

# Mycar Dream

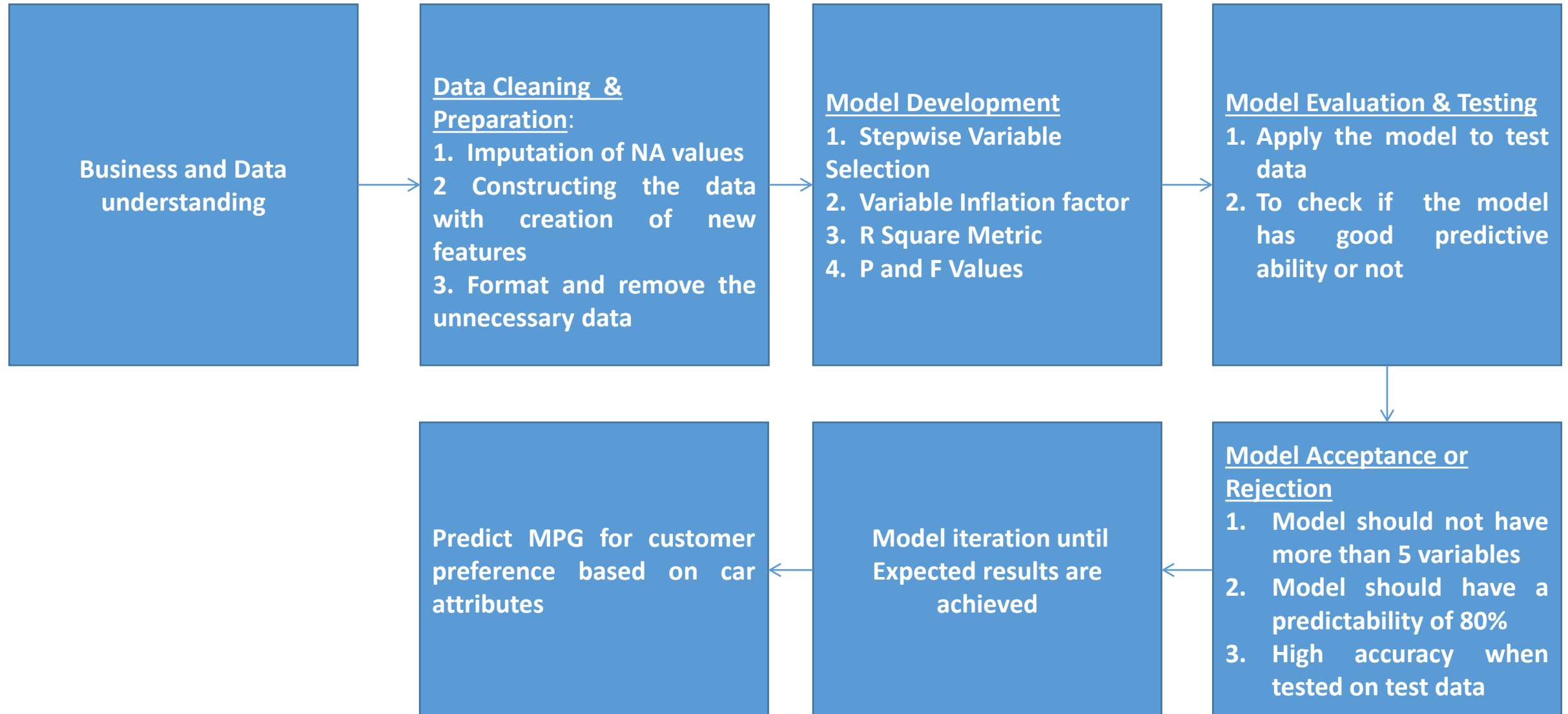
## **FINDING THE ATTRIBUTES INFLUENCING THE TENDENCY OF DEFAULTING**

### **AIM:**

- To automate the process of predicting the car mileage for customer preferences, based on the dataset of car features and attributes obtained by Market surveys
- To predict the city-cycle fuel consumption in miles per gallon, in terms of 3 multivalued discrete and 5 continuous attributes

**Business Objective:** The business model of the company is solely based on consumer interest, aiming to provide the most appropriate car to their clients and hence maximise the customer satisfaction.

# Problem solving methodology



## Goals of the Analysis

To develop a predictive model which can follow these three constraints thoroughly:

- The model should not contain more than **5 variables**.
- According to the business needs, **VIF to be less than 2**.
- The model should be highly predictive in nature i.e it should show **80% ( $R^2$ )** of accuracy.

# Predictor Variables

The following are the Predictor variables included in the model:-

<u>Model without transformed version of Predictor</u>	<u>Model with transformed version of Predictor</u>
Weight	Weight
Pontiac	$(\text{Weight})^2$
Cylinders (Binned 2 & 3)	Model Year (Moderately New)
Model Year (Moderately New)	Model Year (New)
Model Year (New)	-

## **R<sup>2</sup> and P Values of the Model without Transformed Variable**

<u>Data</u>	<u>Multiple R<sup>2</sup></u>	<u>Adjusted R<sup>2</sup></u>	P-Value
Train	81.8%	81.46%	2.2e-16
Test	81.68%	80.88%	2.2e-16

**The R<sup>2</sup> for Predicted and Actual MPG values of Test data is 81.36%**

**The predictive ability of the Model is 81.36%**

## **R<sup>2</sup> and P Values of the Model with Transformed Variable**

**The R squared value on the train dataset is 84.43%**

**The R squared value on the test dataset is 85.25%**

**The R squared value for predicted and actual values of MPG for test data is 85.13%**

## Train data Results

```
Call:
lm(formula = MPG ~ Weight + I(Weight^2) + carmileage.Model_yearnew +
    carmileage.Model_yearModeratly.New, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-8.2893 -1.8908  0.1682  1.4559 12.7306

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      25.3657    0.3355   75.595 <2e-16 ***
Weight          -6.4836    0.2160  -30.011 <2e-16 ***
I(Weight^2)       1.8560    0.2003   9.268  <2e-16 ***
carmileage.Model_yearnew -7.1702    0.4710  -15.223 <2e-16 ***
carmileage.Model_yearModeratly.New -4.8714    0.4564  -10.673 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.059 on 273 degrees of freedom
Multiple R-squared:  0.8443,    Adjusted R-squared:  0.8421
F-statistic: 370.2 on 4 and 273 DF,  p-value: < 2.2e-16
```

## Test data Results

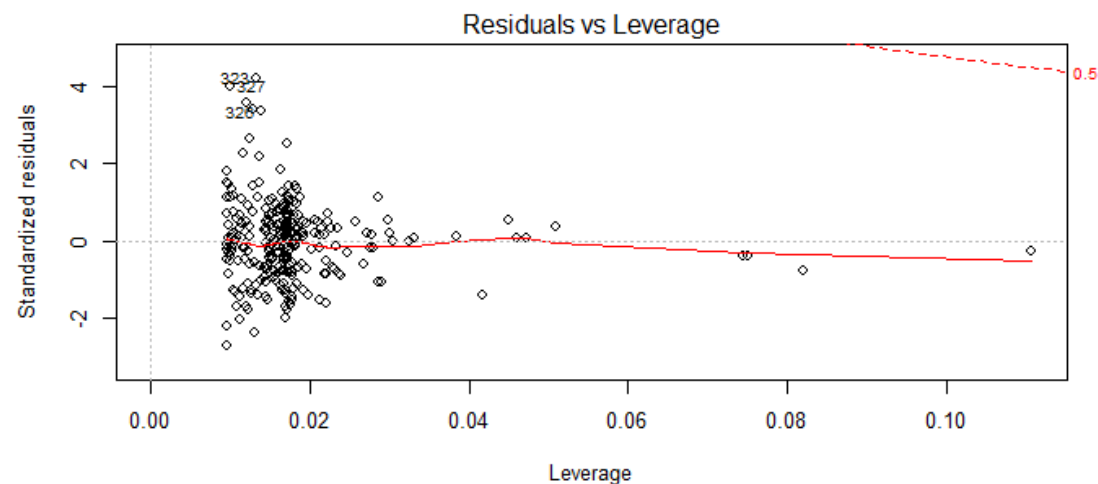
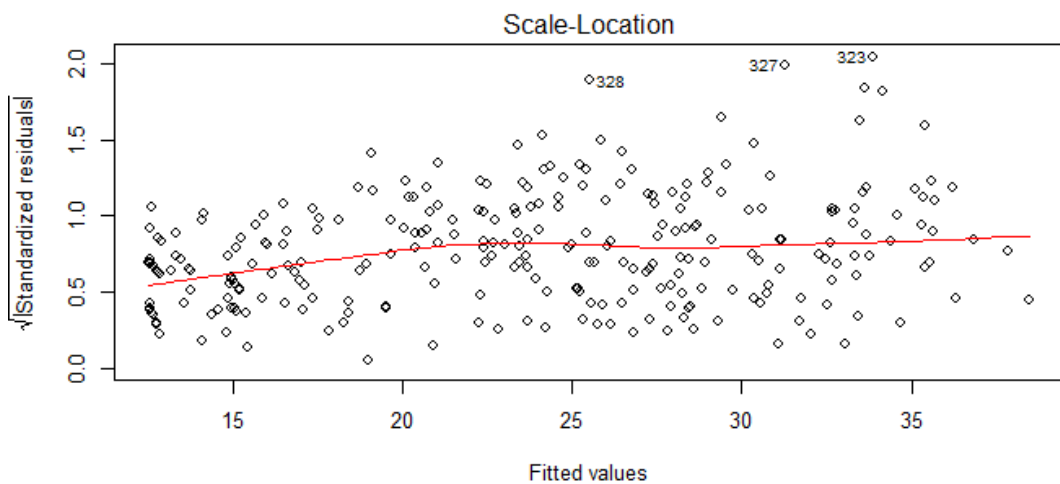
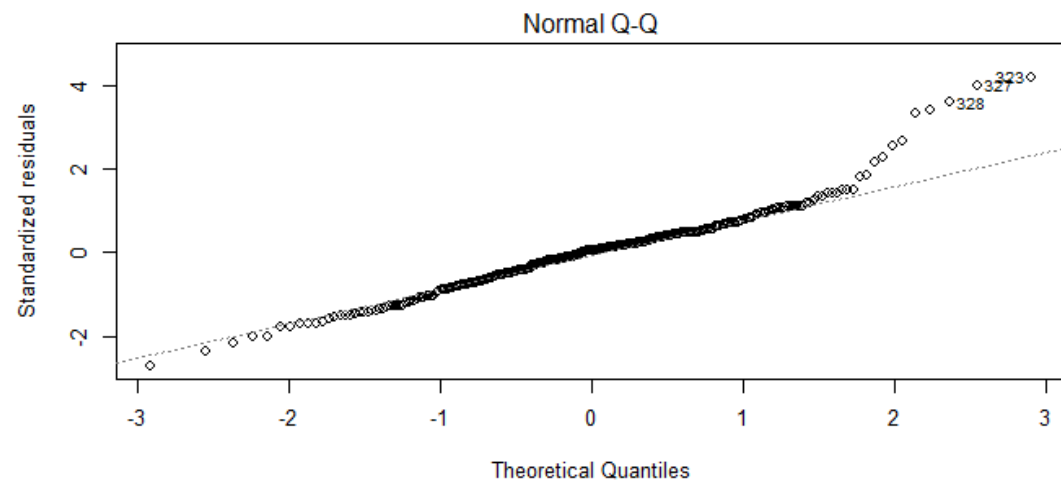
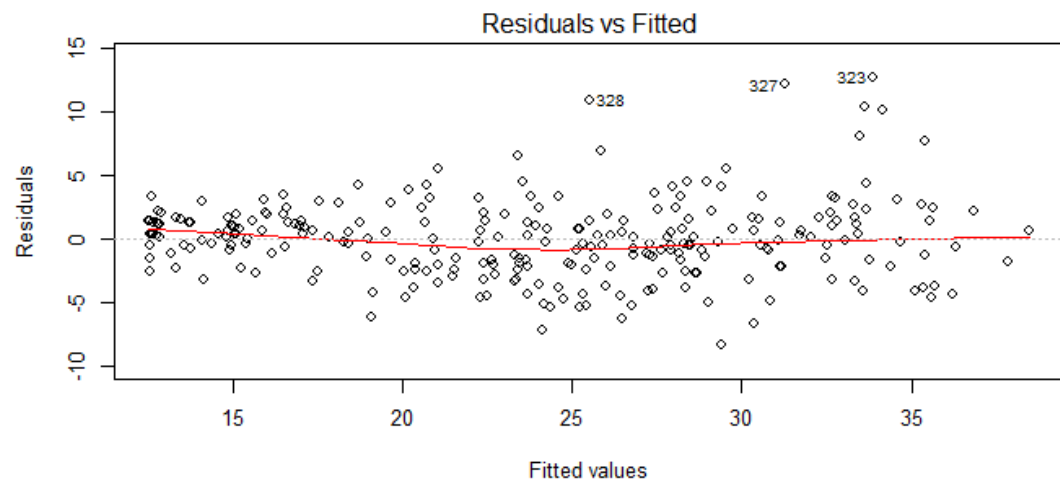
```
Call:
lm(formula = MPG ~ Weight + I(Weight^2) + carmileage.Model_yearnew +
    carmileage.Model_yearModeratly.New, data = test)

Residuals:
    Min       1Q   Median       3Q      Max
-8.6105 -2.0645 -0.0404  1.8724 13.0969

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      25.2524    0.5593   45.151 < 2e-16 ***
Weight          -6.7308    0.3351  -20.089 < 2e-16 ***
I(Weight^2)       1.7955    0.2966   6.054 1.81e-08 ***
carmileage.Model_yearnew -7.1631    0.7519  -9.527 3.27e-16 ***
carmileage.Model_yearModeratly.New -4.4295    0.7288  -6.078 1.63e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.206 on 115 degrees of freedom
Multiple R-squared:  0.8525,    Adjusted R-squared:  0.8474
F-statistic: 166.2 on 4 and 115 DF,  p-value: < 2.2e-16
```

# Residual Vs Fitted values of Final Model





## Recommendation

- The model depends on Weight and Year to a large extent for predicting mileage
- The prediction ability of the model is 85.13%. More data and attributes are required for getting better results of prediction
- The Car Company name which is included in the model is Pontiac. Instead of company name, attributes of car manufactured by the company need to be studied and included in the model for making better prediction.
- If the constraints of VIF less than 2 is not considered than better prediction results are obtained. If better prediction results are expected than the constraint of VIF to be less than 2 may be removed.

## Conclusion

- The model does not consist of more than 5 variables
- The VIF of all variables is less than 2
- The R squared value on the train dataset is 84.43%
- The R squared value on the test dataset is 85.25%
- The R squared value for predicted and actual values of MPG for test data is 85.13%
- As the model has high predictive ability it is accepted