Paras Patel-Day12 Assignment

To try different tests in attrition case study

```
from scipy.stats import mannwhitneyu
```

```
d1 = data[data['Attrition'] == 1]['DistanceFromHome']
d2 = data[data['Attrition'] == 0]['DistanceFromHome']
stat,p = mannwhitneyu(d1,d2)
print(stat,p)
```

```
1295261.0 0.488538986087403
```

H0 -> There is no significant difference in the DistanceFromHome of those who left and those who don't left

HA -> There is significant difference in the DistanceFromHome of those who left and those who don't left.

As p>0.05, we have to reject the null hypothesis and accept the alternative hypothesis.

```
In [112]: from scipy.stats import kruskal
```

```
In [113]: d1 = data[data['BusinessTravel'] == 0]['MonthlyIncome']
          d2 = data[data['BusinessTravel'] == 1]['MonthlyIncome']
          d3 = data[data['BusinessTravel'] == 2]['MonthlyIncome']
          stat,p = kruskal(d1,d2,d3)
          print(stat,p)
```

```
21.51316624477811 2.1304694246114354e-05
```

H0 -> There is no significant difference in the MonthlyIncome of those who non-travel, travel rarely, travel frequently

HA -> There is significant difference in the MonthlyIncome of those who non-travel, travel rarely, travel frequently

As p<0.05, we have to accept the null hypothesis and reject the alternative hypothesis.

```
In [82]: from scipy.stats import ttest_1samp

In [83]: stat,p = ttest_1samp(data.Age,36)
         print(stat,p)

         6.761928469022729 1.5408480508508402e-11
```

H0: There is no significant difference of the mean of age against population mean = 36

H1: There is significant difference of the mean of age against population mean = 36

As p<0.05, we have to accept the null hypothesis and reject the alternative hypothesis.

```
In [86]: from scipy.stats import ttest_ind

In [114]: d1 = data[data['Attrition'] == 0]['MonthlyIncome']
          d2 = data[data['Attrition'] == 1]['MonthlyIncome']
          stat,p = ttest_ind(d1,d2)
          print(stat,p)

          1.9969640177214658 0.045890862744972095
```

H0 -> There is no significant difference in the mean of MonthlyIncome of those who left and those who are there in the company.

H1 -> There is significant difference in the mean of MonthlyIncome of those who left and those who are there in the company.

As p>0.05, we have to reject the null hypothesis and accept the alternate hypothesis.

```
In [126]: d1 = data[data['Attrition'] == 0]['PercentSalaryHike']
          d2 = data[data['Attrition'] == 1]['PercentSalaryHike']
          stat,p = ttest_ind(d1,d2)
          print(stat,p)

          -2.1953267956195552 0.028192446935249534
```

H0 -> There is no significant difference in the mean of PercentSalaryHike of those who left and those who are there in the company.

H1 -> There is significant difference in the mean of PercentSalaryHike of those who left and those who are there in the company.

As p<0.05, we have to accept the null hypothesis and reject the alternate hypothesis.