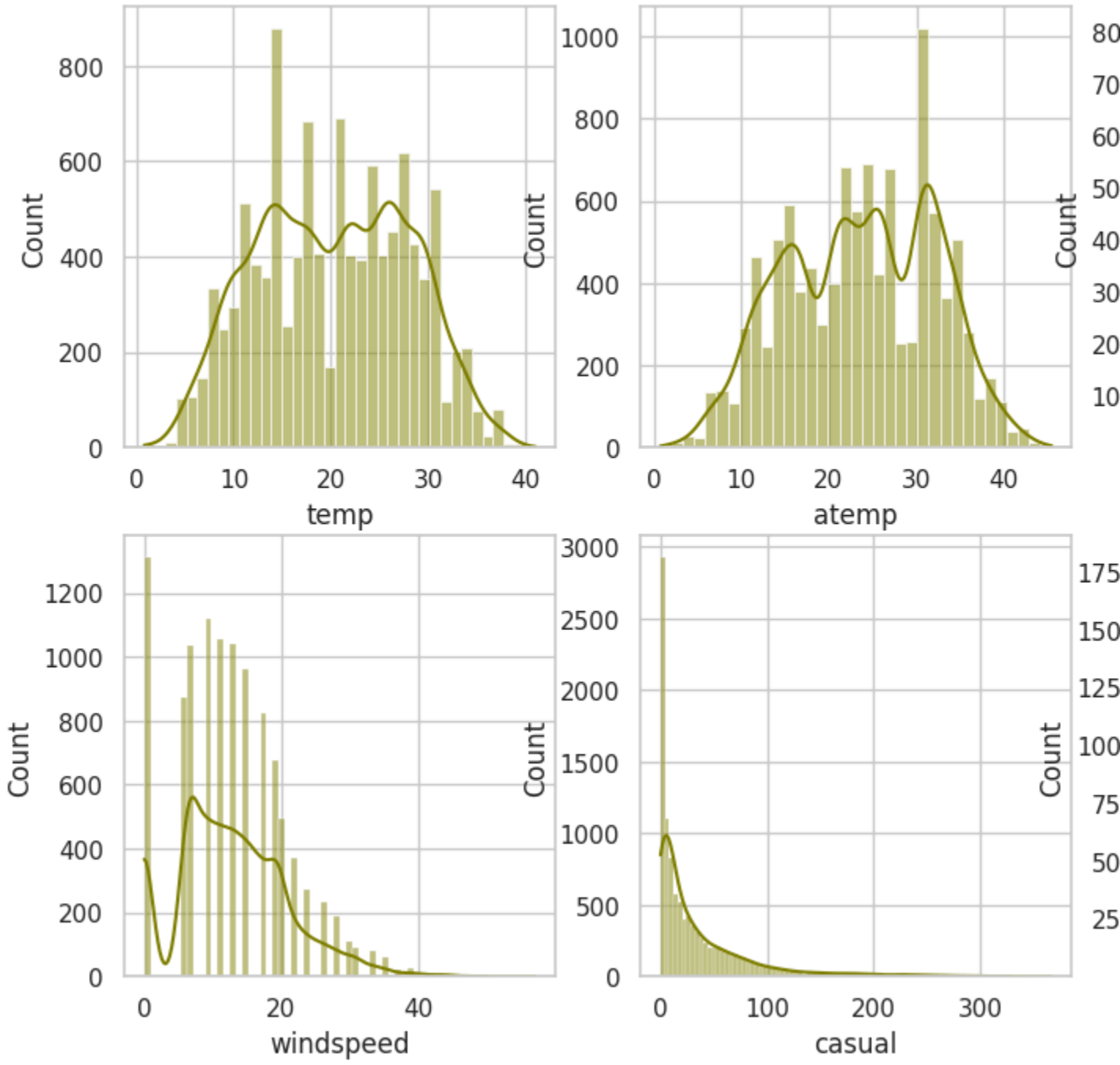
Calculating the IQR for Humidity Q1= 47.0 Q3= 77.0 IQR= 30.0 Upper band= 122.0 Lower band= 2.0 Median 62.0 Total outliers: 22	Calculating the IQR for Windspeed Q1= 7.0015 Q3= 16.9979 IQR= 9.9964 Upper band= 31.9925 Lower band= -7.9931 Median 12.998 Total outliers: 227	Calculating the IQR for Casual Q1= 4.0 Q3= 49.0 IQR= 45.0 Upper band= 116.5 Lower band= -63.5 Median 17.0 Total outliers: 749
Calculating the IQR for Registered Q1= 36.0 Q3= 222.0 IQR= 186.0 Upper band= 501.0 Lower band= -243.0 Median 118.0 Total outliers: 423	Calculating the IQR for Count Q1= 42.0 Q3= 284.0 IQR= 242.0 Upper band= 647.0 Lower band= -321.0 Median 145.0 Total outliers: 300	

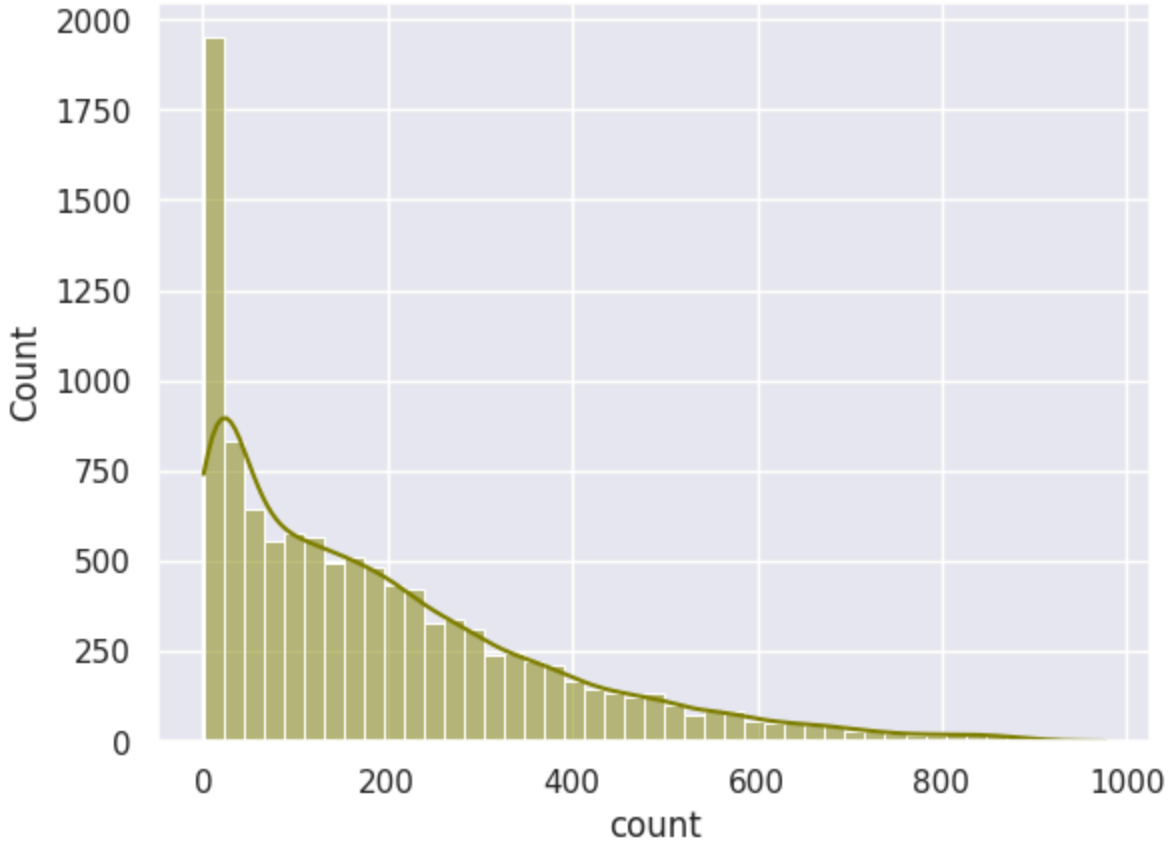
Observations

- Temperature Outliers:
  There are no unusual values (outliers) in the temperature (temp) and perceived temperature (atemp) columns.
- Humidity Outliers:
  There are a few unusual values (outliers) in the humidity column, indicating some data points deviate from the norm.
- Outliers in Several Columns:
  The columns windspeed, casual users, registered users, and the total count of rented bikes (count) all have many unusual values (outliers).
- Percentage of Outliers:
  - Only 0.2% of the humidity values are considered outliers.
  - About 2% of windspeed values are outliers.
  - Approximately 6.8% of casual user counts are considered outliers.
  - About 3.8% of registered user counts are considered outliers.
  - Around 2.7% of the total bike count values are outliers.

Univariate Analysis

Distribution of Numerical Variables

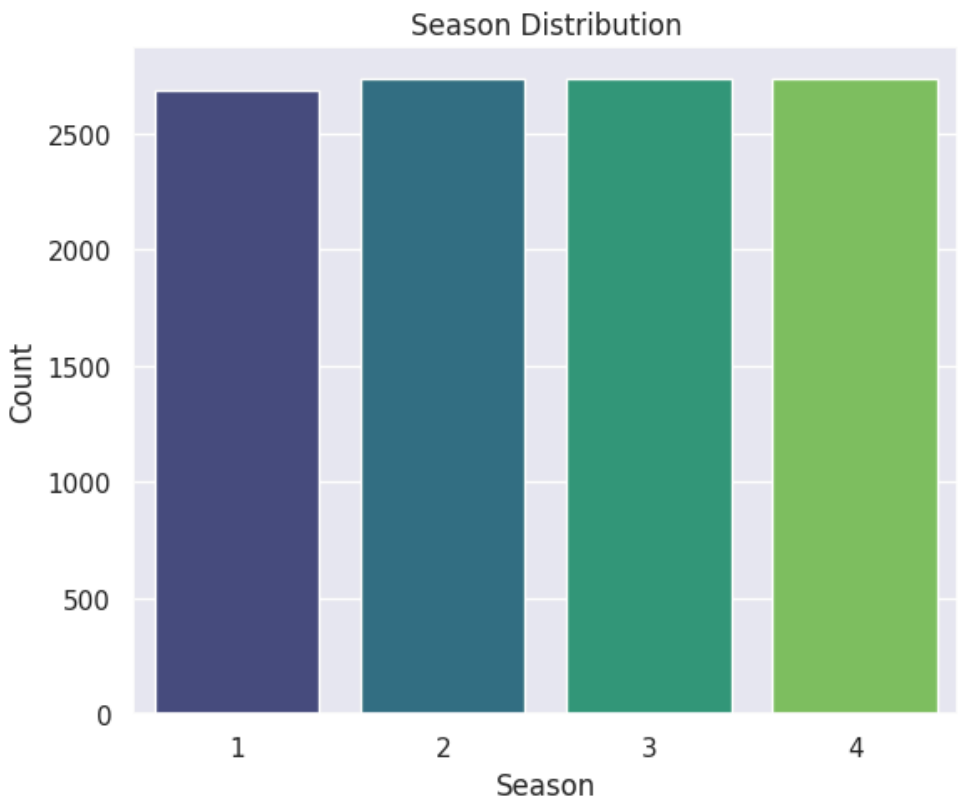


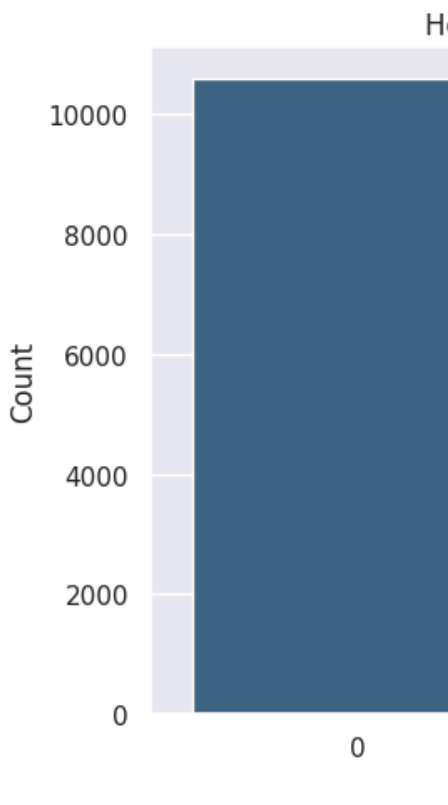


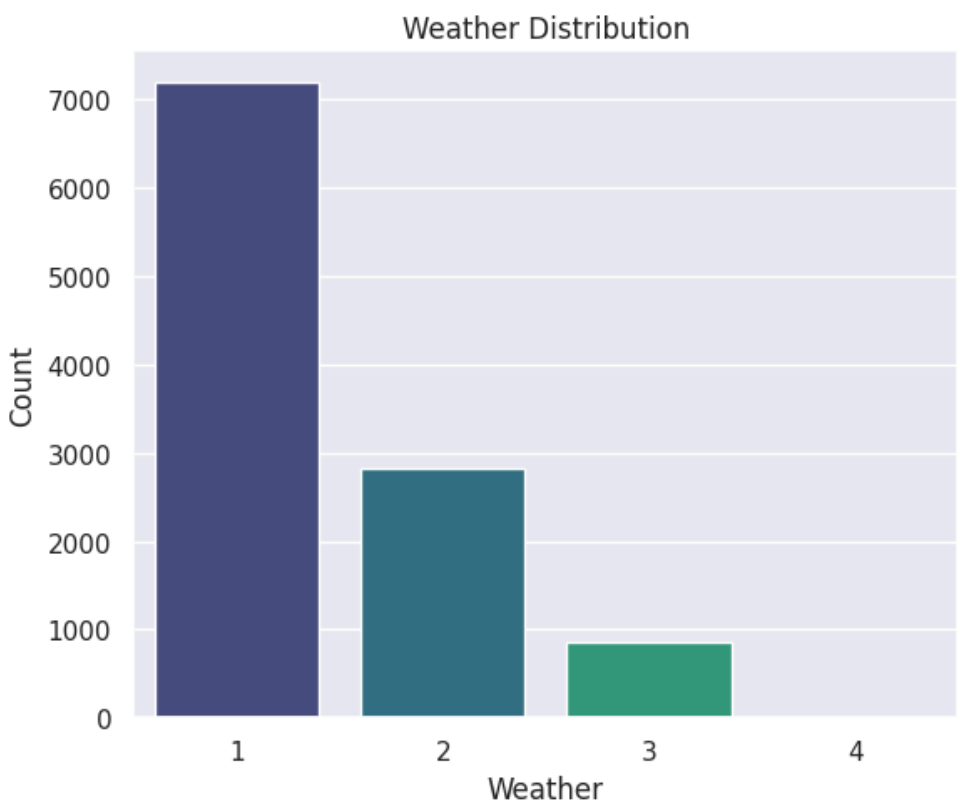
Observations

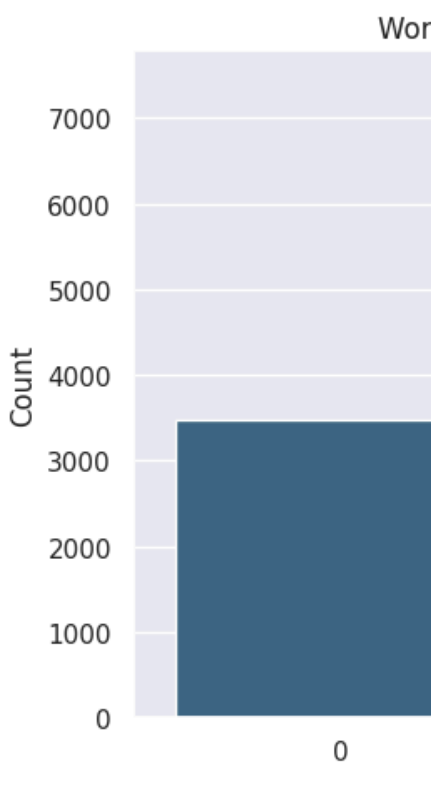
- The distribution of casual, registered, and total counts bears resemblance to a Log Normal Distribution.
- Temperature (temp), apparent temperature (atemp), and humidity exhibit patterns indicative of a Normal Distribution.
- Windspeed conforms to a Binomial Distribution.

Distribution of Categorical Columns

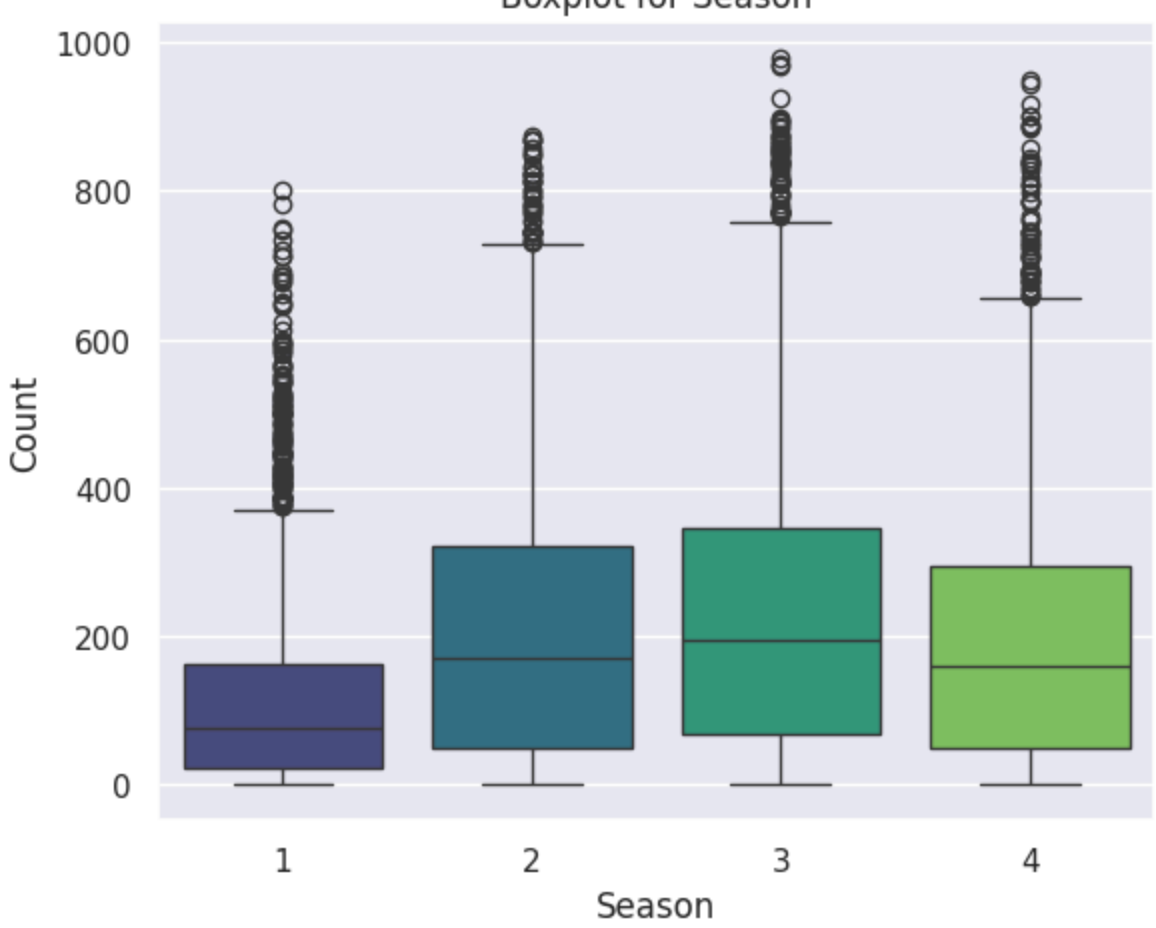

Season Distribution
Count
0
500
1000
1500
2000
2500
1
2
3
4
Season

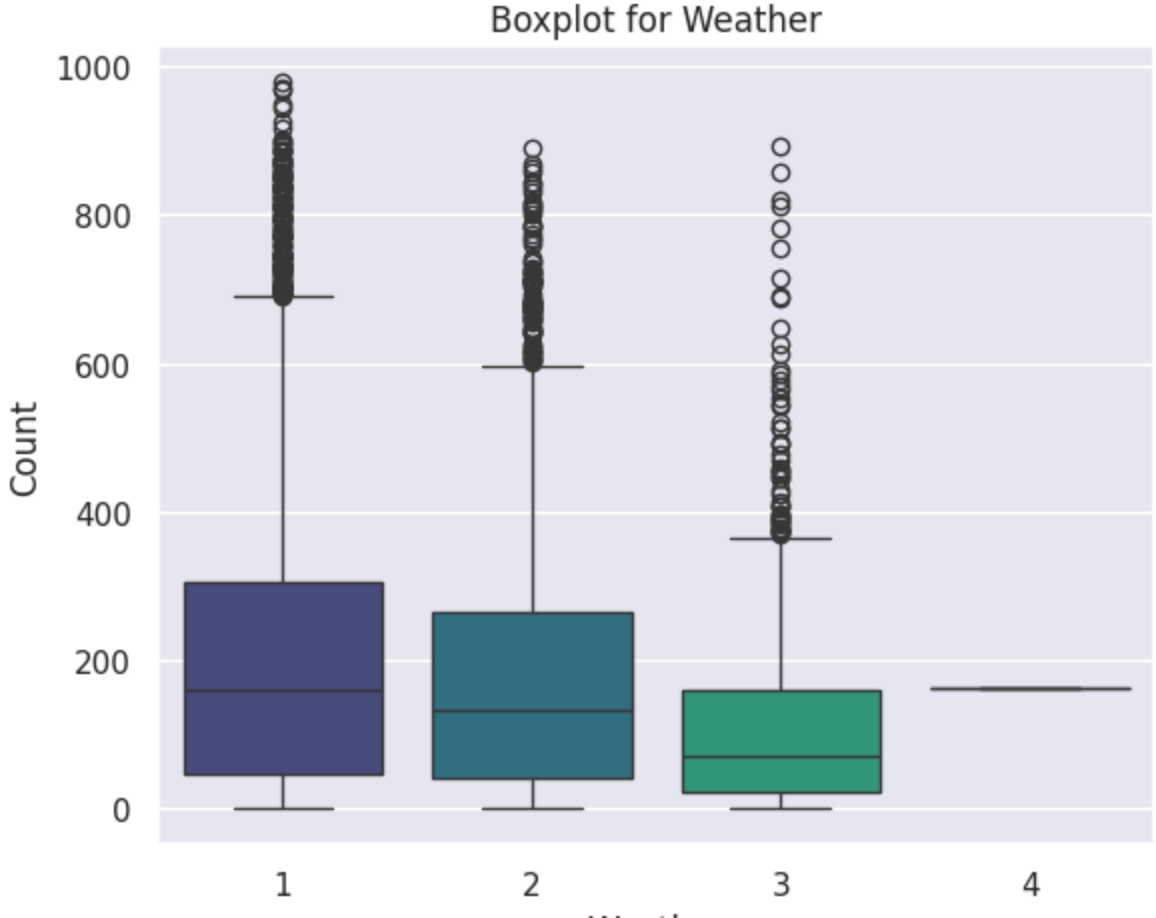

H
Count
0
2000
4000
6000
8000
10000
0

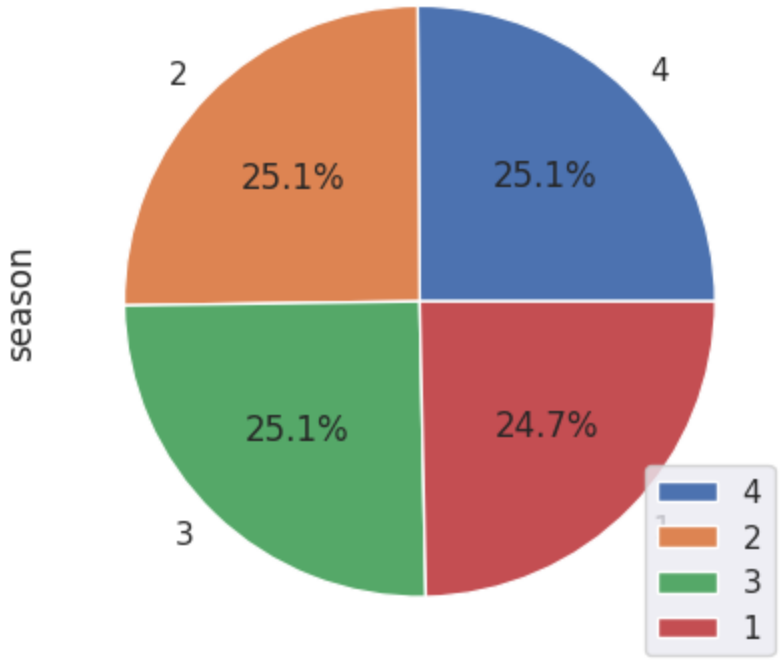

Weather Distribution
Count
0
1000
2000
3000
4000
5000
6000
7000
1
2
3
4
Weather


Wor
Count
0
1000
2000
3000
4000
5000
6000
7000
0

Season


Boxplot for Season
Count
1000
800
600
400
200
0
1
2
3
4
Season


Boxplot for Weather
Count
1000
800
600
400
200
0
1
2
3
4

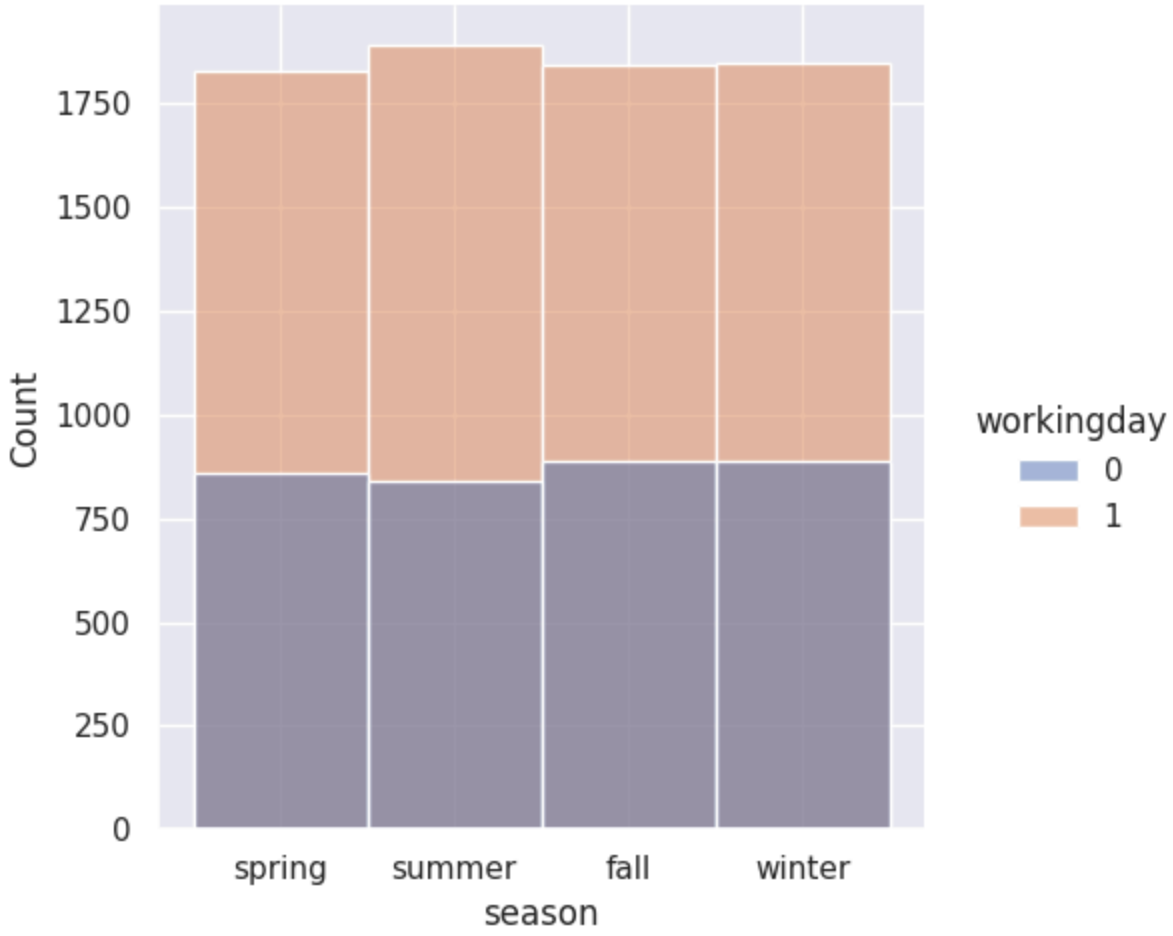


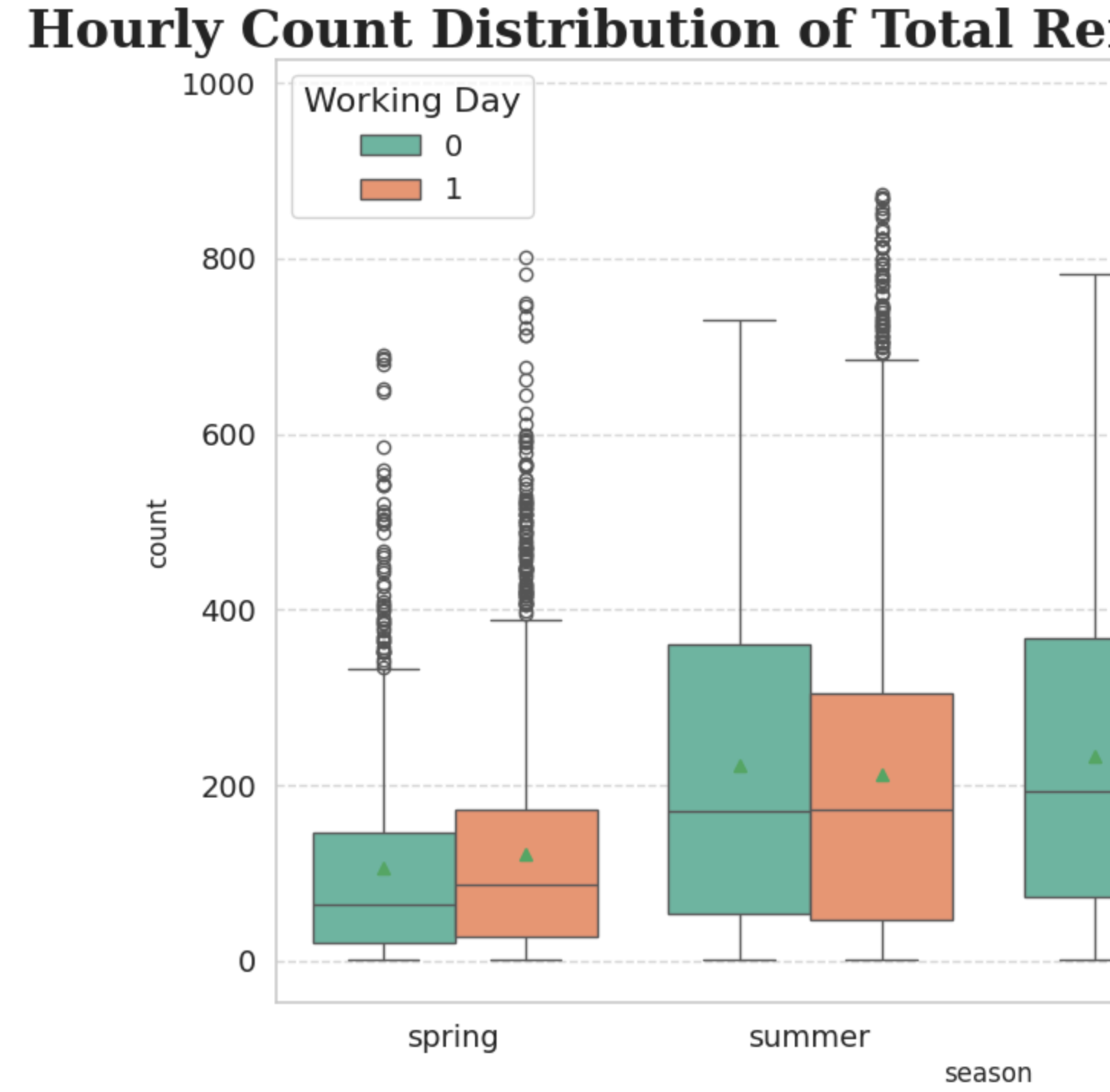
Observations

- The data appears typical and as expected.
- There is an equal distribution of days across each season.
- The number of working days surpasses that of non-working days.
- Predominantly, the weather is characterized by clear conditions.

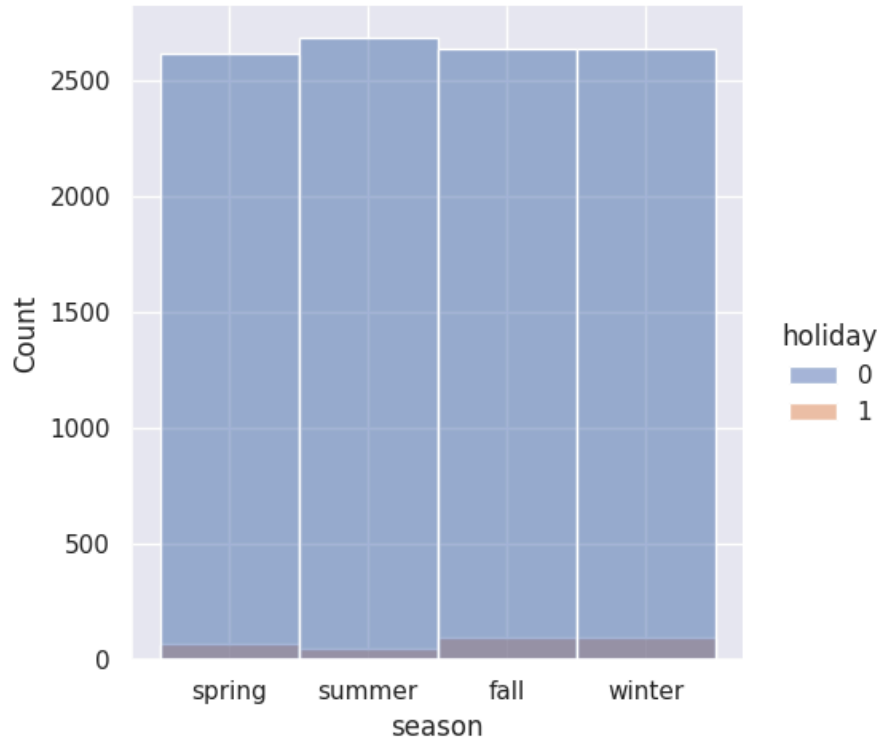
Bi-variate Analysis

Season vs Working Day

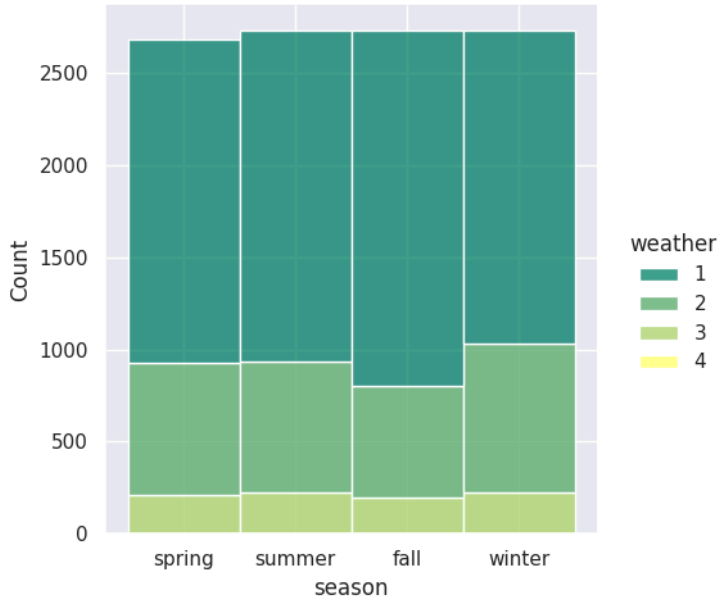

1750
1500
1250
1000
750
500
250
0
Count
spring
summer
fall
winter
season
workingday
0
1



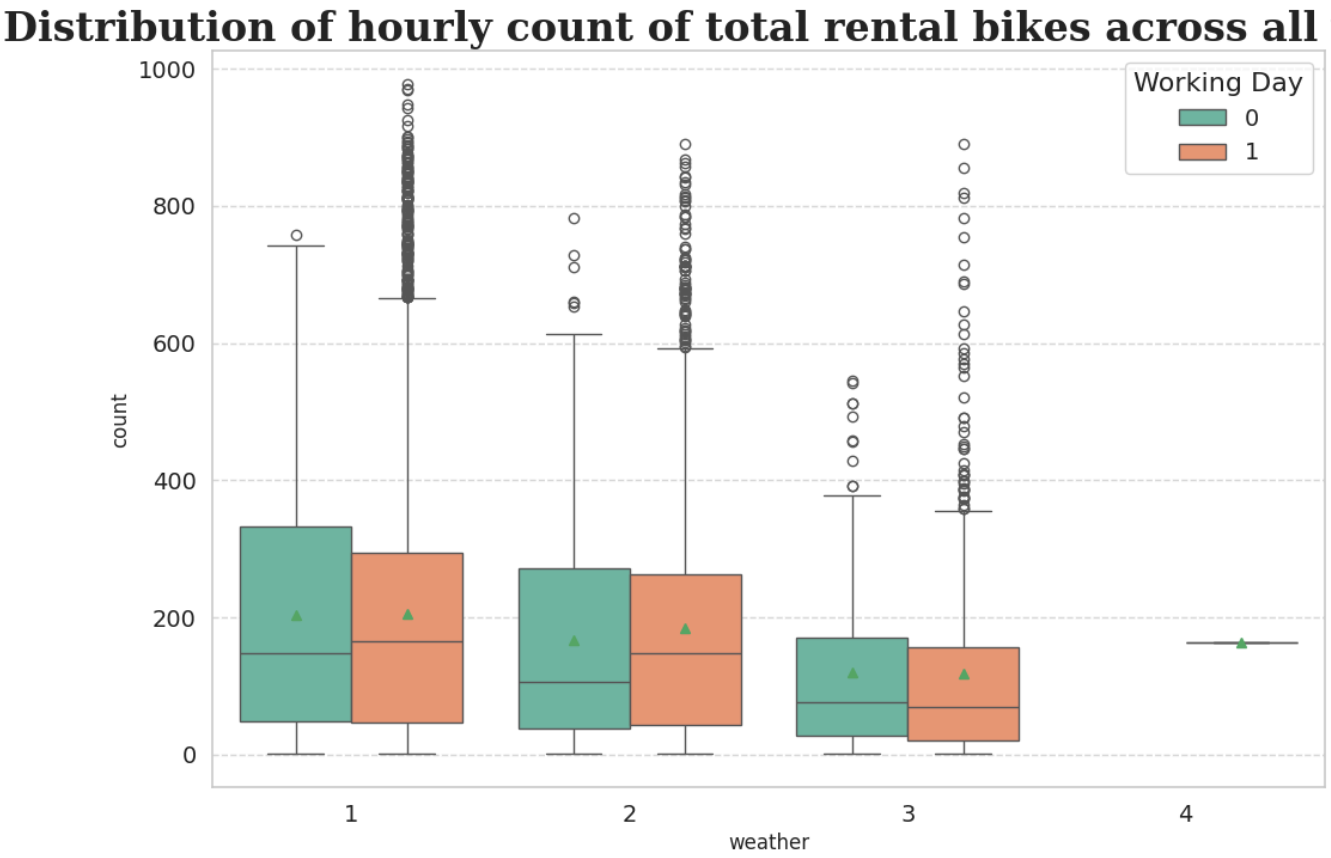
Season vs Holiday Day



Season vs Weather

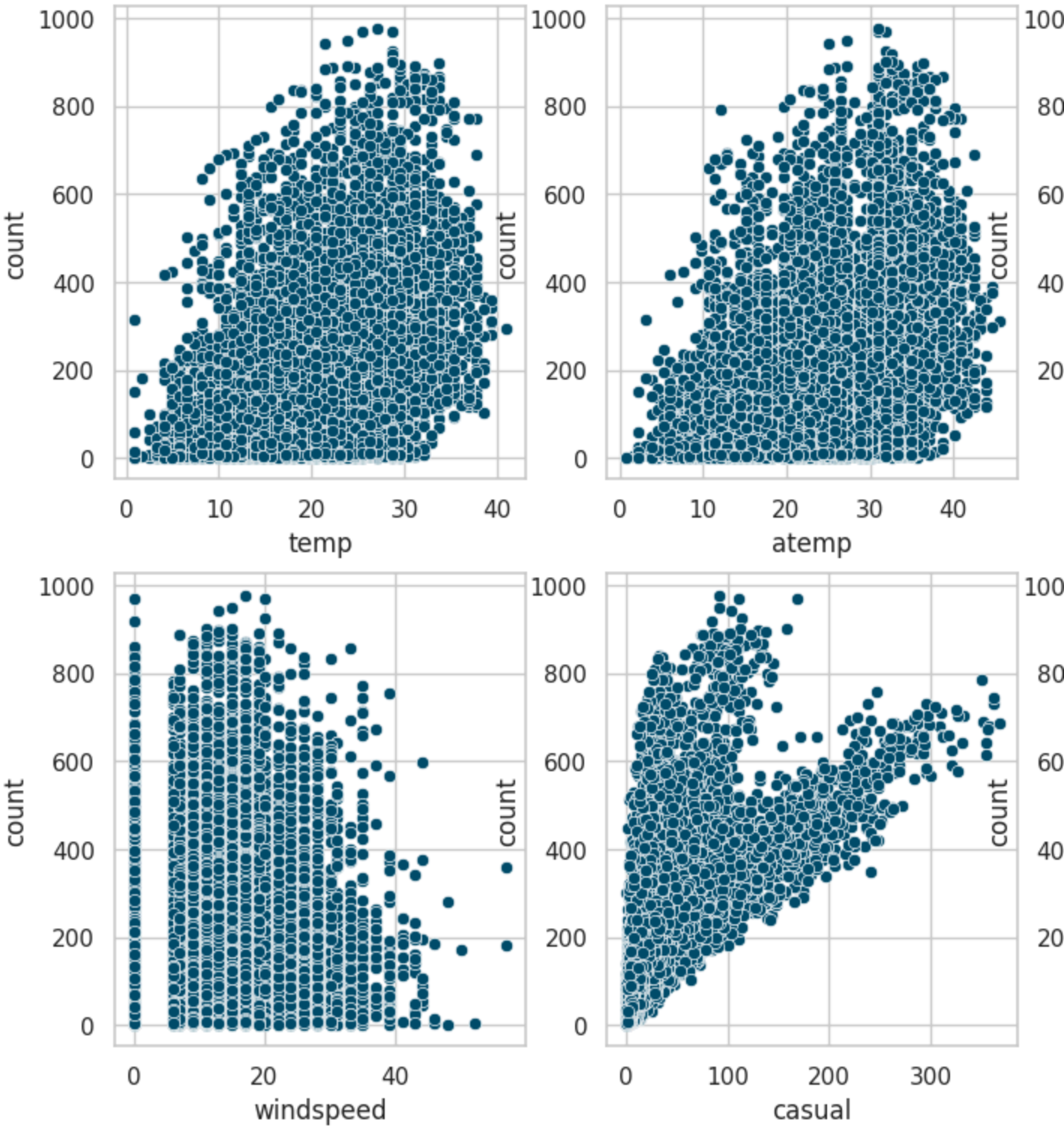


Weather vs Working Day



Observations

- During the summer and fall seasons, there is a higher rental frequency compared to other
- seasons.
- Holidays consistently result in increased bike rentals.
- Notably, on weekends or holidays, there is a slight uptick in bike rentals.
- Conversely, instances of rain, thunderstorms, snow, or fog are associated with a decrease in bike rentals.



Observations

- Bike rentals significantly drop when humidity levels fall below 20.

- Lower temperatures, specifically below 10, correspond to a decrease in the number of rented bikes.
- Elevated windspeeds exceeding 35 result in a reduced number of bike rentals.

Insights

- During the summer and fall seasons, there is a higher rental frequency compared to other
- seasons.
- Holidays consistently result in increased bike rentals.
- Notably, on weekends or holidays, there is a slight uptick in bike rentals.
- Conversely, instances of rain, thunderstorms, snow, or fog are associated with a decrease in bike rentals.
- Bike rentals significantly drop when humidity levels fall below 20.
- Lower temperatures, specifically below 10, correspond to a decrease in the number of rented bikes.
- Elevated windspeeds exceeding 35 result in a reduced number of bike rentals.

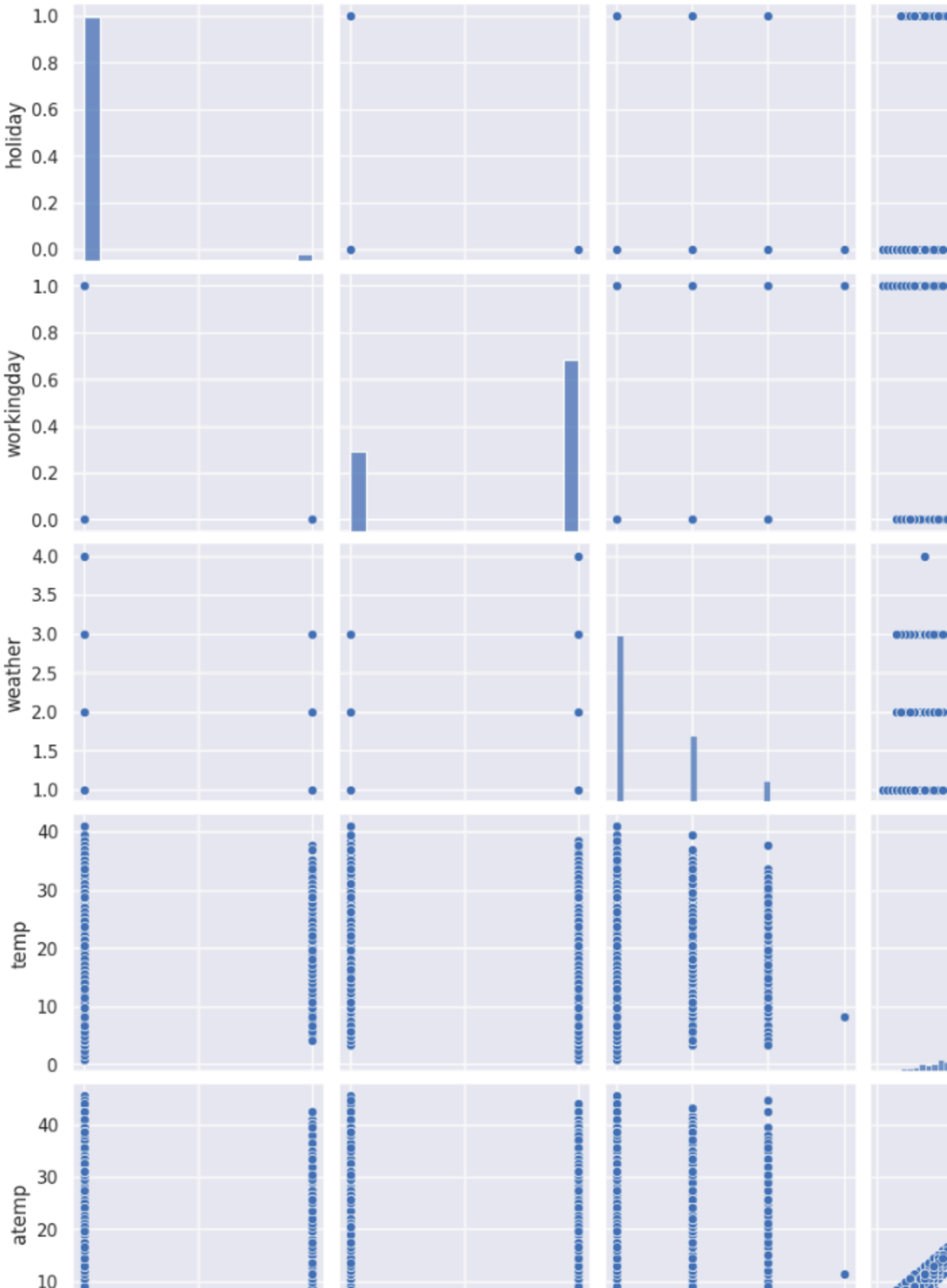
Multivariate Analysis

Establishing relationships between between dependent and independent variables

Heat-Map


holiday
workingday
temp
atemp
humidity
windspeed
casual
registered
count
1 -0.25 0.00029 -0.0052 0.0019 0.0084 0.044 -0.021
-0.25 1 0.03 0.025 -0.011 0.013 -0.32 0.12
0.00029 0.03 1 0.98 -0.065 -0.018 0.47 0.32
-0.0052 0.025 0.98 1 -0.044 -0.057 0.46 0.31
0.0019 -0.011 -0.065 -0.044 1 -0.32 -0.35 -0.27
0.0084 0.013 -0.018 -0.057 -0.32 1 0.092 0.091
0.044 -0.32 0.47 0.46 -0.35 0.092 1 0.5
-0.021 0.12 0.32 0.31 -0.27 0.091 0.5 1
-0.0054 0.012 0.39 0.39 -0.32 0.1 0.69 0.97
holiday workingday temp atemp humidity windspeed casual registered

Pairplot


holiday
workingday
weather
temp
atemp

Correlation data

index	holiday	workingday	temp	atemp
holiday	1.0	-0.2504913911873013	0.00029460339338629 49	-0.0052147782243563 91
workingday	-0.2504913911873013	1.0	0.02996554721480034	0.0246603293210920 47
temp	0.00029460339338629 49	0.02996554721480034	1.0	0.9849481104817075
atemp	-0.0052147782243563 91	0.02466032932109204 7	0.9849481104817075	1.0
humidity	0.0019287112385996 507	-0.0108798451249581 02	-0.0649487709012101 1	-0.0435357090825562 3
windspeed	0.0084087377823242 52	0.013373312551904 14	-0.0178520098613467 9	-0.0574730023281982 14
casual	0.0437989286753464 95	-0.319110963404297	0.46709706412013263	0.46206653642600143
registered	-0.0209556729353309 67	0.11945985076843728	0.31857128033739074	0.3146353862742616
count	-0.0053929844777745 49	0.01159386609157489	0.3944536449672518	0.3897844366269755 4

Observations

- There is a strong correlation between the columns [atemp, temp] and [count, registered], indicating a very high level of association.
- For all other combinations of columns, the correlation is minimal or negligible, suggesting a lack of significant association between those variables.
- There are no columns exhibiting a high positive or negative correlation (0.7 - 0.9), emphasizing the absence of strong linear relationships between any pairs of columns in the dataset.

Checking if there any significant difference between the number of bike rides on Weekdays and Weekends

---

STEP-1 : Set up Null Hypothesis

Null Hypothesis(Ho): Working day has no effect on the number of cycles being rented.
Alternate Hypothesis(Ha): Working day has an effect on the number of cycles being rented.

STEP-2 : Checking for basic assumptions for the hypothesis

We will use the 2-Sample T-Test to test this hypothesis

Before conducting the two-sample T-Test we need to find if the given data groups have the same variance. If the ratio of the larger data groups to the small data group is less than 4:1 then we can consider that the given data groups have equal variance.

Here, the ratio is 34040.70 / 30171.35 which is less than 4:1

STEP-3 : Set a significance level

Significance level (alpha): 0.05

STEP-4 : Calculate test Statistics / p-value

T-Statistic : -1.2096277376026694

P Value : 0.22644804226361348

STEP-5 : Deciding whether to accept or reject the Null Hypothesis.

P-value is greater than 0.05, As the P-value is greater than the predetermined level of significance (alpha), We don't have sufficient evidence to say that working day has an effect on the number of cycles being rented.

Final Outcome:

- The mean hourly count of the total rental bikes is statistically similar for both working and non- working days.

Checking if the demand of bicycles on rent is the same for different Weather conditions

---

STEP-1 : Set up Null Hypothesis

Null Hypothesis(Ho): Mean of cycle rented per hour is same for weather 1, 2 and 3. (ignoring weather 4 as there in only 1 data point)
Alternate Hypothesis(Ha): Mean of cycle rented per hour is not same for season 1,2,3 and 4 are different.

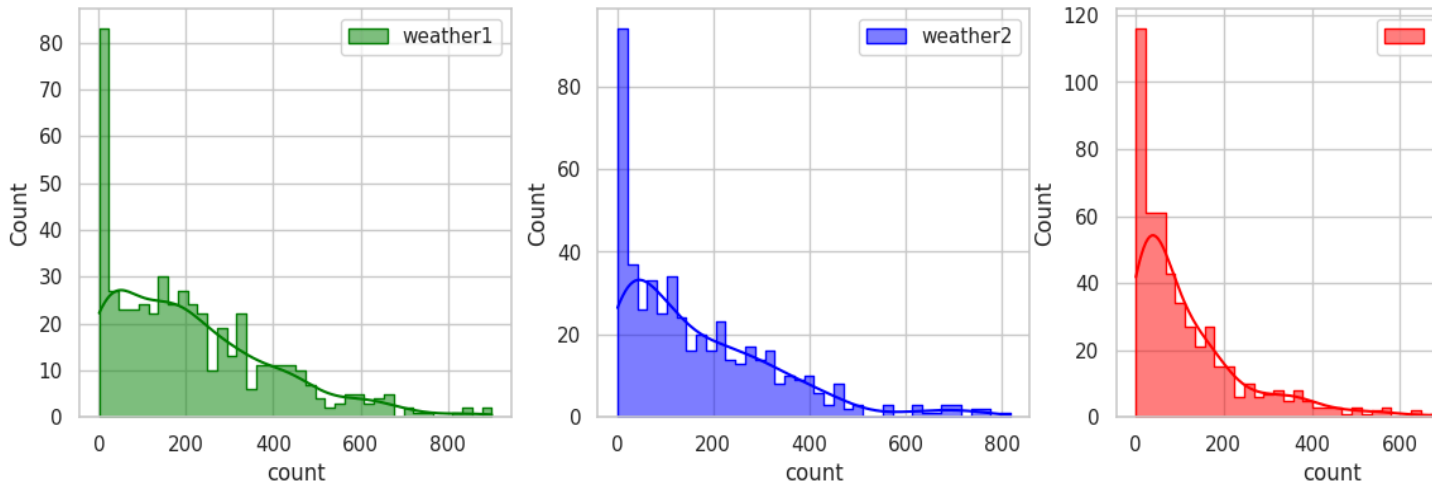
STEP-2 : Checking for basic assumptions for the hypothesis

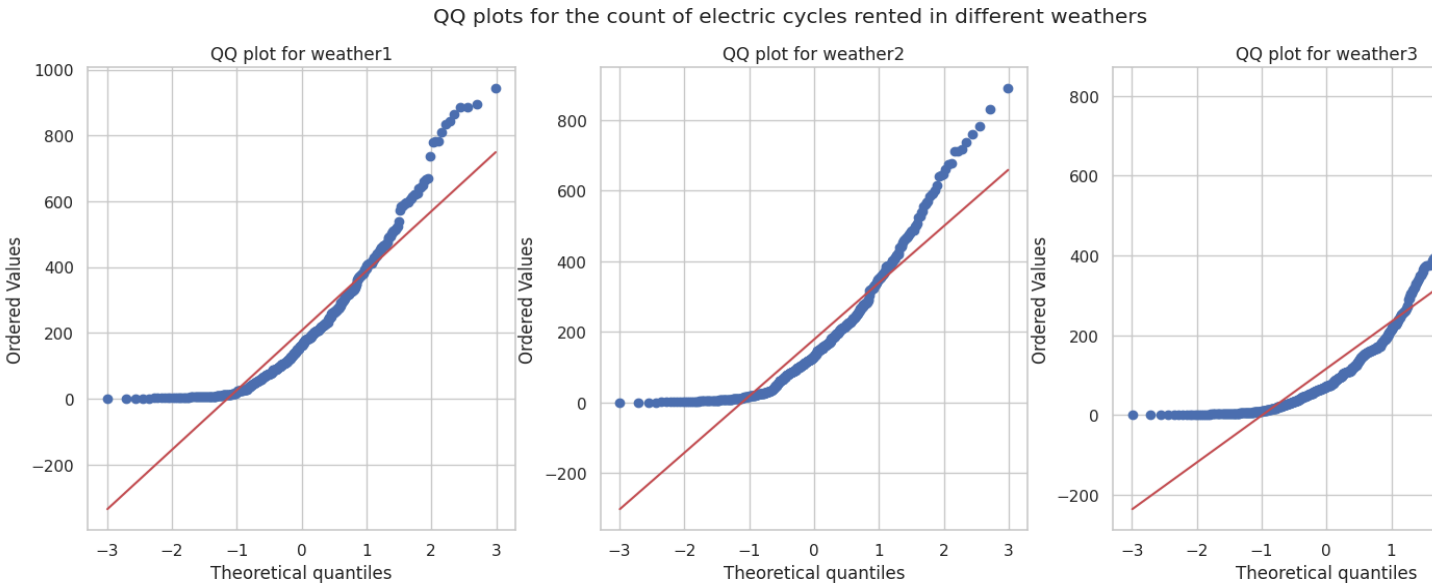
We will use One-way ANOVA to test this hypothesis

Before conducting the One-way ANOVA test, we need to check the assumptions of the test using the normality of histogram and QQ plot.

Normality

QQ Plot





From the above plot we can say that the distributions do not follow normal distribution.

Shapiro-Wilk's Test

Applying Shapiro-Wilk test for normality
alpha: 0.05

Null Hypothesis(Ho): The sample follows normal distribution
Alternate Hypothesis(Ha): The sample does not follow normal distribution

For weather-1

p-value 8.50238582320673e-19
The sample does not follow normal distribution

For weather-2

p-value 1.6130916333821363e-18
The sample does not follow normal distribution

For weather-3

p-value 2.5429566836680963e-27
The sample does not follow normal distribution

- As the P-value is less than 0.05 (alpha) for all the three weathers, We reject the null hypothesis and conclude that the sample does not follow normal distribution.

Equality Variance

Applying Levene's test for checking Homogeneity of Variances

Null Hypothesis(Ho): Homogenous Variance
Alternate Hypothesis(Ha): Non-Homogenous Variance

p-value 2.445146765480774e-13
The samples do not have Homogenous Variance

- As the P-value is less than 0.05 (alpha) for all the three weathers, We reject the null hypothesis and conclude that the sample does not have Homogenous Variance.

Since the samples are not normally distributed and do not have the same variance, f_oneway test cannot be performed here, we can perform its non parametric equivalent test.

Applying Kruskal-Wallis Test for independent samples.

Null Hypothesis(Ho): Mean no. of cycles rented is same for different weather
Alternate Hypothesis(Ha): Mean no. of cycles rented is different for different weather

---

STEP-3 : Set a significance level

Significance level (alpha): 0.05

STEP-4 : Calculate test Statistics / p-value

Test Statistic = [1.36471292e+01 1.83091584e+00 5.37649760e+00 1.56915686e+01

1.08840000e+04 3.70017441e+01 4.14298489e+01 1.83168690e+03

2.80380482e+01 2.84639685e+02 1.73745440e+02 2.04955668e+02]

p value = [1.08783632e-03 4.00333264e-01 6.79999165e-02 3.91398508e-04

0.00000000e+00 9.22939752e-09 1.00837627e-09 0.00000000e+00

8.15859150e-07 1.55338046e-62 1.86920588e-38 3.12206618e-45]

STEP-5 : Deciding whether to accept or reject the Null Hypothesis.

- P-value is less than 0.05.
- As the P-value is less than the predetermined level of significance (alpha), We have sufficient evidence to say that the average number of rental bikes is statistically different for different weathers.

Final Outcome:

- The hourly total number of rental bikes is statistically different for different weathers.

Checking if the demand of bicycles on rent is the same for different Seasons

---

STEP-1 : Set up Null Hypothesis

Null Hypothesis(Ho): Mean of cycle rented per hour is same for season 1,2,3 and 4.
Alternate Hypothesis(Ha): Mean of cycle rented per hour is different for season 1,2,3 and 4.

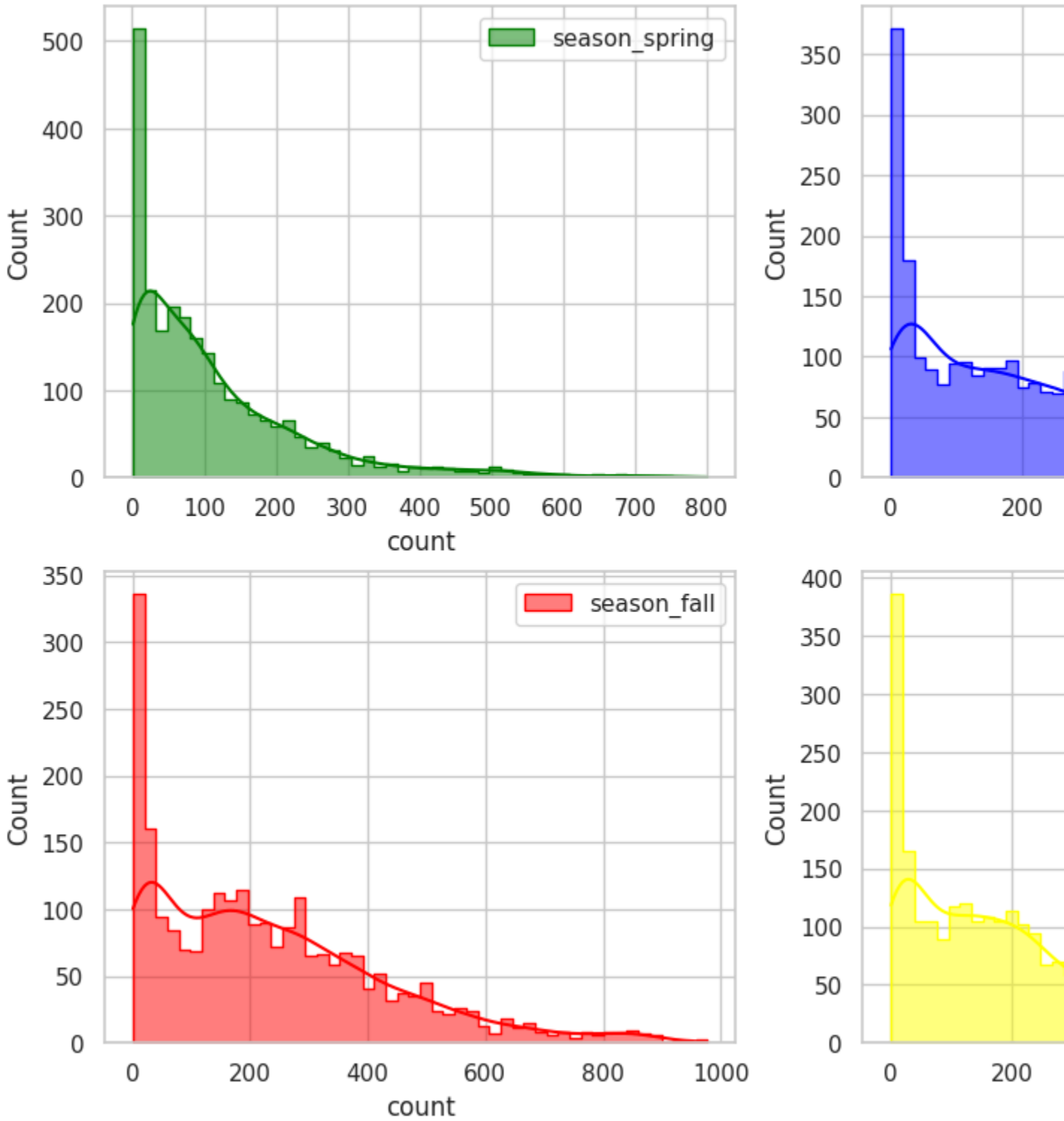
STEP-2 : Checking for basic assumptions for the hypothesis

We will use One-way ANOVA to test this hypothesis

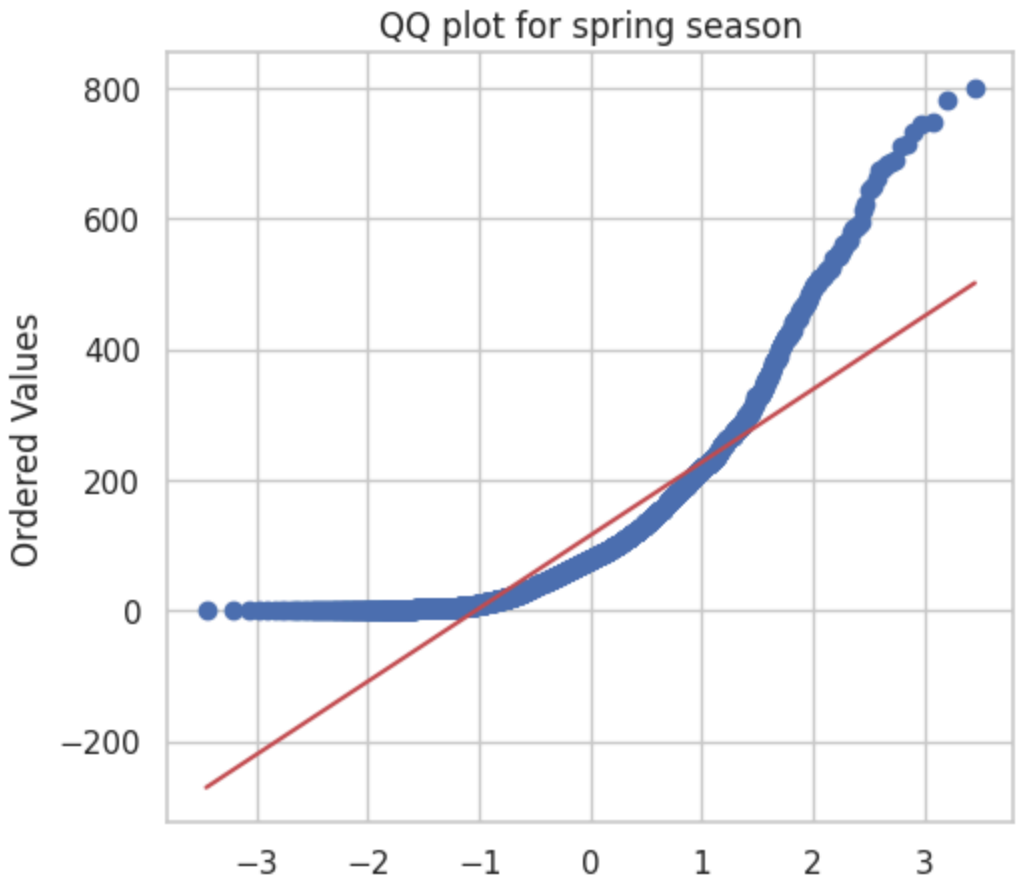
Before conducting the One-way ANOVA test, we need to check the assumptions of the test using the normality of histogram and QQ plot.

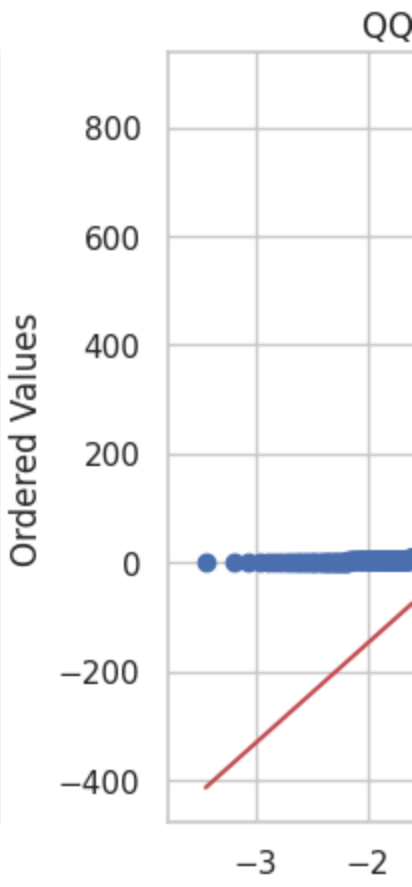
Normality

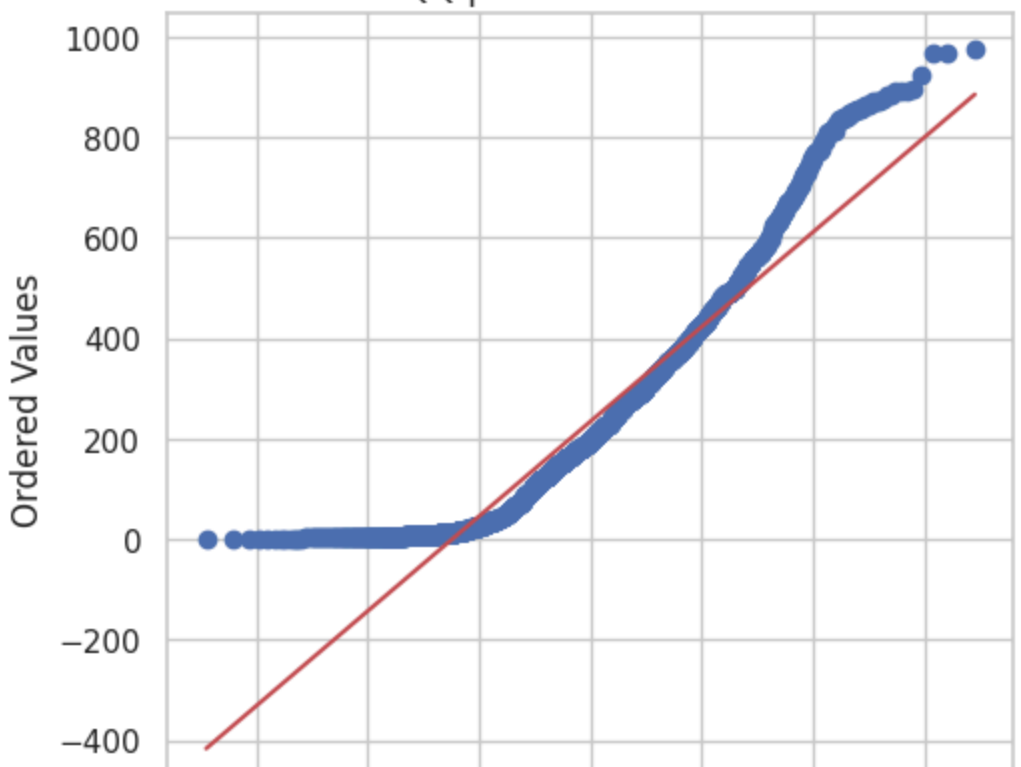
QQ Plot

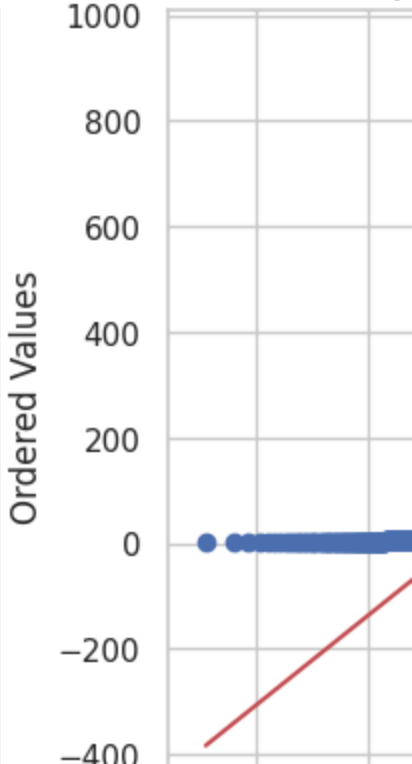


QQ plots for the count of electric cycles rented in di









From the above plot we can say that the distributions do not follow normal distribution.

Shapiro-Wilk's Test

Applying Shapiro-Wilk test for normality
alpha: 0.05

Null Hypothesis(Ho): The sample follows normal distribution
Alternate Hypothesis(Ha): The sample does not follow normal distribution

For Spring

p-value 0.0
The sample does not follow normal distribution

For Summer

p-value 3.4752089811378317e-37
The sample does not follow normal distribution

For Fall

p-value 2.1519158921721582e-35
The sample does not follow normal distribution

For Winter

p-value 5.575758742277047e-38
The sample does not follow normal distribution

- As the P-value is less than 0.05 (alpha) for all the four seasons, We reject the null hypothesis and conclude that the sample does not follow normal distribution.

Equality Variance

Applying Levene's test for checking Homogeneity of Variances

Null Hypothesis(Ho): Homogenous Variance
Alternate Hypothesis(Ha): Non-Homogenous Variance

p-value 1.6144901344317356e-109
The samples do not have Homogenous Variance

- As the P-value is less than 0.05 (alpha) for all the four seasons, We reject the null hypothesis and conclude that the sample does not have Homogenous Variance.

Since the samples are not normally distributed and do not have the same variance, f_oneway test cannot be performed here, we can perform its non parametric equivalent test.

Applying Kruskal-Wallis Test for independent samples.

Null Hypothesis(Ho): Mean no. of cycles rented is same for different seasons
Alternate Hypothesis(Ha): Mean no. of cycles rented is different for different seasons

---

STEP-3 : Set a significance level

Significance level (alpha): 0.05

STEP-4 : Calculate test Statistics / p-value

Test Statistic = 699.6668548181988

p value = 2.479008372608633e-151

Reject the null hypothesis

STEP-5 : Deciding whether to accept or reject the Null Hypothesis.

- P-value is less than 0.05.
- As the P-value is lesser than the predetermined level of significance (alpha), We have sufficient evidence to say that the average number of rental bikes is statistically different for different seasons.

Final Outcome:

- The hourly total number of rental bikes is statistically different for different seasons.

Checking if the Weather conditions are significantly different during different Seasons

Observed values:

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

STEP-1 : Set up Null Hypothesis

Null Hypothesis(Ho): Weather is independent of the season.
Alternate Hypothesis(Ha): Weather is not independent of the season.

STEP-2 : Checking for basic assumptions for the hypothesis

We will use chi-square test to test this hypothesis

STEP-3 : Set a significance level

Significance level (alpha): 0.05

STEP-4 : Calculate test Statistics / p-value

chi-square test statistic: 49.158655596893624

P_value: 1.549925073686492e-07

degrees of freedom: 9

Weather conditions are significantly different during different Seasons

---

STEP-5 : Deciding whether to accept or reject the Null Hypothesis.

- P-value is less than 0.05.
- Since p-value is less than the alpha 0.05, We reject the Null Hypothesis. meaning that Weather conditions are significantly different during different Seasons.

Final Outcome:

- There is statistically significant dependency of weather 1, 2, 3 on season based on the average hourly total number of bikes rented.

Insights

- **Temperature and Bike Counts:**
  The temperature columns, feeling temperature and temperature, as well as the bike count columns, count and registered, show a strong connection. This means they are closely related to each other.
- **Other Column Combinations:**
  For all other pairs of columns, the connection or correlation is very weak. This suggests that those variables are not significantly linked to each other.
- **No Strong Linear Relationships:**
  None of the column pairs have a high positive or negative correlation (between 0.7 and 0.9). This indicates there are no strong linear relationships between any two columns in the dataset.
- **Bike Counts on Working and Non-Working Days:**
  On average, the number of rental bikes per hour is similar on both working and non-working days.
- **Bike Counts in Different Weathers:**
  The total number of rental bikes per hour varies significantly with different weather conditions.
- **Bike Counts in Different Seasons:**

The total number of rental bikes per hour also varies significantly across different seasons.

- **Weather and Seasonal Dependency:**
  The average hourly bike rental numbers show a statistically significant relationship between weather conditions (1, 2, 3) and seasons.

Recommendations

- **Strategic Promotions:**
  - Implement weather-based promotions
  - offering discounts during clear and cloudy days when bike rentals are highest, attracting more customers to Yulu.
- **Flexible Pricing and Inventory:**
  - Introduce time-based pricing to balance demand, with lower rates during off-peak hours.
  - Optimize inventory based on seasonal demand, ensuring ample bikes during summer and fall and adjusting during low-demand months.
- **Customer-Centric Approach:**
  - Encourage customer feedback to enhance services and tailor offerings to meet customer expectations.
  - Leverage social media for marketing, showcasing biking experiences, engaging with customers.
  - Running targeted campaigns for increased bookings.
