Predict number of Bike in sharing systems

Seoul

Nada Fadhel AL-Towerky

Final Project

Choosing the dataset

1

API tools in Twitter

It took more than 3 days to verify my account.

2

Squid game tweets.

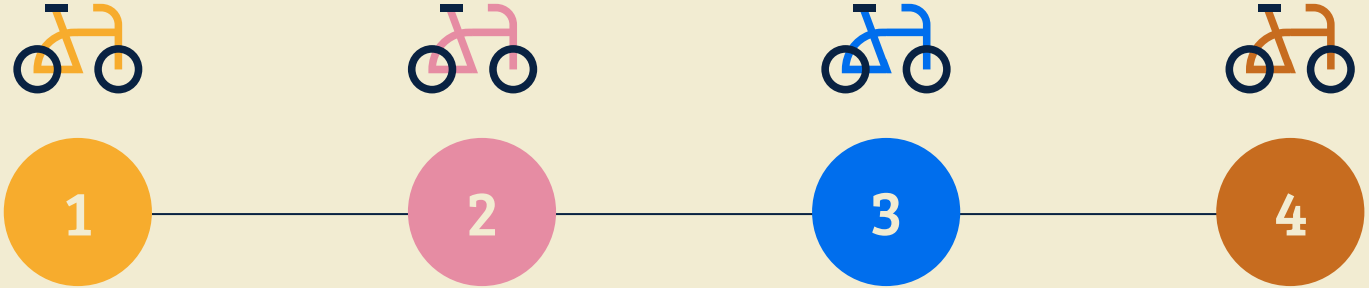
Form kiggle – hard in second week

3

Seoul Bike sharing system

Its good dataset because I I care about the weather and the environment in general because I work in the Ministry of Environment

Introduction

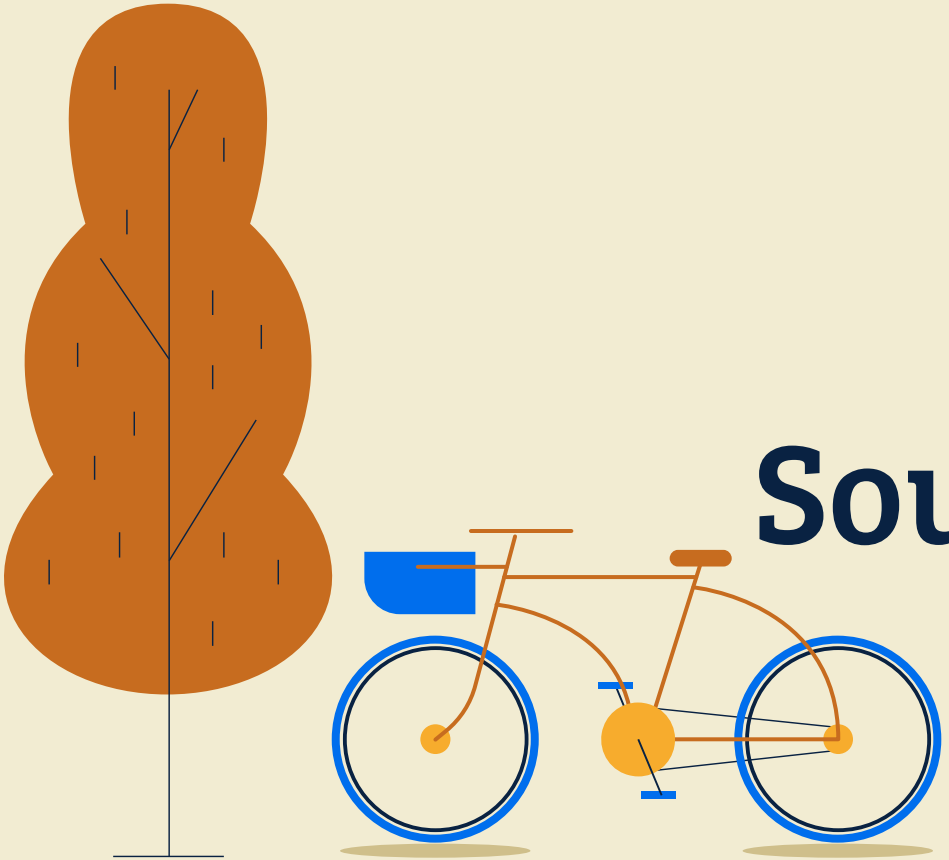


Source of dataset

What is sharing system

Describe the data frame

See the table



1

Source of dataset

UCI

2

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis.

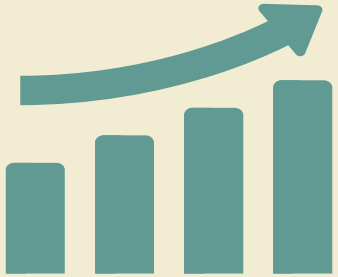
3 Describe the data frame

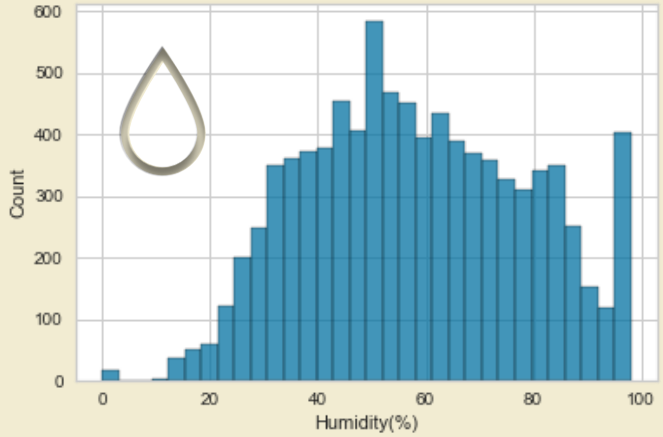
- 8760 entries
- 14 features
  - Date
  - Rented Bike count
  - Hour
  - Temperature
  - Humidity
  - Windspeed
  - Visibility
  - Dew point temperature
  - Solar radiation
  - Rainfall
  - Snowfall
  - Seasons
  - Holiday
  - Functional Day

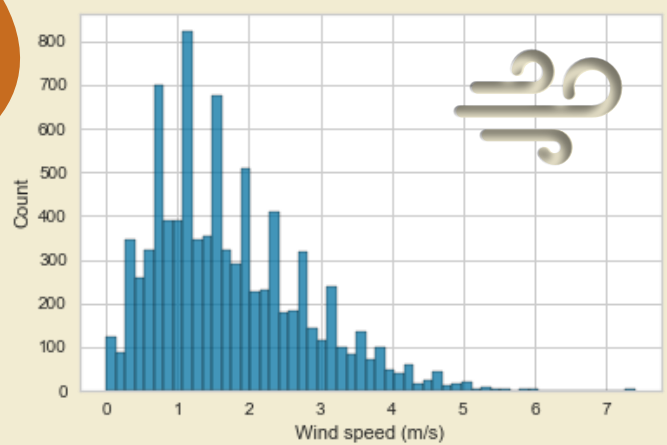
4 The Dataframe

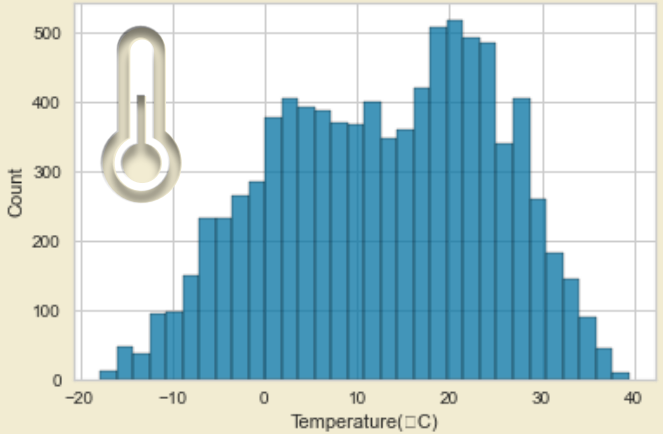
Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
09/09/2018	1662	14	28.2	29	2.0	2000	8.4	3.09	0.0	0.0	Autumn	No Holiday	Yes
24/06/2018	2094	18	29.8	35	3.3	1805	12.6	1.15	0.0	0.0	Summer	No Holiday	Yes
10/05/2018	0	14	21.2	47	3.7	1142	9.4	3.22	0.0	0.0	Spring	No Holiday	No
28/01/2018	187	16	-2.1	27	4.0	2000	-18.6	0.52	0.0	0.0	Winter	No Holiday	Yes
08/07/2018	2266	17	27.4	44	1.7	2000	14.0	1.17	0.0	0.0	Summer	No Holiday	Yes

Plot



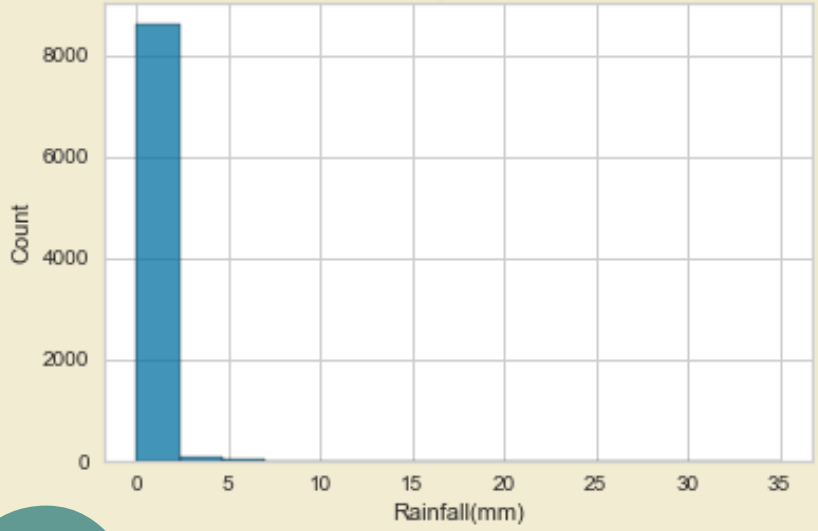


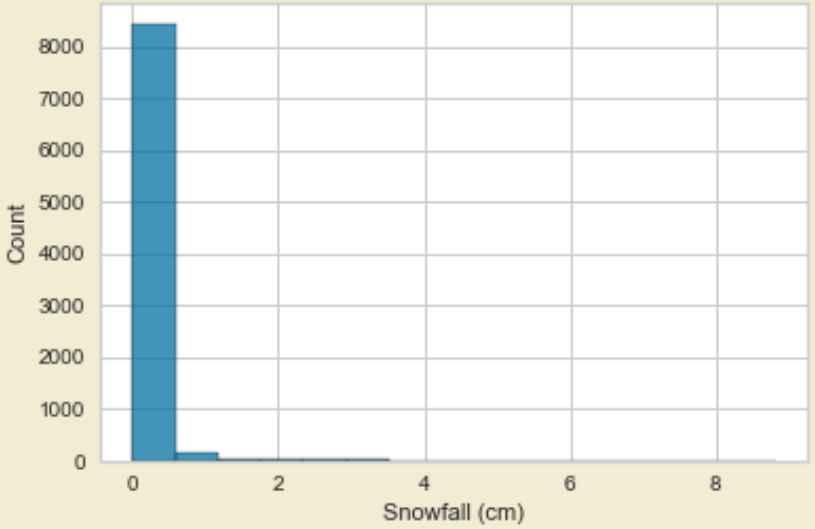




As shown the hist of humidity and temp and wend speed looks Normal

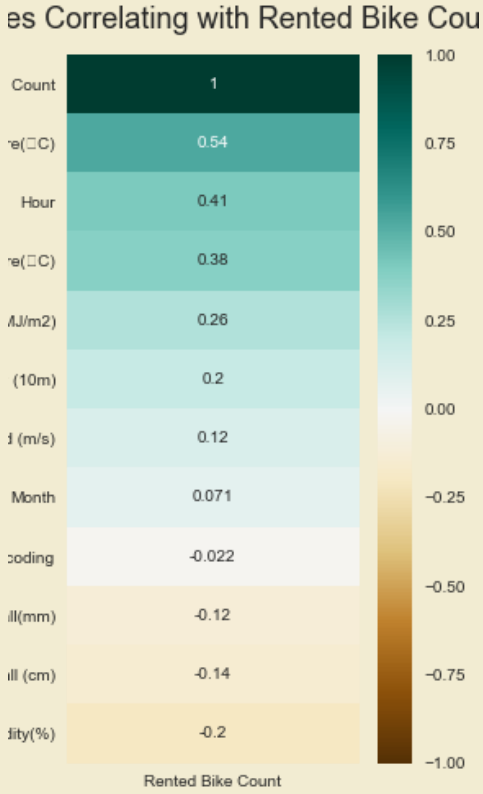
Rain and Snow



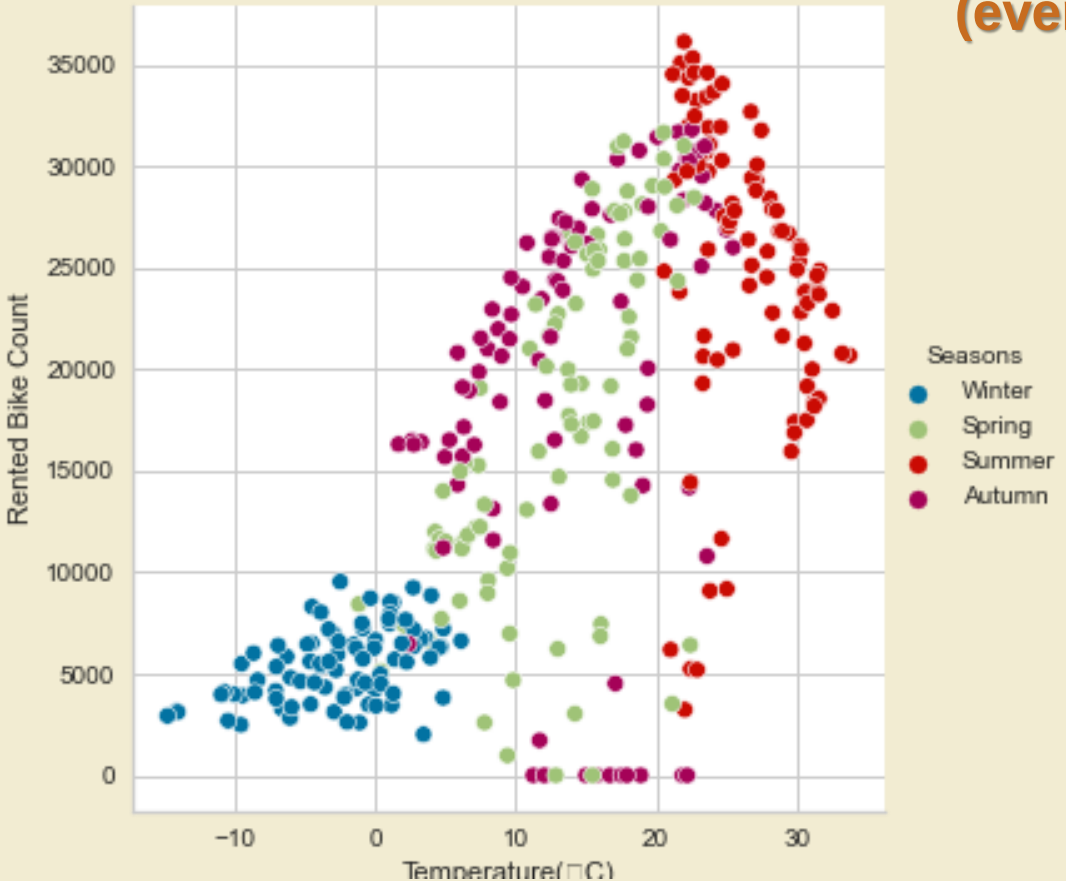


Snowfall and Rainfall are highly skewed

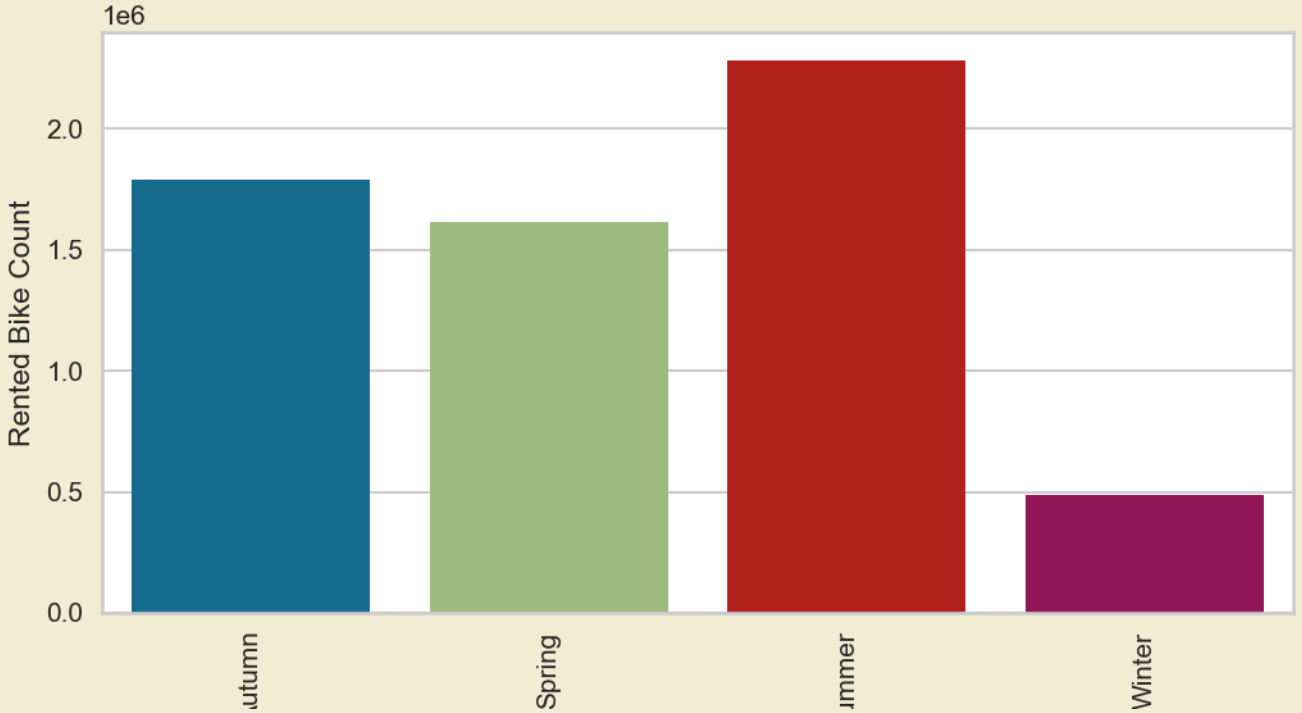
Correlation Heatmap



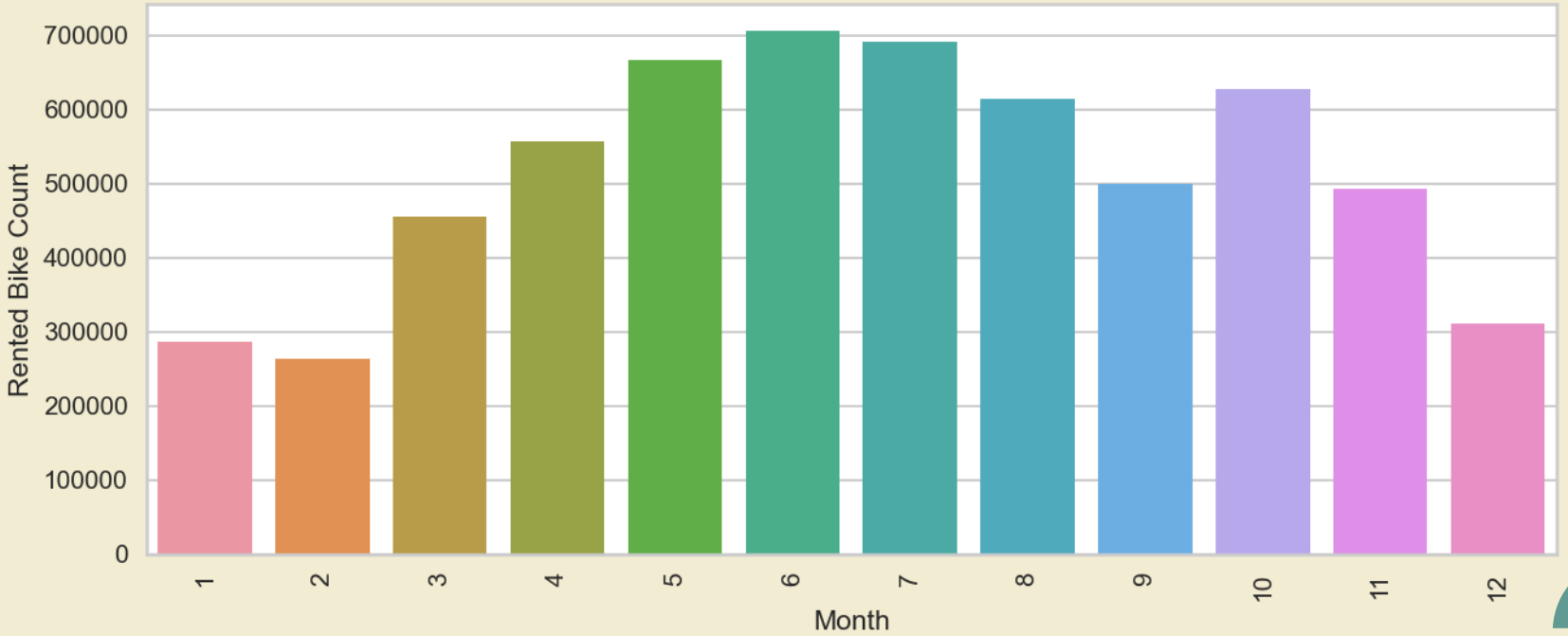
Scatter gorupby date (every dot is day)



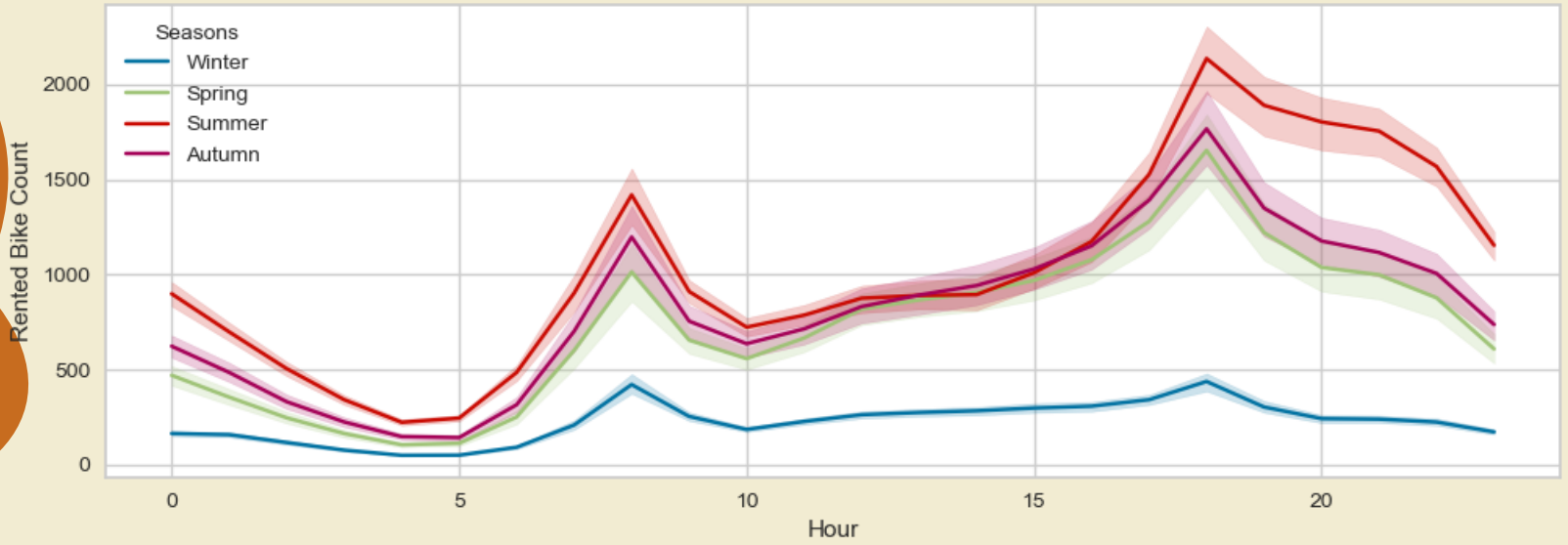
Summer have high rented bike
(the mean temp in Autumn higher than spring)



Months


Rented Bike Count
700000
600000
500000
400000
300000
200000
100000
0
1
2
3
4
5
6
7
8
9
10
11
12
Month

Most of bike rented for work



Modeling

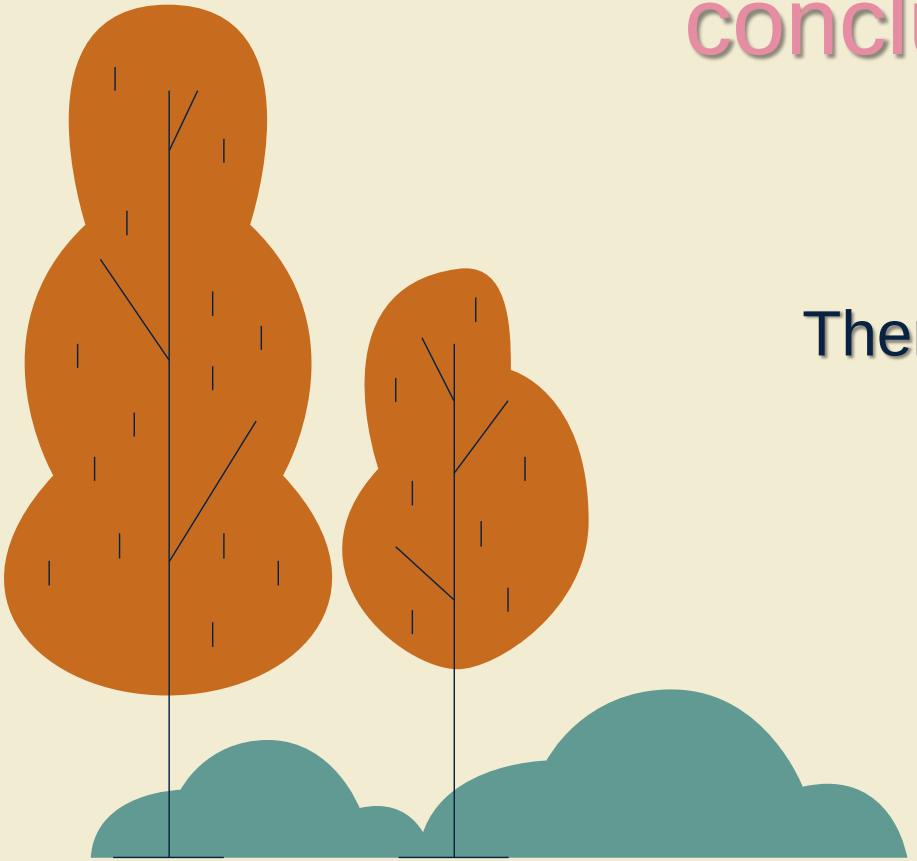
Regular Linear Regression

R-squared 0.755

Lasso Regression

R-squared 0.462

conclusion



There is overfitting