Assignment 6

@author : @ruhend (Mudigonda Himansh)

AP19110010169

Correlation

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [3]:
```
low_memory=False
```

In [4]:
```
df = pd.read_csv("marketing_campaign.csv",delimiter='\t')
print ('dataset: %s'%(str(df.shape)))
```

```
dataset: (2240, 29)
```

In [5]:
```
df
```

Out[5]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Re
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	
...	...	...	...	...	...	...	...	...	
2235	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	
2236	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	
2237	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	
2238	8235	1956	Master	Together	69245.0	0	1	24-01-2014	
2239	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	

2240 rows × 29 columns

In [6]:
```
MntWines = df['MntWines']
Min = MntWines.min()
Max = MntWines.max()
Diff = Max-Min
df['MntWines'] = df['MntWines'].apply(lambda x: (x-Min)/ Diff)
```

In [7]:
```
df
```

Out[7]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Re
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	
...	...	...	...	...	...	...	...	...	
2235	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	
2236	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	
2237	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	
2238	8235	1956	Master	Together	69245.0	0	1	24-01-2014	
2239	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	

2240 rows × 29 columns

In [8]:
```
cols=df.shape[1]
rows=df.shape[0]
cols_array = df.columns
for col1 in range(cols-1):
    for col2 in range(cols-1):
        column_1= df[cols_array[col1]]
        column_2= df[cols_array[col2]]
        if col1<col2:
            if isinstance(column_1[1],str) or isinstance(column_2[1],str):
                pass
            else:
#                print(cols_array[col1],cols_array[col2])
#                print(column_1.dtype)
                correlation = column_1.corr(column_2)
                print("Corelation between ",cols_array[col1], "and",cols_ar
```

```
51962
Corelation between  MntFishProducts and AcceptedCmp1  :  0.260762097916

8391
Corelation between  MntFishProducts and AcceptedCmp2  :  0.002576704636
068111
Corelation between  MntFishProducts and Complain  :  -0.020952843383501
44
Corelation between  MntFishProducts and Z_CostContact  :  nan
Corelation between  MntFishProducts and Z_Revenue  :  nan
Corelation between  MntSweetProducts and MntGoldProds  :  0.36972430038
763765
Corelation between  MntSweetProducts and NumDealsPurchases  :  -0.12010
02669155361
Corelation between  MntSweetProducts and NumWebPurchases  :  0.34854428
3083634
Corelation between  MntSweetProducts and NumCatalogPurchases  :  0.4909
239277343412
Corelation between  MntSweetProducts and NumStorePurchases  :  0.448755
95282137365
```

In [9]:
```
cols=df.shape[1]
rows=df.shape[0]
cols_array = df.columns
col1 = cols
for col2 in range(cols):
    column_1= df[cols_array[col1-1]]
    column_2= df[cols_array[col2-1]]

    if isinstance(column_1[1],str) or isinstance(column_2[1],str):
        pass
    else:
#               print(cols_array[col1],cols_array[col2])
#               print(column_1.dtype)
        correlation = column_1.corr(column_2)
        print("Correlation between ",cols_array[col1-1], "and",cols_array[c
```

```
Correlation between  Response and ID  :  0.9999999999999999
Correlation between  Response and Year_Birth  :  -0.021967817648744604
Correlation between  Response and Education  :  0.021325214441978196
Correlation between  Response and Kidhome  :  0.13304666375157911
Correlation between  Response and Teenhome  :  -0.08000778224709966
Correlation between  Response and Dt_Customer  :  -0.15444559037757166
Correlation between  Response and MntWines  :  -0.1984366465192585
Correlation between  Response and MntFruits  :  0.2472544989721133
Correlation between  Response and MntMeatProducts  :  0.1252888081082247
Correlation between  Response and MntFishProducts  :  0.2363351203783053
Correlation between  Response and MntSweetProducts  :  0.1113307947911551
4
Correlation between  Response and MntGoldProds  :  0.11737190061543223
Correlation between  Response and NumDealsPurchases  :  0.139850136947491
52
Correlation between  Response and NumWebPurchases  :  0.00223831326892120
8
Correlation between  Response and NumCatalogPurchases  :  0.1487295850305
9924
Correlation between  Response and NumStorePurchases  :  0.220810419247429
42
Correlation between  Response and NumWebVisitsMonth  :  0.039363443506241
05
Correlation between  Response and AcceptedCmp3  :  -0.003986598715538831
Correlation between  Response and AcceptedCmp4  :  0.25425828324929817
Correlation between  Response and AcceptedCmp5  :  0.17701860249675638
Correlation between  Response and AcceptedCmp1  :  0.32663394462255435
Correlation between  Response and AcceptedCmp2  :  0.2939815268524724
Correlation between  Response and Complain  :  0.16929266118503408
Correlation between  Response and Z_CostContact  :  -0.001706963954660382
1
Correlation between  Response and Z_Revenue  :  nan
Correlation between  Response and Response  :  nan
```