In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
df = pd.read_csv(r"C:\Users\NITISH SINGH\Downloads\data.xlsx - Sheet1.csv")
df.head()
```

Out[2]:

| | Unnamed: 0 | ID | Salary | DOJ | DOL | Designation | JobCity | Gender | DOB | 10percentage | ... | ComputerScience | MechanicalEngg | ElectricalEngg | TelecomEngg | CivilEngg | c |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | train | 203097 | 420000.0 | 6/1/12 0:00 | present | senior quality engineer | Bangalore | f | 2/19/90 0:00 | 84.3 | ... | -1 | -1 | -1 | -1 | -1 | |
| **1** | train | 579905 | 500000.0 | 9/1/13 0:00 | present | assistant manager | Indore | m | 10/4/89 0:00 | 85.4 | ... | -1 | -1 | -1 | -1 | -1 | |
| **2** | train | 810601 | 325000.0 | 6/1/14 0:00 | present | systems engineer | Chennai | f | 8/3/92 0:00 | 85.0 | ... | -1 | -1 | -1 | -1 | -1 | |
| **3** | train | 267447 | 1100000.0 | 7/1/11 0:00 | present | senior software engineer | Gurgaon | m | 12/5/89 0:00 | 85.6 | ... | -1 | -1 | -1 | -1 | -1 | |
| **4** | train | 343523 | 200000.0 | 3/1/14 0:00 | 3/1/15 0:00 | get | Manesar | m | 2/27/91 0:00 | 78.0 | ... | -1 | -1 | -1 | -1 | -1 | |

5 rows × 39 columns

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Unnamed: 0             3998 non-null   object 
 1   ID                     3998 non-null   int64  
 2   Salary                 3998 non-null   float64
 3   DOJ                    3998 non-null   object 
 4   DOL                    3998 non-null   object 
 5   Designation            3998 non-null   object 
 6   JobCity                3998 non-null   object 
 7   Gender                 3998 non-null   object 
 8   DOB                    3998 non-null   object 
 9   10percentage           3998 non-null   float64
 10  10board                3998 non-null   object 
 11  12graduation           3998 non-null   int64  
 12  12percentage           3998 non-null   float64
 13  12board                3998 non-null   object 
 14  CollegeID              3998 non-null   int64  
 15  CollegeTier            3998 non-null   int64  
 16  Degree                 3998 non-null   object 
 17  Specialization         3998 non-null   object 
 18  collegeGPA             3998 non-null   float64
 19  CollegeCityID          3998 non-null   int64  
 20  CollegeCityTier        3998 non-null   int64  
 21  CollegeState           3998 non-null   object 
 22  GraduationYear         3998 non-null   int64  
 23  English                3998 non-null   int64  
 24  Logical                3998 non-null   int64  
 25  Quant                  3998 non-null   int64  
 26  Domain                 3998 non-null   float64
 27  ComputerProgramming    3998 non-null   int64  
 28  ElectronicsAndSemicon  3998 non-null   int64  
 29  ComputerScience        3998 non-null   int64  
 30  MechanicalEngg         3998 non-null   int64  
 31  ElectricalEngg         3998 non-null   int64  
 32  TelecomEngg            3998 non-null   int64  
 33  CivilEngg              3998 non-null   int64  
 34  conscientiousness      3998 non-null   float64
 35  agreeableness          3998 non-null   float64
 36  extraversion           3998 non-null   float64
 37  nueroticism            3998 non-null   float64
 38  openess_to_experience  3998 non-null   float64
dtypes: float64(10), int64(17), object(12)
memory usage: 1.2+ MB
```

In [4]:

```
df.shape
```

Out[4]:

```
(3998, 39)
```

In [5]:

```
df.columns
```

Out[5]:

```
Index(['Unnamed: 0', 'ID', 'Salary', 'DOJ', 'DOL', 'Designation', 'JobCity',
       'Gender', 'DOB', '10percentage', '10board', '12graduation',
       '12percentage', '12board', 'CollegeID', 'CollegeTier', 'Degree',
       'Specialization', 'collegeGPA', 'CollegeCityID', 'CollegeCityTier',
       'CollegeState', 'GraduationYear', 'English', 'Logical', 'Quant',
       'Domain', 'ComputerProgramming', 'ElectronicsAndSemicon',
       'ComputerScience', 'MechanicalEngg', 'ElectricalEngg', 'TelecomEngg',
       'CivilEngg', 'conscientiousness', 'agreeableness', 'extraversion',
       'nueroticism', 'openess_to_experience'],
      dtype='object')
```

In [6]:

```
df.nunique()
```

Out[6]:

```
Unnamed: 0                  1
ID                       3998
Salary                    177
DOJ                        81
DOL                        67
Designation               419
JobCity                   339
Gender                      2
DOB                      1872
10percentage              851
10board                   275
12graduation               16
12percentage              801
12board                   340
CollegeID                1350
CollegeTier                 2
Degree                      4
Specialization             46
collegeGPA               1282
CollegeCityID            1350
CollegeCityTier             2
CollegeState               26
GraduationYear             11
English                   111
Logical                   107
Quant                     138
Domain                    243
ComputerProgramming        79
ElectronicsAndSemicon      29
ComputerScience            20
MechanicalEngg             42
ElectricalEngg             31
TelecomEngg                26
CivilEngg                  23
conscientiousness         141
agreeableness             149
extraversion              154
nueroticism               217
openess_to_experience     142
dtype: int64
```

In [10]:

```
df = df.drop(columns=['Unnamed: 0', 'ID', 'CollegeID', 'CollegeCityID'])
```

In [11]:

```
df.head()
```

Out[11]:

| | Salary | DOJ | DOL | Designation | JobCity | Gender | DOB | 10percentage | 10board | 12graduation | ... | ComputerScience | MechanicalEngg | ElectricalEngg | TelecomEngg | CivilE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 420000.0 | 6/1/12 0:00 | present | senior quality engineer | Bangalore | f | 2/19/90 0:00 | 84.3 | board ofsecondary education,ap | 2007 | ... | -1 | -1 | -1 | -1 | |
| **1** | 500000.0 | 9/1/13 0:00 | present | assistant manager | Indore | m | 10/4/89 0:00 | 85.4 | cbse | 2007 | ... | -1 | -1 | -1 | -1 | |
| **2** | 325000.0 | 6/1/14 0:00 | present | systems engineer | Chennai | f | 8/3/92 0:00 | 85.0 | cbse | 2010 | ... | -1 | -1 | -1 | -1 | |
| **3** | 1100000.0 | 7/1/11 0:00 | present | senior software engineer | Gurgaon | m | 12/5/89 0:00 | 85.6 | cbse | 2007 | ... | -1 | -1 | -1 | -1 | |
| **4** | 200000.0 | 3/1/14 0:00 | 3/1/15 0:00 | get | Manesar | m | 2/27/91 0:00 | 78.0 | cbse | 2008 | ... | -1 | -1 | -1 | -1 | |

5 rows × 35 columns

In [ ]: