**ATTRITION PROJECT**

* column presents in data sheet

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

a.columns

Out[5]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

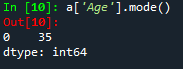
'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],

dtype='object')



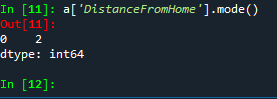
Mean distance from home of employee is 9.2 km



More no of employee work in company

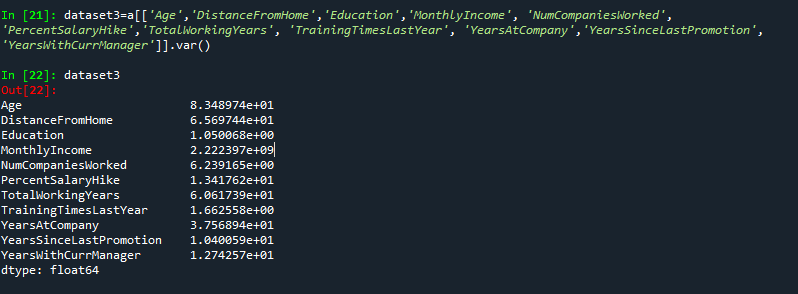


Mean age employee present in company is approx. 37

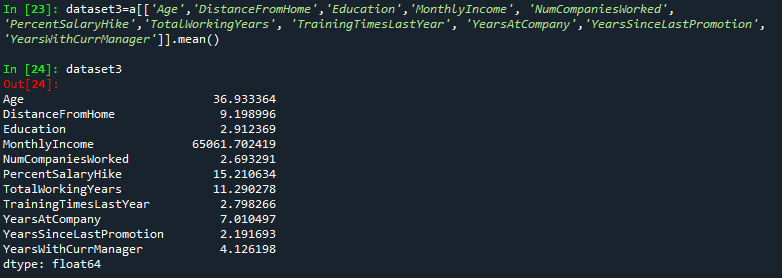


Maximum employee distance from home is 2km

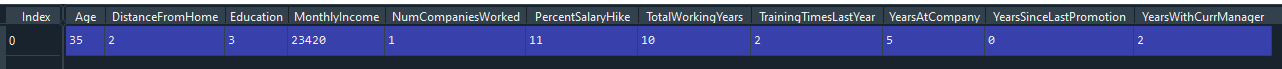
VARIANCE of data



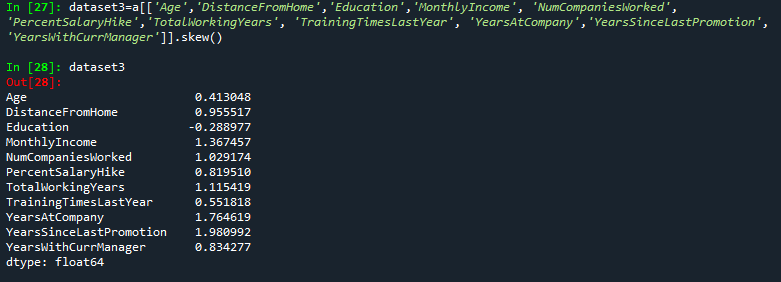
Mean of all univeriant



Mode of data



Skewness of data



Columns present in data set

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

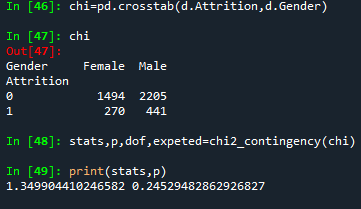
'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],

dtype='object')

1. **Chisquare test (use to check dependency of variable, variable should be categorical)**

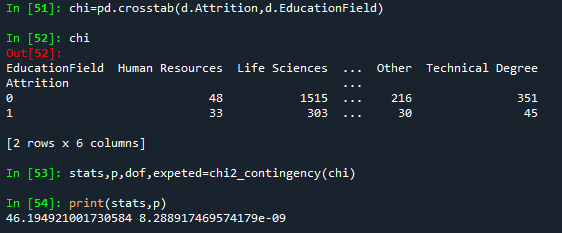
****

H0: there is no releation ship between gender & attration

H1: there is releation ship between gender & attriation of employee

The p value is grater than 0.05 so H0 is proof

**2. Chisquare test (use to check dependency of variable, variable should be categorical)**

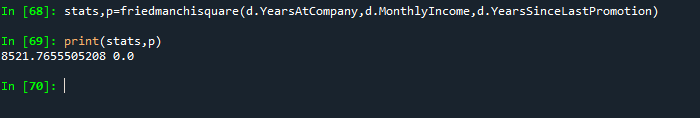


H0: there is no releation ship between in Attriation & education field

Ha: there is releation ship between in Attriation & education field

The p value is less than 0.05 so alternative hypothesis is true

**3. fridman test (compare more than 2 paired variable )**

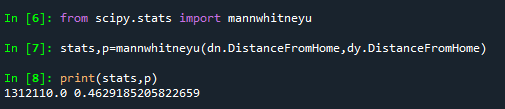
****

H0: there is no significance relation between how many YearAtCompany ,MonthlyIncome & YearSinceLastPromotion

Ha : : there is significance relation between how many YearAtCompany ,MonthlyIncome & YearSinceLastPromotion

P value is less than 0.05 so Ha is true

**4. Mannwhiney test (compare 2 independent sample )**



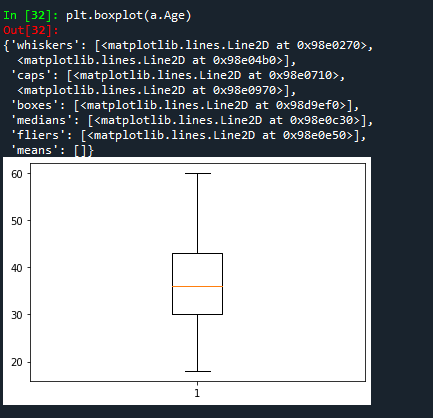
dn=(data set where Attrition is NO)

dy=(data set where Attrition is Yes)

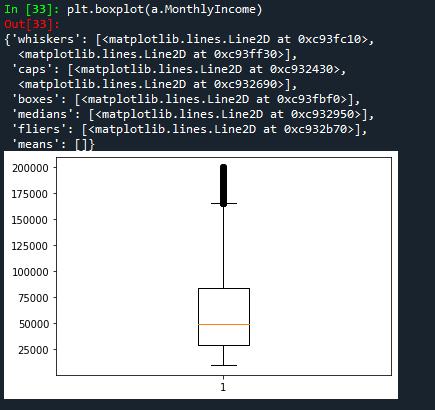
H0: there is no significance between Attrition yes Distance from home to Attrition no Distance from home

H1: there is significance between Attrition yes Distance from home to Attrition no Distance from home

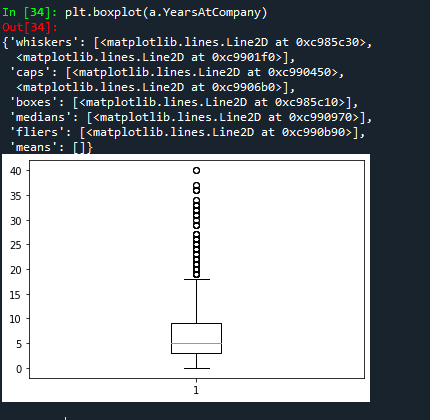
P value is grater than 0.5 so H0 is true



Age is normally distributed without any outliers



Monthly Income is Right skewed with several outliers less no of emplotee that get more salary



Years at company is also Right Skewed with several outliers observed.

The inference of the above analysis are as follows: Attrition & DistanceFromHome:

As r = -0.009, there’s low negative correlation between Attrition and DistanceFromHome

As the P value of 0.518 is > 0.05, we are accepting H0 and hence there’s no significant

correlation between Attrition & DistanceFromHome Attrition & MonthlyIncome:

As r = -0.031, there’s low negative correlation between Attrition and MonthlyIncome As the P value of 0.038 is < 0.05,

we are accepting Ha and hence there’s significant correlation between Attrition & MonthlyIncome Attrition & TotalWorkingYears:

As r = -0.17, there’s low negative correlation between Attrition and TotalWorkingYears

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & TotalWorkingYears Attrition & YearsAtCompany:

As r = -0.1343, there’s low negative correlation between Attrition and YearsAtCompany

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & YearsAtCompany Attrition & YearsWithCurrManager:

As r = -0.1561, there’s low negative correlation between Attrition and YearsWithCurrManager As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & YearsWithCurrManager