upGrad

#LifeKoKaroLift

Bike Sharing Assignment

- By Veena Bind

16-03-2022

upGrad

- The objectives for the assignment
- Model Building
- Model Evaluation

What is BoomBikes ?

A bike-sharing system in which bikes are made available for shared use to individuals on a short term basis for a price or free. It 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: The objective is 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.
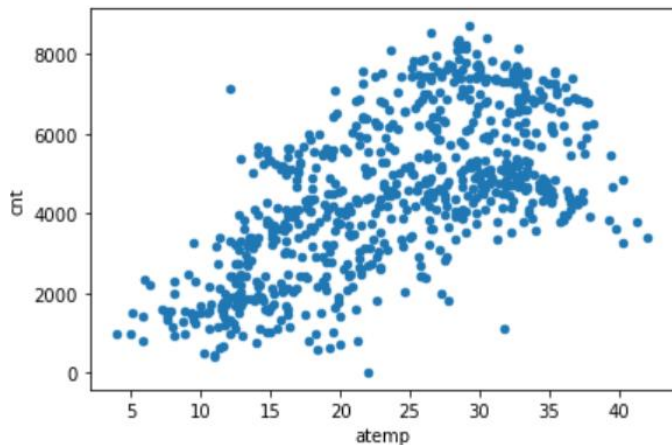
Requirements

- Which variables are significant in predicting the demand for shared bikes.
- How well those variables describe the bike demands.

What you need to do?

- Create a linear model that describe the effect of various features on demand.
- Th model should be interpretable so that the management can understand it.

Scatter plot of 'cnt' and 'atemp'

Correlation of independent variable with dependent variable 'cnt':

```
season          0.404584
yr              0.569728
mnth            0.278191
holiday        -0.068764
weekday         0.067534
workingday      0.062542
weathersit     -0.295929
temp            0.627044
atemp           0.630685
hum            -0.098543
windspeed      -0.235132
casual          0.672123
registered      0.945411
```

Modelling building:

Final params:

```
const         -0.005784
yr             0.233696
workingday     0.054355
atemp          0.540858
windspeed     -0.133705
Mon            0.064770
Light Rain    -0.196500
Sunny          0.083696
Aug            0.064985
Sep            0.116346
summer         0.098699
winter         0.131978
dtype: float64
```

Modelling building:

Final summary:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   cnt   R-squared:                       0.833
Model:                           OLS   Adj. R-squared:                  0.830
Method:                Least Squares   F-statistic:                     226.4
Date:               Wed, 16 Mar 2022   Prob (F-statistic):          9.15e-186
Time:                       21:11:56   Log-Likelihood:                 495.82
No. Observations:                510   AIC:                            -967.6
Df Residuals:                    498   BIC:                            -916.8
Df Model:                         11
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0058      0.019     -0.299      0.765      -0.044       0.032
yr             0.2337      0.008     28.251      0.000       0.217       0.250
workingday     0.0544      0.011      4.829      0.000       0.032       0.076
atemp          0.5409      0.023     23.332      0.000       0.495       0.586
windspeed     -0.1337      0.025     -5.264      0.000      -0.184      -0.084
Mon            0.0648      0.014      4.468      0.000       0.036       0.093
Light Rain    -0.1965      0.025     -7.772      0.000      -0.246      -0.147
Sunny          0.0837      0.009      9.460      0.000       0.066       0.101
Aug            0.0650      0.017      3.935      0.000       0.033       0.097
Sep            0.1163      0.016      7.054      0.000       0.084       0.149
summer         0.0987      0.011      8.839      0.000       0.077       0.121
winter         0.1320      0.011     12.345      0.000       0.111       0.153
==============================================================================
Omnibus:                      59.652   Durbin-Watson:                   2.043
Prob(Omnibus):                 0.000   Jarque-Bera (JB):              158.202
Skew:                         -0.582   Prob(JB):                     4.44e-35
Kurtosis:                      5.468   Cond. No.                         12.7
==============================================================================
```
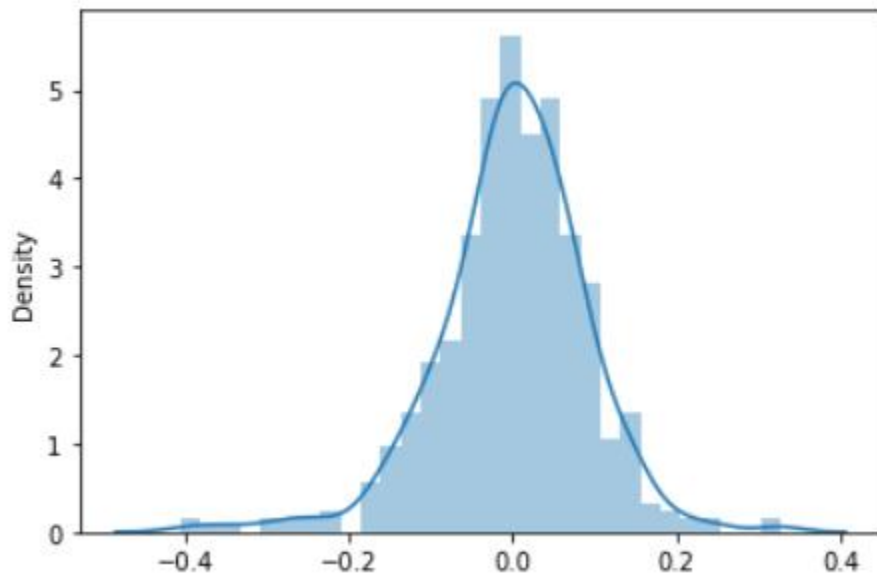
Modelling building:

Final VIF:

	features	VIF
2	atemp	6.87
1	workingday	3.95
3	windspeed	3.58
6	Sunny	2.71
0	yr	2.02
9	summer	1.82
4	Mon	1.67
7	Aug	1.55
10	winter	1.49
8	Sep	1.30
5	Light Rain	1.12

Residual Analysis on train data set:

Observations:

- R squared of train data set - 0.833
- R squared of test data set - 0.799
- Difference between train data set and test data set is ~0.034
- ~0.034 or 3.4% is less than 5% or 0.05. Hence it can be said it is a good model.

Conclusions:

- It was observed feel-like temparate (atemp), and windspeed are predictor variable for bike rentals.
- Based on day of week, it was observed "monday" have high bike rentals.
- Based on month of the year - August and September seems more favorable for bike rentals.
- Based on season - 'summer' and 'winter' are favorable for bike rentals.
- Based on weather situation - Sunny *(Clear, Few clouds, Partly cloudy, Partly cloudy*) &   Light Rain *(Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds)* are favorable season for bike rentals.
- Working days are also predictive variable for bike rentals

upGrad
#LifeKoKaroLift

Thank You!