

# Bike Sharing

BY – ROHIT PANDEY



## Problem Statement

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 the current market situation. In such as attempt, BoomBikes aspires to understand the demand for shared bikes among the people. 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 profit.



## Business Goal

The Company wants to know –

- i). Variables that are significant in predicting the demand.
- ii). How well those variables describe the bike demand.

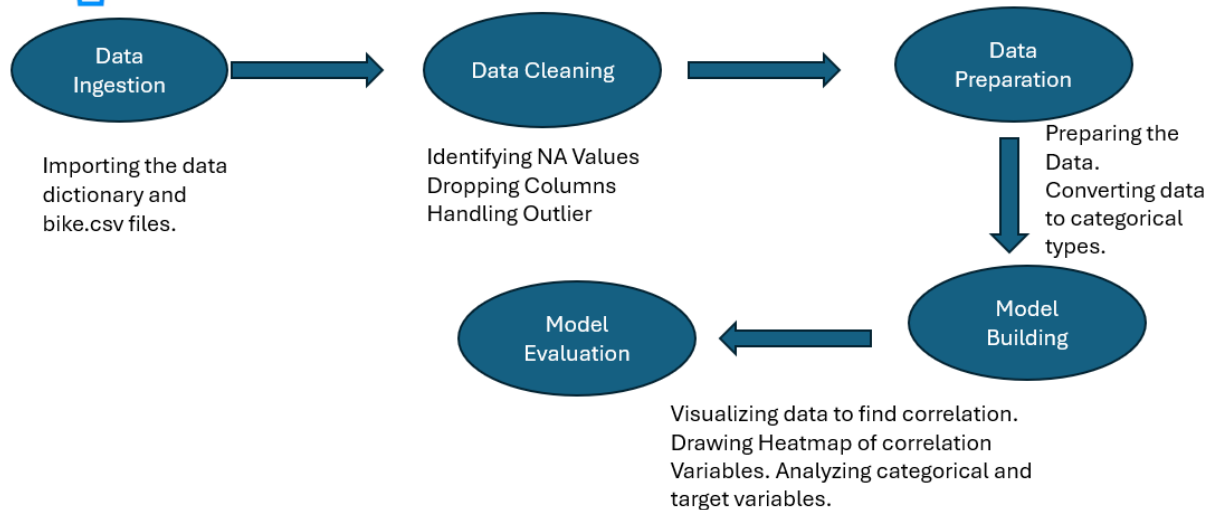
## Solution Context

For this business problem we need to understand the data. The data anomalies needs to be rectified. We need to model the demand of the bikes with the variables identified.

- We need to prepare the data by analyzing the datasets and data dictionary.
- Prepare the data by Data by identifying variables.
- Build the data model to identify which variables are significant for the demand of the shared bikes.
- Evaluate the data model to meet the demand level and customer expectations.



## Approach



# Analysis



## Dropping the columns

	season	yr	mnth	holiday	weekday	workingday	weathersit	temp
season	1.000000e+00	-2.248195e-17	8.310321e-01	-0.010868	-0.003081	0.013762	0.021306	0.333361
yr	-2.248195e-17	1.000000e+00	-2.470170e-17	0.008195	-0.005466	-0.002945	-0.050322	0.048789
mnth	8.310321e-01	-2.470170e-17	1.000000e+00	0.018905	0.009523	-0.004688	0.045613	0.219083
holiday	-1.086804e-02	8.195345e-03	1.890483e-02	1.000000	-0.101962	-0.252948	-0.034395	-0.028764
weekday	-3.081199e-03	-5.466399e-03	9.522969e-03	-0.101962	1.000000	0.035800	0.031112	-0.000168
workingday	1.376178e-02	-2.945396e-03	-4.687953e-03	-0.252948	0.035800	1.000000	0.060236	0.053470
weathersit	2.130636e-02	-5.032247e-02	4.561335e-02	-0.034395	0.031112	0.060236	1.000000	-0.119503
temp	3.333607e-01	4.878919e-02	2.190833e-01	-0.028764	-0.000168	0.053470	-0.119503	1.000000
atemp	3.420139e-01	4.721519e-02	2.264302e-01	-0.032703	-0.007539	0.052940	-0.120559	0.991696
hum	2.082196e-01	-1.125471e-01	2.249368e-01	-0.015662	-0.052290	0.023202	0.090277	0.128565
windspeed	-2.296059e-01	-1.162435e-02	-2.080131e-01	0.006257	0.014283	-0.018666	0.039769	-0.158186
cnt	4.045838e-01	5.697285e-01	2.781909e-01	-0.068764	0.067534	0.062542	-0.295929	0.627044
days_old	4.135867e-01	8.660262e-01	4.982507e-01	0.016275	-0.000023	-0.004619	-0.021862	0.150738

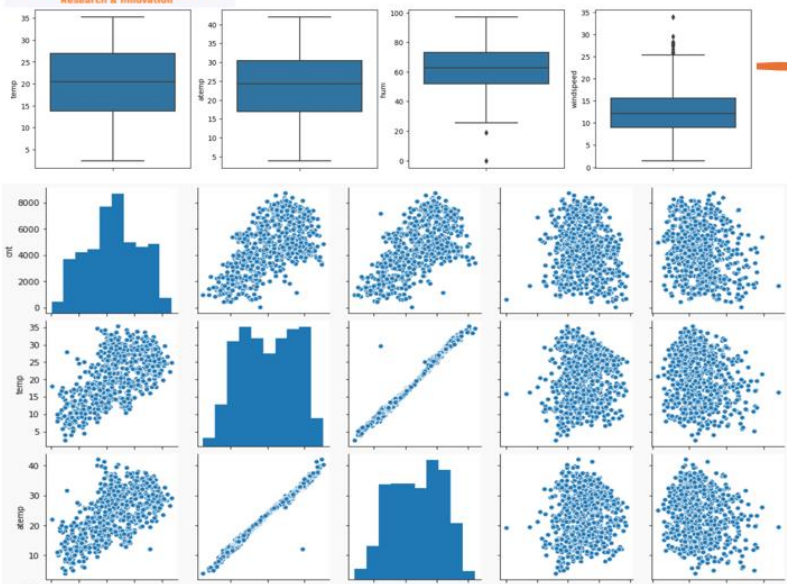
## Handling Missing Values

```
instant      0
dteday      0
season      0
yr          0
mnth       0
holiday     0
weekday     0
workingday  0
weathersit  0
temp        0
atemp       0
hum         0
windspeed  0
casual      0
registered  0
cnt         0
dtype: int64
```

## Handling Outliers

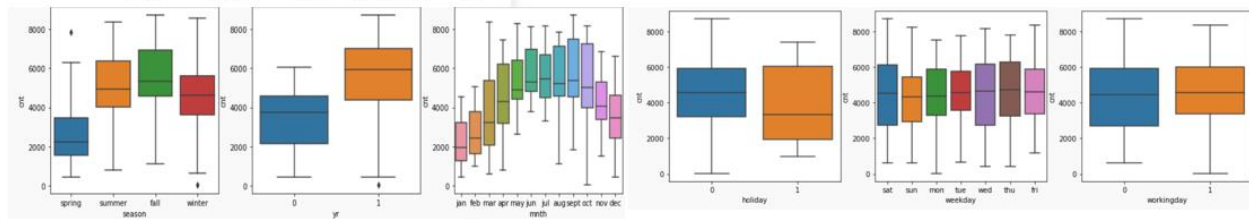
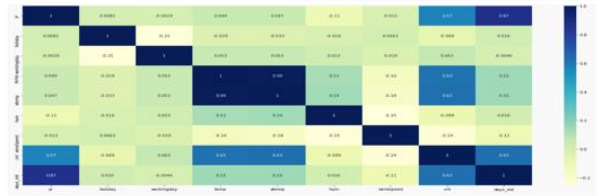
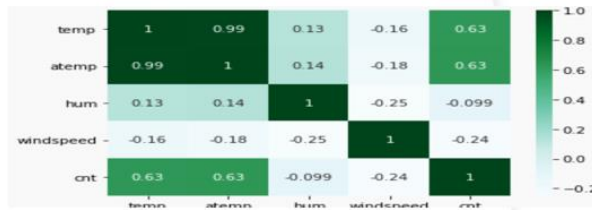
```
instant      730
dteday      730
season       4
yr           2
mnth        12
holiday      2
weekday      7
workingday   2
weathersit   3
temp        498
atemp       689
hum         594
windspeed   649
casual      605
registered   678
cnt         695
dtype: int64
```

# Identifying Linear Regression

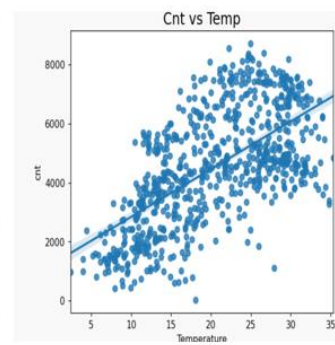
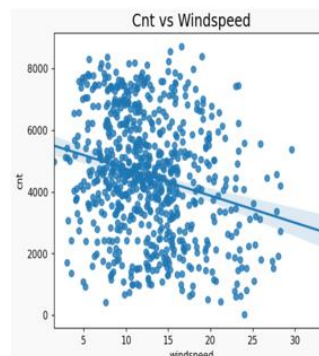




## Heatmap of Correlation



## Categorical Vs Target Variables



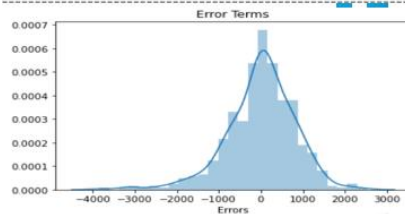




# Models



OLS Regression Results			
Dep. Variable:	cnt	R-squared:	0.836
Model:	OLS	Adj. R-squared:	0.831
Method:	Least Squares	F-statistic:	167.9
Date:	Sun, 24 Jan 2021	Prob (F-statistic):	4.52e-183
Time:	21:24:26	Log-Likelihood:	-4131.1
No. Observations:	511	AIC:	8294.
Df Residuals:	495	BIC:	8362.
Df Model:	15		
Covariance Type:	nonrobust		



OLS Regression Results			
Dep. Variable:	cnt	R-squared:	0.836
Model:	OLS	Adj. R-squared:	0.831
Method:	Least Squares	F-statistic:	180.2
Date:	Sun, 24 Jan 2021	Prob (F-statistic):	3.59e-184
Time:	21:24:28	Log-Likelihood:	-4131.2
No. Observations:	511	AIC:	8292.
Df Residuals:	496	BIC:	8356.
Df Model:	14		
Covariance Type:	nonrobust		

## Conclusion

Temperature had positive correlation with demand of bikes.

On holidays there is decrease in demand. There is continuous growth in demand till June. September month has highest.

With increase in humidity demand gets decreases.

Count has correlation with Temp and Atemp.

Difference between R-Square and Adjusted R-Square values are minimal hence, it would be safe to say no additional parameters can be removed.

Variables which can be used to predict the demand are holiday, temp, windspeed, hum, season, months

