

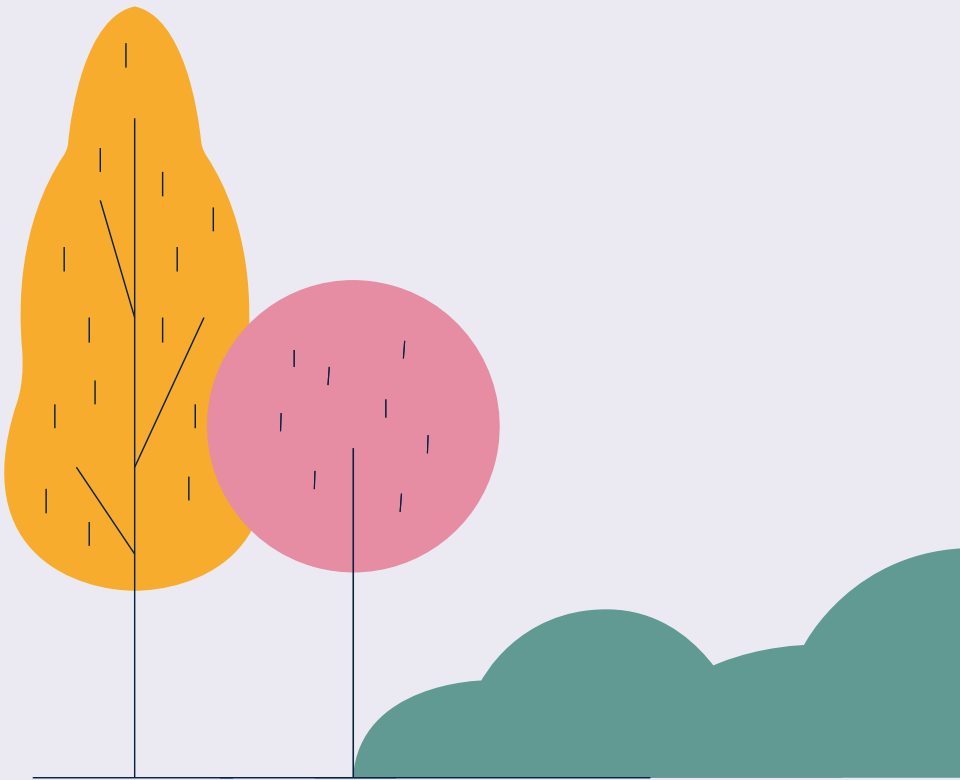
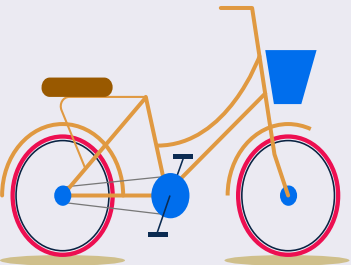
Bike sharing





Overview

Capital Bike Sharing dataset contains information related to the bike sharing program of Washington DC for the 2011 and 2012 years. To review the data set attributes, visit the [data set page on UCI](#).





data!

The dataset contains 17 attributes , 17389 instances

Problem statements



For predict

Prediction of bike rental count hourly based on the environmental and seasonal settings



For visualize

- Bike years highest season
- Top demand hours for Bikes
- Humidity and temperature effect on bike rental

Data cleaning

Check any outlier and handle it

Check if have any inconsistent



Check Missing & duplicate data

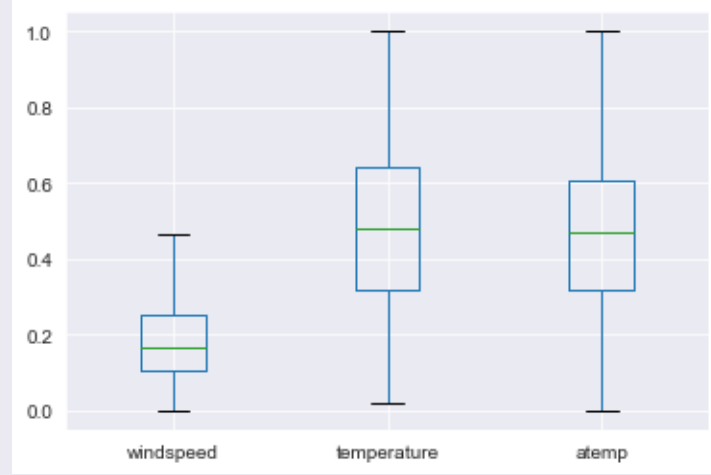
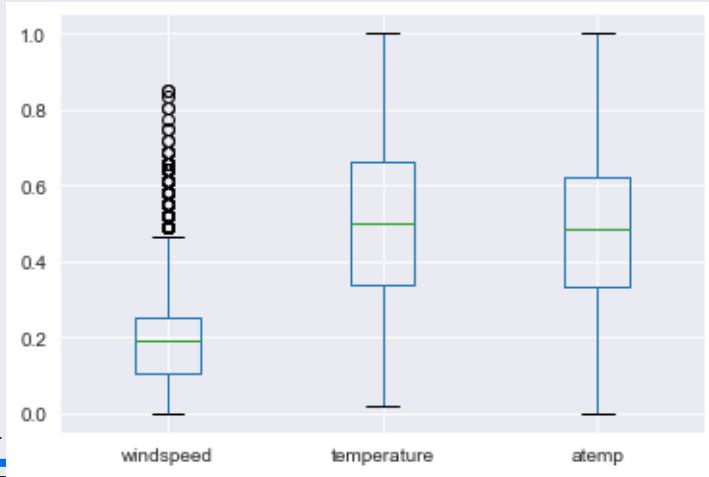
Rename column

Map data nice to visualize

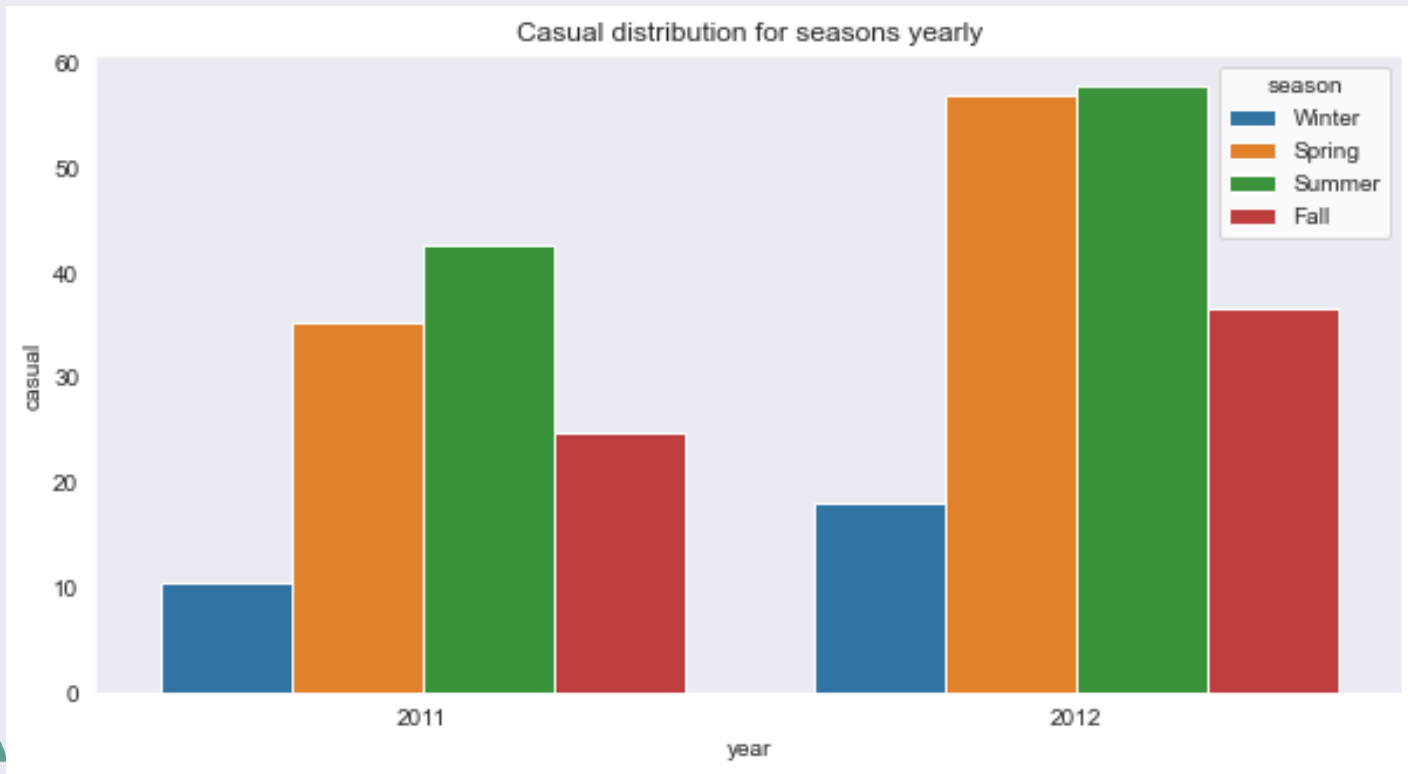
Sample from dataset after cleaning

rec_id	datetime	season	year	month	hr	is_holiday	weekday	is_workingday	weather	temperature	atemp	humidity	windspeed	casual	registered
3288	2011-05-21	Spring	2011	5	21	No	Sunday	No	Clear	0.62	0.6061	0.61	0.1045	77	150
2103	2011-04-02	Spring	2011	4	11	No	Sunday	No	Cloudy	0.40	0.4091	0.47	0.1642	72	126
2664	2011-04-25	Spring	2011	4	21	No	Tuesday	Yes	Clear	0.62	0.6061	0.69	0.2537	41	149
1602	2011-03-12	Winter	2011	3	5	No	Sunday	No	Clear	0.22	0.2273	0.69	0.1940	0	2
7687	2011-11-21	Fall	2011	11	21	No	Tuesday	Yes	Rain	0.40	0.4091	0.87	0.1642	11	103

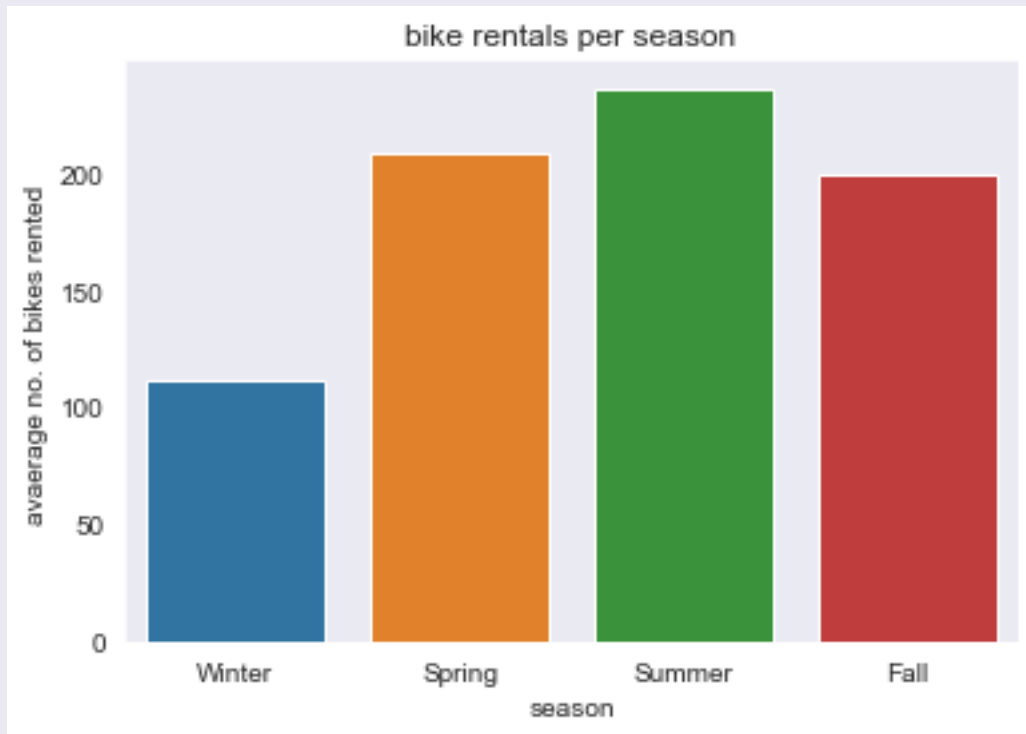
check outlier



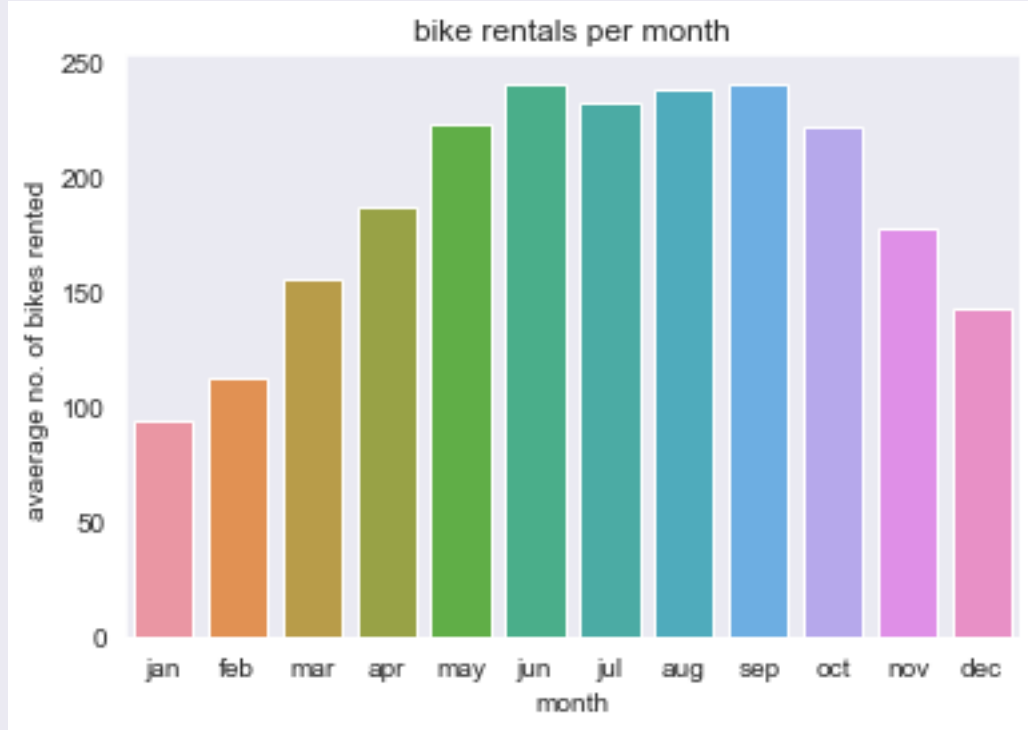
Casual distribution for season yearly



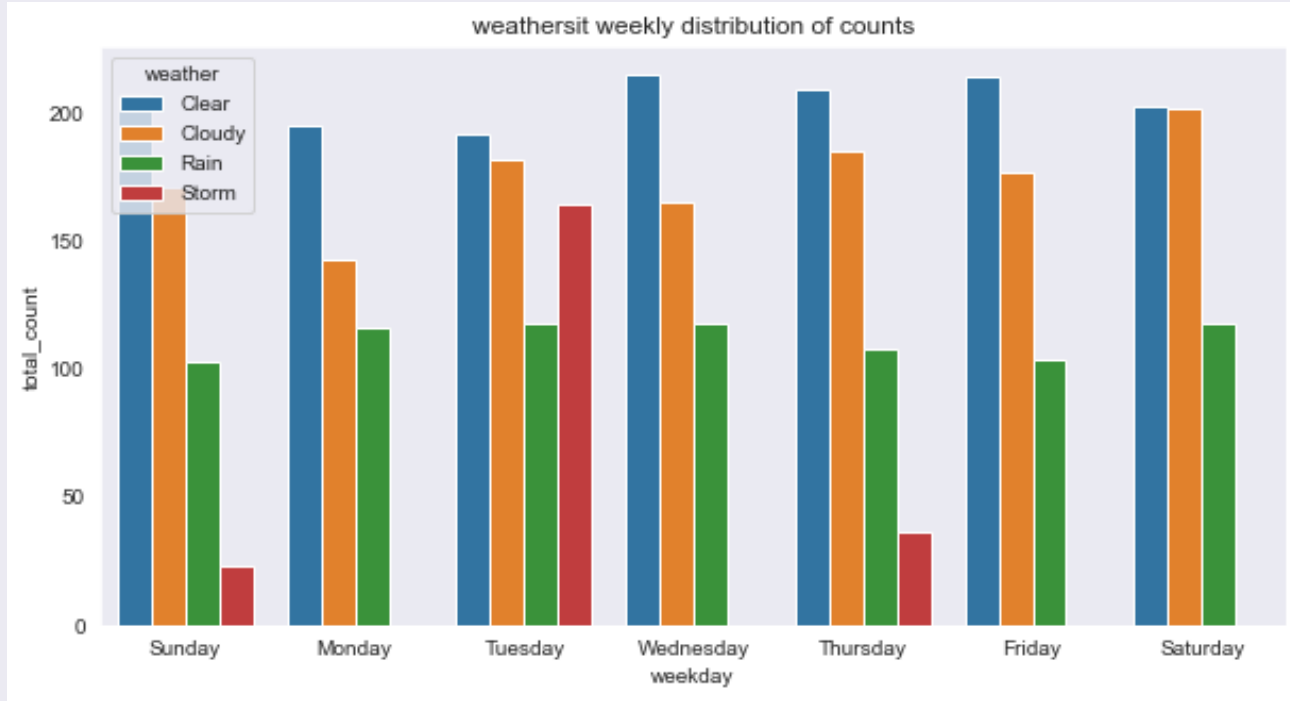
Bike rental per season



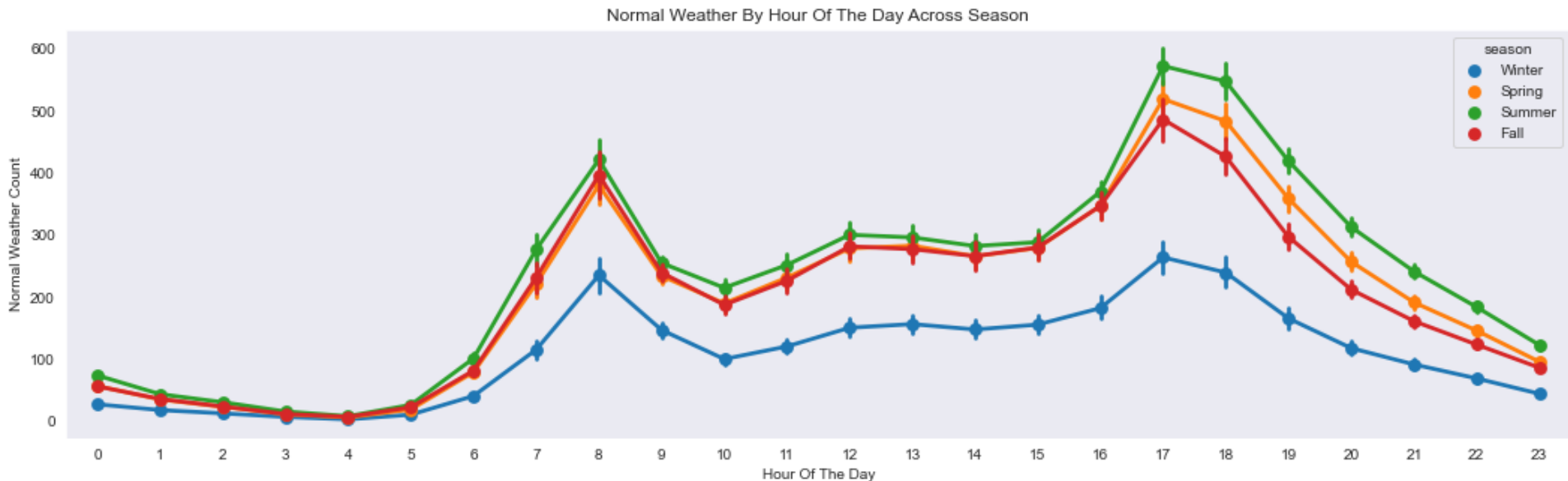
Bike rental per month



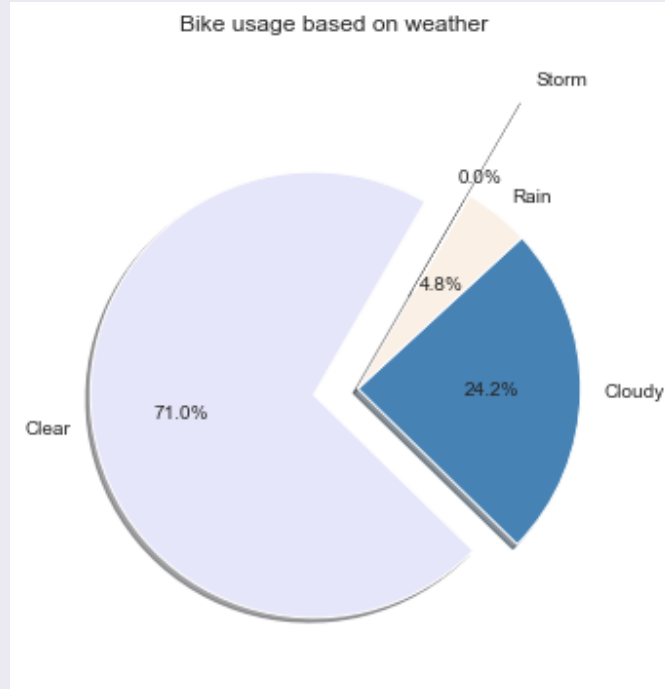
weather weekly distribution of counts



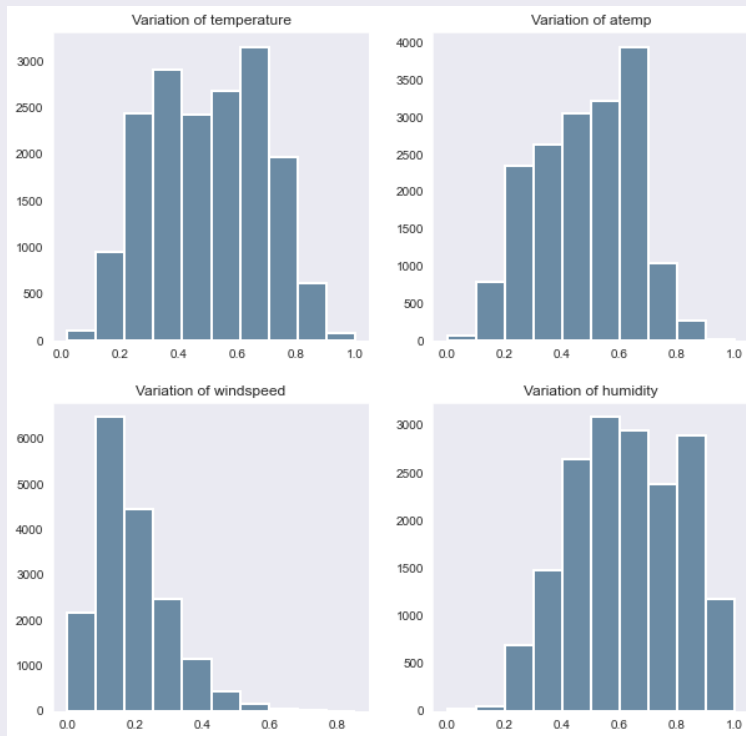
Normal weather By hour of the day across season



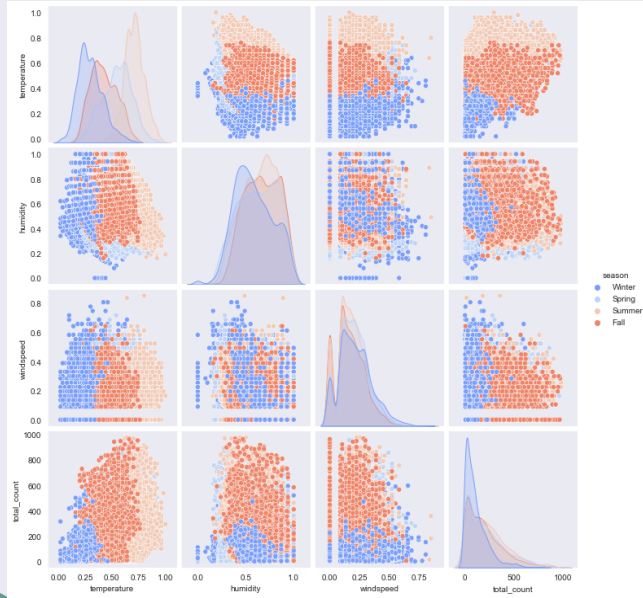
Bike usage on weather



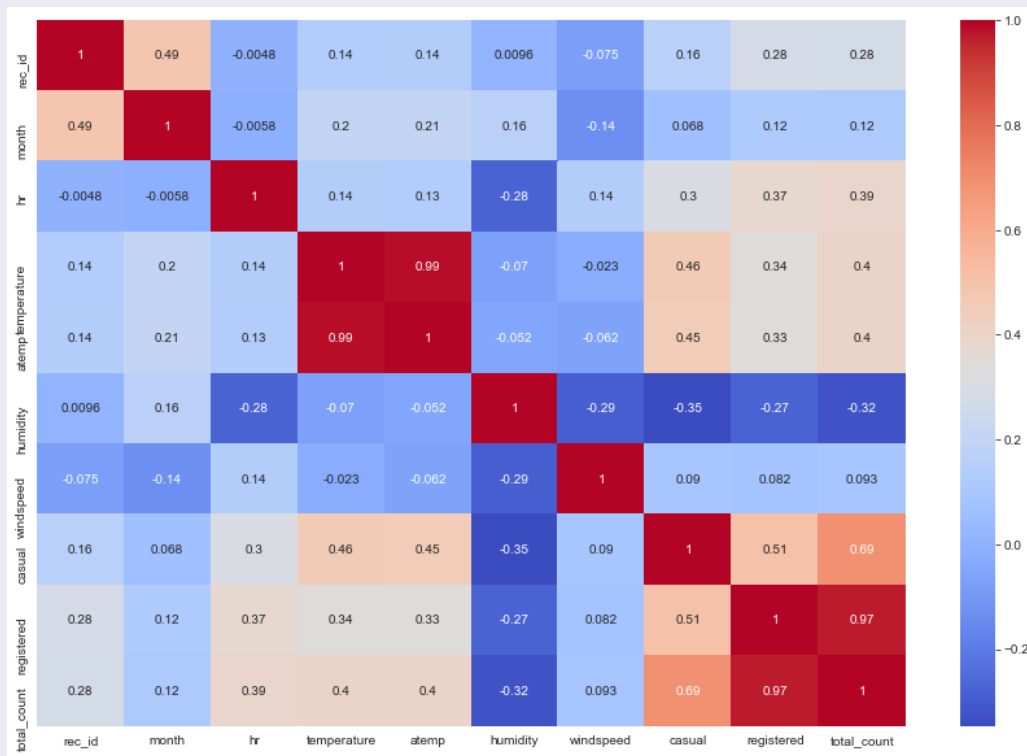
Humidity and temperature effect



Winds speed after fill by 0



Visualize correlation



Drop unnecessary column





Feature engineering

- Encode categorical column

Hour & season
Drop first column

-Fill zero value in windspeed

by KNN neighbors value





Feature engineering

-Scaler numerical column

By standard scaler

- Split data Cross validation

By 5 Kfolds





Models



R square

Linear Regression

0.69

Ridge Regression

0.69

Lasso Regression

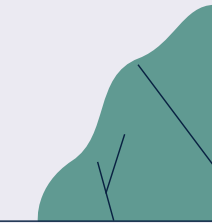

0.69

Polynomial degree (2)

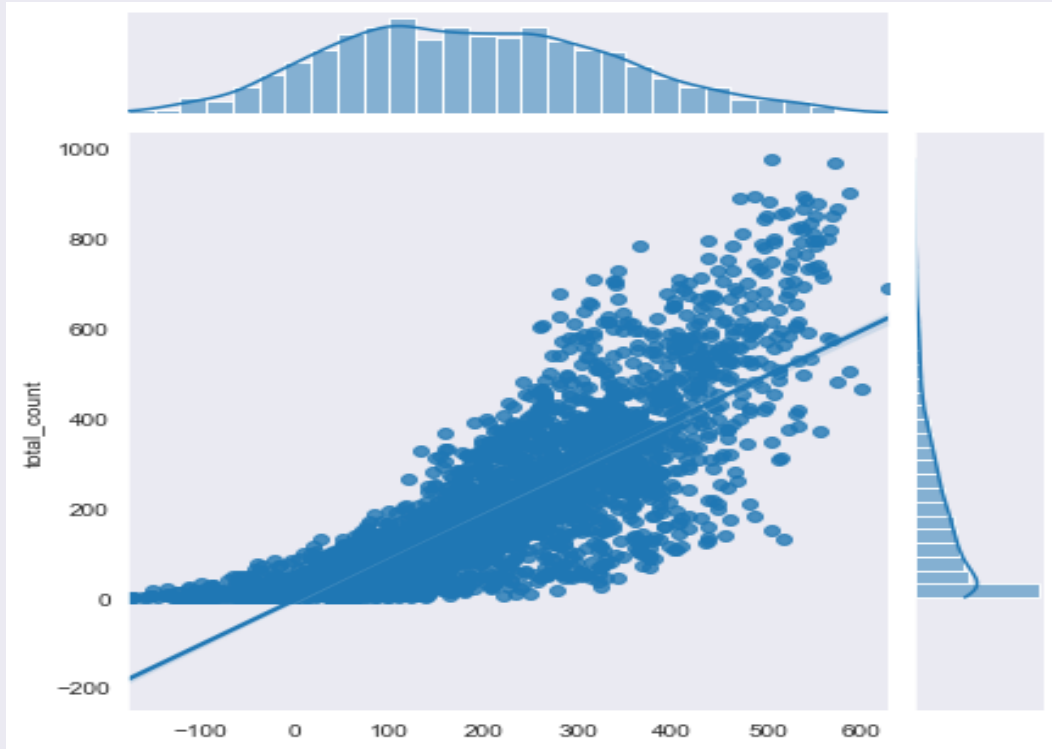
0.90

Predict score

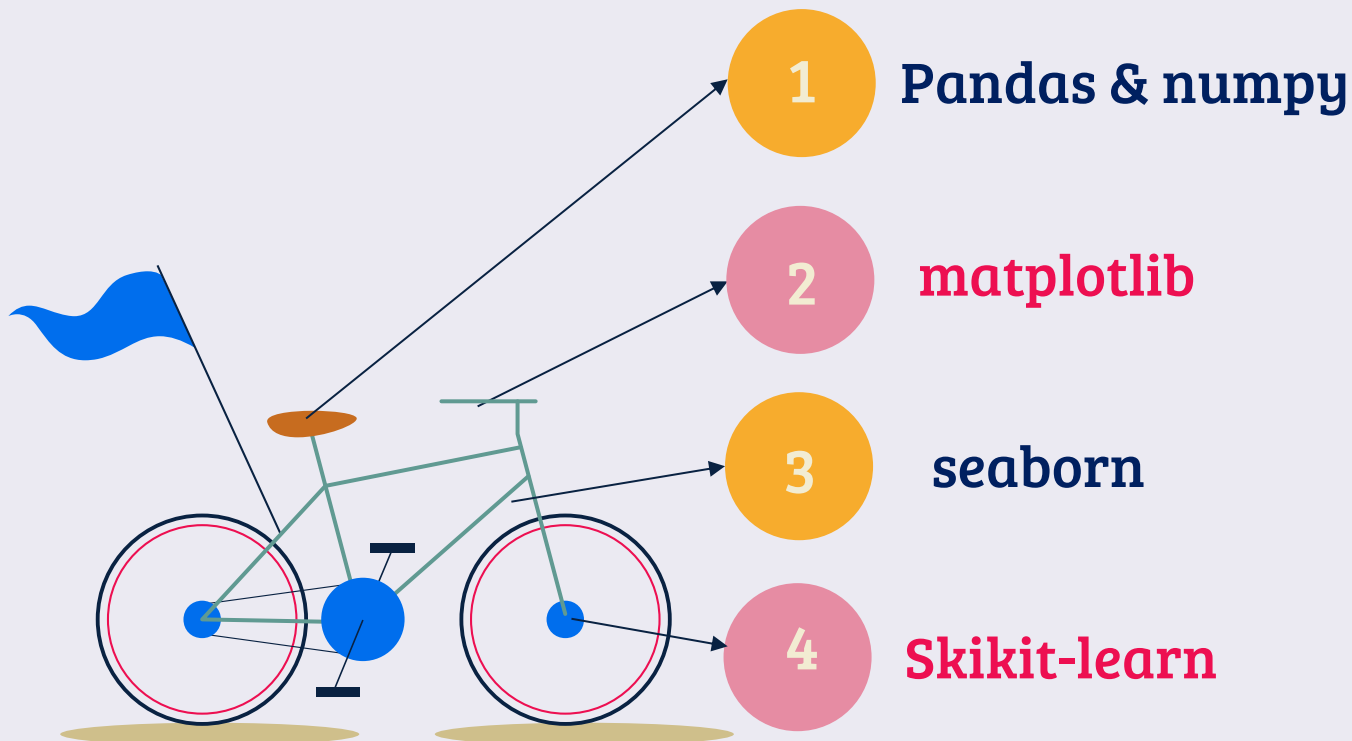
0.674




Result



Tools





“The best way to predict your
future is to create it.”

Thanks

Do you have any questions?