Bike-Share Usage in London Network Analysis

This manuscript (permalink) was automatically generated from uiceds/cee-492-term-project-fall-2022-jiaotonguniv@ee10f0e on October 30, 2022.

Authors

- **Mulin Wan**
  · mulin-wan
  CEE, University of Illinois Urbana-Champaign

- **Jingwen Yao**
  · jingwenyaooo · Yaojune
  CEE, University of Illinois Urbana-Champaign

- **Yunze Guo**
  · cyfcx2
  CEE, University of Illinois Urbana-Champaign

- **Bo-Yang Wang**
  · byw-5
  CEE, University of Illinois Urbana-Champaign

Abstract

1.0 Introduction

Description

1.1 Dataset:

For this project, we are going to focus on the usage of bike sharing in London. The data come from two datasets, "London and Taipei Bike-Share Data" and "London bike sharing dataset."

1.1.1 London and Taipei Bike-Share Data

The raw data is collected from the respective cities open data sites.

London

The data has been reformatted into CSV in order to be easier to use and compare. The content remains unchanged.

This dataset would be the main dataset for the project as it contains every single bike rental in London in the duration. This gives the columns in the data comprise:

Table 1: Discription of London and Taipei Bike-Share

Object	Description

Object	Description
rental_id	id of people who rent the bike
duration	duration of rental
bike_id	id of bike
end_rental_date_time	date and time of end rental
end_station_id	id of end station
end_station_name	name of end station
start_rental_date_time	date and time of start rental
start_station_id	id of end station
start_station_name	name of start station
start_rental_date_time	date and time of start rental

1.1.2 London bike sharing dataset

This dataset is playing a more supporting role, as it helped providing informations on weather conditions. Although the timespan doesn't overlap too much with the first dataset (January 4th 2015 to January 3rd 2017), it gives the idea to incorporate weather conditions into consideration. In the future we may try to find weather datasets that fit the first dataset better to help analyzing.

The data has been formatted into CSV in order to be easier to use and compare. This gives The columns in the data comprise:

Table 2: Description of London Bike-Share

Object	Description
timestamp	timestamp field for grouping the data
cnt	the count of a new bike shares
t1	real temperature in Celsius
t2	temperature in Celsius "feels like"
hum	humidity in percentage
windspeed	wind speed in km/h
isholiday	boolean field - 1 holiday / 0 non holiday - refers to bank holidays
isweekend	boolean field - 1 if the day is weekend / 0 if a working day
season	category (0-spring; 1-summer; 2-autumn; 3-winter)
weathercode	different weather condition

Table 3: Description of weathercode

weathercode	Description
1	clear; mostly clear but have some values with haze/fog
2	scattered clouds / few clouds

weathercode	Description
3	broken clouds
4	clear; cloudy
7	clear; light rain shower / rain / light rain
10	clear; rain with thunderstorm
26	snowfall
90	freezing fog

In addition to season and isweekend, from the timestamp feature we can extract many separate time features - day of the week (as one scaled column or as seven columns of ismonday, istuesday etc.), month number, day of the month, week number, hour, minute. In combination with external data, we could add is_light for after dawn times and is_schoolholiday to match London school holiday times.

1.1.3 Link of dataset:

[London and Taipei bikeshare]

[London bikeshare]

1.2 Proposal "SPECIFIC"

Recently, bike-sharing in big cities has become an important part of residents' daily life, and its role in urban transportation system has never been more significant. Looking at the spatiotemporal bike-sharing data in London, we could explore patterns, describe variations, or model the data in many different ways. From the two datasets, we may have a chance to take a peek at the residents' bike-renting behavior through many angles.

"PREVIOUS" work has shown that weather is a key driver for variation in usage. [1] [2] By utilizing those datasets to analyze how extreme weather events like winter rains in London will affect bike-sharing system, it is safe to draw some conclusions to guide the process of making contingency plans. The locations of start-trip and end-trip is also considered to have the potential of revealing hot spots of bike-renting usage.

The result will be able to offer some suggestions for the decision maker of bike-sharing companies about the arrangement of bike density in different blocks, distribution between urban and rural areas and methods to tackle extreme weather conditions.

2.0 Exploratory Data Analysis

In this section, most of factors attaches a bar chart indicating the distribution of the new bike sharing system in London. In order to achieve the goals of analysing the influencial factors on the new bike share distribution, applying the process of Exploratory Data Analysis (EDA) is a way to develop a better understanding of this dataset and promote the project.

(Hoping there will be more discription and introduction)

2.1 Data Wrangling

(delete me if u think there is no need for more discription, But for me, I guess there could have some intro or description...)

2.1.1 Different factors on the count of new bike shares

There are different factors in the database such as time, windspeed, season, and so on. The group is interested in these arguments and try to find the impact of them on the count of new bike shares(cnt).

Data cleaning process

The csv file with the London bike sharing dataset stored in is generic and need to be restructured before it can be used to anlysis its raw forms which are those variable connecting to factors influencing the number of new bike shares and its distribution.

For the dataset of London Bike Share, it contains 17414 variables (columns) and 10 factors in its original form. Due to the factor called timestamp in original form is hard to analysis the bike share distribution with time.So shift it into five more specific factors into its final form. Factors went from 10 to 14 which as shown in the Table 4. Besides, to fully analysis the distribution of bike share happened in London, using the final form of datasets to arrange 7 subtables using for analysis and visualization.

Table 4: Description of New London Bike-Share form

Object	Description
cnt	the count of a new bike shares
t1	real temperature in Celsius
t2	temperature in Celsius “feels like”
hum	humidity in percentage
windspeed	wind speed in km/h
weathercode	different weather condition
isholiday	boolean field - 1 holiday / 0 non holiday - refers to bank holidays
isweekend	boolean field - 1 if the day is weekend / 0 if a working day
season	category (0-spring; 1-summer; 2-autumn; 3-winter)
date	the date of the bike share happens in the data form of Dates
year	the year of the bike share happens in the data form of Int
hour	the specific time to be observed of the bike share happens in the data form of Time
dayofweek	distiguis the date of the bike share happens in the day of the week
month	the month of the bike share happens in the data form of Int

(delete me except u think those below subtables is useful for our documents content, as they print here is to help teamates have better understanding of this documents)

	hour	cnt_mean
1	00:00:00	215.367
2	01:00:00	128.002
3	02:00:00	81.0077
4	03:00:00	53.25
5	04:00:00	52.2481
6	05:00:00	116.052
7	06:00:00	604.655
8	07:00:00	1969.52
9	08:00:00	3863.92
10	09:00:00	1997.61
⋮ more		
24	23:00:00	434.894

	hour	cnt_mean
1	00:00:00	479.811
2	01:00:00	383.262
3	02:00:00	276.441
4	03:00:00	197.434
5	04:00:00	126.337
6	05:00:00	97.2537
7	06:00:00	120.556
8	07:00:00	213.184
9	08:00:00	415.786
10	09:00:00	787.391
⋮ more		
24	23:00:00	451.654

mean cnt of general weekdays with the distribution of hour

mean cnt of general weekends with the distribution of hour

	hour	Monday	Tuesday	Wednesday	Thursday	Friday
1	00:00:00	212.471	168.381	203.117	212.952	281.52
2	01:00:00	119.663	101.952	116.816	126.269	176.382
3	02:00:00	74.6058	65.4038	68.3204	83.5673	113.647
4	03:00:00	49.2404	42.3558	45.9223	53.8173	75.4851
5	04:00:00	52.3558	46.5769	47.4078	51.8365	63.3366
6	05:00:00	112.221	119.0	115.796	117.615	115.614
7	06:00:00	536.808	657.267	639.106	645.817	542.578
8	07:00:00	1759.97	2128.92	2098.15	2102.15	1752.61
9	08:00:00	3504.8	4098.2	4045.17	4094.26	3572.18
10	09:00:00	1802.52	2004.48	2042.6	2144.37	1995.84
⋮ more						
24	23:00:00	312.219	378.214	1960.66	497.408	555.51

mean cnt of day of week in weekdays with the distribution of hour

	hour	Saturdau	Sunday
1	00:00:00	460.118	499.125
2	01:00:00	355.735	410.26
3	02:00:00	255.267	297.204
4	03:00:00	180.178	214.192
5	04:00:00	121.594	130.942
6	05:00:00	93.1089	101.279
7	06:00:00	123.825	117.317
8	07:00:00	248.835	177.875
9	08:00:00	493.314	339.75
10	09:00:00	873.184	702.423
⋮ more			
24	23:00:00	566.515	335.667

mean cnt of day of week in weekends with the distribution of hour

	hour	cnt_mean_in_holiday	cnt_mean_not_in_holiday
1	00:00:00	347.812	289.316
2	01:00:00	238.75	199.77
3	02:00:00	155.812	135.861
4	03:00:00	115.75	93.7574
5	04:00:00	87.375	72.9943
6	05:00:00	75.4375	111.508
7	06:00:00	104.75	474.782
8	07:00:00	170.25	1498.01
9	08:00:00	301.25	2941.16
10	09:00:00	570.75	1677.37
⋮ more			
24	23:00:00	290.75	443.027

mean cnt of wether in holidays with the distribution of hour

	hour	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12
1	00:00:00	187.855	181.393	190.452	229.4	322.323	371.949	446.508	421.194	336.825	317.984	222.339	253.419
2	01:00:00	131.161	119.107	119.803	150.85	213.419	243.78	303.823	286.0	232.596	244.048	161.949	193.032
3	02:00:00	88.8226	83.4286	87.7288	97.1833	146.984	162.492	206.29	191.694	157.895	161.855	117.051	127.426
4	03:00:00	64.6935	60.8214	64.95	70.35	101.565	113.586	136.823	135.484	108.193	108.726	78.3729	83.8689
5	04:00:00	52.2419	48.5893	52.6833	57.4333	77.9032	89.6207	103.306	103.919	82.386	83.0	62.7458	63.4754
6	05:00:00	79.5323	81.3393	85.1667	91.4333	116.645	135.793	151.516	141.823	130.263	119.806	105.0	88.4426
7	06:00:00	358.145	377.877	406.918	450.6	483.097	545.644	597.377	555.161	559.897	488.484	449.283	329.79
8	07:00:00	1143.84	1256.19	1349.82	1469.03	1485.44	1704.27	1767.82	1600.84	1776.0	1570.97	1474.02	1040.69
9	08:00:00	2420.06	2664.18	2656.07	2812.8	2808.61	3260.92	3256.53	3014.34	3369.62	3161.35	3006.48	2190.52
10	09:00:00	1327.29	1430.16	1450.95	1568.87	1678.85	1899.24	1983.68	1790.53	1882.97	1795.37	1656.83	1374.42
⋮ more													
24	23:00:00	257.339	274.321	296.77	360.717	476.919	567.717	704.032	628.803	509.491	465.583	346.373	376.048

mean cnt of different month with the distribution of hour

	month	cnt_mean
1	1	784.692
2	2	836.28
3	3	921.295
4	4	1112.95
5	5	1275.1
6	6	1383.35
7	7	1543.7
8	8	1463.12
9	9	1356.6
10	10	1217.55
11	11	965.868
12	12	845.516

mean cnt with the distribution of season

(delete me except u think those above subtables is useful for our documents content, as they print here is to help teamates have better understanding of this documents)

2.2 Analysis and Visualization

2.2.1 London Bike Share

In this section, adapt mean cnt (the count of a new bike shares) to analysis how the new bike share demand as this mean cnt could better present the demand levels as possible.

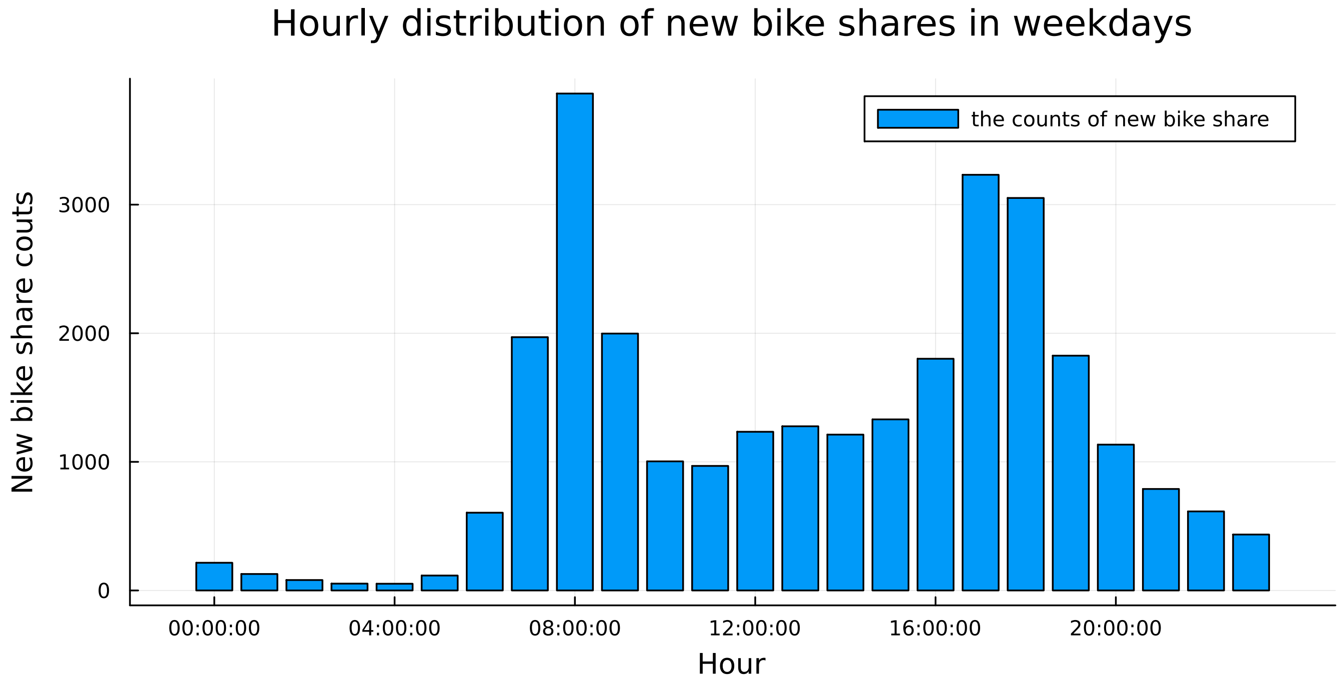


Figure1: mean cnt of general weekdays with the distribution of hour

This figure shows how the new bike share demand distribute in different hours in a weekday. According to the image, it can be clearly concluded that during the working day, the change of demand for borrowed cars shows a double-peaked image during the day, and the peak period of commuting can greatly affect the number of new borrowed cars, and the maximum peak of borrowed cars in a day appears around 8:30 which is higher than3000, and the minimum peak appears around 3:30 which is alomost 0.

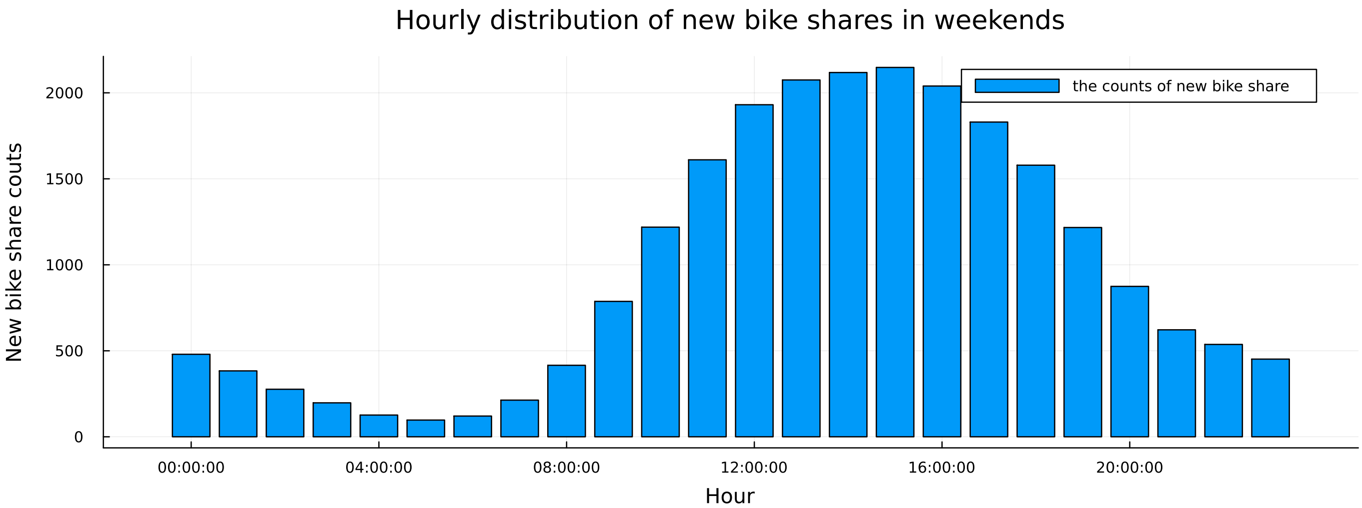


Figure2: mean cnt of general weekends with the distribution of hour

This figure shows how the new bike share demand distribute in different hours in a weekend day. Based on the images, it is clear that during the weekend, the demand for rental cars varies in a single wave throughout the day, with the demand for travel in London mainly between 11:00 and 20:00, peaking around 15:00 and reaching a

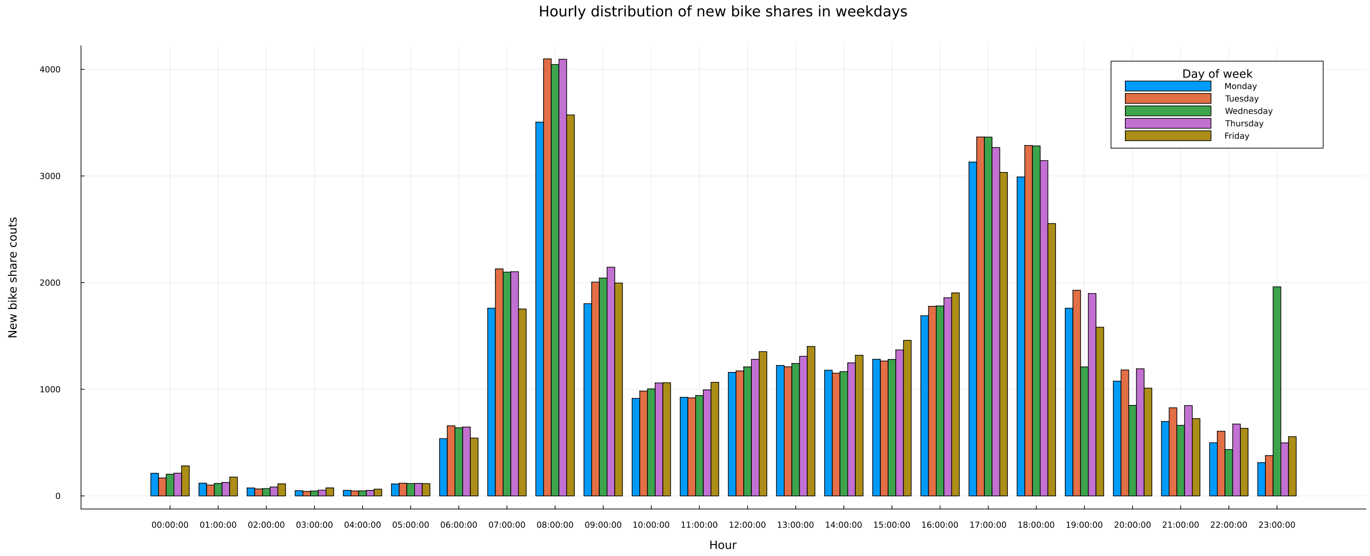


Figure3: mean cnt of day of week in weekdays with the distribution of hour

In this figure, the overall trend is consistent with the trend in Figure1, but according to this figure, we can conclude that the overall number of new bike shares on Monday and Friday is relatively low in each time period. It is worth noting that while the maximum peak in Figure1 is greater than 3000 but less than 4000, in this figure Tuesday, Wednesday and Thursday not only reach 4000, but also Tuesday and Thursday are the two days with higher relative quantities. In addition, the number of new bike shares is extraordinarily high at 22:00 on Wednesday, which is not only different from the trend in Figure1, but even almost three times higher than on other weekdays at the same moment.

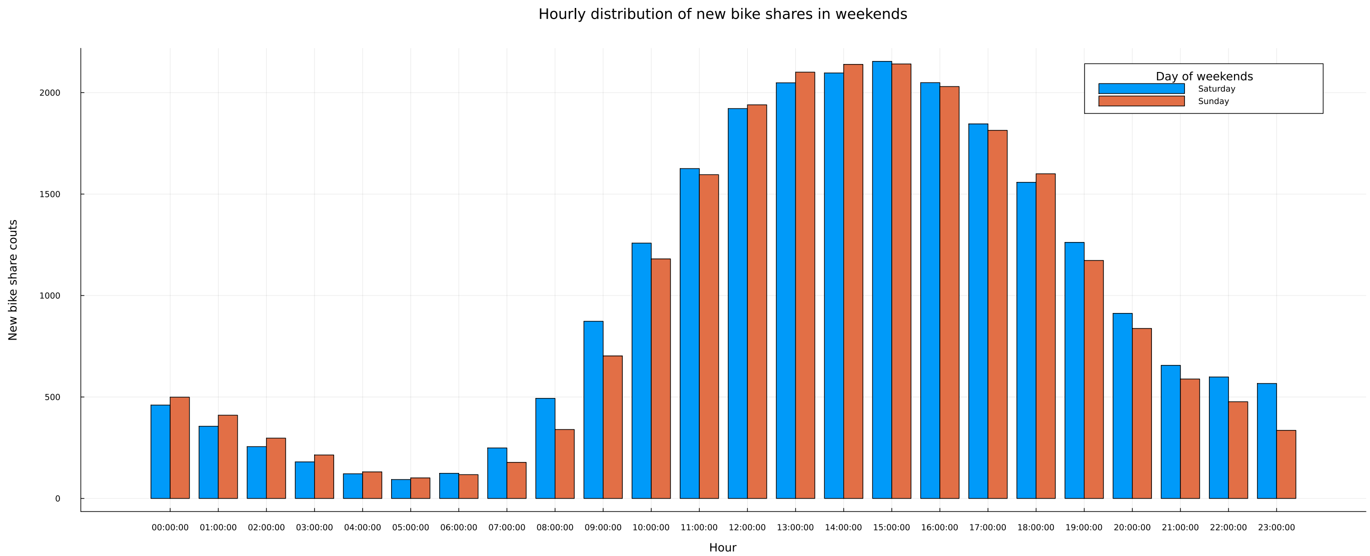


Figure4: mean cnt of day of week in weekends with the distribution of hour

The figure shows that the overall difference of new bike share between the two days of the weekend is not big, but the new bike share from 6:00 to 15:00 on Saturday is higher than that on weekdays. This phenomenon shows that although people’s demand for bike share in London is roughly the same on both days of the weekend, people’s travel demand is higher during the daytime on Saturday.

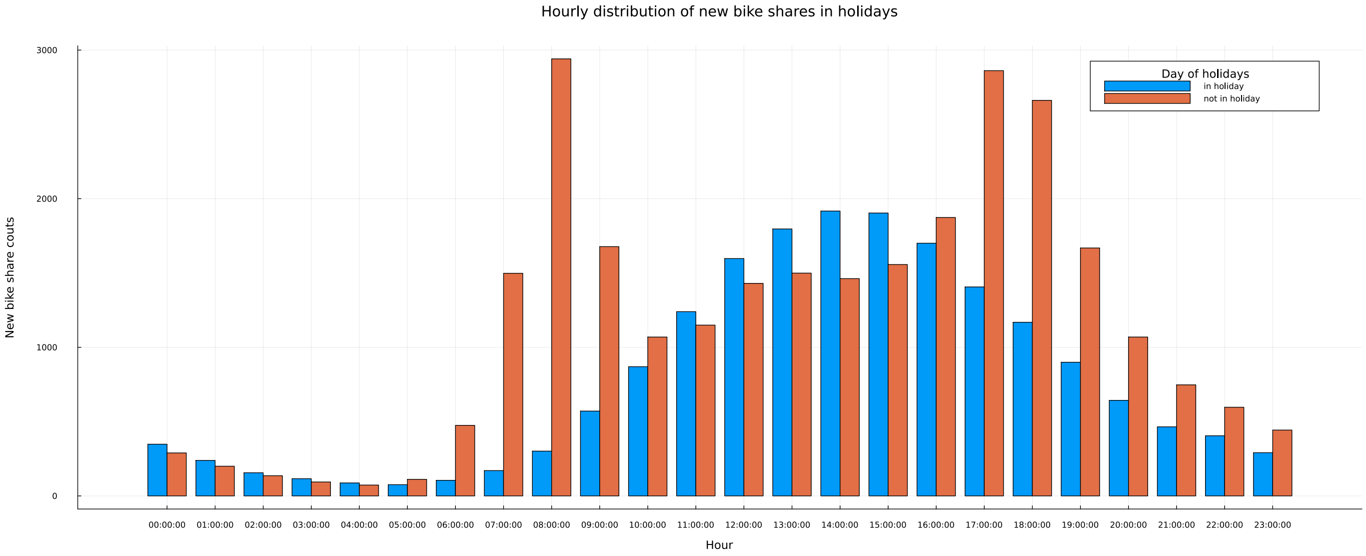


Figure5: mean cnt of in holidays with the distribution of hour

In this figure, it can be found that the change in the distribution of people's new bike share during holidays is similar to the image of the change in the distribution of people's new bike share on weekends in Figure 2, and those not during holidays are similar to the image of the change in the distribution of people's new bike share on weekdays in Figure 1. In addition, the number of people's new bike shares in London during holidays is higher than that of weekdays during midday and late afternoon. This indicates that during weekdays, people's bike share demand is influenced by work commuting, i.e., it reaches a double peak during the peak commuting period; while during holidays, people's bike share demand peaks at noon and after noon, and even during noon (mealtime) is higher than that during weekdays, which can be guessed that people tend to go out for gathering during noon on holidays.

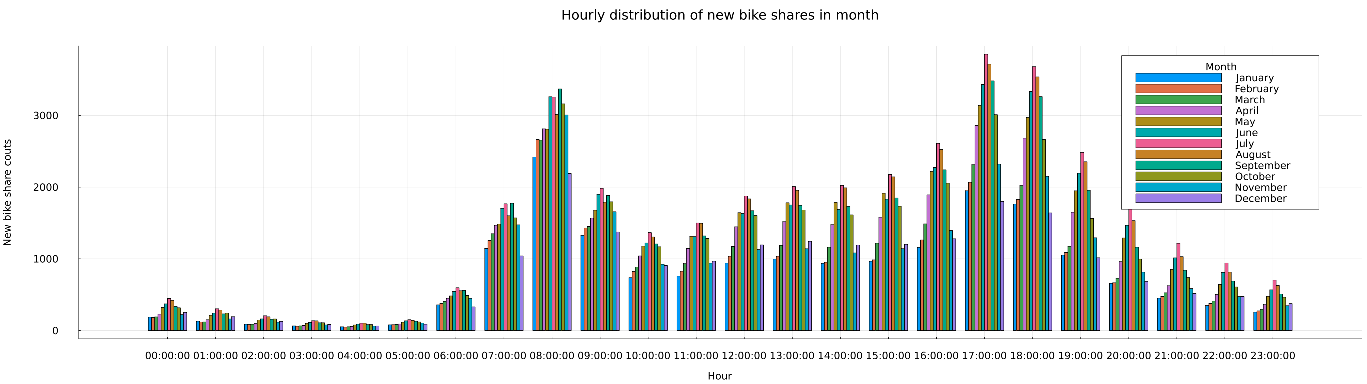


Figure6: mean cnt of each month with the distribution of hour

In this figure, the overall distribution trend is similar to that in figure1 (hour-by-hour new bike share distribution over the workday), with large differences between months, especially between the April to October period and the November to March period. In addition, it can be seen that in the first half of the year, the overall demand for new bike share from 6:00 to 17:00 in London is higher than the level in the second half of the year.

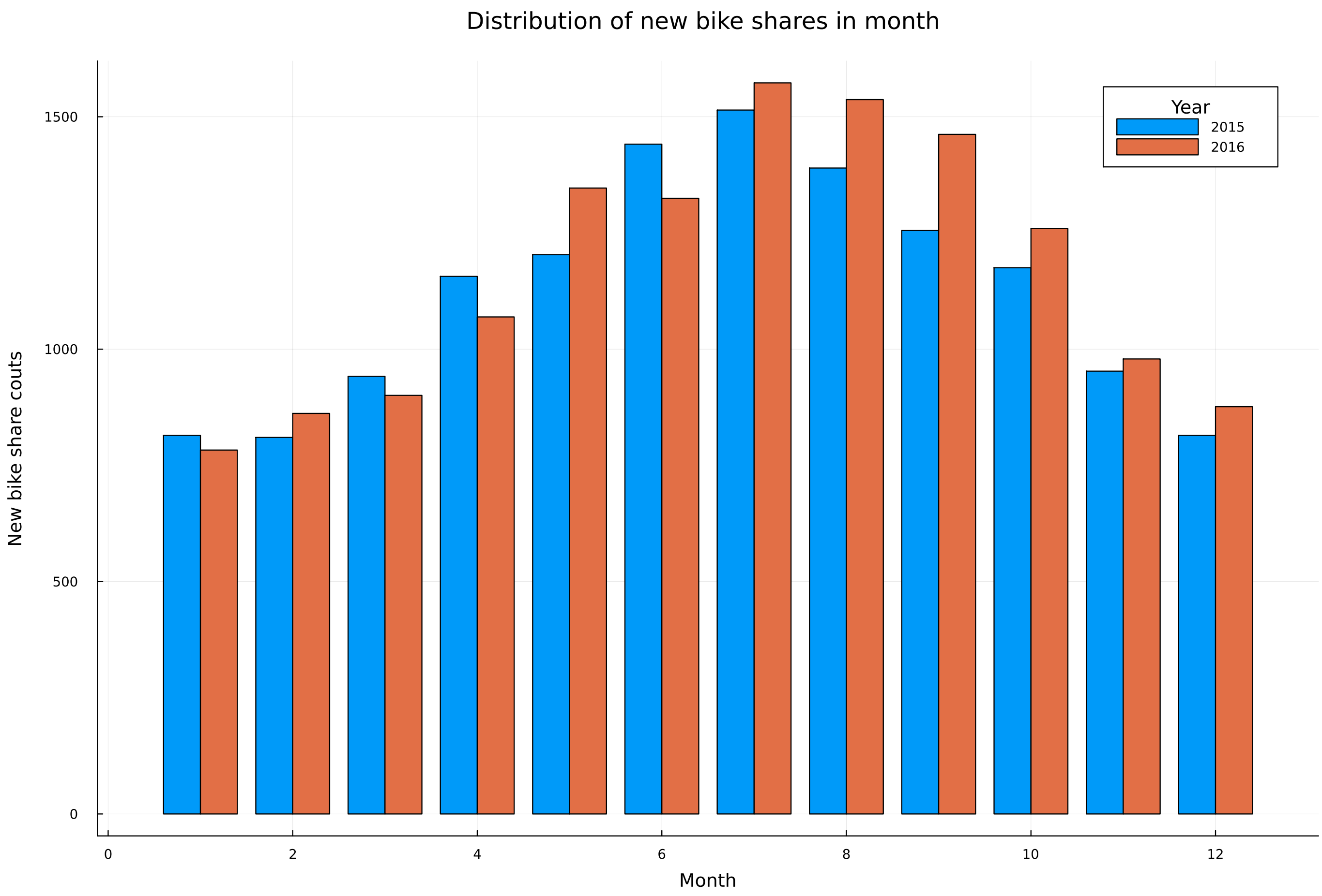


Figure7: mean cnt with the distribution of different month

According to this figure, it can be seen that among all time periods, the new bike share is relatively high from April to October, especially from June to August.

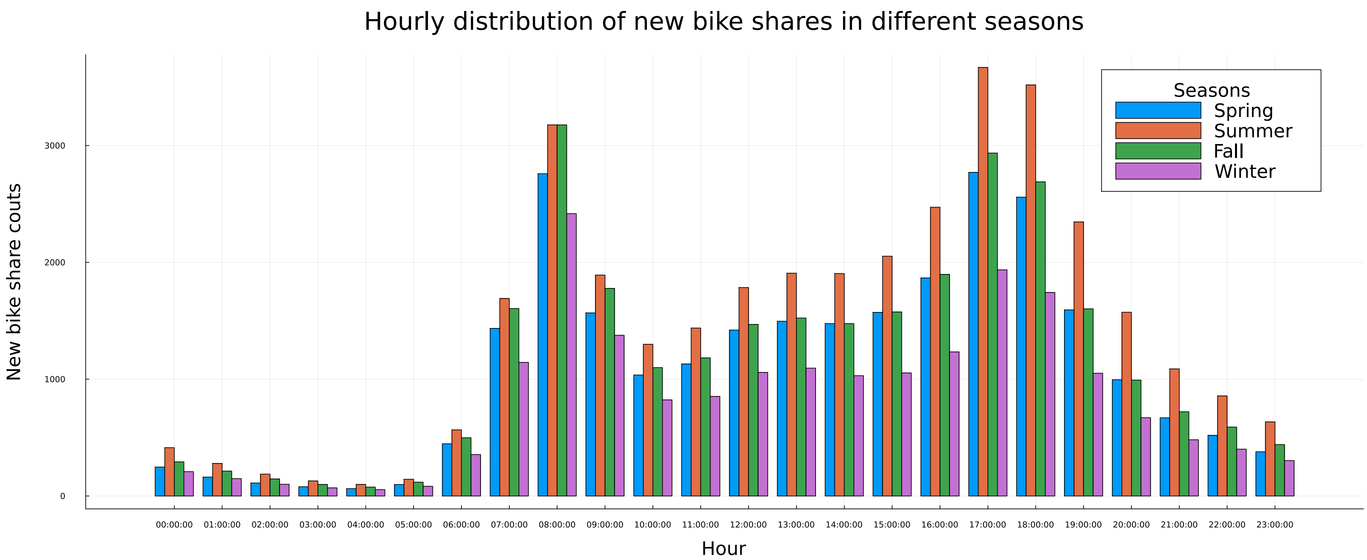


Figure8: mean cnt of each season with the distribution of hour

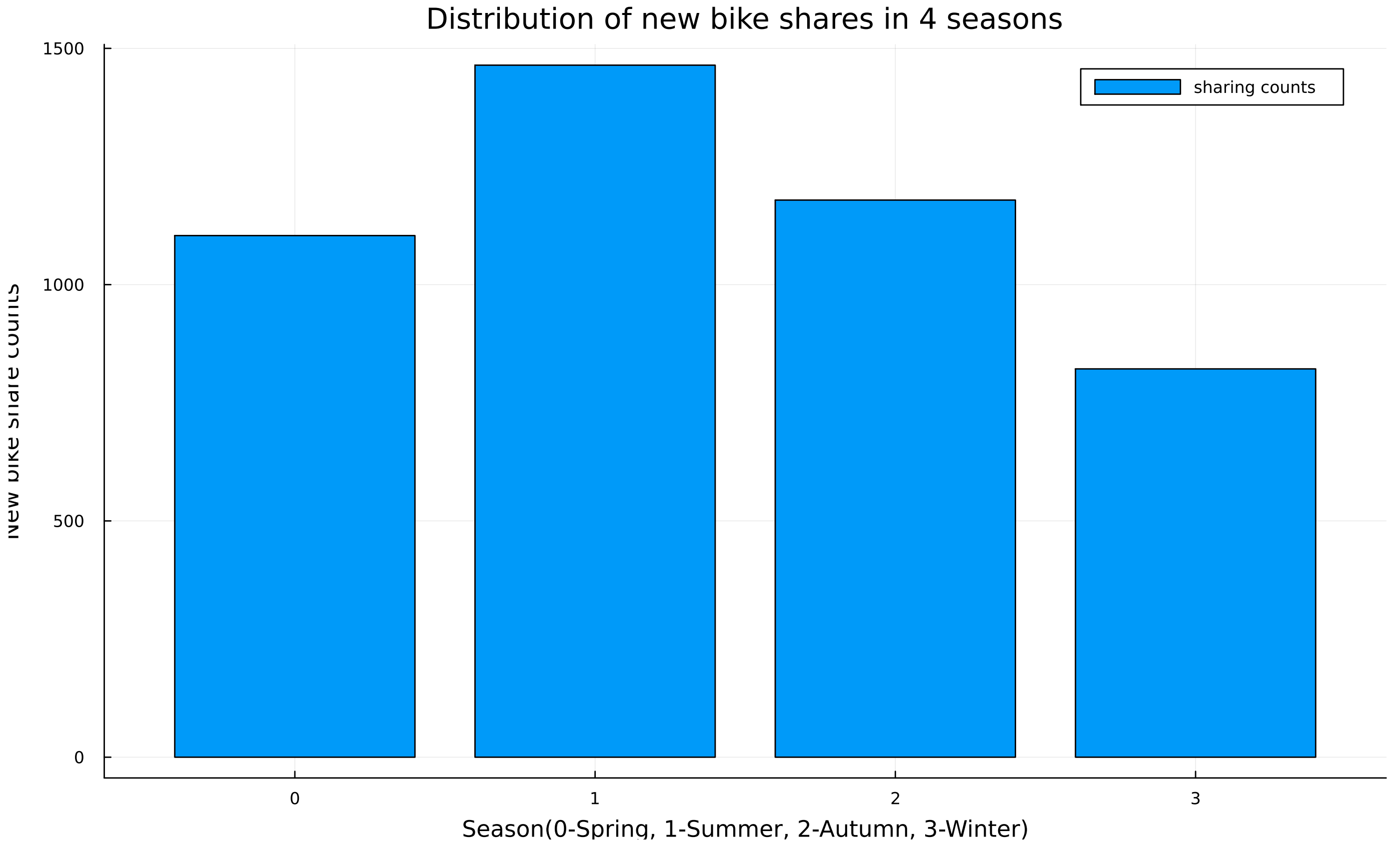


Figure9: mean cnt of each season

According to these two images, it can be seen that the demand for new bike share in London is relatively higher in summer and autumn overall, especially in summer. Winter is undoubtedly the lowest, but in this image it can be seen that the demand for new bike share is lower in spring than in autumn. It can be guessed that people are more willing to take bike share in the season of relatively higher temperature, and autumn is more suitable for bike share than spring in London area. In addition, people's bike share during the off-duty peak period is significantly higher during summer than not only during other seasons, but also during their off-duty peak period, a scenario that differs from all the weekday-related images described above.

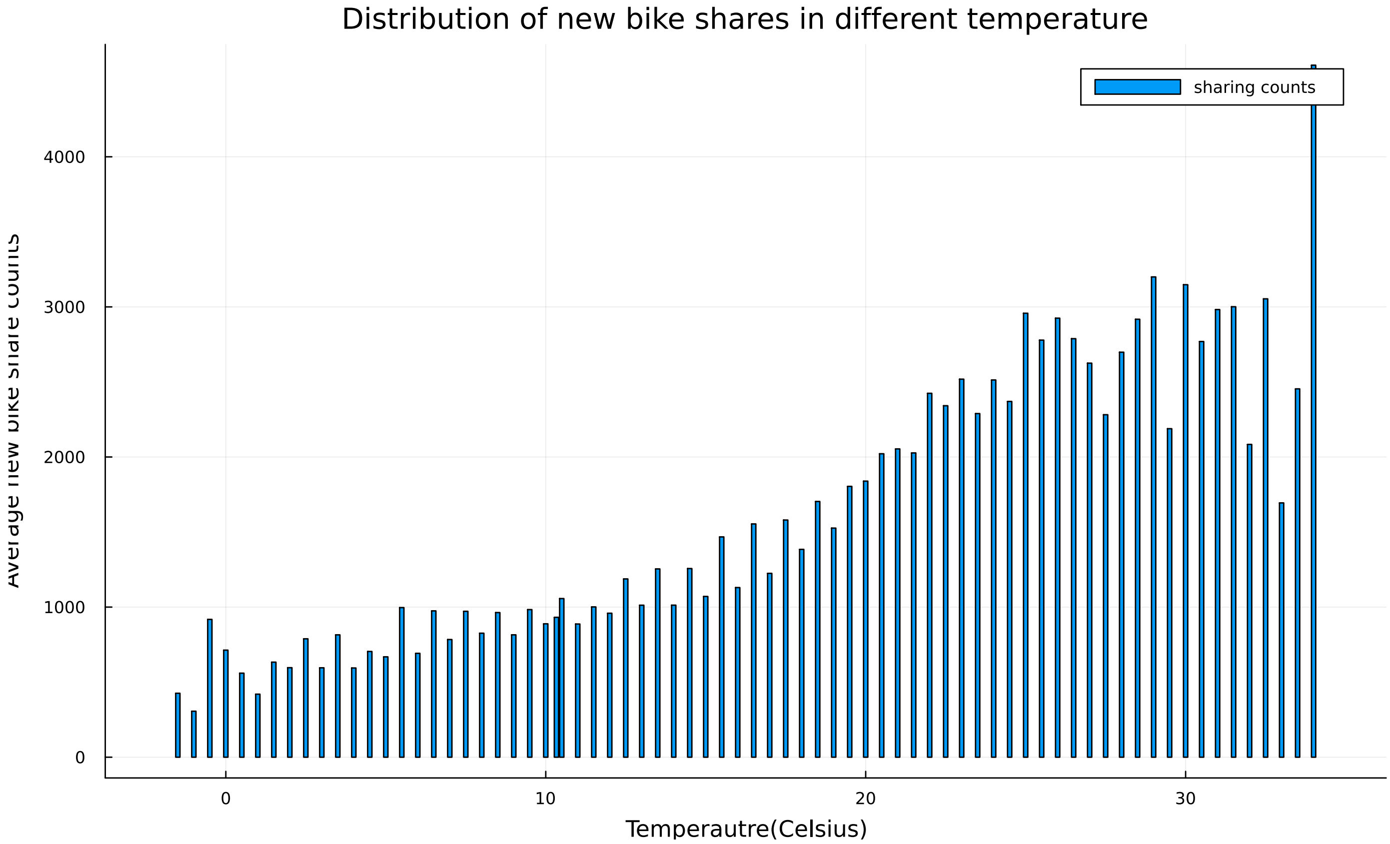


Figure10: mean cnt of each temperature

This image shows the change of new bike share in different temperaure. The number of sharing bikes increases with temperature totally. The minimum result is below zero Celsius degree and the maximum one is more than 30 Celsius degrees. However, this trend is not a strictly positive correlation. If the highest value in the plot is neglected, it can be found that the average counts comes down after the temperature is more than about 30 Celsius dergees.

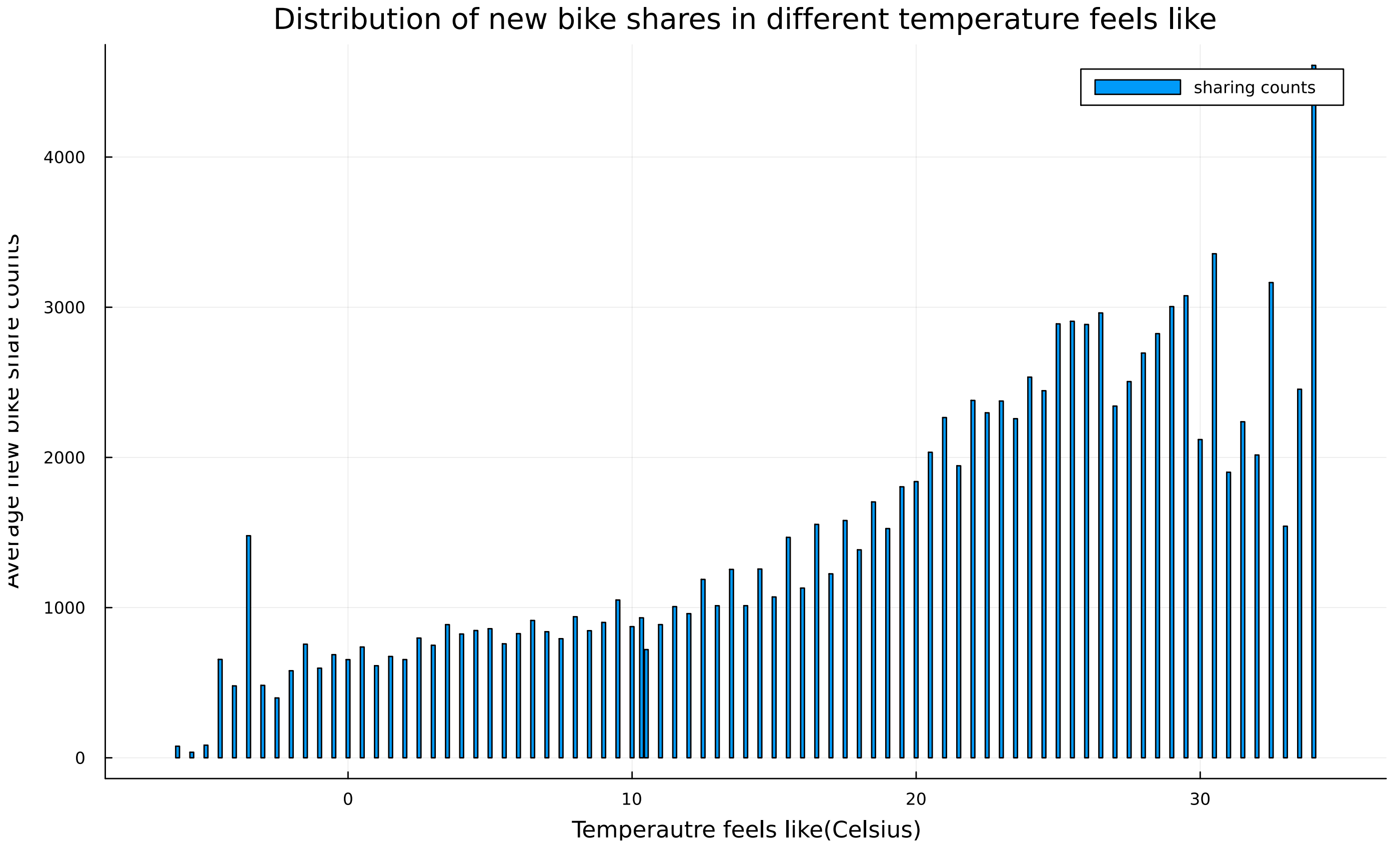


Figure11: mean cnt of each temperature “feels like”

As the dataframe offers both the data of temperature and temperature “feels like”, it is possible to compare the impact of apparent temperature. The lowest and highest apparent temperature affect more on the bike sharing counts, comparing to the real temparature. It can be concluded that users are less willing to choose riding when the temperature is too high or too low.

Figure12: mean cnt of holidays and weekdays

This figure consists of holidays and weekdays. It is obviously that the weekdays has about one third higher average bike sharing counts than holidays, which proves that users choose the sharing bike more frequently on weekdays.

Figure13: pie chart of distribution under 7 weather conditions

This pie chart clearly reveals the percentage of people utilizing bike sharing system. Basically, when the weather is good, people tend to utilize the bike sharing system more often, which is easy to understand. But after deeper investigations, evidences show that there is not a huge difference between “cloudy”, “light rain” and “thunderstorm”. It can be interpreted as when it is cloudy, people tend to feel a little bit reluctant to go outing compared with those warm and sunny days. But when those pedestrians are caught in the rain with no access to a car, they tend to have the thought of going home quickly. In this case, many of them would choose to ride a bike no matter it is “light rain” or “thunderstorm”.

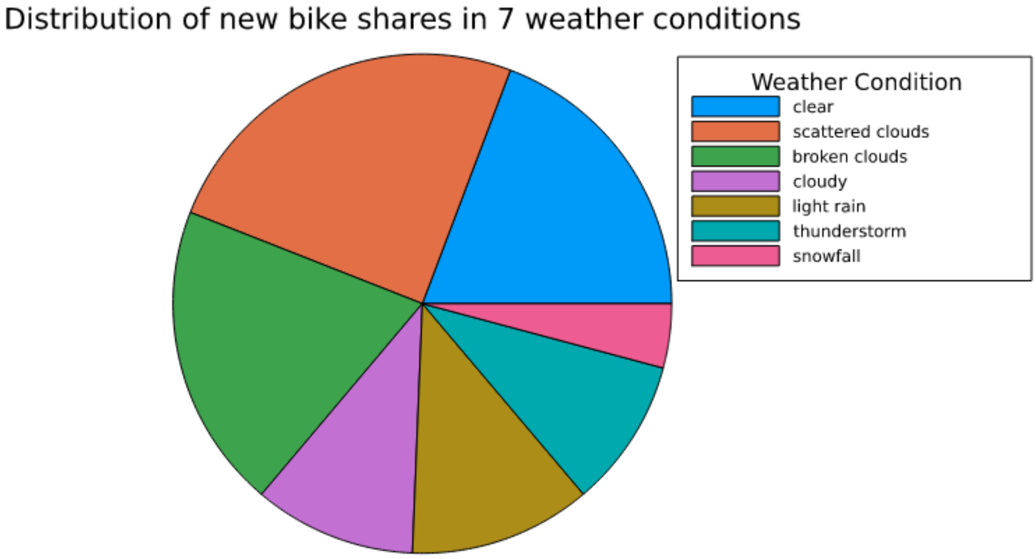
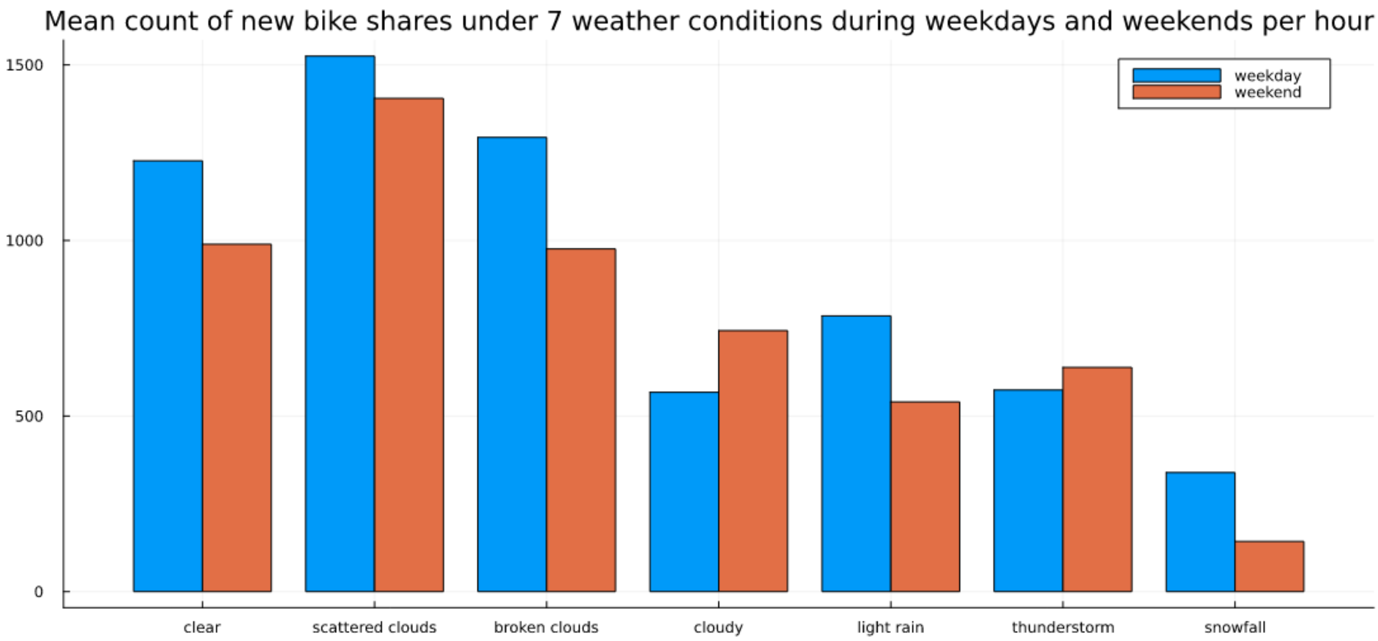


Figure14: mean cnt of distribution under 7 weather conditions during weekdays and weekends per hour

This groupedbar chart shows the mean of people utilizing bike sharing system during weekdays and weekends. Most of the time, people tend to utilize bike sharing system more often during weekdays. However, when it comes to “cloudy” and “thunderstorm”, people who ride shared bikes in the weekends are slightly more than that in the weekdays. It can be interpreted that when it is cloudy during weekdays, people may choose cars or buses to avoid potential rains. But when it comes to weekends, people tend to cherish every chance to go outing, even is has the posibility to rain , resulting in an increase in shared bike usage. But when it is snowing, it becomes extremely hard to go outing, there is a dramatic decrease in bike usage.



Predictive Modeling

References

[1] AndersOhrn (2020) Bike-share usage in London and Taipei Network, Kaggle. Available at: https://www.kaggle.com/datasets/ajohrn/bikeshare-usage-in-london-and-taipei-network (Accessed: October 24, 2022).

[2] Mavrodiev, H. (2019) London Bike Sharing Dataset, Kaggle. Available at: https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset/discussion?resource=download (Accessed: October 24, 2022).