

Bike Sharing Assignment

Satyanarayana.k

Problem Statement

- 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.
- A US bike-sharing provider BoomBikes 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.
- They have contracted a consulting company to understand the factors on which the demand for these shared bikes depends. Specifically, they want to understand the factors affecting the demand for these shared bikes in the American market. The company wants to know:
 - Which variables are significant in predicting the demand for shared bikes.
 - How well those variables describe the bike demands
- Based on various meteorological surveys and people's styles, the service provider firm has gathered a large dataset on daily bike demands across the American market based on some factors.

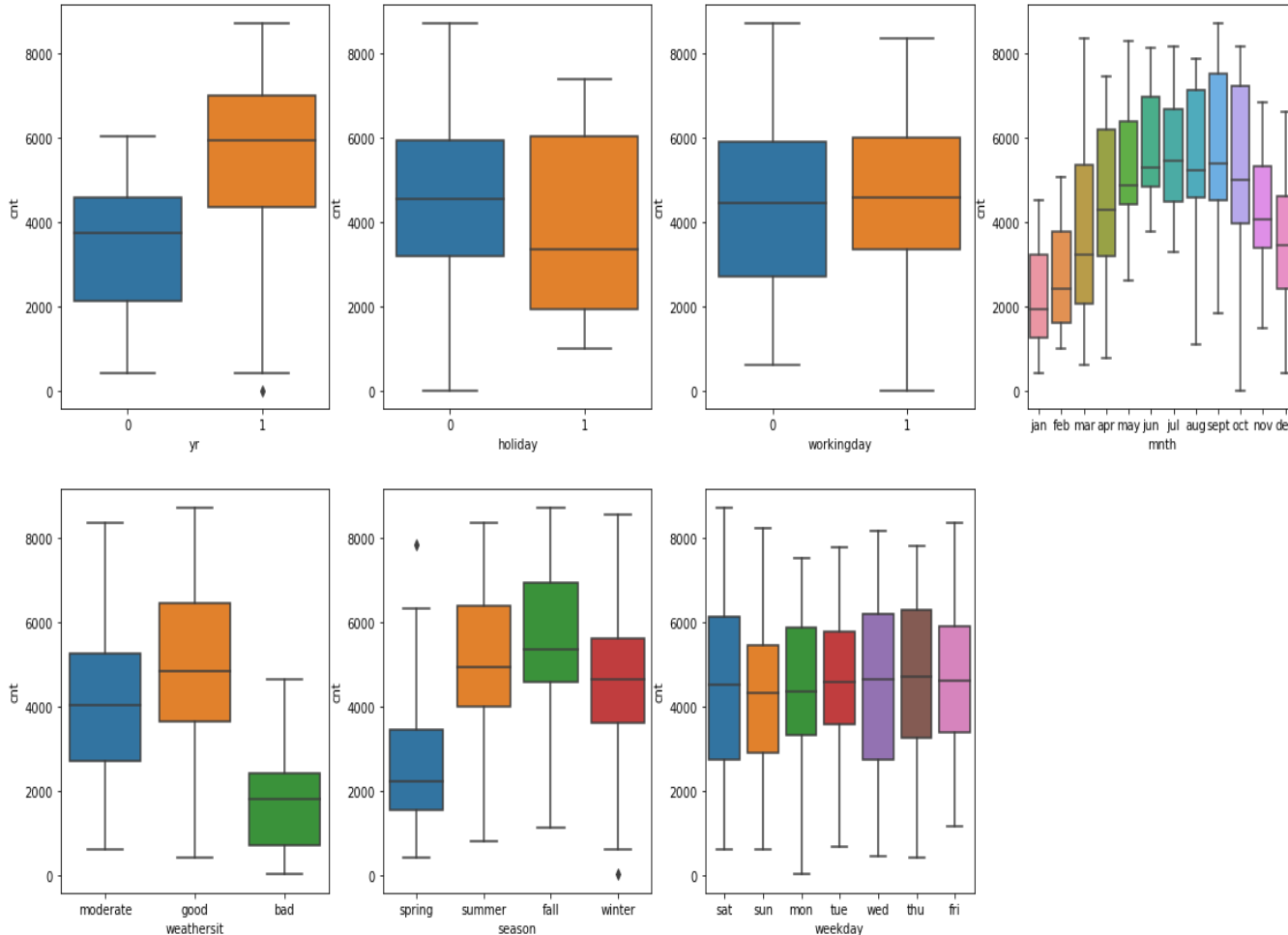
Business Goal

- You are required to model the demand for shared bikes with the available independent variables. It will be used by the management to understand how exactly the demands vary with different features. They can accordingly manipulate the business strategy to meet the demand levels and meet the customer's expectations. Further, the model will be a good way for management to understand the demand dynamics of a new market.

Steps for the process

- Step 1: EDA
- Step 2: Data visualization
- Step 3: Training the model
- Step 4: Residual analysis
- Step 5: predicting and evaluating
- Step 6: Conclusion

EDA & Data visualization



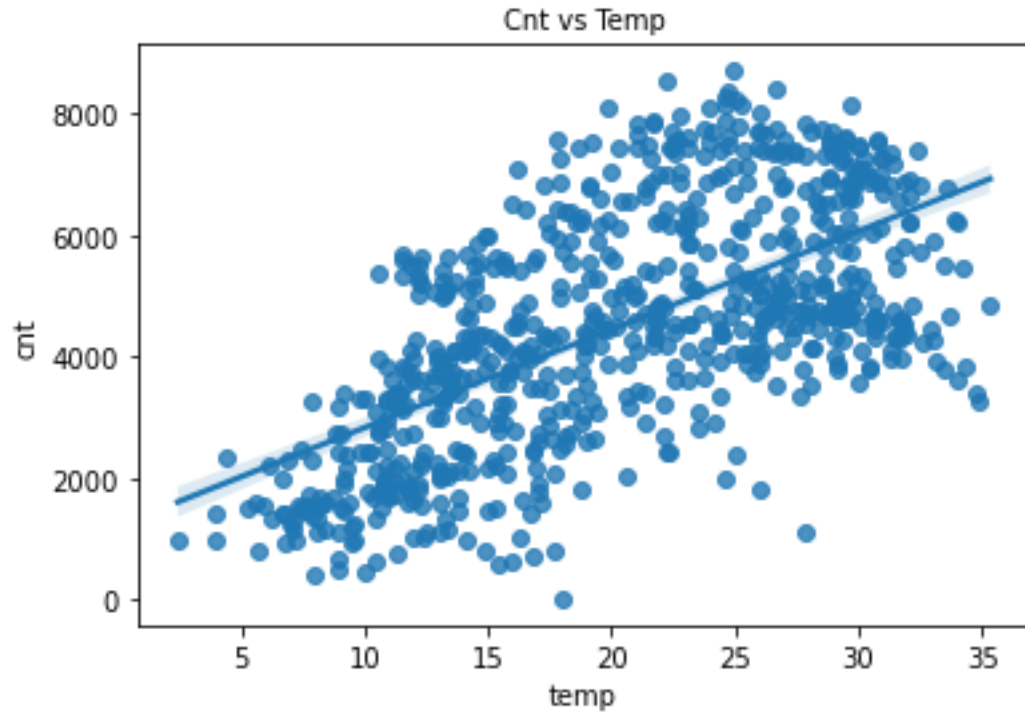
- 1.in fig1 : As we can see demand increase for 2019 when we compared with 2018
- 2.in fig2 : The bike are used more in regular days when we compared with holidays
- 3.in fig3 : Both are in same demand
- 4.in fig4 : In the month of may to oct the demand is very high due to summer season
- 5.in fig5 : Good weather means more demand
- 6.in fig6 : fall and summer season are having more demand
- 7.in fig7 : avg number of demand is in fixed range only

Heat map

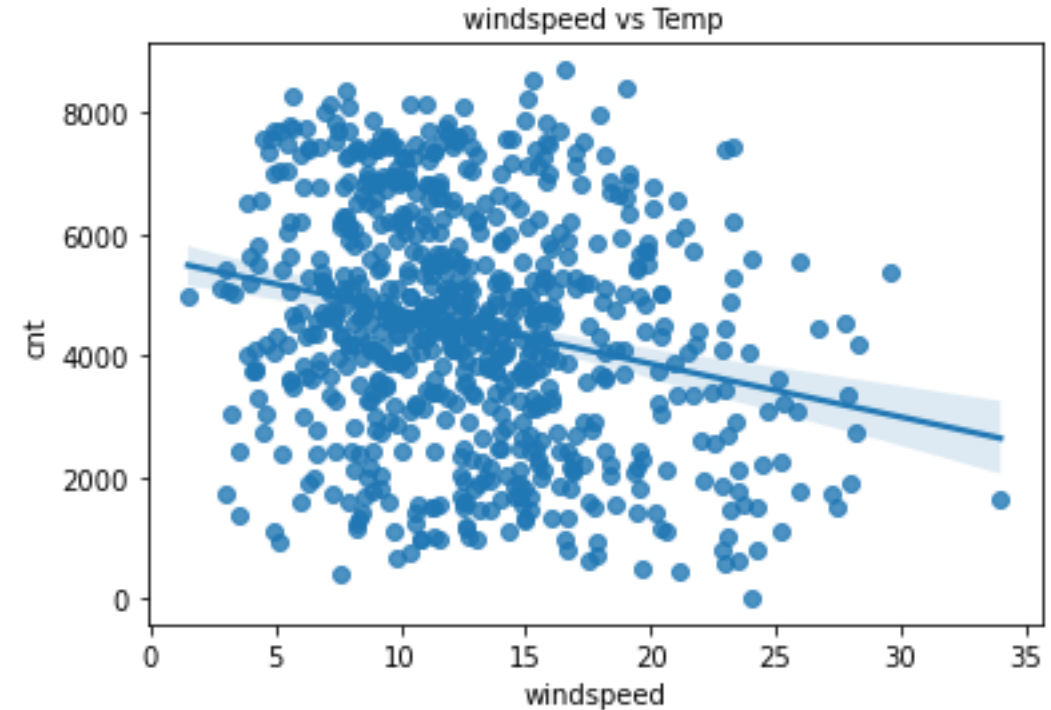


Heatmap correlation with data set

Scatter plot

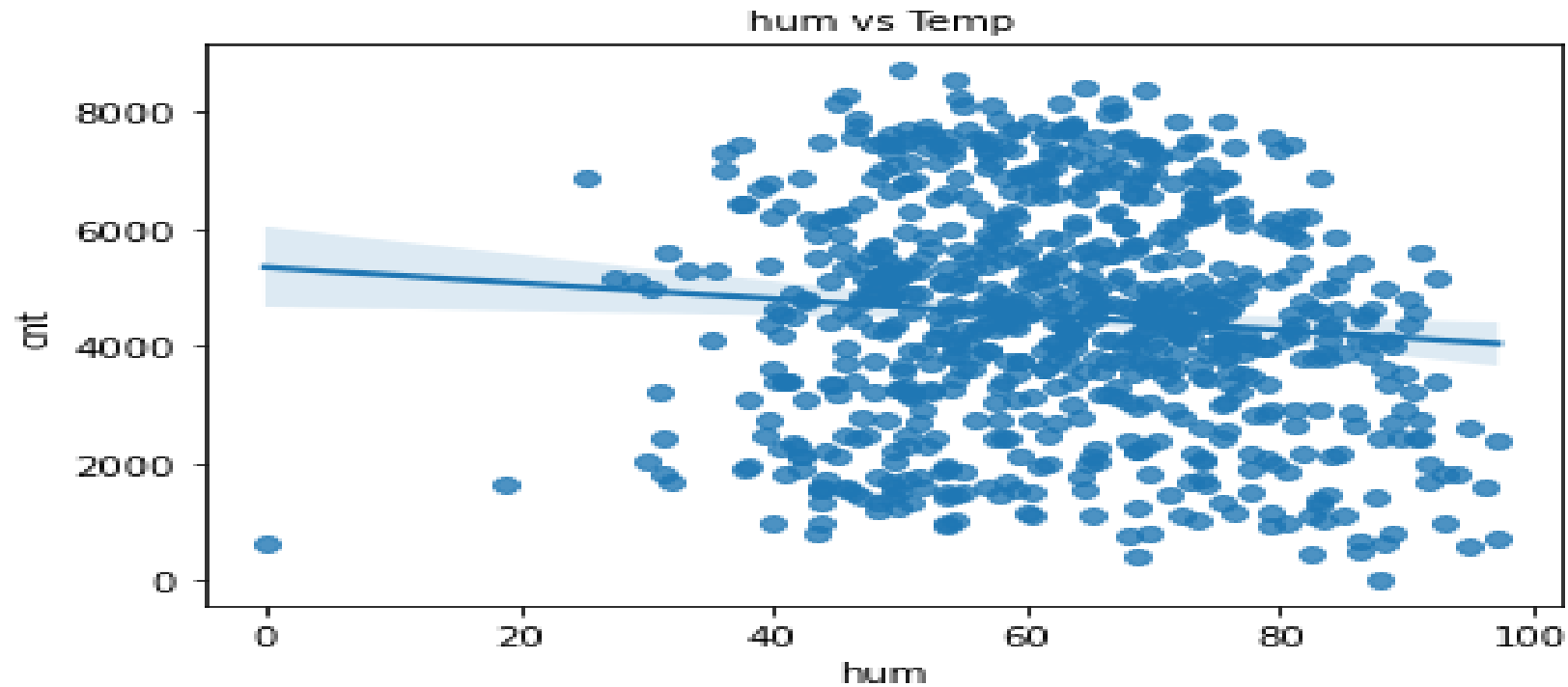


- Scatter plot for count vs temp.
- We can see that count is linearly increasing with temp indicating linear relation.



- Scatter plot for windspeed vs Temp.
- we can see count decreasing with increase in windspeed.

Scatter plot



- Scatter plot for Hum vs Temp.
- we can see count decreasing with increase in humidity.

Model

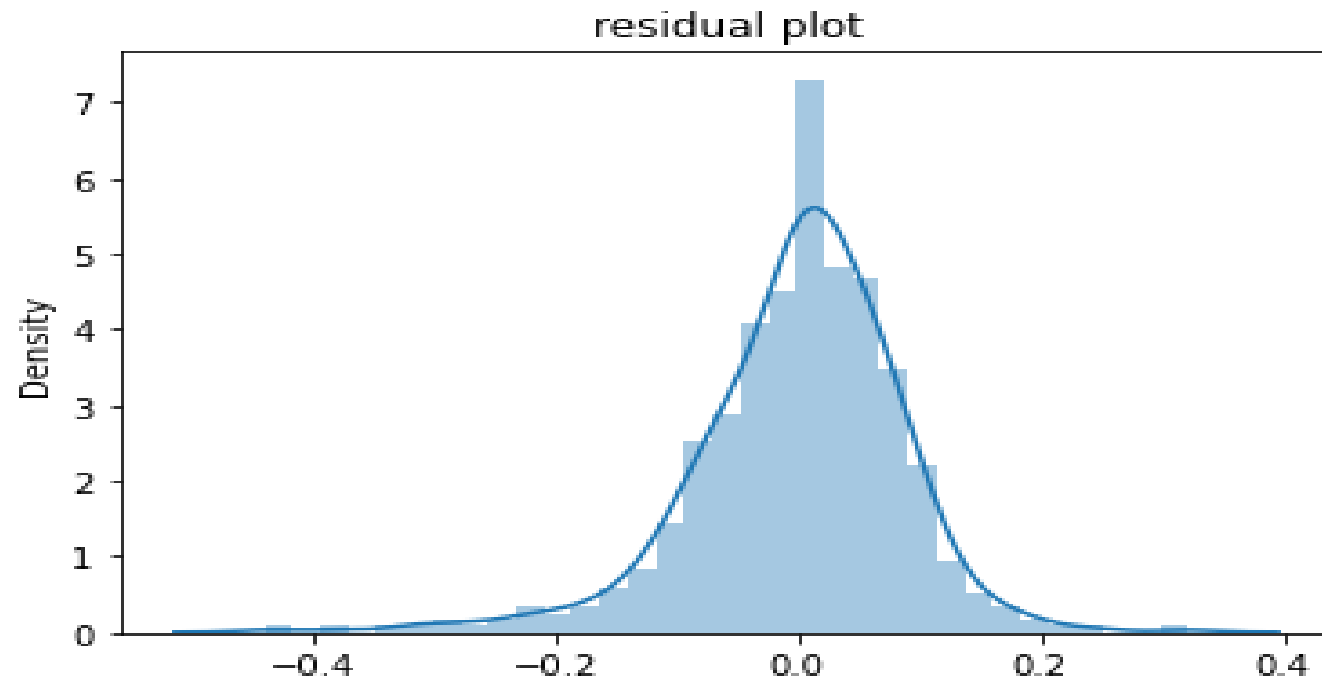
- In the final train model r-square is 0.85 \approx 85 %
- Adj. R-square is 0.84 \approx 84%

OLS Regression Results

Dep. Variable:		cnt	R-squared:		0.853
Model:		OLS	Adj. R-squared:		0.847
Method:		Least Squares	F-statistic:		149.3
Date:		Tue, 08 Feb 2022	Prob (F-statistic):		7.88e-190
Time:		20:40:39	Log-Likelihood:		529.69
No. Observations:		510	AIC:		-1019.
Df Residuals:		490	BIC:		-934.7
Df Model:		19			
Covariance Type:		nonrobust			

Residual analysis

- Errors are normally distributed here with mean 0. So everything seems to be fine



predict on test set

- The prediction came around 0.802 == 80%
- It is good model for build it.

predict on test set

```
|:  ► y_test_pred = lr_model.predict(X_test_sm)  
    r2_score(y_true=y_test, y_pred=y_test_pred)
```

```
: [73]: 0.8023857533285474
```

Conclusion

- Company should focus on expanding business during Spring.
- Company should focus on expanding business during may to oct
- Based on previous data the number are increased more when we compared with 2019.
- Number of users once situation comes back to normal as we can see in 2019.
- There would be less bookings during Light Snow or Rain & Holidays, they could probably use this time to service the bikes without having business impact.
- Temperature is the is good means business is good.