

Final Project

(Data Visualization)

(Ibrahim Khan Jabarkhail)

2023-10-25

1 Introduction

As part of the final project, I will be using the Bike Sharing Dataset. The dataset contains the hourly and daily count of rental bikes between 2011 and 2012 in Capital bike sharing system with the relevant weather and seasonal information. I found this dataset on the UCL Machine Learning Repository.

Unlike buses or subways, bike-sharing systems track the exact travel times, starting points, and destinations. This makes bike-sharing like a network of virtual sensors, helping us understand how people move around the city. By watching this data, we can likely spot major city events. In the following text, I am describing, analyzing and visualizing the dataset.

1.1 My Research Question

Below are my research questions:

What are the peak usage times for bike-sharing systems?

This question will guide an analysis focusing on understanding demand patterns, which are critical for resource allocation, pricing strategies, and operational planning.

How do weather conditions influence bike-sharing habits?

Here, I am investigating external factors affecting demand. This analysis can help in understanding risks and planning for contingencies.

Can we predict bike-sharing demand?

A predictive modeling approach could help in optimizing the bike-sharing operations, improving customer satisfaction, and driving revenue.

1.2 Description of the data

In this task, I will look at the structure of the dataset and store the data in R. The data is available in CSV form and hence I used the `read.csv` command to read the file. It is a multivariate dataset with 16 features or variables.

```

##   instant      dteday season yr mnth hr holiday weekday workingday weathersit
## 1      1 2011-01-01      1  0   1  0      0        6          0          1
## 2      2 2011-01-01      1  0   1  1      0        6          0          1
## 3      3 2011-01-01      1  0   1  2      0        6          0          1
## 4      4 2011-01-01      1  0   1  3      0        6          0          1
## 5      5 2011-01-01      1  0   1  4      0        6          0          1
## 6      6 2011-01-01      1  0   1  5      0        6          0          2
##   temp  atemp  hum  windspeed  casual  registered  cnt
## 1 0.24 0.2879 0.81    0.0000        3         13 16
## 2 0.22 0.2727 0.80    0.0000        8         32 40
## 3 0.22 0.2727 0.80    0.0000        5         27 32
## 4 0.24 0.2879 0.75    0.0000        3         10 13
## 5 0.24 0.2879 0.75    0.0000        0          1  1
## 6 0.24 0.2576 0.75    0.0896        0          1  1

```

Next, I will describe the variables in the dataset given as under:

Variable	Description
instant	Record index
dteday	Date of the rentals
season	Season (1:spring, 2:summer, 3:fall, 4:winter)
yr	Year (0: 2011, 1: 2012)
mnth	Month (1-12)
holiday	Whether day is holiday (0: no, 1: yes)
weekday	Day of the week (0-6)
workingday	Whether day is working day (0: no, 1: yes)
weathersit	Weather situation (1: clear, 2: mist, 3: light rain/snow, 4: heavy rain/snow)
temp	Normalized temperature in Celsius
atemp	Normalized feeling temperature in Celsius
hum	Normalized humidity
windspeed	Normalized wind speed
casual	Count of casual users
registered	Count of registered users
cnt	Total count of bike rentals

2 Data Visualization or Analysis

2.1 Time Series Chart of Bicycle Rented

First, I will draw a time series chart to show number of bikes rented over time. This time series chart shows the count of bike rentals over time. There are noticeable patterns and fluctuations. We observe certain peaks and troughs that may correspond with seasons, suggesting a seasonal impact on bike rentals as shown in figure 1. Moreover, there seems to be a general upward trend, indicating growing popularity or expansion of the service.

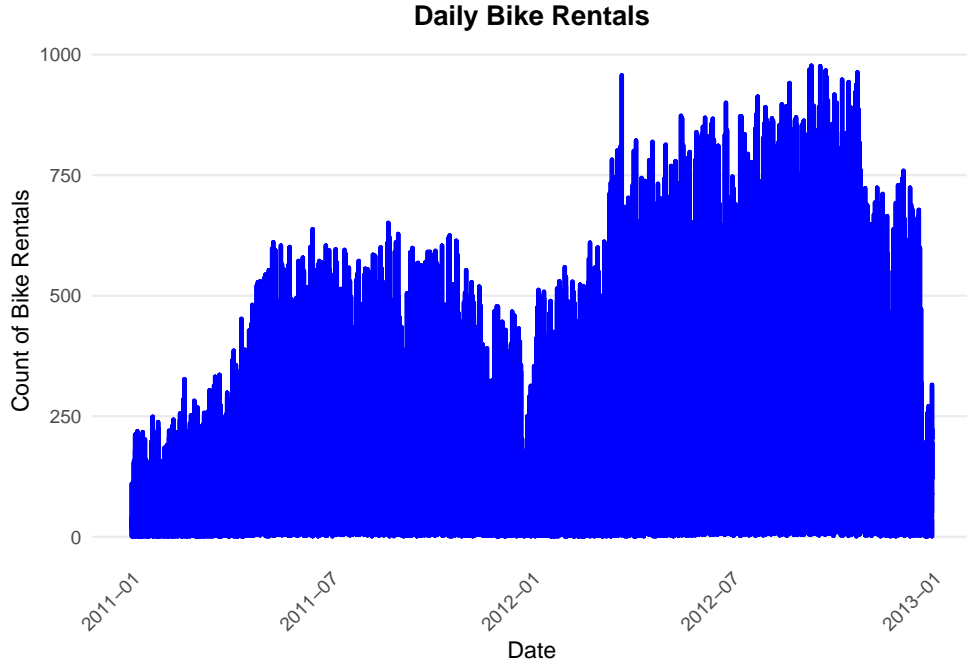


Figure 1: time series data for rental bikes

2.2 Average bike rentals per hour

tals per hour. This graph indicates the average number of bike rentals according to the hour of the day. We see significant spikes during morning and evening hours, aligning with typical commuting times. This suggests a substantial use of the bike-sharing service for commuting purposes as shown in figure 2. The relatively low usage during the late-night and early morning hours (0-5) is also consistent with what one would expect.

2.3 Seasonal breakdown of rental bike

Next, I will present the visualised pattern for each season. The box plot per season reveals a clear pattern: the highest number of rentals occurs during seasons 2 and 3, which represent summer and fall, respectively as shown in figure 3. This is logical, as the weather is generally more conducive to outdoor activities like biking during these seasons. Season 4, representing winter, shows a significant decrease in rentals, likely due to the colder weather.

2.4 Rental bikes in different weathers:

As expected, the weather has a considerable influence on bike rentals. Clear or slightly cloudy weather is ideal, showing the highest number of rentals. As the weather worsens i.e., with rain or snow, the number of rentals decreases, with particularly adverse weather leading to a significant reduction in bike usage. This pattern underscores the

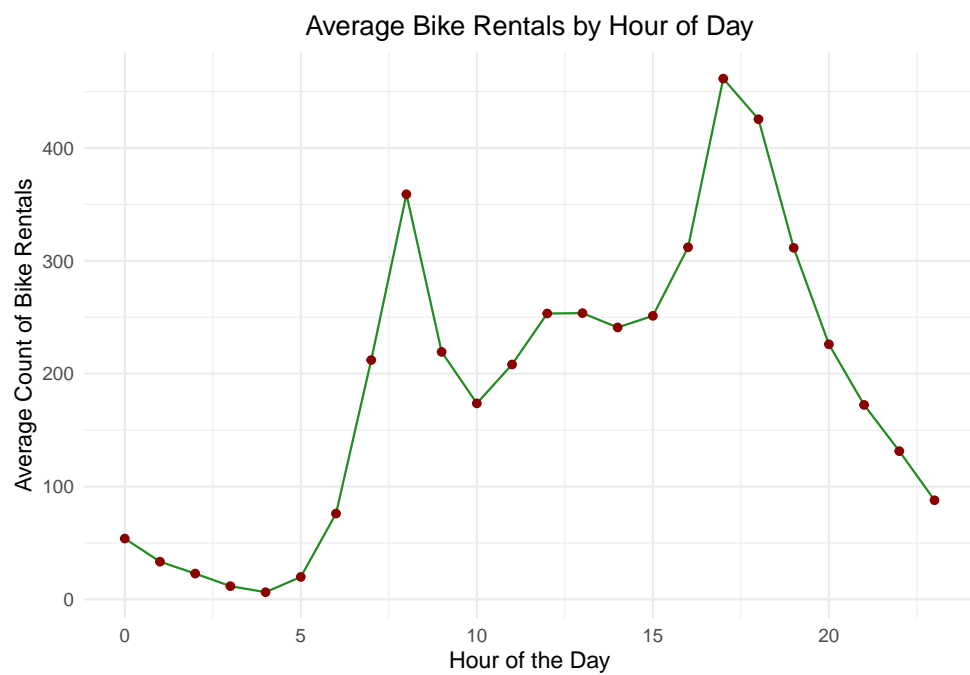


Figure 2: Average bike rentals per hour

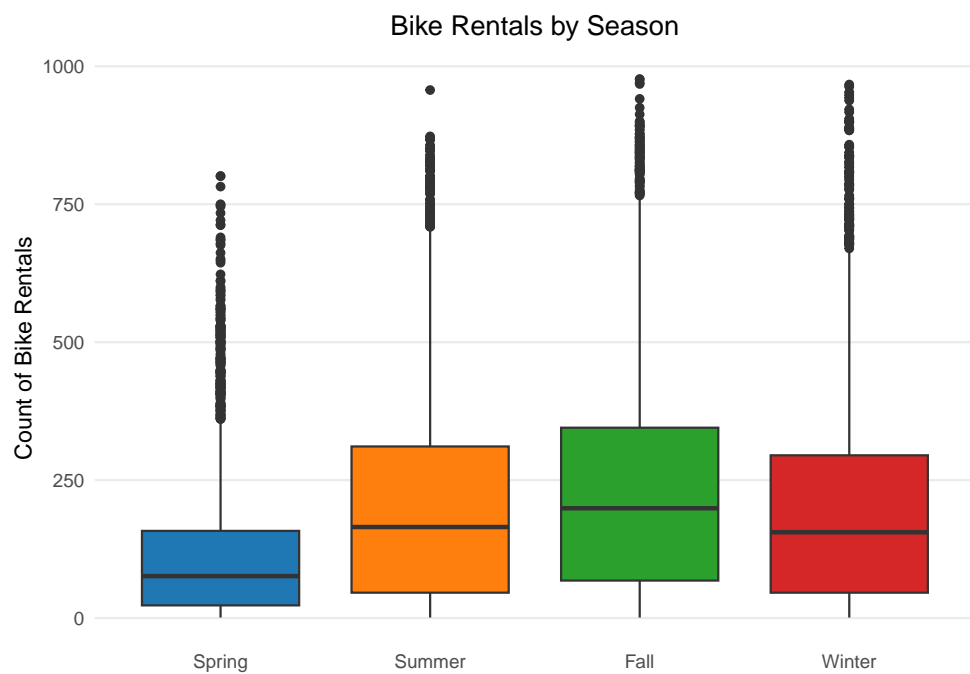


Figure 3: Bike Rentals in different seasons

importance of weather conditions in influencing consumer behavior regarding outdoor activities.

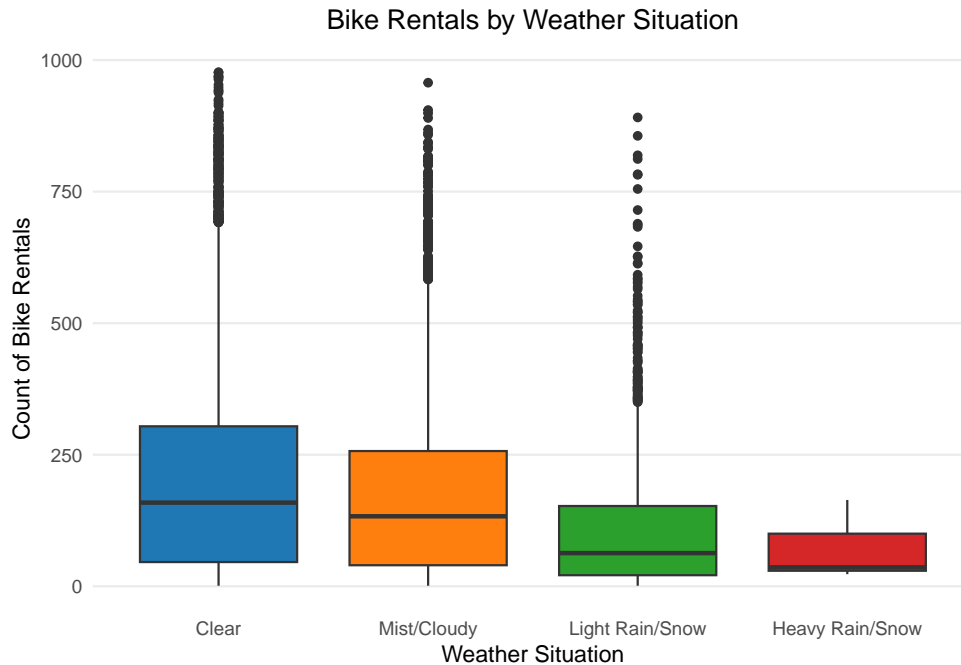


Figure 4: Bike Rentals in different weathers

2.5 On different weekdays

This bar chart contrasts the average number of bike rentals between weekdays and weekends. The rentals are marginally higher on weekends. This could be due to several factors: people may engage more in recreational activities on weekends or use bike-sharing for sightseeing and errands beyond regular commuting. The bike-sharing system, thus, serves diverse needs beyond just workday transportation.

2.6 Comparison between casual and registered bike users

This visualization presents a comparison between casual and registered users over time. Registered users consistently contribute the bulk of the bike rentals, with casual users contributing a smaller and more variable portion as shown in figure 6. This trend suggests a stable, predictable demand from registered users, possibly commuters, and a more variable demand from casual users, likely influenced by factors such as weather, tourism, and leisure activities.

2.7 Monthly breakdown of rental bike sharing

The monthly trends line graph reveals seasonal rental patterns. There's clear uptick in rentals during warmer months (May through September), coinciding with summer and

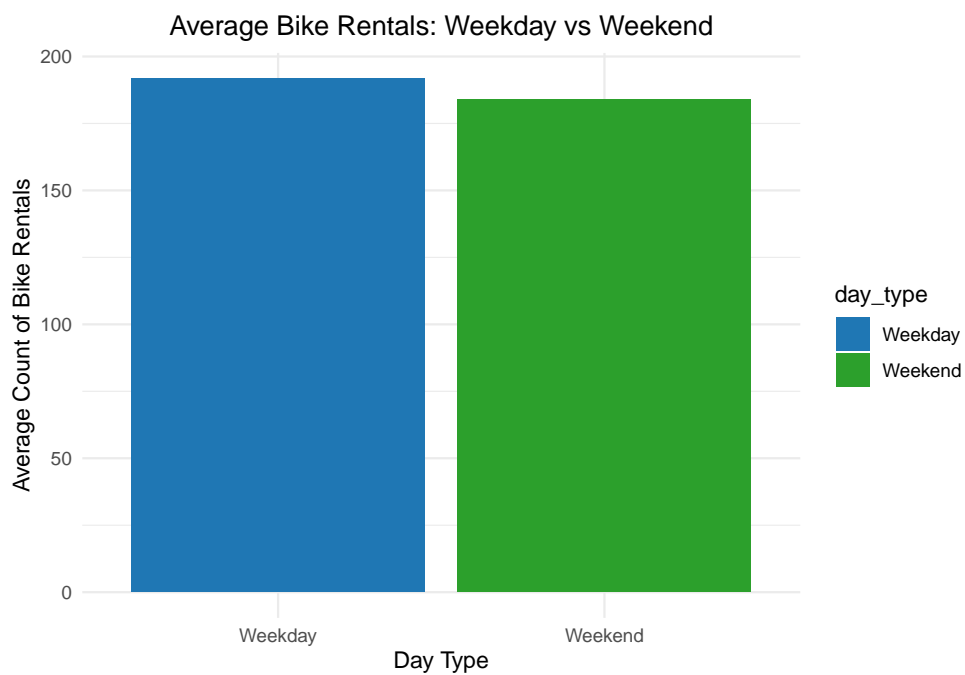


Figure 5: Bike Rentals comparison on weekend and weekdays

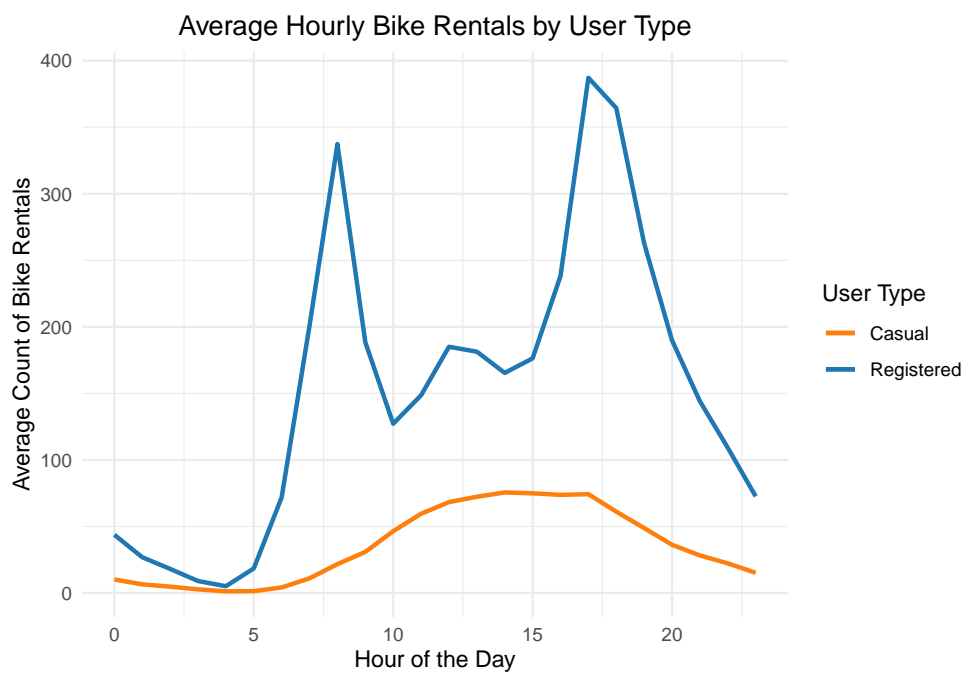
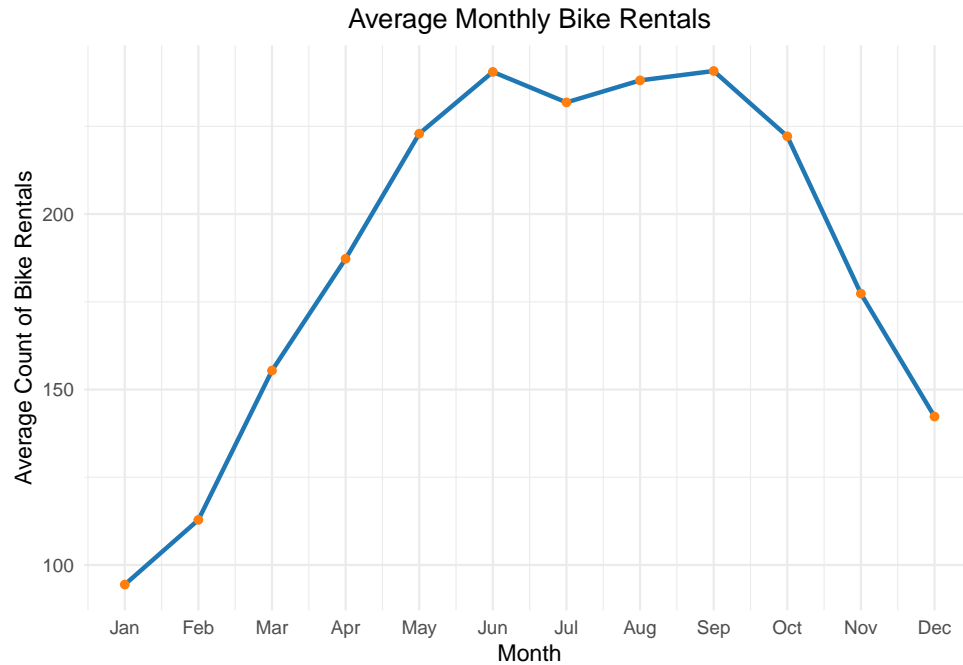


Figure 6: Bike Rentals based on those registered and casual

early fall, while the colder months see significant decrease. This seasonality suggests that weather plays a critical role in bike rental frequency. The service's popularity during the warmer season indicates a potential area for targeted marketing strategies.



3 Conclusion

The data reveals several key insights into the bike-sharing service:

Seasonal Influence: There's a strong seasonal pattern in bike rentals, with demand peaking during milder weather (summer and fall) and decreasing in harsher conditions (winter).

Daily Patterns: Bike rentals are heavily influenced by the time of day, with clear spikes during typical commuting hours. This suggests the service is not only used for leisure but significantly for commuting as well.

Weather Impact: Not surprisingly, weather conditions significantly influence rental patterns, with clear weather promoting higher usage and adverse weather leading to decreased demand.

User Types: Registered users provide a steady demand, likely representing daily commuters and consistent users, while casual usage varies more, possibly influenced by external factors such as weather and holiday seasons.