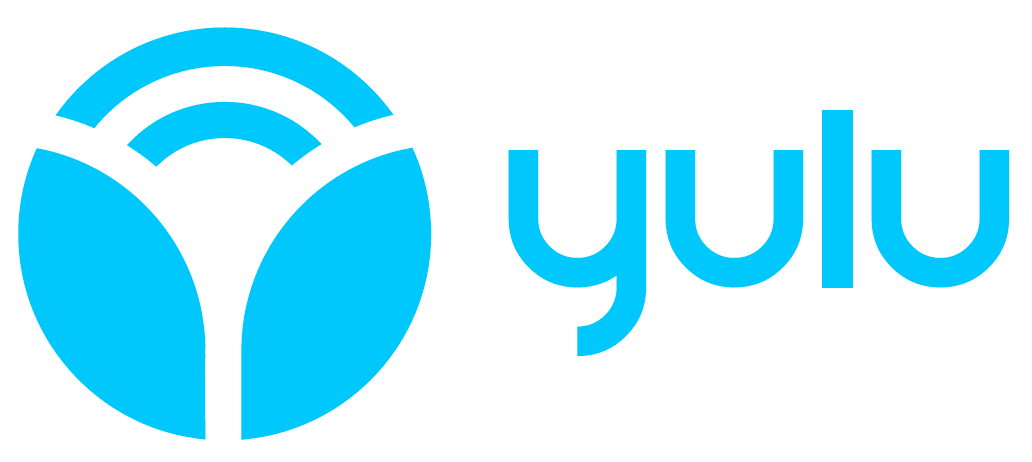
**Business Case: Yulu - Hypothesis Testing**



# Introduction

Yulu is India’s leading micro-mobility service provider, which offers unique vehicles for the daily commute. Starting off as a mission to eliminate traffic congestion in India, Yulu provides the safest commute solution through a user-friendly mobile app to enable shared, solo, and sustainable commuting.

Yulu zones are located at all the appropriate locations (including metro stations, bus stands, office spaces, residential areas, corporate offices, etc) to make those first and last miles smooth, affordable, and convenient!

**Objectives:**

In response to Yulu's concern regarding the decline in revenues and the need to understand the factors affecting the demand for shared electric cycles in the Indian market, our study aims to provide the following insights:

1. Identifying Significant Predictors
2. Understanding the Relationship Between Variables and Demand

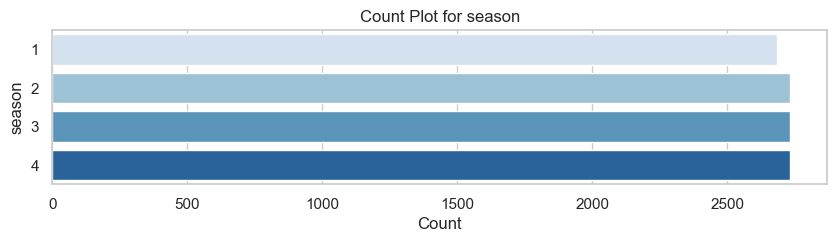
Ultimately, our study will offer Yulu a data-driven understanding of the key factors affecting demand, and provide recommendations for leveraging these factors to optimize operations and improve revenue in the shared electric cycle market in India.

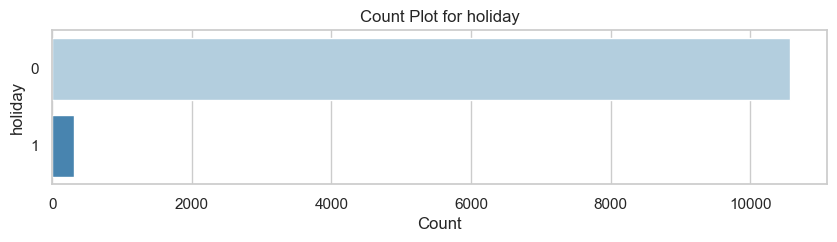
# Let Us Understand Data:

The dataset under analysis provides hourly bike rental data with a total of 10,886 entries across 12 columns, offering a detailed view of bike rental patterns and the factors that influence demand. The data captures both casual and registered bike rentals, alongside environmental and temporal factors, including weather conditions, temperature, humidity, windspeed, and various time-based variables such as season, holiday, and working day.

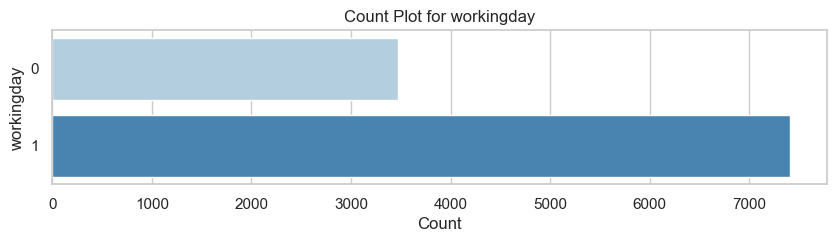
Key statistics reveal important insights into user behaviour and environmental influences on bike rentals. The average humidity level is approximately 61.9%, while wind speeds average around 12.8 km/h. Rental counts exhibits significant variability, highlighting the dynamic nature of bike usage across different conditions. This variation, alongside seasonal and time-of-day patterns, provides valuable information for understanding customer behaviour and optimizing bike-sharing operations. The data offers a comprehensive foundation for analysing the factors affecting bike rental demand, with implications for improving service delivery and revenue generation in the shared electric bike market.

**Let’s deeper look into each kind of Columns:**

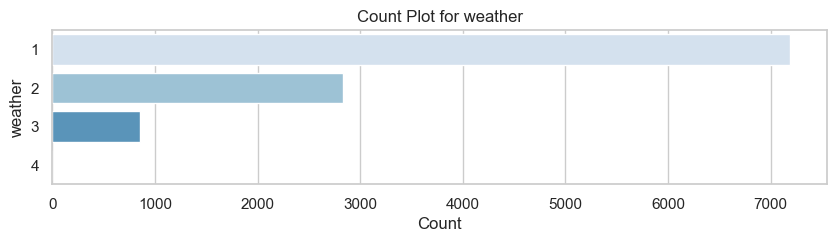
**Categorical Variable**The dataset is evenly distributed across **seasons**, with slightly more records for winter (4) (2734) and slightly fewer for spring (1) (2686).



Non-**holidays** (0) dominate the dataset, comprising 97.1% of records (10575), while only 2.9% (311) correspond to holidays.

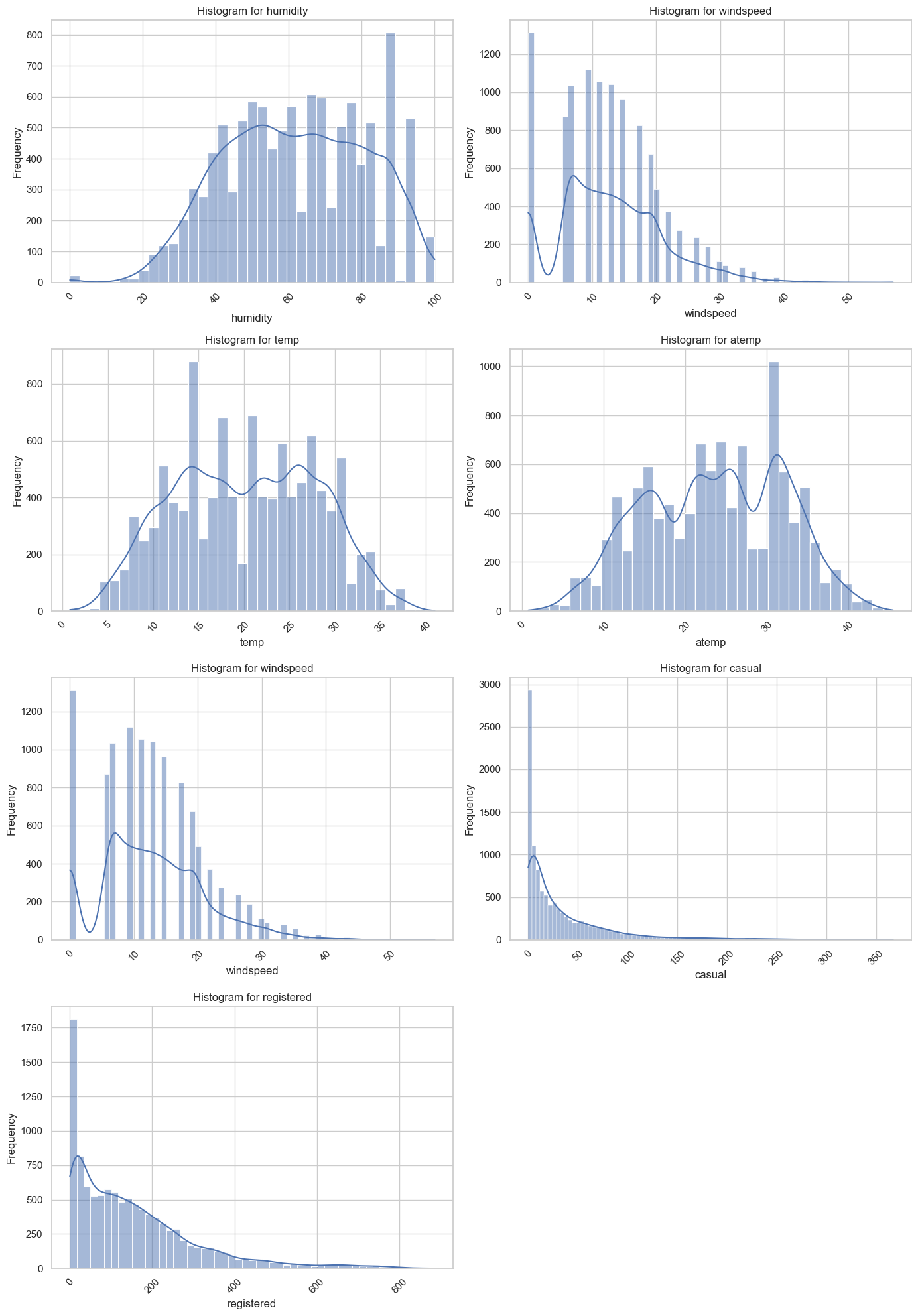


The majority of the days are **working days** (1), making up 68.1% of the dataset (7412), while non-working days (0) represent 31.9% (3474).



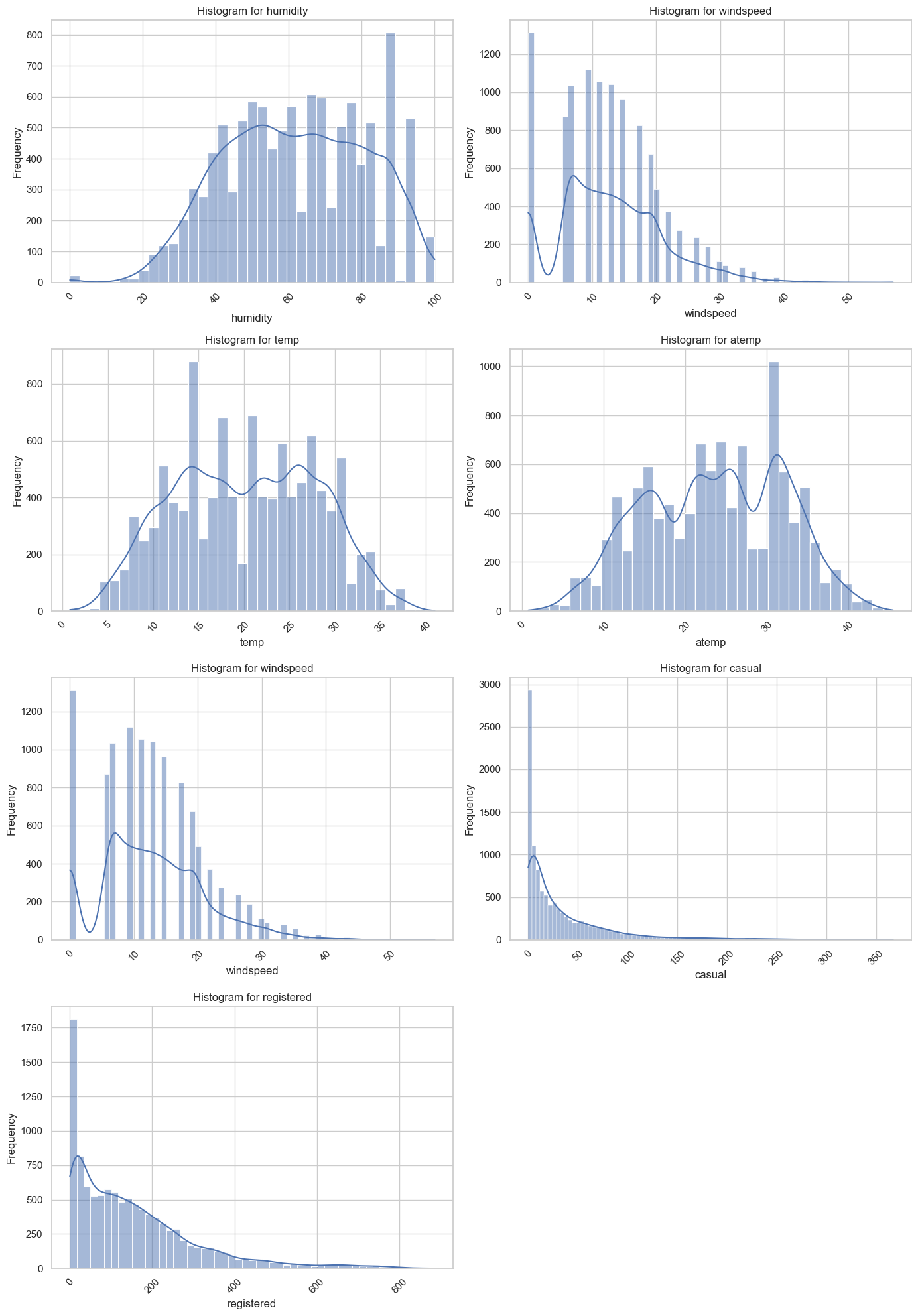
* The most common **weather** condition is Clear or Partly Cloudy (1), accounting for 68.1% of the data (7192).
* Mist-related **weather** (2) follows at 26.8% (2834), while Light Snow or Rain (3) is rare at 8.1% (859).
* Extreme **weather** conditions (4) are exceptionally rare, with just one recorded instance.

**Continuous Variables:**



**Humidity**: The distribution of humidity is nearly uniform, with peaks observed between 50-60%. Humidity values are widely spread, ranging from 0% to 100%, reflecting varying environmental conditions that could influence bike rental demand.

**Windspeed**: Windspeed data exhibits a strong right-skewed distribution, with the majority of wind speeds concentrated below 20 km/h. Only a few observations exceed wind speeds of 30 km/h, indicating that extreme wind conditions are rare but may still impact bike usage during those times.



**Temperature (Temp)**: The temperature data follows a slightly bimodal distribution, with the majority of values falling between 10°C and 30°C. This suggests that most bike rentals occur under moderate temperature conditions, which may affect user comfort and demand.

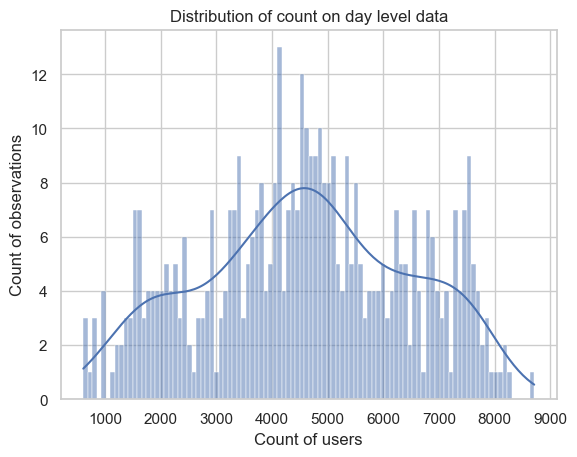
**Feels-like Temperature (Atemp)**: The distribution of the "feels-like" temperature mirrors the temperature data, but with a more pronounced peak around 20-30°C. This indicates that most users experience temperatures within this range, suggesting a clustering of rental demand during these conditions.

**Casual Users (Casual)**: The distribution of casual users is heavily right-skewed, with most values falling below 50 rentals. This points to a smaller casual user base, likely driven by occasional users who may rent bikes during favourable weather or specific events.

**Registered Users (Registered)**: The distribution for registered users also exhibits right-skewness, with most values concentrated below 200 rentals. This suggests that while the registered user base is larger, the frequency of rentals gradually declines as the number of rentals increases.

Analysing the Impact of Variables on Bike Rental Demand (Count):

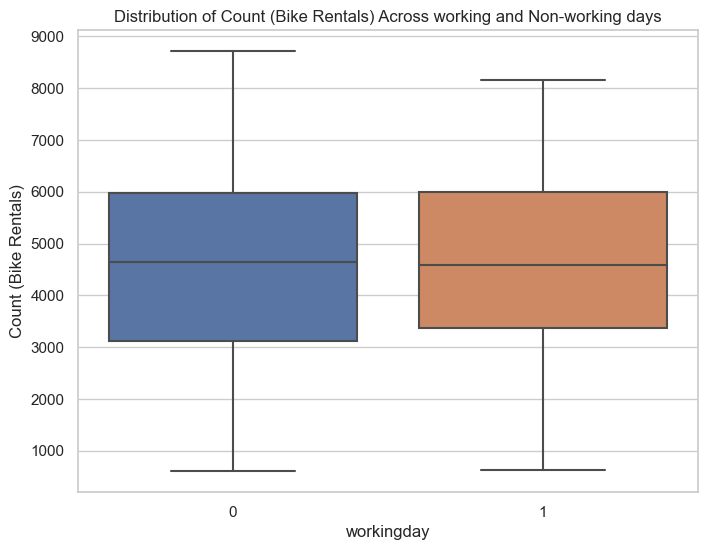
**Before conducting the bivariate analysis, let's first check whether the dependent variable follows a normal distribution:**



Given that the data appears to follow a normal distribution, parametric tests, such as t-tests and ANOVA, are appropriate for analysing the differences between groups and examining the effects of independent variables on the dependent variable.

**Categorical Variable:**

**Working day:**



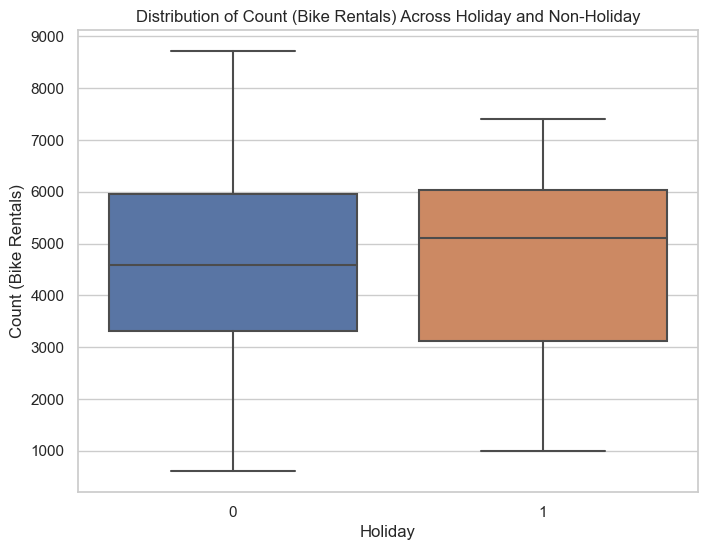
T-Test Analysis: Relationship Between Count and Working Day

The t-test was conducted to assess whether there is a significant difference in the mean count of users between working and non-working days.

* Null Hypothesis (H₀): The mean count of users on working days is equal to the mean count on non-working days, indicating no significant difference.
* Alternative Hypothesis (H₁): The mean count of users on working days is greater than on non-working days, suggesting that working days experience higher rental demand.

The t-statistic value was 0.44, and the p-value was 0.33. Since the p-value is greater than the standard significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is no significant difference in the number of rentals between working and non-working days, suggesting that working days do not consistently lead to higher bike rentals.

**Holiday:**

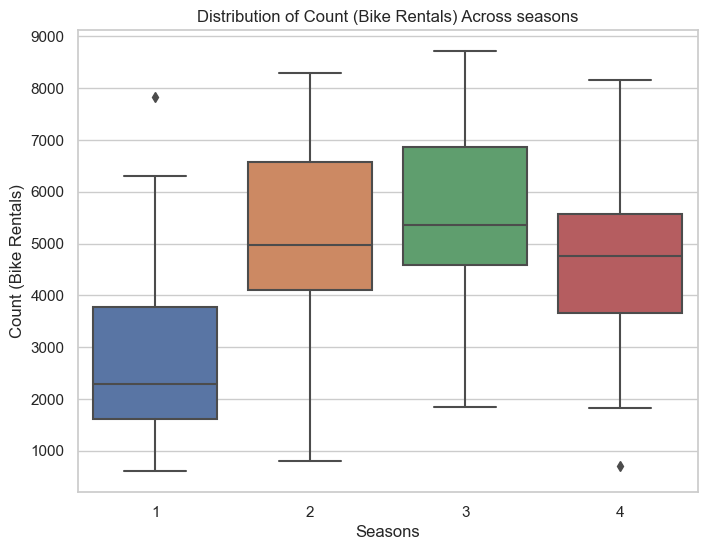
T-Test Analysis: Relationship Between Count and Holiday

The t-test was conducted to evaluate whether there is a significant difference in the mean count of users between holidays and non-holidays.

* Null Hypothesis (H₀): The mean count of users on holidays is equal to the mean count on non-holidays, suggesting no significant difference.
* Alternative Hypothesis (H₁): The mean count of users on holidays is greater than on non-holidays, implying that holidays experience fewer rentals.

The t-statistic value was -0.25, and the p-value was 0.40. Since the p-value is greater than the conventional significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is no significant difference in the number of rentals between holidays and non-holidays, suggesting that holidays do not result in consistently fewer bike rentals.

**Season**

ANOVA Analysis: Relationship Between Count and Season

Before conducting the ANOVA test, **Levene's Test** was performed to assess the equality of variances across the different seasons.

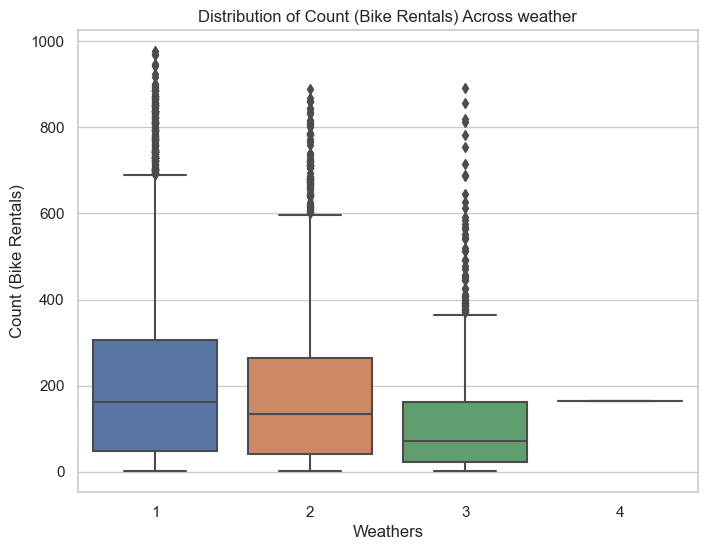
* Null Hypothesis for Levene's Test (H₀): There is no significant difference in variance between the different seasons (variances are equal across all groups).
* Alternative Hypothesis for Levene's Test (H₁): There is a significant difference in variance between at least two of the seasons (variances are not equal across the groups).

The **Levene's test statistic** was 1.51, with a p-value of 0.21. Since the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating that the variances are equal across the groups, and we can proceed with the ANOVA test.

For the **ANOVA test**:

* Null Hypothesis (H₀): The mean count of bike rentals is the same for all seasons, indicating no significant difference between seasons.
* Alternative Hypothesis (H₁): The mean count of bike rentals differs for at least one season.

The F-statistic was 80.05, and the p-value was extremely small (1.51e-41), which is well below the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant difference in bike rentals across the seasons.

**Weather**

Kruskal-Wallis Test: Relationship Between Count and Weather

Before conducting the Kruskal-Wallis test, **Levene's Test** was performed to assess the equality of variances across the different weather conditions.

* Null Hypothesis for Levene's Test (H₀): There is no significant difference in variance between the different weather conditions (variances are equal across all groups).
* Alternative Hypothesis for Levene's Test (H₁): There is a significant difference in variance between at least two of the weather conditions (variances are not equal across the groups).

The Levene's test statistic was 54.85, and the p-value was 3.50e-35. Since the p-value is extremely small (much smaller than 0.05), we reject the null hypothesis, indicating that the variances are significantly different across the weather conditions. Therefore, we consider using the Kruskal-Wallis test, as it is a non-parametric test that does not assume equal variances.

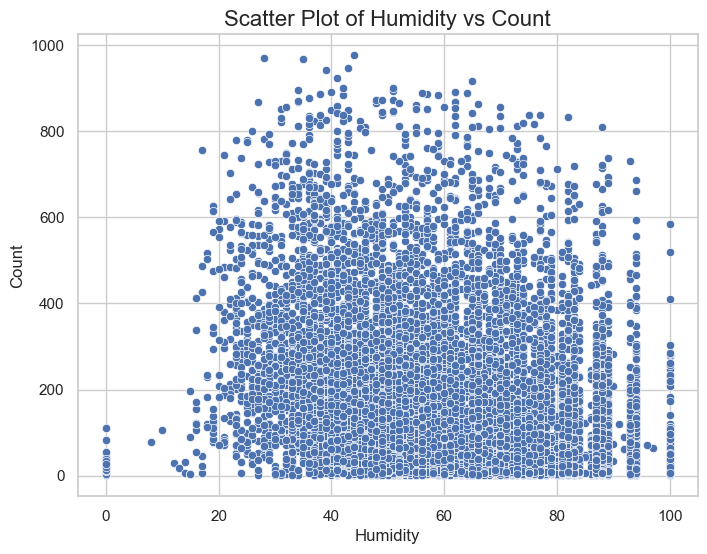
For the **Kruskal-Wallis test**:

* Null Hypothesis (H₀): The medians of the rental counts are equal across all weather conditions, meaning there is no significant difference in rentals based on weather.
* Alternative Hypothesis (H₁): At least one weather condition has a different median rental count, implying a significant difference between weather groups.

The Kruskal-Wallis H-statistic was 205.00, and the p-value was 3.50e-44. Since the p-value is very small and less than 0.05, we reject the null hypothesis. This suggests that there is a significant difference in bike rentals across different weather conditions, with at least one weather condition resulting in a notably different rental count.

**Continuous Variables:**

**Humidity**



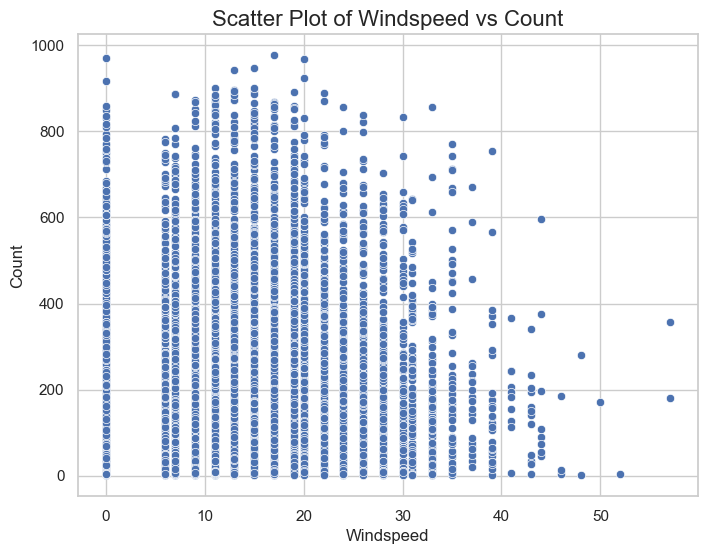
Pearson Correlation Test: Relationship Between Humidity and Count

* Null Hypothesis (H₀): There is no linear relationship between Humidity and Count.
* Alternative Hypothesis (H₁): There is a significant linear relationship between Humidity and Count.

The Pearson correlation coefficient is r=−0.317r = -0.317, which indicates a moderate negative linear relationship between Humidity and Count. This suggests that as humidity increases, the count of bike rentals tends to decrease.

The p-value is extremely small (approximately 0), which confirms that the relationship is statistically significant. However, the correlation is moderate in strength, suggesting that while humidity plays a role in influencing rental counts, other factors may also contribute to the variation in demand for bikes.

**Windspeed**

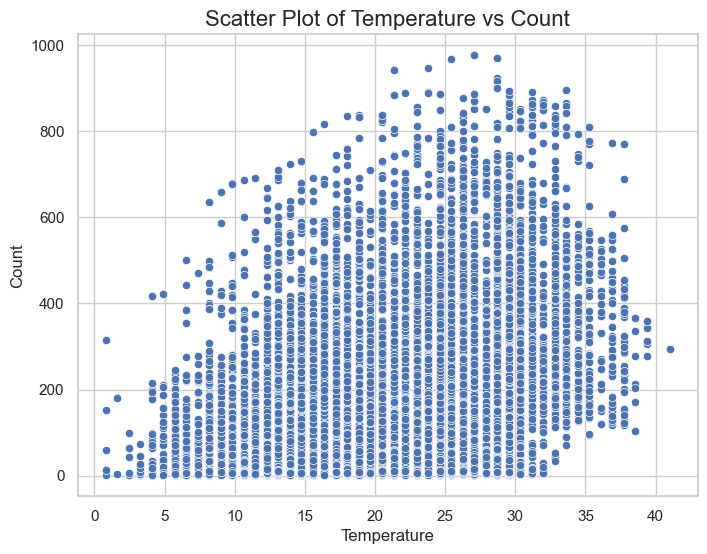
Pearson Correlation Test: Relationship Between Windspeed and Count

* **Null Hypothesis (H₀)**: There is no linear relationship between Windspeed and Count.
* **Alternative Hypothesis (H₁)**: There is a significant linear relationship between Windspeed and Count.

The Pearson correlation coefficient is r=0.101r = 0.101r=0.101, which indicates a weak positive linear relationship between Windspeed and Count. This suggests that as windspeed increases, the count of bike rentals slightly increases.

The extremely small p-value (approximately 0) confirms that this relationship is statistically significant. However, the weak correlation indicates that windspeed has minimal influence on bike rental counts, and other factors are likely contributing to the demand.

**Temperature**

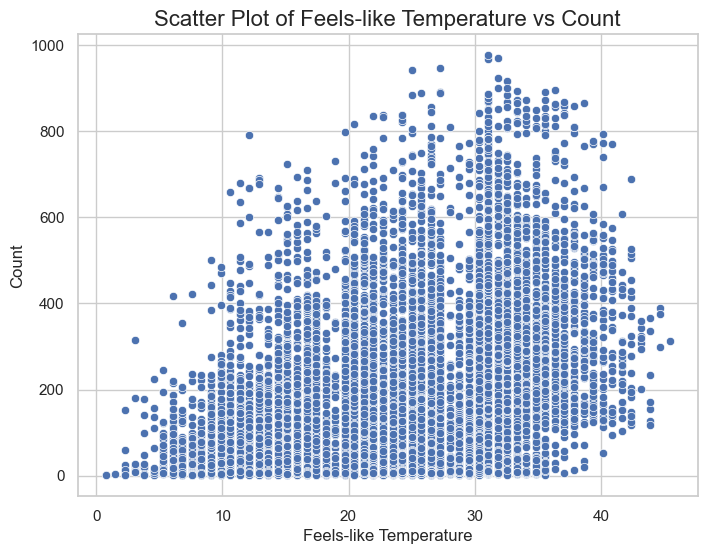
Pearson Correlation Test: Relationship Between Temperature and Count

* **Null Hypothesis (H₀)**: There is no linear relationship between Temperature and Count.
* **Alternative Hypothesis (H₁)**: There is a significant linear relationship between Temperature and Count.

The Pearson correlation coefficient is r=0.394r = 0.394, indicating a moderate positive linear relationship between Temperature and Count. This suggests that as temperature increases, the count of bike rentals tends to increase as well.

The p-value is 0, which confirms that this relationship is statistically significant. This suggests that temperature has a meaningful influence on the count of bike rentals, although other factors may also contribute to variations in demand.

**Feels-like Temperature**

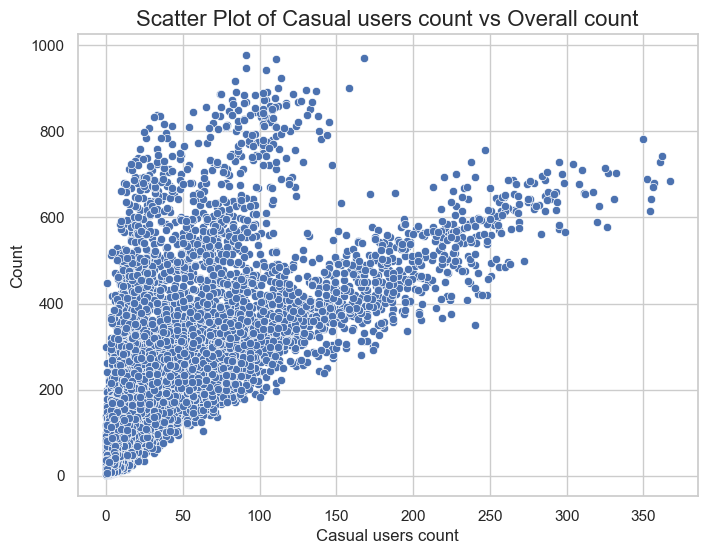
Pearson Correlation Test: Relationship Between Feels-Like Temperature and Count

* Null Hypothesis (H₀): There is no linear relationship between Feels-Like Temperature and Count.
* Alternative Hypothesis (H₁): There is a significant linear relationship between Feels-Like Temperature and Count.

The Pearson correlation coefficient is r=0.390r = 0.390, indicating a moderate positive linear relationship between Feels-Like Temperature and Count. This means that as the feels-like temperature increases, the count of bike rentals tends to increase as well.

The p-value is 0, which confirms that this relationship is statistically significant. This suggests that feels-like temperature has a moderate influence on the count of bike rentals, although other factors are likely to contribute to variations in demand.

**Casual Users count**



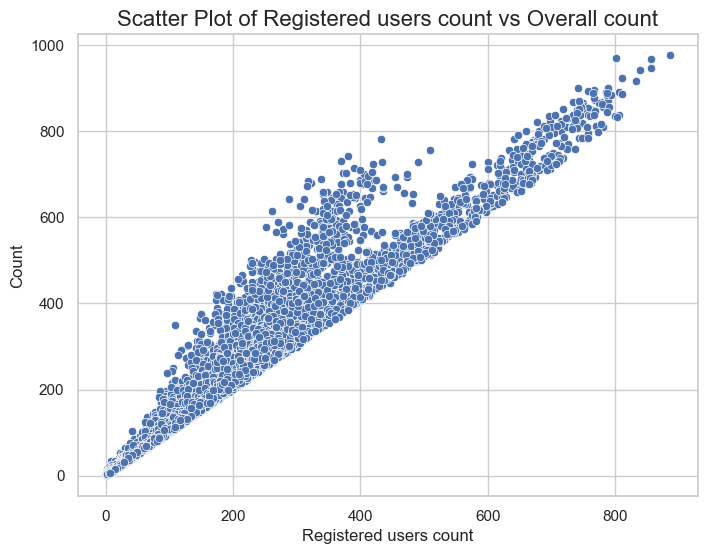
Pearson Correlation Test: Relationship Between Casual User Count and Overall User Count

* **Null Hypothesis (H₀)**: There is no linear relationship between Casual User Count and Overall User Count.
* **Alternative Hypothesis (H₁)**: There is a significant linear relationship between Casual User Count and Overall User Count.

The Pearson correlation coefficient is r=0.690r = 0.690r=0.690, indicating a strong positive linear relationship between the Casual User Count and the Overall User Count. This means that as the Casual User Count increases, the Overall User Count tends to increase as well.

The p-value is 0, which confirms that this relationship is statistically significant. This suggests that the association between Casual User Count and Overall User Count is not due to random chance. Casual users likely play a major role in influencing the Overall User Count, although other factors may still contribute to variations in the overall count.

**Registered Users count**

Pearson Correlation Test: Relationship Between Registered User Count and Overall User Count

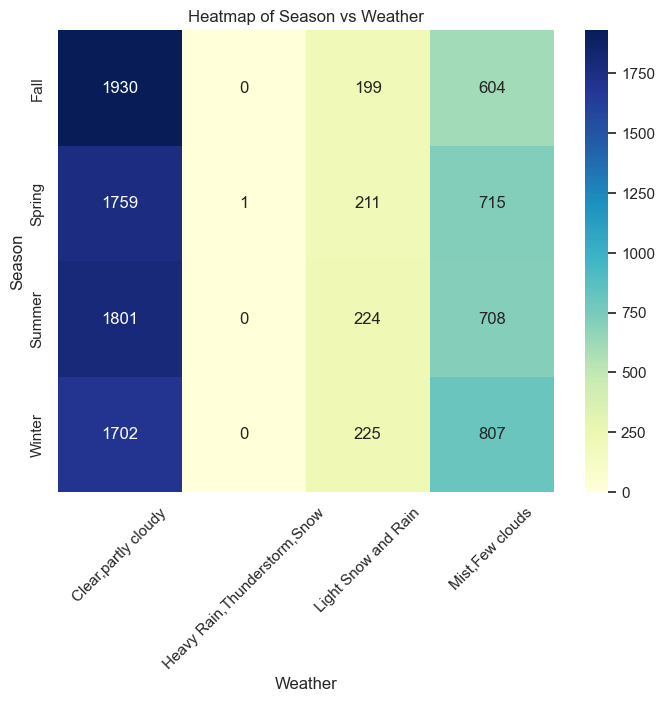
* **Null Hypothesis (H₀)**: There is no linear relationship between Registered User Count and Overall User Count.
* **Alternative Hypothesis (H₁)**: There is a significant linear relationship between Registered User Count and Overall User Count.

The Pearson correlation coefficient is r=0.9709r = 0.9709, indicating a very strong positive linear relationship between the Registered User Count and the Overall User Count. This suggests that as the number of Registered Users increases, the Overall Count also tends to increase.

The p-value is 0, confirming that this relationship is statistically significant. This means that the association between the Registered User Count and the Overall Count is highly unlikely to have occurred by chance. It implies that the count of Registered Users plays a dominant role in determining the Overall Count, with a very strong correlation between the two.

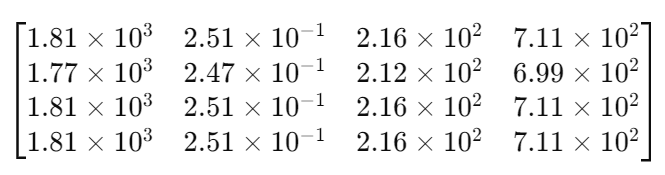
**Now, let's explore the relationships between the independent variables:**

**Dependence of weather on seasons**

 **Chi-Square Test: Relationship Between Seasons and Weather**

* Null Hypothesis (H₀): There is no significant relationship between seasons and weather conditions (they are independent).
* Alternative Hypothesis (H₁): There is a significant relationship between seasons and weather conditions (they are dependent).

Chi-Square Statistic: 49.16  
P-value: 1.55×10−71.55 x 10^{-7}  
Degrees of Freedom: 9  
Expected Frequencies:

**Conclusion:**

Since the p-value is extremely small (p≈0), we reject the null hypothesis. There is a significant relationship between 'season' and 'weather', meaning that the weather conditions and the seasons are not independent of each other. This suggests that the season may influence the types of weather conditions observed.

Business Recommendations:

**Key Business Recommendations:**

1. **Prioritize Registered Users to Boost Demand**:
   * **Insight**: The very strong correlation between registered user counts and overall bike rental counts (r = 0.97) suggests that registered users are a key driver of demand.
   * **Recommendation**: Yulu should focus on initiatives to increase the number of registered users, such as offering loyalty programs, referral incentives, and subscription-based models, as this will have the most significant impact on overall demand.
2. **Leverage Weather Data for Demand Forecasting**:
   * **Insight**: Weather conditions, particularly temperature and humidity, have a significant effect on bike rentals.
   * **Recommendation**: Yulu should integrate real-time weather data into its demand forecasting and fleet management. Adjust bike availability based on favorable weather conditions (e.g., warm temperatures and moderate humidity), and offer targeted promotions on days with high demand potential due to weather.
3. **Optimize Operations Based on Seasonality**:
   * **Insight**: Seasonal differences in bike rental demand are significant, with higher demand in spring and summer, and lower demand in fall and winter.
   * **Recommendation**: Yulu should optimize fleet size and operational capacity based on seasonality. During peak seasons (spring/summer), increase bike availability and consider running seasonal promotions. In off-peak months (fall/winter), adjust pricing strategies and reduce fleet size as needed.
4. **Encourage Casual Users with Weather-Specific Promotions**:
   * **Insight**: Casual users show a moderate correlation with overall rentals, and weather conditions strongly influence demand.
   * **Recommendation**: Yulu should create targeted campaigns to attract casual users, especially on warmer days or during favourable weather conditions, through discounts, special offers, or event-based promotions. This would help maximize casual rentals when demand spikes.
5. **Tailor Marketing Based on Season-Weather Insights**:
   * **Insight**: Significant relationships exist between seasons and weather conditions, which affect bike rental patterns.
   * **Recommendation**: Yulu should tailor its marketing campaigns based on both season and weather. For example, offer incentives or discounts during periods of warm weather in summer or cold, yet mild weather in the off-season, to encourage rentals despite seasonal changes.