**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: ANANDAKRISHNAN K V**

**Batch Id: 19042021**

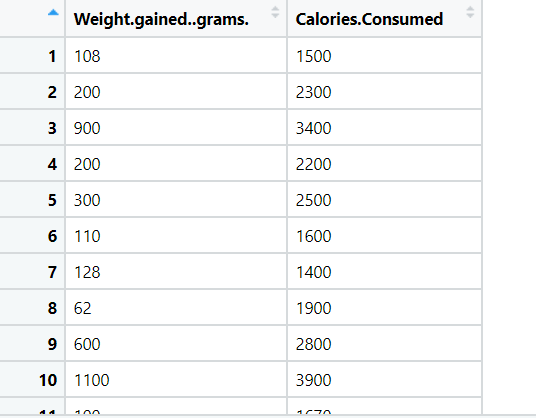
**Topic: Preliminaries for Data Analysis**

**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.



**Ans:-**

**Python code:-**

####################### Transformation techniques ###################

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats as stats

import pylab

### import data set "calories\_consumed" as Dataframe ##

p = pd.DataFrame(calories\_consumedcsv)

p0 = pd.DataFrame(p.iloc[:,0])

p0 = p.iloc[:,0]

p1 = p.iloc[:,1]

p0.skew() ## skewness

p0.kurt() ## excess kurtosis

## custom function for calling qqplot # histogram for a certain variable

def diagnostic\_plots(variable):

plt.figure(figsize=(15,6))

plt.subplot(1,2,1)

plt.hist(variable,bins=10)

plt.subplot(1,2,2)

stats.probplot(variable,dist='norm',plot=plt)

plt.show()

###### Analysis on weight gained in grams ;; ordinary performance curve #####

diagnostic\_plots(p0)

print("skewnes is:",p0.skew(), " ;; excess kurtosis is:",p0.kurt()) ## skewness && excess kurtosis

###### Analysis on calories consumed ;; ordinary performance curve #####

diagnostic\_plots(p1)

print("skewnes is:",p1.skew(), " ;; excess kurtosis is:",p1.kurt()) ## skewness && excess kurtosis

###### Analysis on weight gained in grams ;; logarithmic transformation #####

p0\_log = np.log(p0) ### if p0 is zero then log(0) is not work. thats why add 1

diagnostic\_plots(p0\_log)

print("skewnes is:",p0\_log.skew(), " ;; excess kurtosis is:",p0\_log.kurt()) ## skewness && excess kurtosis

###### Analysis on calories consumed ;; logarithmic transformation #####

p1\_log = np.log(p1) ### if p0 is zero then log(0) is not work. thats why add 1

diagnostic\_plots(p1\_log)

print("skewnes is:",p1\_log.skew(), " ;; excess kurtosis is:",p1\_log.kurt()) ## skewness && excess kurtosis

###### Analysis on weight gained in grams ;; reciprocal transformation #####

p0\_rec = 1/(p0+1) ### if p0 is zero then 1/0 is not work. thats why add 1

diagnostic\_plots(p0\_rec)

print("skewnes is:",p0\_rec.skew(), " ;; excess kurtosis is:",p0\_rec.kurt()) ## skewness && excess kurtosis

###### Analysis on calories consumed ;; reciprocal transformation #####

p1\_rec = 1/(p1+1) ### if p0 is zero then 1/0 is not work. thats why add 1

diagnostic\_plots(p1\_rec)

print("skewnes is:",p1\_rec.skew(), " ;; excess kurtosis is:",p1\_rec.kurt()) ## skewness && excess kurtosis

###### Analysis on weight gained in grams ;; square root transformation #####

p0\_sqrrt = np.sqrt(p0)

diagnostic\_plots(p0\_sqrrt)

print("skewnes is:",p0\_sqrrt.skew(), " ;; excess kurtosis is:",p0\_sqrrt.kurt()) ## skewness && excess kurtosis

###### Analysis on calories consumed ;; square root transformation #####

p1\_sqrrt = np.sqrt(p1)

diagnostic\_plots(p1\_sqrrt)

print("skewnes is:",p1\_sqrrt.skew(), " ;; excess kurtosis is:",p1\_sqrrt.kurt()) ## skewness && excess kurtosis

###### Analysis on weight gained in grams ;; exponential transformation #####

p0\_exp = p0\*\*(1/5) ### p0\*\*(1/5) performs slmost similar performance as exponential function

diagnostic\_plots(p0\_exp)

print("skewnes is:",p0\_exp.skew(), " ;; excess kurtosis is:",p0\_exp.kurt()) ## skewness && excess kurtosis

###### Analysis on calories consumed ;; exponential transformation #####

p1\_exp = p1\*\*(1/5) ### p0\*\*(1/5) performs slmost similar performance as exponential function

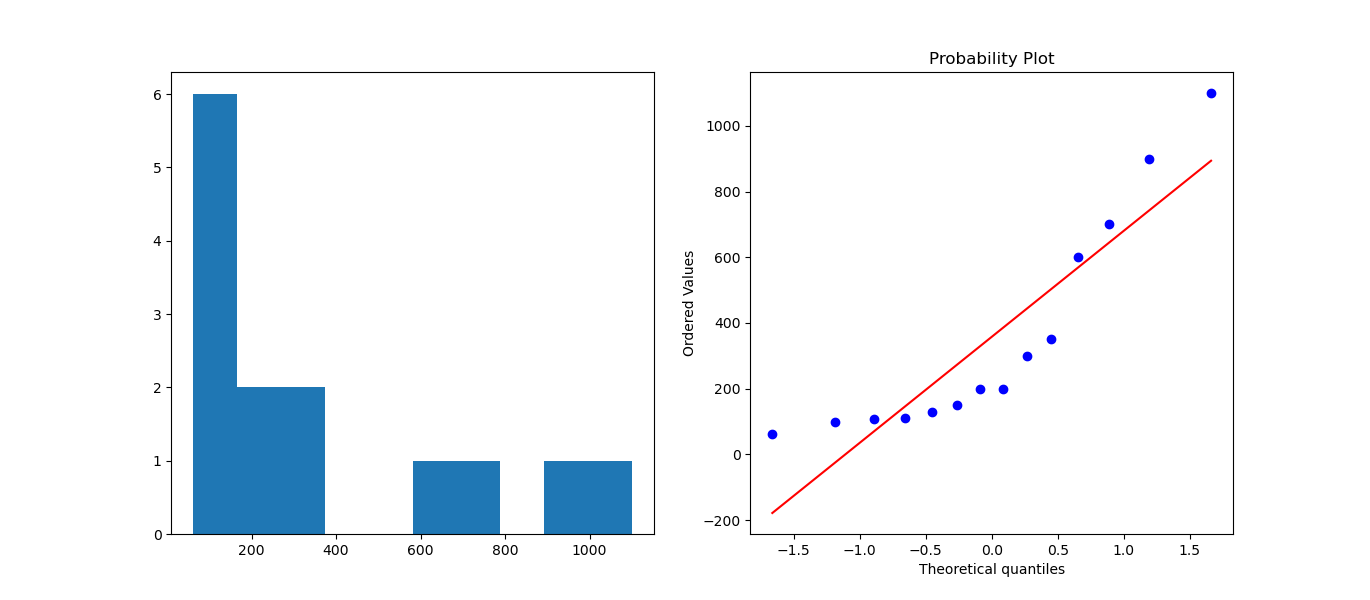
diagnostic\_plots(p1\_exp)

print("skewnes is:",p1\_exp.skew(), " ;; excess kurtosis is:",p1\_exp.kurt()) ## skewness && excess kurtosis

**Analysis:-**

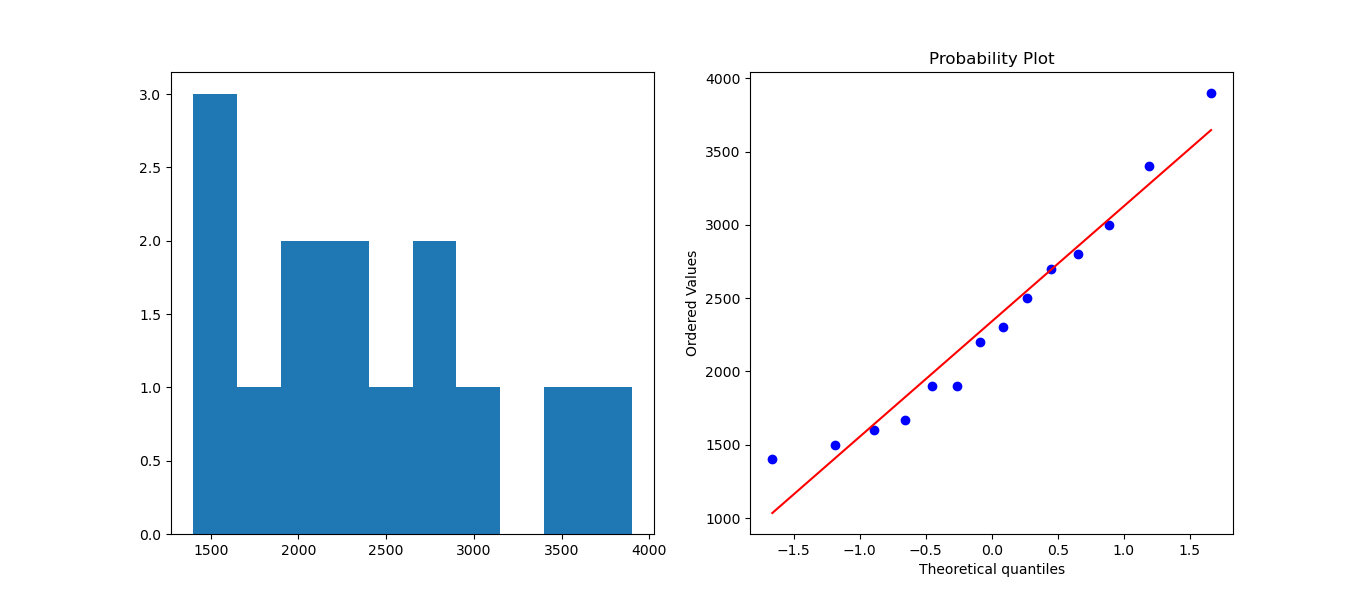
**Histogram & probability plot of datas for weight gained(grams) & calories consumed in different transformation form**

**a)Primary analysis on data without applying any transformation transformation**

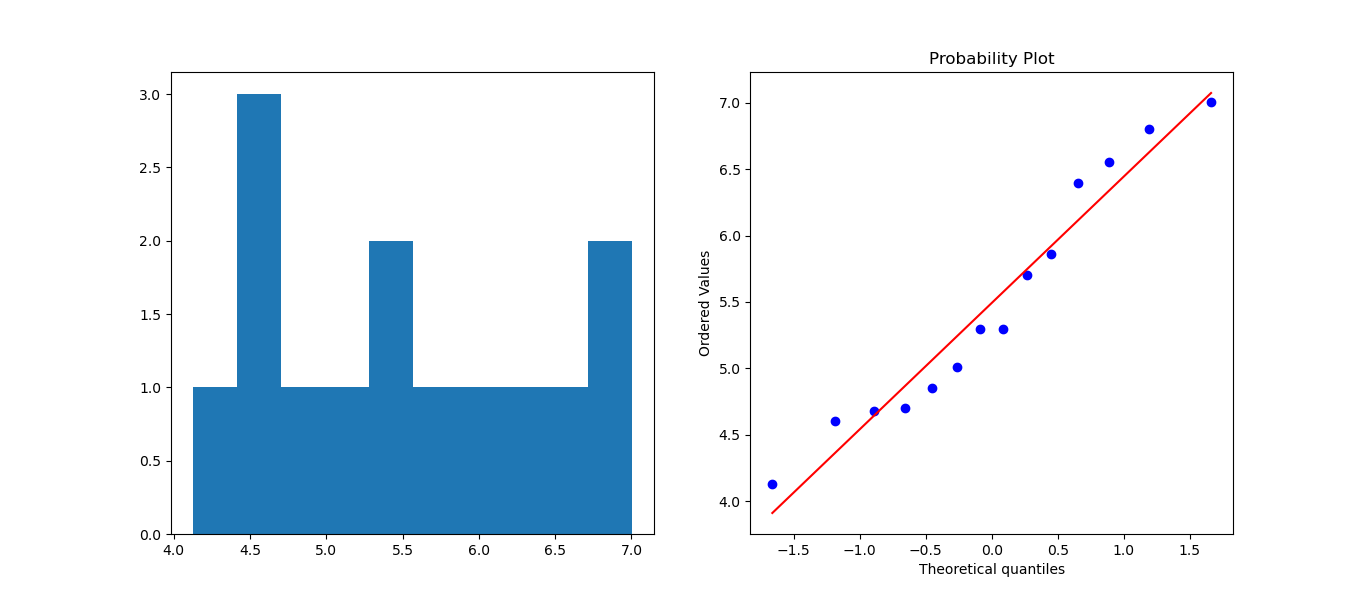
**1) weight gained(grams)**

**skewnes is: 1.2557366483972048 ;; excess kurtosis is: 0.4312724433726336**

**2)calories consumed**

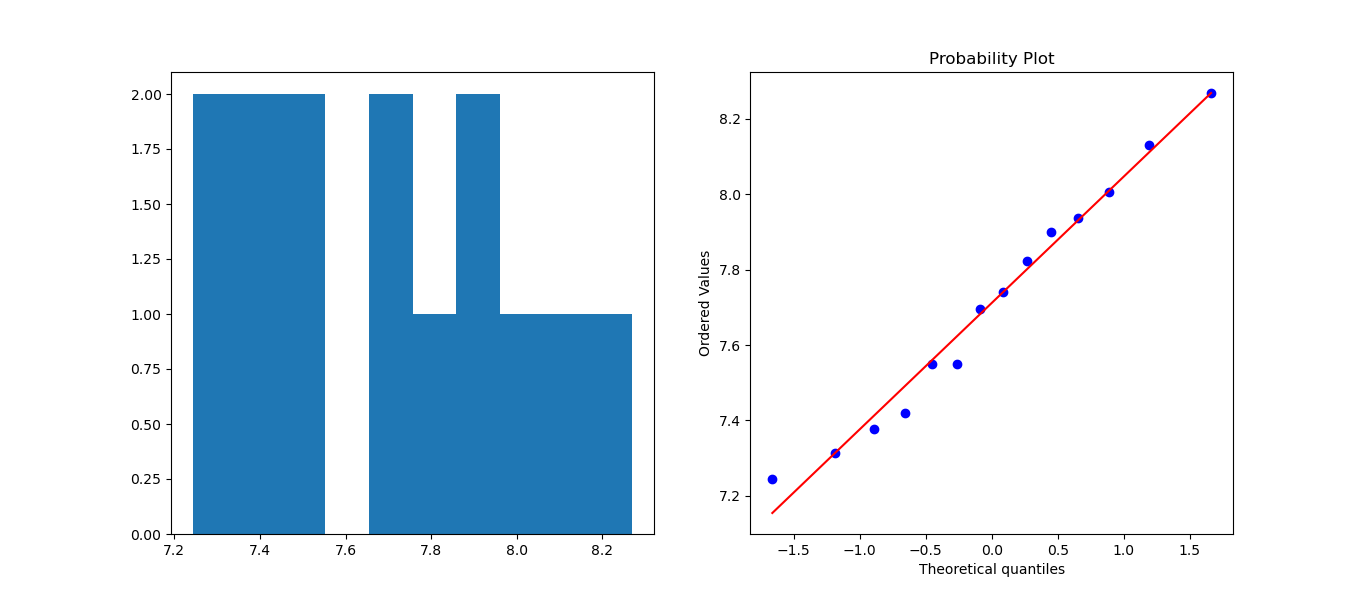
****

**skewnes is: 0.6549299573588712 ;; excess kurtosis is: -0.29048129735135975**

1. **Logarithmic Transformation**
2. **Weight gained (grams) **

**skewnes is: 0.360642241971288 ;; excess kurtosis is: -1.1215243221757696**

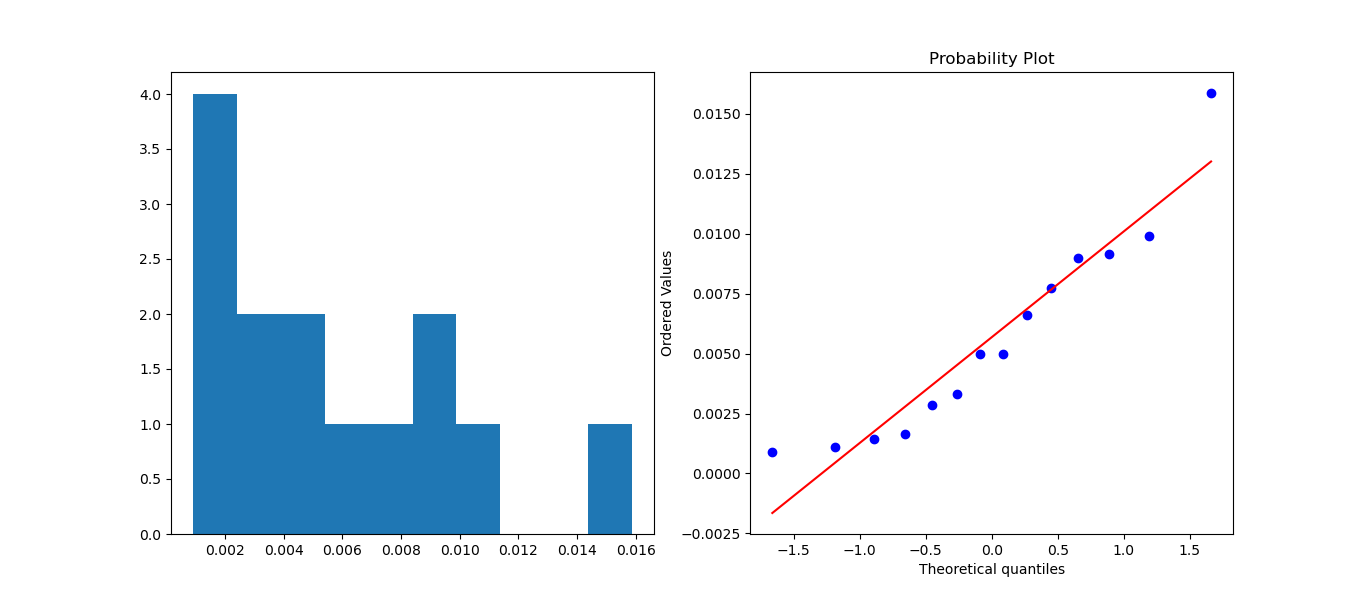
1. **Calories consumed**

****

**skewnes is: 0.1619823036541557 ;; excess kurtosis is: -0.9901699454643298**

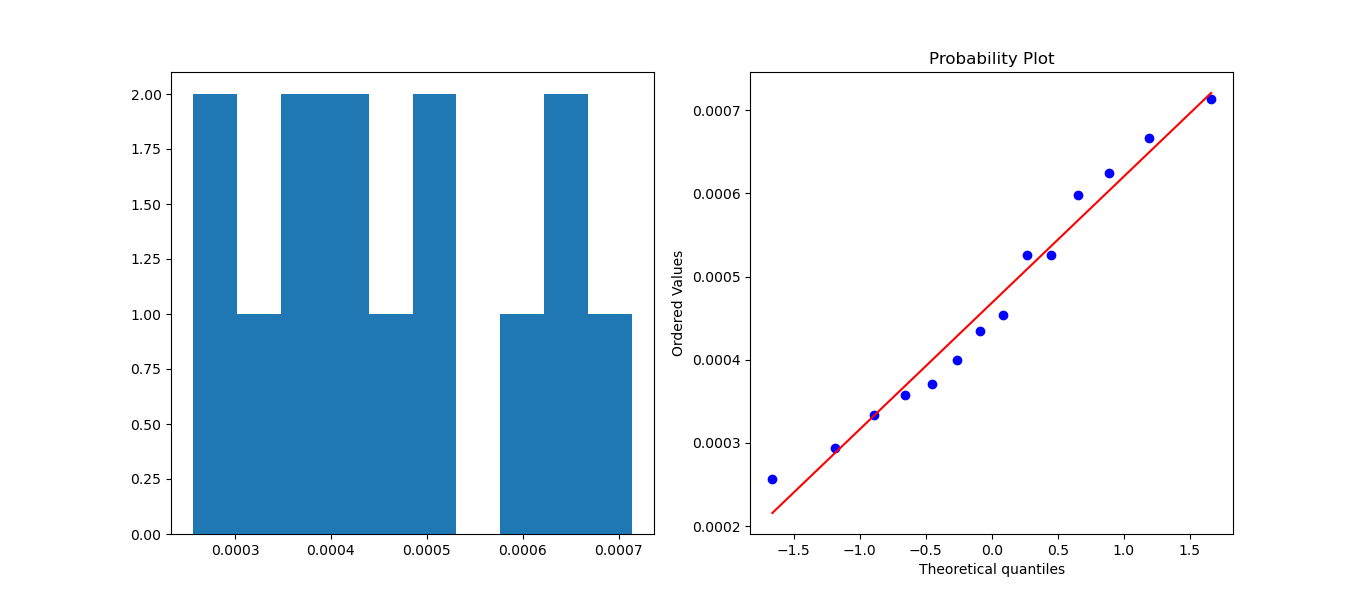
**c) Reciprocal Transformation**

**1) weight gained (grams)**

****

**skewnes is: 0.9338348775532672 ;; excess kurtosis is: 0.7370395489961563**

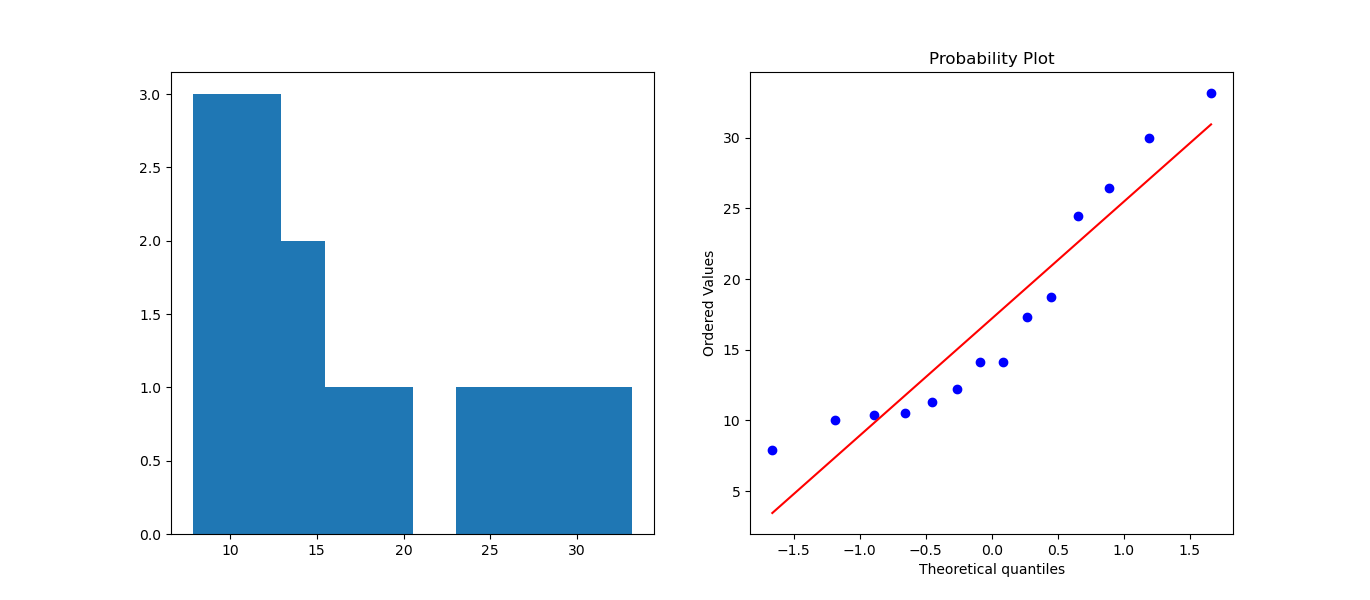
1. **Calories consumed**

****

**skewnes is: 0.2735643215342622 ;; excess kurtosis is: -1.0925193017402082**

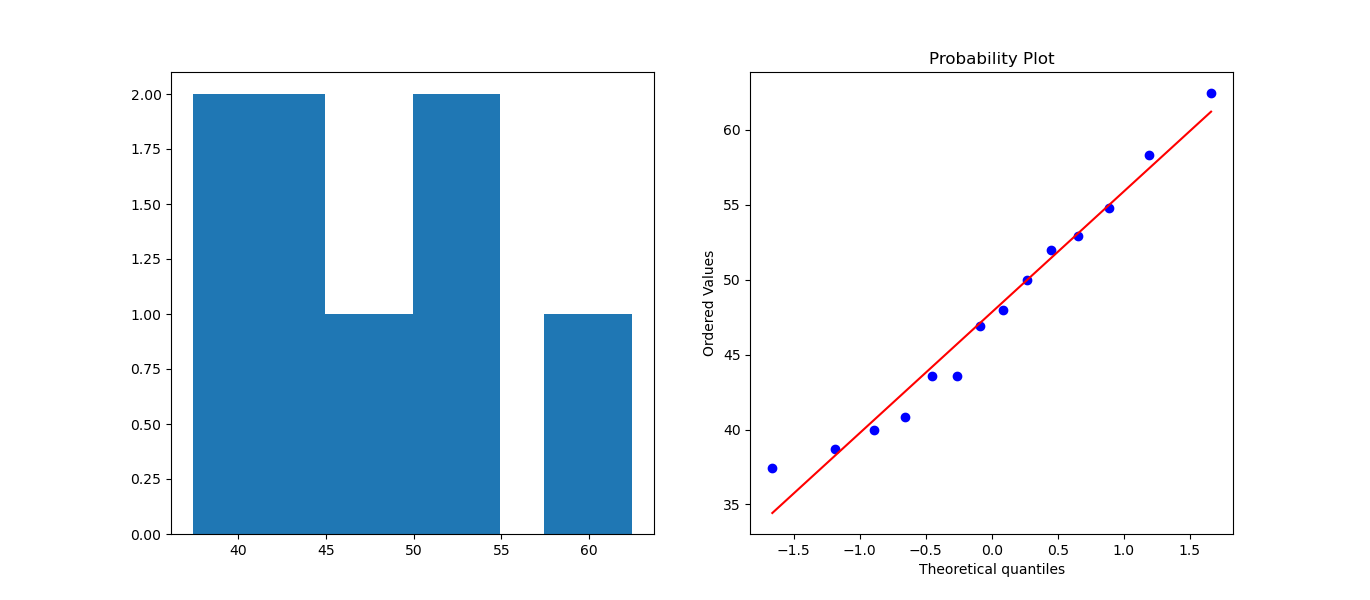
**d) Square root transformation**

**1) weight gained (grams)**

****

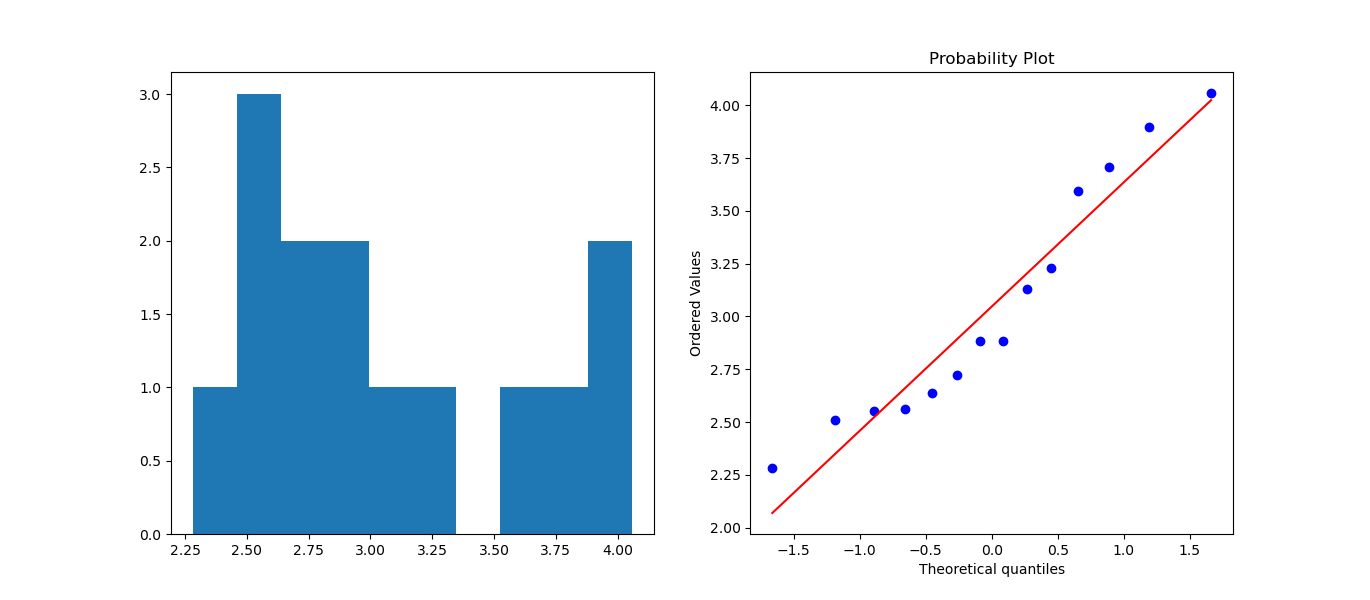
**skewnes is: 0.8515277693713486 ;; excess kurtosis is: -0.5885902251692228**

**2) calories consumed**

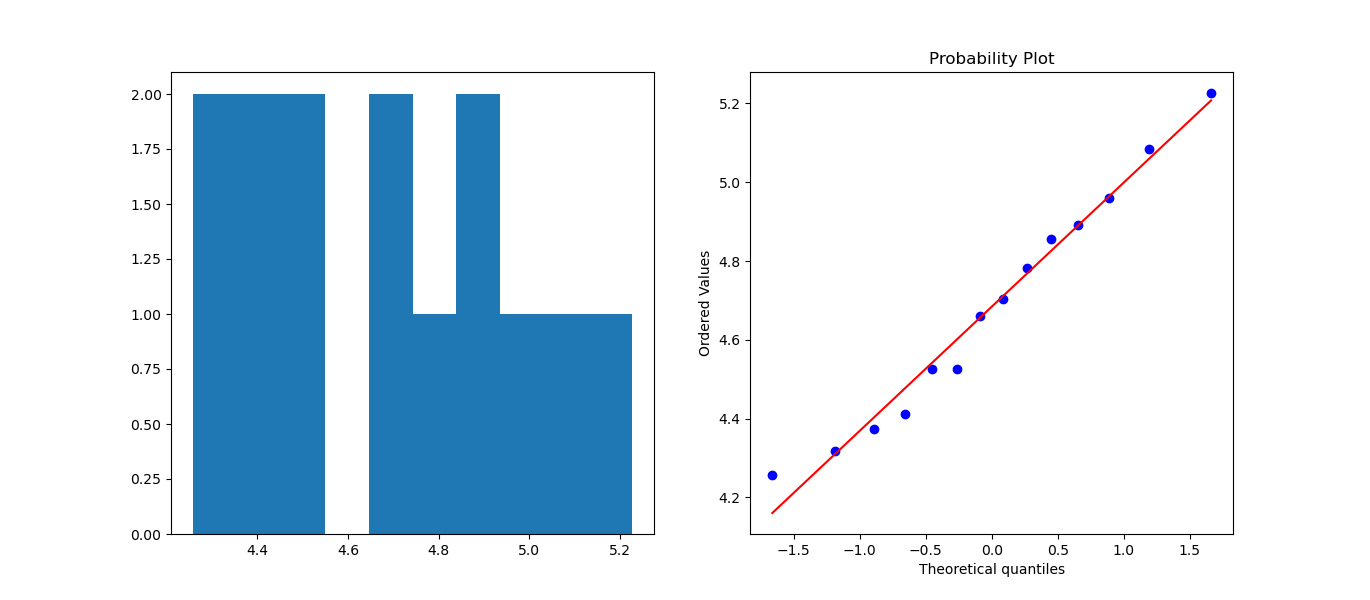
****

**skewnes is: 0.4025584590024192 ;; excess kurtosis is: -0.7208968279503383**

**e) exponential Transformation**

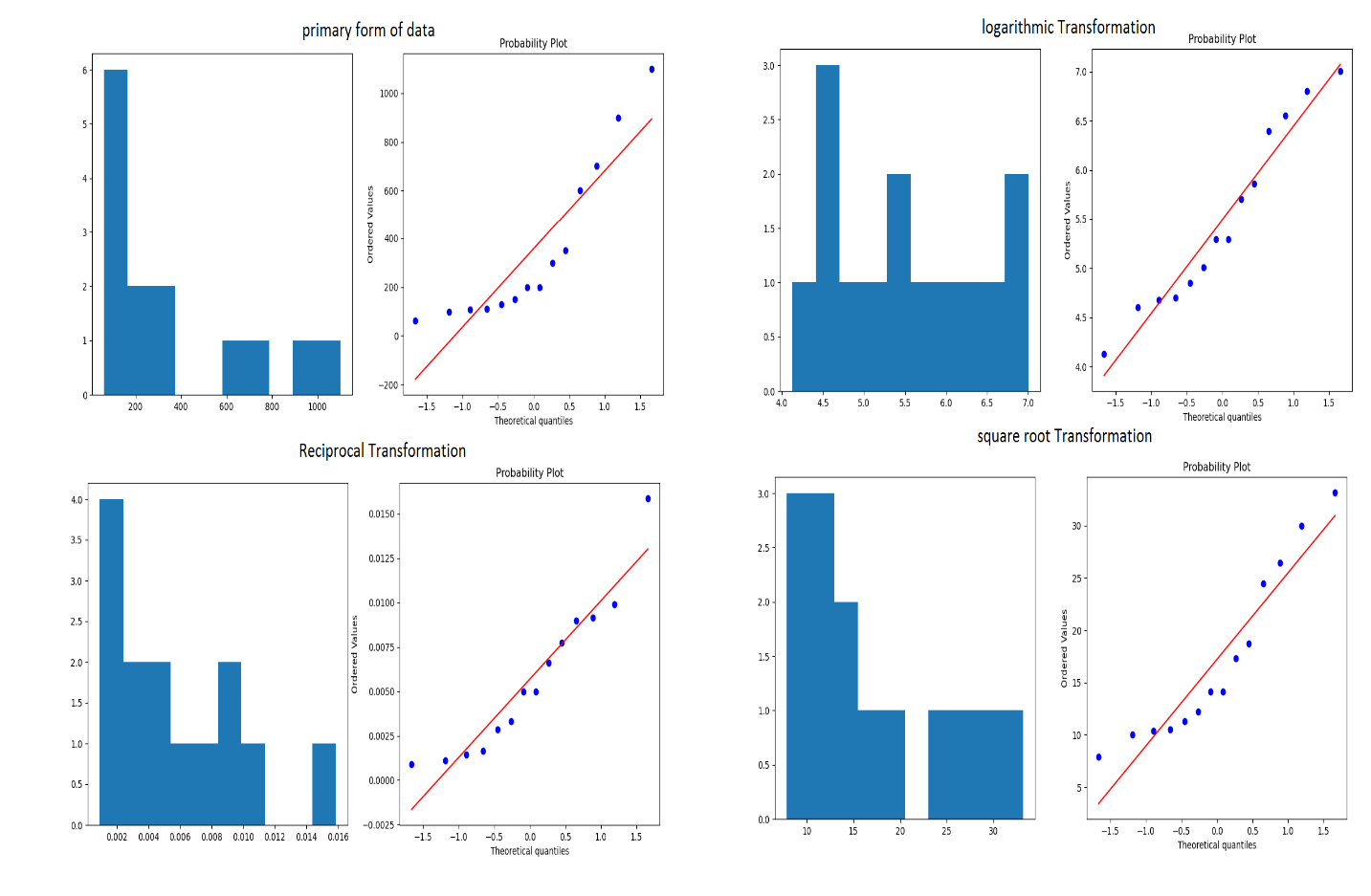
**1) weight gained(grams) **

**skewnes is: 0.5696581968945099 ;; excess kurtosis is: -0.9872804104362713**

1. **Calories consumed **

**skewnes is: 0.2565670168636527 ;; excess kurtosis is: -0.9014187751003706**

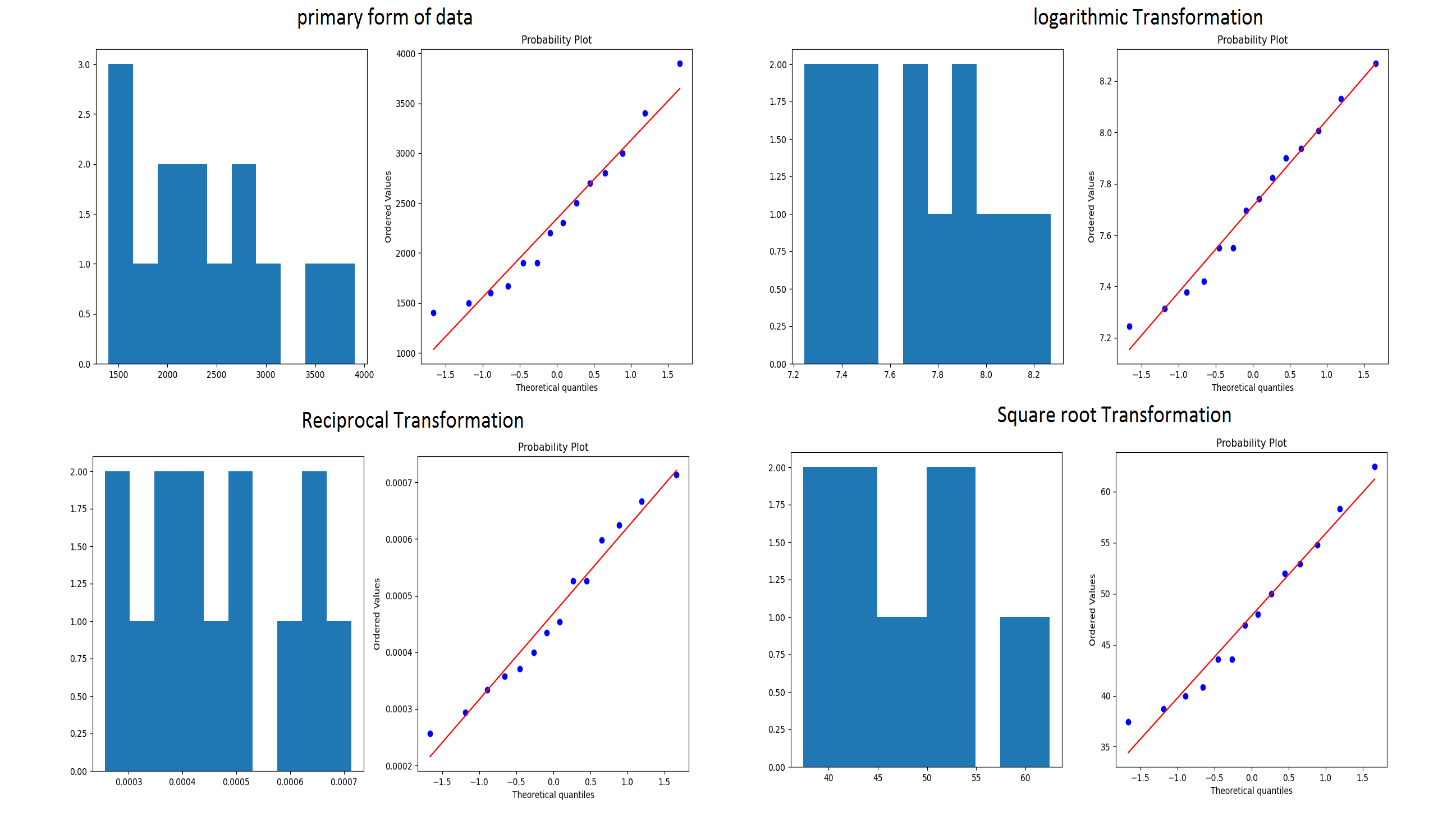
**Over all analysis on weight gained(grams) data:-**

****

From over all analysis on ‘weight gained’ data its showing better performance on reciprocal transformation as because histogram is seemingly have almost symmetrical bell shape(indicating normal characteristic) and from probability plot (qqplot) showing most of the datas are concentrated on qqline, that’s have diagonal in shape.

Skewness and excess kurtosis of perticular transformation is 0.9338348775532672 ((positive skewness=> means> median. means the distribution curve will have right skew)), 0.7370395489961563(( +ve value indicating high peakness => luptokurtic)) respectively. Both are having value less than “1” & greater than “-1”, that indicating normality.

**Over all analysis on calories consumed data:-**

****

Here Primary form of data itself showing better performance by provinding symmetrical or normal bell shaped historical plot curve. And in qqplot almost all the datas are concentrated on qqline, that’s diagonal in shape.

Skewness and excess kurtosis of perticular transformation is **0.6549299573588712** ((positive skewness=> means> median. means the distribution curve will have right skew)), **-0.29048129735135975** (( -ve value indicating low peakness => widely distributed =>platykurtic)) respectively. Both are having value less than “1” & greater than “-1”, that indicating normality.

**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