



**Female**

153

330

1281

In [22]:

**import pandas as pd**

s=pd.read\_csv("general\_data.csv")

**1)Chi square**

**H0:There is no significance difference between Gender and Business Travel**

1. **1:There is significance difference between Gender and Business Travel**
2. n [23]:

**from scipy.stats import** chi2\_contingency

ctable=pd.crosstab(s.Gender,s.BusinessTravel)

ctable

Out[23]:

**BusinessTravel Non-Travel Travel\_Frequently Travel\_Rarely**

**Gender**

**Male** 297 501 1848

In [24]:

stats,p,dof,expeted=chi2\_contingency(ctable)

stats,p

Out[24]:

(7.929887577835395, 0.01896910285626416)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie***

***There is significance difference between Gender and Business Travel***

**2)**

**H0:There is no significance difference between MaritalStatus and JobLevel**

1. **1:There is significance difference between MaritalStatus and JobLevel**
2. n [27]:

**from scipy.stats import** chi2\_contingency

ctable=pd.crosstab(s.MaritalStatus,s.JobLevel)

ctable

Out[27]:

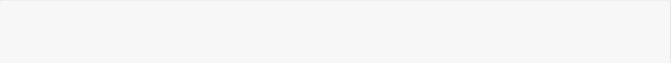
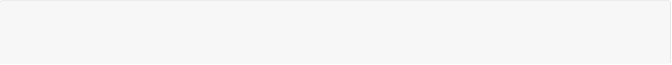
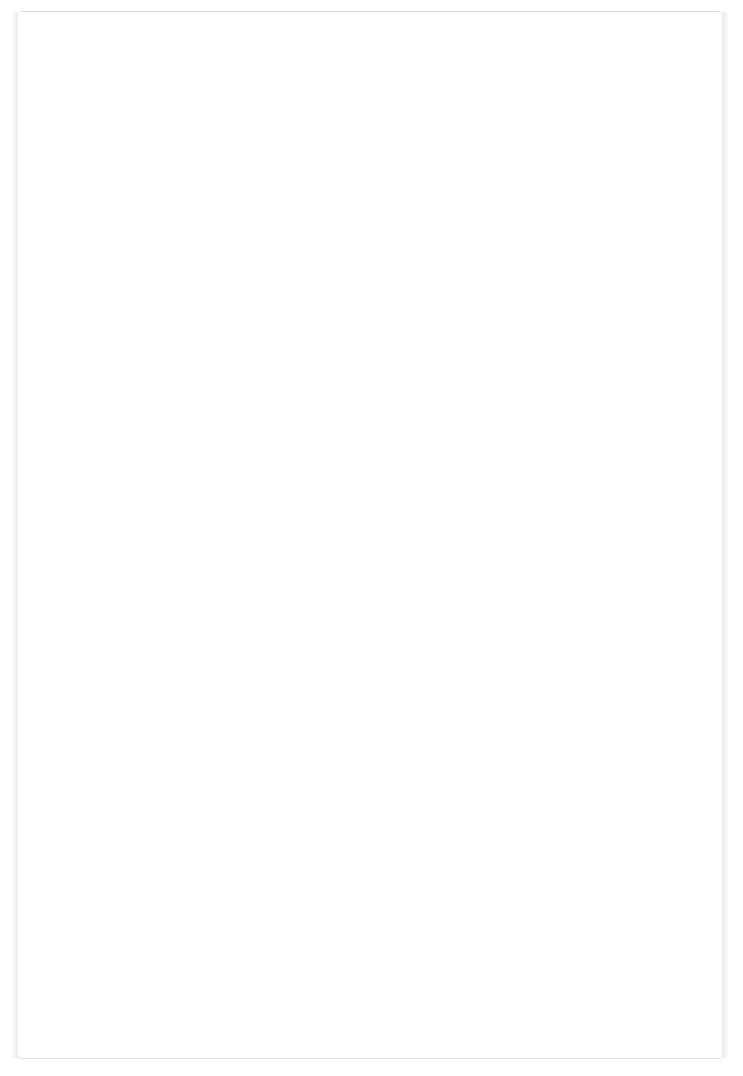
**JobLevel 1 2 3 4 5**

**MaritalStatus**

357 345 153 81 45

**Married** 738 720 321 156 84

534 537 180 81 78



In [28]:

stats,p,dof,expeted=chi2\_contingency(ctable)

stats,p

Out[28]:

(18.091153906719665, 0.020553378716182978)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie,***

***There is significance difference between MaritalStatus and JobLevel***

**3)Friedman Test**

**H0:There is no significance difference between NumCompaniesWorked ,YearsAtCompany and Total working years**

**H1:There is significance difference between NumCompaniesWorked and YearsAtCompany and Total working years** In [29]:

**from scipy.stats import** friedmanchisquare

stats,p =friedmanchisquare(s .NumCompaniesWorked,s .YearsAtCompany,s.TotalWorkingYears) stats,p

Out[29]:

(5100.159963317181, 0.0)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie,***

***H1:There is significance difference between NumCompaniesWorked and YearsAtCompany and Total working years* 4) one sample**

**H0:There is no significance difference between sample mean and population mean of MonthlyIncome**

1. **1:There is significance difference between sample mean and population mean of MonthlyIncome**
2. n [30]:

**from scipy.stats import** ttest\_1samp

stats,p =ttest\_1samp(s.MonthlyIncome, 83800)

stats,p

Out[30]:

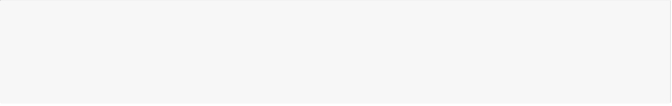
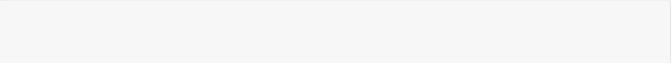
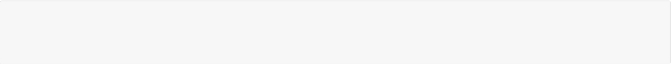
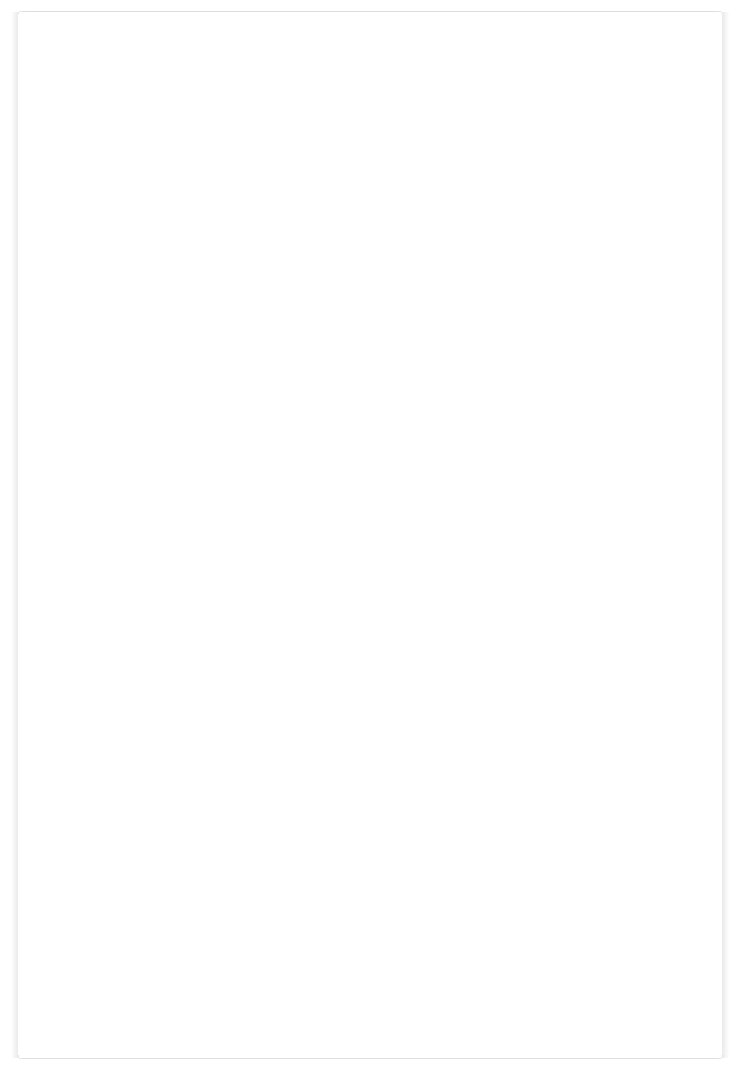
(-26.482898802308437, 1.5129543303048855e-143)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie,***

***There is significance difference between sample mean and population mean of MonthlyIncome***

**5)Wilcoxon**

**H0:There is no significance difference between and Training Times Last Year and distance from home**



**H1:There is significance difference between and Training Times Last Year and distance from home**

1. n [31]:

**from scipy.stats import** wilcoxon

stats,p =wilcoxon(s.TrainingTimesLastYear,s .DistanceFromHome)

stats,p

Out[31]:

(822562.5, 0.0)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie,***

1. ***1:There is significance difference between and Training Times Last Year and distance from home***
2. n [32]:

syes =s[s['Attrition' ]=='Yes']

sno=s[s['Attrition' ]=='No' ]

**6)Mann Whitney Test**

**H0:There is no significance difference between Distance from Home of employees where attrition is yes and where attrition is zero**

1. **1:There is significance difference between Distance from Home of employees where attrition is yes and where attrition is zero**
2. n [33]:

**from scipy.stats import** mannwhitneyu

stats,p =mannwhitneyu(syes.DistanceFromHome,sno .DistanceFromHome)

stats,p

Out[33]:

(1312110.0, 0.4629185205822659)

***The value of p>0.05 hence we reject alternative hypothesis and accept null hypothesis ie,***

***There is no significance difference between Distance from Home of employees where attrition is yes and where attrition is zero***

**7)Kruskal Wallis Test**

**H0:There is no significance difference between Years at company where business Travel is Travel frequently ,Travel Rarely ,Non Travel**

1. **1:There is significance difference between Years at company where business Travel is Travel frequently ,Travel Rarely ,Non Travel**
2. n [34]:

st1=s[s['BusinessTravel']=='Travel\_Frequently']

st2=s[s['BusinessTravel']=='Travel\_Rarely']

st3=s[s['BusinessTravel']=='Non-Travel']

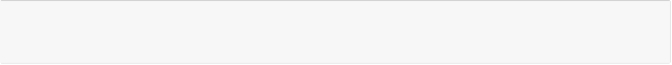
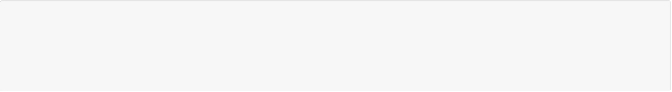
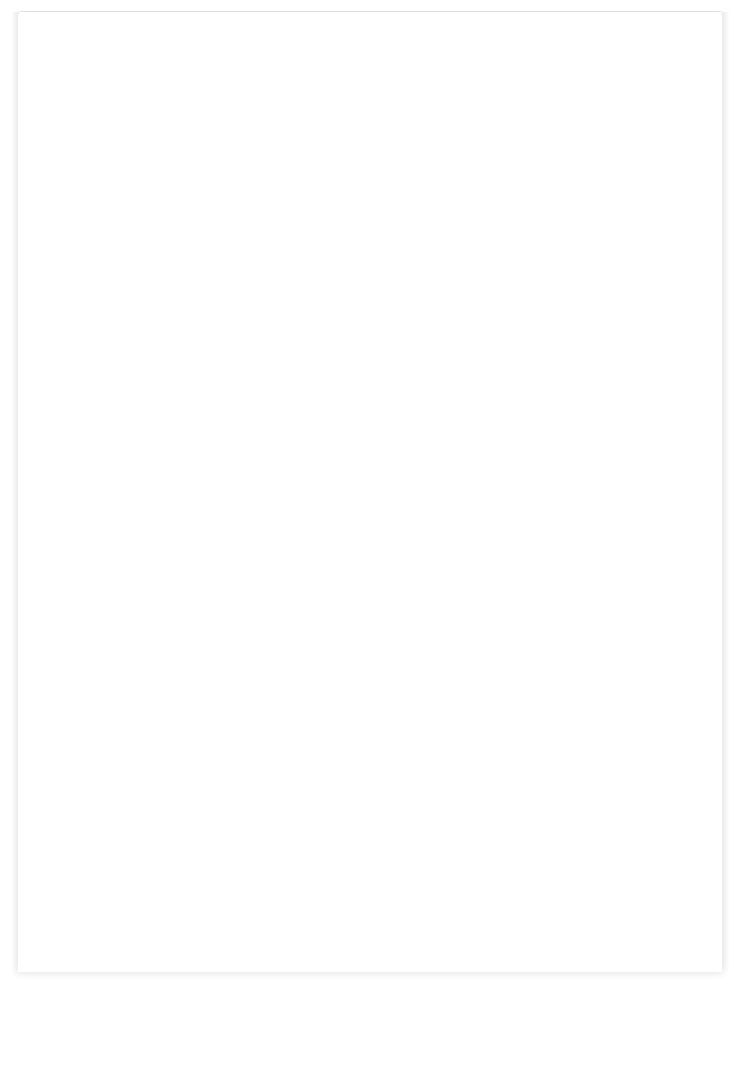
**from scipy.stats import** kruskal

stats,p =kruskal(st1 .YearsAtCompany,st2.YearsAtCompany,st3.YearsAtCompany)

stats,p

Out[34]:

(2.187341926985394, 0.33498451788533107)



***The value of p>0.05 hence we reject alternative hypothesis and accept null hypothesis ie,***

***There is significance difference between Years at company where business Travel is Travel frequently ,Travel Rarely ,Non Travel***

**8)Two sample independent t test**

**H0:There is no significance difference between the mean of Years since Last promotion of Female and Male**

1. **1:There is significance difference between the mean of Years since Last promotion of Female and Male**
2. n [38]:

sm1=s[s['Gender']=='Male']

sf=s[s['Gender']=='Female']

**from scipy.stats import** ttest\_ind

stats,p =ttest\_ind(sm1.YearsSinceLastPromotion,sf.YearsSinceLastPromotion)

stats,p

Out[38]:

(-1.4484914214456694, 0.14755078954927858)

***The value of p>0.05 hence we reject alternative hypothesis and accept null hypothesis ie,***

***There is significance difference between the mean of Years since Last promotion of Female and Male***

**9) two sample paired t test**

**H0:There is no significance difference between the mean of Monthly income Year and Years at company when attrition='Yes'**

1. **1:There is significance difference between the mean of Monthly income Year and Years at company when attrition='Yes'**
2. n [37]:

**from scipy.stats import** ttest\_rel

stats,p =ttest\_rel(s .MonthlyIncome,s.YearsAtCompany)

stats,p

Out[37]:

(91.73768218543985, 0.0)

***The value of p<0.05 hence we reject null hypothesis and accept alternative hypothesis ie,***

***H1:There is significance difference between mean of Monthly income Year and Years at company when***

***attrition='Yes'***

In [ ]: