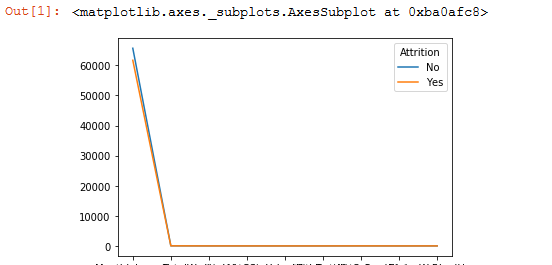
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | | Median | | Mode | | Skewness | | Kurt | | Variance | |
| Attrition | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| MonthlyIncome | 65672.6 | 61682.62 | 49300 | 49080 | 23420 | 25590 | 1.34 | 1.54 | 0.9 | 1.66 | 2.254E+09 | 2.006E+09 |
| Age | 37.56 | 33.61 | 36 | 32 | 34 | 29 | 0.41 | 0.71 | -0.41 | -0.07 | 78.96 | 93.62 |
| TotalWorkingYears | 11.86 | 8.26 | 10 | 7 | 10 | 1 | 1.07 | 1.68 | 0.67 | 3.7 | 60.26 | 51.32 |
| YearsAtCompany | 7.37 | 5.13 | 6 | 3 | 5 | 1 | 1.66 | 2.67 | 3.34 | 9.46 | 37.14 | 35.3 |
| YearsSinceLastPromotion | 2.23 | 1.95 | 1 | 1 | 0 | 0 | 1.95 | 2.21 | 3.42 | 4.78 | 10.46 | 9.91 |
| YearsWithCurrManager | 4.37 | 2.85 | 3 | 2 | 2 | 0 | 0.8 | 1.03 | 0.14 | 0.24 | 12.91 | 9.85 |
| PercentSalaryHike | 15.16 | 15.48 | 14 | 14 | 11 | 13 | 0.83 | 0.76 | -0.28 | -0.42 | 13.21 | 14.25 |
| NumCompaniesWorked | 2.65 | 2.94 | 2 | 1 | 1 | 1 | 1.06 | 0.87 | 0.14 | -0.55 | 6.05 | 7.18 |
| DistanceFromHome | 9.23 | 9.01 | 7 | 7 | 1 | 2 | 0.96 | 0.96 | -0.26 | -0.04 | 66.72 | 60.41 |

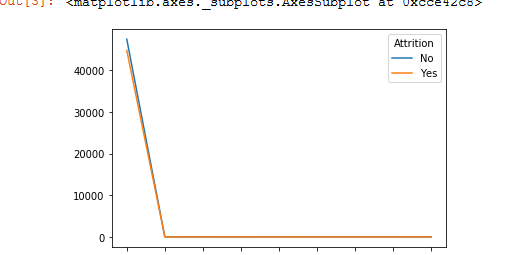
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Std | | 25% | | 50% | | 75% | | IQR | |
| Attrition | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| MonthlyIncome | 47472.8 | 44792.07 | 29360 | 28440 | 49300 | 49080 | 86060 | 71040 | 56700 | 42600 |
| Age | 8.89 | 9.68 | 31 | 28 | 36 | 32 | 43 | 39 | 12 | 11 |
| TotalWorkingYears | 7.76 | 7.16 | 6 | 3 | 10 | 7 | 16 | 10 | 10 | 7 |
| YearsAtCompany | 6.09 | 5.94 | 3 | 1 | 6 | 3 | 10 | 7 | 7 | 6 |
| YearsSinceLastPromotion | 3.23 | 3.15 | 0 | 0 | 1 | 1 | 3 | 2 | 3 | 2 |
| YearsWithCurrManager | 3.59 | 3.14 | 2 | 0 | 3 | 2 | 7 | 5 | 5 | 5 |
| PercentSalaryHike | 3.63 | 3.78 | 12 | 12 | 14 | 14 | 18 | 18 | 6 | 6 |
| NumCompaniesWorked | 2.46 | 2.68 | 1 | 1 | 2 | 1 | 4 | 5 | 3 | 4 |
| DistanceFromHome | 8.17 | 7.77 | 2 | 2 | 7 | 7 | 14 | 15 | 12 | 13 |

Above are the various statistics.

Graph of mean for Attrition Y and N



Graph of Std Deviation of the statistical values of Attrition Yes and No

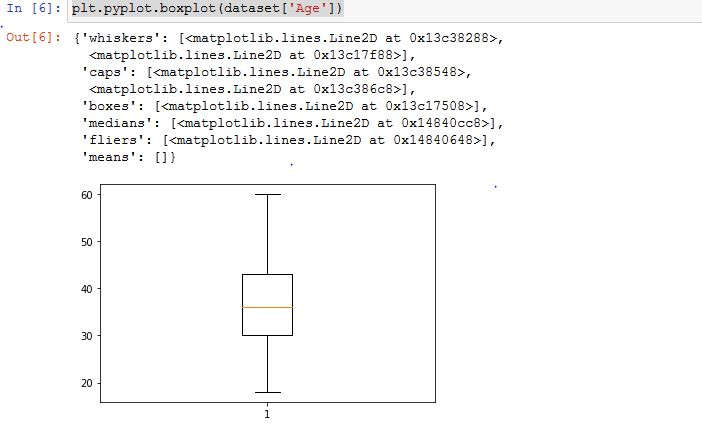


Inference 1 : There is no major difference in the statistics depending on Attrition. We will check further which parameter influences attrition

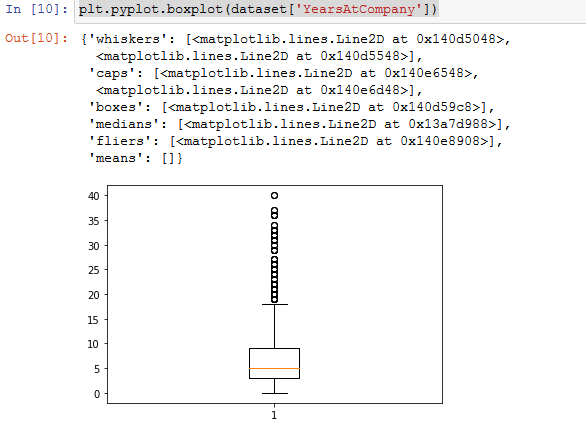
**Outliers**

**Box Plots of various parameters**

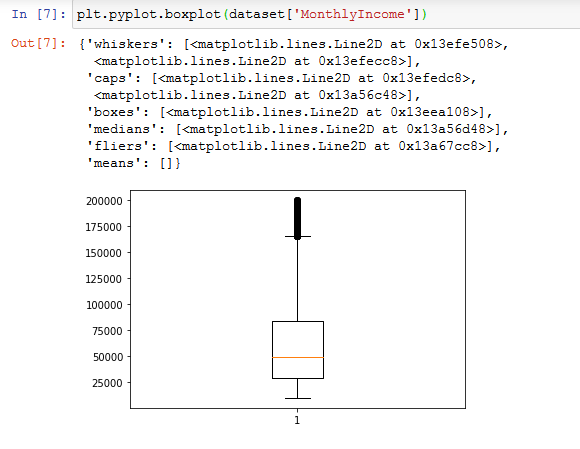
**Age**

****

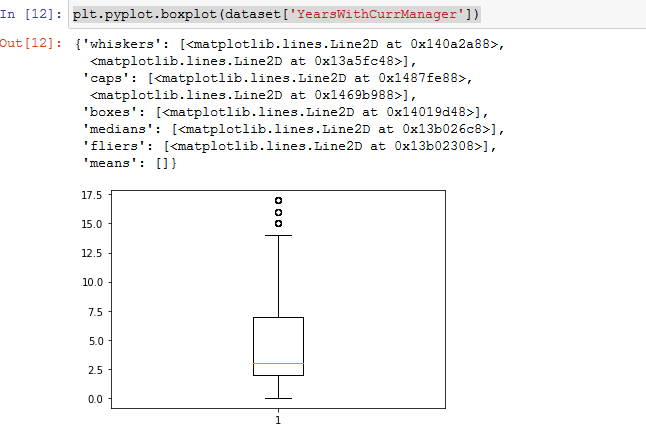
**There are no outliers in Age**

****

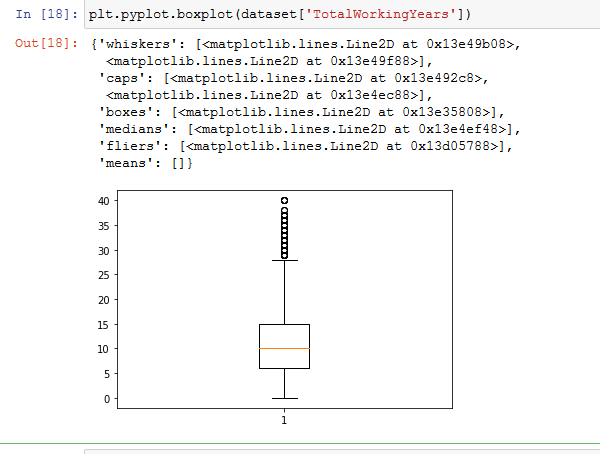
**Years at company has several outliers**

****

**Monthly Income is rightly skewed with many outliers**

****

**YearswithCurrentManager is rightly skewed with few outliers**

****

**Total working years is rightly skewed with many outliers**

**Correlation Table**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Age** | **Attrition** | **DistanceFromHome** | **Education** | **JobLevel** | **MonthlyIncome** | **NumCompaniesWorked** | **PercentSalaryHike** |
| **Age** | 1 | -0.159205 | 0.006963 | -0.035706 | -0.002884 | -0.044314 | 0.298528 | -0.033137 |
| **Attrition** | -0.159205 | 1 | -0.00973 | -0.015111 | -0.01029 | -0.031176 | 0.042232 | 0.032533 |
| **DistanceFromHome** | 0.006963 | -0.00973 | 1 | -0.008638 | -0.037329 | -0.021607 | -0.013818 | 0.038125 |
| **Education** | -0.035706 | -0.015111 | -0.008638 | 1 | 0.045746 | 0.00641 | -0.016219 | -0.040531 |
| **EmployeeID** | 0.008649 | -0.004729 | -0.001097 | -0.00968 | -0.003303 | 0.007338 | -0.001114 | -0.004456 |
| **JobLevel** | -0.002884 | -0.01029 | -0.037329 | 0.045746 | 1 | 0.047316 | -0.009844 | 0.010973 |
| **MonthlyIncome** | -0.044314 | -0.031176 | -0.021607 | 0.00641 | 0.047316 | 1 | -0.02071 | 0.004325 |
| **NumCompaniesWorked** | 0.298528 | 0.042232 | -0.013818 | -0.016219 | -0.009844 | -0.02071 | 1 | 0.030825 |
| **PercentSalaryHike** | -0.033137 | 0.032533 | 0.038125 | -0.040531 | 0.010973 | 0.004325 | 0.030825 | 1 |
| **StockOptionLevel** | -0.031753 | -0.006839 | 0.011169 | 0.001261 | 0.000993 | 0.02693 | 0.01713 | 0.012548 |
| **TotalWorkingYears** | 0.680419 | -0.170111 | 0.009361 | -0.010712 | -0.036934 | -0.033758 | 0.238052 | -0.018717 |
| **TrainingTimesLastYear** | -0.027308 | -0.049431 | -0.009001 | 0.010472 | -0.0325 | 0.050112 | -0.032239 | -0.037392 |
| **YearsAtCompany** | 0.311309 | -0.134392 | 0.031684 | 0.00608 | -0.064219 | 0.000995 | -0.11782 | -0.029707 |
| **YearsSinceLastPromotion** | 0.216513 | -0.033019 | 0.00229 | 0.02249 | -0.060811 | 0.065219 | -0.036618 | -0.029542 |
| **YearsWithCurrManager** | 0.202089 | -0.156199 | 0.021584 | 0.005358 | -0.055251 | 0.024304 | -0.10948 | -0.040864 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **StockOptionLevel** | **TotalWorkingYears** | **TrainingTimesLastYear** | **YearsAtCompany** | **YearsSinceLastPromotion** | **YearsWithCurrManager** |
| -0.031753 | 0.680419 | -0.027308 | 0.311309 | 0.216513 | 0.202089 |
| -0.006839 | -0.170111 | -0.049431 | -0.134392 | -0.033019 | -0.156199 |
| 0.011169 | 0.009361 | -0.009001 | 0.031684 | 0.00229 | 0.021584 |
| 0.001261 | -0.010712 | 0.010472 | 0.00608 | 0.02249 | 0.005358 |
| -0.014254 | -0.001061 | -0.010191 | 0.004086 | 0.000256 | 0.008579 |
| 0.000993 | -0.036934 | -0.0325 | -0.064219 | -0.060811 | -0.055251 |
| 0.02693 | -0.033758 | 0.050112 | 0.000995 | 0.065219 | 0.024304 |
| 0.01713 | 0.238052 | -0.032239 | -0.11782 | -0.036618 | -0.10948 |
| 0.012548 | -0.018717 | -0.037392 | -0.029707 | -0.029542 | -0.040864 |
| 1 | 0.002693 | -0.069902 | 0.007886 | 0.019063 | 0.017757 |
| 0.002693 | 1 | -0.040705 | 0.626876 | 0.403995 | 0.458425 |
| -0.069902 | -0.040705 | 1 | -0.007894 | 0.016121 | -0.01327 |
| 0.007886 | 0.626876 | -0.007894 | 1 | 0.618409 | 0.769212 |
| 0.019063 | 0.403995 | 0.016121 | 0.618409 | 1 | 0.510224 |
| 0.017757 | 0.458425 | -0.01327 | 0.769212 | 0.510224 | 1 |

**Although not very high , Attrition seems to have a negative co-relation with the parameters in the order. Higher the below paramaters, lower the attrition rate**

|  |  |
| --- | --- |
| Total Working Years | -0.170111 |
| Age | -0.159205 |
| Years with Current Manager | -0.156199 |
| Years ay Company | -0.134392 |

**Statistical Tests**

**Mann-Whitney Tests**

**## H0 There is no significance difference in the total working years for Attrition(Y)and Attrition(N)**

stats,p=mannwhitneyu(ds\_test\_Yes['TotalWorkingYears'],ds\_test\_No['TotalWorkingYears'])

print(stats,p)

print(p<0.05)

908192.0 1.3683394202436042e-39

True

## Ha accepted - There is significance difference in the total working years for Attrition(Y)and Attrition(N)

**## H0 There is no significance difference in the Age for Attrition(Y)and Attrition(N)**

**## Ha There is significance difference in the Age for Attrition(Y)and Attrition(N)**

stats,p=mannwhitneyu(ds\_test\_Yes['Age'],ds\_test\_No['Age'])

print(stats,p)

print(p<0.05)

##As p<0.05 , H0 is rejected

## Ha accepted There is significance difference in the Age for Attrition(Y)and Attrition(N)

**## H0 There is no significance difference in the DistanceFromHome for Attrition(Y)and Attrition(N)**

**## Ha There is significance difference in the DistanceFromHome for Attrition(Y)and Attrition(N)**

stats,p=mannwhitneyu(ds\_test\_Yes['DistanceFromHome'],ds\_test\_No['DistanceFromHome'])

print(stats,p)

print(p<0.05)

**##As p<0.05 , H0 is accepted**

**## H0 There is no significance difference in the YearsWithCurrManager for Attrition(Y)and Attrition(N)**

**## Ha There is significance difference in the YearsWithCurrManager for Attrition(Y)and Attrition(N)**

stats,p=mannwhitneyu(ds\_test\_Yes['YearsWithCurrManager'],ds\_test\_No['YearsWithCurrManager'])

print(stats,p)

print(p<0.05)

957253.5 1.2365483142169853e-31

True

## Ha accepted - There is significance difference in the YearsWithCurrManager for Attrition(Y)and Attrition(N)

**## H0 There is no significance difference in the YearsAtCompany for Attrition(Y)and Attrition(N)**

**## Ha There is significance difference in the YearsAtCompany for Attrition(Y)and Attrition(N)**

from scipy.stats import mannwhitneyu

stats,p=mannwhitneyu(ds\_test\_Yes['YearsAtCompany'],ds\_test\_No['YearsAtCompany'])

print(stats,p)

print(p<0.05)

923238.0 6.047598261692858e-37

True

##As p<0.05 , H0 is rejected

## Ha There is significance difference in the YearsAtCompany for Attrition(Y)and Attrition(N)

**## H0 There is no significance difference in the YearsSinceLastPromotion for Attrition(Y)and Attrition(N)**

**## Ha There is significance difference in the YearsSinceLastPromotion for Attrition(Y)and Attrition(N)**

stats,p=mannwhitneyu(ds\_test\_Yes['YearsSinceLastPromotion'],ds\_test\_No['YearsSinceLastPromotion'])

print(stats,p)

print(p<0.05)

1209366.0 0.0002021180346719736

True

##As p<0.05 , H0 is rejected

## Ha There is significance difference in the YearsSinceLastPromotion for Attrition(Y)and Attrition(N)

**Chi-Square test**

##H0- there is no dependency of JobRole and Attrition

##Ha- there is dependency between JobRole and Attrition

from scipy.stats import chi2\_contingency

ds\_chi\_cont\_table=pd.crosstab(dataset['JobRole'],dataset['Attrition'])

stats,p,dof,expected=chi2\_contingency(ds\_chi\_cont\_table)

print(p<0.05)

True

##Ha- there is dependency between JobRole and Attrition

## H0- there is no relation between Gender and Attrition

from scipy.stats import chi2\_contingency

ds\_chi\_cont\_table=pd.crosstab(dataset['Gender'],dataset['Attrition'])

stats,p,dof,expected=chi2\_contingency(ds\_chi\_cont\_table)

print(p<0.05)

## H0 accepted

## H0- there is no relation between Business Travel and Attrition

## Ha- there is relation between Business Travel and Attrition

from scipy.stats import chi2\_contingency

ds\_chi\_cont\_table=pd.crosstab(dataset['BusinessTravel'],dataset['Attrition'])

stats,p,dof,expected=chi2\_contingency(ds\_chi\_cont\_table)

print(p<0.05)

## H0 rejected

## Ha- there is relation between Business Travel and Attrition

**## H0- there is no relation between Marital Status and Attrition**

**## Ha- there is relation between Marital Status and Attrition**

from scipy.stats import chi2\_contingency

ds\_chi\_cont\_table=pd.crosstab(dataset['MaritalStatus'],dataset['Attrition'])

stats,p,dof,expected=chi2\_contingency(ds\_chi\_cont\_table)

print(p<0.05)

## H0 rejected

## Ha- there is relation between Marital Status and Attrition

**Inference**

From the statistical , correlational and inferential tests the observations are :

There is negative correlation between Attrition and

Total WOrking Years - Less the total experience , more the possibility of attrition

Age - Lesser the age , more the possibility of attrition

Years with company - More the number of years with a manager, lesser the possibility of attrition

Years with company - More the number of years in company , lesser the possibility of attrition

Job role also is a deciding factor for Attrition

There is no particular relation between Gender and Attrition