**Attrition Assignment**

* **Importing the file**

import pandas as pad

data1=pad.read\_csv('general\_data.csv')

* **to read first 5 employees**

data1.head()

Out[9]:

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]

* **To check columns**

data1.columns

Out[11]:

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')

* **To check null records in data if present.**

data1.isnull()

Out[12]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 False False ... False False

1 False False ... False False

2 False False ... False False

3 False False ... False False

4 False False ... False False

... ... ... ... ...

4405 False False ... False False

4406 False False ... False False

4407 False False ... False False

4408 False False ... False False

4409 False False ... False False

[4410 rows x 24 columns]

* **To delete records that contains null**

data1.dropna()

Out[13]:

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

... ... ... ... ...

4404 29 No ... 1 5

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

[4382 rows x 24 columns]

* **Checking duplicate records if exist**

data1.duplicated()

Out[14]:

0 False

1 False

2 False

3 False

4 False

4405 False

4406 False

4407 False

4408 False

4409 False

Length: 4410, dtype: bool

* **Remove duplicate records**

data1.drop\_duplicates()

Out[15]:

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

... ... ... ... ...

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

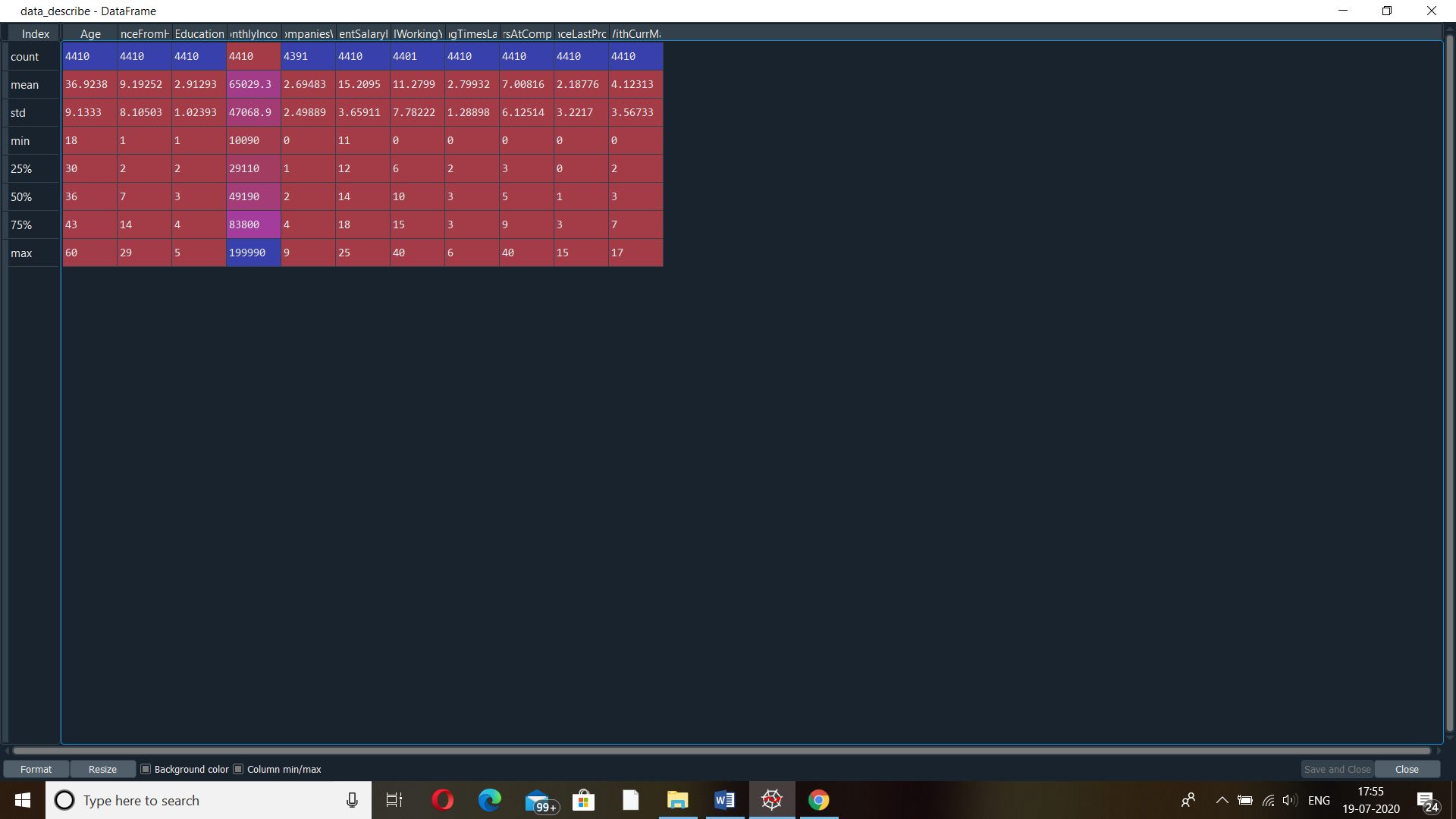
4408 42 No ... 7 8

4409 40 No ... 3 9

[4410 rows x 24 columns]

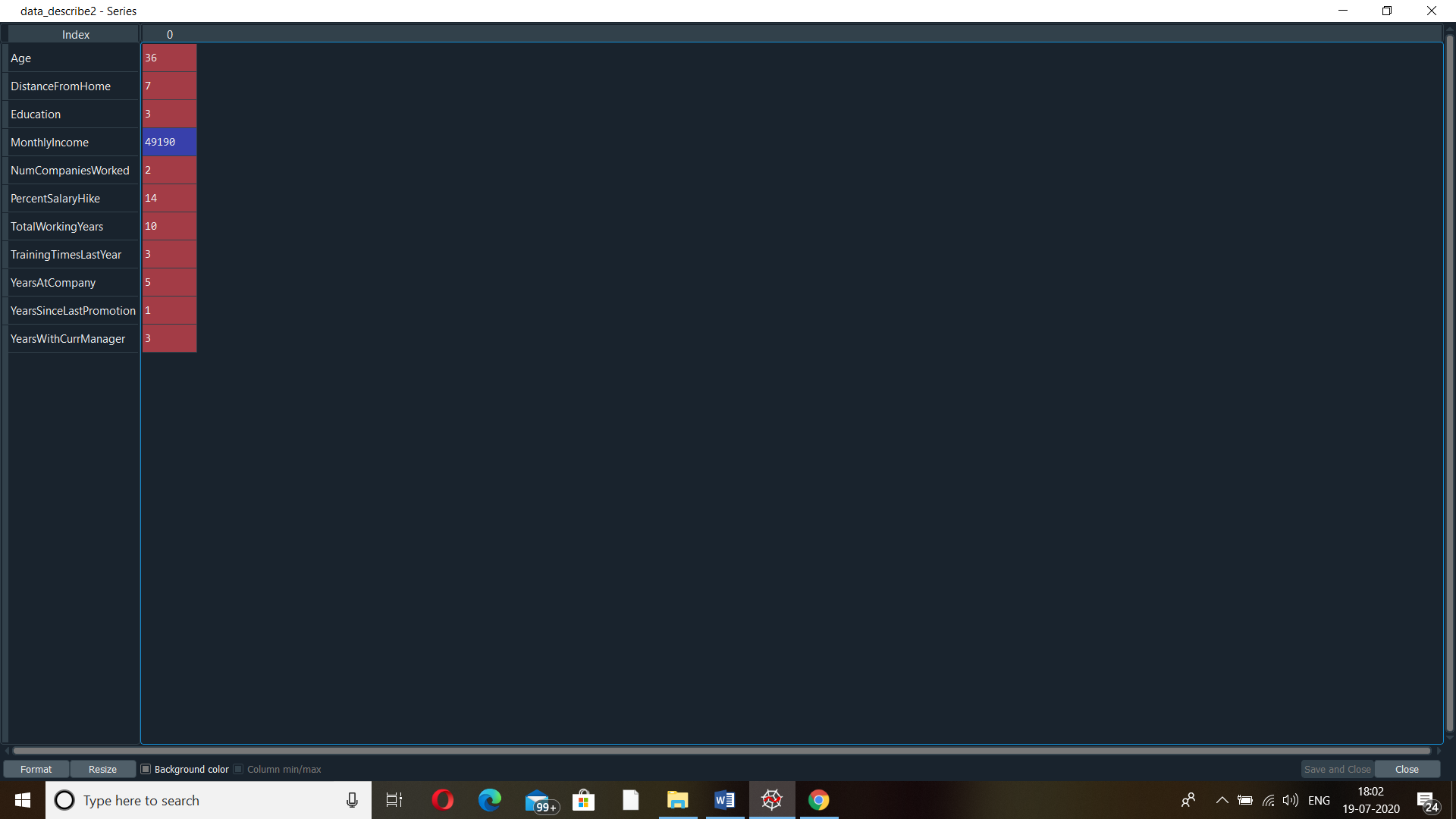
* **Univariate analysis**

data\_describe=data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()



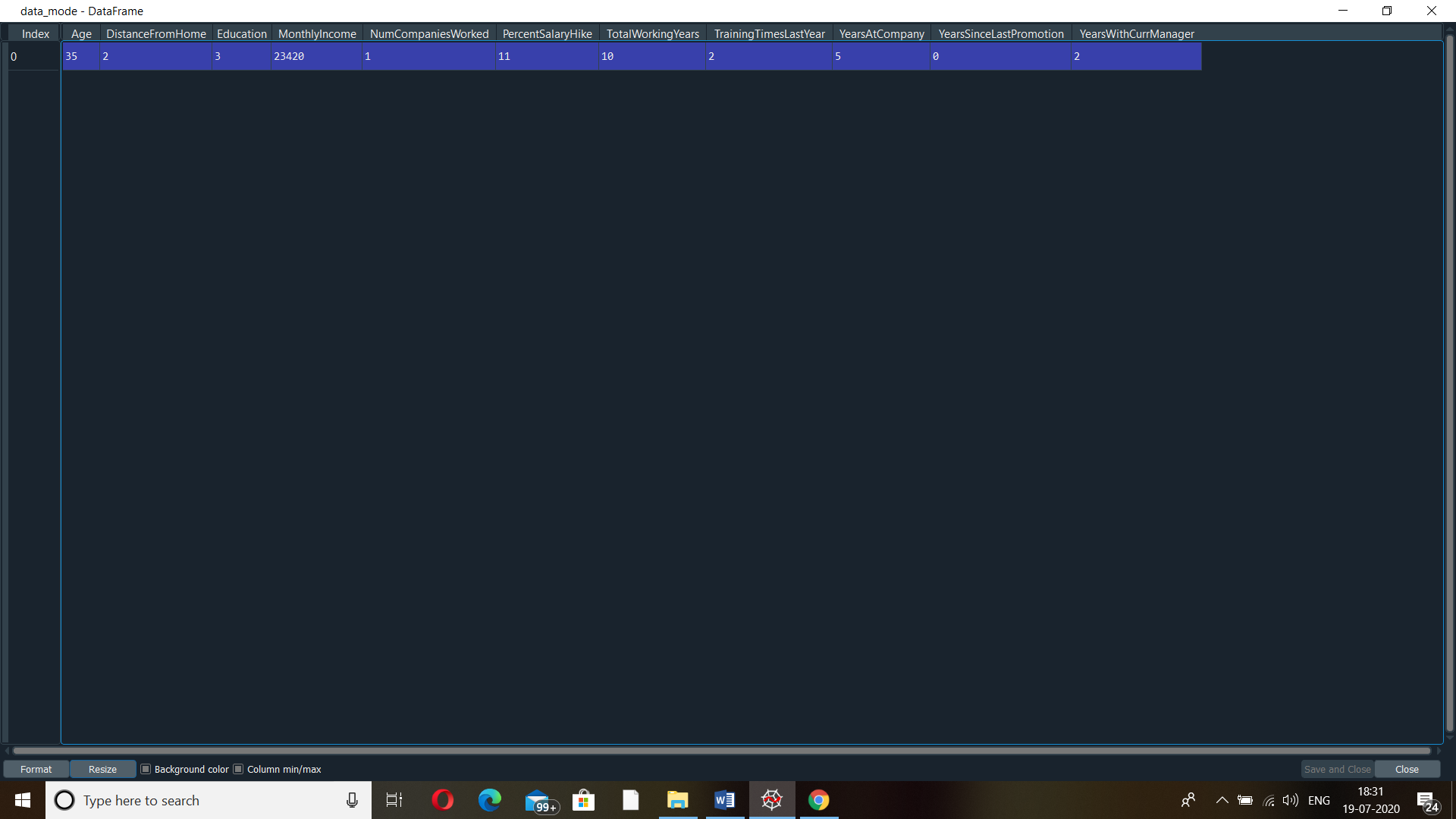
* **Median()**

data\_describe2=data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].median()



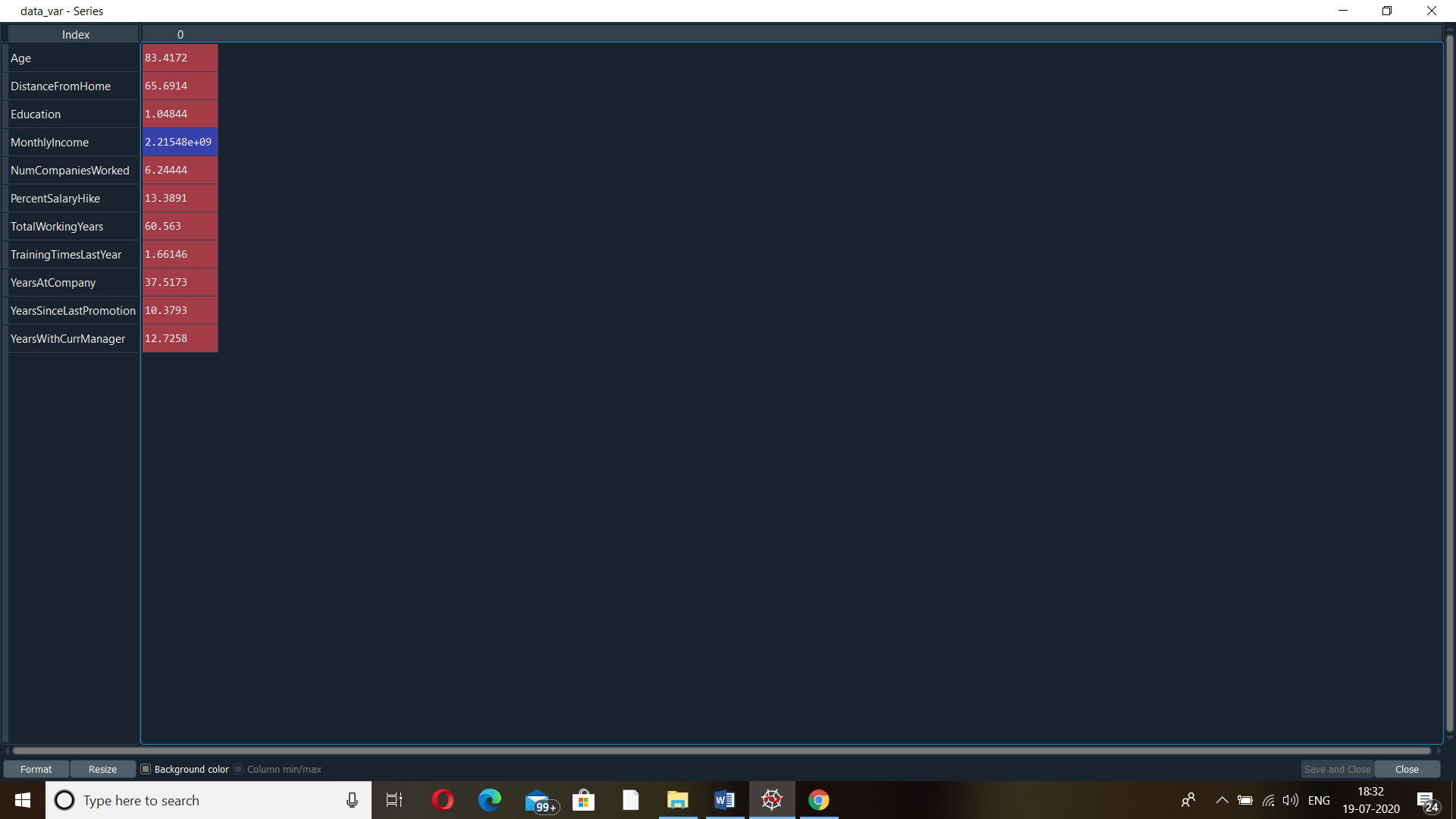
* **Mode()**

data\_mode=data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()



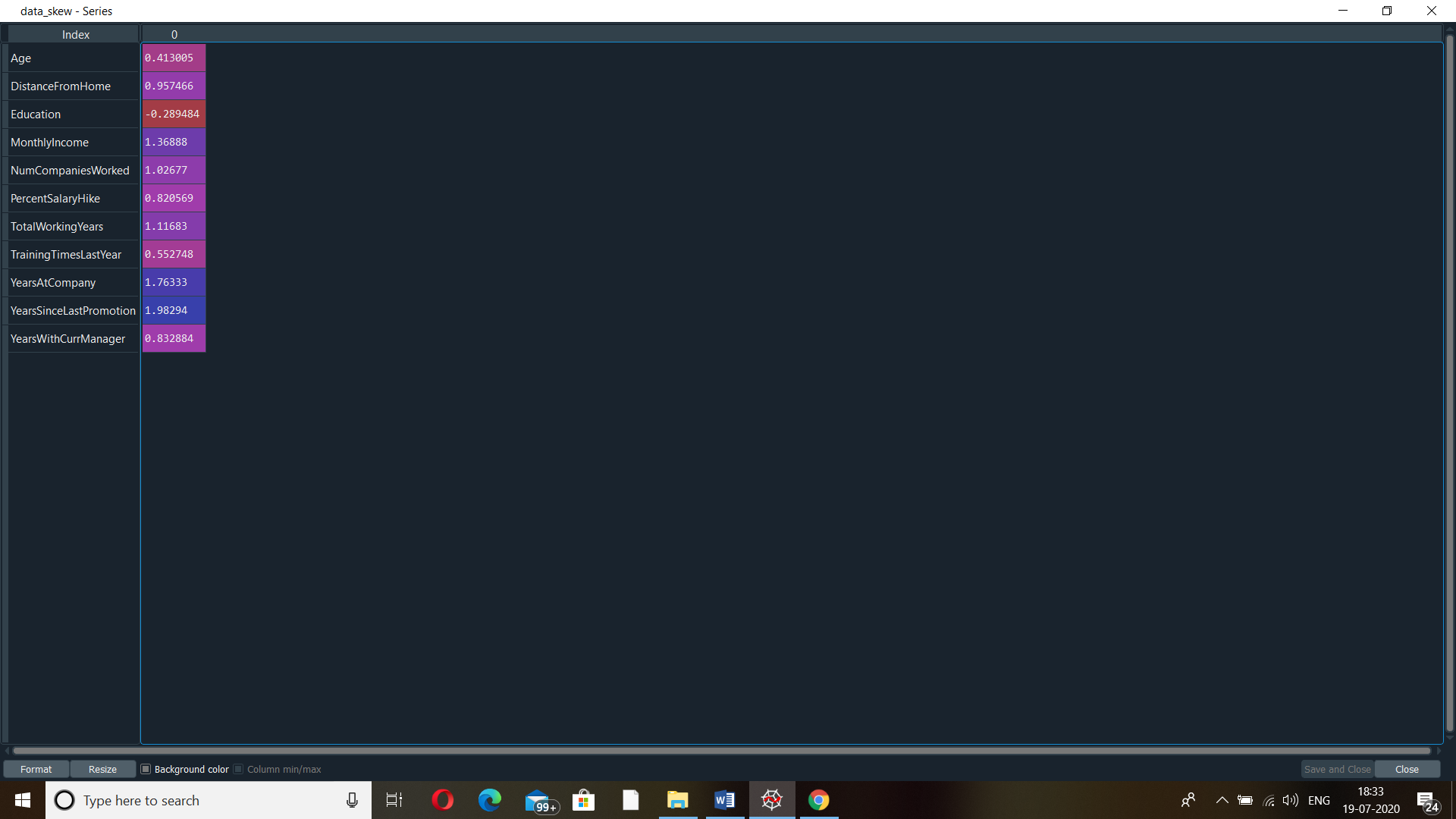
* **Variance**

data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].var()



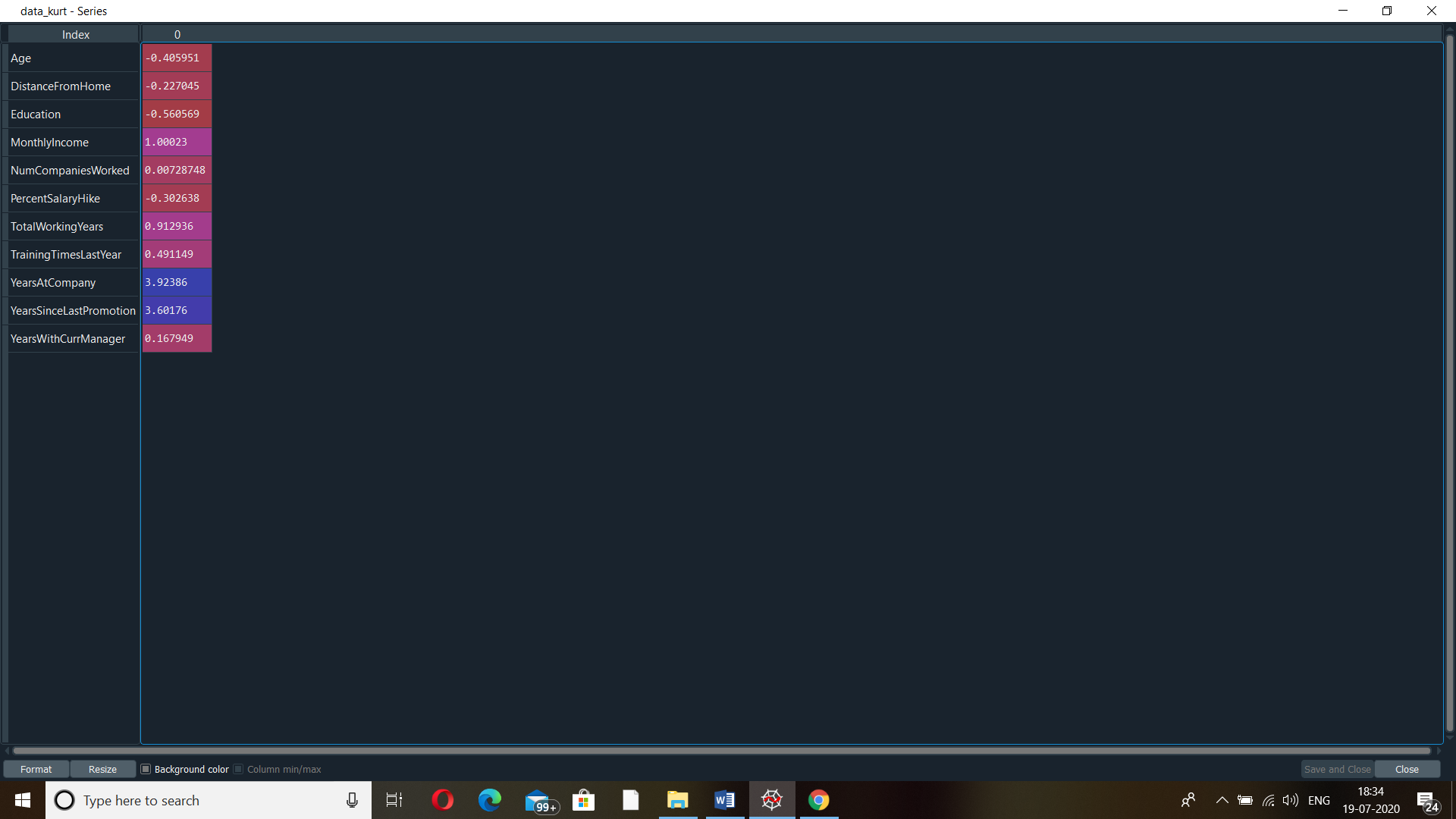
* **Skew**

data\_skew=data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].skew()



* **Kurtosis**

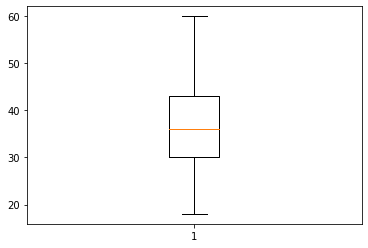
data\_kurt=data1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].kurt()



* **Boxplot (AGE)**

import matplotlib.pyplot as ploting

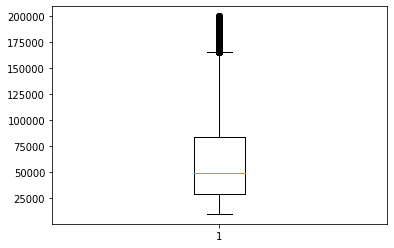
ploting.boxplot(data1.age)

****

**Here plot showing normal distribution .**

* **Boxplot ( monthly income)**

ploting.boxplot(data1.MonthlyIncome)

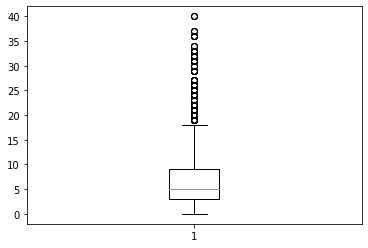
****

Outliers are present here in this plot which show s employees who sre getting good salary comparing others .

The other thing to observe is mean grester than median that means **right-skewed distribution.**

* **Boxplot (years at company)**

ploting.boxplot(data1.YearsAtCompany)

****

**Here in this plot outliers are present this means some employees working from long time in the company**

**Inference from the analysis**

1. Age Box plot showing normal distribution so not making any difference in the analysis .
2. In the case of the monthly income and years of working Boxplot we can see outliers .
3. Positive values of kurtosis in case of monthlyincome and years at company  indicate that is Leptokurtic distribution.