Business Case: Yulu - Hypothesis Testing

Introduction

In response to Yulu's recent decline in revenues, this analysis aims to uncover the key factors influencing the demand for shared electric cycles in the Indian market. The primary goal is to provide actionable insights that can inform Yulu's strategic decisions and help optimize their operations. To achieve this, a systematic approach was employed, combining data exploration, statistical analysis, and clustering techniques to identify patterns in user demand.

The report begins with a thorough data examination to understand the distribution of key variables and their relationships with the target variable (total demand). This is followed by a detailed analysis of seasonal trends, weather influences, and user segmentation. Using these insights, the report provides tailored recommendations for Yulu to improve customer engagement, optimize resource allocation, and enhance profitability.

The subsequent sections outline the findings, highlight the most significant factors affecting demand, and present strategies for Yulu to leverage these insights effectively.

Analysis Approach

To comprehensively understand the factors affecting demand for Yulu's shared electric cycles, a structured and data-driven analysis approach was employed. The following techniques and methodologies were applied:

Exploratory Data Analysis (EDA)

Univariate Analysis: Conducted to examine the distribution of individual variables, such as temperature, humidity, and demand (count). This provided initial insights into central tendencies, variability, and potential outliers.
Bivariate Analysis: Explored relationships between variables (e.g., temperature vs. demand, weather vs. demand) to identify key factors influencing the target variable.

Hypothesis Testing

2-Sample T-Test: Assessed whether there is a significant difference in demand on working days versus non-working days.
ANOVA (Analysis of Variance): Examined the effect of categorical variables such as weather and season on the number of cycles rented, determining whether the mean demand differs significantly across groups.
Chi-Square Test: Evaluated the independence between categorical variables, specifically testing whether weather conditions are dependent on the season.

Correlation Analysis

Identified significant relationships between numerical features (e.g., temperature, humidity) and the target variable (demand). This helped determine which features are most strongly associated with changes in demand.

Clustering Analysis

Applied clustering techniques (e.g., K-means) to segment the demand data based on temperature and usage patterns. This enabled the identification of distinct customer segments with different preferences and behaviors.

Time Series Analysis

Analyzed temporal patterns in the data, identifying trends and seasonality in demand over time. This provided insights into growth potential and cyclical fluctuations.

1.Importing Data and Understanding:

Code:

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

yulu_data = pd.read_csv('Downloads/yulu.csv')
yulu_data.head()

Output:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	13	16
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	32	40
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	27	32
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3	10	13
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0	1	1

yulu_data.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
```

yulu_data.isnull().sum()

```
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
```

Insights:

No Null values Data is clean and ready to go.

yulu_data.describe()

[7]:

	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
count	10886.000000	10886.000000	10886.000000	10886.000000	10886.00000	10886.000000	10886.000000	10886.000000	10886.000000	10886.000000	10886.000000
mean	2.506614	0.028569	0.680875	1.418427	20.23086	23.655084	61.886460	12.799395	36.021955	155.552177	191.574132
std	1.116174	0.166599	0.466159	0.633839	7.79159	8.474601	19.245033	8.164537	49.960477	151.039033	181.144454
min	1.000000	0.000000	0.000000	1.000000	0.82000	0.760000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	0.000000	1.000000	13.94000	16.665000	47.000000	7.001500	4.000000	36.000000	42.000000
50%	3.000000	0.000000	1.000000	1.000000	20.50000	24.240000	62.000000	12.998000	17.000000	118.000000	145.000000
75%	4.000000	0.000000	1.000000	2.000000	26.24000	31.060000	77.000000	16.997900	49.000000	222.000000	284.000000
max	4.000000	1.000000	1.000000	4.000000	41.00000	45.455000	100.000000	56.996900	367.000000	886.000000	977.000000

Insights:

Seasonal Influence: Demand likely fluctuates across seasons (season ranges from 1 to 4), with lower usage expected during extreme weather conditions (e.g., high temperatures or monsoon rains).

Impact of Holidays and Weekdays: Only 2.86% of the data points are holidays (holiday mean = 0.028), while 68% are working days (working day mean = 0.68).
Higher demand is expected on holidays and weekends, suggesting leisure-driven usage, while working days may see higher commuter usage.

Weather Effect: The weather variable ranges from 1 (clear) to 4 (heavy rain/snow), with an average of 1.42, indicating most observations are during clear or moderate weather.
Poor weather (rain/snow) slightly reduces demand (correlation with count is -0.13).

Temperature Impact: Average temperature (temp) is 20.23°C, and feels-like temperature (atemp) is 23.66°C.
There is a moderate positive correlation (0.39) between temperature and demand, suggesting higher demand during moderate temperatures, but extremes may deter users.

Humidity and Wind speed: Average humidity is 61.89% with high variability (max = 100%), while average wind speed is 12.8, ranging up to 56.99.
High humidity negatively correlates with demand (-0.32), and strong winds may reduce comfort, impacting usage.

User Segment Analysis: Registered users dominate the total demand with a mean of 155.55, compared to 36.02 for casual users.
This highlights potential to increase casual user engagement.

Demand Variability:
The demand (count) has a mean of 191.57 and a high standard deviation of 181.14, indicating significant fluctuations (min = 1, max = 977).
This suggests strong influence from external factors like weather conditions and time of day.

Recommendation:

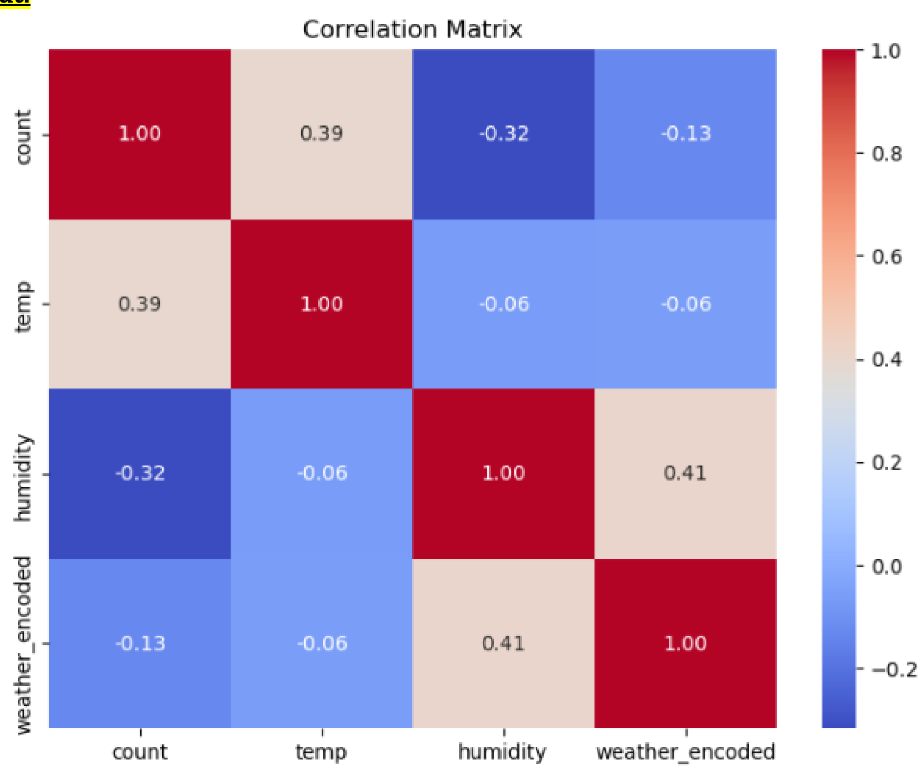
Seasonal Strategies: Adjust fleet availability based on seasonal demand trends, focusing on peak seasons.
Target Promotions: Offer discounts and promotions on holidays and weekends to attract leisure users.
Weather-Based Adjustments: Monitor forecasts and adapt pricing or service availability during adverse weather.
Temperature Campaigns: Promote usage during moderate weather days and explore dynamic pricing during temperature extremes.
Enhance User Conversion: Develop campaigns to convert casual users into registered users for sustained demand.
Monitor Demand Fluctuations: Use real-time data to adjust operational strategies in response to sudden changes in demand patterns.

2. Correlation analysis between demand (count) and features such as temperature (temp), humidity (humidity), and weather (weather).

Code:

```
df['weather_encoded'] = df['weather'].astype('category').cat.codes

# Select features for correlation analysis
correlation_features = ['count', 'temp', 'humidity', 'weather_encoded']

# Calculate the correlation matrix
correlation_matrix = df[correlation_features].corr()

# Display the correlation matrix
print(correlation_matrix)

# Visualize the correlation matrix using a heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```

Output:




```
                    count      temp  humidity  weather_encoded
count            1.000000  0.394454 -0.317371        -0.128655
temp             0.394454  1.000000 -0.064949        -0.055035
humidity        -0.317371 -0.064949  1.000000         0.406244
weather_encoded -0.128655 -0.055035  0.406244         1.000000
```

Insights:

Temperature Boosts Demand: There's a moderate positive correlation (0.39) between temperature and demand. Users prefer riding when the weather is warm and comfortable.

Humidity Lowers Demand: A moderate negative correlation (-0.32) indicates that higher humidity reduces demand, likely due to discomfort while cycling in humid conditions.

Weather Impacts Slightly: Poorer weather conditions (rain/snow) show a weak negative effect on demand (-0.13), suggesting that adverse weather slightly deters users.

Weather and Humidity Linked: There's a moderate positive correlation (0.41) between weather conditions and humidity, implying rainy or snowy weather often coincides with higher humidity.

Recommendations:

Seasonal Strategy: Since demand increases with temperature and decreases with humidity, Yulu should focus on promoting its services during moderate weather conditions. They can also consider increasing the availability of cycles during warmer, less humid periods.

Pricing Adjustments: Implement dynamic pricing strategies based on weather conditions. For instance, offering discounts during rainy or humid weather might help maintain demand.

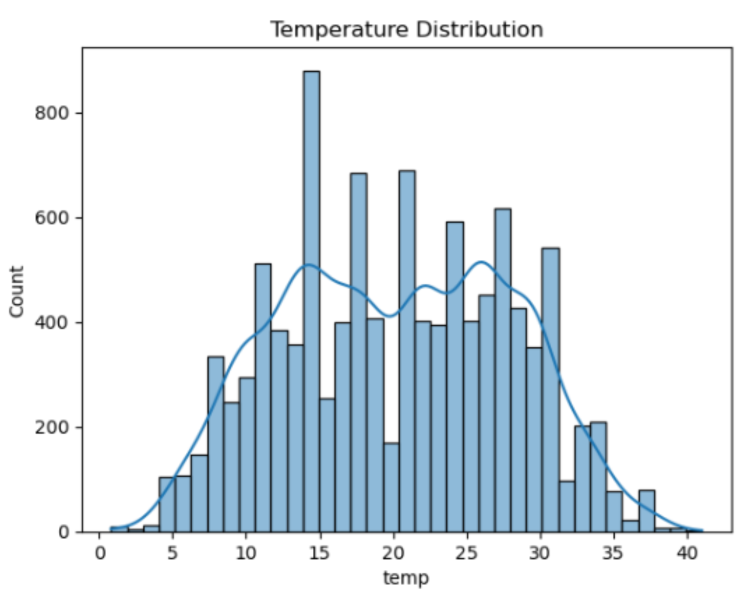
User Engagement: Communicate with users about favorable weather forecasts and encourage usage during such times, potentially boosting demand during peak weather conditions.

3.Univariate Analysis.
Plotting distribution of continuous variables (temp, atemp, humidity, etc.) using histograms.

Code:

```
sns.histplot(yulu_data['temp'], kde=True)
plt.title('Temperature Distribution')
plt.show()
```

Output:



Insights:

Bimodal Distribution:
The temperature data appears to have a bimodal distribution, with two peaks observed around **15°C** and **25°C**.
Interpretation: These peaks could correspond to preferred temperature ranges for cycling, with higher demand during moderate weather.

Moderate Temperature Prevalence: The majority of observations fall between **10°C and 30°C**, indicating that Yulu's service is primarily used during moderate temperatures, which are comfortable for cycling.

Lower Usage in Extreme Temperatures: The distribution tapers off at the extremes (below **5°C** and above **35°C**), suggesting lower demand during very cold or very hot weather.

Actionable Insight: Yulu might see reduced demand during extreme weather, and should adjust fleet availability or pricing accordingly.

Potential for Targeted Marketing: Given the higher frequency of observations around **15°C** to **25°C**, this is an optimal temperature range for marketing campaigns or promotions, as users are more likely to engage during these comfortable weather conditions.

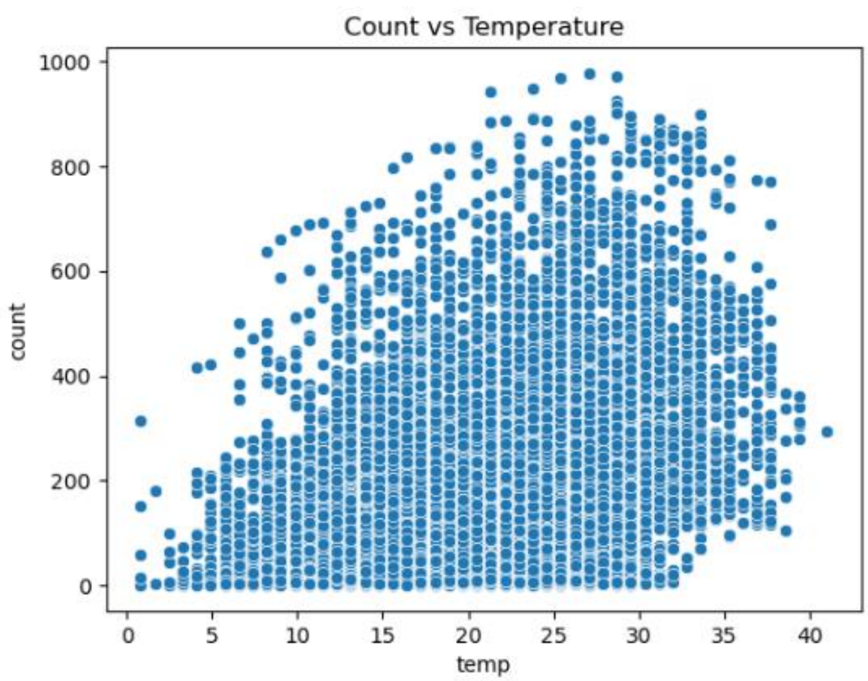
Overall, the distribution suggests that Yulu's shared electric cycle usage is heavily influenced by moderate, comfortable temperatures, with demand likely decreasing during extreme hot or cold periods.

4.Bivariate Analysis (Exploring Relationships)

Exploring the relationship between dependent (count) and independent variables:

Code:

```
sns.scatterplot(x='temp', y='count', data=yulu_data)
plt.title('Count vs Temperature')
plt.show()
```



Insights:

Positive Relationship: The plot shows a general upward trend, indicating a **positive correlation** between temperature and demand (count). As the temperature increases, the demand for electric cycles also tends to rise.

Peak Usage at Moderate Temperatures: The demand appears to peak between **20°C and 30°C**, with several observations reaching counts close to **1000** in this temperature range.

Interpretation: This suggests that users prefer cycling during moderate temperatures, as it is likely more comfortable for outdoor activities.

Decline in Demand at Extreme Temperatures: Beyond **30°C**, there is a noticeable decline in demand. Similarly, at very low temperatures (<10°C), the demand is lower.

Actionable Insight: Extreme weather (too hot or too cold) negatively impacts user comfort, reducing the likelihood of cycle rentals.

High Variability in Demand: There is significant variability in demand across all temperature ranges, especially between **15°C and 30°C**, where counts range from low to high. This suggests that while temperature is a key factor, other variables (e.g., humidity, day type, or weather conditions) also influence demand.

Recommendations:

Optimize Fleet During Peak Temperatures: Increase fleet availability and marketing efforts during moderate temperatures (20°C to 30°C), where demand is highest.

Dynamic Pricing During Extreme Weather: Implement pricing adjustments during extremely hot (>30°C) or cold (<10°C) weather to balance demand and incentivize users.

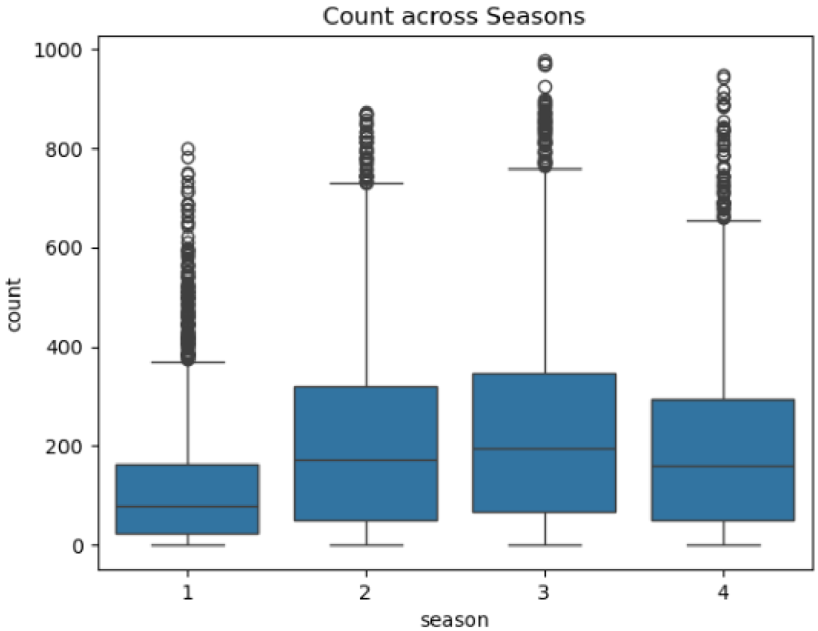
Further Analysis: Investigate additional factors (e.g., humidity, wind speed, holidays) that might explain the variability in demand, especially at moderate temperatures.

5.Boxplot for categorical variables:

Code:

```
sns.boxplot(x='season', y='count', data=yulu_data)
plt.title('Count across Seasons')
plt.show()
```

Output:



Insights:

Seasonality: There appears to be some seasonality in the demand for Yulu's services.
The demand for cycles seems to be higher in seasons 2 and 3 compared to seasons 1 and 4.

Outliers: There are numerous outliers in the data, especially in seasons 2 and 3.
These outliers could be due to various reasons such as special events, promotions, or anomalies in the data collection process.

Distribution: The distribution of demand across seasons appears to be right-skewed.
This indicates that there are more days with lower demand compared to days with higher demand.

Recommendations:

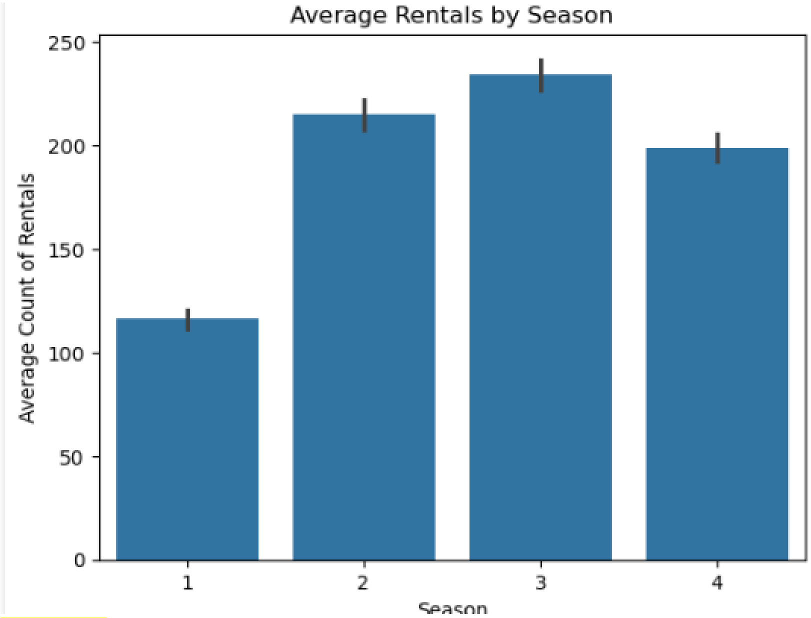
Yulu could leverage the seasonal variation in demand by implementing targeted marketing campaigns and promotions during periods of lower demand.
They could also adjust their fleet size and pricing strategies to align with seasonal fluctuations.

6. Season and Count of Rentals:
Objective: To analyze the relationship between different seasons and the rental demand.
Approach: A barplot to compare rental counts across the four seasons (spring, summer, fall, winter).

Code:

```
sns.barplot(data=yulu_data, x='season', y='count', estimator='mean')
plt.title('Average Rentals by Season')
plt.xlabel('Season')
plt.ylabel('Average Count of Rentals')
```

Output:



Insights:

Seasonality:
The bar chart clearly illustrates the seasonal variation in demand for Yulu's services.
The demand for cycles is significantly higher in season 3 compared to other seasons.
This could be attributed to factors like weather conditions, holidays, or specific events.

Average Rentals:
The average number of rentals per season is highest in season 3, followed by season 2, season 4, and season 1.
This indicates that season 3 is the peak season for Yulu's operations.

Recommendations:

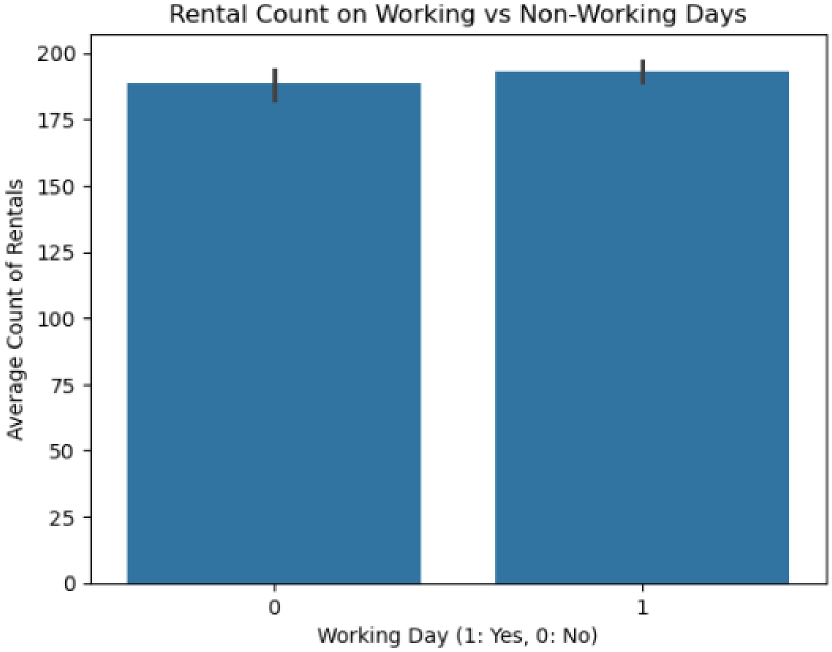
Seasonal Adjustments:
Yulu can leverage the seasonal variation in demand by implementing targeted marketing campaigns and promotions during periods of lower demand.
They could also adjust their fleet size and pricing strategies to align with seasonal fluctuations.

Marketing and Promotions:
Focusing on marketing and promotional efforts during season 3 can help capitalize on the peak demand and attract new customers.

Capacity Planning:
Yulu should ensure that they have sufficient capacity to meet the increased demand during peak seasons.
This includes having enough cycles available and a well-trained workforce to handle the influx of customers.

7. Rental Count on Working vs Non-Working Days
Code:
```
sns.barplot(data=yulu_data, x='workingday', y='count', estimator='mean')
plt.title('Rental Count on Working vs Non-Working Days')
plt.xlabel('Working Day (1: Yes, 0: No)')
plt.ylabel('Average Count of Rentals')
```

Output:



Insights:

Working Days vs. Non-Working Days:
The bar chart shows that the average number of rentals is nearly identical on working days and non-working days. This suggests that Yulu's demand is not significantly impacted by whether it's a working day or a non-working day.

Consistent Demand:
The consistent demand across working and non-working days indicates that Yulu's service is in demand throughout the week.
This is a positive finding as it suggests a stable and reliable revenue stream.

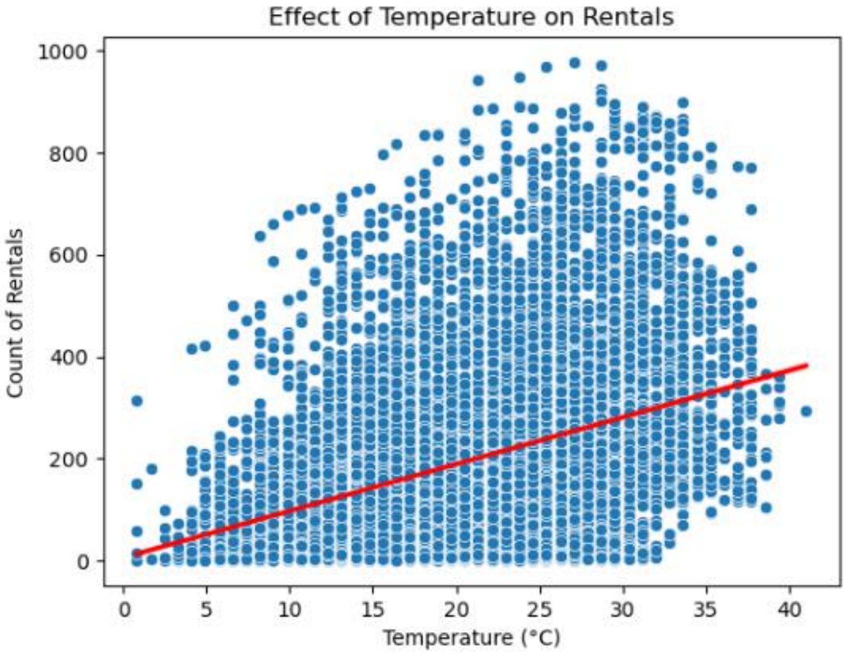
Marketing and Promotions:
Yulu can consider implementing targeted marketing campaigns and promotions on specific days of the week to further boost demand.
For example, they could offer discounts or special deals on weekdays to attract more customers during off-peak hours.

8. Temperature vs. Count of Rentals:
Objective: To understand how temperature affects the number of electric cycle rentals.
Approach: A scatterplot to see the correlation between temperature and the count of rentals. A regression line can also be added to observe trends.

Code:

```
sns.scatterplot(data=yulu_data, x='temp', y='count')
sns.regplot(data=yulu_data, x='temp', y='count', scatter=False, color='red')  # Add regression line
plt.title('Effect of Temperature on Rentals')
plt.xlabel('Temperature (°C)')
plt.ylabel('Count of Rentals')
```

Output:



Insights:

Positive Correlation:
The scatter plot reveals a clear positive correlation between temperature and the number of rentals.
This indicates that as the temperature increases, the demand for Yulu's services also tends to increase.

Linear Trend:
The red line overlaid on the scatter plot represents a linear regression line, which captures the overall trend in the data.
The upward slope of the line confirms the positive relationship between temperature and rentals.

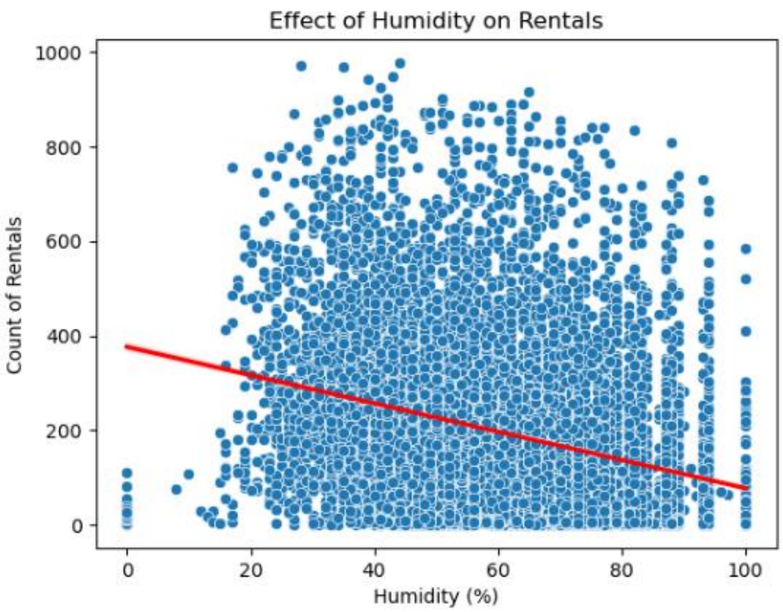
Recommendations:

Temperature-Based Strategies:
Yulu can leverage temperature data to make informed decisions about their operations.
For example, they could increase the availability of cycles in areas with higher temperatures or during periods of warmer weather.

Marketing and Promotions:
Yulu could implement targeted marketing campaigns and promotions during periods of favorable weather conditions to attract more customers.

9. Humidity vs. Count of Rentals:
Objective: To see how humidity levels impact electric cycle demand.
Approach: A scatterplot with a regression line to observe any relationship between humidity and rental count.

Code:
```
sns.scatterplot(data=yulu_data, x='humidity', y='count')
sns.regplot(data=yulu_data, x='humidity', y='count', scatter=False, color='red')
plt.title('Effect of Humidity on Rentals')
plt.xlabel('Humidity (%)')
plt.ylabel('Count of Rentals')
```



Insights:
Negative Correlation:
The scatter plot reveals a clear negative correlation between humidity and the number of rentals.
This indicates that as the humidity increases, the demand for Yulu's services tends to decrease.

Linear Trend:
The red line overlaid on the scatter plot represents a linear regression line, which captures the overall trend in the data.
The downward slope of the line confirms the negative relationship between humidity and rentals.

Recommendation:
Humidity-Based Strategies:
Yulu can leverage humidity data to make informed decisions about their operations.
For example, they could increase the availability of cycles in areas with lower humidity or during periods of lower humidity.

Marketing and Promotions:
Yulu could implement targeted marketing campaigns and promotions during periods of lower humidity to attract more customers.

Hypothesis Testing:
Q. Test to determine if there is a significant difference between the number of bike rides on weekdays and weekends:

Formulate the Hypotheses-->>
Null Hypothesis (H0): There is no significant difference between the number of bike rides on weekdays and weekends.
Alternate Hypothesis (H1): There is a significant difference between the number of bike rides on weekdays and weekends.

Select an Appropriate Test
Since we are comparing the means of two independent groups (weekdays and weekends)**, a 2-sample independent t-test is appropriate for this analysis.**

Set a Significance Level
Significance Level (α)**: 5% (0.05)**

Calculating Test Statistics / p-value

Code:
```
from scipy import stats
t_stat, p_value = stats.ttest_ind(weekday_rides, weekend_rides, equal_var=False)
print(f"T-Statistic: {t_stat:.4f}, P-Value: {p_value:.4f}")
```

Output:

```
T-Statistic: 1.2363, P-Value: 0.2164
```

```
alpha = 0.05  # Significance level
if p_value <= alpha:
   print("Reject the null hypothesis: There is a significant difference in bike rides between weekdays and weekends.")
else:
   print("Fail to reject the null hypothesis: No significant difference in bike rides between weekdays and weekends.")
```

Output and Interpretation:

```
Fail to reject the null hypothesis: No significant difference in bike rides between weekdays and weekends.
```

Explanation:

The **null hypothesis (H0)** states that there is no significant difference in the number of bike rides between weekdays and weekends.

If the **p-value** is greater than the significance level (α = 0.05), we **fail to reject the null hypothesis**. This means that there isn't enough evidence to support the alternate hypothesis (which claims a difference exists).

Insights:

Consistent Demand Across Weekdays and Weekends:
The demand for bike rides does not show a statistically significant difference between weekdays and weekends. This implies that users tend to rent bikes at a similar rate irrespective of the day of the week.

Possible Reasons for Consistency:
This trend may be indicative of a user base that uses bikes for both commuting and leisure, maintaining similar demand throughout the week.
Cities with high demand for micro-mobility for daily commutes (work/school) and recreational activities might explain this pattern.

Operational Efficiency:
Since demand is relatively consistent across the week, it offers a predictable pattern that can optimize resource allocation, such as staffing, maintenance schedules, and distribution of bicycles.

Recommendations:

Promotional Campaigns:
Focus campaigns on maintaining engagement with all users throughout the week, without differentiating heavily between weekdays and weekends.

Inventory Management:
Allocate resources (bikes, support personnel, etc.) evenly across the week, optimizing for daily usage patterns instead of concentrating solely on weekends.

User Retention Strategies:
Since demand is consistent, investing in customer loyalty programs or personalized recommendations may be more impactful than day-based offers.

Impact on Service Optimization:
With consistent demand, it may be beneficial to maintain a balanced supply of bicycles across all days rather than overloading efforts to stock up for weekends specifically.
It can lead to cost-saving strategies by ensuring resources aren't disproportionately allocated based on an assumed difference in usage.

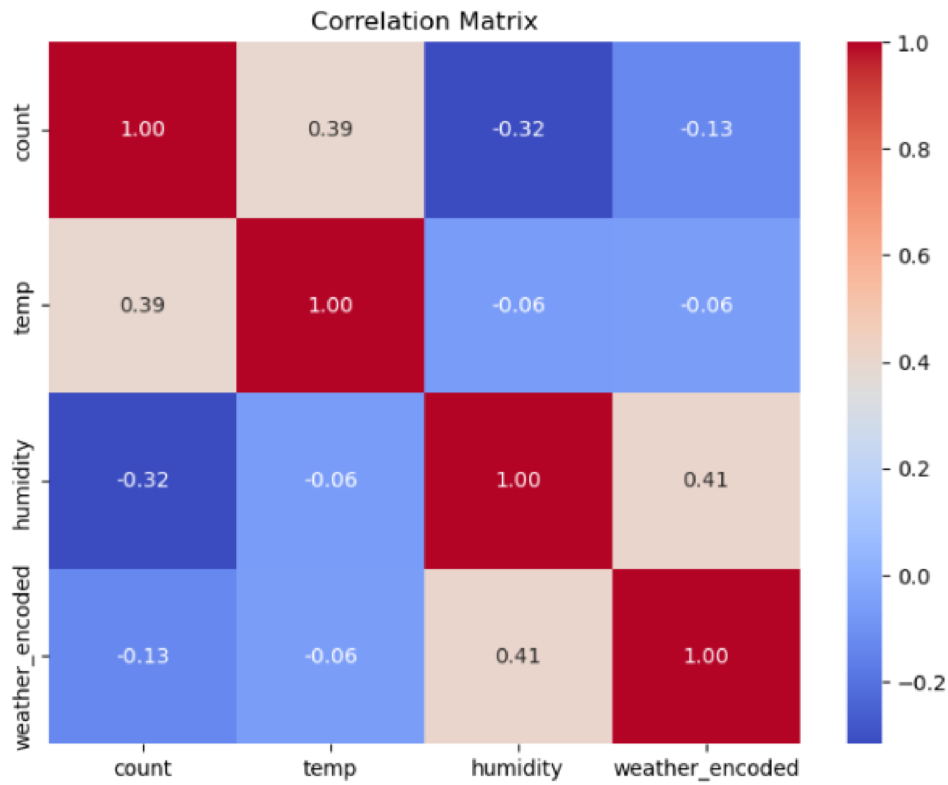
10.Correlation analysis between demand (count) and features such as temperature (temp), humidity (humidity), and weather (weather).

Code:

```
correlation_features = ['count', 'temp', 'humidity', 'weather_encoded']

# Calculate the correlation matrix
correlation_matrix = df[correlation_features].corr()

# Display the correlation matrix
print(correlation_matrix)

# Visualize the correlation matrix using a heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```

Output:

	count	temp	humidity	weather_encoded
count	1.000000	0.394454	-0.317371	-0.128655
temp	0.394454	1.000000	-0.064949	-0.055035
humidity	-0.317371	-0.064949	1.000000	0.406244
weather_encoded	-0.128655	-0.055035	0.406244	1.000000



Insights:

Temperature Boosts Demand: There's a moderate positive correlation (0.39) between temperature and demand. Users prefer riding when the weather is warm and comfortable.

Humidity Lowers Demand: A moderate negative correlation (-0.32) indicates that higher humidity reduces demand, likely due to discomfort while cycling in humid conditions.

Weather Impacts Slightly: Poorer weather conditions (rain/snow) show a weak negative effect on demand (-0.13), suggesting that adverse weather slightly deters users.

Weather and Humidity Linked: There's a moderate positive correlation (0.41) between weather conditions and humidity, implying rainy or snowy weather often coincides with higher humidity.

Recommendations:

Seasonal Strategy: Since demand increases with temperature and decreases with humidity, Yulu should focus on promoting its services during moderate weather conditions. They can also consider increasing the availability of cycles during warmer, less humid periods.
Pricing Adjustments: Implement dynamic pricing strategies based on weather conditions. For instance, offering discounts during rainy or humid weather might help maintain demand.
User Engagement: Communicate with users about favorable weather forecasts and encourage usage during such times, potentially boosting demand during peak weather conditions.

11.The demand for bicycles on rent is the same for different weather conditions

Formulate Hypotheses

Null Hypothesis (H0): The demand for bicycles on rent (measured by count) is the same across different weather conditions.

Alternate Hypothesis (H1): The demand for bicycles on rent (measured by count) is different across at least one pair of weather conditions.

Select an Appropriate Test

One-Way ANOVA Test: This test is appropriate because it compares the means of count across more than two groups (different weather conditions).

Code:

```
weather_groups = [df[df['weather'] == x]['count'] for x in df['weather'].unique()]
levene_stat, levene_pval = stats.levene(*weather_groups)
print(f'Levene's Test Statistic: {levene_stat}, p-value: {levene_pval}')
```

Output:

```
Levene's Test Statistic: 54.85106195954556, p-value: 3.504937946833238e-35
```

One-Way ANOVA Test:

```
model = ols('count ~ C(weather)', data=df).fit()
anova_results = anova_lm(model)
print(anova_results)

p_value = anova_results['PR(>F)'][0]
alpha = 0.05
if p_value <= alpha:
  print("Reject the null hypothesis: Significant differences exist between weather conditions.")
else:
  print("Fail to reject the null hypothesis: No significant differences between weather conditions.")
```

Output:

```
Reject the null hypothesis: Significant differences exist between weather conditions.
```

Assessment of Steps and Findings:

Levene's Test for Homogeneity of Variances:

Test Statistic: 54.85

p-value: 3.50e-35 (a very small value)

Interpretation:

Since the p-value from Levene's test is much smaller than the significance level ($\alpha = 0.05$), it indicates that the variances across different weather groups are not equal (i.e., the assumption of homogeneity of variance is violated).

Although ANOVA is robust to some deviations from homogeneity of variances, this result suggests you should be cautious in interpreting the ANOVA results and consider follow-up tests or use non-parametric alternatives if necessary. Nevertheless, it's reasonable to proceed with the analysis while keeping this in mind.

One-Way ANOVA Results:
F-statistic: 65.53
p-value: 5.48e-42

Interpretation:
The p-value is extremely small (much less than the significance level of 0.05), meaning we can reject the null hypothesis.
This indicates that there are statistically significant differences in the bicycle rental demand among different weather conditions.

Insights and Recommendations:

Impact of Weather on Bicycle Demand:
There is strong statistical evidence that weather conditions significantly affect the demand for bicycle rentals. This insight suggests that customer demand varies notably depending on weather.

Operational and Strategic Implications:
Forecasting and Inventory Management: Depending on the weather, you may need to adjust inventory (bicycles available for rent) or staffing at rental stations.

Promotional Strategies: Implement targeted promotions or discounts based on weather predictions to encourage rentals during less favorable conditions.

Infrastructure Planning: For unfavorable weather conditions (e.g., rainy or extreme weather), consider providing additional shelters or facilities to attract more customers.

Conclusion:

There is a significant difference in bicycle rental demand based on weather conditions, which suggests that rental operations and marketing strategies should be adaptable to weather changes.

12.Check if weather conditions are significantly different during different seasons
Formulate the Hypotheses
Null Hypothesis (H0): Weather conditions are independent of seasons (i.e., there is no significant difference in weather conditions across different seasons).

Alternative Hypothesis (H1): Weather conditions are dependent on seasons (i.e., there is a significant difference in weather conditions across different seasons).

Test: Chi-square test of independence is appropriate for testing the association between two categorical variables, in this case, 'Weather' and 'Season'.

Creating a Contingency Table
Code:
```
# Creating a contingency table
contingency_table = pd.crosstab(df['weather'], df['season'])
print(contingency_table)
```

Output:

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

The Chi-square test of independence.

Code:

chi2, p, dof, expected = chi2_contingency(contingency_table)

print(f"Chi-square Statistic: {chi2}")
print(f"p-value: {p}")
print(f"Degrees of Freedom: {dof}")
print("Expected Frequencies:")
print(expected)

P value:-1.5499250736864862e-07

```
Chi-square Statistic: 49.158655596893363
p-value: 1.5499250736864862e-07
```

Decide Whether to Accept or Reject the Null Hypothesis-

Code:

if p <= alpha:
print("Reject the null hypothesis: There is a significant difference in weather conditions across different seasons.")
else:
print("Fail to reject the null hypothesis: There is no significant difference in weather conditions across different seasons.")

Output

```
Reject the null hypothesis: There is a significant difference in weather conditions across different seasons.
```

Coclusion:

The contingency table and test results the conclusion that weather conditions differ significantly across different se asons. The extremely small p-value (1.5499e-07) clearly indicates a significant association between weather condit ions and seasons. Thus, Reject the null hypothesis: There is a significant difference in weather conditions across dif ferent seasons.

Insights:

Seasonal Variation in Weather Conditions:
The analysis demonstrates that weather conditions are not evenly distributed across seasons. Some weather conditions are more prevalent in specific seasons, highlighting a clear dependency between weather and seasonal changes.
For instance, **Weather Condition 4** is rare and only appears in **Season 1**. This could point to a unique or extreme weather event or condition during that season.

Implications for Seasonal Planning:
The clear differentiation in weather conditions suggests that activities, demand, or services affected by weather should be tailored to reflect these variations. For example, transportation, agriculture, or tourism sectors might need to adjust their strategies seasonally to address weather-dependent changes in demand or service needs.
If this analysis pertains to rental services (such as bike rentals or shared mobility solutions), demand strategies can be adjusted based on expected weather conditions per season to optimize service availability and resource allocation.

Recommendations:

Season-specific Strategies: Tailor services, marketing, and resource allocation based on the expected weather conditions of each season to optimize efficiency and user satisfaction.
Further Analysis: Conduct more granular studies on what drives changes in weather conditions across different seasons. Analyze the impact of other factors such as temperature, humidity, and windspeed.
Proactive Planning for Rare Events: Investigate the impact of rare weather conditions (e.g., Condition 4 in Season 1) and develop contingency plans to mitigate risks or leverage opportunities arising from such events.

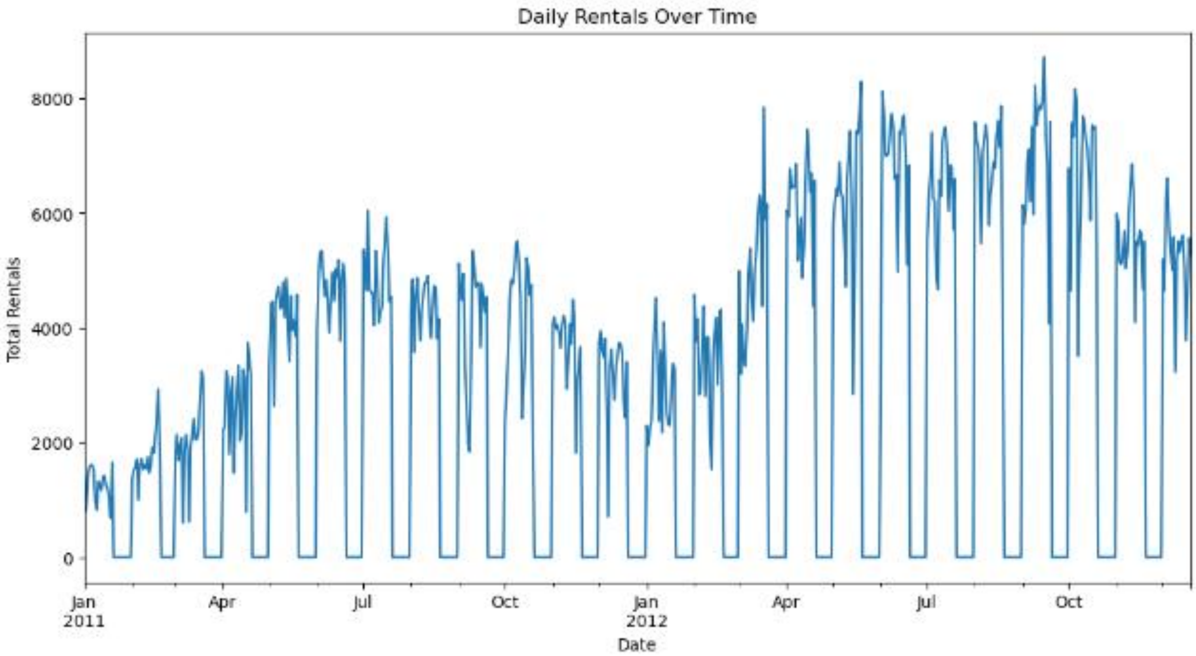
13. Time Series Analysis

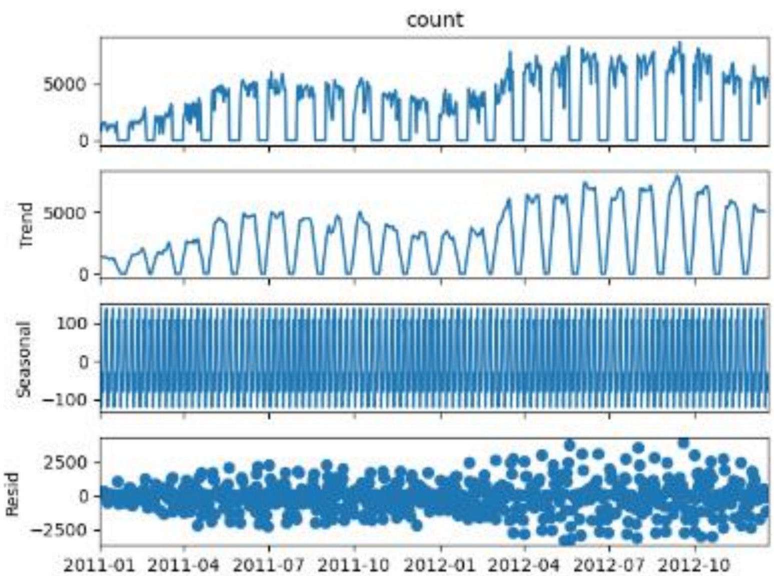
How do bicycle rental demands change over time?

Code:

```
df['datetime'] = pd.to_datetime(df['datetime'])
df.set_index('datetime', inplace=True)

# Resample data to daily rentals (or hourly if preferred)
daily_rentals = df['count'].resample('D').sum()

# Plotting the time series data
import matplotlib.pyplot as plt
plt.figure(figsize=(12, 6))
daily_rentals.plot(title='Daily Rentals Over Time')
plt.xlabel('Date')
plt.ylabel('Total Rentals')
plt.show()

# Decompose the time series
from statsmodels.tsa.seasonal import seasonal_decompose
result = seasonal_decompose(daily_rentals, model='additive')
result.plot()
plt.show()
```

Output:





Insights:

Trend Identification:

There appears to be a **positive trend** in bicycle rentals over the analyzed period, indicating an **increasing popularity of Yulu bicycles**. This trend could suggest that more customers are becoming aware of and using the service over time.

Seasonality:

The plot shows **clear seasonal fluctuations** in rental demand. There are regular peaks and troughs in the data, which suggest that certain periods of the year consistently have higher demand.

The decomposition shows that there is a strong **seasonal component** that can indicate predictable patterns such as increased usage during certain months or seasons (e.g., warmer weather).

Cyclic Patterns:

While there is a noticeable seasonal trend, there may also be **cyclic variations** due to external factors such as public events, holidays, or even economic conditions.

Recommendations:

Seasonal Promotions: Since there is a strong seasonal demand, Yulu can plan **promotions and campaigns** during low-demand periods to encourage higher usage. Conversely, Yulu could offer **premium pricing or additional services** during peak periods.

Demand Forecasting: Use this trend and seasonality to **forecast future demand**, helping Yulu prepare better by optimizing fleet allocation and maintenance schedules to ensure availability when demand is high.

Weather and Climate Adaptation: Given the seasonality, it may also be beneficial to **offer promotions or discounts during weather conditions** that typically see reduced rentals. Alternatively, developing **weather-adaptive offers or incentives** may keep demand steady.

Operational Planning: Plan **staffing, inventory, and operations** around peak times to provide better customer experiences and minimize wait times or availability issues.

Event-Based Marketing: Consider leveraging **special local events or holidays** to drive further engagement during periods that already exhibit spikes in demand.

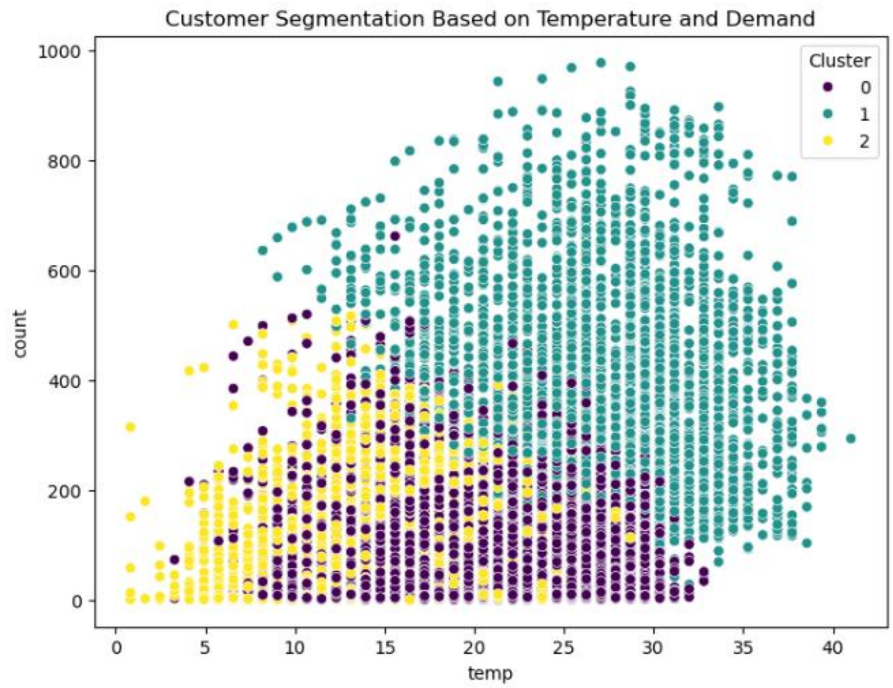
Expanding Reach during High Demand: During high-demand seasons, expanding **geographic reach** or introducing **more bicycles and docking stations** could capture more of the increasing demand.

14.Customer Segmentation Based on Temperature and Demand.

Code:

```
features = df[['temp', 'humidity', 'windspeed', 'count']]
scaler = StandardScaler()
features_scaled = scaler.fit_transform(features)

# Apply k-means
kmeans = KMeans(n_clusters=3, random_state=42)
clusters = kmeans.fit_predict(features_scaled)
df['Cluster'] = clusters

# Visualize the clusters (example for 2 features)
import seaborn as sns
plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x='temp', y='count', hue='Cluster', palette='viridis')
plt.title('Customer Segmentation Based on Temperature and Demand')
plt.show()
```

Output:



Insights and Recommendations:

Cluster 0 (Purple Points) – Low Demand Segment:
Cluster 0 generally represents lower demand and is concentrated at **lower temperatures** (below 15°C).
Insight: Users in this segment may avoid using electric cycles during cold weather, possibly due to discomfort in low temperatures.
Recommendation: Yulu could consider offering special promotions or discounts during colder weather to boost usage among this segment.

Cluster 1 (Yellow Points) – Medium Demand Segment:
This cluster spans across a wider range of temperatures (5°C to 25°C) but is mostly concentrated at lower counts (below 400).
Insight: Users in this segment show moderate usage, likely influenced by other factors beyond temperature (e.g., weather conditions, time of day).
Recommendation: Target this group with personalized marketing or incentives to convert medium-demand users into more frequent users, especially in moderate weather conditions.

Cluster 2 (Teal Points) – High Demand Segment:
Cluster 2 is the largest group, showing high demand, particularly at moderate to warm temperatures (15°C to 35°C). The counts frequently reach values above 600.
Insight: This segment demonstrates strong preference for using electric cycles during comfortable temperatures. It aligns with Yulu's core user base who are likely commuters or leisure users.
Recommendation: Optimize fleet availability during peak temperatures (20°C to 30°C), when demand is highest. Consider introducing loyalty programs for frequent users in this segment.

Temperature and Demand Relationship:
The segmentation shows a clear relationship between temperature and demand. High demand occurs mostly in moderate temperature ranges (20°C to 30°C), while both low and high temperatures (extremes) see reduced demand.

Overall Conclusion and Recommendations based on all the analysis:

The comprehensive analysis of Yulu's shared electric cycle data reveals key drivers of demand, including seasonality, weather conditions, and user preferences. The findings show a clear upward trend in overall demand, signifying a growing adoption of micro-mobility services. However, demand fluctuations are closely tied to external factors like temperature, humidity, and seasonal changes, indicating the need for adaptive strategies. By leveraging tailored marketing efforts, dynamic pricing, and efficient fleet management, Yulu can optimize its operations, better meet customer needs, and address the challenges posed by adverse weather conditions. Strategic focus on enhancing the experience for both casual and registered users will be crucial for sustained growth and revenue recovery.

Actionable Strategic Insights for Yulu

Growing Demand and Seasonal Trends:
There is a consistent upward trend in bicycle rentals, indicating increasing interest in micro-mobility solutions and a growing customer base for Yulu.
Demand exhibits seasonal variation, with higher rentals during warmer months and dips during colder or rainy seasons. This pattern underscores the importance of aligning service offerings with seasonal demand shifts.

Impact of Weather on Demand:
Adverse weather conditions such as rain or extreme temperatures negatively affect rental demand. The predictable influence of weather highlights the need for adaptive strategies in fleet management and resource allocation during unfavorable weather.

Seasonal Differentiation and Strategy:
Rental patterns vary significantly across seasons, suggesting the need for tailored approaches. For instance, offering special promotions during colder seasons or leveraging peak summer demand can enhance user engagement and drive usage.

Segment-Specific Marketing:
Implement customized marketing strategies based on temperature-based customer segmentation. For example, incentivize users in low-demand segments (e.g., Cluster 0) with cold-weather promotions and optimize service for high-demand segments (e.g., Cluster 2) during peak temperature ranges.

Dynamic Pricing Implementation:
Introduce dynamic pricing to maximize demand, especially during off-peak periods like early mornings or cold days. Adjust prices to encourage usage when demand is typically lower, balancing the rental distribution across time periods.

Enhanced Fleet Management for Peak Demand:
Prioritize fleet availability and maintenance during moderate temperatures (20°C to 30°C), where the highest demand is observed. Ensuring optimal service during these periods can maximize revenue and improve customer satisfaction.

Overall Recommendations to Yulu:

Demand-Responsive Operations:
Fleet Optimization: Ensure adequate bicycle availability and maintenance during peak seasons to maximize customer satisfaction and usage.
Predictive Demand Management: Use predictive analytics and demand forecasting to **anticipate high- and low-demand periods** and adapt operational plans accordingly.

Seasonal Campaigns and Offers:
Promotional Campaigns during Low Demand: Offer discounts, loyalty programs, and targeted promotions during off-peak months or unfavorable weather conditions to maintain a baseline demand.
Peak Season Strategies: During high-demand periods, consider **premium service options, higher fleet availability, or bundled offers** to capture maximum value.

Weather-Adaptive Services:
Consider providing **rain gear, heated seats, or weather-proof bicycles** during adverse conditions to encourage continued usage.
Offer real-time weather alerts and adaptive rental pricing based on weather forecasts.

Expanding Coverage and User Base:
Geographic Expansion: Consider expanding to other regions or cities where demand trends may follow similar patterns.
Targeted Marketing Campaigns: Leverage user segmentation data to create **tailored campaigns** that resonate with specific customer profiles and encourage increased usage.

Community and Eco-Friendly Branding:
Emphasize Yulu's commitment to **environmentally-friendly, sustainable transportation** in its branding to further capture eco-conscious customers.
Partner with Local Communities to enhance visibility and adoption through events, group rides, and community engagement initiatives.

Optimized Pricing Models:
Introduce **dynamic pricing strategies** based on demand, weather, and seasonality to ensure optimal revenue while maintaining customer satisfaction.
Consider **subscription-based services or corporate partnerships** to create steady revenue streams.

Jupyter Notebook Analysis

For a detailed view of the full analysis, including code, visualizations please refer to the complete Jupyter notebook available in the PDF format. The notebook documents each step of the analysis process, from data exploration to the final recommendations.

You can access the Jupyter notebook PDF through the **following link**:

Yulu Analysis