import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

dataset1=pd.read\_excel("data\_dictionary.xlsx",sheet\_name=0)

dataset1.head()

Out[5]:

Variable Meaning Levels

0 Age Age of the employee NaN

1 Attrition Whether the employee left in the previous year... NaN

2 BusinessTravel How frequently the employees travelled for bus... NaN

3 Department Department in company NaN

4 DistanceFromHome Distance from home in kms NaN

dataset1.columns

Out[6]: Index(['Variable', 'Meaning', 'Levels'], dtype='object')

#data treatment

dataset1.null()

Traceback (most recent call last):

File "<ipython-input-8-af3957df1c04>", line 1, in <module>

dataset1.null()

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\generic.py", line 5274, in \_\_getattr\_\_

return object.\_\_getattribute\_\_(self, name)

AttributeError: 'DataFrame' object has no attribute 'null'

dataset1.isnull()

Out[9]:

Variable Meaning Levels

0 False False True

1 False False True

2 False False True

3 False False True

4 False False True

5 False False False

6 True True False

7 True True False

8 True True False

9 True True False

10 False False True

11 False False True

12 False False True

13 False False False

14 True True False

15 True True False

16 True True False

17 False False True

18 False False False

19 True True False

20 True True False

21 True True False

22 False False True

23 False False True

24 False False False

25 True True False

26 True True False

27 True True False

28 False False True

29 False False True

30 False False True

31 False False True

32 False False True

33 False False False

34 True True False

35 True True False

36 True True False

37 False False False

38 True True False

39 True True False

40 True True False

41 False False True

42 False False True

43 False False True

44 False False True

45 False False False

46 True True False

47 True True False

48 True True False

49 False False True

50 False False True

51 False False True

dataset1.duplicated()

Out[10]:

0 False

1 False

2 False

3 False

4 False

5 False

6 False

7 False

8 False

9 False

10 False

11 False

12 False

13 False

14 False

15 False

16 False

17 False

18 False

19 True

20 True

21 True

22 False

23 False

24 False

25 True

26 True

27 True

28 False

29 False

30 False

31 False

32 False

33 False

34 False

35 False

36 False

37 False

38 True

39 True

40 True

41 False

42 False

43 False

44 False

45 False

46 True

47 False

48 False

49 False

50 False

51 False

dtype: bool

dataset1.drop\_duplicates()

Out[11]:

Variable ... Levels

0 Age ... NaN

1 Attrition ... NaN

2 BusinessTravel ... NaN

3 Department ... NaN

4 DistanceFromHome ... NaN

5 Education ... 1 'Below College'

6 NaN ... 2 'College'

7 NaN ... 3 'Bachelor'

8 NaN ... 4 'Master'

9 NaN ... 5 'Doctor'

10 EducationField ... NaN

11 EmployeeCount ... NaN

12 EmployeeNumber ... NaN

13 EnvironmentSatisfaction ... 1 'Low'

14 NaN ... 2 'Medium'

15 NaN ... 3 'High'

16 NaN ... 4 'Very High'

17 Gender ... NaN

18 JobInvolvement ... 1 'Low'

22 JobLevel ... NaN

23 JobRole ... NaN

24 JobSatisfaction ... 1 'Low'

28 MaritalStatus ... NaN

29 MonthlyIncome ... NaN

30 NumCompaniesWorked ... NaN

31 Over18 ... NaN

32 PercentSalaryHike ... NaN

33 PerformanceRating ... 1 'Low'

34 NaN ... 2 'Good'

35 NaN ... 3 'Excellent'

36 NaN ... 4 'Outstanding'

37 RelationshipSatisfaction ... 1 'Low'

41 StandardHours ... NaN

42 StockOptionLevel ... NaN

43 TotalWorkingYears ... NaN

44 TrainingTimesLastYear ... NaN

45 WorkLifeBalance ... 1 'Bad'

47 NaN ... 3 'Better'

48 NaN ... 4 'Best'

49 YearsAtCompany ... NaN

50 YearsSinceLastPromotion ... NaN

51 YearsWithCurrManager ... NaN

[42 rows x 3 columns]

dataset3=dataset1[['Age','DistanceFromHome','TrainingTimesLastYear','WorkLifeBalance','YearsAtCompany','YearsSinceLastPromotion','YearsWithCurrManager']].var()

Traceback (most recent call last):

File "<ipython-input-7-5f3ed0e891a4>", line 1, in <module>

dataset3=dataset1[['Age','DistanceFromHome','TrainingTimesLastYear','WorkLifeBalance','YearsAtCompany','YearsSinceLastPromotion','YearsWithCurrManager']].var()

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py", line 2806, in \_\_getitem\_\_

indexer = self.loc.\_get\_listlike\_indexer(key, axis=1, raise\_missing=True)[1]

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexing.py", line 1552, in \_get\_listlike\_indexer

keyarr, indexer, o.\_get\_axis\_number(axis), raise\_missing=raise\_missing

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexing.py", line 1639, in \_validate\_read\_indexer

raise KeyError(f"None of [{key}] are in the [{axis\_name}]")

KeyError: "None of [Index(['Age', 'DistanceFromHome', 'TrainingTimesLastYear', 'WorkLifeBalance',\n 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],\n dtype='object')] are in the [columns]"

box\_plot=dataset1.Age

Traceback (most recent call last):

File "<ipython-input-8-d57db0f3d315>", line 1, in <module>

box\_plot=dataset1.Age

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\generic.py", line 5274, in \_\_getattr\_\_

return object.\_\_getattribute\_\_(self, name)

AttributeError: 'DataFrame' object has no attribute 'Age'

plt.boxplot(box\_plot)

Traceback (most recent call last):

File "<ipython-input-9-fad3be423771>", line 1, in <module>

plt.boxplot(box\_plot)

NameError: name 'box\_plot' is not defined

import matplotlib.pyplot as plt

plt.boxplot(dataset1.Currentsalary)

Traceback (most recent call last):

File "<ipython-input-11-1ec8bb54c444>", line 1, in <module>

plt.boxplot(dataset1.Currentsalary)

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\generic.py", line 5274, in \_\_getattr\_\_

return object.\_\_getattribute\_\_(self, name)

AttributeError: 'DataFrame' object has no attribute 'Currentsalary'