Buisness_Case : Yulu[1]

Table of Contents

Link to Colab - **Buisness_Case_Yulu.ipynb**

[1] Buisness_Case_Yulu.ipynb

1. Problem Statement

Yulu, India's leading micro-mobility service provider, has experienced a decline in revenues and seeks to identify the factors influencing the demand for its shared electric cycles. The goal is to analyze various factors such as weather, season, and working days to understand their impact on rental counts and provide actionable insights to improve demand prediction and business performance.

Objectives

1. **Identify significant predictors:** Determine which factors significantly affect the demand for shared electric cycles.
2. **Quantify impact:** Assess how well these variables describe and predict electric cycle rentals.
3. **Hypothesis testing:** Use statistical methods to evaluate:
    - The effect of working days on the number of rentals.
    - Variations in rentals across seasons and weather conditions.
    - The relationship between weather and season.

2. Exploratory Data Analysis

Dataset Overview:

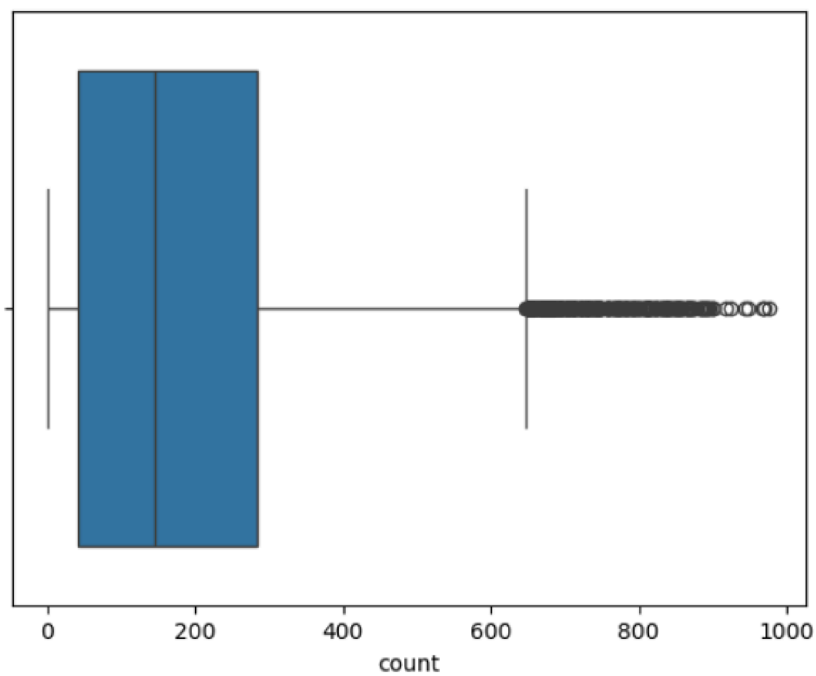
- The dataset contains **10,886 rows** and **12 columns**.
- No missing values are present in the dataset.

Data Types:

- Includes both categorical and numerical variables. Key columns:
  - **Categorical**: `season`, `holiday`, `workingday`, `weather`.
  - **Numerical**: `temp`, `atemp`, `humidity`, `windspeed`, `casual`, `registered`, `count`.

Outlier Analysis:

- Outliers were detected in the `count` column using a boxplot visualization. These may influence analysis and model predictions.

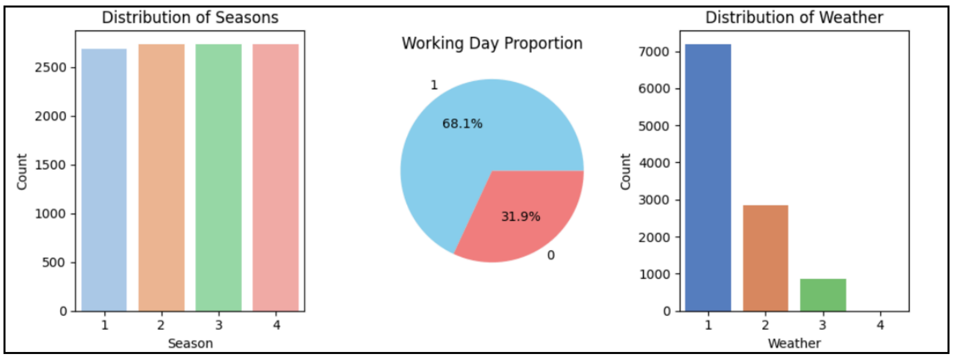


3. **<u>Visual Analysis</u>**

3.a) Univariate Analysis

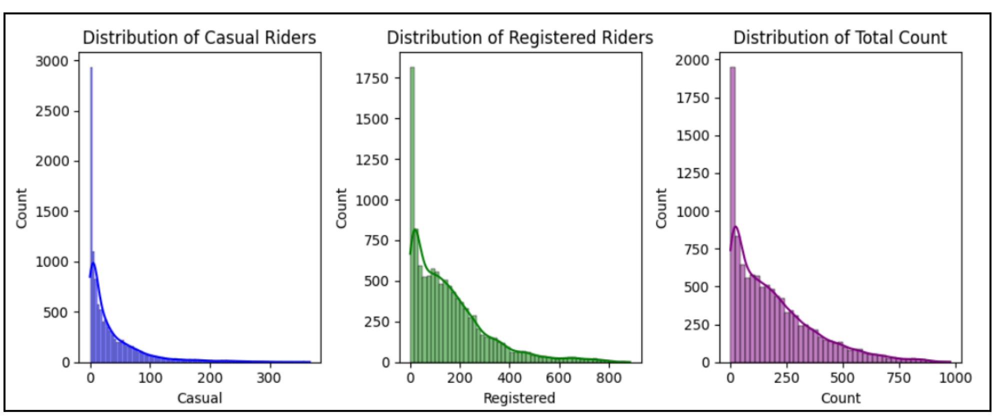
Categorical Variables:

- **Season**: Majority counts observed in certain seasons; visualized via count plot.
- **Working Day**: A nearly balanced split between working and non-working days; shown using a pie chart.
- **Weather**: Clear weather dominates; represented with a count plot.



Continuous Variables:

- **Casual Riders**: Right-skewed distribution, indicating a majority of low counts.
- **Registered Riders**: Higher counts compared to casual riders, with a more balanced spread.
- **Total Count**: Slightly skewed, with outliers noted in the higher range.



3.b) Bivariate Analysis

Count vs. Season:

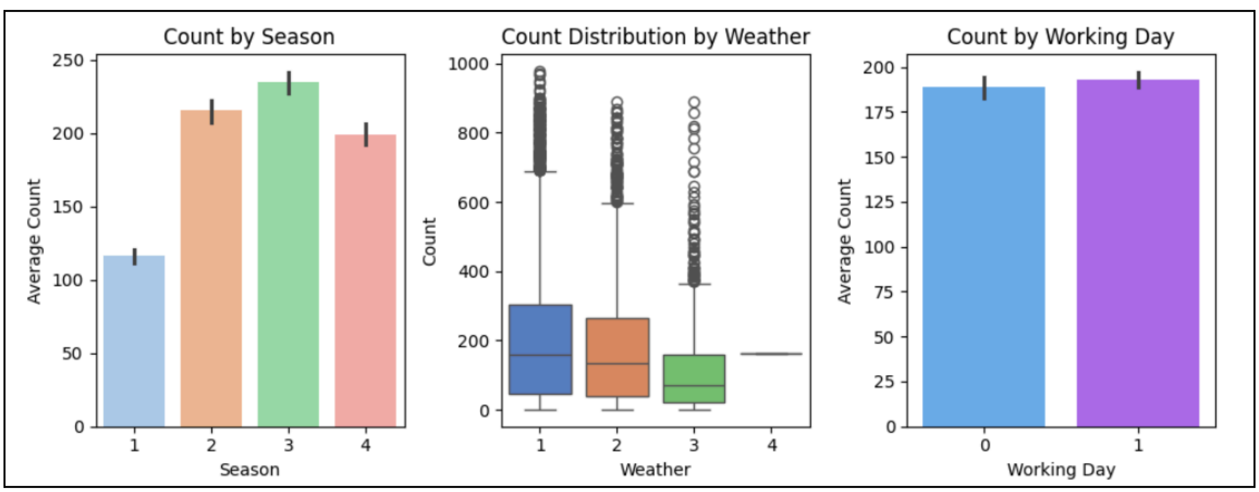
- Count varies across seasons, with some seasons contributing significantly more to total usage.

Count vs. Weather:

- Clear weather conditions correspond to higher counts; adverse weather shows reduced demand.

Count vs. Working Day:

- Counts are generally higher on working days compared to non-working days, reflecting commuter trends.



4. **Hypothesis Testing**

4.a) 2 Pair T-Test

4.a.1) **Working Day has effect on number of electric cycles rented**

1. Objective:

To test if the type of day (working vs. non-working) affects the number of electric cycles rented.

2. Hypotheses:

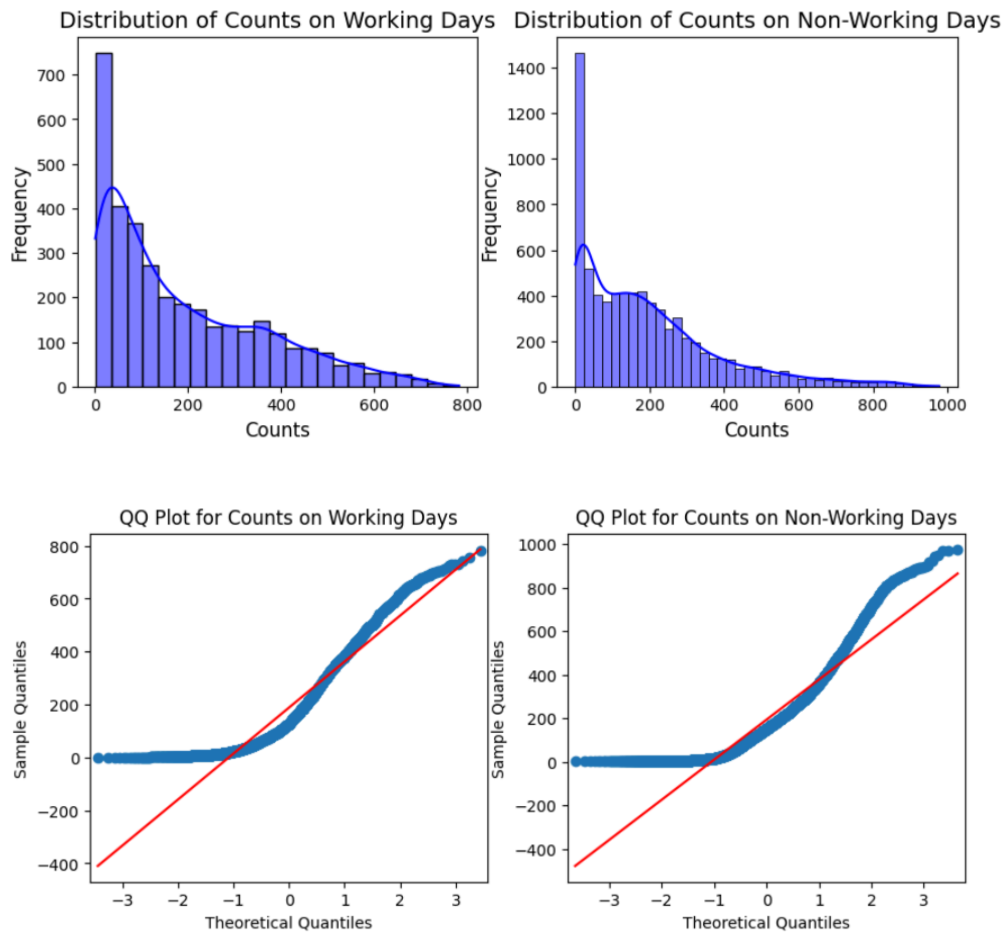
- **Null Hypothesis (H_0)**: The mean rental counts on working days and non-working days are equal.

- **Alternative Hypothesis (H_1)**: The mean rental counts on working days and non-working days are different.

Assumptions and Tests:

1. Normality Check:

- **Hypotheses**:
  - **Null Hypothesis (H_0)**: The rental counts data for both working and non-working days are normally distributed.
  - **Alternative Hypothesis (H_1)**: The rental counts data for both working and non-working days are not normally distributed.
- **Tests Conducted**:
  - **Visual Analysis**:
    - Histograms with KDE and Q-Q plots show right-skewed distributions for both samples.



- **Shapiro-Wilk Test Results**:
  - Working days: $p=4.47\times10^{(-45)}$ (Reject H_0)
  - Non-working days: $p=2.25\times10^{(-61)}$ (Reject H_0)
- **Inference**:
  Data for both samples is not normally distributed.

2. Variance Check:

- **Hypotheses**:
  - **Null Hypothesis (H_0)**: The variances of rental counts for working and non-working days are equal.
  - **Alternative Hypothesis (H_1)**: The variances of rental counts for working and non-working days are not equal.

- **Test Conducted**:
  - Levene's Test:
    $p=0.944>0.05$, Fail to reject H_0. Variances are equal.

3. Mean Comparison (2-Sample T-Test):

- **Hypotheses**:
  - **Null Hypothesis (H_0)**: The means of rental counts for working and non-working days are equal.
  - **Alternative Hypothesis (H_1)**: The means of rental counts for working and non-working days are different.
- **Test Conducted**:
  - T-Test:
    $p=0.226>0.05$, Fail to reject H_0.
- **Inference**:
  The analysis shows no significant difference in rental counts between working and non-working days.

4.b) 1 Way ANNOVA Test

4.b.1) Seasons have an effect on number of cycles rented

1. Objective:

To test if the season has an effect on the number of electric cycles rented.

2. Hypotheses:

- **Null Hypothesis (H_0)**: The mean rental counts across all four seasons are equal.
- **Alternative Hypothesis (H_1)**: At least one season has a different mean rental count.

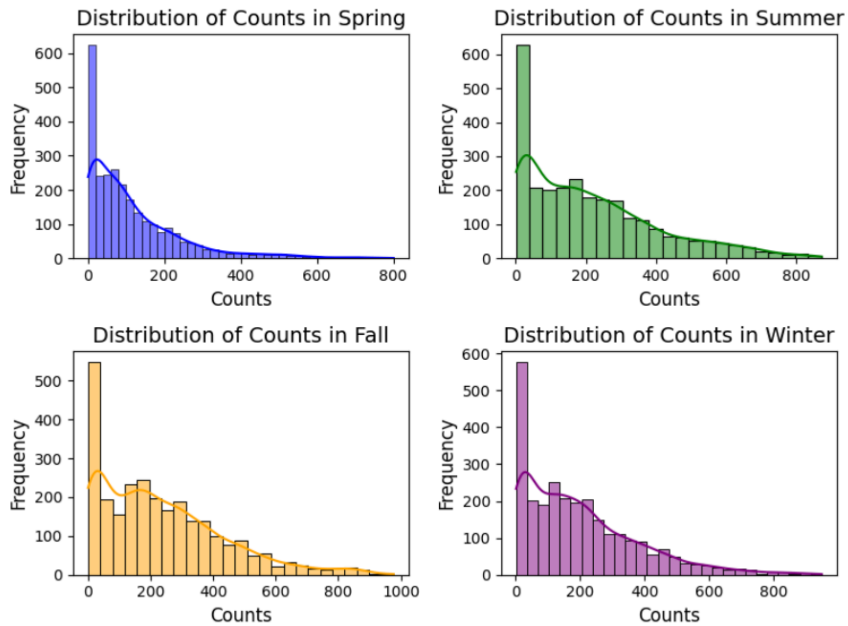
Assumptions and Tests:

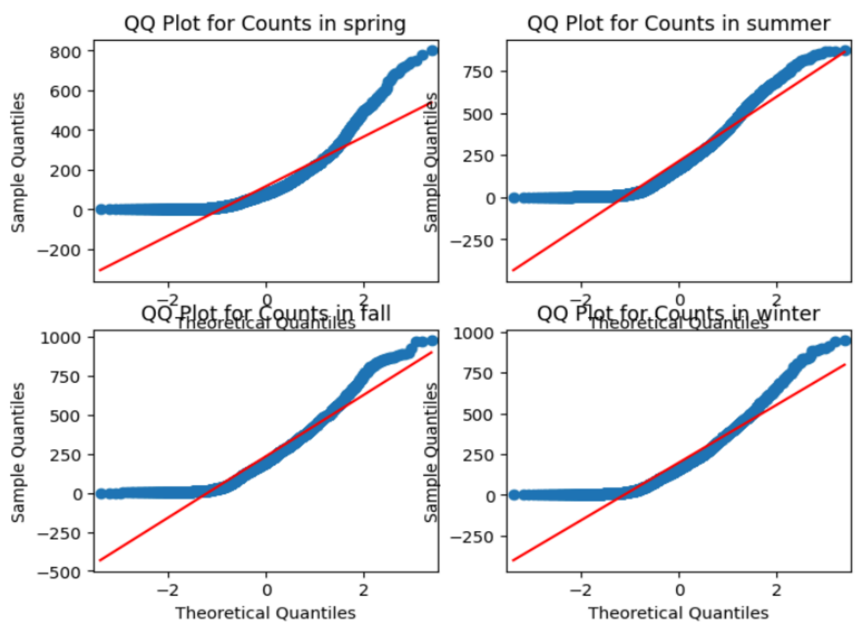
1. Normality Check:

- **Hypotheses**:
  - **H_0**: The rental counts for each season are normally distributed.
  - **H_1**: The rental counts for each season are not normally distributed.

- **Tests Conducted**:

  - **Visual Analysis**:
    - Histograms with KDE and Q-Q plots show right-skewed distributions for all seasons.





- **Shapiro-Wilk Test Results**:
  - Spring: p=8.75×10^(−49), Summer: p=6.04×10^(−39), Fall: p=1.04×10^(−36), Winter: p=1.13×10^(−39)
  - All p-values are <0.05, indicating non-normality.

2. Variance Check:

- **Hypotheses**:
  - **H_0**: Variances of rental counts across seasons are equal.
  - **H_1**: Variances of rental counts across seasons are not equal.
- **Test Conducted**:
  - **Levene's Test**:
    $p=1.01\times10^{(-118)}$, $p<0.05$, Reject H_0. Variances differ across seasons.

3. Mean Comparison (1-Way ANOVA):

- **Hypotheses**:
  - **H_0**: The mean rental counts across all seasons are equal.
  - **H_1**: At least one mean is different.
- **Test Conducted**:
  - **ANOVA Test**:
    $p=6.16\times10^{(-149)}$, $p<0.05$, Reject H_0. Rental counts significantly depend on the season.

4. Non-Parametric Test (Kruskal-Wallis):

- **Reason**: Normality and variance assumptions are violated.
- **Hypotheses**:
  - **H_0**: The median rental counts across all seasons are equal.
  - **H_1**: At least one median is different.
- **Test Conducted**:
  - **Kruskal-Wallis Test**:
    $p=2.48\times10(-151)$, $p<0.05$, Reject H_0. Season significantly affects rental counts.

Conclusion:

Both parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests confirm that the number of cycles rented depends significantly on the season.

4.b.2) Weathers have an effect on number of cycles rented

1. Objective:

To determine if weather conditions significantly impact the number of cycles rented.

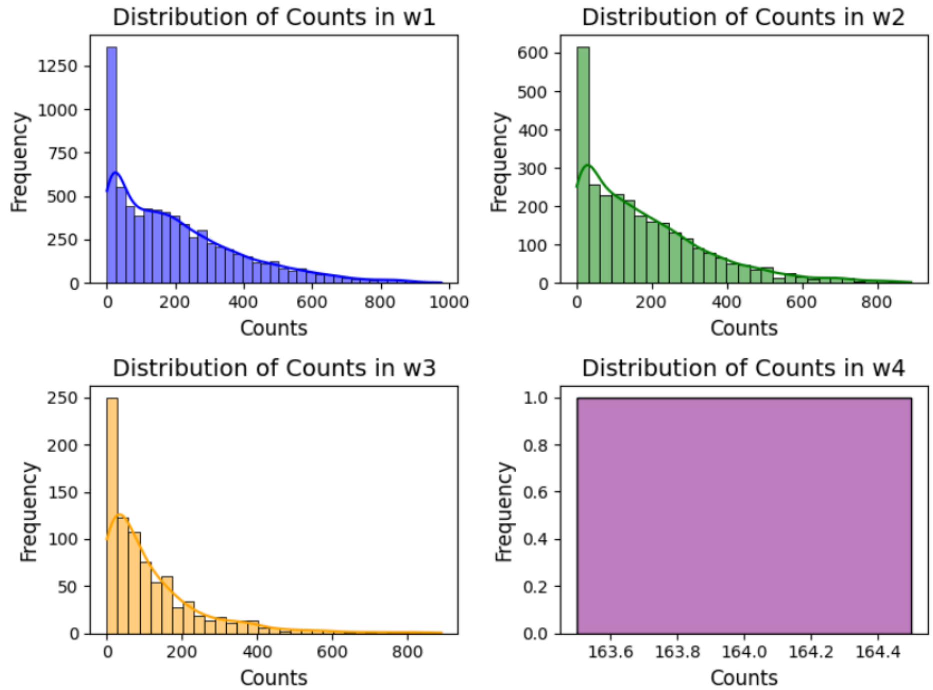
2. Hypotheses:

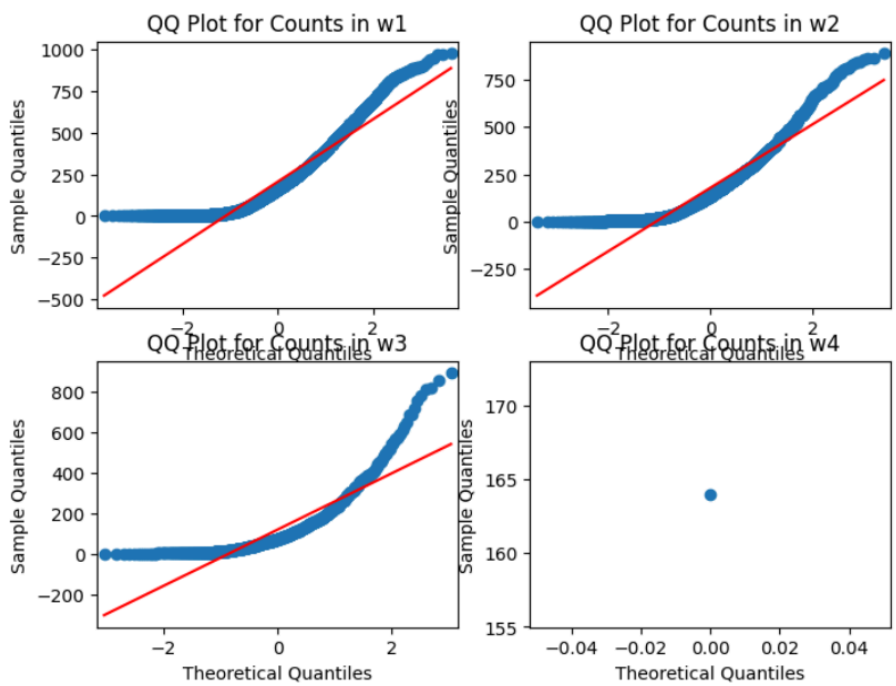
- **Null Hypothesis (H_0)**: The mean rental counts across all weather categories are equal.
- **Alternative Hypothesis (H_1)**: At least one weather category has a different mean rental count.

Assumptions and Tests:

1. Normality Check:

- **Hypotheses**:
  - **H_0**: Rental counts for each weather category are normally distributed.
  - **H_1**: Rental counts for at least one weather category are not normally distributed.
- **Tests Conducted**:
  - **Visual Analysis**:
    - Histograms and Q-Q plots for weather categories w1, w2, and w3 show right-skewed distributions.
    - Normality could not be tested for w4 due to insufficient data (single observation).





- **Shapiro-Wilk Test Results**:
  - p-values for w1, w2, and w3 are <0.05, indicating non-normality.

2. Variance Check:

- **Hypotheses**:
  - **H_0**: Variances of rental counts across weather categories are equal.
  - **H_1**: Variances of rental counts across weather categories are not equal.
- **Tests Conducted**:
  - **Levene's Test**:
    $p=3.50\times10^{(-35)}<0.05$, indicating unequal variances.

3. Mean Comparison (1-Way ANOVA):

- **Hypotheses**:
  - **H_0**: The mean rental counts across all weather categories are equal.
  - **H_1**: At least one mean is different.
- **Tests Conducted**:
  - **ANOVA Test**:
    $p=5.48\times10^{(-42)}$, $p<0.05$, Reject H_0. Weather significantly impacts rental counts.

4. Non-Parametric Test (Kruskal-Wallis):

- **Reason**: Normality and variance assumptions are violated.
- **Hypotheses**:
  - **H_0**: The median rental counts across all weather categories are equal.
  - **H_1**: At least one median is different.
- **Test Conducted**:
  - **Kruskal-Wallis Test**:
    $p=3.50\times10^{(-44)}$, $p<0.05$, Reject H_0. Weather significantly impacts rental counts.

Conclusion:

Both parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests confirm that the number of cycles rented is significantly influenced by weather conditions.

4.c) Chi-Square Test

4.c.1)**Weather is dependent on season (check between 2 predictor variable)**

1. Objective:

To determine if there is a significant relationship between the categorical variables *season* and *weather*.

2. Hypotheses:

- **Null Hypothesis** (H_0): There is no relationship between season and weather (they are independent).
- **Alternative Hypothesis** (H_1): There is a relationship between season and weather (they are dependent).

Methodology:

1. **Data Representation**:
- **Cross-tabulation Table**: A contingency table showing the frequency distribution of weather across seasons:

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

2. **Statistical Test**:

- **Chi-Square Test** for Independence:
  - Statistic: $\chi 2$ = 49.16

- Degrees of Freedom (dof): 9
- Expected Frequencies: Calculated to compare observed frequencies.

Results:

- p-value: $1.55 \times 10^{(-71)}$
- $p<0.05$: Reject the null hypothesis.

Conclusion:

There is a **significant relationship** between season and weather. This suggests that the distribution of weather conditions is dependent on the season.

5. Business Insights

1. **Impact of Season on Bike Rentals**:

  - **Business Insight**: The season significantly impacts the number of bike rentals, with spring and summer having the highest bike rental counts, while winter sees a considerable drop.
    - **Actionable Insight**: Yulu could focus marketing efforts on spring and summer months, offering promotions and targeted advertising campaigns to boost rentals during peak seasons. During winter, Yulu might consider offering winter-specific services or discounts to maintain user engagement.

2. **Weather Conditions and Rentals**:

  - **Business Insight**: Different weather conditions (clear skies, light rain, snow, etc.) influence the number of rentals. Clear weather generally results in higher rentals, while heavy rain and snow lead to fewer bike rentals.
    - Actionable Insight: Yulu can implement dynamic pricing based on weather forecasts, offering discounts on days with unfavorable weather to encourage rentals. Additionally, introducing weather-proof bikes or enhanced services (like covered bike stations) could help mitigate weather-related barriers to rentals.

3. **Optimizing Fleet Management for Seasonal and Weather Variability**:

  - **Business Insight**: As rentals peak in spring and summer and dip in winter, Yulu can optimize fleet distribution.
    - **Actionable Insight**: During peak seasons, increase the availability of bikes in high-demand areas and, conversely, reduce fleet sizes in low-demand areas during the off-seasons (especially winter). This can help optimize operational costs and improve service availability.

4. **Improved User Experience Based on Seasonal and Weather Trends**:

  - **Business Insight**: Seasonal and weather changes significantly affect user behavior and rental patterns.

    - **Actionable Insight**: Yulu could introduce a seasonal subscription model where users can sign up for discounts or guarantees of bike availability during specific seasons. They can also introduce weather notifications, alerting users about favorable conditions for bike rentals, to enhance user engagement.

5. **Strategic Partnerships for Off-Season Engagement**:

- **Business Insight**: The drastic decrease in rentals during the winter months provides an opportunity to explore alternative business models.
    - **Actionable Insight**: Yulu could partner with tourism agencies, shopping malls, or universities to offer bikes as part of off-season promotions (e.g., for winter shopping sprees or city tours). Additionally, offering indoor bike stations or exploring alternative uses for bikes in the off-season could help sustain engagement.

6. **Targeted Advertising Based on Weather Conditions**:

- **Business Insight**: Weather is a key predictor of bike rentals, with adverse weather conditions reducing rental numbers.
    - **Actionable Insight**: Yulu could leverage weather data for targeted advertising or app notifications, encouraging users to rent bikes on sunny days or offer promotions on days when the weather is less than ideal, thereby boosting rentals even during less favorable weather.

By leveraging these insights, Yulu can optimize its operations, enhance customer experience, and boost rentals year-round.

6. Recommendations

- **Seasonal Promotions**: Focus on marketing campaigns and discounts during peak seasons (spring and summer) to maximize rentals.
- **Weather-Based Pricing**: Implement dynamic pricing based on weather conditions, offering discounts during rainy or snowy days to boost rentals.
- **Fleet Optimization**: Adjust bike availability based on seasonal demand, increasing fleet size during peak seasons and reducing it during the off-season.
- **User Engagement**: Introduce seasonal subscriptions or loyalty programs with benefits tied to weather or season, enhancing user retention.
- **Partnerships for Off-Season**: Collaborate with tourism and retail businesses to offer bikes during low-rent seasons, maintaining engagement throughout the year.
- **Targeted Notifications**: Use weather forecasts for app notifications encouraging rentals during good weather or offering incentives during adverse weather conditions.