**TRANSFORMATIONS**

Instructions:

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: Navya.P**

**Batch Id: WDEO171220**

**Topic: Preliminaries for Data Analysis**

**Problem Statement:**

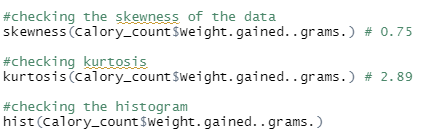
Everything will revolve around the data in Analytics world. Proper data will help you to make useful predictions which improve your business. Sometimes the usage of original data as it is does not help to have accurate solutions. It is needed to convert the data from one form to another form to have better predictions. Explore on various techniques to transform the data for better model performance. you can go through this link:

<https://360digitmg.com/mindmap-data-science>

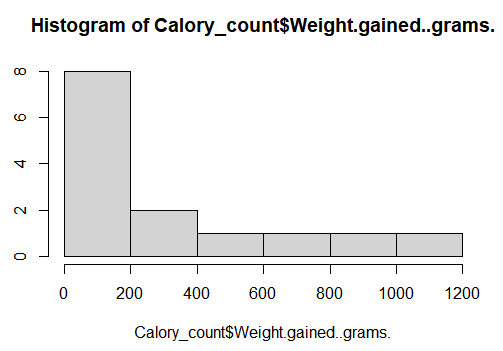
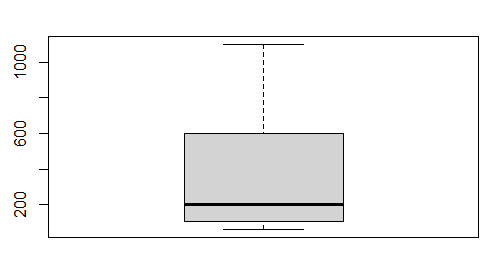
1. Prepare the dataset by performing the preprocessing techniques, to have the data which improve model performance.

The given Calory\_count data set has two columns.

For Weight gained (grams) column, the initial data has skewness of 0.75 and kurtosis of 2.89.

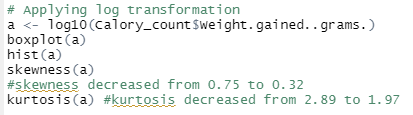


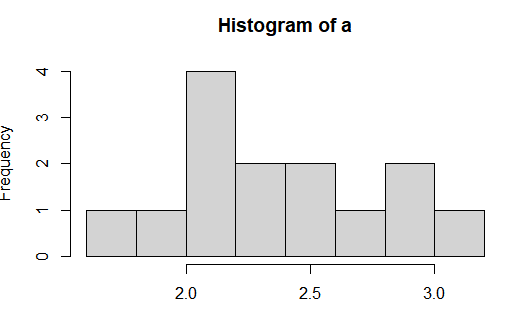
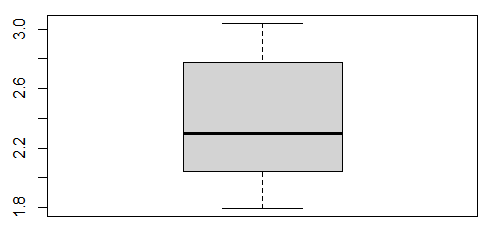


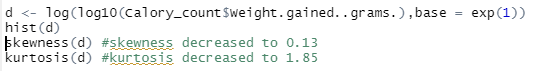
The different transformation techniques are applied on the dataset such that skewness reduces near to zero and the histogram replicates the normal distribution curve.

Out of all transformations done and checked.. log transformation has shown the skewness and kurtosis near to normal distribution.

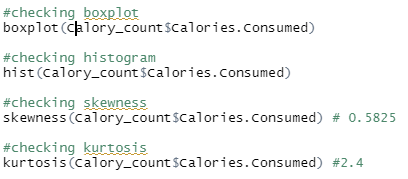


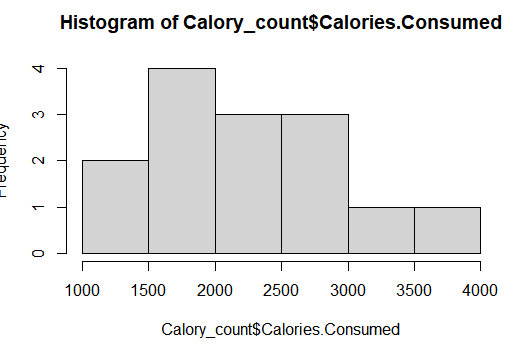
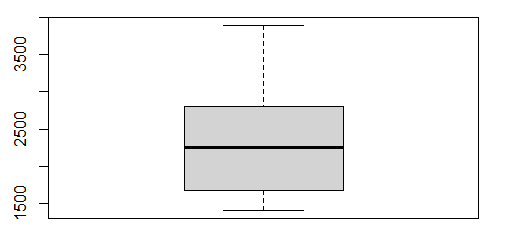
 

Applying double log, gave the following result.



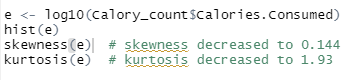
Checking the calories\_Count – calories\_Consumed column

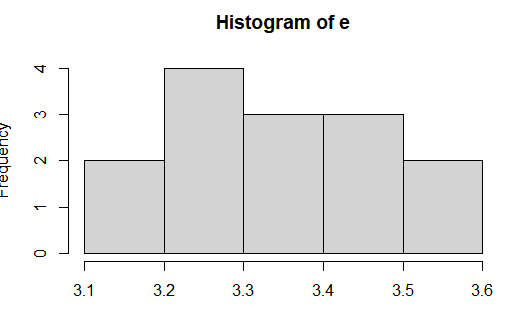
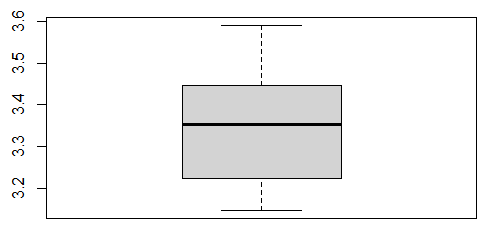


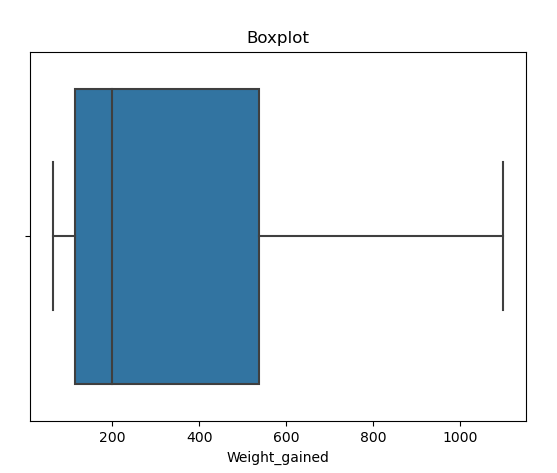
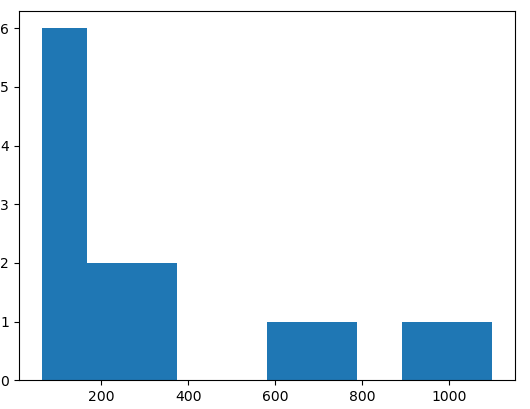
The different transformation techniques are applied on the dataset such that skewness reduces near to zero and the histogram replicates the normal distribution curve.

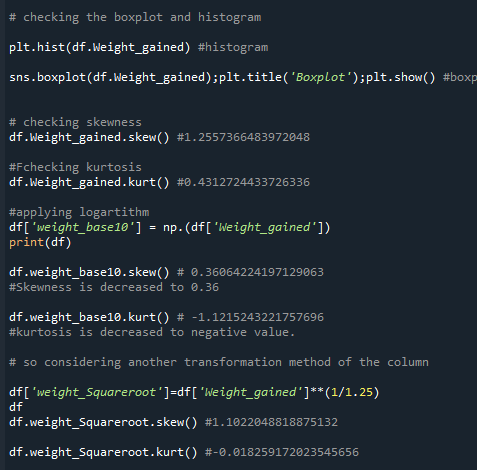
Out of all transformations done and checked.. log transformation has shown the skewness and kurtosis near to normal distribution.

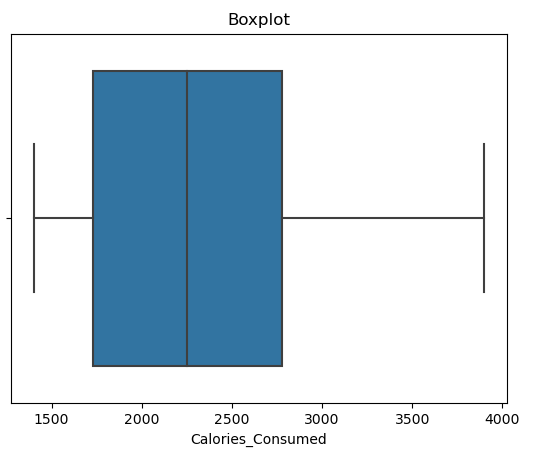
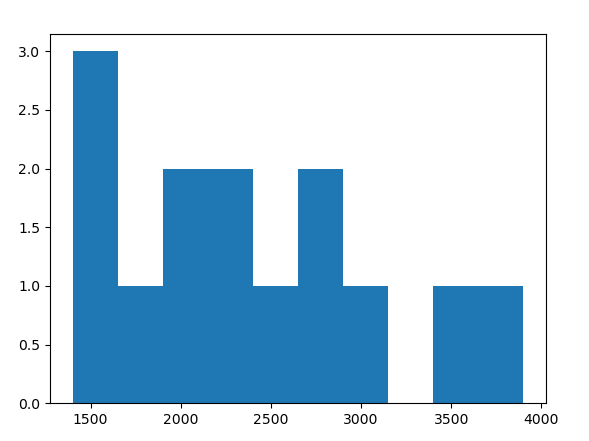


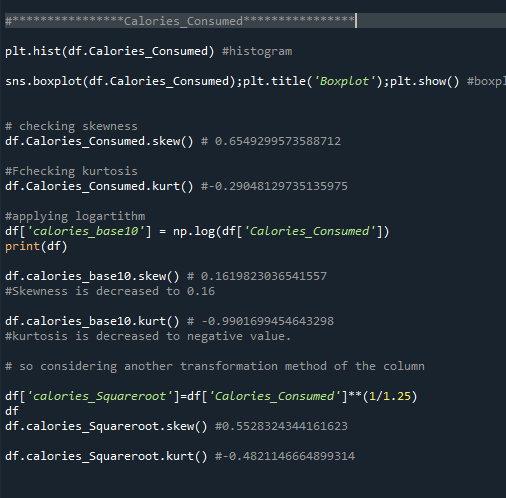
 











**Hints:**

For each assignment, the solution should be submitted in the below format

1. Work on each feature to create a data dictionary as displayed in the image displayed below:
2. Hint: Refer to calories\_consumed.csv dataset
3. Research and perform all possible steps for obtaining solution
4. All the codes (executable programs) should execute without errors
5. Code modularization should be followed
6. Each line of code should have comments explaining the logic and why you are using that function