import pandas as pd

data = pd.read\_csv("E:\Latha\LathaSKPIMCS\Machine Learning\Class\Practical\Practical\_1\RawData.csv")

print(data.columns)

print(len(data.columns))

print(len(data))

print(data.dtypes)

print(data.isnull().values.any())

print(data.isnull().values.any())

print("\nTotal empty cells by column :\n", data.isnull().sum(), "\n\n")

print("\n\nNumber of Unique Locations : ", len(data['Location'].unique()))

print("\n\nNumber of Unique Salaries : ", len(data['Salary'].unique()))

print(len(data['Salary'].unique()))

print("\n\nUnique Salaries:\n", data['Salary'].unique())

#Cleaning the experience

exp = list(data.Experience)

print(exp)

a=len(exp)

min\_ex = []

max\_ex = []

for i in range(len(exp)):

exp[i] = exp[i].replace("yrs","").strip()

min\_ex.append(int(exp[i].split("-")[0].strip()))

max\_ex.append(int(exp[i].split("-")[1].strip()))

print(min\_ex)

print(max\_ex)

#Attaching the new experiences to the original dataset

data["minimum\_exp"] = min\_ex

data["maximum\_exp"] = max\_ex

data

#Label encoding location and salary

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

data['Location'] = le.fit\_transform(data['Location'])

data['Salary'] = le.fit\_transform(data['Salary'])

print(data['Location'])

print(data['Salary'])

print(data)

Index=data['Index']

Company=data['Company']

Location = data['Location']

Salary = data['Salary']

minimum\_exp = data['minimum\_exp']

maximum\_exp = data['maximum\_exp']

# dictionary of lists

dict = {'Index': Index, 'Company': Company, 'Location': Location, 'Salary':Salary, 'minimum\_exp':minimum\_exp, 'maximum\_exp':maximum\_exp}

df = pd.DataFrame(dict)

# saving the dataframe

df.to\_csv('E:\Latha\LathaSKPIMCS\Machine Learning\Class\Practical\Practical\_1\File4.csv')

#Read New dataset

data = pd.read\_csv("E:\Latha\LathaSKPIMCS\Machine Learning\Class\Practical\Practical\_1\File4.csv")