AI-ML ASSIGNMENT DAY-7(statistical learning project)

Python 3.6.3 |Anaconda custom (32-bit)| (default, Oct 15 2017, 07:29:16) [MSC v.1900 32 bit (Intel)]

Type "copyright", "credits" or "license" for more information.

IPython 6.1.0 -- An enhanced Interactive Python.

import pandas as pd

dataset=pd.read\_excel("3. Descriptive Statistics.xlsx",sheetname=0)

dataset.head()

Out[3]:

   ID Gender Birth Date  Education  JobCategory  CurrentSalary  After6Months  \

0   3      f 1929-07-26         12            1          21450       16725.0

1   4      f 1947-04-15          8            1          21900       17550.0

2   8      f 1966-06-05         12            1          21900       15825.0

3   9      f 1946-01-23         15            1          27900       20325.0

4  10      f 1946-02-13         12            1          24000       18750.0

   SalBegin  Job Time  Prev Exep  Minority

0     12000        98        381         0

1     13200        98        190         0

2      9750        98          0         0

3     12750        98        115         0

4     13500        98        244         0

dataset['CurrentSalary'].mean()

Out[4]: 34419.56751054852

dataset['CurrentSalary'].median()

Out[5]: 28875.0

dataset['CurrentSalary'].mode()

Out[6]:

0    30750

dtype: int64

dataset['CurrentSalary'].std()

Out[7]: 17075.66146458606

dataset['CurrentSalary'].var()

Out[8]: 291578214.45314944

dataset['CurrentSalary'].describe()

Out[9]:

count       474.000000

mean      34419.567511

std       17075.661465

min       15750.000000

25%       24000.000000

50%       28875.000000

75%       36937.500000

max      135000.000000

Name: CurrentSalary, dtype: float64

dataset['CurrentSalary'].skew()

Out[10]: 2.1246062823682506

dataset['CurrentSalary'].kurt()

Out[11]: 5.3778223961367821

import matplotlib.pyplot as plt

plt.boxplot(dataset['CurrentSalary'])

Out[13]:

{'boxes': [<matplotlib.lines.Line2D at 0xa7771d0>],

 'caps': [<matplotlib.lines.Line2D at 0xa2cab90>,

  <matplotlib.lines.Line2D at 0xa77c510>],

 'fliers': [<matplotlib.lines.Line2D at 0xa77c790>],

 'means': [],

 'medians': [<matplotlib.lines.Line2D at 0xa77c7f0>],

 'whiskers': [<matplotlib.lines.Line2D at 0xa777870>,

  <matplotlib.lines.Line2D at 0xa777e50>]}

plt.scatter(dataset['CurrentSalary'],dataset['After6Months'])

Out[14]: <matplotlib.collections.PathCollection at 0x24294d0>