import pandas as pd

df = pd.read\_csv("file:///C:/Users/OMKAR/Downloads/general\_data.csv")

df.columns

Out[3]:

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')

df.head(10)

Out[4]:

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 46 No ... 7 7

6 28 Yes ... 0 0

7 29 No ... 0 0

8 31 No ... 7 8

9 25 No ... 1 5

[10 rows x 24 columns]

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 4410 entries, 0 to 4409

Data columns (total 24 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Age 4410 non-null int64

1 Attrition 4410 non-null object

2 BusinessTravel 4410 non-null object

3 Department 4410 non-null object

4 DistanceFromHome 4410 non-null int64

5 Education 4410 non-null int64

6 EducationField 4410 non-null object

7 EmployeeCount 4410 non-null int64

8 EmployeeID 4410 non-null int64

9 Gender 4410 non-null object

10 JobLevel 4410 non-null int64

11 JobRole 4410 non-null object

12 MaritalStatus 4410 non-null object

13 MonthlyIncome 4410 non-null int64

14 NumCompaniesWorked 4391 non-null float64

15 Over18 4410 non-null object

16 PercentSalaryHike 4410 non-null int64

17 StandardHours 4410 non-null int64

18 StockOptionLevel 4410 non-null int64

19 TotalWorkingYears 4401 non-null float64

20 TrainingTimesLastYear 4410 non-null int64

21 YearsAtCompany 4410 non-null int64

22 YearsSinceLastPromotion 4410 non-null int64

23 YearsWithCurrManager 4410 non-null int64

dtypes: float64(2), int64(14), object(8)

memory usage: 827.0+ KB

df[['Age','MonthlyIncome','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager','MonthlyIncome']].mode()

Out[6]:

Age MonthlyIncome ... YearsWithCurrManager MonthlyIncome

0 35 23420 ... 2 23420

[1 rows x 6 columns]

df[['Age','MonthlyIncome','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager','MonthlyIncome']].mean()

Out[7]:

Age 36.923810

MonthlyIncome 65029.312925

YearsAtCompany 7.008163

YearsSinceLastPromotion 2.187755

YearsWithCurrManager 4.123129

MonthlyIncome 65029.312925

dtype: float64

df[['Age','MonthlyIncome','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager','MonthlyIncome']].median()

Out[8]:

Age 36.0

MonthlyIncome 49190.0

YearsAtCompany 5.0

YearsSinceLastPromotion 1.0

YearsWithCurrManager 3.0

MonthlyIncome 49190.0

dtype: float64

df[['Age','MonthlyIncome','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager','MonthlyIncome']].skew()

Out[9]:

Age 0.413005

MonthlyIncome 1.368884

YearsAtCompany 1.763328

YearsSinceLastPromotion 1.982939

YearsWithCurrManager 0.832884

MonthlyIncome 1.368884

dtype: float64

df[['Age','MonthlyIncome','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager','MonthlyIncome']].kurt()

Out[10]:

Age -0.405951

MonthlyIncome 1.000232

YearsAtCompany 3.923864

YearsSinceLastPromotion 3.601761

YearsWithCurrManager 0.167949

MonthlyIncome 1.000232

dtype: float64

df.describe()

Out[11]:

Age ... YearsWithCurrManager

count 4410.000000 ... 4410.000000

mean 36.923810 ... 4.123129

std 9.133301 ... 3.567327

min 18.000000 ... 0.000000

25% 30.000000 ... 2.000000

50% 36.000000 ... 3.000000

75% 43.000000 ... 7.000000

max 60.000000 ... 17.000000

[8 rows x 16 columns]

desc = df.describe()

import matplot.pyplot as plt

Traceback (most recent call last):

File "<ipython-input-13-9669cae5d6fe>", line 1, in <module>

import matplot.pyplot as plt

ModuleNotFoundError: No module named 'matplot'

import matplotlib.pyplot as plt

plt.boxplot(df.MonthlyIncome)

Out[15]:

{'whiskers': [<matplotlib.lines.Line2D at 0x1fa393d1608>,

<matplotlib.lines.Line2D at 0x1fa393eac88>],

'caps': [<matplotlib.lines.Line2D at 0x1fa393ead88>,

<matplotlib.lines.Line2D at 0x1fa393eae48>],

'boxes': [<matplotlib.lines.Line2D at 0x1fa3445e388>],

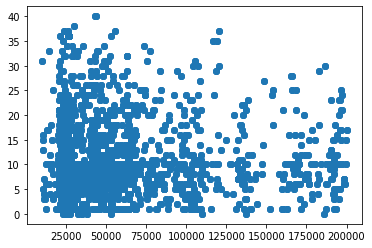
'medians': [<matplotlib.lines.Line2D at 0x1fa393f7f48>],

'fliers': [<matplotlib.lines.Line2D at 0x1fa393f7ac8>],

'means': []}

￼

plt.scatter(df.MonthlyIncome,df.TotalWorkingYears)



plt.hist(df.MonthlyIncome)

Out[17]:

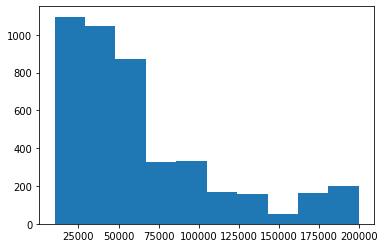
(array([1095., 1047., 870., 327., 330., 168., 156., 54., 162.,

201.]),

array([ 10090., 29080., 48070., 67060., 86050., 105040., 124030.,

143020., 162010., 181000., 199990.]),

<a list of 10 Patch objects>)



sol = df[df['MonthlyIncome']>175000]

solution = df[df['MonthlyIncome']<175000]

solution.describe()

Out[22]:

Age ... YearsWithCurrManager

count 4167.000000 ... 4167.000000

mean 36.963283 ... 4.111591

std 9.115799 ... 3.546045

min 18.000000 ... 0.000000

25% 30.000000 ... 2.000000

50% 36.000000 ... 3.000000

75% 43.000000 ... 7.000000

max 60.000000 ... 17.000000

[8 rows x 16 columns]