Boom Bike – A Bike Sharing Company

- Vasundhara L Kadekodi

Business Context:

A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it. This bike can then be returned to another dock belonging to the same system.

Problem Statement:

A US bike-sharing provider Boom Bikes has recently suffered considerable dips in their revenues due to the ongoing Corona pandemic. The company is finding it very difficult to sustain in the current market scenario. So, it has decided to come up with a mindful business plan to be able to accelerate its revenue as soon as the ongoing lockdown comes to an end, and the economy restores to a healthy state.

In such an attempt, BoomBikes aspires to understand the demand for shared bikes among the people after this ongoing quarantine situation ends across the nation due to Covid-19. They have planned this to prepare themselves to cater to the people's needs once the situation gets better all around and stand out from other service providers and make huge profits.

Objective

The goal is to build a linear model to analyse the demand for shared bikes with the available independent variables. So that we can understand how exactly the demands vary with different features, based on the analysis, we can make the business strategy to meet the demand levels and meet the customer's expectations.

Steps Used

Fetch & Analyze data

Pre-Processing the data

Building a prediction Model

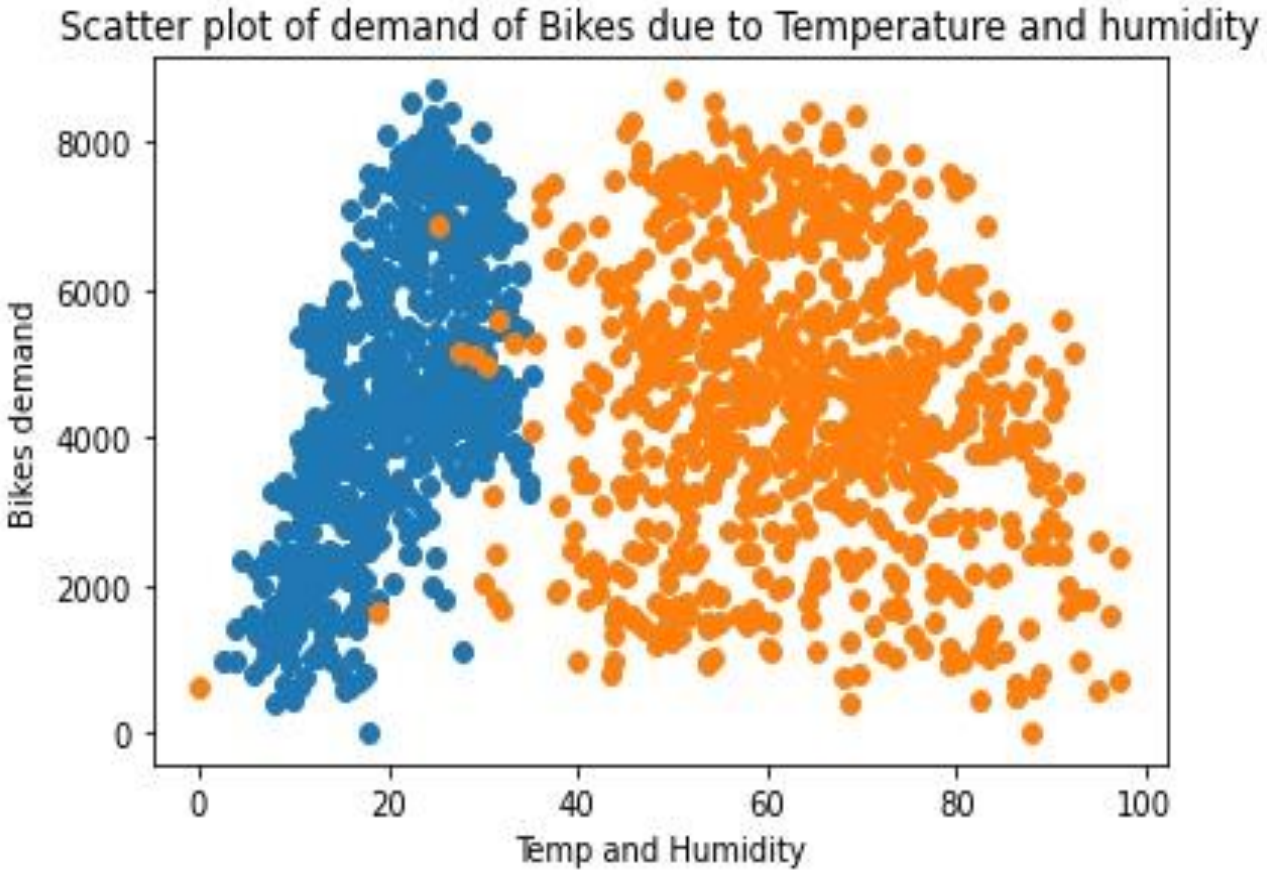
Validate the Model

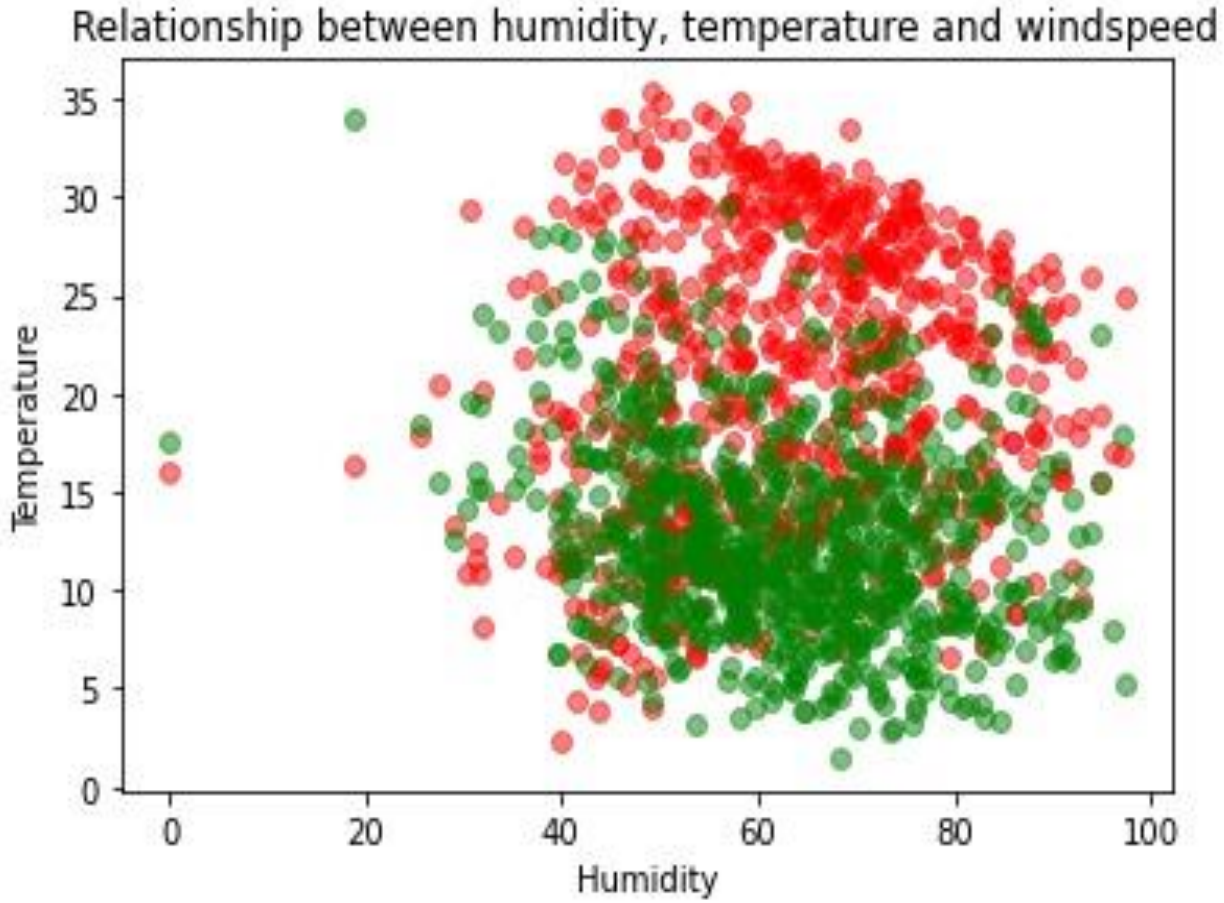
Predicting values for test data

Tools Used:

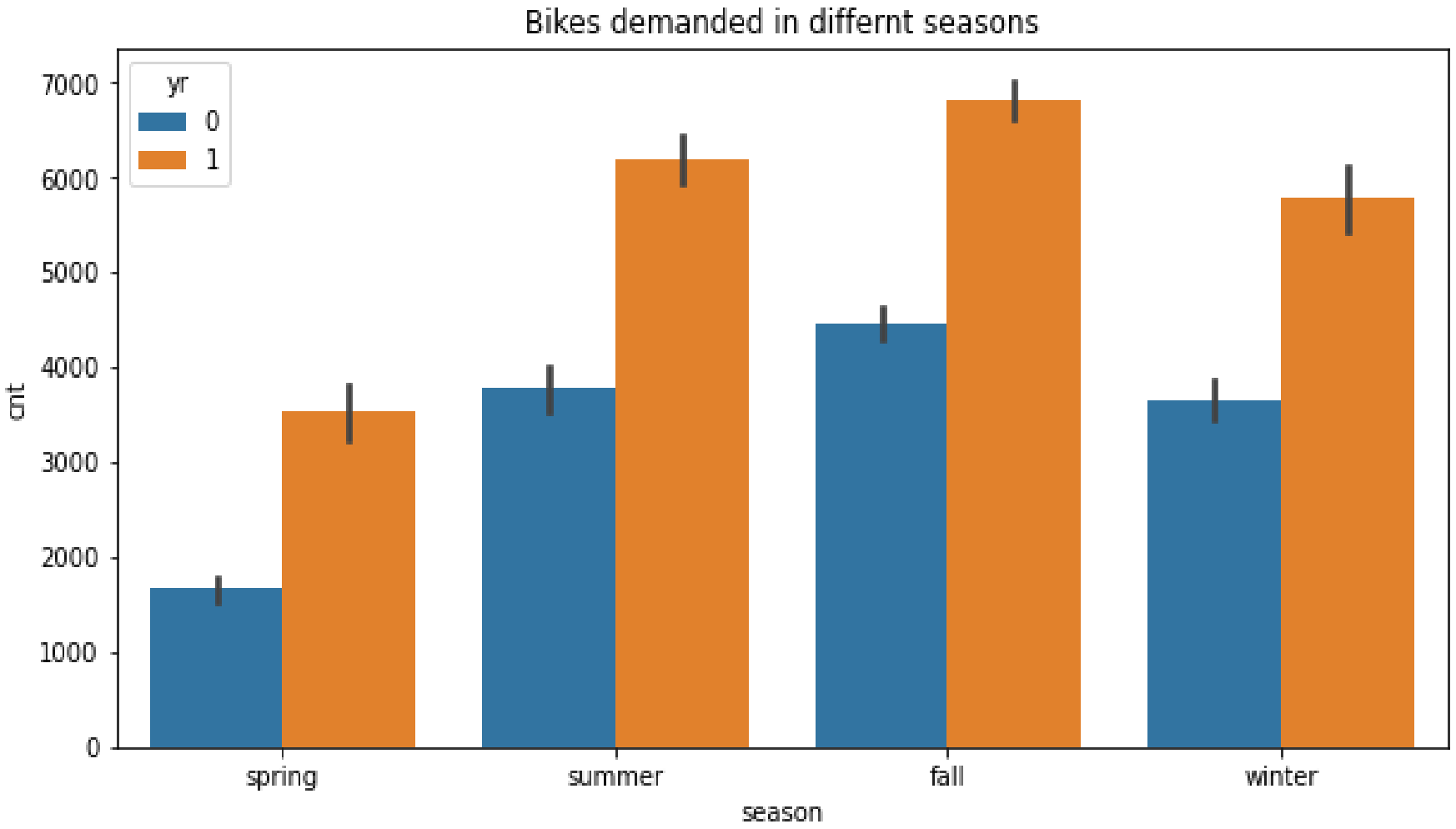
python

Relationship between Temperature, Humidity and Windspeed



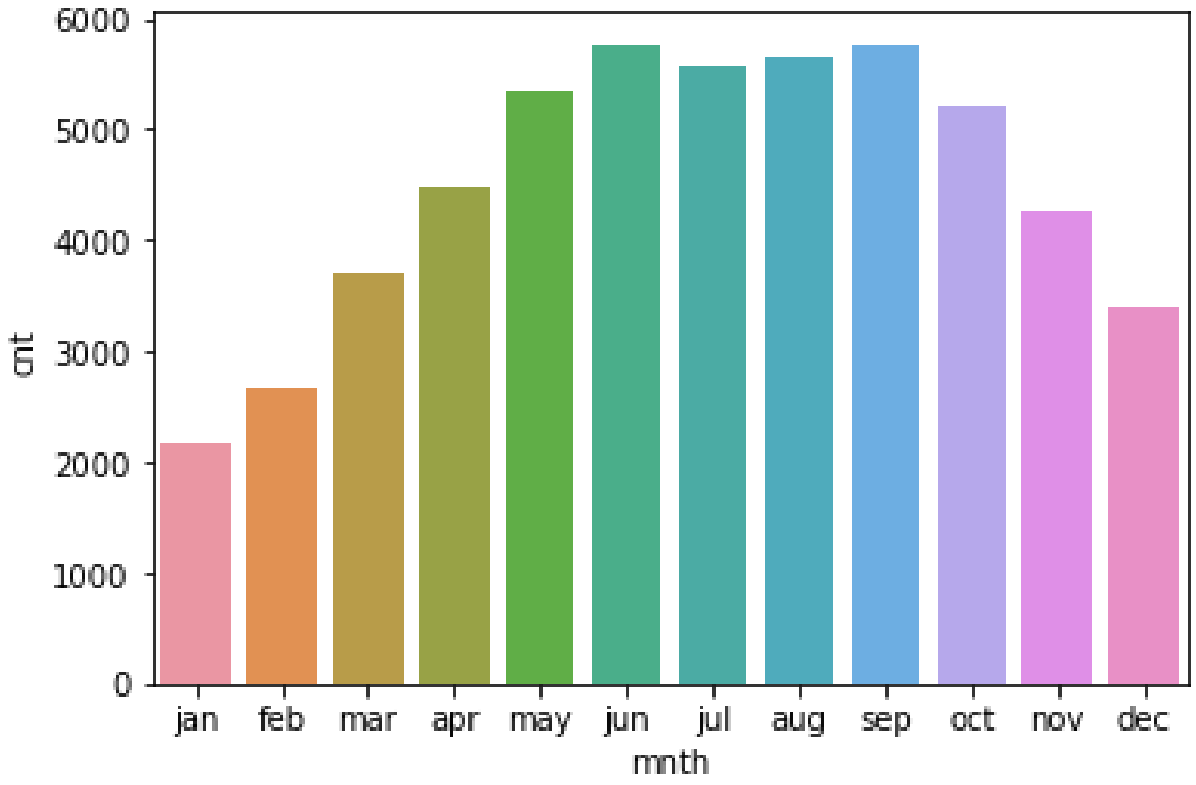


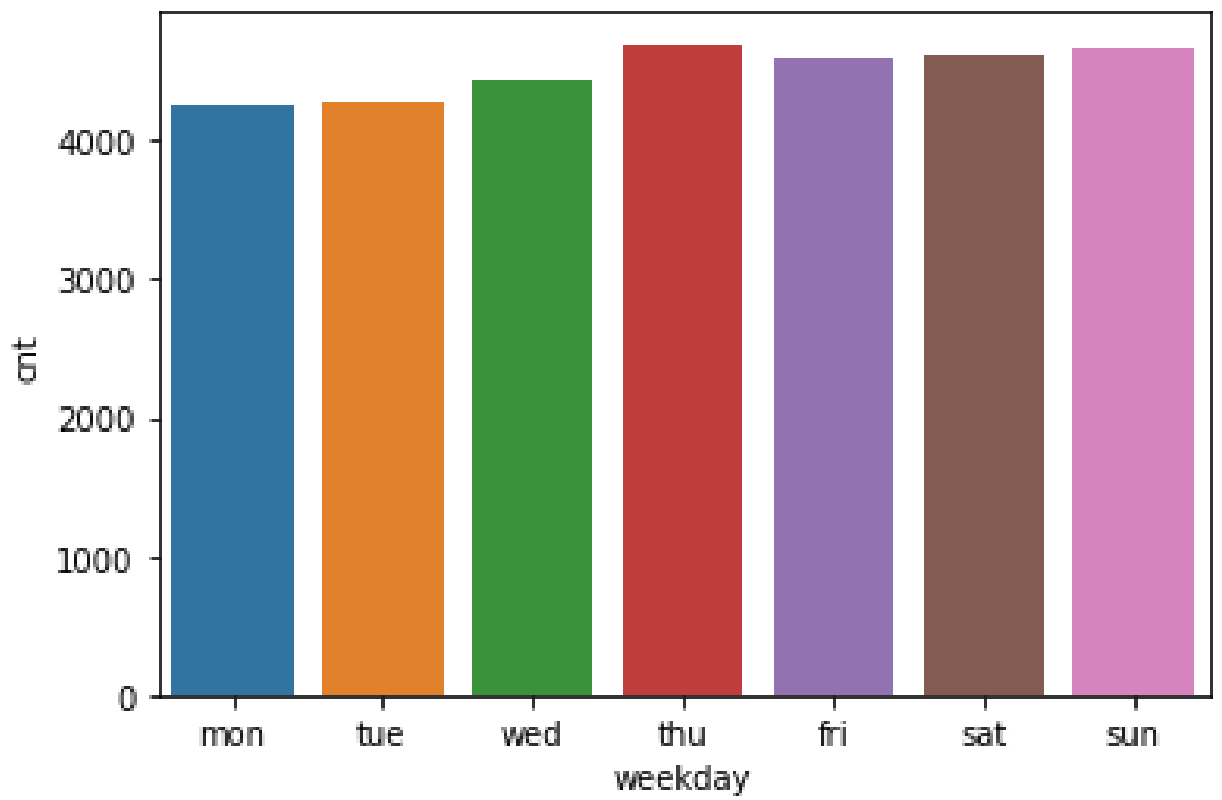
Bikes demand in different seasons and different years



It has been found that the highest demand for bicycles in both 2019 and 2018 is during the fall season, and the lowest demand in spring

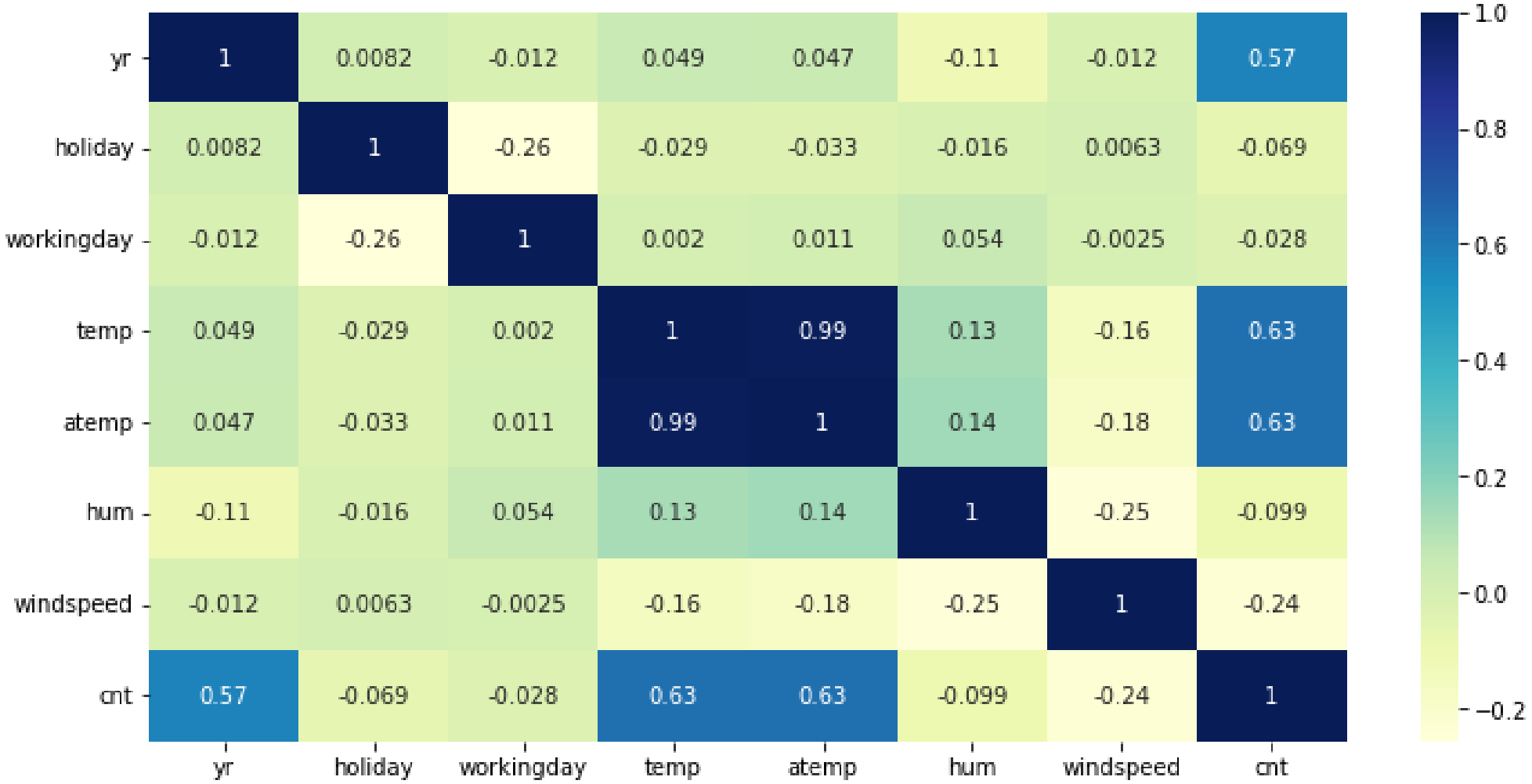
Bikes demand in different months and weeks





- Renters used the rental bikes more on Thursday and Sunday than on other weekdays.
- A large number of users used the rental bike in June, July, August and September (i.e. rainy season).

Correlation



According to the graph, there is a very strong correlation between the variables 'atemp' and 'temp' i.e. 0.99, and they are also moderately correlated with the target variable "cnt".

Final Model
(Linear Regression)

OLS Regression Results

Dep. Variable:	cnt	R-squared:	0.836
Model:	OLS	Adj. R-squared:	0.832
Method:	Least Squares	F-statistic:	230.4
Date:	Tue, 03 Jan 2023	Prob (F-statistic):	2.40e-187
Time:	21:21:26	Log-Likelihood:	499.17
No. Observations:	510	AIC:	-974.3
Df Residuals:	498	BIC:	-923.5
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1996	0.030	6.758	0.000	0.142	0.258
yr	0.2335	0.008	28.361	0.000	0.217	0.250
holiday	-0.0980	0.026	-3.761	0.000	-0.149	-0.047
temp	0.4915	0.033	14.798	0.000	0.426	0.557
windspeed	-0.1480	0.025	-5.893	0.000	-0.197	-0.099
spring	-0.0669	0.021	-3.167	0.002	-0.108	-0.025
summer	0.0453	0.015	2.971	0.003	0.015	0.075
winter	0.0831	0.017	4.818	0.000	0.049	0.117
light Rain	-0.2852	0.025	-11.536	0.000	-0.334	-0.237
mist	-0.0816	0.009	-9.301	0.000	-0.099	-0.064
jul	-0.0524	0.019	-2.811	0.005	-0.089	-0.016
sep	0.0767	0.017	4.511	0.000	0.043	0.110

Omnibus:	59.298	Durbin-Watson:	2.041
Prob(Omnibus):	0.000	Jarque-Bera (JB):	135.189
Skew:	-0.628	Prob(JB):	4.41e-30
Kurtosis:	5.187	Cond. No.	17.3

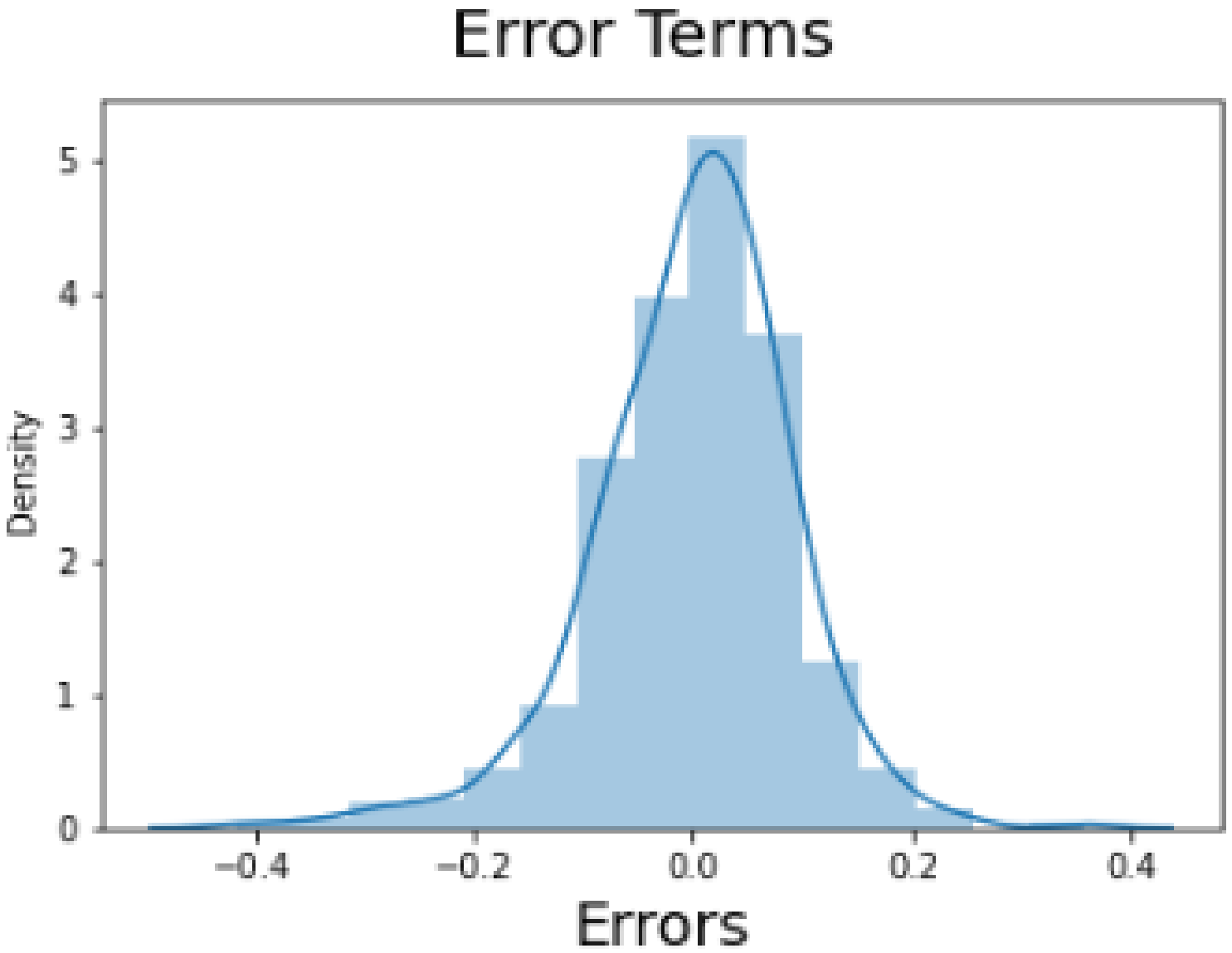
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

	Features	VIF
2	temp	5.09
3	windspeed	4.60
5	summer	2.21
4	spring	2.08
0	yr	2.07
6	winter	1.79
9	jul	1.58
8	mist	1.55
10	sep	1.34
7	light Rain	1.08
1	holiday	1.04

This model is Significant as per below observations

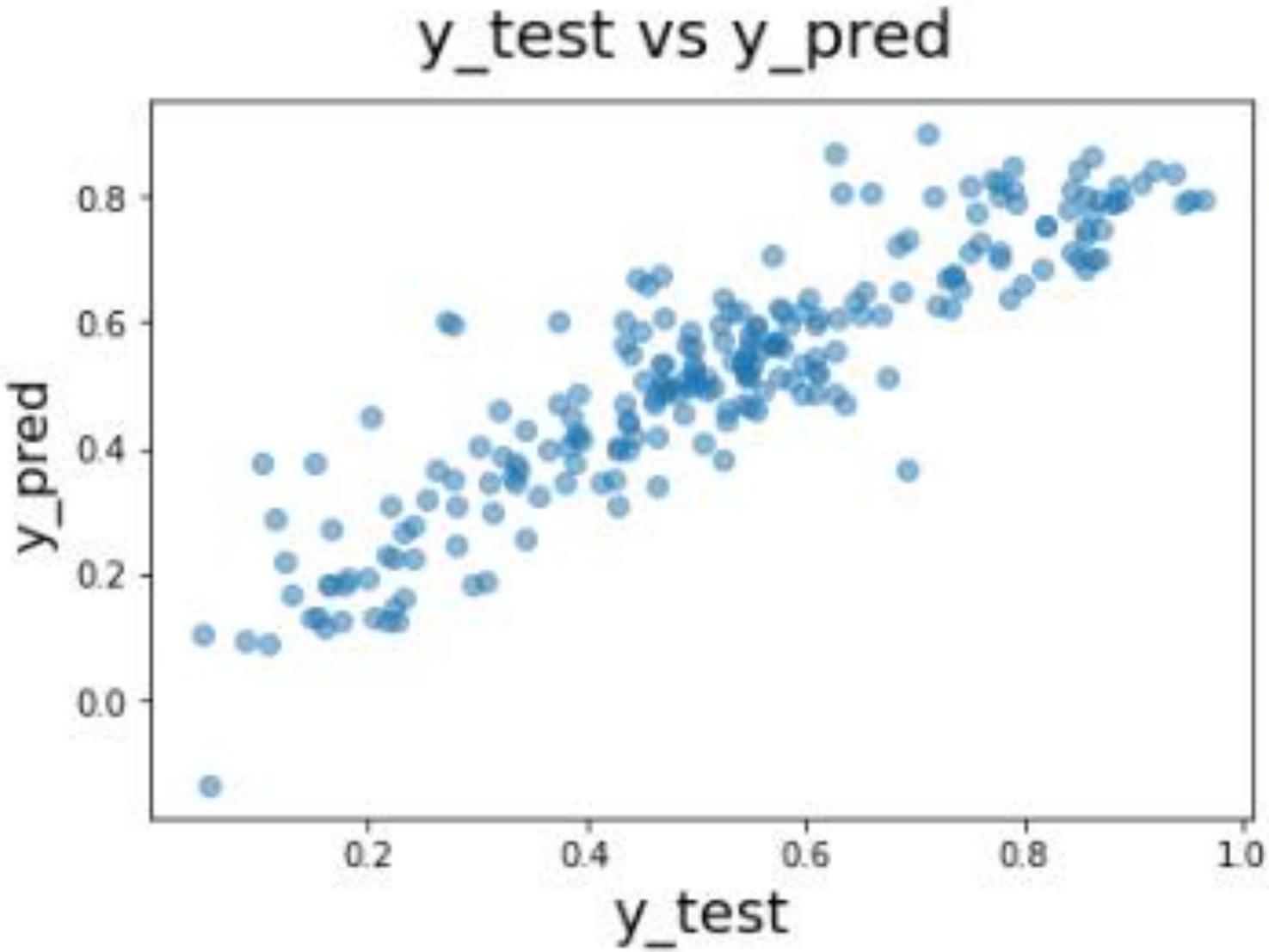
- ❖ All the p-values are less than 0.05
- ❖ All the VIF values are lies in the acceptance region i.e 1 to 5
- ❖ R square value is 0.836
- ❖ Adjusted R square value is 0.832

Residual Analysis of the train data



This histogram indicates that the residuals are normally distributed, thus supporting our assumption that linear regression is valid.

Best Fit Line



$$cnt = 0.1996 + (0.2334 * yr) - (0.0980 * holiday) + (0.4915 * temp) - (0.1479 * windspeed) - (0.0664 * spring) + (0.0452 * summer) + (0.0830 * winter) - (0.2851 * lightrain) - (0.0815 * mist) - (0.05241 * jul) + (0.0766 * sep)$$

Solution Developed

As per the final model, the top 3 predictor variables that influence bike booking are:

- ✓ Temperature (Temp): A coefficient value of ‘0.491508’ indicated that a temperature has significant impact on bike rentals
- ✓ Light Rain: A coefficient value of ‘-0.2852’ indicated that the light snow and rain deters people from renting out bikes
- ✓ Year (yr): A coefficient value of ‘0.2335’ indicated that year wise the rental numbers are increasing

Suggestions

It is recommended to give utmost importance to these three variables while planning to achieve maximum bike rental booking.

As high temperature and good weather positively impacts bike rentals, it is recommended that bike availability and promotions to be increased during summer months to further increase bike rentals.

THANK YOU