**Table of Contents**

[EXECUTIVE SUMMARY 3](#_Toc197376404)

[INTRODUCTION 4](#_Toc197376405)

[1. Business background 4](#_Toc197376406)

[2. Business challenge 4](#_Toc197376407)

[3. Dataset Overview 4](#_Toc197376408)

[4. Objectives 6](#_Toc197376409)

[METHODOLOGY AND DATA PRE-PROCESSING 6](#_Toc197376410)

[1. Data Pre-processing 6](#_Toc197376411)

[2. Outlier Detection 8](#_Toc197376412)

[3. Data Post-processing 9](#_Toc197376413)

[4. Methodology 10](#_Toc197376414)

[BIKE-SHARING TRAVEL CHARACTERISTICS ANALYSIS 12](#_Toc197376415)

[1. User Background 12](#_Toc197376416)

[2. Correlation analysis 13](#_Toc197376417)

[3. Bike-Sharing Demand: Time Characteristics Analysis 14](#_Toc197376418)

[4. Bike-Sharing Demand: Meteorology Characteristics Analysis 17](#_Toc197376419)

[PREDICTIVE MODEL 19](#_Toc197376420)

[1. Primary Approaches 19](#_Toc197376421)

[2. Final Model 20](#_Toc197376422)

[3. Model Evaluation 23](#_Toc197376423)

[4. Statistical Analysis 24](#_Toc197376424)

[5. Scenarios Development and Interpretation 26](#_Toc197376425)

[CONCLUSION 31](#_Toc197376426)

[RECOMMENDATIONS 31](#_Toc197376427)

# EXECUTIVE SUMMARY

In response to growing demand for sustainable urban mobility, Capital Bikeshare (CaBi) has become an integral part of Washington D.C.’s transportation network. However, as the system expands, it faces challenges in aligning its operations with diverse user behaviors, fluctuating demand, and bike availability. This report investigates bike usage patterns across these segments through an in-depth analysis of cleaned and structured data, focusing on key factors such as seasonality, temperature, working days, and user type.

The analysis separates users into casual and registered groups, revealing stark behavioral contrasts. Registered users show consistent weekday usage, particularly during commuting hours, and minimally influenced by weather conditions, indicating routine-driven behaviour. In contrast, casual users are more active on weekends and holidays, with their demand highly dependent on favorable weather and temperature. Using a range of visualization techniques and predictive modeling approaches, these tools confirm these patterns and uncover seasonal peaks, especially during summer, highlighting the risk of service imbalance and underutilization during off-peak periods.

To address these challenges, two key strategies are proposed. A time-specific rebalancing approach is recommended to ensure adequate bike availability during peak commuting hours. Additionally, a seasonal maintenance plan is introduced to align operation resources with demand trends. Thus, those two help to ensure availability, optimize costs, and enhance user experience while supporting sustainable transport goals.

# INTRODUCTION

## Business background

Bike-sharing systems are essential for sustainable urban transportation, reducing carbon emissions and alleviating urban traffic (Shaheen et al. 2010). Established in 2010, Capital Bikeshare provides extensive service in Washington, D.C., accommodating both registered commuters and casual users (Capital Bikeshare n.d.). In light of the increasing popularity of cycling due to environmental consciousness, customer demand varies with weather conditions, time of day, and seasons, creating challenges for Capital Bikeshare.

## Business challenge

Driven by rising sustainability initiatives, urban policies that encourage low-carbon transportation, and changes in mobility patterns, has led to a heightened demand for bike-share services (Barbour and Mannering 2023; Shaheen et al. 2010), offering an opportunity for providers such as Capital Bikeshare to expand in Washington D.C. However, these systems struggle to operate, including uneven bike distribution during peak usage periods (Kim 2024), seasonal demand fluctuations (Godavarthy and Taleqani 2017), and disruption from different weather conditions (Kumar 2021). If not addressed, these issues can lead to user dissatisfaction, lower ridership, and increased running expenses, especially worrisome as bike-sharing networks grow under strong urban mobility competition.

## Dataset Overview

A summary table of the original variables is provided to understand the dataset, thereby determining well-articulated objectives for the upcoming analysis (Table 1).

**Table 1.** Variable Summary (UCI Irvins 2025)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable name** | **Role** | **Type** | **Description** | **Units** |
| instant | ID | Integer | Record Index | - |
| dteday | Feature | Date | Date | - |
| season | Feature | Categorical | 1: Winter, 2: Spring, 3: Summer, 4: Fall | - |
| yr | Feature | Categorical | Year (0: 2011, 1: 2012) | - |
| mnth | Feature | Categorical | Month (1 to 12) | - |
| hr | Feature | Categorical | Hour (0 to 23) | - |
| holiday | Feature | Binary | Whether the day is a holiday (from DC holiday schedule) | - |
| weekday | Feature | Categorical | Day of the week | - |
| workingday | Feature | Binary | 1 if the day is neither weekend nor holiday, otherwise 0 | - |
| weathersit | Feature | Categorical | 1: Clear, Few clouds, Partly cloudy; 2: Mist/Cloudy;  3: Light Snow/Rain; 4: Heavy Rain/Snow | - |
| temp | Feature | Continuous | Normalized temperature in Celsius: | °C |
| atemp | Feature | Continuous | Normalized feeling temperature in Celsius: | °C |
| hum | Feature | Continuous | Normalized humidity (divided by 100) | - |
| windspeed | Feature | Continuous | Normalized wind speed (divided by 67) | - |
| casual | Other | Integer | Count of casual users | - |
| registered | Other | Integer | Count of registered users | - |
| cnt | Target | Integer | Count of total rental bikes including both casual and registered | - |

## Objectives

Given the related business problems of a bikeshare service, the report aims to determine the customer profiles of the casual and registered users, including the persona and customer behaviours. It is crucial because the users’ distinctive usage purposes and schedules influence the company’s operations and logistics plans to meet customers’ demand. Further, the paper attempts to predict the usage at specific times of day, determine the high and low seasons to develop impactful operational strategies.

# METHODOLOGY AND DATA PRE-PROCESSING

## Data Pre-processing

Data pre-processing is a vital step enabling accurate analysis and improving analytical models’ performance by ensuring data is consistent, well-structured, and valid (Maharana et al. 2022; Fan et al, 2021). The chosen dataset comprises 17 variables and 17,379 rows without missing or duplicated values.

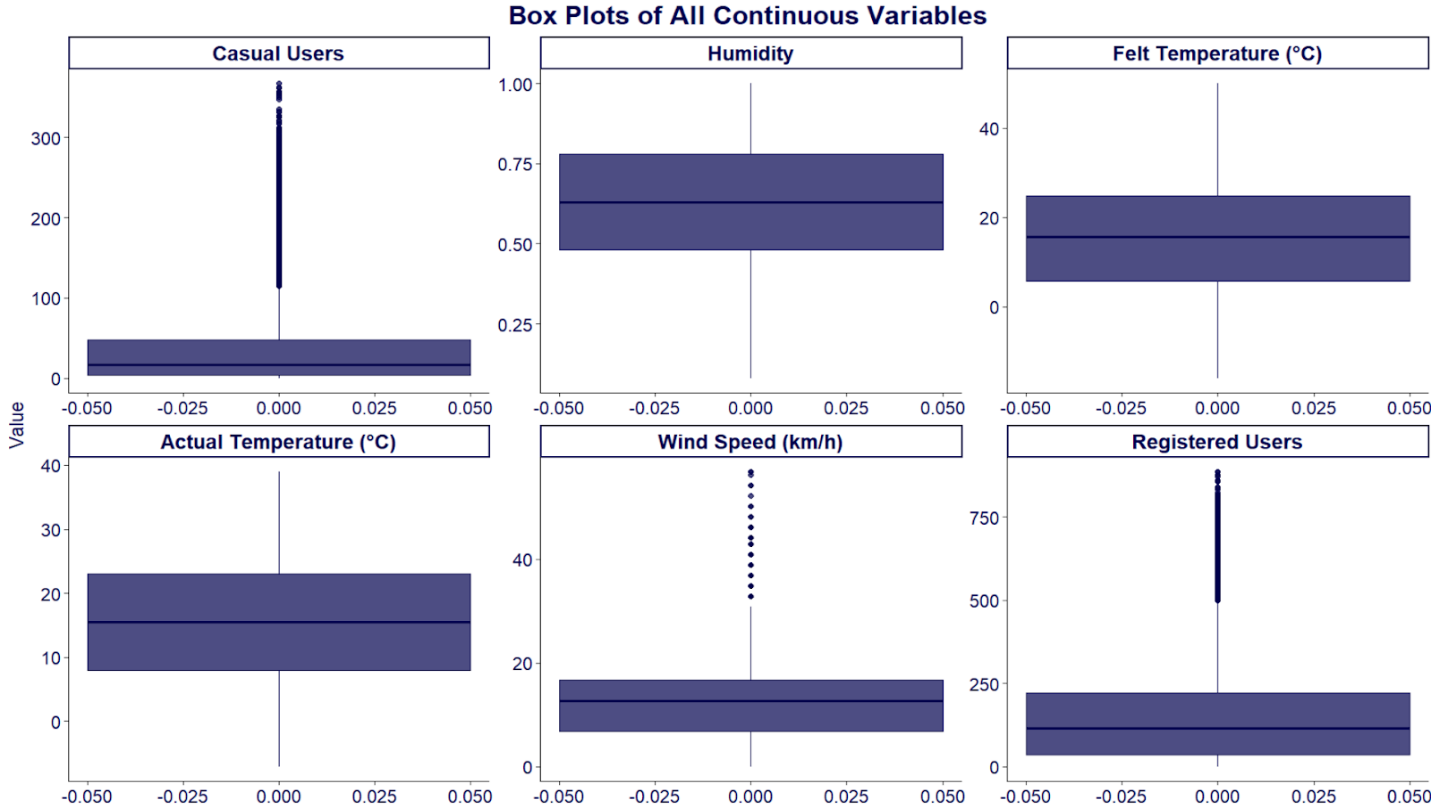
To meet the dataset selection requirement, time-series variables are removed from the dataset (Table 2).

**Table 2.** Data Preprocessing (R Script 1)

|  |  |
| --- | --- |
| **Data Preprocessing** | **Justification** |
| **Deletion of Time-series Variables** | The dataset contains the variable `instant` representing an hour in a day, collected from 1/1/2011 to 1/1/2012, and ‘dteday’, which captures the date information, making it a time-series variable in nature.  The analysis aims to explore relationships between variables with limited emphasis on the temporal trends. The deletion is to prevent the analysis from being time-series focused, ensuring analytical simplicity. Therefore, the variables `instant`, `dteday` are removed from the dataset.  `mnth` and `yr` are also removed to match the purpose of treating the data as a cross-sectional dataset. |
| **Deletion of Erroneous Data** | Twenty-two instances of the `hum` (humidity) variable recorded 0.00 humidity levels, which is physically unrealistic under natural outdoor conditions (Alger 2015). The error can result in misunderstanding and flaws in the analysis and decision-making. Since these erroneous entries accounted for only 0.13% of the data, which is far below the 5% threshold commonly considered acceptable for deletion without introducing bias (Yeatts and Martin 2015). Therefore, the incorrect observations are removed from the data to ensure the reliability and accuracy of the analysis. |
| **Variable Recoding** | The dataset has `workingday`, `holiday` and `weekday`, representing the date type of the recorded data points. By definition, a holiday is not a working day. It means that the `workingday` variable is derived from the `holiday` variable. The dependency can introduce bias to visualization and prediction. Meanwhile, `weekday` is independent from `holiday` because any weekday can be a holiday.  Therefore, `workingday` is removed from the data and `weekday` is renamed to `weekend`. The Monday to Friday observations are recoded to 0 (weekday) and, the Saturday and Sunday observations are recoded to 1 (weekend). |
| **Categorizing `hr` observations** | To ensure the simplicity of the analysis, `hr` observations are grouped into 6 distinctive hour blocks. The interval of each hour is set based on the behavioral pattern driving bike usage demand.  ·       Late Night: (0 - 4) - Minimal activity, sleep period.  ·       Early Morning: (5 - 6) - Pre-commute activity starts.  ·       Morning Peak: (7 - 9) - Commute/start of day rush.  ·       Midday: (10 - 16) - Core working/leisure hours, including lunch.  ·       Afternoon Peak: (17 - 19) - End of work/evening commute rush.  ·       Evening: (20 - 23) - Post-commute leisure, dinner time. |
| **Data Conversion** | The dataset includes categorical variables (`season`, `holiday`, `weathersit`, `weekend`, `yr`, `mnth`) that are already in numeric form but represent categories rather than quantities. Categorical variables are converted to factors to preserve their categorical nature.  Regression models require a proper distinction between categorical and continuous numerical variables to correctly interpret their influence. Treating categorical variables as continuous could lead to inaccurate model estimations (Rhemtulla et al. 2012). |
| **Variables Transformation** | Because scale variables, including `temp`, `atemp`, and `windspeed`, cause difficulties when interpreting the result, they are transformed back to the normal variables following their scaling formulas. Transformed variables are renamed to `raw\_temp`, `raw\_felt\_temp` and `raw\_windspeed`. |

## Outlier Detection

There are no outliers in `raw\_temp`, ‘raw\_felt\_temp’ and ‘hum’, whereas ‘raw\_windspeed’, ‘casual’, and ‘registered’ are opposite (Figure 1). However, the outliers present naturally occurring weather factors and usage demand rather than error. For example, fluctuations in 'raw\_windspeed' are often induced by weather dynamics (Zou and Djokic 2020), whereas rental numbers can normally increase over the weekends, holidays, or pleasant weather (Kinoshita et al. 2024; Shi et al. 2023). Further, outliers in usage variables may result from the demand spikes in special events. Therefore, removing these outliers would eliminate useful variation, capturing real behavioral and environmental patterns. As a result, they are retained to maintain data integrity as well as modeling accuracy.



**Figure 1.** Boxplots of each continuous variable (R Script 2)

## Data Post-processing

After preprocessing, the dataset consisted of 17,357 rows and 11 columns, with no missing values (Table 3).

**Table 3.** Data Post-processing (R-script 2)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Data types** | **Level** | **Missing** |
| **season** | Factor | 1 - Winter  2 - Spring  3 - Summer  4 - Fall | 0 |
| **holiday** | Factor | 0 - No  1 - Yes | 0 |
| **weathersit** | Factor | 1 - Clear, Few clouds, Partly cloudy  2 - Mist/Cloudy  3 - Light Snow/Rain  4 - Heavy Rain/Snow | 0 |
| **hum** | Numeric | Null | 0 |
| **casual** | Numeric | Null | 0 |
| **registered** | Numeric | Null | 0 |
| **raw\_temp** | Numeric | Null | 0 |
| **raw\_felt\_temp** | Numeric | Null | 0 |
| **raw\_windspeed** | Numeric | Null | 0 |
| **weekend** | Factor | 0 - No  1- Yes | 0 |
| **hr\_cat** | Numeric | Null | 0 |

## Methodology

To support data visualization and predictive modeling, the dataset was analyzed using several methods detailed in Table 4. The process relied on R programming packages, comprising of ‘readr’, ‘tidyverse’, ‘naniar’, ‘gtExtras’, ‘patchwork’, ‘gridExtra’, ‘skimr’, ‘VIM’, ‘corrplot’, ‘ggplot2’, ‘grid’, and ‘car’ are used to conduct the analysis.

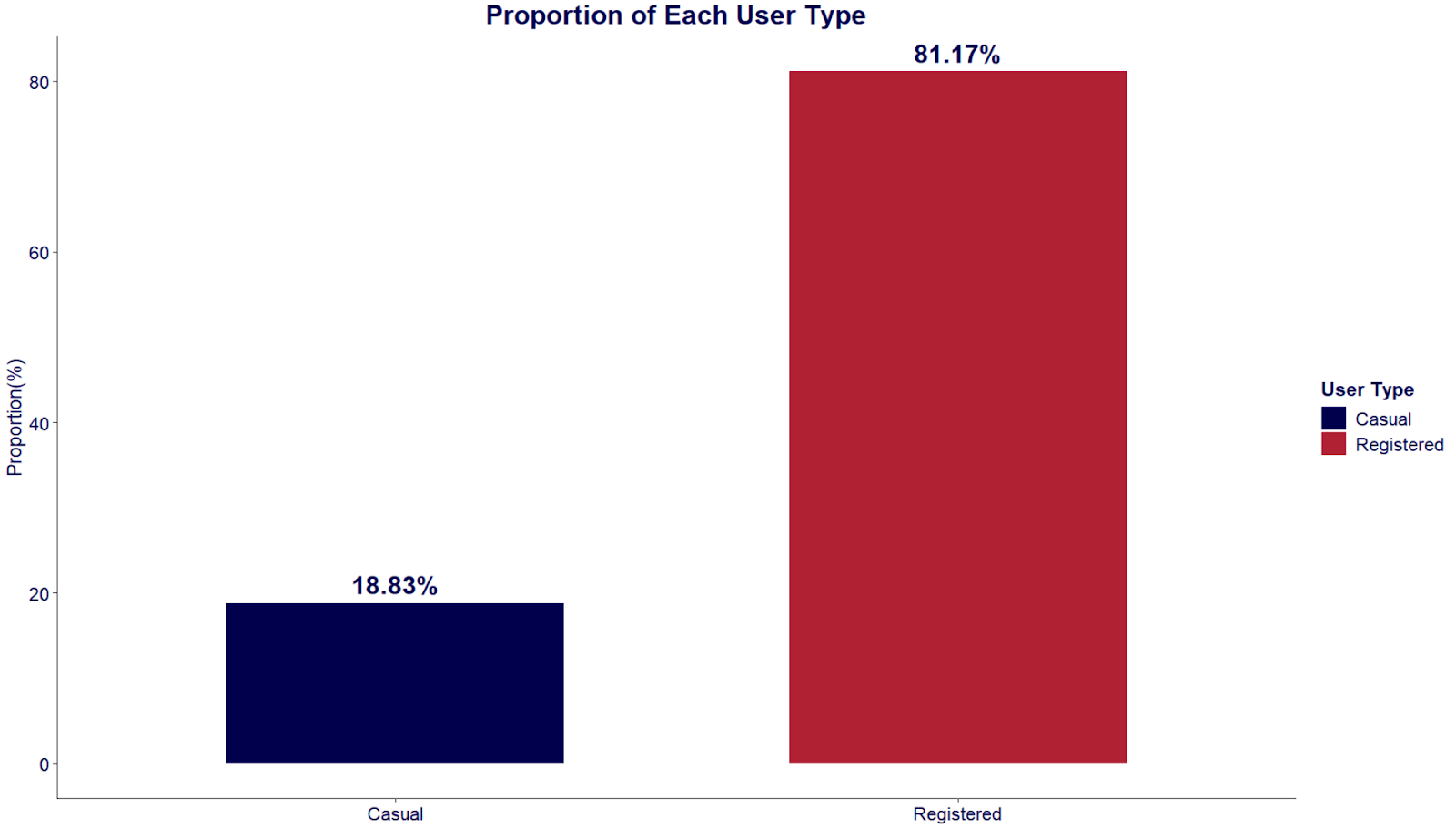
**Table 4**. Methodology of Data Analysis

|  |  |  |
| --- | --- | --- |
| **Method/Measurement** | | **Rationale** |
| **Measurement of central tendency** | Median | Due to the existence of outliers and the heavy skewness of `casual` and `registered`, the median is selected to represent the differences between variables instead of the mean (Laerd Statistic n.d.). The median is insensitive to the outliers, thereby interpreting the variability between variables more accurately. |
| **Measurement of Association** | Correlation Plot | Because the goal is to create multiple linear regression models, selecting IVs having linear relationships with the DVs is crucial to formulate highly accurate prediction models. Correlation analysis is a suitable method to determine the relationship between the variables, thereby selecting reliable and statistically significant variables for the analysis.  Given that the dataset has both numeric and ordinal variables, the Spearman correlation is utilized instead of the Pearson correlation (Laerd Statistics n.d.). Spearman correlation exhibits the monotonic relationship between variables based on the rank of the variables rather than the data value. As a result, the mechanism is robust to outliers and captures the association between variables with higher accuracy. Further, Spearman correlation is applicable to ordinal variables, where Pearson cannot, making it more robust to the current dataset, thereby helping select suitable IVs for regression analysis (Laerd Statistics n.d.). |
| **Data Visualization** | Box plot | Box plots are the primary means of visualization throughout the report. Box plots illustrate the summary of central tendency, the spread and the skewness of data, thereby demonstrating the differences between categories in categorical variables. Further, box plots are useful in visualizing the presence of outliers, thereby determining irregularities such as extreme outliers and erroneous values. |
|  | Column Chart | The column chart is the primary means of visualization in the analysis, allowing comparisons of usage between different seasons, day types, weather situations and hour blocks. |
| **Prediction Model** | Multiple Linear Regression | Given that both `casual` and `registered` variables are continuous variables, multiple linear regression is most suitable to predict their dependence on IVs in the dataset. Further, the regression’s coefficients can determine the degrees of change in the DVs when the IVs change (Laerd Statistics n.d.).. Therefore, the analysis can create different scenarios based on the models, and then have a deeper understanding of the business problems to make impactful decisions. |

# BIKE-SHARING TRAVEL CHARACTERISTICS ANALYSIS

## User Background

The critical aspect of users is the dominance of registered users, accounting for 81.17% of total recorded trips (Figure 2). The differences between the two user types are defined by their access method: registered users commit to a long-term plan via the annual and 30-day membership, while casual users pay for short-term day passes (Capital Bikeshare 2011). The fundamental distinction in the commitment levels suggests a potential divergent usage pattern and schedule across temporal and weather-related variables. Further, the findings may support the analysis to determine the customer persona, thereby building suitable plans to satisfy the users.

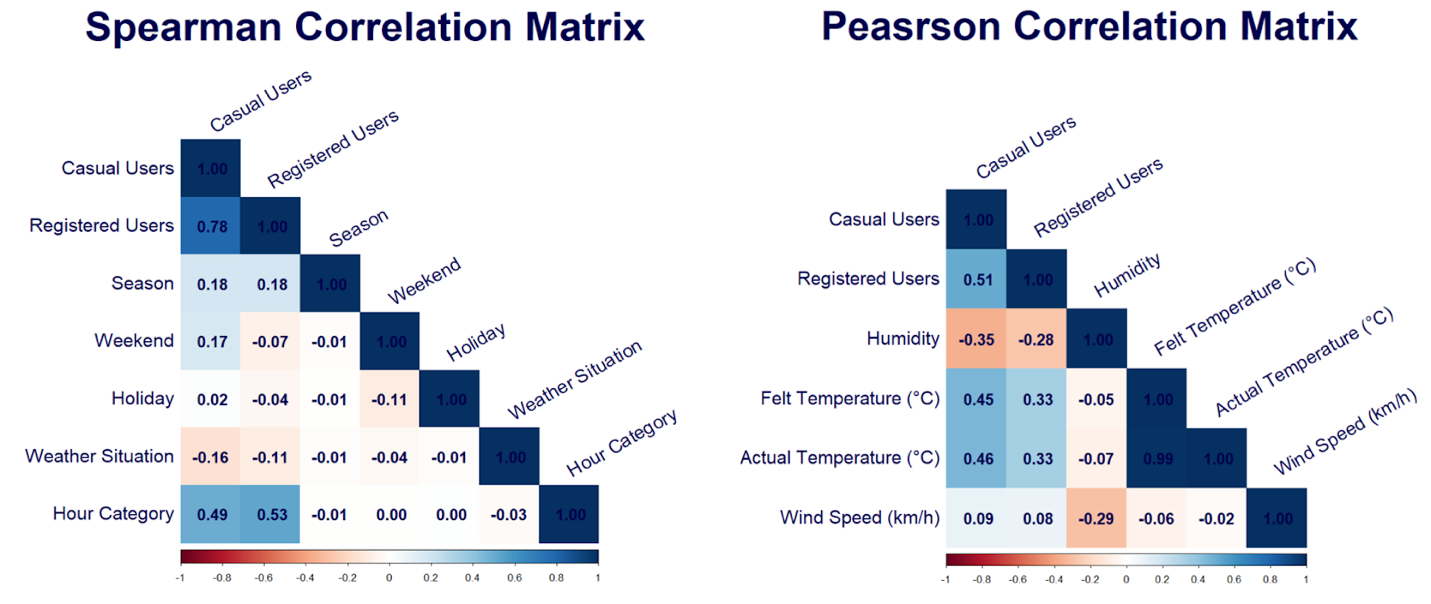


**Figure 2.** Column Chart of Proportion of Each User Type

## Correlation analysis

Focusing on two fundamental variables, `registered` (registered users) and `casual` (casual users), the correlation analysis provides insights into the monotonic relationships with different influencing factors (Figure 3). Strong and statistically significant positive correlation between `hr\_cat` and both casual (0.49) and registered (0.53) emphasizes the importance of hour blocks. `weekend` also highlights the contrast between casual (0.17) and registered (-0.07) users, indicating their fundamental differences (Appendix 1).

`raw\_felt\_temp` exhibits a positive correlation with casual (0.46) and registered  (0.33), implying a higher sensitivity to weather changes among casual users. In contrast, `hum` has a significant negative relationship with both users, indicating a subtle role in predicting bike usage. Although the correlations between `weathersit` and usage `variables are weak, it has a practical importance in exploring the impacts of season fluctuations on bike demand.

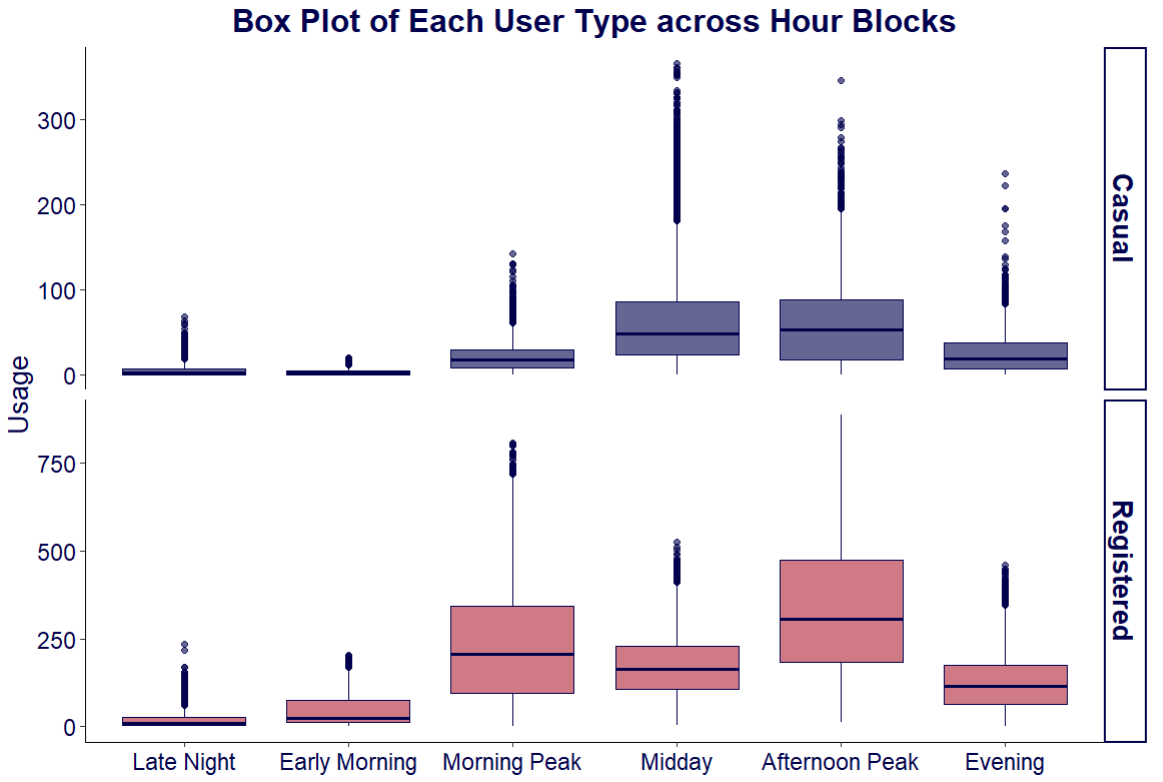


**Figure 3.** Correlation Heatmap of Bike-Sharing Demand and Related Variables (R Script 3)

## Bike-Sharing Demand: Time Characteristics Analysis

**Relationship with Hour Blocks**

The ascending medians and IQRs shown by the box plots validate the positive relationship between ‘hr\_cat’ and the two usage variables. However, the drop after the Midday for casual users, and after the Afternoon Peak for registered users prevents the continuous increase through hour blocks. The pattern violates the perfect monotonicity, thereby lowering the Spearman coefficients. Further, Late Night and Early Morning show a negligible median compared to other hour blocks in both user types, confirming the times of minimal activity. The boxplots also reveal that casual users’ demand usually peaks at Midday, while registered usage surges during Morning and Afternoon peaks. This pattern reflects a preference for leisure hours among casual users, whereas registered users have a structured, work-related travel pattern. Overall, the distinctive trends emphasize the significance of `hr\_cat` in predicting bike usage of each user.

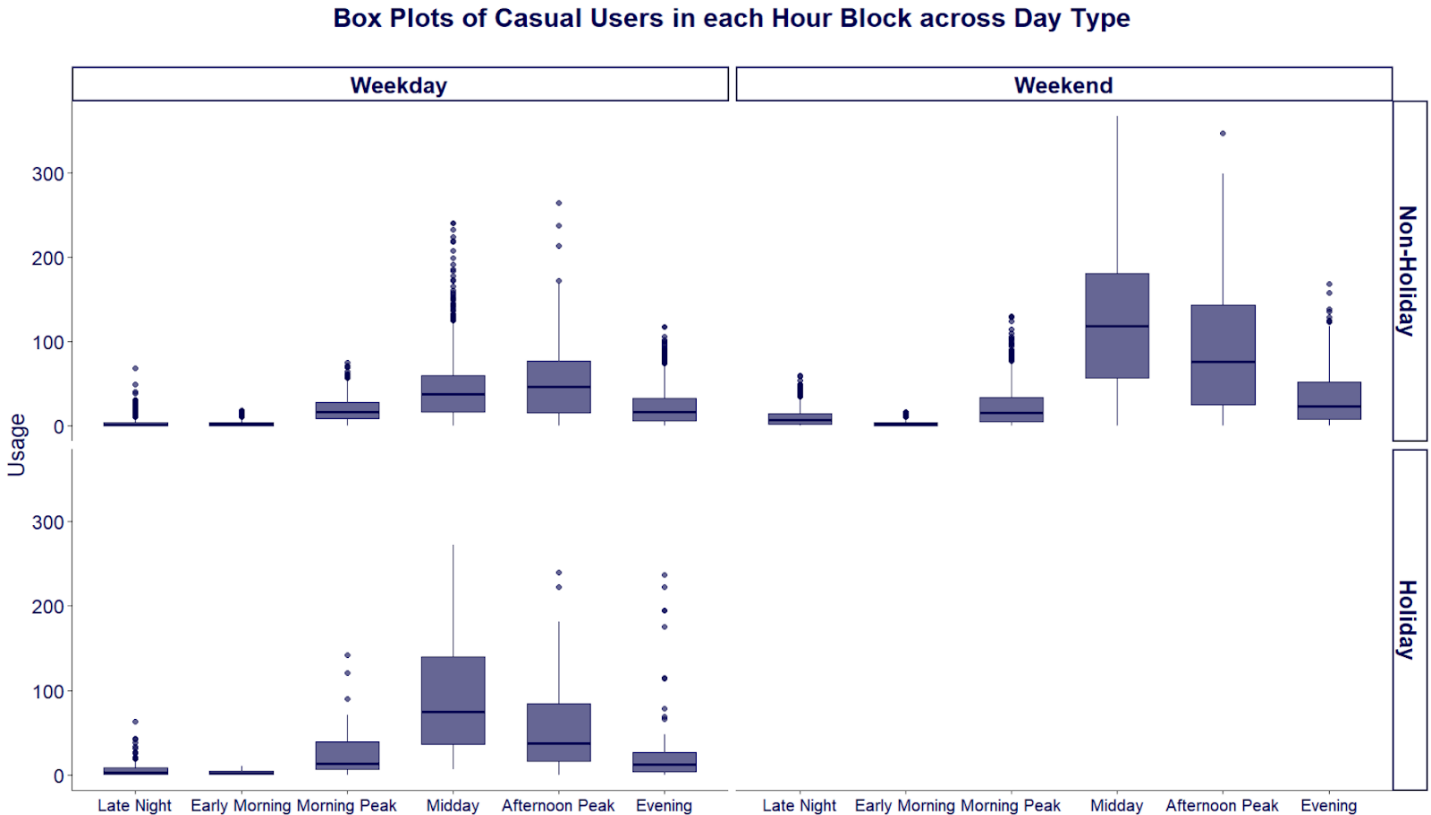


**Figure 4.** Hourly Usage of Capital Bikeshare by Hour Blocks (R Script 4)

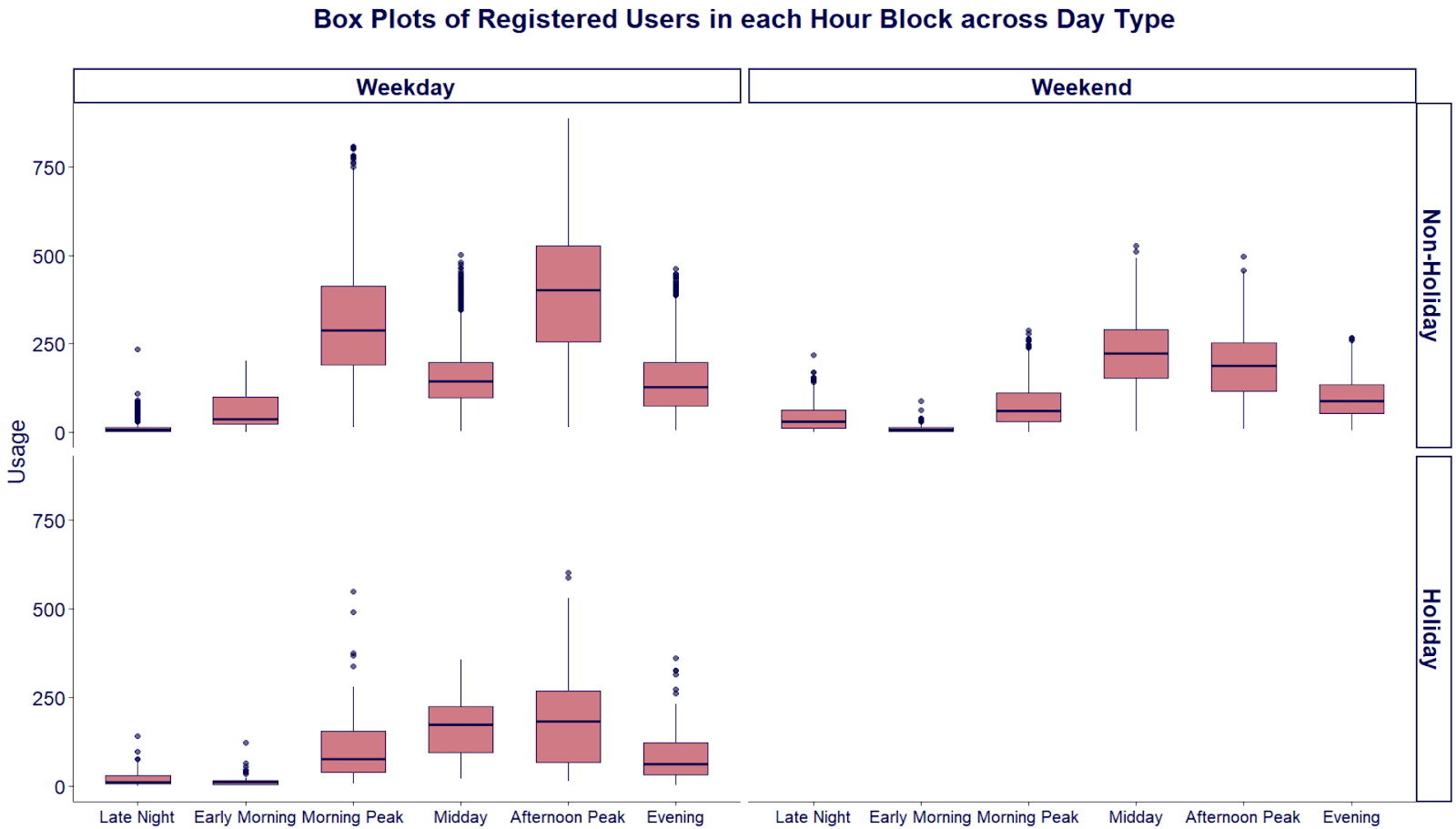
**Relationship with Weekend and Holiday**

Regarding casual users, the medians of non-holiday weekends and holiday weekdays are significantly higher than non-holiday weekdays during Midday and Afternoon Peak. Further, the holiday demonstrates a larger IQR in Midday, which indicates an increase in usage compared to non-holiday weekdays. Although there is a change in the magnitude of usage, the overall usage distributions across hour blocks change modestly. In contrast, the distributions of registered users noticeably fluctuate between different day types. Non-holiday weekend usage significantly drops during peak hours, whereas it slightly increases during Midday compared to non-holiday weekends. Further, holidays are associated with drops in demand compared to non-holiday weekdays, which is proven by the significant reduction in medians and IQRs.

In short, `weekend` has associations with the fluctuations of registered users in each hour block. `weekend` appears to be inconsiderably correlated with casual users in each hour block. `holiday` is expected to be associated with the changes in magnitude of both user types. The findings consolidate the assumption of differences in usage purposes between user types.



**Figure 5.** Box Plots of Casual Users in each Hour Block across Day Type (R Script 4)

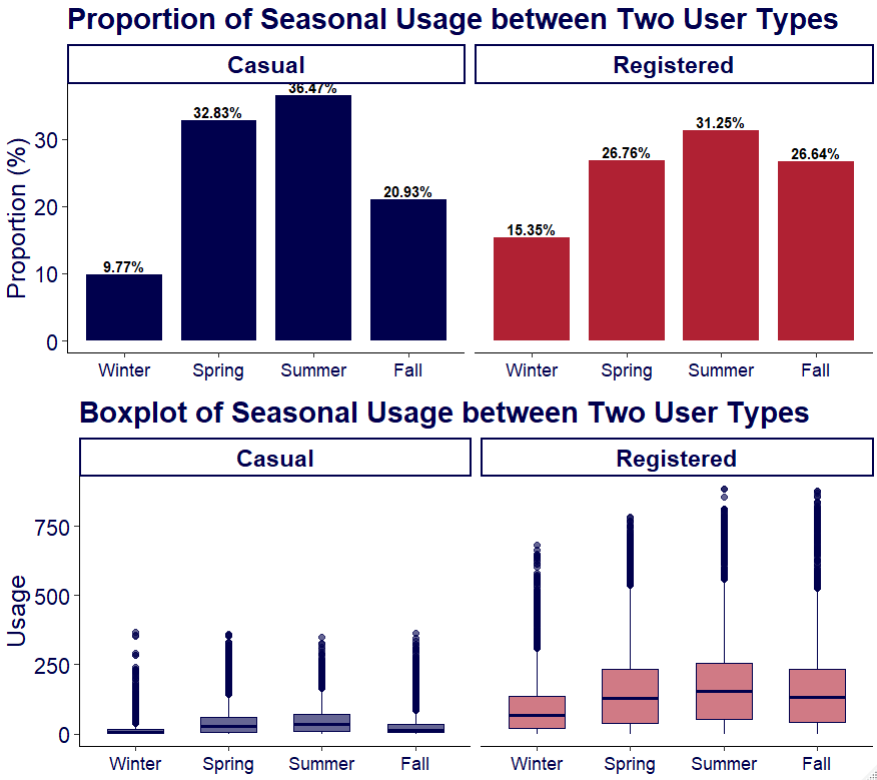


**Figure 6.** Box Plots of Registered Users in each Hour Block across Day Type (R Script 4)

## Bike-Sharing Demand: Meteorology Characteristics Analysis

**Relationship with Seasons**

Casual users primarily cluster in warmer seasons, particularly summer and spring, accounting for about 70% of total usage, while winter usage fell sharply to only 20.93% (Figure 7). The registered usage is distributed more evenly, with the highest proportion in Summer (31.25%), followed by Spring (26.76%) and Fall (26.64%). Due to shorter daylight hours and lower temperatures, the cycling activities are hindered, leading to the lowest proportion of usage in both groups (Gebhart and Noland 2014). Further, the boxplots of each user type demonstrate the changes in median, increasing from Winter to Summer, and dropping during Fall, which aligns with the positive Spearman coefficients. The findings suggest the importance of `season` categories in predicting the bike usage of both users.

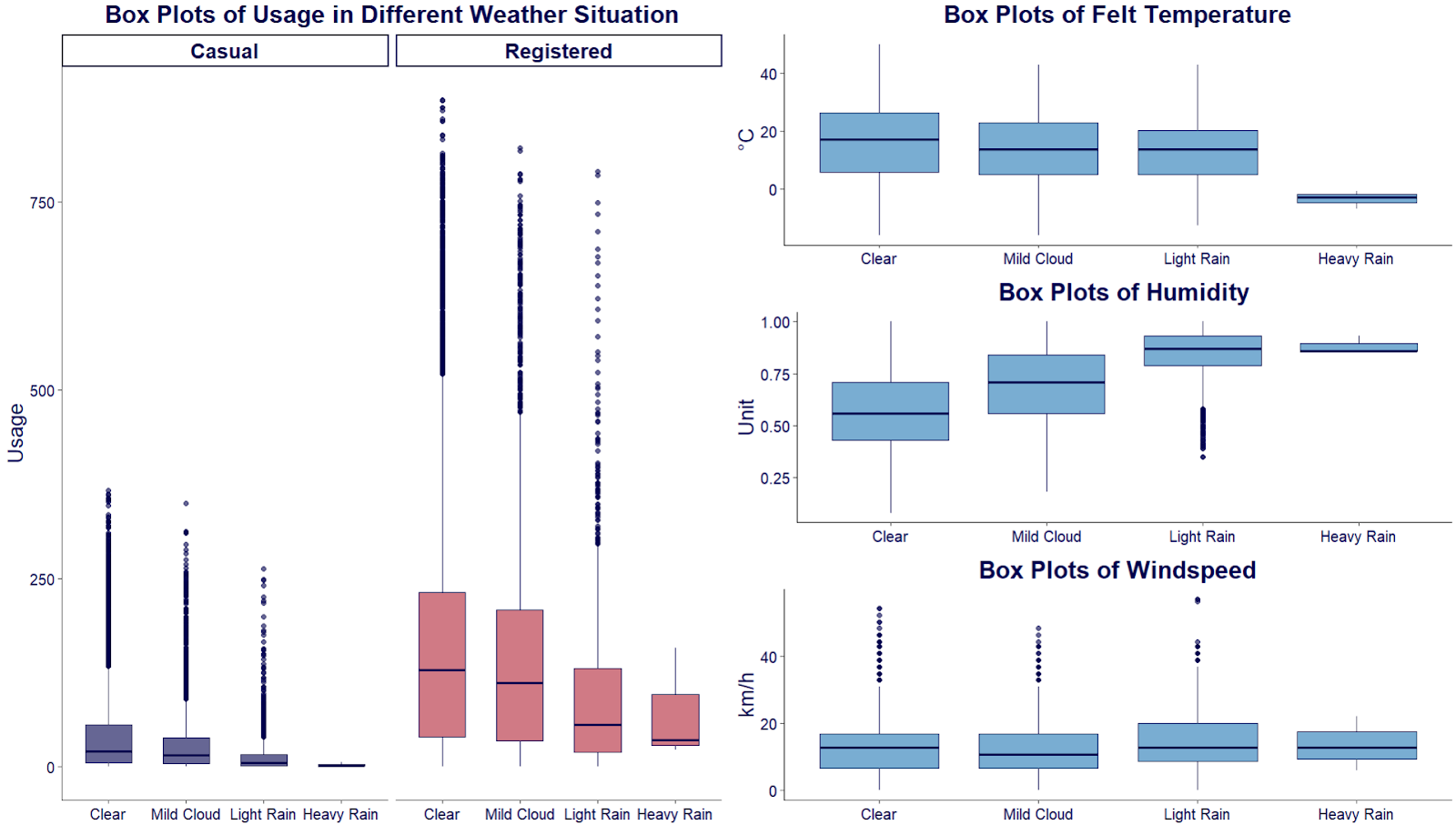


**Figure 7.** Column charts of seasonal proportion and box plots of seasonal usage of each user type (R Script 4)

**Relationship with Weather Conditions, Humidity and Felt Temperature**

Figure 8 confirms the negative relationship between usage variables and `weathersit` categories. The median usage of both user types decreases when the weather worsens, from clear conditions to light rain/snow. These findings provide the roles of related weather factors such as humidity, temperature and windspeed, thereby facilitating the exploration of their association with bike usages. Further, Figure also reveals distinctive patterns of `hum` and `raw\_felt\_temp` across the weather conditions. The clear weather has the lowest median humidity and the highest median felt temperature compared to the adverse weather conditions, such as light rain or snow. Regarding windspeed, its changes in median across different weather conditions are closely equal. Therefore, it is difficult to validate the positive relationship with bike usage variables.

The visual evidence supports the explored relationships from the correlation analysis because higher usage falls under the most favourable weather conditions with lower humidity and higher felt temperature. Therefore, the findings suggest that `weathersit`, `hum` and `raw\_felt\_temp` are potentially important predictors for the regressions. In contrast, the unclear association of `raw\_windspeed` results in its lower importance in the regression.



**Figure 8.** Box Plot of Usage, Felt Temperature, Humidity and Windspeed across Weather Conditions (R Script 4)

# PREDICTIVE MODEL

## Primary Approaches

Recognizing the distinctions between Casual and Registered users, the analysis creates separate prediction models for each user type. Further, the prediction involves multiple variables, including both numeric and categorical variables, to ensure the high accuracy of the model. Table 5 summarises the selected variables for the models with adequate justification.

**Table 5.**Primary approaches (R Script 5)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Casual User** | **Registered User** | **Justification** |
| `season` | Involved | Involved | `season` has a weak but statistically significant correlation with both user types. Further, involving `season` will help create different scenarios. |
| `hr\_cat` | Involved | Involved | A moderate coefficient indicates a positive linear relationship between times of day and usage demand. Further, visualizations also demonstrate the difference in mean between hour blocks, suggesting their influence on bike demand. |
| `weekend` | Involved | Involved | The visualizations shows noticeable distinctions between weekends and weekdays in the demand of each user type. Therefore, `weekend` is included in the regression to fully reflect the difference between user types and create meaningful scenarios. |
| `weathersit` | Involved | Involved | `weathersit` shows a weak negative correlation with both user types, indicating that the more extreme the weather, the lower the demand. |
| `hum` | Involved | Involved | `has` has a weak to moderate relationship with both user types’ demand, thereby playing a subtle role in the prediction. Further, `hum` will effectively display the impacts of weather factors on the usage fluctuations. |
| `atemp` | Involved | Involved | `atemp` is selected due to its moderate correlation with usage demand. Further, felt temperature reflects the perceived weather comfort of humans better than the actual temperature. |
| `hr\_cat`:`weekend` | Not involved | Involved | Registered user demand has distinctive differences between weekends and weekdays across hour blocks. Therefore, an interaction term is added to the registered user’s model. In contrast, although the interaction between hour blocks and day types exists, it contributes inconsiderably to the goodness of fit of the casual user’s model,  adding unnecessary complexity to the model. |

## Final Model

Because both `casual` and `registered` are heavily right-skewed, natural log transformation is applied to increase normality of the data. Log transformation increases the normality of data, making the residuals’ distribution more symmetrical, thereby increasing the accuracy of the model (Changyong et al. 2014).

**Casual User Model** (R Script 5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Estimate** | **t-value** | **P-value** | |
| (Intercept) | 0.672 | 23.939 | 0.000\*\*\* | |
| **hr** |  | | |
| Early morning (5-6) | -0.127 | -5.654 | 0.000\*\*\* | |
| Morning peak (7-9) | 1.471 | 75.306 | 0.000\*\*\* | |
| Midday (10-16) | 2.108 | 119.440 | 0.000\*\*\* | |
| Afternoon peak (17-19) | 2.056 | 98.309 | 0.000\*\*\* | |
| Evening (20-23) | 1.379 | 75.695 | 0.000\*\*\* | |
| **weekend** (yes) | 0.676 | 55.870 | 0.000\*\*\* | |
| **weathersit** |  | | |
| 2 (mist/cloudy) | -0.024 | -1.783 | 0.075 | |
| 3 (light snow/rain) | -0.561 | -24.694 | 0.000\*\*\* | |
| 4 (heavy rain/snow) | -0.272 | -0.656 | 0.512 | |
| **season** |  | | |
| 2 (Spring) | 0.495 | 25.431 | 0.000\*\*\* | |
| 3 (Summer) | 0.234 | 9.550 | 0.000\*\*\* | |
| 4 (Fall) | 0.441 | 25.924 | 0.000\*\*\* | |
| **holiday** (yes) | 0.458 | 13.974 | 0.000\*\*\* | |
| **hum** | -0.814 | -22.199 | 0.000\*\*\* | |
| **attempt** | 0.053 | 66.103 | 0.000\*\*\* | |
| **Adjusted R-squared:** 0.769 | | | | |
| **ln(casual+1) =** 0.672 − 0.127\*Early Morning + 1.471\*Morning Peak​ + 2.108\*Midday​ + 2.056\*Afternoon Peak ​ + 1.379\*Evening​ − 0.676\*Working Day − 0.024\*Cloudy  − 0.561\*Light Snow/Rain − 0.272\*Heavy Snow/Rain ​ + 0.495\*Spring + 0.234\*Summer + 0.441\*Fall + 0.457\*Holiday − 0.814\*Humidity + 0.053\*Raw Felt Temperature | | | | |

**Registered User Model** (R Script 5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Estimate** | **t-value** | **P-value** | |
| (Intercept) | 2.079 | 81.853 | 0.000\*\*\* | |
| **hr** |  | | |
| Early morning (5-6) | 1.603 | 69.203 | 0.000\*\*\* | |
| Morning peak (7-9) | 3.389 | 167.151 | 0.000\*\*\* | |
| Midday (10-16) | 2.501 | 140.071 | 0.000\*\*\* | |
| Afternoon peak (17-19) | 3.441 | 161.457 | 0.000\*\*\* | |
| Evening (20-23) | 2.473 | 131.469 | 0.000\*\*\* | |
| **weekend** (yes) | 1.169 | 50.564 | 0.000\*\*\* | |
| **weathersit** |  | | |
| 2 (mist/cloudy) | -0.008 | -0.661 | 0.508 | |
| 3 (light snow/rain) | -0.432 | -21.776 | 0.000\*\*\* | |
| 4 (heavy rain/snow) | 0.116 | 0.321 | 0.748 | |
| **season** |  | | |
| 2 (Spring) | 0.206 | 12.112 | 0.000\*\*\* | |
| 3 (Summer) | 0.152 | 7.101 | 0.000\*\*\* | |
| 4 (Fall) | 0.476 | 32.072 | 0.000\*\*\* | |
| **holiday** (yes) | -0.319 | -11.149 | 0.000\*\*\* | |
| **hum** | -0.615 | -19.197 | 0.000\*\*\* | |
| **attempt** | 0.027 | 38.577 | 0.000\*\*\* | |
| **hr:weekend** |  | | |
| Early morning: weekend 1 | -2.731 | -63.185 | 0.000\*\*\* | |
| Morning peak: weekend 1 | -2.684 | -71.490 | 0.000\*\*\* | |
| Midday: weekend 1 | -0.783 | -25.981 | 0.000\*\*\* | |
| Afternoon peak: weekend 1 | -1.875 | -49.986 | 0.000\*\*\* | |
| Evening: weekend 1 | -1.482 | -42.978 | 0.000\*\*\* | |
| **Adjusted R-squared:** 0.799 | | | | |
| **ln(registered + 1) =** 2.079 + 1.603\*EarlyMorning + 3.389\*MorningPeak + 2.501\*Midday + 3.441\*AfternoonPeak + 2.473\*Evening + 1.169\*Weekend - 0.008\*Mist\_Cloudy - 0.432\*LightSnow\_Rain + 0.116\*HeavySnow\_Rain + 0.206\*Spring + 0.152\*Summer + 0.476\*Fall - 0.319\*Holiday - 0.615\*Humidity + 0.027\*FeltTemperature - 2.731\*EarlyMorning:Weekend - 2.684\*MorningPeak:Weekend - 0.783\*Midday:Weekend - 1.875\*AfternoonPeak:Weekend - 1.482\*Evening:Weekend | | | | |

## Model Evaluation

All variables of both models are statistically significant following the significance threshold, except the cloudy weather (2) and heavy rain/snow (4). Statistically insignificant variables are usually removed from the regression to ensure accurate prediction. However, they are unremovable because they are derived from `weathersit`, thereby being ignored when building scenarios.

Predictors of the casual and registered user regression model can explain 76.9% and 79.9%, respectively, which are indicated by the adjusted R-squared of each model. The results indicate a very high reliability of the models in predicting bike usage in different contexts. Regarding the interaction between `hr\_cat` and `weekend`, while it significantly improves the accuracy of the registered user model by 9.2%, its impact on the other one is insignificant, with only 1.8% improvement (Appendix 1).

## Statistical Analysis

**Baseline model analysis**

The effects of predictors on the changes in demand of both models are compared to the defined baseline, in which all variables are equal to zero. The baseline scenario of the two models is bike usage at Late Nights on Weekdays with Clear Weather in the Winter, assuming 0% humidity and 0 degrees Celsius felt temperature. The models are able to predict the growth of demand and the estimated average usage in a specific condition, following the equations (1) and (2):

Therefore, the baseline models predict there is 1 casual user and 8 registered on average under the baseline condition.

**Time-dependent variables**

Firstly, `hr\_cat` is the most important factor in predicting the bike usage of both user types. The casual demand during midday, afternoon peak and morning peak hours is expected to increase by 723%, 681% and 335%, respectively, compared to the baseline. Conversely, registered users usually use bikes during morning and afternoon peak hours, which predicts the demand to grow by 2863% and 3021%, respectively.

Secondly, the variable `weekend` is always associated with an increase in casual users, indicated by a positive coefficient. In contrast, the relationship with registered users is more complex. With the interaction term, the changes in registered users’ demand in a specific hour block on weekends are calculated by equation (3). Therefore, registered demand during non-midday blocks is expected to decrease substantially in comparison with weekdays.

Further, the contrast between casual and registered users on holidays is also highlighted. While casual usage reliably increases on holidays by 58%, registered usage is predicted to drop by 37%. From the analysis of time-dependent variables, their correlations are varied between casual and registered groups, suggesting the difference in demographics and usage purposes between the 2 groups

**Meteorology variables**

Regarding seasonal and weather factors, seasons also have a major and positive relationship with the changes in demand. All seasons are reliable demand boosters for both user types. For casual users, the pure seasonal boost is largest in Spring, Fall, and Summer. For registered users, Fall has the strongest boost, followed by Spring and Summer. However, while Summer has the most observed usage in the visualization, its coefficients in the models are significantly smaller than those of the others. This discrepancy can be attributed to the differing roles of visualization and regression analysis. Seasonal averages in the visualization reflect aggregate demand influenced by multiple overlapping factors, including weather conditions. Conversely, the regression model isolates the effect of each variable when keeping the other predictors constant. Therefore, the low coefficients reflect the more direct contribution of weather factors to Summer demand. Further, it reflects the strong pure seasonal effect of Fall on the demand growth.

In the visualization, both casual and registered usages start to increase from Winter to Summer and then drop in Fall. Tucker and Gilliland (2007) also support that Spring, Summer and even the early Fall have the highest level of physical activity, while decreasing gradually at the end of the seasonal cycle.

The extremity of the weather condition suppresses usage demand when keeping other factors constant. Light rain/snow leads to a substantial decrease in demand because people tend to choose safer means of transportation, such as bus, train, or car, during bad weather. Evidently, even a small hint of rain can result in a significant drop in cycling rate (Böcker et al. 2013). Moreover, the positive relationship with `raw\_felt\_temp` and negative association with `hum` found in the visualization are validated by the regression. Comparing the demand changes of both user types in relation to `hum`, `raw\_felt\_temp` and `weathersit`, the casual model’s coefficients are more significant than those of the registered model. Therefore, casual users appear to be more sensitive to the fluctuation of weather conditions.

## Scenarios Development and Interpretation

The analysis highlights the importance of daily rhythm in shaping bikeshare demand, with specific time blocks aligning with typical human activities such as commuting, lunch breaks, and leisure. Moreover, behavioral differences between casual and registered users reflect user personas shaped by different habits, motivations, and sensitivity to time and weather conditions.

**Time-dependent scenarios**

We define the constant variables used across all time-dependent scenarios by fixing summer season on a non-holiday, while humidity and felt temperature are their mean values during summer (0.633 and 27.4, respectively). Weather conditions are kept clear with no mist, rain, or snow. Only time-dependent variables, like hour blocks and weekday/weekend status, will vary across scenarios.

In table 1, the afternoon peak yields the highest engagement among registered members (412), followed closely by the morning peak (391) and midday (161) (Table 6), suggesting that weekdays, especially during commuting hours, demand is driven largely by registered users.

Aligning with the statistical analysis finding, Tables 6,7 reveal that registered users have much lower usage on weekends compared to weekdays, suggesting this user type is more engaged with the system on working days. These opposing patterns suggest differing user demographics and mobility needs: registered users are likely routine commuters, as evidenced by peak usage during weekday peak periods, while casual users tend to use the service for leisure during weekend midday hours.

**Table 6.** Time-dependent scenarios with Weekday (R Script 6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | **1** | | **2** | | **3** | |
| **User Type** | Casual | Registered | Casual | Registered | Casual | Registered |
| Weekend | 0 | | | | | |
| Summer | **1** | | | | | |
| hum | **0.633** | | | | | |
| attempt | **27.4** | | | | | |
| MorningPeak | **1** | | 0 | | 0 | |
| Midday | 0 | | **1** | | 0 | |
| AfternoonPeak | 0 | | 0 | | **1** | |
| MorningPeak: weekend | 0 | | 0 | | 0 | |
| Midday: weekend | 0 | | 0 | | 0 | |
| AfternoonPeak: weekend | 0 | | 0 | | 0 | |
| **Results** | **13** | **391** | **26** | **161** | **24** | **412** |

**Table 7.** Time-dependent scenarios with Weekend (R Script 6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | **4** | | **5** | | **6** | |
| **User Type** | Casual | Registered | Casual | Registered | Casual | Registered |
| Weekend | **1** | | | | | |
| Summer | **1** | | | | | |
| hum | **0.633** | | | | | |
| attempt | **27.4** | | | | | |
| MorningPeak | **1** | | 0 | | 0 | |
| Midday | 0 | | **1** | | 0 | |
| AfternoonPeak | 0 | | 0 | | **1** | |
| MorningPeak: weekend | 0 | **1** | 0 | | 0 | |
| Midday: weekend | 0 | | 0 | **1** | 0 | |
| AfternoonPeak: weekend | 0 | | 0 | | 0 | **1** |
| **Results** | **27** | **86** | **51** | **237** | **49** | **203** |

Casual customers mainly use bikes for leisure-friendly periods, indicating a preference for flexibility, comfort, and recreational purposes rather than routine commuting, suggesting that they are often tourists or residents taking leisure or spontaneous trips. This inference is supported in Wöhner (2022) study, who noted that casual users travel more on non-work purposes like shopping and leisure. More recent paper by Kim (2024) further confirms that Capital Bikeshare casual users predominantly use the service for tourism and sightseeing purposes. Based on observed user behavior, casual users are most likely to visit public parks and tourist landmarks, places conducive to tourism and leisurely exploration.

Registered users show structured, routine-based travel patterns. Their usage peaks on weekday commute hours and notably declines on weekends and holidays, suggesting that bikeshare use among this group is driven primarily by work or educational obligations, demonstrating consistency (Willberg et al. 2021; Wang and Lindsey 2019). Tucker and Gilliland (2007) noted the same pattern: working populations tend to engage in schedule-bound physical activity. Similarly, Kim (2024) found long-term bikeshare users mainly used the system for commuting, with minimal weekend usage. This registered user persona indicates frequent travel between residential areas, office districts, major transit hubs, and university zones, locations associated with structured daily routines.

Weekend midday and afternoon peak hours generate the highest levels of usage. Midday is the most active period, with average casual and registered users reaching 51 and 237 trips, respectively (Scenario 5). These patterns reflect despite registered reduced commuting activity as most of them are off work, they are likely to travel for non-work purposes such as shopping and leisure  (Motte-Baumvol et al., 2024; Wöhner, 2022). This along with casual users are generally more active due to leisure and tourism (O’Brien et al. 2014; Fu et al. 2024), could lead to problems with service congestion, imbalances at docking stations, and a lack of available bikes, all of which could make users unsatisfied.

**Seasonal-dependent scenarios**

We assume a fixed status of afternoon peak hour, non-weekend, non-holiday, and clear weather, while humidity and felt temperature will vary by season, with mean values specific to each season; thus, the impact of seasonal changes will be concentrated.

In table 8, both registered and casual users reach their highest usage in summer (25 and 442) (Scenario 8). Conversely, during winter, when spring, summer, and fall indicators are 0 (Scenario 10), the activity of both user types is low (5 and 192). This highlights distinct seasonal patterns shaped by climate. Notably, the sharper decline of casual users is witnessed during winter compared to summer.

**Table 8.** Seasonal-dependent scenarios (R Script 6)

| **Scenario** | **7** | | | **8** | | **9** | | **10** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **User Type** | Casual | Registered | | Casual | Registered | Casual | Registered | Casual | Registered |
| Spring | **1** | | | 0 | | 0 | | 0 | |
| Summer | 0 | | | **1** | | 0 | | 0 | |
| Fall | 0 | | | 0 | | **1** | | 0 | |
| hum | 0.627 | | | 0.633 | | 0.667 | | 0.584 | |
| attempt | 18.4 | | | 27.4 | | 11.4 | | 3.63 | |
| **Results** | **20** | | **342** | **25** | **442** | **12** | **362** | **5** | **192** |

Bike usage peaks in spring and summer, seasons associated with comfortable conditions, validating weather-driven demand patterns, which are further justified in Bean et al. (2021) study across different cities, indicating that bike-sharing usage peaks during seasons with favorable weather conditions. With sharper decline in usage, casual users are more weather-sensitive  (Jaber and Csonka 2023a; Thomas et al. 2013), registered users exhibit a more balanced seasonal distribution, reflecting riding patterns of commuting or daily routines, which is less influenced by seasonality (Nosal and Miranda-Moreno 2014). This is consistent with Böcker et al. (2013) study, noting that non-routine travelers' outdoor physical activities and travel behaviors are heavily influenced by weather conditions.

Together, these insights point to distinct high- and low-traffic season periods shaped by climate. Summer represents an excessive demand period that could lead to bike shortages, station congestion, and added pressure on bike quality deterioration. Winter, on the other hand, creating underutilization risks as well as inefficiency in the use of resources.

# CONCLUSION

This report examined usage trends within the Capital Bikeshare system by analyzing patterns across different times of day, seasons, and weather conditions. The results revealed clear differences in behavior between casual and registered users. Registered users generally adhere to a structured routine using the system predominantly during weekday commuting periods. Their usage exhibits relative stability across seasons and is minimally influenced by weather variations. Conversely, casual users typically utilize rides during midday, especially on weekends and holidays, showing greater sensitivity to external factors such as temperature and humidity.

# RECOMMENDATIONS

|  |  |  |  |
| --- | --- | --- | --- |
| **Recommendation** | **Justification** | **Strategic Implications** | |
| **1. Time-Based Bike Rebalancing & Station Placement Optimization** | Registered users and casual users have different cycling time patterns. A suitable vehicle allocation strategy is needed to prevent the risk of undeserving both user types, enhancing the company's revenue. (Jaber and Csonka 2023b). Additionally, being unable to catch up with customer demand can deprive the opportunity to leverage the growing demand for health-conscious and sustainable mobility options.  Jaber et al. (2024) indicate that stations' alignment based on user routines improves their accessibility and encourages long-term adoption, reinforcing the need to synchronize station placement and size with specific customer demands for time and space, highlighted through regression, persona mapping, and scenarios. | With the persona analysis, these time patterns map to location clusters: registered users frequent transport hubs, business districts, and office towers, while casual users gravitate toward parks and tourist landmarks, underscoring the need for differentiated operational planning.  **Bike Rebalancing Action Plan – Weekday Operations**   1. **Morning Peak (6:00–9:00 AM)**   As this time block witnessed a high traffic volume originating from registered users, bikes are distributed to residential zones to facilitate outbound commuter activities.   1. **Midday Reinforcement of High-Traffic Zones (10:00 AM–4:00 PM)**   Relocate 70% of bikes from underutilized stations (Type A: residential, far from tourist and leisure zone) to high demand stations (Type B: office district and leisure area), while retaining a minimal operational buffer.  This step matches the casual demand, which is significantly higher during midday hours and nearby leisure destinations and prepare for the afternoon peak demand (Bachand-Marleau et al. 2012).   1. **Afternoon Return Allocation from Closing tourist spots  (4:00–5:00 PM)**   Bikes left in tourist spots after venue closure might be underutilized due to the reduction of casual users. Thus, bikes will be transfered from early-closing cultural or tourist sites to office stations in preparation for the afternoon peak commuter flows.   1. **Evening Rebalancing (8:00–11:00 PM)**   Shift inventory from low-activity peripheral zones to central business and entertainment districts using a 7:3 redistribution ratio, facilitating the recreation and nightlife use in downtown area (Faghih-Imani et al. 2014).   1. **Overnight Positioning (Post 11:00 PM)**   Deploy unused bikes to residential areas for early morning commuter readiness.  **Weekend Operations**   1. **Leisure-Oriented Allocation (All Day)**   Bike allocation concentrates within 3-5 km of the CBD, leverage on leisure and tourist activity zones. | |
|  | | |
| **2. Seasonal Maintenance Strategy** | The analysis shows significant differences between seasonality, including a sharp contrast between the summer peak period and the winter low-demand, which poses risks of resource underutilization in winter and operational strain during summer. Since the bikes continuously operate year-round, the quality deteriorates, which results in accident risk from equipment failure, threatening both customer safety and brand reputation, particularly when users come to the service looking for a safe and comfortable cycling experience (Maioli et al. 2019).  In response, Ahmadi et al. 2024 suggesting ultilizing off-peak periods for equipment maintenance, minimizing service disruption while optimizing readiness for peak season. | Based on the findings of demand rising significantly in spring, peaking in summer, and reducing gradually in fall and winter, we proposed a maintenance plan customized to minimize operation disruption.  **Optimize the Rotation Maintenance Cycle**  Maintenance is strategically concentrated in winter, when the overall demand is at its lowest, less than half of the peak season. This allows efficient maintenance without affecting user availability as long as it prepares the system for the high-demand spring and summer period.  **The cycle involves 3 main phases.**   1. **Initial Maintenance at Leisure, tourist stations**   As the finding indicates that casual users are more weather-sensitive, casual demand reduced much more significantly compared with registered during off-peak season. This pattern can clearly be observed at table 5 when casual usage drops by over 90% in winter, while registered user activity declines by approximately 57% compared to spring.  Fleet size at casual dominant stations of leisure and tourist use is therefore significantly reduced. Bikes from these stations are withdrawn for preventive check-ups and maintenance with a small buffer of 20% to maintain basic service availability.   1. **Office and Residential stations as recipients of maintained vehicles**   Maintained bikes are later redistributed to office and residential stations, ensuring availability for registered users using these stations for commuting traffic.   1. **Unmaintained Bikes from Office and Residential Stations Extraction**   The same number of bikes at the Office and residential stations that have not yet been serviced, especially those nearing usage thresholds, are then redistributed to leisure and tourist stations to enter the next rotation of the maintenance cycle. This strategy ensures continuous operations while maintaining system quality standards. | |
|  | | |

**REFERENCES**

Ahmadi SA, Ghasemi P and Ehmke JF (2024) ‘Optimizing urban bike-sharing systems: a stochastic mathematical model for infrastructure planning’, *Central European Journal of Operations Research*, doi:10.1007/s10100-024-00950-z.

Alger M (2015) *Mike Alger: Is 0% humidity possible?*, Reno Gazette Journal website, accessed 4 May 2025. https://www.rgj.com/story/news/2015/08/29/mike-alger-humidity-possible/71349608/

Bachand-Marleau J, Lee BHY and El-Geneidy AM (2012) ‘Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use’, *Transportation Research Record: Journal of the Transportation Research Board*, 2314(1):66–71, doi:10.3141/2314-09.

Barbour N and Mannering F (2023) ‘Intended cycling frequency and the role of happiness and environmental friendliness after COVID-19’, *Scientific Reports*, 13(1):636, doi:10.1038/s41598-023-27681-6.

Bean R, Pojani D and Corcoran J (2021) ‘How does weather affect bikeshare use? A comparative analysis of forty cities across climate zones’, *Journal of Transport Geography*, 95:103155, doi:10.1016/j.jtrangeo.2021.103155.

Böcker L, Dijst M and Prillwitz J (2013) ‘Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review’, *Transport Reviews*, 33(1):71–91, doi:10.1080/01441647.2012.747114.

Capital Bikeshare (2011) *Capital Bikeshare 2011 Member Survey Report,* Capital Bikeshare website, accessed 5 May 2025. https://capitalbikeshare.com/system-data

Capital Bikeshare (n.d.) *All About Capital Bikeshare*, Capital Bikeshare website, accessed 4 May 2025. https://capitalbikeshare.com/how-it-works

Changyong F, Hongyue W, Naiji L, Tian C, Hua H, Ying L. and Xin M (2014) ‘Log-transformation and its implications for data analysis’, *Shanghai archives of psychiatry*, *26*(2): 105.

Faghih-Imani A, Eluru N, El-Geneidy AM, Rabbat M and Haq U (2014) ‘How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal’, *Journal of Transport Geography*, 41:306–314, doi:10.1016/j.jtrangeo.2014.01.013.

Fan C, Chen M, Wang X, Wang J and Huang B (2021) ‘A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data’, *Frontiers in Energy Research*, 9:652801, doi:10.3389/fenrg.2021.652801.

Fu C, Yen BTH and Yeh C-J (2024) ‘The analysis of the shared bike usage pattern: Application of survival model to Taiwan YouBike’, *Asian Transport Studies*, 10:100125, doi:10.1016/j.eastsj.2024.100125.

Gebhart K and Noland RB (2014) ‘The impact of weather conditions on bikeshare trips in Washington, DC’, *Transportation*, 41(6):1205–1225, doi:10.1007/s11116-014-9540-7.

Godavarthy RP and Rahim Taleqani A (2017) ‘Winter bikesharing in US: User willingness, and operator’s challenges and best practices’, *Sustainable Cities and Society*, 30:254–262, doi:10.1016/j.scs.2017.02.006.

Jaber A and Csonka B (2023a) ‘How Do Land Use, Built Environment and Transportation Facilities Affect Bike-Sharing Trip Destinations?’, *Promet*, 35(1):119–132, doi:10.7307/ptt.v35i1.67.

Jaber A and Csonka B (2023b) ‘Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models’, *Alexandria Engineering Journal*, 77:1–13, doi:10.1016/j.aej.2023.06.087.

Jaber A, Ashqar H and Csonka B (2024) ‘Determining the Location of Shared Electric Micro-Mobility Stations in Urban Environment’, *Urban Science*, 8(2):64, doi:10.3390/urbansci8020064.

Kim K (2024) ‘Discovering spatiotemporal usage patterns of a bike-sharing system by type of pass: a case study from Seoul’, *Transportation*, 51(4):1373–1407, doi:10.1007/s11116-023-10371-7.

Kinoshita S, Bando Y and Sayama H (2024) ‘Spatio-Temporal Differences in Bike Sharing Usage: A Tale of Six Cities’, doi:10.48550/ARXIV.2412.19294.

Kumar D (2021) ‘Meteorological barriers to bike rental demands: A case of Washington D.C. using NCA approach’, *Case Studies on Transport Policy*, 9(2):830–841, doi:10.1016/j.cstp.2021.04.002.

Laerd Statistics (n.d.) *Measures of Central Tendency,* Laerd Statistics website, accessed 5 May 2025. https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php

Laerd Statistics (n.d.) *Multiple Regression Analysis using SPSS Statistics,* Laerd Statistics website, accessed 5 May 2025. https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php

Laerd Statistics (n.d.) *Spearman's Rank-Order Correlation using SPSS Statistics,* Laerd Statistics website, accessed 5 May 2025. https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php#:~:text=The%20Spearman%20rank%2Dorder%20correlation,letter%20%CF%81%2C%20pronounced%20rho

Lee M, Hwang S, Park Y and Choi B (2022) ‘Factors affecting bike-sharing system demand by inferred trip purpose: Integration of clustering of travel patterns and geospatial data analysis’, *International Journal of Sustainable Transportation*, 16(9):847–860, doi:10.1080/15568318.2021.1943076.

Maharana K, Mondal S and Nemade B (2022) ‘A review: Data pre-processing and data augmentation techniques’, *Global Transitions Proceedings*, 3(1):91–99, doi:10.1016/j.gltp.2022.04.020.

Maioli HC, De Carvalho RC and De Medeiros DD (2019) ‘SERVBIKE: Riding customer satisfaction of bicycle sharing service’, *Sustainable Cities and Society*, 50:101680, doi:10.1016/j.scs.2019.101680.

Motte-Baumvol B, Schwanen T and Bonin O (2024) ‘Telework and the day-to-day variability of travel behaviour: The specificities of Fridays’, *Transportation Research Part D: Transport and Environment*, 132:104245, doi:10.1016/j.trd.2024.104245.

Nosal T and Miranda-Moreno LF (2014) ‘The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts’, *Transportation Research Part A: Policy and Practice*, 66:213–225, doi:10.1016/j.tra.2014.04.012.

O’Brien O, Cheshire J and Batty M (2014) ‘Mining bicycle sharing data for generating insights into sustainable transport systems’, *Journal of Transport Geography*, 34:262–273, doi:10.1016/j.jtrangeo.2013.06.007.

Rhemtulla M, Brosseau-Liard PÉ and Savalei V (2012) ‘When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions.’, *Psychological Methods*, 17(3):354–373, doi:10.1037/a0029315.

Shaheen SA, Guzman S and Zhang H (2010) ‘Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future’, *Transportation Research Record: Journal of the Transportation Research Board*, 2143(1):159–167, doi:10.3141/2143-20.

Shi Y, Zhang L, Lu S and Liu Q (2023) ‘Short-Term Demand Prediction of Shared Bikes Based on LSTM Network’, *Electronics*, 12(6):1381, doi:10.3390/electronics12061381.

Thomas T, Jaarsma R and Tutert B (2013) ‘Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling’, *Transportation*, 40(1):1–22, doi:10.1007/s11116-012-9398-5.

Tucker P and Gilliland J (2007) ‘The effect of season and weather on physical activity: A systematic review’, *Public Health*, 121(12):909–922, doi:10.1016/j.puhe.2007.04.009.

UCI Irvins (2013) *Bike Sharing,* UCI Irvins website, accessed 5 May 2025. https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset

Wang H and Noland RB (2021) ‘Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City’, *Transport Policy*, 106:262–270, doi:10.1016/j.tranpol.2021.04.004.

Wang J and Lindsey G (2019) ‘Neighborhood socio-demographic characteristics and bike share member patterns of use’, *Journal of Transport Geography*, 79:102475, doi:10.1016/j.jtrangeo.2019.102475.

Wang K, Yan X, Zhu Z and Chen X (Michael) (2024) ‘Understanding bike-sharing usage patterns of members and casual users: A case study in New York City’, *Travel Behaviour and Society*, 36:100793, doi:10.1016/j.tbs.2024.100793.

Willberg E, Salonen M and Toivonen T (2021) ‘What do trip data reveal about bike-sharing system users?’, *Journal of Transport Geography*, 91:102971, doi:10.1016/j.jtrangeo.2021.102971.

Wöhner F (2022) ‘Work flexibly, travel less? The impact of telework and flextime on mobility behavior in Switzerland’, *Journal of Transport Geography*, 102:103390, doi:10.1016/j.jtrangeo.2022.103390.

Yeatts SD and Martin RH (2015) ‘What Is Missing From My Missing Data Plan?’, *Stroke*, 46(6), doi:10.1161/STROKEAHA.115.007984.

Zou M and Djokic SZ (2020) ‘A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements’, *Energies*, 13(16):4228, doi:10.3390/en13164228.

**APPENDIX**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Coefficient** | **t-value** | **p-value** |
| registered and hr\_cat | 0.53 | N/A | 0.00 |
| casual and hr\_cat | 0.49 | N/A | 0.00 |
| registered and weather\_sit | -0.11 | N/A | 0.00 |
| casual and weather\_sit | -0.16 | N/A | 0.00 |
| registered and season | 0.18 | N/A | 0.00 |
| casual and season | 0.18 | N/A | 0.00 |
| registered and season | -0.07 | N/A | 0.00 |
| casual and season | 0.17 | N/A | 0.00 |
| registered and hum | -0.28 | -38.33 | 0.00 |
| casual and hum | -0.30 | -45.92 | 0.00 |
| registered and raw\_felt-temp | 0.33 | 46.28 | 0.00 |
| casual and raw\_felt\_temp | 0.45 | 66.92 | 0.00 |
| registered and raw\_felt-temp | 0.08 | 10.88 | 0.00 |
| casual and raw\_felt\_temp | 0.09 | 11.97 | 0.00 |

**Appendix 1.** Table of Correlation Tests’ Results (R Script 3)