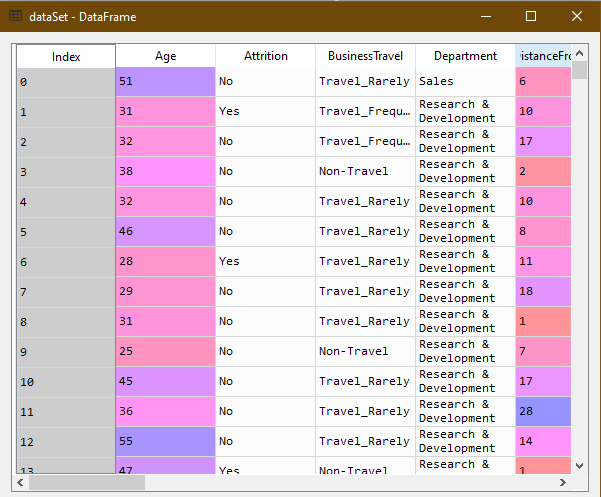
# Day 7 Assignment Solution

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

dataSet = pd.read\_csv('general\_data.csv')



dataSet.head()

Out[27]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 51 No ... 0 0

1 31 Yes ... 1 4

2 32 No ... 0 3

3 38 No ... 7 5

4 32 No ... 0 4

[5 rows x 24 columns]

dataSet.isnull()

* No null value found

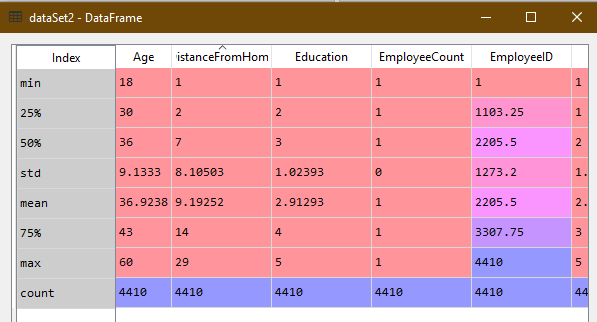
dataSet.duplicates()

* No duplicated found

dataSet.drop\_duplicates()

* No row dropped

dataSet2 = dataSet.describe()



dataSet[['Age','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()

Out[29]:

Age DistanceFromHome ... YearsSinceLastPromotion YearsWithCurrManager

0 35 2 ... 0 2

[1 rows x 11 columns]

plt.boxplot(dataSet.Age) # No outliers found

Out[37]:

{'whiskers': [<matplotlib.lines.Line2D at 0x294ed5a9470>,

<matplotlib.lines.Line2D at 0x294ed5a97b8>],

'caps': [<matplotlib.lines.Line2D at 0x294ed5a9b00>,

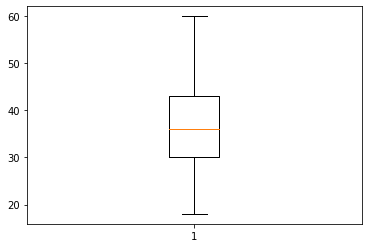
<matplotlib.lines.Line2D at 0x294ed5a9e48>],

'boxes': [<matplotlib.lines.Line2D at 0x294ed5a9048>],

'medians': [<matplotlib.lines.Line2D at 0x294ed5a9f28>],

'fliers': [<matplotlib.lines.Line2D at 0x294ed801518>],

'means': []}



# Inference from the analysis:

* All the variables have positive skewness except Education which has negative skewness.
* Age, DistanceFromHome, Education, PercentSalaryHike are platykurtic. Rest all other variables are leptokurtic.