## Part 2 Exploratory Data Analysis

2.1 DC dataset

In the first part, we will explore ‘hour.csv’ file. This data set records the hourly weather conditions and the number of shared bikes used between 2011 and 2012.

2.1.1 Basic analysis

We did not find na values in the dataset, which means we can use the dataset directly for analysis.

2.1.2 How many people use bike sharing in each season?

The count of season bike using shows below:

|  |  |  |
| --- | --- | --- |
| Season Id | Season Name | Number of using |
| 1 | Spring | 471,348 |
| 2 | Summer | 918,589 |
| 3 | Autumn | 1,061,129 |
| 4 | Winter | 841,613 |

A screenshot of a social media post

Description automatically generated

As shown in the figure, most people using shared bicycles in autumn, followed by summer, winter, and spring.

2.1.3 How many people use bike sharing on holidays and non-holidays?

We use averages for comparison, because the number of holidays and non-holidays is different.

The count of holidays and non-holidays bike using shows below:

|  |  |  |
| --- | --- | --- |
| Holiday variable | Whether it is a holiday | Number of using |
| 0 | Non-holiday | 190.42858 |
| 1 | Holiday | 156.87000 |

A screenshot of a cell phone

Description automatically generated

As shown in the figure, the average number of bikes used by people during holidays and non-holidays is very close, indicating that people will often use bikes during holidays or non-holidays.

2.1.4 How many people use bike sharing on different weather type?

The count of different weather bike using shows below:

|  |  |  |
| --- | --- | --- |
| Weather id | Weather type | Number of using |
| 1 | Clear, Few clouds, Partly cloudy, Partly cloudy | 2,338,173 |
| 2 | Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist | 795,952 |
| 3 | Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds | 158,331 |
| 4 | Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog | 223 |

A screenshot of a cell phone

Description automatically generated

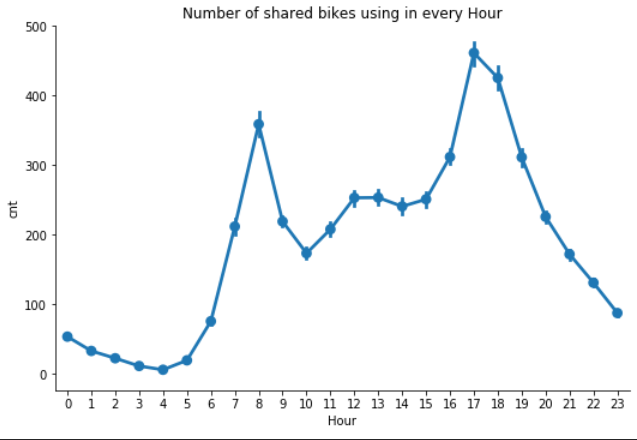
As shown in the figure, the better the weather, the more people use it.

2.1.5 How many people use bike sharing in each hour?

A screenshot of a cell phone

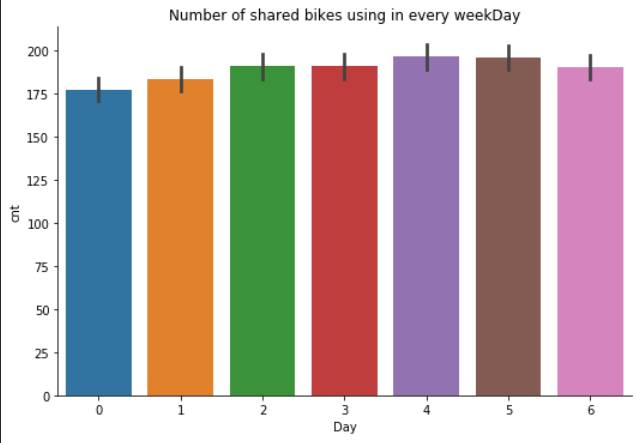
Description automatically generated

As shown in the figure, in daily life, the peaks of people's use of shared bicycles are 8 AM and 5-6 PM, with 261,001 and 336,860 to 309,772 people, respectively.



We can better observe this phenomenon in the line chart.

2.1.6 How many people use bike sharing in each weekday?

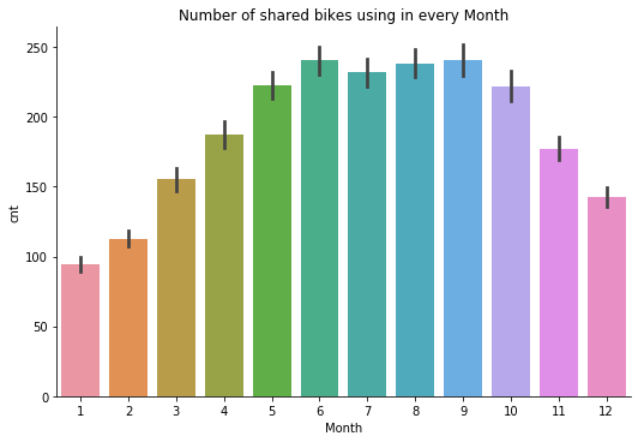


As shown in the figure, the number of shared bicycles used by people is very close every day, indicating that in daily life, shared bicycles are used every day.

2.1.7 How many people use bike sharing in each month?

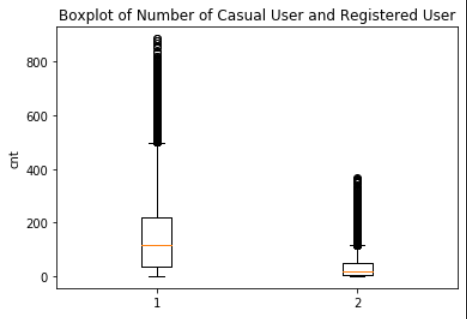
The count of month bike using shows below:

|  |  |  |  |
| --- | --- | --- | --- |
| Month | Count | Month | Count |
| 1 | 134,933 | 7 | 344,948 |
| 2 | 151,352 | 8 | 351,194 |
| 3 | 228,920 | 9 | 345,991 |
| 4 | 269,094 | 10 | 322,352 |
| 5 | 331,686 | 11 | 254,831 |
| 6 | 346,342 | 12 | 211,036 |



As shown in the figure , the peak period of bicycle sharing is concentrated in June to September.

2.1.7 Observe registered users and occasional registered users



A picture containing screenshot

Description automatically generated

1 for registered users, 2 for accidental registered users

It can be seen from the box and violin charts that the number of registered users is far more than the number of accidental registrations, indicating that many people often use shared bicycles.

2.1.8 How many people use bike sharing on different weather condition?

A screenshot of a video game

Description automatically generated

After comprehensively comparing the analysis results of weather conditions, we can conclude that the most people use bicycle sharing when the temperature is high, the more people use it when the wind speed is low, and the most people use it when the humidity is high. But when these factors reach extreme values, the number of sharing bicycles people used is the least.

2.1.9 correlation matrix of DC dataset

A close up of a logo

Description automatically generated

Through this figure, we use appropriate features for modeling

2.2 DC2017 dataset

In the first part, we will explore ‘hour2017.csv’ file. This data set records the hourly weather conditions and the number of shared bikes used in 2017.

2.2.1 How many people use bike sharing in each hour?

A screenshot of a cell phone

Description automatically generated

As shown in the figure, in daily life, the peaks of people's use of shared bicycles are 8 AM and 5 PM, with 299,108and 415,574 people, which is very similar to data recorded in 2011 and 2012.

A close up of a map

Description automatically generated

We can better observe this phenomenon in the line chart.

2.2.2 How many people use bike sharing on different weather condition?

A screenshot of a cell phone

Description automatically generated

In the 2017 record, the impact of weather conditions on the use of shared bicycles has not changed much. Compared with the changes in 2011-2012, people will use shared bicycles more when the wind speed is moderate.