Marketing Anlysis

In [10]:

```
import pandas as pd
import numpy as np
from datetime import date
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("Company_Data.csv",index_col=0)
data
```

Out[10]:

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recei
0	1970	Graduation	Single	84835.0	0	0	6/16/14	
1	1961	Graduation	Single	57091.0	0	0	6/15/14	
2	1958	Graduation	Married	67267.0	0	1	5/13/14	
3	1967	Graduation	Married	32474.0	1	1	05-11-2014	
4	1989	Graduation	Single	21474.0	1	0	04-08-2014	
...	...	...	...	...	...	...	...	
2232	1976	PhD	Single	66476.0	0	1	03-07-2013	
2233	1977	Master	Married	31056.0	1	0	1/22/13	
2234	1976	Graduation	Single	46310.0	1	0	12-03-2012	
2235	1978	Graduation	Married	65819.0	0	0	11/29/12	
2236	1969	PhD	Married	94871.0	0	2	09-01-2012	

2237 rows × 30 columns

Simple Random Sampling

In [2]:

```
Sample = data.sample(n=30,replace=False)
Sample
```

Out[2]:

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recei
488	1954	Graduation	Single	50002.0	0	1	6/17/13	
1772	1982	Master	Married	66664.0	0	0	9/25/12	
934	1976	Master	Single	75012.0	0	0	1/25/14	
987	1951	Master	Married	78497.0	0	0	12-01-2013	
121	1962	Graduation	Married	71434.0	0	1	9/18/13	
1337	1993	Graduation	Single	79244.0	0	0	12/19/12	
1432	1992	Basic	Single	7500.0	1	0	12-03-2012	
1033	1975	Master	Married	37284.0	1	1	3/29/13	
732	1955	Master	Single	41769.0	0	1	2/13/13	
1211	1978	Master	Married	51381.5	0	0	08-12-2012	
2120	1961	Master	Married	61923.0	0	2	7/26/13	
262	1983	Graduation	Married	51381.5	1	0	11/15/13	
1864	1979	Graduation	Married	17688.0	1	0	1/13/13	
897	1970	Master	Married	51381.5	0	1	04-01-2013	
1762	1982	Graduation	Married	28718.0	1	0	7/13/13	
1537	1951	Master	Single	57530.0	0	1	10-12-2013	
343	1961	PhD	Married	66426.0	0	1	10-02-2013	
1785	1983	Graduation	Single	24072.0	1	0	4/16/13	
594	1974	Graduation	Single	19346.0	1	0	1/30/14	
1471	1971	Graduation	Married	70886.0	1	0	10/14/13	
617	1988	Graduation	Married	84219.0	0	0	02-02-2014	
433	1977	Master	Married	55842.0	0	1	5/23/14	
1753	1973	PhD	Single	54466.0	1	1	02-09-2014	
1361	1949	Graduation	Single	70165.0	0	0	9/22/13	
1144	1959	Graduation	Married	54984.0	0	1	5/24/14	
420	1980	Graduation	Married	37406.0	1	0	10-10-2013	
1130	1963	Master	Single	48721.0	1	1	5/13/13	
2115	1979	Basic	Married	24594.0	1	0	12-10-2013	
1467	1974	Master	Married	63206.0	0	0	3/20/14	
558	1969	Graduation	Married	56796.0	0	1	2/16/13	

30 rows × 30 columns

In [ ]:

```
```

Systematic Random Sampling

In [3]:

```
# Ordering data by "Year_Birth"
df = data.sort_values(by=["Year_Birth"],ascending=False)
n = len(df)
```

In [4]:

```
print("Select value from 0 to 9 : ")
x = int(input())
lst = []
i=x
while i<n:
    lst.append(i)
    i = i+x
df.iloc[lst]
```

```
Select value from 0 to 9 :
4
```

Out[4]:

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recei
1466	1995	Graduation	Single	34824.0	0	0	3/26/14	
686	1994	Graduation	Single	95529.0	0	0	12-03-2012	
1493	1993	Graduation	Single	74293.0	0	0	05-04-2014	
1542	1992	Graduation	Married	63207.0	0	0	1/20/13	
930	1992	Basic	Single	18746.0	1	0	5/29/14	
...	...	...	...	...	...	...	...	
472	1945	PhD	Married	70356.0	0	0	11-05-2012	
1922	1944	PhD	Single	55614.0	0	0	11/27/13	
1207	1943	PhD	Single	48948.0	0	0	02-01-2013	
1348	1943	PhD	Single	57513.0	0	0	07-06-2013	
2169	1940	PhD	Single	51141.0	0	0	07-08-2013	

559 rows × 30 columns

In [ ]:

```
```

Cluster Sampling

In [82]:

```
import random
n = pd.unique(df['Country'])
n =n.tolist()
print("Countries are : ",n)
x = random.sample(n,1)
print("Randomly selected Country :" , x[0])
df2 =df.loc[data["Country"]==x[0]]
df2
```

```
Countries are :  ['SP', 'CA', 'US', 'AUS', 'GER', 'IND', 'SA', 'ME']
Randomly selected Country : US
```

Out[82]:

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recei
2	1958	Graduation	Married	67267.0	0	1	5/13/14	
8	1954	PhD	Married	65324.0	0	1	01-11-2014	
10	1947	Master	Married	81044.0	0	0	12/27/13	
49	1960	Graduation	Single	48904.0	0	1	12-02-2012	
93	1969	PhD	Married	69476.0	0	0	9/30/13	
...	...	...	...	...	...	...	...	
2214	1960	PhD	Single	50611.0	0	1	10-04-2012	
2215	1960	PhD	Single	50611.0	0	1	10-04-2012	
2216	1970	Graduation	Single	83273.0	1	2	9/25/12	
2231	1973	Graduation	Single	78901.0	0	1	9/17/13	
2232	1976	PhD	Single	66476.0	0	1	03-07-2013	

109 rows × 30 columns

In [ ]:

```
```

Stratified Random Sampling

A sample of 100 is collected

In [22]:

```
strata1 = df.loc[ (df["Marital_Status"] == "Married") ]
strata2 = df.loc[ (df["Marital_Status"] == "Single") ]
Sample1 = strata1.sample(n=50,replace=False)
Sample2 = strata2.sample(n=50,replace=False)
frames = [Sample1, Sample2]

result = pd.concat(frames)
display(result)
```

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recei
1528	1953	Graduation	Married	34587.0	1	1	5/16/14	
621	1964	Graduation	Married	75236.0	0	1	11/25/13	
2039	1978	PhD	Married	27683.0	1	0	08-04-2012	
746	1951	PhD	Married	53312.0	0	0	08-09-2013	
1117	1957	PhD	Married	37633.0	1	1	9/20/12	
...	...	...	...	...	...	...	...	
1807	1947	PhD	Single	68117.0	0	1	03-10-2013	
300	1982	Master	Single	75777.0	0	0	07-04-2013	
2216	1970	Graduation	Single	83273.0	1	2	9/25/12	
1181	1970	Master	Single	29548.0	1	0	8/29/13	
1644	1971	PhD	Single	60504.0	0	1	02-07-2014	

100 rows × 30 columns

In [ ]:

```

```

MULTI STAGE SAMPLING

A sample of 40 need to be collected

In [67]:

```
import random
df = data
strata1 = df.loc[ (df["Marital_Status"] == "Married") ]
strata2 = df.loc[ (df["Marital_Status"] == "Single") ]
lst = ["s1","s2"]
n = random.sample(lst, 1)
print(n)
if n[0]=="s1":
    cluster1 = strata1.loc[ (df["Education"] == "Graduation") ]
    cluster2 = strata1.loc[ (df["Education"] == "PhD") ]
    cluster3 = strata1.loc[ (df["Education"] == "Master") ]
    cluster4 = strata1.loc[ (df["Education"] == "Basic") ]
    S1 = cluster1.sample(n=10,replace=False)
    S2 = cluster2.sample(n=10,replace=False)
    S3 = cluster3.sample(n=10,replace=False)
    S4 = cluster4.sample(n=10,replace=False)
    frames = [S1, S2, S3,S4]


elif n[0]=="s2":
    cluster1 = strata2.loc[ (df["Education"] == "Graduation") ]
    cluster2 = strata2.loc[ (df["Education"] == "PhD") ]
    cluster3 = strata2.loc[ (df["Education"] == "Master") ]
    cluster4 = strata2.loc[ (df["Education"] == "Basic") ]
    S1 = cluster1.sample(n=10,replace=False)
    S2 = cluster2.sample(n=10,replace=False)
    S3 = cluster3.sample(n=10,replace=False)
    S4 = cluster4.sample(n=10,replace=False)
    frames = [S1, S2, S3,S4]

result = pd.concat(frames)
result
```

```
['s2']
```

Out[67]:

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	R
2066	1975	Graduation	Single	33955.0	1	0	08-09-2013	
280	1975	Graduation	Single	33249.0	1	0	2/20/13	
456	1965	Graduation	Single	4861.0	0	0	6/22/14	
1786	1954	Graduation	Single	76773.0	0	0	2/25/13	
1604	1955	Graduation	Single	57959.0	0	1	05-02-2013	
854	1972	Graduation	Single	51813.0	1	1	04-11-2013	
1325	1988	Graduation	Single	20518.0	1	0	5/18/14	
130	1954	Graduation	Single	72071.0	0	1	2/14/13	
1758	1975	Graduation	Single	83829.0	0	0	10-08-2013	
1120	1952	Graduation	Single	69142.0	0	1	6/29/14	
1753	1973	PhD	Single	54466.0	1	1	02-09-2014	

S.No	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	R
1103	1985	PhD	Single	37929.0	0	0	7/30/13	
1697	1966	PhD	Single	78420.0	0	0	6/29/13	
1922	1944	PhD	Single	55614.0	0	0	11/27/13	
2131	1977	PhD	Single	34487.0	1	1	09-06-2012	
2176	1972	PhD	Single	86857.0	0	0	09-12-2012	
881	1956	PhD	Single	78028.0	0	1	09-10-2012	
1941	1966	PhD	Single	46734.0	1	2	10/21/13	
560	1963	PhD	Single	52278.0	0	1	1/25/13	
225	1975	PhD	Single	74165.0	0	0	05-01-2013	
218	1968	Master	Single	50014.0	1	0	1/22/14	
765	1981	Master	Single	36143.0	1	0	3/30/14	
698	1954	Master	Single	42607.0	0	1	11/17/13	
1250	1951	Master	Single	49389.0	1	1	8/29/13	
157	1972	Master	Single	46423.0	1	1	9/18/13	
97	1949	Master	Single	47570.0	1	1	5/29/13	
2175	1952	Master	Single	28457.0	0	0	10/28/12	
52	1978	Master	Single	52195.0	2	1	05-12-2014	
868	1958	Master	Single	75342.0	0	1	05-06-2013	
992	1966	Master	Single	89572.0	0	0	9/15/12	
1335	1982	Basic	Single	24367.0	1	0	3/20/13	
528	1978	Basic	Single	26487.0	1	0	5/20/13	
590	1986	Basic	Single	8940.0	1	0	8/22/12	
417	1984	Basic	Single	25707.0	1	0	1/18/14	
2098	1987	Basic	Single	15038.0	1	0	1/29/13	
1951	1986	Basic	Single	14906.0	0	0	12/22/12	
930	1992	Basic	Single	18746.0	1	0	5/29/14	
956	1979	Basic	Single	16014.0	1	1	3/17/13	
739	1976	Basic	Single	9548.0	1	0	08-08-2012	
1432	1992	Basic	Single	7500.0	1	0	12-03-2012	

In [ ]:

In [ ]:

In [ ]:

```

```

In [ ]:

```

```