```
#WQD7005 – DATA MINING
#INSTRUCTOR : PROF. DR. TEH YING WAH
#ZAIMIE AZMIN BIN ZAINUL ABIDIN (OLD MATRIX : WQD190018 /NEW MATRIX : 17202336)
#ONLINE ASSESSMENT : DATA CLEANSING TASK
#DATE : 08 MAY 2020
```

Step to Data Cleansing

1. Crawl Data from website (mywebscrapBrentFinal.csv) - previous assignment

2. Read csv file

3. Get Data detail, shape
   print(df.shape)

   Output :
   Rows = 3675
   Columns = 5

4. Check the following condition
   i. Column headers are not value - **not related**
   ii. Multiple variables are store in one column - **not related**
   iii. Multiple types of observational units are stored in the same table - **not related**

5. Check null data
   print(df.isnull().sum())

   Output :
   Date 0
   Closing Price 0
   Open 49
   Daily High 28
   Daily Low 30

There are null attributes in column Open, Daily High and Daily Low.

6. Check how many rows are not affected by the null value
print(df.dropna(how='**any**').shape)

Total Rows = 3675
Total Rows not affected = 3624
Total Rows affected = 51 Rows

There are only 51 rows affected that has null value. This can be remove or replace the null value with mean as mention below.

7. Check mean for all columns
print(df.mean())

   Closing Price 78.408506
   Open 78.281627
   Daily High 79.315898
   Daily Low 77.422038

8. Fill in the null value for each column with mean
df2=df.fillna(df.mean())

9. Save dataframe into new csv file
df2.to_csv('new_mywebscrapBrentFinal.csv')