```
import pandas as pd
from scipy.stats import ttest_ind
import scipy.stats as stats
import numpy as np

df=pd.read_csv('Social_Network_Ads.csv')

df
```

```
      User ID  Gender  Age  EstimatedSalary  Purchased
0    15624510    Male   19            19000          0
1    15810944    Male   35            20000          0
2    15668575  Female   26            43000          0
3    15603246  Female   27            57000          0
4    15804002    Male   19            76000          0
..        ...     ...  ...              ...        ...
395  15691863  Female   46            41000          1
396  15706071    Male   51            23000          1
397  15654296  Female   50            20000          1
398  15755018    Male   36            33000          0
399  15594041  Female   49            36000          1

[400 rows x 5 columns]
```

```
df.describe()
```

```
            User ID         Age  EstimatedSalary   Purchased
count  4.000000e+02  400.000000       400.000000  400.000000
mean   1.569154e+07   37.655000     69742.500000    0.357500
std    7.165832e+04   10.482877     34096.960282    0.479864
min    1.556669e+07   18.000000     15000.000000    0.000000
25%    1.562676e+07   29.750000     43000.000000    0.000000
50%    1.569434e+07   37.000000     70000.000000    0.000000
75%    1.575036e+07   46.000000     88000.000000    1.000000
max    1.581524e+07   60.000000    150000.000000    1.000000
```

Q2

a) POPULATION MEAN OF MALE AND FEMALE AGES

```
m1 = np.array(df[df['Gender']=='Male']['Age'])

m2 =np.array(df[df['Gender']=='Female']['Age'])

print(f"MEAN OF MALE IS {np.mean(m1)}")

MEAN OF MALE IS 36.86734693877551

print(f"MEAN OF FEMALE IS {np.mean(m2)}")
```

```
MEAN OF FEMALE IS 38.411764705882355
```

b) POPULATION OF VARIANCE OF AGES OF MALE AND FEMALE

```
var_male=np.array(df[df['Gender']=='Male']['Age'])
print(f"VARIANCE OF MALE IS {np.var(var_male)}")
```

```
VARIANCE OF MALE IS 100.35995418575592
```

```
var_female=np.array(df[df['Gender']=='Female']['Age'])
print(f"VARIANCE OF FEMALE IS {np.var(var_female)}")
```

```
VARIANCE OF FEMALE IS 117.34025374855825
```

Q3

Proportion of males whose ads were purchased and not purchased.

```
P1_Male=len(df[df['Gender']=='Male'][df['Purchased']==1])/
len(df[df['Gender']=='Male'])
```

```
/var/folders/kf/vbsk0d312vv9vb_sf9v4ghz00000gn/T/
ipykernel_24576/574178490.py:1: UserWarning: Boolean Series key will
be reindexed to match DataFrame index.

P1_Male=len(df[df['Gender']=='Male'][df['Purchased']==1])/len(df[df['G
ender']=='Male'])
```

Male Purchased Ads

```
P1_Male
```

```
0.336734693877551
```

Male Not Purchased Ads

```
Q1_Male={1-P1_Male}
```

```
Q1_Male
```

```
{0.6632653061224489}
```

Q4

Proportion of females whose ads were purchased and not purchased

```
P1_Female=len(df[df['Gender']=='Female'][df['Purchased']==1])/
len(df[df['Gender']=='Female'])
```

```
/var/folders/kf/vbsk0d312vv9vb_sf9v4ghz00000gn/T/
ipykernel_24576/3831493596.py:1: UserWarning: Boolean Series key will
be reindexed to match DataFrame index.
```

```
P1_Female=len(df[df['Gender']=='Female'][df['Purchased']==1])/len(df[d
f['Gender']=='Female'])
```

Female Purchased Ads

```
P1_Female
```

```
0.37745098039215685
```

Female Not Purchased Ads

```
Q1_Female={1-P1_Female}
```

```
Q1_Female
```

```
{0.6225490196078431}
```

Q5

Take a random sample

```
Success=np.array([196,204])
Total_Sample_Size=np.array([400,400])
```

```
from statsmodels.stats.proportion import proportions_ztest
z_stat, p_value =
proportions_ztest(Success,Total_Sample_Size,alternative='two-sided')
```

```
z_stat,p_value
if p_value>0.05:
    print(f'p_value is {p_value}')
```

```
p_value is 0.5716076449533312
```

As we can see that the p_value is greater than 0.05 so we will take the size of the random sample as equal.

```
Male = df[df['Gender']=='Male']
```

```
Male_Random_Sample = Male.sample(n=40,random_state=1)
```

```
len(Male_Random_Sample)
```

```
40
```

```
Female = df[df['Gender']=='Female']
```

```
Female_Random_Sample = Female.sample(n=40,random_state=1)
```

```
len(Female_Random_Sample)
```

```
40
```

Q6

HYPOTHESIS TESTING

a) Proportion of Male ads Purchased = Proportion of Female ads Purchased

Ho: Proportion of Male and Female are equal p1=p2

HA: Proportion of Male and Female are not equal p1!=p2

```
m_purchased = Male_Random_Sample[Male_Random_Sample['Purchased']==1]
len(m_purchased)
```

```
14
```

```
f_purchased =
Female_Random_Sample[Female_Random_Sample['Purchased']==1]

len(f_purchased)
```

```
16
```

```
Success = np.array([14,16])
Sample_Size = np.array([40,40])

z_stat, p_value =
proportions_ztest(Success,Sample_Size,alternative='two-sided')

z_stat,p_value
if p_value>0.05:
    print(f'{p_value} Accept Null Hypothesis')
else:
        print('Reject Null Hypothesis')
```

```
0.6441672226837099 Accept Null Hypothesis
```

b) Proportion of Male ads not Purchased = Proportion of Female ads not Purchased

HO = P1=P2

HA = P1!=P2

```
m_unpurchased=Male_Random_Sample[Male_Random_Sample['Purchased']==0]
```

```
f_unpurchased=Female_Random_Sample[Female_Random_Sample['Purchased']==
0]


len(m_unpurchased)
```

```
26
```

```
len(f_unpurchased)
```

```
24
```

```
unsucess=np.array([26,24])
Sample_Size=np.array([40,40])

z_stat, p_value =
proportions_ztest(unsucess,Sample_Size,alternative='two-sided')

z_stat,p_value
if p_value>0.05:
    print(f'{z_stat,p_value} Accept Null Hypothesis')
else:
        print('Reject Null Hypothesis')
```

```
(0.46188021535170104, 0.6441672226837099) Accept Null Hypothesis
```

c) Equal Variances of Ages of Male and Female

HO: $\sigma 1 = \sigma 2$

HA: $\sigma 1 != \sigma 2$

```
x1 = Male_Random_Sample['Age']
x2 = Female_Random_Sample['Age']

X1=np.array(x1)
X2=np.array(x2)

def f_test(x, y):
    X1=np.array(x1)
    X2=np.array(x2)
    f = np.var(X1, ddof=1)/np.var(X2, ddof=1)
    dfn = X1.size-1
    dfd = X2.size-1
    p = 1-stats.f.cdf(f, dfn, dfd)
    if p>0.05:
        print(f'The value of p is {p}, So we will accept the Null
Hypothesis')
    else:
        print(f'The value of p is {p}, So we will reject the Null
Hypothesis')
```

```
f_test(X1,X2)
```

```
The value of p is 0.5208720264310946, So we will accept the Null Hypothesis
```

d) Based on the results of 'c', the equality of means of ages of males and females.

According to the result of c, we denote that the variance of both random samples are same so the mean of ages of males and females will also be same.

Q7

Confidence Intervals

The confidence interval was taken 95%

a: Both male and female will equally purchase ads.

b: Proportion of male and female are equal in terms of not purchasing ads

c: Variances of male and female ages are equal

d: Means of male and female ages are equal

Q8

SUMMARY

While observing the above analysis, we have got to know that male and female have same buying patterns.The rate of ads purchased and not purchased are almost same for both the genders. We will accept the Null Hypothesis H0 and reject the Alternate Hypothesis HA.

Q9

Title: Marketing Analysis on Social Network