BIKE SHARING CASE STUDY

Ravi Pandey
Neha Bansal

Bike Sharing Case Study

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.

Business Objective:

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.

Bike Sharing Case Study

Observations:

There are no missing value and no duplicate data found in dataset.

Removing Columns:

- instant : index value
- dteday : We already have separated columns for month and year hence ignoring this deteday column
- Casual & Registered: : cnt is a dervied column for both casual and registerd hence we can ignore both columns

Univariate Analysis:

- Most of the bike booking happening in season 3
- Most of the bike booking month range between 6 to 8
- Bike booking is happening when there is holiday value=0
- Most of the Bike booking happening when thers is a working day
- Weekday is independent, All the days have marginly same count

Bike Sharing Case Study

Observations:

Bivariate Analysis:

- Over 5K booking is happening on the Season3, Season2 and Season 4 wherease we observe that less then 3.5k booking is happening on season 1
- Over 4k boking happening on the range between 4 to 10.
- Over 4k booking is only happening in weathersit 1.
- Most of the Bike booking happening when thers is a working day
- Weekday is independent, All the days have marginly same count

Correlation Analysis:

temp and atemp are highly corelated with each other

Bike Sharing Case Study

Final Linear Regression Model:

Feature Selection:

- atemp and temp are highly correlated and their VIF and p-value are also high, So dropped atemp
- Dropped Hum due to highest VIF
- dropped mnth_2, mnth_5, mnth_7, Mnth_10, mnth _12, Weekday_1, Weekday_4, Weekday_5 due to high p value

Observation:

- In Final model, there is very low Multicollinearity between the predictors and all p values are significant

```
print(checkVIF(x_new))
```

	Features	VIF
2	temp	4.88
1	workingday	4.16
3	windspeed	3.88
0	yr	1.97
7	weekday_6	1.69
4	season_2	1.60
8	weathersit_2	1.52
5	season_4	1.40
6	mnth_9	1.20
9	weathersit_3	1.09

OLS Regression Results

Dep. Variable:	cnt	R-squared:	0.824
Model:	OLS	Adj. R-squared:	0.820
Method:	Least Squares	F-statistic:	267.4
Date:	Wed, 09 Feb 2022	Prob (F-statistic):	2.07e-208
Time:	18:07:50	Log-Likelihood:	547.93
No. Observations:	584	AIC:	-1074.
Df Residuals:	573	BIC:	-1026.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0817	0.019	4.410	0.000	0.045	0.118
yr	0.2393	0.008	30.093	0.000	0.224	0.255
workingday	0.0443	0.011	4.102	0.000	0.023	0.066
temp	0.5584	0.019	29.489	0.000	0.521	0.596
windspeed	-0.1414	0.026	-5.473	0.000	-0.192	-0.091
season_2	0.0855	0.010	8.608	0.000	0.066	0.105
season_4	0.1372	0.010	13.364	0.000	0.117	0.157
mnth_9	0.0884	0.015	5.722	0.000	0.058	0.119
weekday_6	0.0622	0.014	4.394	0.000	0.034	0.090
weathersit_2	-0.0767	0.009	-8.850	0.000	-0.094	-0.060
weathersit_3	-0.2947	0.025	-11.817	0.000	-0.344	-0.246

Omnibus:	75.638	Durbin-Watson:	1.877
Prob(Omnibus):	0.000	Jarque-Bera (JB):	153.849
Skew:	-0.744	Prob(JB):	3.91e-34
Kurtosis:	5.027	Cond. No.	12.2

Bike Sharing Case Study

Final Linear Regression Model:

- Hypothesis testing states that:

$$H0 : B1 = B2 = \ldots = Bn = 0$$

$$H1 : Bi != 0$$

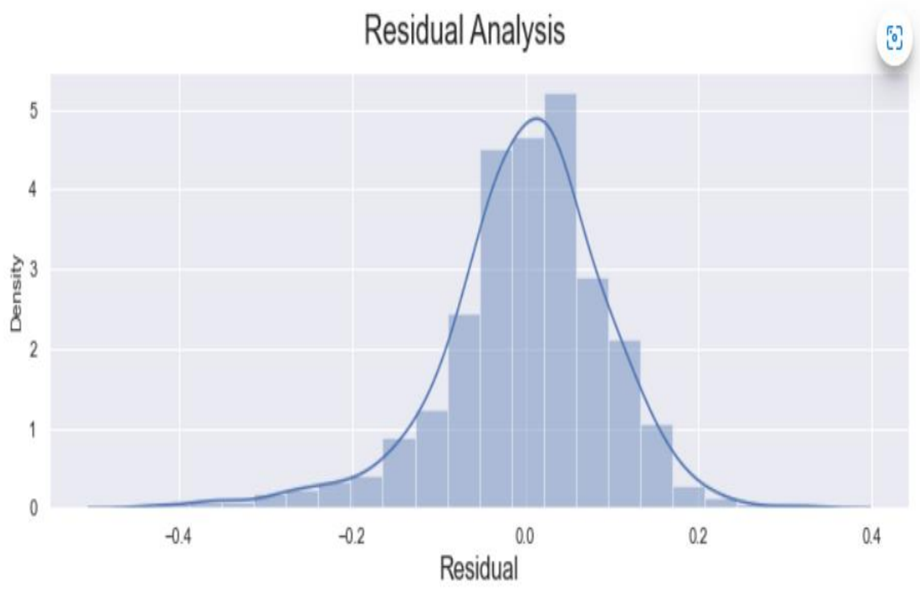
Our coef are not equal to 0. Thus, **We REJECT the NULL HYPOTHESIS**

- Linear Regression Assumption:
  Residuals are normally distributed, So assumption for Linear Regression is Valid

- Goodness of Model:
  Adjusted R^2 for both Train and Test are near to each other, So we can state that model is "GOOD"

$$TrainModel\,R^2 = 0.824$$
$$TrainModel\,Adjusted\,R^2 = 0.820$$

$$Predicted\,R^2 = 0.826$$
$$Predicted\,Adjusted\,R^2 = 0.813$$



Thank You