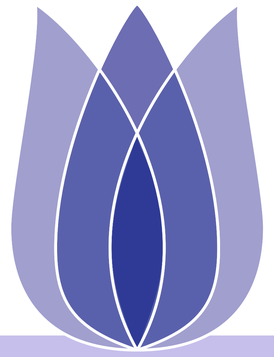
Flip00 Presentation

Tao Yu

Xi'an Shiyou University

(None)



Content

Introduction

Introduction

Defn

Introduction to Bike Sharing Project.

- The influence of weather, time, humidity, wind speed, seasons, holidays and other factors on bicycle usage.
- Analyze and predict bicycle usage.

Data analysis and processing

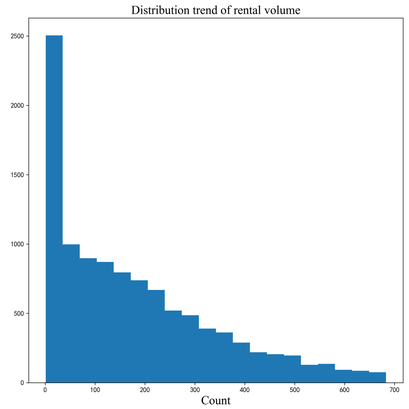
Data analysis

Data	season	holiday	weather	humidity	windspeed	registered	count
count	10886.00	10886.00	10886.00	10886.00	10886.00	10886.00	10886.00
mean	2.506614	0.028569	1.418427	61.886460	12.799395	155.552177	191.574132
std	1.116174	0.166599	0.633839	19.245033	8.164537	151.039033	181.144454
min	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	1.000000	47.000000	7.001500	36.000000	42.000000
50%	3.000000	0.000000	1.000000	62.000000	12.998000	118.000000	145.000000
75%	4.000000	0.000000	2.000000	77.000000	16.997900	222.000000	284.000000
max	4.000000	1.000000	4.000000	100.000000	56.996900	886.000000	977.0000

- From the figure, the standard deviation of the rental volume we need to forecast is very large.
- So let's look at the distribution.

Data processing 1

- The whole distribution inclines seriously and needs to be dealt with in order to avoid over fitting in the end.



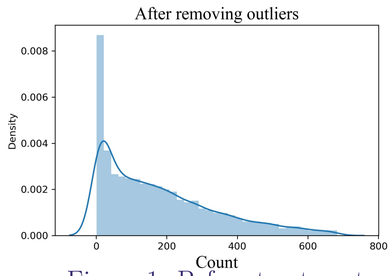


Figure 1: Before treatment

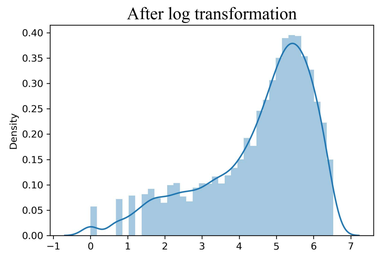


Figure 2: After treatment

- After the conversion, the distribution of the graph is not so severely inclined, and the difference is also smaller.

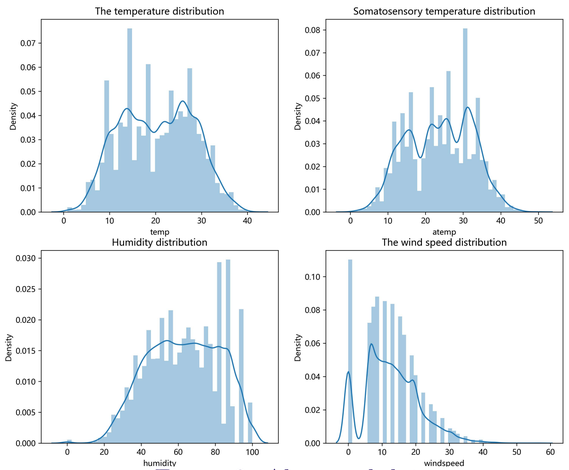


Figure 3: Abnormal data

- There are some gaps between wind speed 1-6. It is speculated that some wind speed data is missing, but the missing wind speed is filled with 0 in the data.

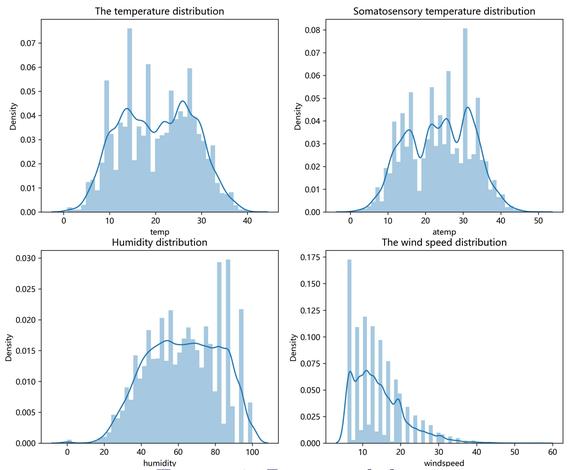


Figure 4: Processed data

- It can be seen that the feature distribution after filling is relatively normal.

Data visualization

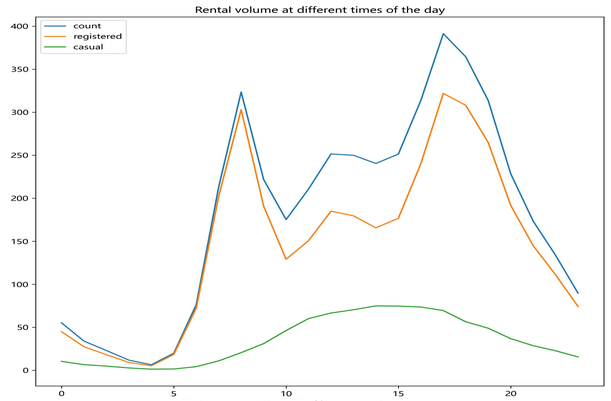


Figure 5: Effect picture

- 7-8 o'clock in the morning, 5-6 o'clock in the afternoon, respectively, the morning peak and evening peak, in line with the actual situation.

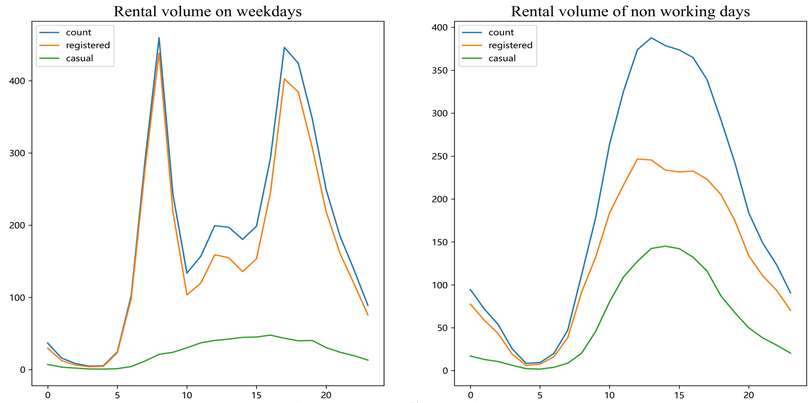


Figure 6: Effect picture

- By comparing the trend of weekdays and non weekdays, we find that the peak time of commuting is obvious on weekdays, while on non weekdays, people prefer to go out after 2-3 PM.

- Come to the conclusion
  - During the week, Saturdays have the highest rental volume.
  - Presumably, this is the day when people spend the most time and enjoy going out, and the number of non-registered users is also the highest.

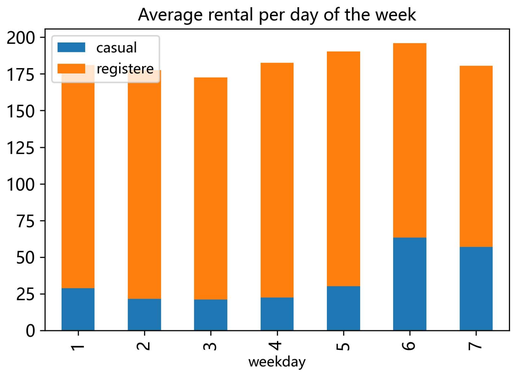


Figure 7: Effect picture

The next stage of work

Work

- Analyze the influence of weather factors on rental volume.
- Feature processing and selection.
  - The correlation analysis of each factor to the rental volume.
- Build and evaluate models.
  - Create training subsets and test subsets.
  - Select the optimal parameter.
- Generate predictions for bike rentals.