

Amcat-eda-by-mora-vamshikrishna

February 23, 2024

1 EDA on AMCAT Data

1.1 importing libraries

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# We have many column names, By default Pandas show some column Names.
# To see as many columns as possible by using pd.set_option('display.
    ↪max_columns', None)
pd.set_option('display.max_columns', None)

# To avoid Warnings
import warnings
warnings.filterwarnings("ignore")
```

1.2 Loading DATA

```
[2]: # loading the data
df = pd.read_csv('data/data.xlsx - Sheet1.csv')

df.shape
```

```
[2]: (3998, 39)
```

```
[3]: df.columns
```

```
[3]: 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 the data set we have one Garbage Column 'Unnamed: 0', we can't use this column in Analysis.

Drop 'Unnamed: 0' column

```
[4]: data = df.iloc[:, 1:]
```

```
[5]: print(f'Rows: {data.shape[0]} \nColumns: {data.shape[1]}\n')
data.info()
```

```
Rows: 3998
Columns: 38
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to
3997 Data columns (total 38
columns):
 #   Column           Non-Null Count Dtype  
 ---  --  
 0   ID               3998 non-null int64 
 1   Salary            3998 non-null float64
 2   DOJ              3998 non-null object 
 3   DOL              3998 non-null object 
 4   Designation       3998 non-null object 
 5   JobCity           3998 non-null object 
 6   Gender             3998 non-null object 
 7   DOB               3998 non-null object 
 8   10percentage     3998 non-null float64
 9   10board            3998 non-null object 
 10  12graduation      3998 non-null int64 
 11  12percentage     3998 non-null float64
 12  12board            3998 non-null object 
 13  CollegeID          3998 non-null int64 
 14  CollegeTier         3998 non-null int64 
 15  Degree             3998 non-null object 
 16  Specialization      3998 non-null object 
 17  collegeGPA          3998 non-null float64
 18  CollegeCityID        3998 non-null int64 
 19  CollegeCityTier       3998 non-null int64 
 20  CollegeState          3998 non-null object 
 21  GraduationYear        3998 non-null int64 
 22  English              3998 non-null int64 
 23  Logical              3998 non-null int64 
 24  Quant                3998 non-null int64 
 25  Domain                3998 non-null float64
 26  ComputerProgramming    3998 non-null int64 
 27  ElectronicsAndSemicon 3998 non-   int64 
null
```

```

28 ComputerScience      3998 non-      int64
                           null
29 MechanicalEngg       3998 non-      int64
                           null
30 ElectricalEngg       3998 non-      int64
                           null
31 TelecomEngg         3998 non-      int64
                           null
32 CivilEngg            3998 non-      int64
                           null
33 conscientiousness    3998 non-      float64
                           null
34 agreeableness        3998 non-      float64
                           null
35 extraversion          3998 non-      float64
                           null
36 nueroticism          3998 non-      float64
                           null
37 openness_to_experience 3998 non-      float64
                           null
dtypes: float64(10), int64(17), object(11)
memory usage: 1.2+ MB

```

[6]: data.head()

```

[6]:      ID      Salary      DOJ      DOL      Designation \
0 203097  420000.0 6/1/12 0:00  present senior quality engineer
1 579905  500000.0 9/1/13 0:00  present   assistant manager
2 810601  325000.0 6/1/14 0:00  present   systems engineer
3 267447 1100000.0 7/1/11 0:00  present senior software
                                         engineer
4 343523  200000.0 3/1/14 0:00  3/1/15 0:00                  get

      JobCity Gender      DOB 10percentage \
0 Bangalore   f 2/19/90 0:00      84.3
1 Indore     m 10/4/89 0:00      85.4
2 Chennai     f 8/3/92 0:00      85.0
3 Gurgaon     m 12/5/89 0:00      85.6
4 Manesar     m 2/27/91 0:00      78.0

      10board 12graduation 12percentage \
0                               board ofsecondary education,ap      2007  95.8
1                               cbse      2007  85.0
2                               cbse      2010  68.2
3                               cbse      2007  83.6
4                               cbse      2008  76.8

      12board CollegeID CollegeTier      Degree
\ 0 board of intermediate education,ap 1141  2 B.Tech/B.E. 1 cbse

```

```

5807 2 B.Tech/B.E. 2 cbse 64      2 B.Tech/B.E. 3 cbse 6920 1
B.Tech/B.E.
4                               cbse      11368          2 B.Tech/B.E.

                           Specialization collegeGPA CollegeCityID \
0                  computer engineering    78.00        1141
1 electronics and communication engineering 70.06        5807
2                  information technology   70.00         64
3                  computer engineering    74.64        6920
4 electronics and communication engineering 73.90        11368
CollegeCityTier CollegeState GraduationYear English Logical Quant \
0          0 Andhra Pradesh    2011  515  585  525
1          0 Madhya Pradesh    2012  695  610  780
2          0 Uttar Pradesh     2014  615  545  370
3          1 Delhi            2011  635  585  625
4          0 Uttar Pradesh     2012  545  625  465

                           Domain ComputerProgramming ElectronicsAndSemicon ComputerScience \
0 0.635979 445 -1 -1 1 0.960603 -1 466 -1 2 0.450877 395 -1 -1
3 0.974396 615 -1 -1 4 0.124502 -1 233 -1

MechanicalEngg ElectricalEngg TelecomEngg CivilEngg conscientiousness \
0      -1           -1           -1           -1           0.9737
1      -1           -1           -1           -1           -
2      -1           -1           -1           -1           0.7335
3      -1           -1           -1           -1           0.0464
4      -1           -1           -1           -1           -
                           agreeableness extraversion nueroticism
                           openness_to_experience
0      0.8128       0.5269      1.35490     -0.4455 1  0.3789
1      1.2396      -0.10760     0.8637 2     1.7109   0.1637
2      -0.86820    0.6721 3     0.3448     -0.3440   -0.40780
3      -0.9194
4      -0.2793      -1.0697     0.09163          -0.1295

```

[7]: np.round(data.describe())

```

[7]:      ID   Salary 10percentage 12graduation 12percentage \
count  3998.0  3998.0      3998.0      3998.0      3998.0
mean   663795.0 307700.0      78.0       2008.0      74.0
std    363218.0 212737.0      10.0        2.0       11.0
min    11244.0   35000.0      43.0       1995.0      40.0
25%   334284.0 180000.0      72.0       2007.0      66.0
50%   639600.0 300000.0      79.0       2008.0      74.0
75%   990480.0 370000.0      86.0       2009.0      83.0

```

max	1298275.0	4000000.0	98.0	2013.0	99.0
CollegeID CollegeTier collegeGPA CollegeCityID CollegeCityTier \					
count	3998.0	3998.0	3998.0	3998.0	3998.0
mean	5157.0	2.0	71.0	5157.0	0.0
std	4802.0	0.0	8.0	4802.0	0.0
min	2.0	1.0	6.0	2.0	0.0
25%	494.0	2.0	66.0	494.0	0.0
50%	3879.0	2.0	72.0	3879.0	0.0
75%	8818.0	2.0	76.0	8818.0	1.0
max	18409.0	2.0	100.0	18409.0	1.0
GraduationYear English Logical Quant Domain ComputerProgramming \					
count	3998.0	3998.0	3998.0	3998.0	3998.0
mean	2012.0	502.0	502.0	513.0	1.0
std	32.0	105.0	87.0	122.0	0.0
min	0.0	180.0	195.0	120.0	-1.0
25%	2012.0	425.0	445.0	430.0	0.0
50%	2013.0	500.0	505.0	515.0	1.0
75%	2014.0	570.0	565.0	595.0	1.0
max	2017.0	875.0	795.0	900.0	1.0
ElectronicsAndSemicon ComputerScience MechanicalEngg ElectricalEngg \					
count	3998.0	3998.0	3998.0	3998.0	3998.0
mean	95.0	91.0	23.0	16.0	
std	158.0	175.0	98.0	88.0	
min	-1.0	-1.0	-1.0	-1.0	
25%	-1.0	-1.0	-1.0	-1.0	
50%	-1.0	-1.0	-1.0	-1.0	
75%	233.0	-1.0	-1.0	-1.0	
max	612.0	715.0	623.0	676.0	
TelecomEngg CivilEngg conscientiousness agreeableness extraversion \					
count	3998.0	3998.0	3998.0	3998.0	3998.0
mean	32.0	3.0	-0.0	0.0	0.0
std	105.0	37.0	1.0	1.0	1.0
min	-1.0	-1.0	-4.0	-6.0	-5.0
25%	-1.0	-1.0	-1.0	-0.0	-1.0
50%	-1.0	-1.0	0.0	0.0	0.0
75%	-1.0	-1.0	1.0	1.0	1.0

max	548.0	516.0	2.0	2.0	3.0
	nueroticism				
	openess_to_experience				
count	3998.0		3998.0		
mean	-0.0		-0.0		
std	1.0		1.0		
min	-3.0		-7.0		
25%	-1.0		-1.0		
50%	-0.0		-0.0		
75%	1.0		1.0		
max	3.0		2.0		

1.3 Data Cleaning

[8]: # finding irregularities in the data

```
def irregularities(df):
```

```
    for i in df:
```

```
        print('*'*20, i, '*'*20)
```

```
        print(data[i].unique())
```

```
        print()
```

```
irregularities(data)
```

```
***** ID *****
```

```
[203097 579905 810601 ... 355888 947111 324966]
```

```
***** Salary *****
```

```
[ 420000. 500000. 325000. 1100000. 200000. 300000. 400000. 600000.
230000. 450000. 270000. 350000. 250000. 120000. 320000. 190000.
180000. 335000. 435000. 345000. 145000. 220000. 530000.
340000. 360000. 215000. 80000. 330000. 380000. 110000. 205000.
95000. 390000. 60000. 240000. 525000. 305000. 150000. 310000.
455000. 800000. 100000. 280000. 445000. 315000. 370000.
275000. 1500000. 425000. 470000. 460000. 510000. 480000.
170000. 640000. 225000. 440000. 1200000. 675000. 105000.
195000. 385000. 235000. 615000. 290000. 140000. 405000.
1860000. 375000. 430000. 660000. 70000. 410000. 550000. 35000.
115000. 415000. 265000. 285000. 245000.
395000. 560000. 700000. 185000. 160000. 625000. 85000. 135000.
785000. 210000. 155000. 355000. 535000. 690000. 260000. 1110000.
1000000. 505000. 475000. 715000. 820000. 90000. 720000. 2600000.
515000. 55000. 495000. 65000. 655000. 545000. 520000. 645000.
1025000. 775000. 490000. 1300000. 3500000. 910000. 570000. 255000.
```

130000. 175000. 730000. 555000. 465000. 680000. 165000. 630000.
365000. 1050000. 2000000. 860000. 125000. 50000. 580000. 485000.
4000000. 2020000. 650000. 45000. 610000. 760000. 585000. 620000.
870000. 2050000. 540000. 144000. 605000. 1320000. 755000. 880000.
3000000. 75000. 295000. 40000. 575000. 565000. 2500000. 2300000.
590000. 950000. 1800000. 725000. 930000. 750000. 705000. 1745000.
850000. 845000. 670000. 1030000. 770000. 900000. 1210000. 810000.
925000.]

***** DOJ *****
['6/1/12 0:00' '9/1/13 0:00' '6/1/14 0:00' '7/1/11 0:00' '3/1/14 0:00'
'8/1/14 0:00' '7/1/14 0:00' '7/1/13 0:00' '4/1/11 0:00' '8/1/11 0:00'
'12/1/13 0:00' '1/1/14 0:00' '8/1/13 0:00' '9/1/14 0:00' '11/1/10 0:00'
'8/1/12 0:00' '10/1/13 0:00' '9/1/12 0:00' '1/1/11 0:00' '2/1/15 0:00'
'11/1/14 0:00' '12/1/11 0:00' '10/1/14 0:00' '1/1/15 0:00' '3/1/13 0:00'
'10/1/10 0:00' '1/1/13 0:00' '6/1/11 0:00' '4/1/14 0:00' '5/1/12 0:00'
'10/1/12 0:00' '4/1/15 0:00' '3/1/12 0:00' '6/1/13 0:00' '9/1/09 0:00'
'11/1/13 0:00' '7/1/10 0:00' '2/1/14 0:00' '6/1/15 0:00' '5/1/14 0:00' '12/1/14
0:00' '11/1/11 0:00' '7/1/15 0:00' '5/1/13 0:00' '3/1/11 0:00'
'3/1/15 0:00' '7/1/12 0:00' '10/1/11 0:00' '4/1/10 0:00' '4/1/13 0:00'
'12/1/10 0:00' '2/1/13 0:00' '9/1/11 0:00' '2/1/12 0:00' '1/1/12 0:00'
'12/1/12 0:00' '9/1/10 0:00' '4/1/12 0:00' '11/1/12 0:00' '5/1/15 0:00'
'6/1/10 0:00' '2/1/11 0:00' '8/1/10 0:00' '5/1/10 0:00' '5/1/11 0:00'
'8/1/04 0:00' '11/1/08 0:00' '6/1/09 0:00' '2/1/10 0:00' '11/1/09 0:00'
'3/1/10 0:00' '11/1/15 0:00' '1/1/06 0:00' '8/1/15 0:00' '1/1/10 0:00'
'12/1/15 0:00' '9/1/07 0:00' '6/1/91 0:00' '7/1/07 0:00' '6/1/07
0:00' '2/1/07 0:00']

***** DOL *****
['present' '3/1/15 0:00' '5/1/15 0:00' '7/1/15 0:00' '4/1/15 0:00'
'10/1/14 0:00' '9/1/14 0:00' '6/1/14 0:00' '9/1/12 0:00' '12/1/13 0:00'
'6/1/15 0:00' '10/1/13 0:00' '1/1/15 0:00' '4/1/14 0:00' '6/1/13 0:00'
'3/1/12 0:00' '7/1/14 0:00' '2/1/13 0:00' '1/1/14 0:00' '4/1/13 0:00'
'7/1/12 0:00' '5/1/14 0:00' '9/1/13 0:00' '2/1/15 0:00' '1/1/12 0:00'
'8/1/15 0:00' '8/1/14 0:00' '12/1/15 0:00' '12/1/14 0:00' '5/1/12 0:00'
'3/1/11 0:00' '7/1/11 0:00' '2/1/14 0:00' '12/1/11 0:00' '10/1/15 0:00'
'11/1/14 0:00' '3/1/14 0:00' '11/1/11 0:00' '5/1/13 0:00' '7/1/13 0:00'
'11/1/13 0:00' '1/1/11 0:00' '5/1/11 0:00' '2/1/12 0:00' '11/1/12 0:00'
'6/1/12 0:00' '8/1/13 0:00' '3/1/05 0:00' '3/1/13 0:00' '10/1/12 0:00'
'2/1/11 0:00' '2/1/10 0:00' '1/1/13 0:00' '6/1/11 0:00' '9/1/15 0:00'
'4/1/12 0:00' '8/1/12 0:00' '4/1/11 0:00' '10/1/11 0:00' '11/1/15 0:00'
'12/1/12 0:00' '9/1/11 0:00' '8/1/10 0:00' '8/1/11 0:00' '6/1/09 0:00'
'3/1/08 0:00' '10/1/10 0:00']

***** Designation *****

['senior quality engineer' 'assistant manager' 'systems engineer'

'senior software engineer' 'get' 'system engineer'
'java software engineer' 'mechanical engineer' 'electrical engineer'
'project engineer' 'senior php developer' 'senior systems engineer'
'quality assurance engineer' 'qa analyst' 'network engineer'
'product development engineer' 'associate software developer'
'data entry operator' 'software engineer' 'developer'
'electrical project engineer' 'programmer analyst' 'systems analyst'
'ase' 'telecommunication engineer' 'application developer'
'ios developer' 'executive assistant' 'online marketing manager'
'documentation specialist' 'associate software engineer'
'management trainee' 'site manager' 'software developer' '.net developer'
'production engineer' 'jr. software engineer'
'trainee software developer' 'ui developer' 'assistant system engineer'
'android developer' 'customer service' 'test engineer' 'java developer'
'engineer' 'recruitment coordinator' 'technical support engineer'
'data analyst' 'assistant software engineer' 'faculty'
'entry level management trainee' 'customer service representative'
'software test engineer' 'firmware engineer' 'php developer'
'research associate' 'research analyst' 'quality engineer' 'programmer'
'technical support executive' 'business analyst' 'web developer' 'application
engineer' 'project coordinator' 'engineer trainee'
'sap consultant' 'quality analyst' 'marketing coordinator'
'system administrator' 'senior engineer' 'business development managerde'
'network administrator' 'technical support specialist'
'business development executive' 'junior software engineer'
'asp.net developer' 'graduate engineer trainee' 'field engineer'
'assistant professor' 'trainee software engineer'
'senior software developer' 'quality assurance automation engineer'
'design engineer' 'telecom engineer' 'quality control engineer'
'hardware engineer' 'hr recruiter' 'sales associate' 'junior engineer'
'associate engineer' 'maintenance engineer' 'sales engineer'
'human resources associate' 'mobile application developer'
'electronic field service engineer' 'process associate'
'field service engineer' 'it support specialist'
'software development engineer' 'business process analyst'
'operation engineer' 'electrical designer' 'marketing assistant'
'sales executive' 'admin assistant' 'senior java developer'
'account executive' 'oracle dba' 'rf engineer'
'embedded software engineer' 'programmer analyst trainee'
'technical engineer' 'operations executive' 'trainee engineer'
'recruiter' 'lecturer' '.net web developer' 'marketing executive'
'operations assistant' 'associate manager' 'electrical design engineer'
'systems administrator' 'client services associate' 'it analyst'
'senior developer' 'cad designer' 'business technology analyst'
'asst. manager' 'service engineer' 'executive recruiter'

'planning engineer' 'associate technical operations' 'web designer'
'software architect' 'software quality assurance tester' 'seo trainee'
'process engineer' 'software quality assurance analyst' 'designer'
'business systems consultant' 'business development manager'
'junior research fellow' 'technical recruiter' 'operations analyst'
'quality assurance test engineer' 'linux systems administrator'
'software trainee' 'entry level sales and marketing'
'electrical field engineer' 'windows systems administrator'
'junior software developer' 'python developer'
'web application developer' 'assistant systems engineer'
'javascript developer' 'operation executive' 'performance engineer'
'technical writer' 'operations engineer and jetty handling'
'lead engineer' 'portfolio analyst' 'associate system engineer'
'mechanical design engineer' 'product engineer'
'network security engineer' 'operations manager' 'technical lead'
'operations' 'quality assurance tester' 'automation engineer'
'data scientist' 'quality associate' 'manual tester' 'sr. engineer'
'embedded engineer' 'service and sales engineer'
'telecom support engineer' 'engineer- customer support' 'cloud engineer'
'branch manager' 'business analyst consultant' 'technology lead'
'software trainee engineer' 'dcs engineer' 'junior manager' 'ux designer'
'clerical' 'hr generalist' 'database administrator'
'senior design engineer' 'seo' 'assistant engineer' 'marketing analyst' 'it
executive' 'salesforce developer' 'software tester' 'sql dba'
'junior engineer product support' 'manager' 'senior business analyst'
'c# developer' 'implementation engineer' 'executive hr'
'executive engineer' 'sharepoint developer' 'system analyst'
'sales management trainee' 'senior project engineer' 'it recruiter'
'software engineer analyst' 'desktop support technician'
'continuous improvement engineer' 'process advisor' 'etl developer'
'sales and service engineer' 'project manager' 'training specialist'
'product manager' 'staffing recruiter' 'assistant programmer'
'quality controller' 'mis executive' 'game developer'
'digital marketing specialist' 'principal software engineer'
'software developer' 'senior mechanical engineer'
'technical operations analyst' 'service coordinator' 'testing engineer'
'technical assistant' 'sap abap consultant' 'seo engineer'
'project assistant' 'talent acquisition specialist'
'sales account manager' 'software engineer trainee'
'customer service manager' 'help desk analyst' 'general manager'
'engineering manager' 'senior network engineer'
'field based employee relations manager' 'phone banking officer'
'support engineer' 'associate test engineer' 'technology analyst'
'network support engineer' 'it business analyst' 'junior system analyst'
'senior .net developer' 'secretary' 'research engineer'

'quality assurance auditor' 'process executive'
'lecturer & electrical maintenance' 'office coordinator' 'hr manager'
'html developer' 'sales support' 'front end web developer'
'administrative support' 'territory sales manager'
'project administrator' 'environmental engineer' 'web designer and seo'
'information security analyst' 'field business development associate'
'operational executive' 'administrative coordinator'
'senior risk consultant' 'desktop support engineer' 'cad drafter'
'noc engineer' 'industrial engineer' 'it engineer'
'human resources intern' 'senior quality assurance engineer'
'clerical assistant' 'software enginner' 'quality assurance'
'delivery software engineer' 'graphic designer'
'sales development manager' 'visiting faculty'
'business intelligence analyst' 'team lead'
'operational excellence manager' 'sales & service engineer' 'web intern'
'full stack developer' 'database developer' 'sr. database engineer'
'graduate apprentice trainee' 'software engineer associate'
'technical analyst' 'executive engg' 'it technician'
'business system analyst' 'process control engineer'
'technical consultant' 'business office manager'
'quality control inspector' 'product design engineer'
'manufacturing engineer' 'seo executive' 'sap analyst'
'software engineere' 'financial service consultant' 'co faculty'
'software analyst' 'desktop support analyst' 'graduate engineer'
'engineering technician' 'it assistant' 'marketing manager'
'human resource assistant' 'hr assistant' 'product developer'
'customer support engineer' 'quality control inspection technician'
'gis/cad engineer' 'senior web developer' 'sql developer'
'research staff member' 'sap abap associate consultant' 'associate qa'
'corporate recruiter' 'project management officer'
'business systems analyst' 'software programmer' 'help desk technician'
'sales manager' 'catalog associate' 'assistant store manager'
'software engg' 'it developer' 'apprentice' 'business consultant'
'controls engineer' 'ruby on rails developer' 'risk consultant'
'account manager' 'professor' 'assistant administrator' 'civil engineer'
'educator' 'service manager' 'teradata dba'
'full-time loss prevention associate' 'junior recruiter'
'associate developer' 'assistant electrical engineer' 'shift engineer'
'dotnet developer' 'rf/dt engineer' 'human resources analyst'
'software test engineerte' 'junior .net developer' 'java trainee'
'maintenance supervisor' 'r&d engineer' 'front end developer'
'engineer-hws' 'operations engineer' 'senior research fellow'
'web designer and joomla administrator' 'enterprise solutions developer'
'information technology specialist' 'site engineer'
'graduate trainee engineer' 'quality assurance analyst' 'cnc programmer'

'financial analyst' 'system engineer trainee' 'sap mm consultant'
'assistant system engineer trainee' 'qa trainee' 'teradata developer'
'hr executive' 'senior programmer' 'software test engineer (etl)'
'associate software engg' 'supply chain analyst' 'sales trainer'
'software executive' 'team leader' 'assistant system engineer - trainee'
'seo analyst' 'risk investigator' 'executive administrative assistant'
'program manager' 'r & d' 'sap functional consultant'
'website developer/tester' 'software designer' 'sales coordinator'
'qa engineer' 'aircraft technician' 'customer care executive'
'senior test engineer' 'program analyst trainee'
'electrical controls engineer' 'trainee decision scientist' 'editor'
'bss engineer' 'dba' 'software eng' 'computer faculty'
'recruitment associate' 'logistics executive' 'quality consultant'
'senior sales executive' 'db2 dba' 'test technician'
'it operations associate' 'software engineering associate'
'research scientist' 'jr. software developer']

***** JobCity *****

['Bangalore' 'Indore' 'Chennai' 'Gurgaon' 'Manesar' 'Hyderabad' 'Banglore'
'Noida' 'Kolkata' 'Pune' '-1' 'mohali' 'Jhansi' 'Delhi' 'Hyderabad '
'Bangalore ' 'noida' 'delhi' 'Bhubaneswar' 'Navi Mumbai' 'Mumbai'
'New Delhi' 'Mangalore' 'Rewari' 'Gaziabaad' 'Bhiwadi' 'Mysore' 'Rajkot'
'Greater Noida' 'Jaipur' 'noida ' 'HYDERABAD' 'mysore' 'THANE'
'Maharajganj' 'Thiruvananthapuram' 'Punchkula' 'Bhubaneshwar' 'Pune '
'coimbatore' 'Dhanbad' 'Lucknow' 'Trivandrum' 'kolkata' 'mumbai'
'Gandhi Nagar' 'Una' 'Daman and Diu' 'chennai' 'GURGOAN' 'vsakhapttnam'
'pune' 'Nagpur' 'Bhagalpur' 'new delhi - jaisalmer' 'Coimbatore'
'Ahmedabad' 'Kochi/Cochin' 'Bankura' 'Bengaluru' 'Mysore ' 'Kanpur '
'jaipur' 'Gurgaon ' 'bangalore' 'CHENNAI' 'Vijayawada' 'Kochi' 'Beawar' 'Alwar'
'NOIDA' 'Greater noida' 'Siliguri ' 'raipur' 'gurgaon' 'Bhopal'
'Faridabad' 'Jodhpur' 'udaipur' 'Muzaffarpur' 'Kolkata' 'Bulandshahar'
'Haridwar' 'Raigarh' 'Visakhapatnam' 'Jabalpur' 'hyderabad' 'Unnao'
'KOLKATA' 'Thane' 'Aurangabad' 'Belgaum' 'gurgoan' 'Dehradun' 'Rudrapur'
'Jamshedpur' 'vizag' 'Nouda' 'Dharamshala' 'Banagalore' 'Hissar' 'Ranchi'
'BANGALORE' 'Madurai' 'Gurga' 'Chandigarh' 'Australia' ' Chennai'
'CHEYYAR' 'Mumbai ' 'sonepat' 'Ghaziabad' 'Panthagar' 'Siliguri'
'mumbai ' 'Jagdalpur' 'Chennai ' 'angul' 'Baroda' ' ariyalur' 'Jowai'
'Kochi/Cochin, Chennai and Coimbatore' 'bhubaneswar' 'Neemrana' 'VIZAG'
'Tirupathi' 'Lucknow ' 'Ahmedabad ' 'Bhubneshwar' 'Noida ' 'pune '
'Calicut' 'Gandhinagar' 'LUCKNOW' 'Dubai' 'bengaluru' 'MUMBAI'
'Ahmednagar' 'Nashik' 'New delhi' 'Bellary' 'Ludhiana' 'New Delhi '
'Muzaffarnagar' 'BHOPAL' 'Gurgoan' 'Gagret' 'Indirapuram, Ghaziabad'
'Gwalior' 'new delhi' 'TRIVANDRUM' 'Chennai & Mumbai' 'Rajasthan'
'Sonipat' 'Bareli' 'Kanpur' 'Hospete' 'Miryalaguda' ' mumbai' 'Dharuhera'
'lucknow' 'meerut' 'dehradun' 'Ganjam' 'Hubli' 'bangalore ' 'NAVI MUMBAI'

'ncr' 'Agra' 'Trichy' 'kudankulam ,tarapur' 'Ongole' 'Sambalpur'
'Pondicherry' 'Bundi' 'SADULPUR,RAJGARH,DISTT-CHURU,RAJASTHAN' 'AM'
'Bikaner' 'Vadodara' 'BAngalore' 'india' 'Asansol' 'Tirunelvelli'
'Ernakulam' 'DELHI' 'Bilaspur' 'Chandrapur' 'Nanded' 'Dharmapuri'
'Vandavasi' 'Rohtak' 'trivandrum' 'Nagpur ' 'Udaipur' 'Patna' 'banglore'
'indore' 'Salem' 'Nasikcity' 'Gandhinagar ' 'Technopark, Trivandrum'
'Bharuch' 'Tornagallu' 'Raipur' 'Kolkata ' 'Jaspur' 'Burdwan'
'Bhubaneswar ' 'Shimla' 'ahmedabad' 'Gajiabaad' 'Jammu' 'Shahdol'
'Muvattupuzha' 'Al Jubail,Saudi Arabia' 'Kalmar, Sweden' 'Secunderabad'
'A-64, sec-64, noida' 'Ratnagiri' 'Jhajjar' 'Gulbarga'
'hyderabad(bhadurpally)' 'Nalagarh' 'Chandigarh ' 'Jaipur '
'Jeddah Saudi Arabia' ' Delhi' 'PATNA' 'SHAHDOL' 'Chennai, Bangalore'
'Bhopal ' 'Jamnagar' 'PUNE' 'Tirupati' 'Gonda' 'jamnagar' 'chennai '
'orissa' 'kharagpur' 'Trivandrum ' 'Navi Mumbai , Hyderabad' 'Joshimath'
'chandigarh' 'Bathinda' 'Johannesburg' 'kala amb ' 'Karnal' 'LONDON'
'Kota' 'Panchkula' 'Baddi HP' 'Nagari' 'Mettur, Tamil Nadu ' 'Durgapur'
'pondi' 'Surat' 'Kurnool' 'kolhapur' 'Madurai ' 'GREATER NOIDA' 'Bhilai'
' Pune' 'hderabad' 'KOTA' 'thane' 'Vizag' 'Bahadurgarh'
'Rayagada, Odisha' 'kakinada' 'GURGAON' 'Varanasi' 'punr' 'Nellore'
'patna' 'Meerut' 'hyderabad ' 'Sahibabad' 'Howrah' 'BHUBANESWAR'
'Trichur' 'Ambala' 'Khopoli' 'keral' 'Roorkee' 'Greater NOIDA'
'Navi mumbai' 'ghaziabad' 'Allahabad' 'Delhi/NCR' 'Panchkula ' 'Ranchi '
'Jalandhar' 'manesar' 'vapi' 'PILANI' 'muzzafarpur' 'RAS AL KHAIMAH'
'bihar' 'singaruli' 'KANPUR' 'Banglore ' 'pondy' 'Mohali' 'Phagwara'
' Mumbai' ' bangalore' 'GURAGAON' 'Baripada' 'MEERUT' 'Yamuna Nagar'
'shahibabad' 'sampla' 'Guwahati' 'Rourkela' 'Banaglore' 'Vellore' 'Dausa'
'latur (Maharashtra)' 'NEW DELHI' 'kanpur' 'Mainpuri' 'karnal' 'Dammam'
'Haldia' 'sambalpur' 'RAE BARELI' 'ranchi' 'jaipur' 'BANGLORE' 'Patiala'
'Gorakhpur' 'new dehli' 'BANGALORE ' 'Ambala City' 'Karad' 'Rajpura'
'Pilani' 'haryana' 'Asifabadbanglore']

***** Gender *****

['f' 'm']

***** DOB *****

['2/19/90 0:00' '10/4/89 0:00' '8/3/92 0:00' ... '2/28/86 0:00'
'6/22/90 0:00' '4/15/87 0:00']

***** 10percentage *****

[84.3 85.4 85. 85.6 78. 89.92 86.08 92. 90. 77. 88.6 81.
83.6 60.8 81.2 69.54 85.8 65. 79. 91.2 75.67 92.5 70. 88.
86.8 90.88 76.8 84. 64. 77.2 87. 73. 71. 91.5 52.33 66.6
75. 91.4 59. 84.5 95.8 69.6 86.6 81.92 66.5 80. 85.2 89.25
58.4 90.8 89.88 63.3 72. 83. 85.3 60.66 89.37 84.4 89. 86.
70.6 93.44 76. 86.4 84.83 72.3 69. 73.33 86.16 56. 79.25 88.66

80.8	81.16	83.5	55.6	84.6	62.	68.4	51.	79.6	63.4	86.53	76.18
91.43	76.17	82.4	58.	90.04	60.4	74.23	59.8	82.3	84.2	60.	82.8
82.6	74.89	8.8	88.67	64.66	77.8	89.5	64.56	91.12	73.08	78.33	81.33
89.7	87.4	92.66	76.87	91.3	77.9	76.2	77.78	65.6	65.8	67.75	91.6
78.66	78.4	61.	90.4	58.6	82.2	82.	90.1	86.17	83.89	76.7	88.2
80.6	91.	74.4	79.28	84.96	92.8	79.4	66.8	79.8	65.3	94.6	83.33
80.83	67.23	86.2	55.86	62	69.13	89.12	87.5	66.4	80.33	75.2	50.6
81.1	60.14	78.83	75.8	77.66	64.48	67.	75.6	85.28	71.5	93.6	93.33
78.8	82.16	77.65	56.5	79.83	57.	89.42	72.8	86.3	77.5	71.2	80.2
73.6	68.	74.7	69.2	65.33	96.48	82.5	91.8	93.4	68.5	73.4	72.2
71.8	66.33	71.67	70.2	90.27	67.67	53.4	76.16	65.71	79.5	87.2	80.4
88.4	90.6	89.2	81.6	86.5	77.16	72.33	79.89	75.4	72.83	88.33	78.88
95.2	89.33	75.3	69.8	52.	58.2	90.5	70.83	62.13	74.5	63.	73.37
83.4	92.6	78.6	76.4	86.88	66.	70.67	65.1	68.14	92.2	93.5	82.83
81.66	90.15	44.16	94.4	64.4	85.17	70.1	88.25	94.8	77.4	66.66	81.03
44.	45.6	87.8	72.6	79.86	84.67	48.	53.3	71.66	68.8	78.15	83.43
86.9	84.8	75.06	86.66	70.9	81.12	67.5	78.2	71.06	71.28	62.1	90.56
87.6	71.3	88.8	86.09	67.72	76.6	68.6	63.43	70.4	67.6	73.8	55.5
74.67	92.16	83.66	62.86	49.87	11	77.53	88.5	61.9	79.2	83.8	79.33
67.3	83.2	72.5	77.67	94.2	59.33	87.63	89.76	84.14	88.17	59.6	64.3
75.04	82.9	73.5	73.67	77.7	87.69	83.16	71.32	61.6	61.8	85.06	91.71
75.46	74.8	67.4	90.3	86.15	64.7	69.7	82.33	58.83	75.36	76.5	66.67
67.2	80.88	88.88	71.4	65.2	96.	61.2	81.4	68.66	65.56	86.7	63.2
62.2	68.2	77.85	83.3	56.78	83.04	71.1	76.53	74.83	59.3	73.2	86.33
64.83	72.1	61.1	86.83	70.8	94.	80.3	64.6	84.1	84.93	92.83	89.6
80.5	83.82	77.57	51.2	66.46	82.67	61.4	69.4	90.24	75.75	90.83	89.1
87.33	83.1	88.34	91.67	87.3	48.8	75.38	55.52	71.05	77.63	92.47	93.
68.33	79.16	85.33	80.42	78.25	95.52	87.86	89.4	94.7	97.12	93.94	
59.57	80.53	81.9	63.6	66.16	62.5	69.5	80.93	82.7	79.29	81.5	
62.34	83.68	70.66	60.1	77.38	68.83	94.43	85.53	88.09	74.2	85.73	
72.4	67.7	79.78									
81.3	79.37	85.5	87.83	70.33	80.64	58.7	60.2	77.81	85.67	89.9	79.75
75.73	87.68	60.5	81.38	75.52	48.5	88.3	82.1	85.18	81.8	90.46	70.5
79.52	89.29	61.75	78.67	71.6	80.13	81.67	95.81	14	72.16	87.04	88.64
82.56	90.01	63.66	65.17	91.84	92.1	43.65	23	92.48	82.88	73.07	58.56
85.83	67.34	85.75	80.7	79.23	70.26	52.7	75.86	90.16	90.2	78.5	58.9
80.32	67.16	73.06	92.93	85.76	87.52	88.36	81.73	60.7	87.7	79.85	68.9
73.83	64.8	77.84	74.14	64.28	92.4	73.94	63.33	70.06	88.7	89.3	83.83
91.33	80.07	72.17	63.5	69.17	67.42	84.16	76.64	78.13	61.69	81.7	
82.46	64.57	78.1	66.3	59.71	80.67	77.88	88.53	93.38	52.93	78.17	
72.36	84.75										
86.67	77.6	74.3	62.4	65.16	59.2	69.85	79.68	65.4	94.72	71.86	81.25
87.62	54.	85.92	74.33	82.28	75.1	69.73	92.12	70.3	76.33	80.1	76.67
77.83	57.67	83.14	50.91	21	81.83	78.16	80.14	93.3	61.63	73.73	80.46
76.48	82.25	56.16	57.88	87.07	67.1	87.73	77.12	64.53	86.46	82.62	53.06

74.62 76.66 45.33 69.69 53.75.33 74.28 74.6 74.88 74.53 46.24 80.15
 86.13 85.72 51.36 78.53 53.8 86.1 84.43 76.36 77.86 88.83 88.18 79.14
 54.5 84.33 78.3 77.44 87.31 58.16 61.3 77.33 75.12 72.12 65.26
 83.56 50.5 82.27 67.36 87.16 75.83 78.44 91.9 68.3 92.64 58.17
 83.46 88.04
 70.58 71.17 84.13 64.62 66.15 67.8 57.78 56.2 61.5 94.1 74.16 78.93
 70.56 85.16 82.26 71.13 65.66 71.71 86.37 88.57 59.05 79.66 89.39 95.54
 55.3 70.14 87.23 85.38 86.92 95.04 95.6 60.83 71.33 94.16 78.57 80.16
 70.25 82.13 88.76 51.6 70.76 57.5 68.67 74.18 68.25 71.04 64.43 82.24
 90.06 67.12 75.85 87.81 76.93 65.5 92.3 50.66 61.83 63.16 64.2 53.6
 83.69 80.04 78.86 70.61 62.8 67.06 65.85 94.24 63.8 75.77 94.5 71.83
 91.1 91.52 90.08 93.16 86.34 88.1 97.62.93 76.96 85.46 78.08 66.7
 95.86 92.09 81.46 81.86 61.73 77.22 46.8 79.3 72.75 93.8 93.67 75.16
 72.45 85.62 86.85 93.28 58.33 64.33 75.62 54.8 58.5 91.68 69.3 62.6
 65.41 91.62 88.16 92.32 69.83 83.67 69.92 52.5 56.4 54.6 78.61 69.53
 75.71 71.84 70.16 69.66 95.4 84.62 91.53 85.66 61.57 80.63 69.33 62.3
 87.87 70.75 87.58 58.8 62.88 73.66 97.76 88.44 54.83 67.25 90.76 82.75
 75.66 91.04 90.58 91.86 73.1 73.3 69.1 51.83 62.26 65.67 68.65 51.42
 78.28 80.58 55.33 91.17 74.9 93.43 90.81 89.89 62.67 90.26 62.15 70.08
 87.88 72.67 93.2 60.46 71.07 46.54.4 86.12 72.15 71.85 49.9 83.75
 71.73 90.33 86.14 66.2 88.75 67.33 57.14 67.76 82.66 77.92 79.38 67.17
 89.17 79.67 96.8 71.37 82.87 89.44 71.76 57.7 89.23 79.06 83.25 61.86
 89.56 82.14 70.27 59.7 93.07 79.9 64.5 66.85 69.16 52.09 78.72]

***** 10board *****

['board ofsecondary education,ap' 'cbse' 'state board' 'mp board bhopal'
 'icse' 'karnataka secondary school of examination' 'up'
 'karnataka state education examination board' 'ssc'
 'kerala state technical education' '0' 'bseb'
 'state board of secondary education, andhra pradesh' 'matriculation'
 'gujarat state board' 'karnataka state board' 'wbbse'
 'maharashtra state board' 'icse board' 'up board'
 'board of secondary education(bse) orissa'
 'little jacky matric higher secondary school' 'uttar pradesh board'
 'bsc,orissa' 'mp board' 'upboard' 'matriculation board' 'j & k bord'
 'rbse' 'central board of secondary education' 'pseb' 'jkbose'
 'haryana board of school education, (hbse)' 'metric' 'ms board' 'kseeb'
 'stateboard' 'maticulation' 'karnataka secondary education board'
 'mumbai board' 'sslc' 'kseb' 'board secondary education' 'matric board'
 'board of secondary education' 'west bengal board of secondary education'
 'jharkhand secondary examination board,ranchi' 'u p board' 'bseb,patna' 'hsc' 'bse'
 'sss pune' 'karnataka education board (keeb)' 'kerala'
 'state board of secondary education(ssc)' 'gsheb' 'up(allahabad)'
 'nagpur' 'don bosco maatriculation school'
 'karnataka state secondary education board' 'maharashtra'

'karnataka secondary education board'
'himachal pradesh board of school education'
'certificate of middle years program of ib'
'karnataka board of secondary education'
'board of secondary education rajasthan' 'uttarakhand board' 'ua'
'board of secendary education orissa'
'karantaka secondary education and examination borad' 'hbse'
'kseeb(karnataka secondary education examination board)'
'cbse[gulf zone]' 'hbse' 'state(karnataka board)'
'jharkhand accademic council'
'jharkhand secondary examination board (ranchi)'
'karnataka secondary education examination board' 'delhi board'
'mirza ahmed ali baig' 'jseb' 'bse, odisha' 'bihar board'
'maharashtra state(latur board)' 'rajasthan board' 'mpboard' 'upbhsie'
'secondary board of rajasthan' 'tamilnadu matriculation board'
'jharkhand secondary board'
'board of secondary education, andhara pradesh' 'up baord' 'state'
'board of intermediate education'
'state board of secondary education, andhra pradesh'
'up board , allahabad' 'stjosephs girls higher sec school,dindigul'
'maharashtra board' 'education board of kerala' 'board of ssc'
'maharashtra state board pune' 'board of school education harayana'
'secondary school cerfificate' 'maharashtra sate board' 'ksseb'
'bihar examination board, patna' 'latur'
'board of secondary education, rajasthan' 'state borad hp' 'cluny'
'bsepatna' 'up borad' 'ssc board of andrapradesh' 'matric' 'bse,orissa'
'ssc-andhra pradesh' 'mp' 'karnataka education board' 'mhsbse'
'karnataka sslc board bangalore' 'karnataka' 'u p'
'secondary school of education' 'state board of karnataka'
'karnataka secondary board' 'andhra pradesh board ssc'
'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district'
'hse,orissa' 'national public school' 'nagpur board'
'jharkhand academic council' 'bsemp'
'board of secondary education, andhra pradesh'
'board of secondary education orissa'
'board of secondary education,rajasthan(rbse)'
'board of secondary education,ap'
'board of secondary education, andhra pradesh'
'jawahar navodaya vidyalaya' 'aisse'
'karnataka board of higher education' 'bihar' 'kerala state board'
'cicse' 'tn state board' 'kolhapur divisional board, maharashtra'
'bharathi matriculation school' 'uttaranchal state board' 'wbbse'
'mp state board' 'seba(assam)' 'anglo indian' 'gseb' 'uttar pradesh'
'ghseb' 'board of school education uttarakhand' 'msbshse,pune' 'tamilnadu state
board' 'kerala university'

'uttaranchal shiksha avam pariksha parishad'
'bse(board of secondary education)' 'bright way college, (up board)'
'school secondary education, andhra pradesh'
'secondary state certificate'
'maharashtra state board of secondary and higher secondary education,pune'
'andhra pradesh state board' 'stmary higher secondary' 'cgbse'
'secondary school certificate' 'rajasthan board ajmer' 'mpbse'
'pune board' 'cbse' 'board of secondary education,orissa'
'maharashtra state board,pune' 'up bord'
'kiran english medium high school' 'state board (jac, ranchi)'
'gujarat board' 'state board ' 'sarada high scchool'
'kalaimagal matriculation higher secondary school' 'karnataka board'
'maharastra board' 'sslc board' 'ssc maharashtra board'
'tamil nadu state' 'uttrakhand board'
'bihar secondary education board,patna'
'haryana board of school education'
'sri kannika parameswari highier secondary school, udumalpet'
'ksseb(karnataka state board)' 'nashik board'
'jharkhand secondary education board' 'himachal pradesh board'
'maharashtra satate board'
'maharashtra state board mumbai divisional board'
'dav public school,hehal' 'state board of secondary education, ap'
'rajasthan board of secondary education' 'hsce'
'karnataka secondary education' 'board of secondary education,odisha'
'maharashtra nasik board'
'west bengal board of secondary examination (wbbse)'
'holy cross matriculation hr sec school' 'cbsc' 'apssc' 'bseb patna'
'kolhapur' 'bseb, patna' 'up board allahabad' 'biharboard'
'nagpur board,nagpur' 'pune' 'gyan bharati school' 'rbse,ajmer'
'board of secondaray education' 'secondary school education' 'state bord'
'jbse,jharkhand' 'hse' 'madhya pradesh board'
'bihar school examination board'
'west bengal board of secondary eucation' 'state boardmp board '
'icse board , new delhi' 'board of secondary education (bse) orissa'
'maharashtra state board for ssc' 'board of secondary school education'
'latur board' "stmary's convent inter college" 'nagpur divisional board'
'ap state board' 'cgbse raipur' 'uttranchal board' 'ksbe'
'central board of secondary education, new delhi'
'bihar school examination board patna' 'cbse board' 'sslc,karnataka'
'mp-bse' 'up bound' 'dav public school sec 14'
'board of school education haryana'
'council for indian school certificate examination' 'aurangabad board'
'j&k state board of school education'
'maharashtra state board of secondary and higher secondary education'
'maharashtra state boar of secondary and higher secondary education'

'ssc regular' 'karnataka state examination board' 'nasik'
'west bengal board of secondary education' 'up board, allahabad'
'bseb , patna'
'state board - west bengal board of secondary education : wbbse'
'maharashtra state board of secondary & higher secondary education'
'delhi public school' 'karnataka secondary eduction'
'secondary education board of rajasthan' 'maharashtra board, pune'
'rbse (state board)' 'apsche' 'board of secondary education'
'board of high school and intermediate education uttarpradesh' 'kea'
'board of secondary education - andhra pradesh'
'ap state board for secondary education' 'seba'
'punjab school education board, mohali' 'jharkhand acedemic council'
'hse,board' 'board of ssc education andhra pradesh' 'up-board'
'bse,odisha']

***** 12graduation *****
[2007 2010 2008 2009 2006 2011 2005 1995 2004 2012 2003 2002 2001
1998 2013 1999]

***** 12percentage *****
[95.8 85. 68.2 83.6 76.8 87. 67.5 91. 91.2 72.2 83.7 86.
69.83 62.4 79.9 64.43 74.8 66.66 64.8 62.2 84.63 74.4 95.2 43.42
90. 82.8 82.5 83. 68. 74. 92. 86.1 84.4 68.4 61. 93.8
85.4 67. 89.66 68.6 60. 73. 87.7 87.16 82. 71.4 75. 61.46
49.5 72.5 81.5 78. 90.1 70.1 95.4 61.1 49. 79. 88.3 77.8 94.4
86.67 73.2 64. 77. 78.67 72. 78.2 95. 82.4 60.2 62.6
70. 71.33 70.08 56. 80. 84.33 87.9 65. 68.5 94.2 66. 88.
76.6 50.8 66.5 76.4 65.8 48. 63. 71.55 86.33 71.3 57.6 83.4 75.16
77.77 60.25 89.6 86.4 80.4 59. 73.6 63.6 66.6 86.8 79.6
87.58 81.4 89. 62. 47. 77.2 71.2 54. 67.6 91.9 63.1 69.
68.46 85.2 78.4 82.2 95.6 81.33 88.9 82.75 72.6 66.67 70.2 61.5
70.6 79.4 61.8 95.5 80.7 60.4 77.14 75.2 81.2 80.8 88.88 83.9
65.2 83.1 80.6 70.16 90.91 84.7 68.55 59.16 78.83 59.9 75.6 87.8
79.2 80.3 82.1 65.5 84.3 64.4 91.6 95.3 69.8 86.9 73.4 56.9
86.7 64.7 80.5 61.6 90.7 81. 57. 92.6 78.6 71. 71.5 70.4 89.5
76.66 80.1 54.4 80.9 84.8 69.45 93.4 56.8 91.5 90.67 64.5
96.1 62.8 94.1 89.7 76. 73.8 78.3 96.6 60.17 75.4 72.4 52.
77.4 69.6 56.2 78.43 82.7 74.45 76.2 68.8 78.8 50. 67.4 95.41
84. 71.08 94.5 67.75 87.1 81.8 80.2 83.58 62.83 69.4 86.5 67.67
74.2 66.4 55. 61.2 64.6 43.12 61.57 84.5 83.8 89.1 74.6 81.1 80.83
67.8 76.5 87.83 69.9 88.4 58.55 83.33 82.6 69.2 63.2 88.7
62.66 61.7 96.7 62.5 79.57 93. 78.16 74.92 94.16 51.4 97.1 94.9
87.4 63.4 87.66 65.4 68.67 58. 68.66 67.7 91.3 79.8 81.6 67.25
55.2 65.6 64.3 51.63 93.7 69.04 50.2 96.5 77.7 90.2 72.67 73.83
71.9 82.83 82.08 77.38 71.8 66.8 71.25 53.76.62 72.77 79.83 77.3

81.17	81.12	62.16	67.68	57.8	85.3	83.67	75.69	75.9	55.02	77.5	88.83
75.5	92.3	67.2	62.3	72.48	59.2	72.8	60.13	94.3	89.8	81.3	82.02
76.89	60.16	77.6	69.33	87.5	74.5	46.	74.88	68.33	84.67	87.2	66.3
58.6	85.36	68.7	55.54	68.89	82.3	58.2	48.34	54.2	92.7	71.01	53.8
56.6	66.77	71.6	90.6	82.66	89.08	93.6	72.3	69.7	84.17	92.1	90.4
92.67	64.33	62.81	67.17	95.1	68.83	97.8	88.1	48.8	60.05	69.07	89.3
86.3	94.6	86.25	79.23	97.	84.75	89.2	77.54	60.8	84.6	86.2	65.16
94.33	74.67	94.7	92.2	90.33	73.61	53.6	58.4	78.86	76.44	78.66	97.5
62.26	69.84	78.13	60.83	51.3	57.67	67.9	56.3	88.6	52.5	76.33	94.
64.88	80.75	71.66	50.3	76.67	67.57	64.45	69.5	72.15	59.77	61.4	72.1
90.25	97.4	89.4	57.5	70.8	56.4	96.2	88.77	96.	83.2	69.17	80.33
79.04	59.8	81.66	72.46	92.5	50.5	96.8	93.5	65.33	92.4	72.66	87.33
94.8	60.3	90.3	61.17	82.25	88.5	75.11	70.33	85.8	88.8	45.6	58.33
76.24	45.68.3	89.75	69.67	88.2	65.56	74.12	86.54	57.11	65.66	85.67	
96.3	72.31	70.3	77.86	94.75	69.32	84.9	60.1	86.6	70.15	66.62	79.16
76.7	65.25	93.3	68.15	80.25	76.77	65.9	91.25	54.14	55.55	95.65	
75.91											
66.2	70.66	59.4	58.92	51.	63.7	85.5	70.83	81.7	63.8	75.25	43.
61.16	56.12	60.66	69.58	70.04	79.67	49.67	76.3	75.67	61.3	82.33	88.33
88.45	96.75	84.25	78.5	77.85	82.46	74.3	73.67	60.01	91.33	69.16	91.7
65.83	74.83	58.5	72.17	87.6	75.85	79.87	85.9	85.6	85.7	73.11	88.66
82.56	78.34	89.33	66.15	87.25	69.53	53.17	79.19	77.16	54.08	52.2	83.5
85.57	60.33	67.23	47.6	73.25	69.88	84.53	89.9	79.3	93.1	92.25	91.58
75.14	73.3	77.1	79.5	59.6	50.4	74.89	63.33	77.23	83.43	83.3	95.7
90.8	75.75	47.2	86.31	85.17	63.83	64.57	73.69	66.86	81.75	55.6	62.12
80.58	74.96	70.14	69.54	61.83	72.33	74.33	89.04	71.67	60.36	90.66	60.5
70.26	57.2	87.69	85.56	64.2	45.5	70.5	93.9	57.58	85.1	64.1	79.1
91.83	88.93	46.33	94.17	95.9	81.9	76.12	84.28	55.4	58.04	91.8	74.7
83.25	91.4	54.66	82.9	66.46	71.52	56.17	59.66	62.7	78.25	61.92	68.16
53.83	47.83	92.75	63.44	87.05	92.08	68.69	55.33	89.67	63.77	74.53	60.6
77.83	56.04	70.67	65.1	87.3	56.22	96.4	83.62	74.44	97.9	57.33	71.17
95.08	84.2	97.6	57.83	93.41	79.33	64.83	75.66	75.8	62.33	92.9	92.8
70.92	81.42	90.9	52.7	82.03	73.63	75.1	87.17	73.33	62.23	61.33	52.9
83.88	83.75	83.83	67.74	59.38	68.04	77.81	89.26	74.71	53.2	53.85	65.54
51.6	68.77	86.29	88.75	49.6	77.33	58.8	71.85	66.17	53.44	66.83	93.2
85.33	72.9	57.1	98.7	88.16	54.33	85.42	90.5	74.16	98.2	82.71	90.03
85.23	63.16	96.33	84.09	74.05	81.16	40.60.06	89.16	69.89	61.23	79.82	
65.72	55.8	76.85	69.3	90.17	53.16	61.85	89.58	83.23	67.3	93.33	52.8
73.17	52.34	63.5	77.11	62.11	97.2	78.88	65.12	55.44	53.55	80.22	62.9
55.66	94.91	76.83	64.08	83.34	85.88	63.3	53.4	54.8	54.5	74.25	79.75
60.44	85.66	55.5	51.23	90.83	67.1	83.16	79.7	56.1	68.17	82.53	87.14
63.9	74.14	89.91	77.06	60.42	75.33	96.25	69.12	77.56	86.46	81.26	64.31
71.83	86.91	81.25	54.83	59.11	91.1	81.67	53.33	82.55]			

***** 12board *****

['board of intermediate education,ap' 'cbse' 'state board' 'mp board'
'isc' 'icse' 'karnataka pre university board' 'up' 'p u board, karnataka'
'dept of pre-university education' 'bie' 'kerala state hse board'
'up board' '0' 'bseb' 'chse' 'puc' ' upboard'
'state board of intermediate education, andhra pradesh'
'karnataka state board'
'west bengal state council of technical education' 'wbchse'
'maharashtra state board' 'ssc' 'isc board'
'sda matric higher secondary school' 'uttar pradesh board' 'ibe' 'chsc' 'board of
intermediate' 'isce' 'upboard' 'sbtet'
'hisher seconadry examination(state board)' 'pre university'
'borad of intermediate' 'j & k board'
'intermediate board of andhra pardesh' 'rbse'
'central board of secondary education' 'jkbose' 'hbse'
'board of intermediate education' 'state' 'ms board' 'pue'
'intermediate state board' 'stateboard' 'hsc'
'electonincs and communication(dote)' 'karnataka pu board'
'government polytechnic mumbai , mumbai board' 'pu board'
'baord of intermediate education' 'apbie' 'andhra board'
'tamilnadu stateboard'
'west bengal council of higher secondary education' 'cbse,new delhi'
'u p board' 'intermediate' 'biec,patna'
'diploma in engg (e &tc) tilak maharashtra vidyapeeth' 'hsc pune'
'pu board karnataka' 'kerala' 'gsheb' 'up(allahabad)' 'nagpur'
'st joseph hr sec school' 'pre university board' 'ipe' 'maharashtra'
'kea' 'apsb' 'himachal pradesh board of school education' 'staae board'
'international baccalaureate (ib) diploma' 'nios'
'karnataka board of university' 'board of secondary education rajasthan'
'uttarakhand board' 'ua' 'scte vt orissa' 'matriculation'
'department of pre-university education' 'wbscte'
'preuniversity board(karnataka)' 'jharkhand accademic council' 'bieap'
'msbte (diploma in computer technology)'
'jharkhand acamedic council (ranchi)'
'department of pre-university eduction' 'biec' 'all india board' 'sjrcw'
' board of intermediate' 'msbte' 'sri sankara vidyalaya' 'chse, odisha'
'bihar board' 'maharashtra state(latur board)' 'rajasthan board'
'mpboard' 'state board of technical eduction panchkula' 'upbhsie' 'apbsc'
'state board of technical education and training'
'secondary board of rajasthan'
'tamilnadu higher secondary education board' 'jharkhand academic council'
'board of intermediate education,hyderabad' 'up baord' 'pu' 'dte'
'board of secondary education' 'pre-university'
'board of intermediate education, andhra pradesh' 'up board , allahabad'
'srv girls higher sec school,rasipuram'
'intermediate board of education, andhra pradesh'

'intermediate board examination'
'department of pre-university education, bangalore'
'stmiras college for girls' 'mbose'
'department of pre-university education(government of karnataka)' 'dpue'
'msbte pune' 'board of school education harayana' 'sbte, jharkhand'
'bihar intermediate education council, patna' 'higher secondary'
's j polytechnic' 'latur' 'board of secondary education, rajasthan'
'jyoti nivas' 'pseb' 'biec-patna'
'board of intermediate education, andra pradesh' 'chse,orissa'
'pre-university board' 'mp' 'intermediate board'
'govt of karnataka department of pre-university education'
'karnataka education board' 'board of secondary school of education'
'pu board ,karnataka' 'karnataka secondary education board'
'karnataka sslc' 'board of intermediate ap' 'u p'
'state board of karnataka' 'directorate of technical education,banglore'
'matric board' 'andhpрадеш board of intermediate education'
'stjoseph of cluny matrhrsecsschool,neyveli,cuddalore district' 'bte up'
'scet and vt ,orissa' 'hbse' 'jawahar higher secondary school'
'nagpur board' 'bsemp' 'board of intermediate education, andhra pradesh'
'board of higher secondary orissa'
'board of secondary education,rajasthan(rbse)'
'board of intermediate education:ap,hyderabad' 'science college'
'karnatak pu board' 'aiissce' 'pre university board of karnataka' 'bihar'
'kerala state board' 'uo board' 'cicse' 'karnataka board'
'tn state board' 'kolhapur divisional board, maharashtra'
'jaycee matriculation school'
'board of higher secondary examination, kerala' 'uttaranchal state board'
'intermidiate' 'bciec,patna' 'bice' 'karnataka state' 'state broad'
'wbbhse' 'gseb' 'uttar pradesh' 'ghseb'
'board of school education uttarakhand' 'gseb/technical education board'
'msbshse,pune' 'tamilnadu state board' 'board of technical education'
'kerala university' 'uttaranchal shiksha avam pariksha parishad'
'chse(concil of higher secondary education)'
'bright way college, (up board)' 'board of intermidiate'
'higher secondary state certificate' 'karanataka secondary board'
'maharashtra board' 'andhra pradesh state board' 'cgbse'
'diploma in computers' 'bte,delhi' 'rajasthan board ajmer' 'mpbse'
'pune board' 'state board of technical education' 'gshseb'
'amravati divisional board' 'dote (diploma - computer engg)' 'up bord'
'karnataka pre-university board' 'jharkhand board'
'punjab state board of technical education & industrial training'
'department of technical education' 'sri chaitanya junior kalasala'
'state board (jac, ranchi)' 'gujarat board' 'aligarh muslim university'
'tamil nadu state board' 'hse' 'karnataka secondary education'
'state board ' 'karnataka pre unversity board'

'ks rangasamy institute of technology'
'karnataka board secondary education' 'narayana junior college' 'bteup'
'board of intermediate(bie)' 'hsc maharashtra board' 'tamil nadu state'
'uttrakhand board' 'psbte' 'stateboard/tamil nadu'
'intermediate council patna' 'technical board, punchkula'
'board of intermidiate examination'
'sri kannika parameswari highier secondary school, udumalpet' 'ap board'
'nashik board' 'himachal pradesh board' 'maharashtra satate board'
'andhra pradesh board of secondary education' 'tamil nadu polytechnic'
'maharashtra state board mumbai divisional board'
'department of pre university education' 'dav public school,hehal'
'board of intermediate education, ap'
'rajasthan board of secondary education'
'department of technical education, bangalore' 'chse,odisha'
'maharashtra nasik board'
'west bengal council of higher secondary examination (wbchse)'
'holy cross matriculation hr sec school' 'cbsc' 'pu board karnataka'
'biec patna' 'kolhapur' 'bseb, patna' 'up board allahabad' 'intermideate'
'nagpur board,nagpur' 'diploma(msbte)' 'dav public school'
'pre university board, karnataka' 'ssm srsecschool' 'state bord'
'jstb,jharkhand' 'intermediate board of education' 'mp board bhopal'
'pub' 'madhya pradesh board' 'bihar intermediate education council'
'west bengal council of higher secondary eucation'
'isc board , new delhi' 'mpc'
'certificate for higher secondary education (chse)orissa'
'maharashtra state board for hsc' 'board of intermeadiate education'
'latur board' 'andhra pradesh' 'karnataka pre-university'
'lucknow public college' 'nagpur divisional board'
'ap intermediate board' 'cgbse raipur' 'uttranchal board' 'jiec'
'central board of secondary education, new delhi'
'bihar school examination board patna'
'state board of technical education harayana' 'mp-bse' 'up bound'
'dav public school sec 14'
'haryana state board of technical education chandigarh'
'council for indian school certificate examination'
'jaswant modern school' 'madhya pradesh open school' 'aurangabad board'
'j&k state board of school education'
'diploma (maharashtra state board of technical education)'
'board of technicaleducation ,delhi'
'maharashtra state boar of secondary and higher secondary education'
'hslc (tamil nadu state board)' 'karnataka state examination board'
'puboard' 'nasik' 'west bengal board of higher secondary education'
'up board,allahabad' 'board of intrmediate education,ap' 'cbese'
'karnataka state pre- university board'
'state board - west bengal council of higher secondary education : wbchse'

'maharashtra state board of secondary & higher secondary education'
'biec, patna' 'state syllabus' 'cbse board' 'scte&vt'
'board of intermediate,ap'
'secnior secondary education board of rajasthan'
'maharashtra board, pune' 'rbse (state board)'
'board of intermidiate education,ap'
'board of high school and intermediate education uttarpradesh'
'higher secondary education' 'board fo intermediate education, ap'
'intermedite' 'ap board for intermediate education' 'ahsec'
'punjab state board of technical education & industrial training, chandigarh'
'state board - tamilnadu' 'jharkhand acedemic council'
'scte & vt (diploma)' 'karnataka pu' 'board of intmediate education ap'
'up-board' 'boardofintermediate']

***** CollegeID
***** [1141 5807 64 ... 3572 6327
4883]

***** CollegeTier *****
[2 1]

***** Degree *****
['B.Tech/B.E.' 'MCA' 'M.Tech./M.E.' 'M.Sc. (Tech.)']

***** Specialization *****
['computer engineering' 'electronics and communication engineering'
'information technology' 'computer science & engineering'
'mechanical engineering' 'electronics and electrical engineering'
'electronics & telecommunications'
'instrumentation and control engineering' 'computer application'
'electronics and computer engineering' 'electrical engineering'
'applied electronics and instrumentation'
'electronics & instrumentation eng' 'information science engineering'
'civil engineering' 'mechanical and automation'
'industrial & production engineering'
'control and instrumentation engineering' 'metallurgical engineering'
'electronics and instrumentation engineering' 'electronics engineering'
'ceramic engineering' 'chemical engineering' 'aeronautical engineering'
'other' 'biotechnology' 'embedded systems technology'
'electrical and power engineering' 'computer science and technology'
'mechatronics' 'automobile/automotive engineering' 'polymer technology'
'mechanical & production engineering' 'power systems and automation'
'instrumentation engineering' 'telecommunication engineering'
'industrial & management engineering' 'industrial engineering'
'computer and communication engineering'
'information & communication technology' 'information science'

```
'internal combustion engine' 'computer networking'  
'biomedical engineering' 'electronics' 'computer science']
```

```
***** collegeGPA  
***** [78.    70.06 70. ... 65.05  
74.73 70.42]
```

```
***** CollegeCityID  
***** [1141 5807 64 ... 3572 6327  
4883]
```

```
***** CollegeCityTier  
***** [0 1]
```

```
***** CollegeState *****  
['Andhra Pradesh' 'Madhya Pradesh' 'Uttar Pradesh' 'Delhi' 'Karnataka'  
'Tamil Nadu' 'West Bengal' 'Maharashtra' 'Haryana' 'Telangana' 'Orissa'  
'Punjab' 'Kerala' 'Gujarat' 'Rajasthan' 'Chhattisgarh' 'Uttarakhand'  
'Jammu and Kashmir' 'Jharkhand' 'Himachal Pradesh' 'Bihar' 'Assam' 'Goa'  
'Sikkim' 'Union Territory' 'Meghalaya']
```

```
***** GraduationYear *****  
[2011 2012 2014 2016 2013 2010 2015 2009 2017 0 2007] *****  
English *****  
[515 695 615 635 545 560 590 605 565 495 380 395 485 685 465 455 385 370  
625 575 415 535 580 475 570 430 450 510 425 555 300 505 440 525 420 640  
444 630 665 675 325 405 375 315 710 345 250 350 275 360 265 595 585 520  
500 735 765 335 490 660 355 530 365 655 730 445 720 645 650 875 534 454  
544 295 285 435 464 705 554 745 280 825 290 715 310 215 700 870 305 524  
755 790 800 205 725 780 404 770 805 180 830 795 255 324 775 394  
240 225 850 684 334]
```

```
***** Logical *****  
[585 610 545 625 555 435 670 565 455 605 580 425 520 530 495 445 535 360  
335 510 570 375 405 485 475 525 640 595 560 340 395 415 465 505 385 460  
410 500 645 480 355 450 440 470 255 305 590 630 365 350 325 400 205 655  
295 345 390 665 515 540 680 245 620 420 575 635 554 315 615 215 370 300  
274 685 324 675 650 464 684 275 334 544 454 534 404 795 285 715 700 674  
690 695 394 270 705 310 490 330 280 735 380 290 265 240 195 235 660]
```

```
***** Quant *****  
[525 780 370 625 465 620 380 590 530 545 565 715 470 645 355 515 435 445  
485 270 630 575 405 605 385 695 450 295 430 415 635 475 460 825 500 455  
554 595 495 665 250 310 325 390 510 535 340 440 705 534 400 395 570 750  
330 320 454 365 615 505 425 235 210 585 810 555 735 560 524 690 870 765  
675 520 655 305 725 840 650 375 720 265 280 464 404 800 680 260 674 760
```

```

345 335 165 685 544 215 180 795 200 860 334 285 514 195 494 214 275 315
324 175 684 225 740 805 444 410 135 255 220 755 855 145 245 885 120 900
794 775 745 504 820 150 710 190 185 155 580 394]

```

```

***** Domain *****
[ 0.63597876 0.96060325 0.45087658 0.97439611 0.12450207 -1.
  0.35653649 0.82958460 0.69447933 0.49359639 0.76567358 0.9682375
  0.22948175 0.53838689 0.30840058 0.91139528 0.56326782
  0.86468541 0.64938971 0.74475835 0.88412251 0.88162007
  0.20739217 0.48674701 0.67074315 0.62264292 0.41383826
  0.52592258 0.73579571 0.13044174 0.23780284 0.11213944
  0.37755142 0.06696071 0.08005528 0.92564577 0.84312373
  0.91686996 0.78330354 0.60005718 0.79293628 0.79358061
  0.16563309 0.75537512 0.99990456 0.33878635 0.91077016
  0.98205712 0.84224832 0.01854094 0.05316031 0.94211655
  0.12301673 0.48834798 0.37605959 0.0587928 0.10487136
  0.60064396 0.70409041 0.14478989 0.81941653 0.65576694
  0.02106623 0.44461772 0.83762073 0.72598415 0.95389978
  0.04099931 0.02196911 0.3423149 0.53586282 0.90148957
  0.96177212 0.67964464 0.93839914 0.19015341 0.99546472
  0.99000876 0.97879929 0.41433743 0.8246664 0.49063696
  0.96600692 0.43696265
  0.99225919 0.97629256 0.98546139 0.90915194 0.86372418 0.07454627
  0.66183448 0.93037061 0.95224557 0.96090309 0.97706647 0.55738951
  0.7908818 0.99966434 0.8799152 0.85882669 0.97166349 0.19376844
  0.24545566 0.98374997 0.98720709 0.97952174 0.86873659 0.45901584
  0.02231329 0.94513486 0.27604723 0.53678267 0.42395140 0.29876913
  0.20671062 0.12569018 0.16363093 0.99138693 0.94327216
  0.60703404 0.99139758 0.25577819 0.96221701 0.99123063
  0.87550391 0.02331217 0.44714789 0.14325654 0.84084097
  0.33618507 0.30451091 0.27925851 0.84980271 0.03293702
  0.65410747 0.28677781 0.21678549 0.95868202 0.32874638
  0.98466189 0.99674449 0.99434211 0.70082616 0.99887611
  0.86262547 0.60553355 0.68456507 0.71452936 0.44619825 0.92514031
  0.99526571 0.59828069 0.98652502 0.11166096 0.11655256 0.8957765
  0.1559085 0.99764277 0.88790332 0.88570368 0.99405087
  0.02781507 0.90394078 0.93174969 0.09575376 0.1925871
  0.99868025 0.03114969 0.61061199 0.75837988 0.63958738
  0.27350033 0.0109953 0.21325143
  0.99925029 0.06844575 0.04222313 0.77657832 0.75618011 0.99808656
  0.99982917 0.01656464 0.00815478 0.89984384 0.92879365 0.99769826
  0.9956138 0.9788553 0.03056599 0.04219243 0.18477158 0.92460994
  0.81062051 0.99742783 0.66033499 0.99615597 0.99859145 0.00275015
  0.51986392 0.99729897 0.20273159 0.38009183 0.52173598 0.0139623
  0.24249992 0.29429542 0.40995220 0.80558321 0.24320866 0.96532716
  0.4845907 0.97524659 0.1798739 0.37113867 0.86613982 0.11502273

```

```

0.99991041 0.99179243 0.22105882 0.04173320.06222129 0.83968608
0.41552456 0.25302837 0.99657055 0.97289871 0.99896694 0.99391731
0.52911566 0.55321638 0.01770484 0.91541843 0.29959730.66732645
0.27845741 0.00853725 0.99858824 0.28278814 0.08874741 0.79984821
0.99653553 0.90474069 0.93858826]

```

***** ComputerProgramming *****

```

[445 -1 395 615 645 405 735 385 485 605 495 355 515 545 425 525 455 475
565 535 335 345 465 415 435 155 375 555 305 315 804 285 575 505 195 225
595 275 334 365 685 655 625 585 665 325 235 255 205 494 695 635 215 464
295 394 245 715 265 135 105 524 165 175 125 675 454 745 185 214 145 544
725 840 404 755 705 115 554]

```

***** ElectronicsAndSemicon *****

```

[ -1 466 233 366 324 266 333 356 420 260 228 388 300 292 433 196 200 164
400 484 500 452 516 166 533 566 612 133 548]

```

***** ComputerScience *****

```

[ -1 407 346 376 500 438 315 253 469 192 530 284 223 561 684 592
623 653 130 715]

```

***** MechanicalEngg *****

```

[ -1 469 313 286 253 366 446 206 438 332 393 383 260 561 553 376 526 284
409 473 340 223 420 538 346 435 512 407 580 280 358 500 315 254 616 564
233 306 461 180 606 623]

```

***** ElectricalEngg *****

```

[ -1 484 606 393 500 553 580 446 420 324 388 356 313 633 516 366 612 452
526 548 228 433 473 676 292 660 411 286 340 260 206]

```

***** TelecomEngg *****

```

[ -1 206 313 420 260 393 366 446 324 340 286 473 484 452 233 292 526 153
516 356 548 228 196 164 388 500]

```

***** CivilEngg *****

```

[ -1 320 400 388 260 440 356 292 500 200 300 452 322 340 166 277 516 380
433 280 420 460 480]

```

***** conscientiousness *****

```

[ 0.9737 -0.7335 0.2718 0.0464 -0.881 -0.3027 1.7081 -0.0154 -0.159
-1.308 -2.272 0.1282 0.3555 0.7027 1.7465 1.1336 0.8463 0.8192
-0.1082 -1.0355 -0.4463 0.4155 0.99-3.1994 -0.4173 1.5644 -0.4854
-1.0208 0.3941 -0.8772 0.51 -0.5899 -2.5039 1.2828 0.335 -0.3014
1.8517 -1.1644 -2.2351 0.6646 -0.2628 -1.8825 -1.4517 0.5591 1.4208
-0.7264 -0.5116 -1.7389 0.2009 -0.0696 -2.5811 -2.3134 1.2772 -2.8879

```

```

  1.4374 -1.3447 0.1623 1.7156 -1.9629 -2.457      1.9953 -2.0262 -
  2.1175 -2.7443 -1.4606 0.8578 -1.1901 -0.7651 -0.5719 -2.1698 -1.8083
  1.592
-0.9969 -1.3742 -1.4992 -1.5953 -3.6631 1.1283 -3.6060 1.788 0.2782
-1.6538 -3.3539 -1.1128 -3.3188 0.4285 0.7419 -0.6491 -0.51-0.5236
-0.9653 -3.1752 0.7208 1.3215 -1.6924 -0.1855 -0.6749 0.2318
1.5533 1.0768 1.3686 -0.5332 -0.2632 -1.2287 -1.5765 1.9011 1.0896
                           0.215
-3.4624 0.3836 -3.0448 -2.6007 1.2056 -1.295 -2.7357 -0.0415 1.6692
  0.626 -2.4266 -2.1561 -2.8903 -4.1267 0.4034 -1.9243 -1.3025 -1.5964
  1.7852 0.6696 0.5522 0.8479 -1.0135 -0.1982 -3.8933 -3.5085 -3.7496
-4.0369 -1.977 -3.0315 1.2266 -0.4595 0.8986]

```

***** agreeableness *****

```

[ 0.8128 0.3789 1.7109 0.3448 -0.2793 -0.6201 -0.1054 1.2114 0.5454
  1.1248 0.0328 0.7119 1.9048 1.0449 0.2668 0.9688 -0.5913 -2.1186
  0.8027 1.2028 0.1888 1.3779 -1.8393 0.6568 -0.4536 -0.5213 0.2124
  1.2808 -1.1196 0.2578 -2.4516 1.7488 -0.1206 0.0924 -0.0842 -0.4353
  1.5444 -0.9531 -2.6847 -0.1232 -3.7836 1.4368 0.8784 -1.4526 0.0459
-0.2871 1.0858 1.7878 -0.7866 0.8229 -1.6191 -1.2861 0.6178 -0.2012
  0.5008 -0.9033 1.5538 -4.2831 0.8518 0.1498 -1.9953 -1.3713 -2.9314
-1.2153 -0.7473 -0.5523 -1.0593 0.4934 -1.7856 -1.9521 -5.6151 1.5928
-0.6693 -1.8855 -2.4633 -2.6193 0.7348 -0.8865 -1.6833 1.3198 -5.1156
-0.3183 0.3731 -1.2543 -0.7993 -2.1903 0.7816 0.6009 0.3002 -3.6171
-2.6181 -1.5273 -2.7754 1.5081 -2.1513 -2.2851 1.6708 -0.7863 -2.3073
-3.4506 -0.3684 0.7135 -0.6867 -3.0874 -5.7816 0.3838 0.3123 0.1125
-1.4859 -0.0873 -3.1176 0.4488 0.8993 -3.1264 -3.0094 -2.0733 -3.9501
-2.9511 -1.7223 -0.1374 -4.7826 -1.0905 -0.6504 0.0875 -3.8284 -
  1.4883 0.9117 1.5293 0.6211 -2.7846 -0.1334 0.4395 -1.0203 0.9028
  -1.7056
-3.2434 0.5121 0.8351 0.0762 -1.6313 -2.4243 -1.1373 -3.3994 -2.6583
-0.9884 1.3476 -0.4778 -0.0651 -0.832 ]

```

***** extraversion *****

```

[ 0.5269 1.2396 0.1637 -0.344 -1.0697 -2.2954 -1.0379 0.01-0.6048
-0.9122 0.0914 0.8171 -0.7585 -0.598 0.672 0.7785 -1.0659 1.3933
-0.2714 -1.3599 -1.9881 0.1357 -0.9245 -1.7954 0.0552 -0.0537 1.0859
  0.3174 2.1129 0.4711 0.6248 0.2366 -0.5349 -0.4511 -0.6343 -0.7794
  0.3817 1.8331 -2.2308 -0.6582 -0.2974 -2.6028 -2.4491 -1.2196 2.1617
-0.1988 -0.4891 0.9322 -1.2148 1.7007 1.1437 -0.1437 1.8543 1.547
  0.8809 -1.6807 0.7083 0.5994 1.1074 -1.5776 0.9623 -1.8344 -1.5051
-0.1626 0.926 -3.2176 -1.3733 1.2525 -1.4688 -2.1418 1.6881.1558
-1.6502 0.4906 1.5428 -1.5271.9782 -1.9405 0.2113 -2.3759 0.2075
-2.0856 -0.3803 0.2729 -4.6009 -1.2511 -0.0319 2.3154 -3.525 -0.8157
-0.6355 0.6984 1.3977 -3.0639 2.1234 2.008 -2.3396 -1.9042 1.1804
-2.775 0.73 0.164 -0.824 -2.1219 -0.7068 1.4702 1.4267 -1.1422

```

```

-1.6865 -0.2882 -2.6662 1.0348 -1.0334 -2.9565 -4.2935 -2.7565 0.065
-0.1996 -3.537 -2.0131 -0.4226 -3.8636 1.5791 -0.1408 -1.9408 -0.7026
-2.4485 -2.9102 -0.6339 1.6484 -3.3713 0.3034 1.3614 2.5354 -0.0933
1.9801 0.3292 -0.8703 -1.0116 -0.4899 0.1138 -1.7086 0.6388 -3.9861
-2.521 -2.8113 0.9042 0.2477 -3.8324 -0.3149 -0.2516 -4.4472 -1.7083
-1.2056]

```

***** nueroticism *****

```

[ 1.3549 -0.1076 -0.8682 -0.4078 0.09163 -0.7415 -2.0092 0.1459
  0.9066 0.1798 -0.995 -0.2902 -0.6147 -1.6289 -0.2344 0.06223
  0.7798 -0.4879 0.53231.8249 -0.3612 0.0623 -1.2303 -1.5021
  1.1601 -1.8824 -0.735 0.2727 -2.1998 -1.2486 -1.1218 0.2973
  1.7074 0.26793 -2.2879 1.0024 0.653 -0.4821 -0.6428 -0.349
  1.5404 0.3995 0.64981.794 0.0192 1.0333 0.4148 0.8848
  -0.7603 0.88483 -0.52531.2869 1.4724 -0.8778 -0.1727 0.0035
  -1.1128 1.6672 0.64983 1.1199 -2.3895 -1.7556 -2.136 0.53233
  -0.0552 -0.26087 0.17983 -0.61340.219 0.5262 0.76733
                           0.00353
  -1.3753 -0.29027 1.0611 -0.7015 2.6475 0.7673 1.4136 0.29733
  0.0917 -1.4653 0.7967 2.2949 -1.3478 -0.2609 1.5899 -
                           0.05527
  2.1774 1.4297 -0.3414 -0.78967 -1.1422 2.301 -0.9953 0.9169
  -0.52527 -0.87777 -2.643-0.40777 -0.01 -0.4371 1.58993
                           2.4278
  1.3255 2.9349 0.6204 -2.0529 -1.8179 -1.05407 2.0599 2.0475
  -1.3184 1.85433 -0.5644 -1.11277 -0.365 -1.4066 1.2374 -
  1.9033 1.9424 -2.5163 0.70853 -1.5828 1.67803 1.7662 -1.58287
  2.53
  -0.70157 0.44423 0.4442 2.1743 1.9207 0.3561 -1.7004 -2.5047
  1.1492 0.3756 -1.671 0.8457 -0.17277 -0.99527 -0.76027 -1.34787
  -1.23027 0.7086 1.5018 -1.3255 0.973 -2.1704 1.06113 2.1187
  0.40413 -0.0846 -1.9354 -0.43717 1.41363 3.3525 0.4041 1.76613
  1.14923 -2.2627 -0.64277 -1.46537 1.32553 0.1477 -0.7897 -
  0.61337 1.11983 3.235 0.2679 1.70743 -0.34897 2.4125 1.35483
  0.62043 2.6814 -1.0541 1.8543 -1.1911 0.41483 -0.9659 3.0617
  1.23733
  -0.7496 2.765 1.82493 2.7356 -1.49467 2.4712 -0.5382 2.29493
  -0.5958 2.47123 -1.5899 -0.08457 -1.7591 3.3152 0.35603 2.2068
  -1.3008 -0.96597 -0.8177 0.7493 1.678 1.00233 1.50173 0.97293 1.0747
  1.5578 2.5546 -1.14217 0.6605 2.5593 0.9553 0.2759
  2.0306 ]

```

***** openness_to_experience *****

```

[-0.4455 0.8637 0.6721 -0.9194 -0.1295 -0.8608 -1.0872 1.247 -0.2859
  0.0973 0.0284 -1.2354 1.2528 1.4386 0.3444 -1.3539 -2.7769 -5.0763
  -0.6692 -0.2875 1.1343 -0.0943 -0.7615 0.2889 -0.4776 1.0554 1.8224

```

```

0.6603 -1.4356 -1.359 -3.1602 0.1864 0.5024 -1.244 -0.1543 -
0.6035 -5.477 -1.8189 -2.3937 0.3049 0.8183 0.4805 1.6302 -
2.2021 1.2923 0.9763 -2.9731 -1.0524 -1.6273 1.6082 -5.2679 -
0.169 -1.0774 -3.9605
-0.0506 -0.5245 -1.8673 -0.8799 -0.9984 0.5419 -4.5015 -2.1833 -2.3415
0.1275 -0.643 -1.3934 1.4502 -6.9925 -1.7093 -0.8782 -3.4471 -2.0105
-0.4137 0.4234 -1.5513 -1.1169 -1.425 0.8973 0.0916 -2.5853 -2.0253
0.0679 0.7788 -0.4139 -3.3518 -3.735 -2.3412 -2.9686 1.0031 -4.3099
-0.0844 1.0158 -2.7595 1.3976 -0.4061.4186 -0.4601 -1.0458 -5.8428
-2.0648 0.167 0.7941 1.2121 -3.5434 -6.8009 0.1187 0.585 -3.9266
-2.6572 -3.6051 -3.763 -5.686 -1.1291 -0.0167 -2.8152 -1.8278 -5.6512
0.7631 -5.4595 1.0395 -0.4392 -3.1311 -0.5081 -1.4724 -2.3017 -1.6662
-1.9463 0.9404 -2.4202 -0.2511 0.7906 -1.9234 -6.6092 -7.3757 -0.1521
1.4003 -0.8045 0.376 -1.8386 0.7657 0.7104 -0.4229]

```

1.3.1 Observations

Column Name	Observation
DOL	It have the Value ‘present’, that means the employee is working on the company now. Replace ‘present’ with Today’s Date
JobCity	Contains ‘-1’ in the column Considered to be a Missing Value and City Names are not proper
10board	Contains ‘0’ in the column considered as Missing Value and having high Cardinality. convet to State, CBSE, and ICSE Board
12board	Contains ‘0’ in the column considered as Missing Value and having high Cardinality. convet to State, CBSE, and ICSE Board
CollegeState	Contains ‘Union Territory’ in the column considered as Missing Value. we don’t know
Domain	College state belongs to which Union Territory Contains ‘-1’ in the column Considered to be a Missing Value
Designation	Contains ‘get’ in the column Considered to be a Missing Value

DOJ, DOB columns need to be in Date Time format, convert Data Type

- **DOL column having ‘present’ - Replace with present date**
 - I am considering today as '20/03/2020' (i.e; Before COVID-19)

```
[9]: data['DOJ'] = pd.to_datetime(data['DOJ']).dt.date
      data['DOB'] = pd.to_datetime(data['DOB']).dt.date
```

```
[10]: today = pd.to_datetime('20/03/2020')
      today
```

```
[10]: Timestamp('2020-03-20 00:00:00')
```

```
[11]: data['DOL'].replace('present', today, inplace=True)
```

```
[12]: # I Derived column a 'period' from 'DOJ' & 'DOL' columns, by subtracting the ↵
      ↵'DOJ' from 'DOL' (i.e; DOL - DOJ) in Years
```

```
      data['period'] = pd.to_datetime(data['DOL']).dt.year - pd.
      ↵to_datetime(data['DOJ']).dt.year
```

```
[13]: data['DOB_Year'] = pd.to_datetime(data['DOB']).dt.year
```

```
[14]: data.head()
```

```
[14]:
```

	ID	Salary	DOJ	DOL \
0	203097	420000.0	2012-06-01	2020-03-20 00:00:00
1	579905	500000.0	2013-09-01	2020-03-20 00:00:00
2	810601	325000.0	2014-06-01	2020-03-20 00:00:00
3	267447	1100000.0	2011-07-01	2020-03-20 00:00:00
4	343523	200000.0	2014-03-01	3/1/15 0:00

	Designation	JobCity	Gender	DOB	10percentage \
0	senior quality engineer	Bangalore	f	1990-02-19	84.3
1	assistant manager	Indore	m	1989-10-04	85.4
2	systems engineer	Chennai	f	1992-08-03	85.0
3	senior software engineer	Gurgaon	m	1989-12-05	85.6
4	get	Manesar	m	1991-02-27	78.0

	10board	12graduation	12percentage \
0	board ofsecondary education,ap	2007	95.8
1	cbse	2007	85.0
2	cbse	2010	68.2
3	cbse	2007	83.6
4	cbse	2008	76.8

	12board	CollegeID	CollegeTier	Degree \
--	---------	-----------	-------------	----------

0 board of intermediate education, ap 1141 2 B.Tech/B.E.
 1 cbse 5807 2 B.Tech/B.E.
 2 cbse 64 2 B.Tech/B.E.
 3 cbse 6920 1 B.Tech/B.E.
 4 cbse 11368 2 B.Tech/B.E.

	Specialization	college	GPA	CollegeCityID
0	computer engineering	78.00	1141	
1	electronics and communication engineering	70.06	5807	
2	information technology	70.00	64	
3	computer engineering	74.64	6920	
4	electronics and communication engineering	73.90	11368	

	CollegeCityTier	CollegeState	GraduationYear	English	Logical	Quant
0	0	Andhra Pradesh	2011	515	585	525
1	0	Madhya Pradesh	2012	695	610	780
2	0	Uttar Pradesh	2014	615	545	370
3	1	Delhi	2011	635	585	625
4	0	Uttar Pradesh	2012	545	625	465

	Domain	ComputerProgramming	ElectronicsAndSemicon	ComputerScience
0	0.635979	445	-1	-1
1	0.974396	615	-1	-1
2	0.124502		-1	
3				233
4				-1

	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg	conscientiousness
0	-1	-1	-1	-1	0.9737
1	-1	-1	-1	-1	-
2	-1	-1	-1	-1	0.2718
3	-1	-1	-1	-1	0.0464
4	-1	-1	-1	-1	-
					0.8810

	agreeableness	extraversion	nueroticism	openess_to_experience	period
0	0.8128	0.5269	1.35490	-0.4455	8
1	0.3789	1.2396	-0.10760	0.8637	7
2	1.7109	0.1637	-0.86820	0.6721	6
3	0.3448	-0.3440	-0.40780	-0.9194	9
4	-0.2793	-1.0697	0.09163	-0.1295	1

	DOB_Year
0	1990
1	1989
2	1992
3	1989
4	1991

```
[15]: data['DOB'] = pd.to_datetime(data['DOB'], dayfirst=True)
data['DOJ'] = pd.to_datetime(data['DOJ'], dayfirst=True)
data['DOL'] = pd.to_datetime(data['DOL'], dayfirst=True)

[16]: data.info()
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to
3997 Data columns (total 40
columns):
 #   Column           Non-Null Count Dtype  
 ---  --  
 0   ID               3998 non-null    int64  
      null                
 1   Salary            3998 non-null    float64 
      null                
 2   DOJ              3998 non-null    datetime64[ns]
      null                
 3   DOL              3998 non-null    datetime64[ns]
      null                
 4   Designation       3998 non-null    object  
      null                
 5   JobCity           3998 non-null    object  
      null                
 6   Gender             3998 non-null    object  
      null                
 7   DOB              3998 non-null    datetime64[ns]
      null                
 8   10percentage     3998 non-null    float64 
      null                
 9   10board            3998 non-null    object  
      null                
 10  12graduation      3998 non-null    int64  
      null                
 11  12percentage     3998 non-null    float64 
      null                
 12  12board            3998 non-null    object  
      null                
 13  CollegeID          3998 non-null    int64  
      null                
 14  CollegeTier         3998 non-null    int64  
      null                
 15  Degree             3998 non-null    object  
      null                
 16  Specialization      3998 non-null    object  
      null
```

```

17 collegeGPA          3998 non-      float64
                           null
18 CollegeCityID        3998 non-      int64
                           null
19 CollegeCityTier       3998 non-      int64
                           null
20 CollegeState         3998 non-      object
                           null
21 GraduationYear       3998 non-      int64
                           null
22 English              3998 non-      int64
                           null
23 Logical              3998 non-      int64
                           null
24 Quant                3998 non-      int64
                           null
25 Domain               3998 non-      float64
                           null
26 ComputerProgramming   3998 non-      int64
                           null
27 ElectronicsAndSemicon 3998 non-      int64
                           null
28 ComputerScience       3998 non-nullint64
29 MechanicalEngg        3998 non-nullint64
30 ElectricalEngg        3998 non-nullint64
31 TelecomEngg          3998 non-nullint64
32 CivilEngg             3998 non-nullint64
33 conscientiousness     3998 non-nullfloat64
34 agreeableness         3998 non-nullfloat64
35 extraversion          3998 non-nullfloat64
36 nueroticism           3998 non-nullfloat64
37 openness_to_experience 3998 non-      float64
                           null
38 period                3998 non-nullint32
39 DOB_Year              3998 non-nullint32
dtypes: datetime64[ns](3), float64(10), int32(2), int64(17),
object(8) memory usage: 1.2+ MB

```

12graduation & GraduationYear means Age at completion of 12graduation & GraduationYear respectively

- using this columns we can derive New columns (12GradAge, GradAge) that holding Age of person at 12graduation & GraduationYear

GraduationYear having ‘0’ in the column considered as missing values - impute the missing values using Mode imputation Method

```
[17]: data['GraduationYear'].mode()[0]
```

```
[17]: 2013
```

```
[18]: data['GraduationYear'].replace(0, data['GraduationYear'].mode()[0],  
    ↪inplace=True) data['GraduationYear'] =  
pd.to_datetime(data['GraduationYear']) data['GYear'] =  
pd.to_datetime(data['GraduationYear']).dt.year
```

```
[19]: data[['12graduation', 'GYear']]
```

```
[19]:   12graduation GYear  
0           2007 1970  
1           2007 1970  
2           2010 1970  
3           2007 1970  
4           2008 1970  
...         ... ...  
3993        2006 1970  
3994        2009 1970  
3995        2008 1970  
3996        2010 1970  
3997        2008 1970  
[3998 rows x 2 columns]
```

```
[20]: data['12GradAge'] = abs((pd.to_datetime(data['12graduation']).dt.year) - (pd.  
    ↪to_datetime(data['DOB']).dt.year))  
data['GradAge'] = abs((pd.to_datetime(data['GYear']).dt.year) - (pd.  
    ↪to_datetime(data['DOB']).dt.year))
```

```
[21]: # no.of 0's per column
```

```
(data==0).astype('int').sum()
```

```
[21]: ID                  0  
Salary               0  
DOJ                 0  
DOL                 0  
Designation          0  
JobCity              0  
Gender               0  
DOB                 0  
10percentage        0  
10board              0  
12graduation         0  
12percentage        0  
12board              0  
CollegeID            0
```

```
CollegeTier          0
Degree              0
Specialization      0
collegeGPA          0
CollegeCityID       0
CollegeCityTier     2797
CollegeState         0
GraduationYear      0
English              0
Logical              0
Quant                0
Domain               0
ComputerProgramming 0
ElectronicsAndSemicon 0
ComputerScience      0
MechanicalEngg      0
ElectricalEngg       0
TelecomEngg          0
CivilEngg            0
conscientiousness    0
agreeableness        0
extraversion          0
nueroticism          0
openess_to_experience 0
period               319
DOB_Year             0
GYear                0
12GradAge            0
GradAge              0
dtype:
int64
```

```
[22]: data.isin([-1, 'NaN']).sum()
```

```
[22]: ID                  0
Salary              0
DOJ                 0
DOL                 0
Designation         0
JobCity             0
Gender              0
DOB                 0
10percentage        0
10board             0
12graduation        0
12percentage        0
12board             0
CollegeID           0
```

```
CollegeTier          0
Degree              0
Specialization      0
collegeGPA          0
CollegeCityID       0
CollegeCityTier     0
CollegeState         0
GraduationYear      0
English              0
Logical              0
Quant                0
Domain               246
ComputerProgramming  868
ElectronicsAndSemicon 2854
ComputerScience      3096
MechanicalEngg       3763
ElectricalEngg        3837
TelecomEngg          3624
CivilEngg            3956
conscientiousness    0
agreeableness        0
extraversion          0
nueroticism          0
openess_to_experience 0
period               0
DOB_Year             0
GYear                0
12GradAge            0
GradAge              0
dtype:
int64
```

1.4 Designation Column

```
[23]: data['Designation']
```

```
[23]: 0      senior quality engineer
1      assistant manager
2      systems engineer
3      senior software engineer
4      get
...
3993      software engineer
3994      technical writer
3995  associate software
        engineer
3996      software developer
3997      senior systems engineer
Name: Designation, Length: 3998, dtype: object
```

```
[24]: data[data['Designation']=='get'][['Designation', 'JobCity', 'Specialization']]
```

	Designation	JobCity	Specialization
4	get	Manesar	electronics and communication engineering
332	get	-1	mechanical and automation
350	get	Faridabad	electrical engineering
1717	get	Hyderabad	mechanical engineering
1897	get	Hyderabad	mechanical engineering
2140	get	Indore	mechanical engineering
2318	get	Lucknow	electronics and electrical engineering
2443	get	kharagpur	mechanical engineering
2666	get	Hyderabad	electronics and electrical engineering
2757	get	Nashik	mechanical engineering
3045	get	Sahibabad	electronics and electrical engineering
3126	get	Bhopal	mechanical engineering
3594	get	MEERUT	mechanical engineering
3980	get	haryana	mechanical engineering

Observation:

- Where ever the designation is 'get' - 70% of them belongs to mechanical engineering and Remaining 30% Belongs to electronics and electrical engineering
- Impute 'get' with the mode of mech and eee

```
[25]: # Mechanical Engg.  
mech = data[data['Specialization'].isin(['mechanical and automation',  
                                         'mechanical engineering'])]['Designation'].mode()[0]  
  
# eee  
  
eee = data[data['Specialization'].isin(['electronics and electrical _  
                                         'engineering'])]['Designation'].mode()[0]  
  
print(mech, '\n', eee)
```

production engineer
system engineer

Impute the values into the Designation - 'get'

```
[26]: data.loc[data['Specialization'].isin(['mechanical and automation', 'mechanical _  
                                         'engineering'])]['Designation'].replace('get', mech, inplace=True)  
data['Designation'].replace('get', eee, inplace=True)
```

1.5 JobCity Cloumn

```
[27]: data['JobCity'].value_counts()
```

```
[27]: JobCity
```

```
Bangalore      627  
-1            461  
Noida          368  
Hyderabad     335  
Pune           290  
...  
Tirunelvelli    1  
Ernakulam      1  
Nanded          1  
Dharmapuri     1  
Asifabadbanglore 1  
Name: count, Length: 339, dtype: int64
```

```
[28]: data['JobCity'].replace('-1', 'unknown', inplace=True)  
data['JobCity'] = data['JobCity'].apply(lambda x: x.strip().upper())
```

```
[29]: city = data['JobCity'].unique()
```

```
[30]: city.sort()
```

```
[31]:
```

```
city  
[31]: array(['A-64,SEC-64,NOIDA', 'AGRA', 'AHMEDABAD', 'AHMEDNAGAR',  
           'AL JUBAIL,SAUDI ARABIA', 'ALLAHABAD', 'ALWAR', 'AM', 'AMBALA',  
           'AMBALA CITY', 'ANGUL', 'ARIYALUR', 'ASANSOL', 'ASIFABADBANGLORE',  
           'AURANGABAD', 'AUSTRALIA', 'BADDI HP', 'BAHADURGARH', 'BANAGALORE',  
           'BANAGLORE', 'BANGALORE', 'BANGLORE', 'BANKURA', 'BARELI',  
           'BARIPADA', 'BARODA', 'BATHINDA', 'BEAWAR', 'BELGAUM', 'BELLARY',  
           'BENGALURU', 'BHAGALPUR', 'BHARUCH', 'BHILAI', 'BHIWADI', 'BHPAL',  
           'BHUBANESHWAR', 'BHUBANESWAR', 'BHUBNESHWAR', 'BIHAR', 'BIKANER',  
           'BILASPUR', 'BULANDSHAHAR', 'BUNDI', 'BURDWAN', 'CALICUT',  
           'CHANDIGARH', 'CHANDRAPUR', 'CHENNAI', 'CHENNAI & MUMBAI',  
           'CHENNAI, BANGALORE', 'CHEYYAR', 'COIMBATORE', 'DAMAN AND DIU',  
           'DAMMAM', 'DAUSA', 'DEHRADUN', 'DELHI', 'DELHI/NCR', 'DHANBAD',  
           'DHARAMSHALA', 'DHARMAPURI', 'DHARUHERA', 'DUBAI', 'DURGAPUR',  
           'ERNAKULAM', 'FARIDABAD', 'GAGRET', 'GAJIABAAD', 'GANDHI NAGAR',  
           'GANDHINAGAR', 'GANJAM', 'GAZIABAAD', 'GHAZIABAD', 'GONDA',  
           'GORAKHPUR', 'GREATER NOIDA', 'GULBARGA', 'GURAGAON', 'GURGA',  
           'GURGAON', 'GURGOAN', 'GUWAHATI', 'GWALIOR', 'HALDIA', 'HARIDWAR',  
           'HARYANA', 'HDERABAD', 'HISSAR', 'HOSPETE', 'HOWRAH', 'HUBLI',  
           'HYDERABAD', 'HYDERABAD(BHADURPALLY)', 'INDIA',
```

'INDIRAPURAM', 'GHAZIABAD', 'INDORE', 'JABALPUR', 'JAGDALPUR',
'JAIPUR', 'JALANDHAR', 'JAMMU', 'JAMNAGAR', 'JAMSHEDPUR', 'JASPUR',
'JEDDAH SAUDI ARABIA', 'JHAJJAR', 'JHANSI', 'JODHPUR',
'JOHANNESBURG', 'JOSHIMATH', 'JOWAI', 'KAKINADA', 'KALA AMB',
'KALMAR, SWEDEN', 'KANPUR', 'KARAD', 'KARNAL', 'KERAL',
'KHARAGPUR', 'KHOPOLI', 'KOCHI', 'KOCHI/COCHIN',
'KOCHI/COCHIN, CHENNAI AND COIMBATORE', 'KOLHAPUR', 'KOLKATA',
'KOLKATA', 'KOTA', 'KUDANKULAM, TARAPUR', 'KURNOOL',
'LATUR (MAHARASHTRA)', 'LONDON', 'LUCKNOW', 'LUDHIANA', 'MADURAI',
'MAHARAJGANJ', 'MAINPURI', 'MANESAR', 'MANGALORE', 'MEERUT',
'METTUR, TAMIL NADU', 'MIRYALAGUDA', 'MOHALI', 'MUMBAI',
'MUVATTUPUZHA', 'MUZAFFARNAGAR', 'MUZAFFARPUR', 'MUZZAFARPUR',
'MYSORE', 'NAGARI', 'NAGPUR', 'NALAGARH', 'NANDED', 'NASHIK',
'NASIKCITY', 'NAVI MUMBAI', 'NAVI MUMBAI , HYDERABAD', 'NCR',
'NEEMRANA', 'NELLORE', 'NEW DEHLI', 'NEW DELHI',
'NEW DELHI - JAISALMER', 'NOIDA', 'NOUDA', 'ONGOLE', 'ORISSA',
'PANCHKULA', 'PANTNAGAR', 'PATIALA', 'PATNA', 'PHAGWARA', 'PILANI',
'PONDI', 'PONDICHERRY', 'PONDY', 'PUNCHKULA', 'PUNE', 'PUNR',
'RAE BARELI', 'RAIGARH', 'RAIPUR', 'RAJASTHAN', 'RAJKOT',
'RAJPURA', 'RANCHI', 'RAS AL KHAIMAH', 'RATNAGIRI',
'RAYAGADA, ODISHA', 'REWARI', 'ROHTAK', 'ROORKEE', 'ROURKELA',
'RUDRAPUR', 'SADULPUR, RAJGARH, DISTT-CHURU, RAJASTHAN', 'SAHIBABAD',
'SALEM', 'SAMBALPUR', 'SAMPLA', 'SECUNDERABAD', 'SHAHDOL',
'SHAHIBABAD', 'SHIMLA', 'SILIGURI', 'SINGARULI', 'SONEPAT',
'SONIPAT', 'SURAT', 'TECHNOPARK, TRIVANDRUM', 'THANE',
'THIRUVANANTHAPURAM', 'TIRUNELVELLI', 'TIRUPATHI', 'TIRUPATI',
'TORNAGALLU', 'TRICHUR', 'TRICHY', 'TRIVANDRUM', 'UDAIPUR', 'UNA',
'UNKNOWN', 'UNNAO', 'VADODARA', 'VANDAVASI', 'VAPI', 'VARANASI',
'VELLORE', 'VIJAYAWADA', 'VISAKHAPATNAM', 'VIZAG', 'VSAKHAPTTNAM',
'YAMUNA NAGAR'], dtype=object)

City names are not in proper - Replace Names with correct names

```
[32]: jobcity_mapping = {'A-64,SEC-64,NOIDA':'Noida', 'AGRA':'Agra',
    'AHMEDABAD':'Ahmedabad',
    'AHMEDNAGAR':'Ahmednagar',
    'AL JUBAIL,SAUDI ARABIA':'Al Jubail - Saudi Arabia',
    'ALLAHABAD':'Allahabad',
    'ALWAR':'Alwar',
    'AMBALA':'Ambala',
    'AMBALABALA':'Ambala',
    'AMBALABALA CITY':'Ambala',
    'ANGUL':'Angul',
    'ARIYALUR':'Ariyalur',
    'ASANSOL':'Asansol',
    'AURANGABAD':'Aurangabad',
    'AUSTRALIA':'Australia',
    'BADDI HP':'Baddi',
    'BAHADURGARH':'Bahadurgarh',
    'BANGALORE':'Bengaluru',
    'BANKURA':'Bankura',
    'BARELI':'Bareli',
    'BARIPADA':'Baripada',
    'BARODA':'Vadodara',
    'BATHINDA':'Bathinda',
    'BEAWAR':'Beawar',
    'BELGAUM':'Belagavi',
    'BELLARY':'Ballari',
    'BHAGALPUR':'Bhagalpur',
    'BHARUCH':'Bharuch',
    'BHILAI':'Bhilai',
    'BHIWADI':'Bhiwadi',
    'BHOPAL':'Bhopal',
    'BHUBANESWAR':'Bhubaneswar',
    'BIHAR':'Bihar',
    'BIKANER':'Bikaner',
    'BILASPUR':'Bilaspur',
    'BULANDSHAHAR':'Bulandshahr',
    'BUNDI':'Bundi',
    'BURDWAN':'Bardhaman',
    'CALICUT':'Calicut',
    'CHANDIGARH':'Chandigarh',
    'CHANDRAPUR':'Chandrapur',
    'CHENNAI':'Chennai',
    'CHEYYAR':'Cheyyar',
    'COIMBATORE':'Coimbatore',
    'DAMBALAAN AND DIU':'Daman and Diu',
    'DAMBALAMAMBALA':'Ambala',
    'DAUSA':'Dausa',
    'DEHRADUN':'Dehradun',
```

```
'DELHI':'Delhi',
'DELHI - JAISALMER':'Delhi',
'DHANBAD':'Dhanbad',
'DHARAMBALASHALA':'Dharamshala',
'DHARMAPURI':'Dharmapuri',
'DHARUHERA':'Dharuhera',
'DURGAPUR':'Durgapur',
'ERNAKULAMBALA':'Ernakulam',
'FARIDABAD':'Faridabad',
'GAGRET':'Gagret',
'GAJIABAAD':'Ghaziabad',
'GANDHI NAGAR':'Gandhinagar',
'GANJAMBALA':'Ambala',
'GHAZIABAD':'Ghaziabad',
'GONDA':'Gonda',
'GORAKHPUR':'Gorakhpur',
'GULBARGA':'Kalaburagi',
'GURGAON':'Gurugram',
'GURGAONON':'Gurugram',
'GUWAHATI':'Guwahati',
'GWALIOR':'Gwalior',
'HALDIA':'Haldia',
'HARIDWAR':'Haridwar',
'HARYANA':'Haryana',
'HISSAR':'Hisar',
'HOSPETE':'Hosapete',
'HOWRAH':'Howrah',
'HUBLI':'Hubballi',
'HYDERABAD':'Hyderabad',
'INDIA':'India',
'INDIRAPURAMBALA, GHAZIABAD':'Ghaziabad',
'INDORE':'Indore',
'JABALPUR':'Jabalpur',
'JAGDALPUR':'Jagdalpur',
'JAIPUR':'Jaipur',
'JALANDHAR':'Jalandhar',
'JAMBALAMU':'Jamshedpur',
'JAMBALANAGAR':'Jamshedpur',
'JAMBALASHEDPUR':'Jamshedpur',
'JASPUR':'Jaspur',
'JHAJJAR':'Jhajjar',
'JHANSI':'Jhansi',
'JODHPUR':'Jodhpur',
'JOHANNESBURG':'Johannesburg - South Africa',
'JOSHIMATH':'Joshimath',
'JOWAI':'Jowai',
'KAKINADA':'Kakinada',
```

```
'KALA AMBALAB':'Ambala',
    'KALMAR':'Kalmar - southeast Sweden',
'KANPUR':'Kanpur',
'KARAD':'Karad',
'KARNAL':'Karnal',
'KERAL':'Kerala',
'KHARAGPUR':'Kharagpur',
    'KHOPOLI':'Khopoli',
'KOCHI':'Kochi',
'KOCHI, CHENNAI AND COIMBATORE':'Kochi',
'KOLHAPUR':'Kolhapur',
    'KOLKATA':'Kolkata',
'KOTA':'Kota',
'KUDANKULAMBALA , TARAPUR':'Kudankulam',
'KURNOOL':'Kurnool',
'LATUR':'Latur',
'LONDON':'London - UK',
'LUCKNOW':'Lucknow',
'LUDHIANA':'Ludhiana',
'MADURAI':'Madurai',
'MAHARAJGANJ':'Maharajganj',
    'MAINPURI':'Mainpuri',
'MANESAR':'Manesar',
'MANGALORE':'Mangaluru',
'MEERUT':'Meerut',
    'METTUR, TAMBALAIL NADU':'Mettur',
'MIRYALAGUDA':'Miryalaguda',
'MOHALI':'Mohali',
'MUMBAI':'Mumbai',
    'MUMBAI , HYDERABAD':'Mumbai',
'MUVATTUPUZHA':'Muvattupuzha',
'MUZAFFARPUR':'Muzaffarpur',
'MYSORE':'Mysuru',
    'NAGARI':'Nagari',
'NAGPUR':'Nagpur',
'NALAGARH':'Nalagarh',
'NANDED':'Nanded',
'NASHIK':'Nashik',
'NASICKCITY':'Nashik',
    'NCR':'Delhi',
'NEEMRANA':'Neemrana',
'NELLORE':'Nellore',
'NOIDA':'Noida',
'ODISHA':'Odisha',
'ONGOLE':'Ongole',
    'PANCHKULA':'Panchkula',
'PANTNAGAR':'Pantranagar',
```

```
'PATIALA':'Patiala',
'PATNA':'Patna',
'PHAGWARA':'Phagwara',
'PILANI':'Pilani',
    'PONDICHERRY':'Pondicherry',
'PONDICHERRYCHERRY':'Pondicherry',
'PUNCHKULA':'Panchkula',
'PUNE':'Pune',
    'RAE BARELI':'Raebareli',
'RAIGARH':'Raigarh',
'RAIPUR':'Raipur',
'RAJASTHAN':'Rajasthan',
'RAJKOT':'Rajkot',
    'RAJPURA':'Rajpura',
'RANCHI':'Ranchi',
'RAS AL KHAIMAH':'Ras Al-Kaimah - UAE',
'RATNAGIRI':'Ratnagiri',
'REWARI':'Rewari',
    'ROHTAK':'Rohtak',
'ROORKEE':'Roorkee',
'ROURKELA':'Roorkee',
'RUDRAPUR':'Rudrapur',
'SADULPUR':'Sadulpur',
    'SAHIBABAD':'Sahibabad',
'SALEM':'Salem',
'SAMBALABALPUR':'Sambalpur',
'SAMBALAPLA':'Sambalpur',
    'SECUNDERABAD':'Hyderabad',
'SHAHDOL':'Shahdol',
'SAHIBABAD':'Sahibabad',
'SHIMLA':'Shimla',
'SILIGURI':'Siliguri',
    'SINGARULI':'Singrauli',
'SONEPAT':'Sonipat',
'SURAT':'Surat',
'THANE':'Thane',
'THIRUVANANTHAPURAMBALA':'Thiruvananthapuram',
    'TIRUNELVELLI':'Tirunelveli',
'TIRUPATHI':'Tirupati',
'TORNAGALLU':'Toranagallu',
'TRICHUR':'Thrissur',
'TRICHY':'Tiruchirappalli',
    'TRIVANDRUM':'Thiruvananthapuram',
'UDAIPUR':'Udaipur',
'UNNAO':'Unnao',
'VADODARA':'Vadodara',
    'VANDAVASI':'Vandavasi',
```

```

'VAPI':'Vapi',
'VARANASI':'Varanasi',
'VELLORE':'Vellore',
'VIJAYAWADA':'Vijayawada',
'VISAKHAPATNAM':'Visakhapatnam',
'VISAKHAPATNAMBALA':'Visakhapatnam',
'YAMBALAUNNAO NAGAR':'Ambala'}

for old_value, new_value in jobcity_mapping.items():
    data['JobCity'] = data['JobCity'].str.replace(old_value, new_value, ↵
    regex=False)

```

[33]: len(set(jobcity_mapping.values()))

[33]: 170

[34]: data[data['JobCity']=='A-64,SEC-64,NOIDA']

[34]: Empty DataFrame

Columns: [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, openness_to_experience, period, DOB_Year, GYear, 12GradAge, GradAge] Index: []

[35]: data['JobCity'].unique()

[35]: array(['Bengaluru', 'Indore', 'Chennai', 'Gurugram', 'Manesar', 'Hyderabad', 'BANGLORE', 'Noida', 'Kolkata', 'Pune', 'UNKNOWN', 'Mohali', 'Jhansi', 'Delhi', 'Bhubaneswar', 'NAVI Mumbai', 'Mumbai', 'NEW Delhi', 'Mangaluru', 'Rewari', 'GAZIABAAD', 'Bhiwadi', 'Mysuru', 'Rajkot', 'GREATER Noida', 'Jaipur', 'Thane', 'Maharajganj', 'THIRUVANANTHAPURAM', 'Panchkula', 'BHUBANESHWAR', 'Coimbatore', 'Dhanbad', 'Lucknow', 'Thiruvananthapuram', 'Gandhinagar', 'UNA', 'DAMAN AND DIU', 'GURGOAN', 'VSAKHAPTTNAM', 'Nagpur', 'Bhagalpur', 'NEW Delhi - JAISALMER', 'Ahmedabad', 'Kochi/COCHIN', 'Bankura', 'BENGALURU', 'Kanpur', 'Vijayawada', 'Kochi', 'Beawar', 'Alwar', 'Siliguri', 'Raipur', 'Bhopal', 'Faridabad', 'Jodhpur', 'Udaipur', 'Muzaffarpur', 'Kolkata', 'Bulandshahr', 'Haridwar', 'Raigarh', 'VISAKHAPatnaM', 'Jabalpur', 'Unnao', 'Aurangabad', 'Belagavi', 'Dehradun', 'Rudrapur', 'JAMSHEDPUR', 'VIZAG', 'NOUDA', 'DHARAMSHALA', 'BANAGALORE', 'Hisar', 'Ranchi', 'Madurai', 'GURGA', 'Chandigarh', 'Australia', 'Cheyyar', 'Sonipat', 'Ghaziabad', 'Pantnagar', 'Jagdalpur',

```

'Angul', 'Vadodara', 'Ariyalur', 'Jowai',
'Kochi/COCHIN, Chennai AND Coimbatore', 'Neemrana', 'Tirupati',
'BHUBNESHWAR', 'Calicut', 'GANDHINAGAR', 'DUBAI', 'Ahmednagar',
'Nashik', 'Ballari', 'Ludhiana', 'MUZAFFARNAGAR', 'Gagret',
'INDIRAPURAM, Ghaziabad', 'Gwalior', 'Chennai & Mumbai',
'Rajasthan', 'SONIPAT', 'Bareli', 'Hosapete', 'Miryalaguda',
'Dharuhera', 'Meerut', 'GANJAM', 'Hubballi', 'Agra',
'Tiruchirappalli', 'KUDANKULAM ,TARAPUR', 'Ongole', 'SAMBALPUR',
'Pondicherry', 'Bundi', 'Sadulpur,RAJGARH,DISTT-CHURU,Rajasthan',
'AM', 'Bikaner', 'India', 'Asansol', 'Tirunelveli', 'ERNAKULAM',
'Bilaspur', 'Chandrapur', 'Nanded', 'Dharmapuri', 'Vandavasi',
'Rohtak', 'Patna', 'Salem', 'TECHNOPARK, Thiruvananthapuram',
'Bharuch', 'Toranagallu', 'Jaspur', 'Bardhaman', 'Shimla', 'JAMMU',
'Shahdol', 'Muvattupuzha', 'Al Jubail - Saudi Arabia',
'Kalmar - southeast Sweden, SWEDEN', 'Ratnagiri', 'Jhajjar',
'Kalaburagi', 'Hyderabad(BHADURPALLY)', 'Nalagarh',
'JEDDAH SAUDI ARABIA', 'Chennai, Bengaluru', 'JAMNAGAR',
'TIRUPATI', 'Gonda', 'ORISSA', 'Kharagpur',
'NAVI Mumbai , Hyderabad', 'Joshimath', 'Bathinda',
'Johannesburg - South Africa', 'KALA AMB', 'Karnal', 'London - UK',
'Kota', 'Baddi', 'Nagari', 'METTUR, TAMIL NADU', 'Durgapur',
'PONDI', 'Surat', 'Kurnool', 'Kolhapur', 'Bhilai', 'HDERABAD',
'Bahadurgarh', 'RAYAGADA, Odisha', 'Kakinada', 'Varanasi', 'PUNR',
'Nellore', 'Sahibabad', 'Howrah', 'Thrissur', 'Ambala', 'Khopoli',
'Kerala', 'Roorkee', 'Allahabad', 'Delhi/Delhi', 'Jalandhar',
'Vapi', 'Pilani', 'MUZZAFARPUR', 'Ras Al-Khaimah - UAE', 'Bihar',
'Singrauli', 'PONDY', 'Phagwara', 'GURAGAON', 'Baripada',
'YAMUNA NAGAR', 'SAMPLA', 'Guwahati', 'BANAGLORE', 'Vellore',
'Dausa', 'Latur (MAHARASHTRA )', 'Mainpuri', 'DAMMAM', 'Haldia',
'RAE Bareli', 'Patiala', 'Gorakhpur', 'NEW DEHLI', 'Ambala CITY',
'Karad', 'Rajpura', 'Haryana', 'ASIFABADBANGLORE'], dtype=object)

```

[36]:

```

data[data["JobCity"]=="UNKNOWN"][[ "Designation", "12GradAge", "GradAge", "JobCity",
,"Gender", "10pe ↴mode())

```

[36]: Designation 12GradAge GradAge JobCity Gender 10percentage \

0	software engineer	21.0	21.0	UNKNOWN	m	76.0
1		NaN	NaN	NaN	NaN	78.0
2		NaN	NaN	NaN	NaN	82.0

10board 12percentage 12board Degree \

0	cbse	64.0	cbse	B.Tech/B.E.
---	------	------	------	-------------

1	NaN	66.0	NaN	NaN
---	-----	------	-----	-----

2	NaN	NaN	NaN	NaN
---	-----	-----	-----	-----

Specialization CollegeState \

```
0           electronics and communication engineering
           Uttar Pradesh
1             NaN         NaN
2             NaN         NaN
          Specializati
on 0 electronics and communication
engineering
1             NaN
2             NaN
```

```
[37]: data['JobCity'].mode()
```

```
[37]: 0 Bengaluru
Name: JobCity, dtype: object
```

```
[38]: # we can see the mode JobCity Column is 'Bengaluru'
data["JobCity"].replace("UNKNOWN",'Bengaluru',inplace=True)
```

1.6 10board Column

```
[39]: board_10 = data['10board'].unique()
board_10.sort()
```

```
[40]: board_10
```

```
[40]: array(['0', 'aisse', 'andhra pradesh board ssc',
       'andhra pradesh state board', 'anglo indian', 'ap state board',
       'ap state board for secondary education', 'apsche', 'apssc',
       'aurangabad board', 'bharathi matriculation school', 'bihar',
       'bihar board', 'bihar examination board, patna',
       'bihar school examination board',
       'bihar school examination board patna',
       'bihar secondary education board,patna', 'biharboard',
       'board of secondary education',
       'board of high school and intermediate education uttarpradesh',
       'board of intermediate education',
       'board of school education harayana',
       'board of school education haryana',
       'board of school education uttarakhand',
       'board of secendary education orissa',
       'board of secondaray education', 'board of secondary education',
       'board of secondary education (bse) orissa',
       'board of secondary education - andhra pradesh',
       'board of secondary education orissa',
       'board of secondary education rajasthan',
       'board of secondary education(bse) orissa',
       'board of secondary education, andhra prades',
```

'board of secondary education, rajasthan',
'board of secondary education, andhara pradesh',
'board of secondary education, andhra pradesh',
'board of secondary education, ap',
'board of secondary education, odisha',
'board of secondary education, orissa',
'board of secondary education, rajasthan (rbse)',
'board of secondary school education', 'board of ssc',
'board of ssc education andhra pradesh',
'board of secondary education, ap', 'board secondary education',
'bright way college, (up board)', 'bsc, orissa', 'bse',
'bse (board of secondary education)', 'bse, odisha', 'bse, odisha',
'bse, orissa', 'bseb', 'bseb , patna', 'bseb patna', 'bseb, patna',
'bseb, patna', 'bsemp', 'bsepatna', 'cbse', 'cbse', 'cbse ',
'cbse board', 'cbse [gulf zone]',
'central board of secondary education',
'central board of secondary education, new delhi',
'certificate of middle years program of ib', 'cgbse',
'cgbse raipur', 'cicse', 'cluny',
'council for indian school certificate examination',
'dav public school sec 14', 'dav public school, hehal',
'delhi board', 'delhi public school',
'don bosco maatriculation school', 'education board of kerala',
'ghseb', 'gseb', 'gsheb', 'gujarat board', 'gujarat state board',
'gyan bharati school', 'haryana board of school education',
'haryana board of school education, (hbse)', 'hbse', 'hbse',
'himachal pradesh board',
'himachal pradesh board of school education',
'holy cross matriculation hr sec school', 'hsc', 'hsce', 'hse',
'hse, board', 'hse, orissa', 'icse', 'icse board',
'icse board , new delhi', 'j & k bord',
'j&k state board of school education',
'jawahar navodaya vidyalaya', 'jbse, jharkhand',
'jharkhand academic council', 'jharkhand accademic council',
'jharkhand acedemic council', 'jharkhand secondary board',
'jharkhand secondary education board',
'jharkhand secondary examination board (ranchi)',
'jharkhand secondary examination board, ranchi', 'jkbose', 'jseb',
'kalaimagal matriculation higher secondary school',
'karantaka secondary education and examination borad', 'karnataka',
'karnataka board', 'karnataka board of higher education',
'karnataka board of secondary education',
'karnataka education board', 'karnataka education board (keeb)',
'karnataka secondary board', 'karnataka secondary education',
'karnataka secondary education board',

'karnataka secondary education examination board',
'karnataka secondary eduction',
'karnataka secondary school of examination',
'karnataka secondary education board',
'karnataka sslc board bangalore', 'karnataka state board',
'karnataka state education examination board',
'karnataka state examination board',
'karnataka state secondary education board', 'kea', 'kerala',
'kerala state board', 'kerala state technical education',
'kerala university', 'kiran english medium high school',
'kolhapur', 'kolhapur divisional board, maharashtra', 'ksbe',
'kseb', 'kseeb',
'kseeb(karnataka secondary education examination board)', 'ksseb',
'ksseb(karnataka state board)', 'latur', 'latur board',
'little jacky matric higher secondary school',
'madhya pradesh board', 'maharashtra', 'maharashtra board',
'maharashtra board, pune', 'maharashtra nasik board',
'maharashtra satate board', 'maharashtra sate board',
'maharashtra state boar of secondary and higher secondary education',
'maharashtra state board', 'maharashtra state board for ssc',
'maharashtra state board mumbai divisional board',
'maharashtra state board of secondary & higher secondary education',
'maharashtra state board of secondary and higher secondary education',
'maharashtra state board of secondary and higher secondary
education,pune',
'maharashtra state board pune', 'maharashtra state board,pune',
'maharashtra state(latur board)', 'maharastra board',
'maticulation', 'matric', 'matric board', 'matriculation',
'matriculation board', 'metric', 'mhsbse', 'mirza ahmed ali baig',
'mp', 'mp board', 'mp board bhopal', 'mp state board', 'mp-bse',
'mpboard', 'mpbse', 'ms board', 'msbshse,pune', 'mumbai board',
'nagpur', 'nagpur board', 'nagpur board,nagpur',
'nagpur divisional board', 'nashik board', 'nasik',
'national public school', 'pseb', 'pune', 'pune board',
'punjab school education board, mohali', 'rajasthan board',
'rajasthan board ajmer', 'rajasthan board of secondary education',
'rbse', 'rbse (state board)', 'rbse,ajmer', 'sarada high scchool',
'school secondary education, andhra pradesh', 'seba',
'seba(assam)', 'secondary board of rajasthan',
'secondary education board of rajasthan',
'secondary school cerfificate', 'secondary school certificate',
'secondary school education', 'secondary school of education',
'secondary state certificate',
'sri kannika parameswari highier secondary school, udumalpet',
'ssc', 'ssc board of andrapradesh', 'ssc maharashtra board',

'ssc regular', 'ssc-andhra pradesh', 'sslc', 'sslc board',
'sslc,karnataka', 'sss pune', 'state', 'state board',
'state board ', 'state board (jac, ranchi)',
'state board - west bengal board of secondary education : wbbse',
'state board of karnataka',
'state board of secondary education(ssc)' ,

```

'state board of secondary education, andhra pradesh',
'state board of secondary education, ap',
'state board of secondary education, andhra pradesh',
'state boardmp board ', 'state borad hp', 'state bord',
'state(karnataka board)', 'stateboard',
'stjoseph of cluny matrhrsecsschool,neyveli,cuddalore
district',
'stjosephs girls higher sec school,dindigul',
'stmary higher secondary', "stmary's convent inter college",
'tamil nadu state', 'tamilnadu matriculation board',
'tamilnadu state board', 'tn state board', 'u p', 'u p board',
'ua', 'up', 'up baord', 'up board', 'up board , allahabad',
'up board allahabad', 'up board,allahabad', 'up borad', 'up
bord',
'up bound', 'up(allahabad)', 'up-board', 'upbhsie',
'upboard',
'uttar pradesh', 'uttar pradesh board', 'uttarakhand board',
'uttaranchal shiksha avam pariksha parishad',
'uttaranchal state board', 'uttrakhand board', 'uttranchal
board',
'wbbsce', 'wbbse', 'west bengal board of secondary education',
'west bengal board of secondary education',
'west bengal board of secondary eucation',
'west bengal board of secondary examination (wbbse)'],
dtype=object)

[41]: state_board = ['aisse', 'andhra pradesh board ssc',
'andhra pradesh state board', 'anglo indian', 'ap state
board',
'ap state board for secondary education', 'apsche', 'apssc',
'aurangabad board', 'bharathi matriculation school',
'bihar',
'bihar board', 'bihar examination board, patna',
'bihar school examination board',
'bihar school examination board patna',
'bihar secondary education board,patna', 'biharboard',
'board of secondary education',
'board of high school and intermediate education
uttarpradesh',
'
'
'
'
'
'
```

'board of intermediate education',
'board of school education harayana',
'board of school education haryana',
'board of school education uttarakhand',
'board of secendary education orissa',
'board of secondaray education', 'board of secondary
education',
'board of secondary education (bse) orissa',
'board of secondary education - andhra pradesh',
'board of secondary education orissa',
'board of secondary education rajasthan',
'board of secondary education(bse)
orissa', board of secondary education,
andhra pradesh', board of secondary
education, rajasthan', board of
secondary education, andhara pradesh',
board of secondary education, andhra
pradesh', board of secondary
education,ap',
'board of secondary education,odisha',
'board of secondary education,orissa',
'board of secondary education,rajasthan(rbse)',
'board of secondary school education', 'board of ssc',
'board of ssc education andhra prades',
'board of seconday education,ap', 'board secondary
education',
'bright way college, (up board)', 'bsc,orissa', 'bse',
'bse(board of secondary education)', 'bse, odisha',
'bse,odisha',
'bse,orissa', 'bseb', 'bseb ,patna', 'bseb patna', 'bseb,
patna',
'bseb,patna', 'bsemp', 'bsepatna',
'certificate of middle years program of ib', 'cgbse',
'cgbse raipur', 'cluny',
'council for indian school certificate examination',
'dav public school sec 14', 'dav public school,hehal',
'delhi board', 'delhi public school',
'don bosco maatriculation school', 'education board of
kerala',
'
'
'
'
'

'ghseb', 'gseb', 'gsheb', 'gujarat board', 'gujarat state board',
'gyan bharati school', 'haryana board of school education',
'haryana board of school education, (hbse)', 'hbse', 'hbse',
'himachal pradesh board',
'himachal pradesh board of school education',
'holy cross matriculation hr sec school', 'hsc', 'hsce',
'hse',
'hse,board', 'hse,orissa', 'j & k bord',
'j&k state board of school education',
'jawahar navodaya vidyalaya', 'jbse,jharkhand',
'jharkhand academic council', 'jharkhand accademic council',
'jharkhand acedemic council', 'jharkhand secondary board',
'jharkhand secondary education board',
'jharkhand secondary examination board (ranchi)',
'jharkhand secondary examination board,ranchi', 'jkbose',
'jseb',
'kalaimagal matriculation higher secondary school',
'karantaka secondary education and examination borad',
'karnataka',
'karnataka board', 'karnataka board of higher education',
'karnataka board of secondary education',
'karnataka education board', 'karnataka education board
(keeb)',
'karnataka secondary board', 'karnataka secondary
education',
'karnataka secondary education board',
'karnataka secondary education examination board',
'karnataka secondary eduction',
'karnataka secondary school of examination',
'karnataka secondary education board',
'karnataka sslc board bangalore', 'karnataka state
board', karnataka state education examination
board', karnataka state examination board',
karnataka state secondary education board', 'kea',
'kerala', kerala state board', 'kerala state
technical education', kerala university', 'kiran
english medium high school', 'kolhapur', 'kolhapur
divisional board, maharashtra', 'ksbe',
'
'
'
'
'

'kseb', 'kseeb',
'kseeb(karnataka secondary education examination board)',
'ksseb',
'ksseb(karnataka state board)', 'latur', 'latur board',
'little jacky matric higher secondary school',
'madhya pradesh board', 'maharashtra', 'maharashtra board',
'maharashtra board, pune', 'maharashtra nasik board',
'maharashtra satate board', 'maharashtra sate board',
'maharashtra state boar of secondary and higher secondary
education',
'maharashtra state board', 'maharashtra state board for
ssc',
'maharashtra state board mumbai divisional board',
'maharashtra state board of secondary & higher secondary
education',
'maharashtra state board of secondary and higher secondary
education',
'maharashtra state board of secondary and higher secondary
education,pune',
'maharashtra state board pune', 'maharashtra state
board,pune',
'maharashtra state(latur board)', 'maharastra board',
'maticulation', 'matric', 'matric board', 'matriculation',
'matriculation board', 'metric', 'mhsbse', 'mirza ahmed ali
baig',
'mp', 'mp board', 'mp board bhopal', 'mp state board', 'mp-
bse',
'mpboard', 'mpbse', 'ms board', 'msbshse,pune', 'mumbai
board',
'nagpur', 'nagpur board', 'nagpur board,nagpur',
'nagpur divisional board', 'nashik board', 'nasik',
'national public school', 'pseb', 'pune', 'pune board',
'punjab school education board, mohali', 'rajasthan board',
'rajasthan board ajmer', 'rajasthan board of secondary
education',
'rbse', 'rbse (state board)', 'rbse,ajmer', 'sarada high
scchool',
'school secondary education, andhra pradesh', 'seba',
'seba(assam)', 'secondary board of rajasthan',
'
'
'
'
'

'secondary education board of rajasthan',
'secondary school cerfificate', 'secondary school
certificate',
'secondary school education', 'secondary school of
education',
'secondary state certificate',
'sri kannika parameswari highier secondary school,
udumalpet',
'ssc', 'ssc board of andrapradesh', 'ssc maharashtra board',
'ssc regular', 'ssc-andhra pradesh', 'sslc', 'sslc board',
'sslc,karnataka', 'sss pune', 'state', 'state board',
'state board ', 'state board (jac, ranchi)',
'state board - west bengal board of secondary education :
wbbse',
'state board of karnataka',
'state board of secondary education(ssc)',
'state board of secondary education, andhra
pradesh', state board of secondary education,
ap', state board of secondary education, andhra
pradesh', state boardmp board ', 'state borad
hp', 'state bord',
'state(karnataka board)', 'stateboard',
stjoseph of cluny matrhrsecschool,neyveli,cuddalore district',

'
'
'
'
'

```

'stjosephs girls higher sec school,dindigul',
'stmary higher secondary', "stmary's convent inter
college",
'tamil nadu state', 'tamilnadu matriculation board',
'tamilnadu state board', 'tn state board', 'u p', 'u p
board',
'ua', 'up', 'up baord', 'up board', 'up board , allahabad',
'up board allahabad', 'up board,allahabad', 'up borad', 'up
bord',
'up bound', 'up(allahabad)', 'up-board', 'upbhsie',
'upboard',
'uttar pradesh', 'uttar pradesh board', 'uttarakhand
board',
'uttaranchal shiksha avam pariksha parishad',
'uttaranchal state board', 'uttrakhand board', 'uttranchal
board',
'wbbsce', 'wbbse', 'west bengal board of secondary
education',
'west bengal board of secondary education',
'west bengal board of secondary eucation',
'west bengal board of secondary examination (wbbse)']

cbse_board = ['cbsc', 'cbse', 'cbse ',
    'cbse board', 'cbse[gulf zone]',
    'central board of secondary education',
    'central board of secondary education, new delhi']

icse_board = ['icse', 'icse board', 'icse',
    'icse board , new delhi', 'cicse', 'icse board', 'new
delhi']

[42]: # Replace 10board column with 'State Board', 'Central Board',
'ICSE Board'

    for i in state_board: data['10board'].replace(i,
'State Board', inplace=True)

[43]:     for j in cbse_board:
    data['10board'].replace(j, 'CBSE',
inplace=True)

[44]:     for k in icse_board:
    data['10board'].replace(k, 'ICSE',
inplace=True)

```

```
[45]: data['10board'].unique()

[45]: array(['State Board', 'CBSE', 'ICSE', '0'], dtype=object)
[46]: data['10board'].mode()[0]

[46]: 'State Board'
[47]: data['10board'].replace('0', data['10board'].mode()[0], inplace=True)

[48]: data['10board'].unique()

[48]: array(['State Board', 'CBSE', 'ICSE'], dtype=object)
```

1.7 12board Column

```
[49]: data['12board'] = data['12board'].apply(lambda x: x.strip())

[50]: board_12 = data['12board'].unique()
      board_12.sort()

[51]: board_12

[51]: array(['0', 'ahsec', 'aissce', 'aligarh muslim university',
           'all india board', 'amravati divisional board',
           'andhra pradesh board of intermediate education', 'andhra
           board',
           'andhra pradesh', 'andhra pradesh board of secondary
           education',
           'andhra pradesh state board', 'ap board',
           'ap board for intermediate education', 'ap intermediate
           board',
           'apbie', 'apbsc', 'apsb', 'aurangabad board',
           'baord of intermediate education', 'bciec,patna', 'bice',
           'bie',
           'bieap', 'biec', 'biec patna', 'biec, patna', 'biec,patna',
           'biec-patna', 'bihar', 'bihar board',
           'bihar intermediate education council',
           'bihar intermediate education council, patna',
           'bihar school examination board patna',
           'board fo intermediate education, ap',
           'board of high school and intermediate education
           uttarpradesh',
           'board of higher secondary examination, kerala',
           'board of higher secondary orissa',
           'board of intermeadiate education', 'board of intermediate',
           'board of intermediate ap', 'board of intermediate
           education',
```

'board of intermediate education, andhra pradesh',
'board of intermediate education, ap',
'board of intermediate education, andhra pradesh',
'board of intermediate education, andra pradesh',
'board of intermediate education, ap',
'board of intermediate education, hyderabad',
'board of intermediate education:ap, hyderabad',
'board of intermediate(bie)', 'board of intermediate, ap',
'board of intermidiate', 'board of intermidiate
education, ap',
'board of intermidiate examination',
'board of intmediate education ap',
'board of intrmediate education, ap',
'board of school education harayana',
'board of school education uttarakhand',
'board of secondary education',
'board of secondary education rajasthan',
'board of secondary education, rajasthan',
'board of secondary education, rajasthan(rbse)',
'board of secondary school of education',
'board of technical education',
'board of technical education ,delhi', 'boardofintermediate',
'borad of intermediate', 'bright way college, (up board)',
'bseb',
'bseb, patna', 'bsemp', 'bte up', 'bte,delhi', 'bteup',
'cbese',
'cbsc', 'cbse', 'cbse board', 'cbse,new delhi',
'central board of secondary education',
'central board of secondary education, new delhi',
'certificate for higher secondary education (chse)orissa',
'cgbse',
'cgbse raipur', 'chsc', 'chse',
'chse(concil of higher secondary education)', 'chse, odisha',
'chse,odisha', 'chse,orissa', 'cicse',
'council for indian school certificate examination',
'dav public school', 'dav public school sec 14',
'dav public school,hehal',
'department of pre university education',
'department of pre-university education',
'department of pre-university education(government of
karnataka)',
'department of pre-university education, bangalore',
'department of pre-university eduction',
'department of technical education',

'department of technical education, bangalore',
'dept of pre-university education',
'diploma (maharashtra state board of technical education)',
'diploma in computers',
'diploma in engg (e &tc) tilak maharashtra vidyapeeth',
'diploma(msbte)', 'directorate of technical
education,banglore',
'dote (diploma - computer engg)', 'dpue', 'dte',
'electonincs and communication(dote)', 'ghseb',
'government polytechnic mumbai , mumbai board',
'govt of karnataka department of pre-university education',
'gseb',
'gseb/technical education board', 'gsheb', 'gshseb',
'gujarat board',
'haryana state board of technical education chandigarh',
'hbsc',
'hbse', 'higher secondary', 'higher secondary education',
'higher secondary state certificate', 'himachal pradesh
board',
'himachal pradesh board of school education',
'hisher seconadry examination(state board)',
'holy cross matriculation hr sec school', 'hsc',
'hsc maharashtra board', 'hsc pune', 'hse',
'hslc (tamil nadu state board)', 'ibe', 'icse',
'intermediate',
'intermediate board', 'intermediate board examination',
'intermediate board of andhra pardesh',
'intermediate board of education',
'intermediate board of education, andhra pradesh',
'intermediate council patna', 'intermediate state board',
'intermedite', 'intermideate', 'intermidiate',
'international baccalaureate (ib) diploma', 'ipe', 'isc',
'isc board', 'isc board , new delhi', 'isce', 'j & k board',
'j&k state board of school education', 'jaswant modern
school',
'jawahar higher secondary school', 'jaycee matriculation
school',
'jharkhand academic council',
'jharkhand acamedic council (ranchi)',
'jharkhand accademic council', 'jharkhand acedemic council',
'jharkhand board', 'jiec', 'jkbose', 'jstb,jharkhand',
'jyoti nivas', 'karanataka secondary board', 'karnatak pu
board',
'karnataka board', 'karnataka board of univeristy',

'karnataka board secondary education', 'karnataka education board',
'karnataka pre university board', 'karnataka pre university board',
'karnataka pre-university', 'karnataka pre-university board',
'karnataka pu', 'karnataka pu board',
'karnataka secondary education',
'karnataka secondary education board', 'karnataka sslc',
'karnataka state', 'karnataka state board',
'karnataka state examination board',
'karnataka state pre- university board', 'kea', 'kerala',
'kerala state board', 'kerala state hse board',
'kerala university', 'kolhapur',
'kolhapur divisional board, maharashtra',
'ks rangasamy institute of technology', 'latur', 'latur board',
'lucknow public college', 'madhya pradesh board',
'madhya pradesh open school', 'maharashtra', 'maharashtra board',
'maharashtra board, pune', 'maharashtra nasik board',
'maharashtra satate board',
'maharashtra state boar of secondary and higher secondary education',
'maharashtra state board', 'maharashtra state board for hsc',
'maharashtra state board mumbai divisional board',
'maharashtra state board of secondary & higher secondary education',
'maharashtra state(latur board)', 'matric board',
'matriculation',
'mbose', 'mp', 'mp board', 'mp board bhopal', 'mp-bse',
'mpboard',
'mpbse', 'mpc', 'ms board', 'msbshse,pune', 'msbte',
'msbte (diploma in computer technology)', 'msbte pune',
'nagpur',
'nagpur board', 'nagpur board,nagpur', 'nagpur divisional board',
'narayana junior college', 'nashik board', 'nasik', 'nios',
'p u board, karnataka', 'pre university', 'pre university board',
'pre university board of karnataka',
'pre university board, karnataka', 'pre-university',
'pre-university board', 'preuniversity board(karnataka)',
'psbte',

'pseb', 'pu', 'pu board karnataka', 'pu board',
'pu board ,karnataka', 'pu board karnataka', 'pub',
'puboard',
'puc', 'pue', 'pune board',
'punjab state board of technical education & industrial
training',
'punjab state board of technical education & industrial
training, chandigarh',
'rajasthan board', 'rajasthan board ajmer',

'rajasthan board of secondary education', 'rbse',
'rbse (state board)', 's j polytechnic', 'sbte, jharkhand',
'sbtet', 'science college', 'scte & vt (diploma)',
'scte and vt ,orissa', 'scte vt orissa', 'scte&vt',
'sda matric higher secondary school',
'secnior secondary education board of rajasthan',
'secondary board of rajasthan', 'sjrcw',
'sri chaitanya junior kalasala',
'sri kannika parameswari highier secondary school,
udumalpet',
'sri sankara vidyalaya', 'srv girls higher sec
school,rasipuram',
'ssc', 'ssm srsecschool', 'st joseph hr sec school', 'staae
board',
'state', 'state board of intermediate education, andhra
pradesh',
'state board', 'state board (jac, ranchi)',
'state board - tamilnadu',
'state board - west bengal council of higher secondary
education : wbchse',
'state board of karnataka', 'state board of technical
education',
'state board of technical education and training',
'state board of technical education harayana',
'state board of technical eduction panchkula', 'state
bord',
'state broad', 'state syllabus', 'stateboard',
'stateboard/tamil nadu',
'stjoseph of cluny matrhrsecschool,neyveli,cuddalore
district',
'stmiras college for girls', 'tamil nadu polytechnic',
'tamil nadu state', 'tamil nadu state board',
'tamilnadu higher secondary education board',
'tamilnadu state board', 'tamilnadu stateboard',
'technical board, punchkula', 'tn state board', 'u p', 'u
p board', 'ua', 'uo board', 'up', 'up baord', 'up board',
'up board , allahabad', 'up board allahabad', 'up
board,allahabad',

'
'
'
'
'

```
'up bord', 'up bourd', 'up(allahabad)', 'up-board',
'upbhsie',
'upboard', 'uttar pradesh', 'uttar pradesh board',
'uttarakhand board', 'uttaranchal shiksha avam pariksha
parishad',
'uttaranchal state board', 'uttrakhand board', 'uttranchal
board',
'wbbhse', 'wbchse', 'wbscte',
'west bengal board of higher secondary education',
'west bengal council of higher secondary education',
'west bengal council of higher secondary eucation',
'west bengal council of higher secondary examination
(wbchse)',
'west bengal state council of technical education'],
dtype=object)
```

```
[52]: state_board_12 = ['ahsec', 'aissce', 'aligarh muslim
university', 'all india board', 'amravati divisional
board', 'andhpрадеш board of intermediate education',
'andhra board', 'andhra pradesh', 'andhra pradesh board
of secondary education', 'andhra pradesh state board',
'ap board',
'ap board for intermediate education', 'ap intermediate
board',
'apbie', 'apbsc', 'apsb', 'aurangabad board',
'baord of intermediate education', 'bciec,patna', 'bice',
'bie',
'bieap', 'biec', 'biec patna', 'biec, patna',
'biec,patna',
'biec-patna', 'bihar', 'bihar board',
'bihar intermediate education council',
'bihar intermediate education council, patna',
'bihar school examination board patna',
'board fo intermediate education, ap',
'board of high school and intermediate education
uttarpradesh',
'board of higher secondary examination, kerala',
'board of higher secondary orissa',
'board of intermeadiate education', 'board of
intermediate',
'
'
'
'
```

'board of intermediate ap', 'board of intermediate education',
'board of intermediate education, andhra pradesh',
'board of intermediate education, ap',
'board of intermediate education, andhra pradesh',
'board of intermediate education, andra pradesh',
'board of intermediate education, ap',
'board of intermediate education, hyderabad',
'board of intermediate education:ap, hyderabad',
'board of intermediate(bie)', 'board of intermediate, ap',
'board of intermidiate', 'board of intermidiate education, ap',
'board of intermidiate examination',
'board of intmediate education ap',
'board of intrmediate education, ap',
'board of school education harayana',
'board of school education uttarakhand',
'board of secondary education',
'board of secondary education rajasthan',
'board of secondary education, rajasthan',
'board of secondary education, rajasthan(rbse)',
'board of secondary school of education',
'board of technical education',
'board of technicaleducation ,delhi',
'boardofintermediate',
'borad of intermediate', 'bright way college, (up board)',
'bseb',
'bseb, patna', 'bsemp', 'bte up', 'bte,delhi', 'bteup',
'certificate for higher secondary education (chse)orissa',
'cgbse',
'cgbse raipur', 'chsc', 'chse',
'chse(concil of higher secondary education)', 'chse, odisha',
'chse,odisha', 'chse,orissa', 'cicse',
'council for indian school certificate examination',
'dav public school', 'dav public school sec 14', dav
public school,hehal', 'department of pre university
education', 'department of pre-university education',
'department of pre-university education(government of

',
,

karnataka)', department of pre-university education, bangalore',
'department of pre-university eduction',
'department of technical education',
'department of technical education, bangalore',
'dept of pre-university education',
'diploma (maharashtra state board of technical education)',
'diploma in computers',
'diploma in engg (e &tc) tilak maharashtra vidayapeeth',
'diploma(msbte)', 'directorate of technical education,banglore',
'dote (diploma - computer engg)', 'dpue', 'dte',
'electonincs and communication(dote)', 'ghseb',
'government polytechnic mumbai , mumbai board',
'govt of karnataka department of pre-university education', 'gseb',
'gseb/technical education board', 'gsheb', 'gshseb',
'gujarat board',
'haryana state board of technical education chandigarh',
'hbsc',
'hbse', 'higher secondary', 'higher secondary education',
'higher secondary state certificate', 'himachal pradesh board',
'himachal pradesh board of school education',
'hisher seconadry examination(state board)',
'holy cross matriculation hr sec school', 'hsc',
'hsc maharashtra board', 'hsc pune', 'hse',
'hslc (tamil nadu state board)', 'ibe', 'intermediate',
'intermediate board', 'intermediate board examination',
'intermediate board of andhra pardesh',
'intermediate board of education',
'intermediate board of education, andhra pradesh',
'intermediate council patna', 'intermediate state board',
'intermedite', 'intermideate', 'intermidiate',
'international baccalaureate (ib) diploma', 'j & k board',
'j&k state board of school education', 'jaswant modern school',

'
'
'
'
'

'jawahar higher secondary school', 'jaycee matriculation school',
'jharkhand academic council',
'jharkhand acamedic council (ranchi)',
'jharkhand accademic council', 'jharkhand acedemic council',
'jharkhand board', 'jiec', 'jkbose', 'jstb,jharkhand',
'jyoti nivas', 'karanataka secondary board', 'karnatak pu board',
'karnataka board', 'karnataka board of university',
'karnataka board secondary education', 'karnataka education board',
'karnataka pre university board', 'karnataka pre unversity board',
'karnataka pre-university', 'karnataka pre-university board',
'karnataka pu', 'karnataka pu board',
'karnataka secondary education', karnataka secondary education board', 'karnataka sslc',
karnataka state', 'karnataka state board',
karnataka state examination board', karnataka state pre- university board', 'kea', 'kerala',
kerala state board', 'kerala state hse board',

'
'
'
'
'

'kerala university', 'kolhapur',
'kolhapur divisional board, maharashtra',
'ks rangasamy institute of technology', 'latur', 'latur board',
'lucknow public college', 'madhya pradesh board',
'madhya pradesh open school', 'maharashtra', 'maharashtra
board',
'maharashtra board, pune', 'maharashtra nasik board',
'maharashtra satate board',
'maharashtra state boar of secondary and higher secondary
education',
'maharashtra state board', 'maharashtra state board for hsc',
'maharashtra state board mumbai divisional board',
'maharashtra state board of secondary & higher secondary
education',
'maharashtra state(latur board)', 'matric board',
'matriculation',
'mbose', 'mp', 'mp board', 'mp board bhopal', 'mp-bse',
'mpboard',
'mpbse', 'mpc', 'ms board', 'msbshse,pune', 'msbte',
'msbte (diploma in computer technology)', 'msbte pune',
'nagpur',
'nagpur board', 'nagpur board,nagpur', 'nagpur divisional
board',
'narayana junior college', 'nashik board', 'nasik', 'nios',
'p u board, karnataka', 'pre university', 'pre university
board',
'pre university board of karnataka',
'pre university board, karnataka', 'pre-university',
'pre-university board', 'preuniversity board(karnataka)',
'psbte',
'pseb', 'pu', 'pu board karnataka', 'pu board',
'pu board ,karnataka', 'pu board karnataka', 'pub', 'puboard',
'puc', 'pue', 'pune board',
'punjab state board of technical education & industrial
training',
'punjab state board of technical education & industrial
training',
↳chandigarh',
'rajasthan board', 'rajasthan board ajmer',
'rajasthan board of secondary education', 'rbse',
'rbse (state board)', 's j polytechnic', 'sbte, jharkhand',
'sbtet', 'science college', 'scte & vt (diploma)',
'
'
'

'scte and vt ,orissa', 'scte vt orissa', 'scte&vt',
'sda matric higher secondary school',
'secnior secondary education board of rajasthan',
'secondary board of rajasthan', 'sjrcw',
'sri chaitanya junior kalasala',
'sri kannika parameswari highier secondary school, udumalpet',
'sri sankara vidyalaya', 'srv girls higher sec
school,rasipuram',
'ssc', 'ssm srsecschool', 'st joseph hr sec school', 'staae
board',
'state', 'state board of intermediate education, andhra
pradesh',
'state board', 'state board (jac, ranchi)',
'state board - tamilnadu',
'state board - west bengal council of higher secondary education
:_
↳wbchse', state board of karnataka', 'state board of
technical education', state board of technical
education and training', state board of technical
education harayana',
'state board of technical eduction panchkula', 'state bord',
'state broad', 'state syllabus', 'stateboard',
'stateboard/tamil nadu',
'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district',
'stmiras college for girls', 'tamil nadu polytechnic',
'tamil nadu state', 'tamil nadu state board',
'tamilnadu higher secondary education board',
'tamilnadu state board', 'tamilnadu stateboard',
'technical board, punchkula', 'tn state board', 'u p', 'u p
board', 'ua', 'uo board', 'up', 'up baord', 'up board',
'up board , allahabad', 'up board allahabad', 'up
board,allahabad',
'up bord', 'up bound', 'up(allahabad)', 'up-board', 'upbhsie',
'upboard', 'uttar pradesh', 'uttar pradesh board',
'uttarakhand board', 'uttaranchal shiksha avam pariksha
parishad',
'uttaranchal state board', 'uttrakhand board', 'uttranchal
board',
'wbbhse', 'wbchse', 'wbscte',
'west bengal board of higher secondary education',
'west bengal council of higher secondary education',
'west bengal council of higher secondary eucation',
'west bengal council of higher secondary examination
(wbchse)', 'west bengal state council of technical
education']

```

cbse_board_12 = [ 'cbese',
                  'cbsc', 'cbse', 'cbse board', 'cbse,new delhi',
                  'central board of secondary education',
                  'central board of secondary education, new delhi']

icse_board_12 = [ 'icse', 'ipe', 'isc',
                  'isc board', 'isc board , new delhi', 'isce',
                  'cicse']

[53]: # Replace 12board column with 'State Board', 'Central Board', 'ICSE
Board'

for i in state_board_12: data['12board'].replace(i,
                                                 'State Board', inplace=True)

[54]: for j in cbse_board_12:
       data['12board'].replace(j, 'CBSE',
                               inplace=True)

[55]: for k in icse_board_12:
       data['12board'].replace(k, 'ICSE',
                               inplace=True)

[56]: data['12board'].unique()

[56]: array(['State Board', 'CBSE', 'ICSE', '0'], dtype=object)
[57]: data['12board'].mode()[0]
[57]: 'State Board'

[58]: data['12board'] = data['12board'].replace('0',
                                                data['12board'].mode()[0])

```

1.8 Specialization Column

```

[59]: data['Specialization'].unique()

[59]: array(['computer engineering',
            'electronics and communication engineering',
            'information technology', 'computer science & engineering',
            'mechanical engineering', 'electronics and electrical
            engineering',
            'electronics & telecommunications',
            'instrumentation and control engineering', 'computer
            application',
            'electronics and computer engineering', 'electrical engineering'],

```

```

'applied electronics and instrumentation',
'electronics & instrumentation eng',
'information science engineering', 'civil engineering',
'mechanical and automation', 'industrial & production
engineering',
'control and instrumentation engineering',
'metallurgical engineering',
'electronics and instrumentation engineering',
'electronics engineering', 'ceramic engineering',
'chemical engineering', 'aeronautical engineering', 'other',
'biotechnology', 'embedded systems technology',
'electrical and power engineering',
'computer science and technology', 'mechatronics',
'automobile/automotive engineering', 'polymer technology',
'mechanical & production engineering',
'power systems and automation', 'instrumentation engineering',
'telecommunication engineering',
'industrial & management engineering', 'industrial engineering',
'computer and communication engineering',
'information & communication technology', 'information science',
'internal combustion engine', 'computer networking',
'biomedical engineering', 'electronics', 'computer science'],
dtype=object)

```

```

[60]: specialization_mapping = {'electronics and communication engineering':
: 'ECE',
'computer science & engineering' : 'CSE',
'information technology' : 'CSE',
'computer engineering' : 'CSE',
'computer application' : 'CSE',
'mechanical engineering' : 'MECH',
'electronics and electrical engineering' : 'ECE',
'electronics & telecommunications' : 'ECE',
'electrical engineering' : 'EEE',
'electronics & instrumentation eng' : 'ECE',
'civil engineering' : 'CE',
'electronics and instrumentation engineering' : 'ECE',
'information science engineering' : 'CSE',
'instrumentation and control engineering' : 'ECE',
'electronics engineering' : 'ECE',
'biotechnology' : 'other',
'other' : 'other',
'industrial & production engineering' : 'other',
'chemical engineering' : 'other',
'applied electronics and instrumentation' : 'ECE',
}

```

```

'computer science and technology' : 'CSE',
'telecommunication engineering' : 'ECE',
'mechanical and automation' : 'MECH',
'automobile/automotive engineering' : 'MECH',
'instrumentation engineering' : 'ECE',
'mechatronics' : 'MECH',
'electronics and computer engineering' : 'CSE',
'aeronautical engineering' : 'MECH',
'computer science' : 'CSE',
'metallurgical engineering' : 'other',
'biomedical engineering' : 'other',
'industrial engineering' : 'other',
'information & communication technology' : 'ECE',
'electrical and power engineering' : 'EEE',
'industrial & management engineering' : 'other',
'computer networking' : 'CSE',
'embedded systems technology' : 'ECE',
'power systems and automation' : 'EEE',
'computer and communication engineering' : 'CSE',
'information science' : 'CSE',
'internal combustion engine' : 'MECH',
'ceramic engineering' : 'other',
'mechanical & production engineering' : 'MECH',
'control and instrumentation engineering' : 'ECE',
'polymer technology' : 'other',
'electronics' : 'ECE'}

for old, new in specialization_mapping.items():
    data['Specialization'] = data['Specialization'].replace(old, new)

```

```
[61]: data['Specialization'].unique()
```

```
[61]: array(['CSE', 'ECE', 'MECH', 'EEE', 'CE', 'other'], dtype=object)
```

1.9 'Domain' Column

- Domain Column having positive values between 0 to 1, But, -1 seems to be irrelevant or anomaly is considered as Missing value
- **Impute -1 with Median of the Domain column**

```
[62]: data['Domain'].unique()
```

```
[62]: array([ 0.63597876, 0.96060325, 0.45087658, 0.97439611, 0.12450207,
           -1.        , 0.35653649, 0.8295846 , 0.69447933, 0.49359639,
           0.76567358, 0.9682375 , 0.22948175, 0.53838689, 0.30840058,
```

0.91139528, 0.56326782, 0.86468541, 0.64938971, 0.74475835,
 0.88412251, 0.88162007, 0.20739217, 0.48674701, 0.67074315,
 0.62264292, 0.41383826, 0.52592258, 0.73579571, 0.13044174,
 0.23780284, 0.11213944, 0.37755142, 0.06696071, 0.08005528,
 0.92564577, 0.84312373, 0.91686996, 0.78330354, 0.60005718,
 0.79293628, 0.79358061, 0.16563309, 0.75537512, 0.99990456,
 0.33878635, 0.91077016, 0.98205712, 0.84224832, 0.01854094,
 0.05316031, 0.94211655, 0.12301673, 0.48834798, 0.37605959,
 0.0587928 , 0.10487136, 0.60064396, 0.70409041, 0.14478989,
 0.81941653, 0.65576694, 0.02106623, 0.44461772, 0.83762073,
 0.72598415, 0.95389978, 0.04099931, 0.02196911, 0.3423149 ,
 0.53586282, 0.90148957, 0.96177212, 0.67964464, 0.93839914,
 0.19015341, 0.99546472, 0.99000876, 0.97879929, 0.41433743,
 0.8246664 , 0.49063696, 0.96600692, 0.43696265, 0.99225919,
 0.97629256, 0.98546139, 0.90915194, 0.86372418, 0.07454627,
 0.66183448, 0.93037061, 0.95224557, 0.96090309, 0.97706647,
 0.55738951, 0.7908818 , 0.99966434, 0.8799152 , 0.85882669,
 0.97166349, 0.19376844, 0.24545566, 0.98374997, 0.98720709,
 0.97952174, 0.86873659, 0.45901584, 0.02231329, 0.94513486,
 0.27604723, 0.53678267, 0.4239514 , 0.29876913, 0.20671062,
 0.12569018, 0.16363093, 0.99138693, 0.94327216, 0.60703404,
 0.99139758, 0.25577819, 0.96221701, 0.99123063, 0.87550391,
 0.02331217, 0.44714789, 0.14325654, 0.84084097, 0.33618507,
 0.30451091, 0.27925851, 0.84980271, 0.03293702, 0.65410747,
 0.28677781, 0.21678549, 0.95868202, 0.32874638, 0.98466189,
 0.99674449, 0.99434211, 0.70082616, 0.99887611, 0.86262547,
 0.60553355, 0.68456507, 0.71452936, 0.44619825, 0.92514031,
 0.99526571, 0.59828069, 0.98652502, 0.11166096, 0.11655256,
 0.8957765 , 0.1559085 , 0.99764277, 0.88790332, 0.88570368,
 0.99405087, 0.02781507, 0.90394078, 0.93174969, 0.09575376,
 0.1925871 , 0.99868025, 0.03114969, 0.61061199, 0.75837988,
 0.63958738, 0.27350033, 0.0109953 , 0.21325143, 0.99925029,
 0.06844575, 0.04222313, 0.77657832, 0.75618011, 0.99808656,
 0.99982917, 0.01656464, 0.00815478, 0.89984384, 0.92879365,
 0.99769826, 0.9956138 , 0.9788553 , 0.03056599, 0.04219243,
 0.18477158, 0.92460994, 0.81062051, 0.99742783, 0.66033499,
 0.99615597, 0.99859145, 0.00275015, 0.51986392, 0.99729897,
 0.20273159, 0.38009183, 0.52173598, 0.0139623 , 0.24249992,
 0.29429542, 0.4099522 , 0.80558321, 0.24320866, 0.96532716,
 0.4845907 , 0.97524659, 0.1798739 , 0.37113867, 0.86613982,
 0.11502273, 0.99991041, 0.99179243, 0.22105882, 0.0417332 ,
 0.06222129, 0.83968608, 0.41552456, 0.25302837, 0.99657055,
 0.97289871, 0.99896694, 0.99391731, 0.52911566, 0.55321638,
 0.01770484, 0.91541843, 0.2995973 , 0.66732645, 0.27845741,
 0.00853725, 0.99858824, 0.28278814, 0.08874741, 0.79984821,
 0.99653553, 0.90474069, 0.93858826])

```
[63]: data['Domain'].median()

[63]: 0.6226429158

[64]: data['Domain'].replace(-1, data['Domain'].median(), inplace=True)

[65]: data['Domain'].unique()

[65]: array([0.63597876, 0.96060325, 0.45087658, 0.97439611, 0.12450207,
       0.62264292, 0.35653649, 0.8295846 , 0.69447933,
       0.49359639, 0.76567358, 0.9682375 , 0.22948175,
       0.53838689, 0.30840058,
       0.91139528, 0.56326782, 0.86468541, 0.64938971,
       0.74475835, 0.88412251, 0.88162007, 0.20739217,
       0.48674701, 0.67074315, 0.41383826, 0.52592258,
       0.73579571, 0.13044174, 0.23780284, 0.11213944,
       0.37755142, 0.06696071, 0.08005528, 0.92564577,
       0.84312373, 0.91686996, 0.78330354, 0.60005718,
       0.79293628, 0.79358061, 0.16563309, 0.75537512,
       0.99990456, 0.33878635, 0.91077016, 0.98205712,
       0.84224832, 0.01854094, 0.05316031,
       0.94211655, 0.12301673, 0.48834798, 0.37605959, 0.0587928 ,
       0.10487136, 0.60064396, 0.70409041, 0.14478989,
       0.81941653, 0.65576694, 0.02106623, 0.44461772,
       0.83762073, 0.72598415,
       0.95389978, 0.04099931, 0.02196911, 0.3423149 ,
       0.53586282, 0.90148957, 0.96177212, 0.67964464,
       0.93839914, 0.19015341, 0.99546472, 0.99000876,
       0.97879929, 0.41433743, 0.8246664 , 0.49063696,
       0.96600692, 0.43696265, 0.99225919, 0.97629256,
       0.98546139, 0.90915194, 0.86372418, 0.07454627,
       0.66183448, 0.93037061, 0.95224557, 0.96090309,
       0.97706647, 0.55738951, 0.7908818 , 0.99966434, 0.8799152
       , 0.85882669, 0.97166349, 0.19376844, 0.24545566,
       0.98374997, 0.98720709, 0.97952174, 0.86873659,
       0.45901584, 0.02231329, 0.94513486, 0.27604723,
       0.53678267, 0.4239514 , 0.29876913, 0.20671062,
       0.12569018, 0.16363093, 0.99138693, 0.94327216,
       0.60703404, 0.99139758, 0.25577819, 0.96221701,
       0.99123063, 0.87550391, 0.02331217, 0.44714789,
       0.14325654, 0.84084097, 0.33618507, 0.30451091,
       0.27925851, 0.84980271, 0.03293702, 0.65410747,
       0.28677781, 0.21678549, 0.95868202, 0.32874638,
       0.98466189, 0.99674449,
       0.99434211, 0.70082616, 0.99887611, 0.86262547, 0.60553355,
       0.68456507, 0.71452936, 0.44619825, 0.92514031, 0.99526571,
```

```

0.59828069, 0.98652502, 0.11166096, 0.11655256,
0.8957765 , 0.1559085 , 0.99764277, 0.88790332,
0.88570368, 0.99405087, 0.02781507, 0.90394078,
0.93174969, 0.09575376, 0.1925871 ,
0.99868025, 0.03114969, 0.61061199, 0.75837988, 0.63958738,
0.27350033, 0.0109953 , 0.21325143, 0.99925029, 0.06844575,
0.04222313, 0.77657832, 0.75618011, 0.99808656, 0.99982917,
0.01656464, 0.00815478, 0.89984384, 0.92879365, 0.99769826,
0.9956138 , 0.9788553 , 0.03056599, 0.04219243, 0.18477158,
0.92460994, 0.81062051, 0.99742783, 0.66033499, 0.99615597,
0.99859145, 0.00275015, 0.51986392, 0.99729897, 0.20273159,
0.38009183, 0.52173598, 0.0139623 , 0.24249992, 0.29429542,
0.4099522 , 0.80558321, 0.24320866, 0.96532716, 0.4845907 ,
0.97524659, 0.1798739 , 0.37113867, 0.86613982, 0.11502273,
0.99991041, 0.99179243, 0.22105882, 0.0417332 , 0.06222129,
0.83968608, 0.41552456, 0.25302837, 0.99657055, 0.97289871,
0.99896694, 0.99391731, 0.52911566, 0.55321638, 0.01770484,
0.91541843, 0.2995973 , 0.66732645, 0.27845741, 0.00853725,
0.99858824, 0.28278814, 0.08874741, 0.79984821, 0.99653553,
0.90474069, 0.93858826])

```

- Succesfully imputed -1 with its Median Value

1.10 CollegeState Column

- CollegeState column contains State name as Union Territory with 5 entries. we can Avoid making changes

```
[66]: data['CollegeState'].value_counts()
```

```
[66]: CollegeState
```

Uttar Pradesh	915
Karnataka	370
Tamil Nadu	367
Telangana	319
Maharashtra	262
Andhra Pradesh	225
West Bengal	196
Punjab	193
Madhya Pradesh	189
Haryana	180
Rajasthan	174
Orissa	172
Delhi	162
Uttarakhand	113
Kerala	33

```
Jharkhand          28
Chhattisgarh       27
Gujarat           24
Himachal Pradesh  16
Bihar              10
Jammu and Kashmir 7
Assam              5
Union Territory    5
Sikkim             3
Meghalaya          2
Goa                1
Name: count, dtype: int64
```

2 Analysis Starts Here

2.1 Univariate Analysis

```
[67]: data.head()
```

```
[67]:      ID      Salary        DOJ        DOL      Designation \
0  203097  420000.0 2012-06-01 2020-03-20  senior quality engineer
1  579905  500000.0 2013-09-01 2020-03-20  assistant manager
2  810601  325000.0 2014-06-01 2020-03-20  systems engineer  3  267447
   1100000.0 2011-07-01 2020-03-20  senior software engineer
4  343523  200000.0 2014-03-01 2015-01-03      system engineer

      JobCity  Gender  DOB  10percentage  10board  12graduation \
Bengaluru  f  1990-02-19  84.3  State Board  2007 1  Indore  m  1989-10-
04  85.4  CBSE  2007 2  Chennai  f  1992-08-03  85.0  CBSE  2010
3  Gurugram  m  1989-12-05      85.6  CBSE  2007
4  Manesar  m  1991-02-27      78.0  CBSE  2008

      12percentage  12board CollegeID CollegeTier      Degree \
0  95.8  State Board  1141 2  B.Tech/B.E.  1  85.0  CBSE  5807 2
B.Tech/B.E.  2  68.2  CBSE  64 2  B.Tech/B.E.  3  83.6  CBSE  6920
1  B.Tech/B.E.  4  76.8  CBSE  11368 2  B.Tech/B.E.

Specialization collegeGPA CollegeCityID CollegeCityTier CollegeState \
0            CSE  78.00 1141 0  Andhra Pradesh
1            ECE  70.06 5807 0  Madhya Pradesh
2            CSE  70.00 64 0  Uttar Pradesh
3            CSE  74.64 6920 1  Delhi
```

	ECE	73.90	11368	0	Uttar	Pradesh	GraduationYear	English
0	1970-01-01	00:00:00.000002011	515	585	525	0.635979		
1	1970-01-01	00:00:00.000002012	695	610	780	0.960603		
2	1970-01-01	00:00:00.000002014	615	545	370	0.450877		
3	1970-01-01	00:00:00.000002011	635	585	625	0.974396		
4	1970-01-01	00:00:00.000002012	545	625	465	0.124502		
	ComputerProgramming	ElectronicsAndSemicon	ComputerScience	\				
0	445		-1		-			
1	-1		466		1			
2	395		-1		-			
3	615		-1		1			
4	-1		233		-			
MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg	conscientiousness	\			
0	-1	-1	-1	-1	0.9737			
1	-1	-1	-1	-1	-	0.7335		
2	-1	-1	-1	-1	0.2718			
3	-1	-1	-1	-1	0.0464			
4	-1	-1	-1	-1	-	0.8810		
agreeableness	extraversion	nueroticism	openness_to_experience	period	\			
0	0.8128	0.5269	1.35490	-0.4455	8			
1	0.3789	1.2396	-0.10760	0.8637	7			
2	1.7109	0.1637	-0.86820	0.6721	6			
3	0.3448	-0.3440	-0.40780	-0.9194	9			
4	-0.2793	-1.0697	0.09163	-0.1295	1			
DOB_Year	GYear	12GradAge	GradAge					
0	1990	1970	20	20				
1	1989	1970	19	19				
2	1992	1970	22	22				
3	1989	1970	19	19				
4	1991	1970	21	21				

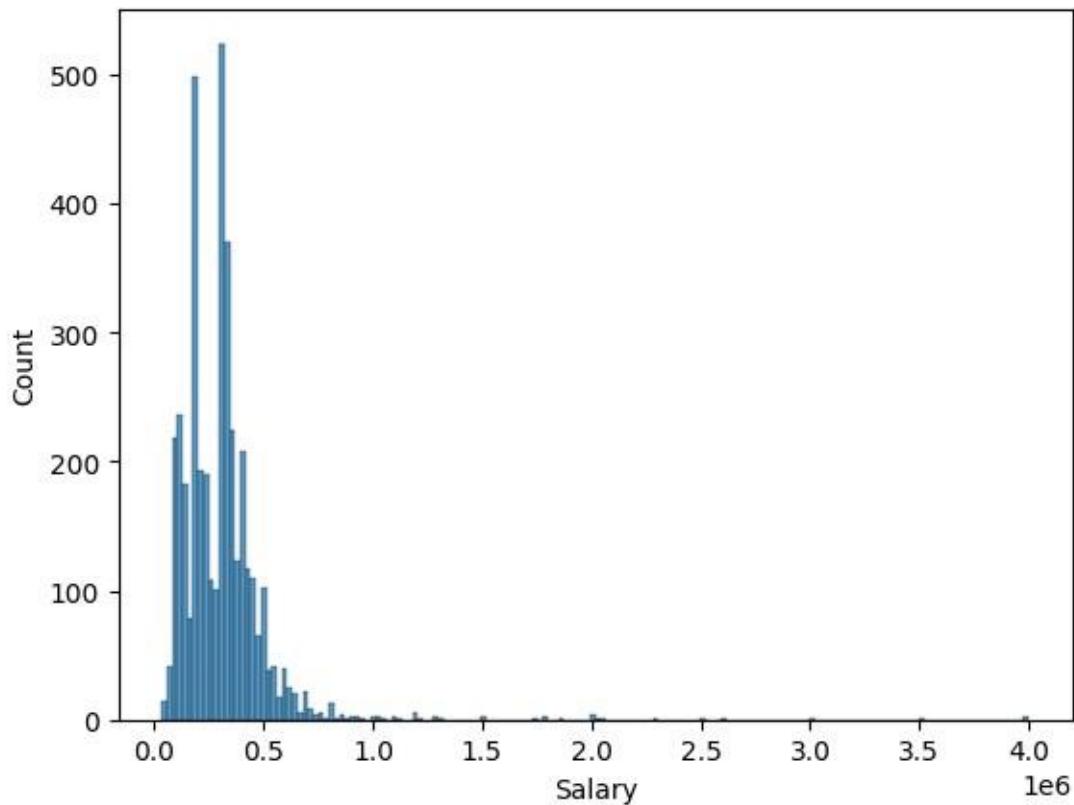
2.2 Salary Column

```
[68]: data['Salary'].min()
```

```
[68]: 35000.0
```

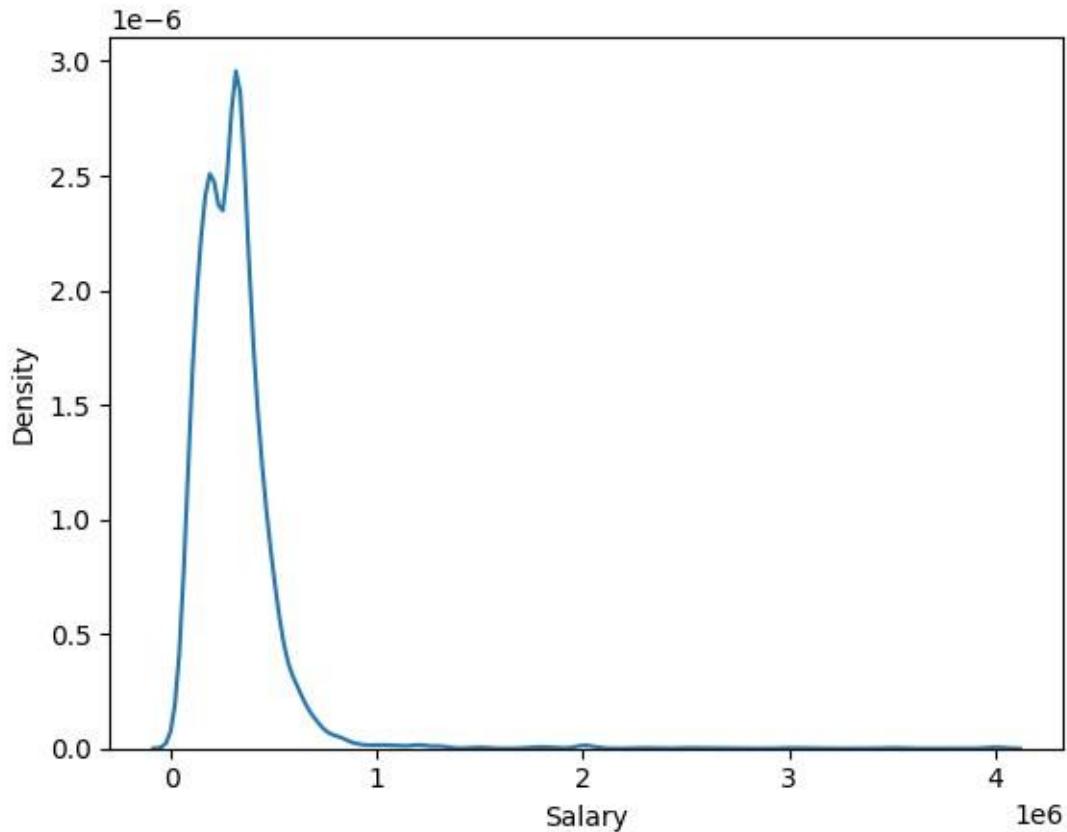
```
[69]: sns.histplot(data=data, x='Salary')
```

```
[69]: <AxesSubplot:xlabel='Salary', ylabel='Count'>
```



```
[70]: sns.kdeplot(data=data, x='Salary')
```

```
[70]: <AxesSubplot:xlabel='Salary', ylabel='Density'>
```

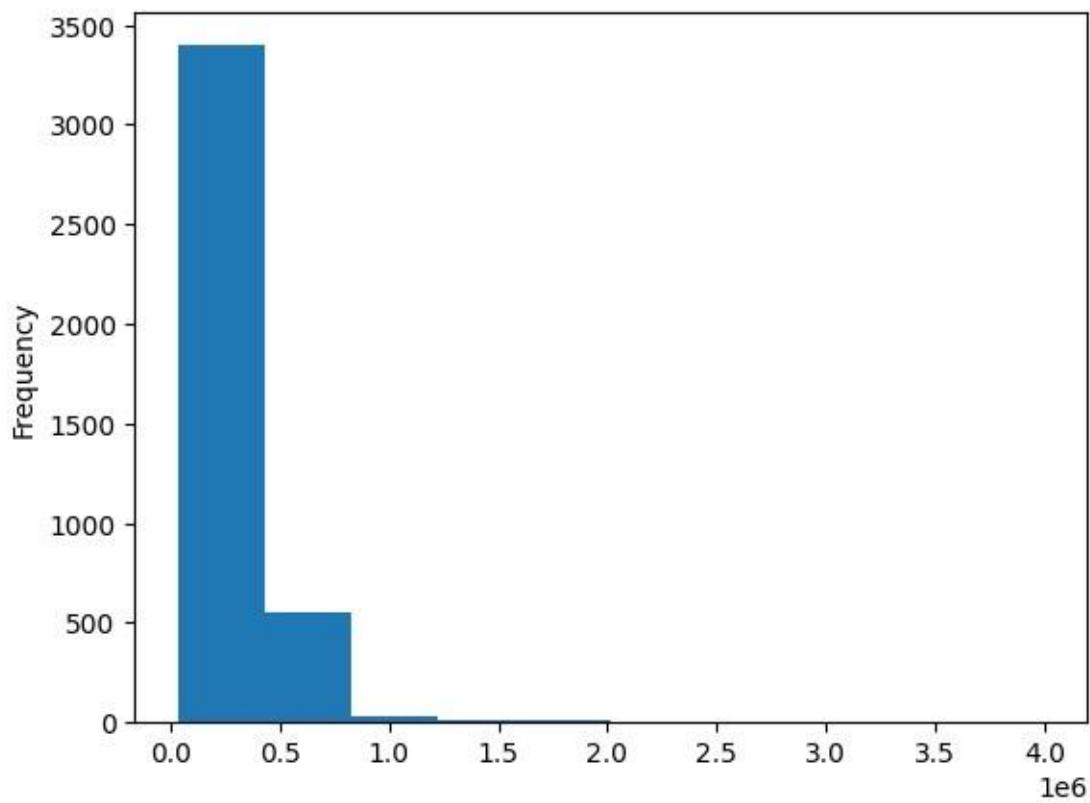


- Salary of people <50000 might enter Monthly salary instead of Yearly Salary

```
[71]: data[data['Salary'] <= 50000]['Salary'] *= 12
```

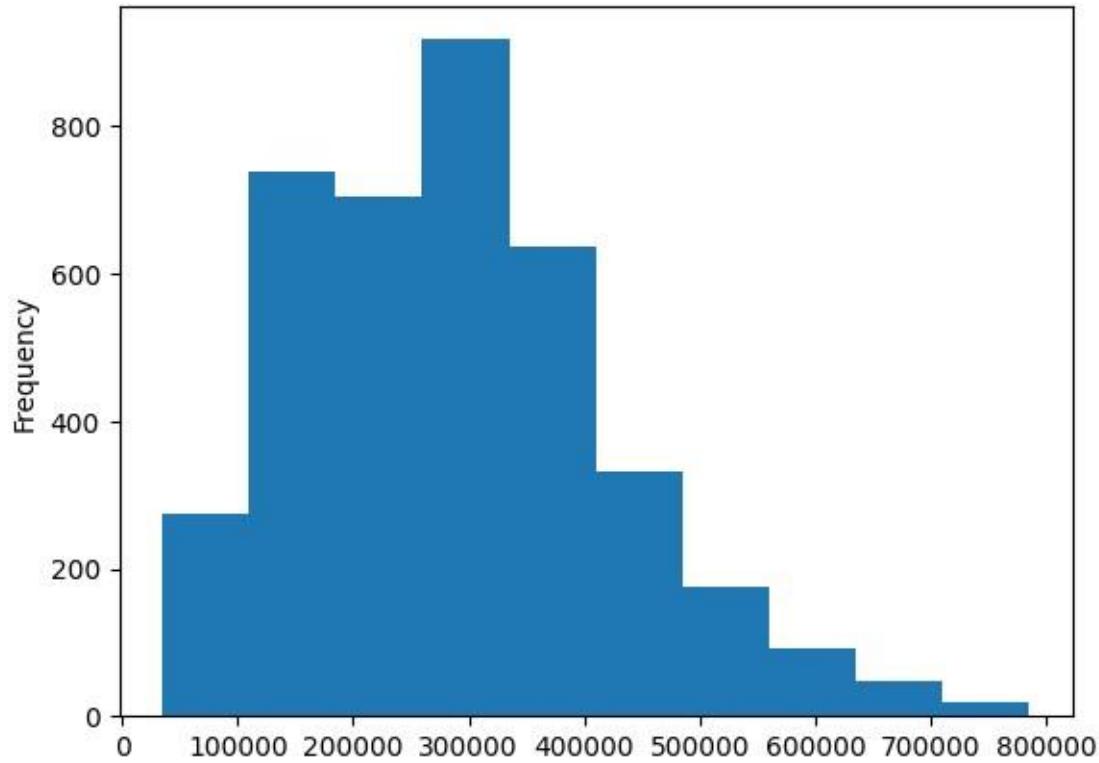
```
[72]: data['Salary'].plot(kind='hist')
```

```
[72]: <AxesSubplot:ylabel='Frequency'>
```



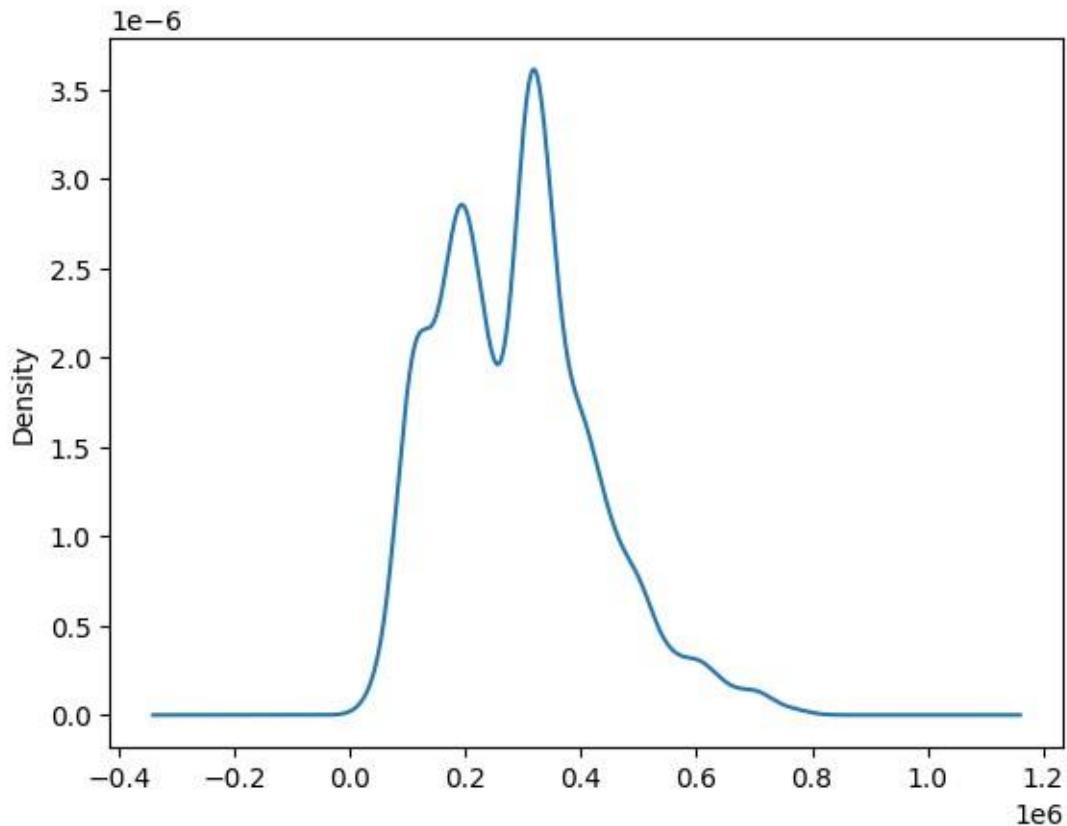
```
[73]: data[data['Salary'] < 800000]['Salary'].plot(kind='hist')
```

```
[73]: <AxesSubplot:ylabel='Frequency'>
```



```
[74]: data[data['Salary'] < 800000]['Salary'].plot(kind='kde')
```

```
[74]: <AxesSubplot:ylabel='Density'>
```



2.3 ‘ComputerProgramming’, ‘ElectronicsAndSemicon’, ‘ComputerScience’, ‘MechanicalEngg’ ‘ElectricalEngg’, ‘TelecomEngg’, ‘CivilEng Columns contain -1 g

[75]: data.columns

```
[75]: Index(['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',
```

```

        'period',    'DOB_Year',    'GYear',    '12GradAge',    'GradAge'],
       dtype='object')

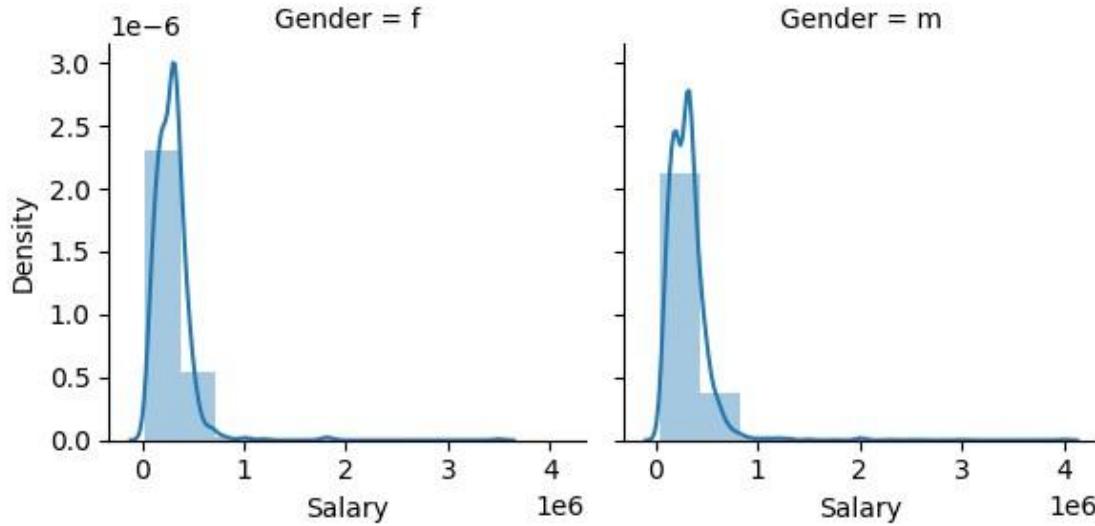
[76]: data[['ComputerProgramming', 'ElectronicsAndSemicon',
'ComputerScience', ...
        'MechanicalEngg', 'ElectricalEngg', 'TelecomEngg', 'CivilEngg']]
```

	ComputerProgramming	ElectronicsAndSemicon	ComputerScience	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg
0	445	-1	-1				
1	-1	466	-1				
2	395	-1	-1				
3	615	-1	-1				
4	-1	233	-1				
...
3993	345	-1	-1				
3994	325	420	-1				
3995	405	-1	-1				
3996	445	-1	438				
3997	435	-1	-1				
	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg			
0	-1	-1	-1	-1			
1	-1	-1	-1	-1			
2	-1	-1	-1	-1			
3	-1	-1	-1	-1			
4	-1	-1	-1	-1			
...
3993	-1	-1	-1	-1			
3994	-1	-1	-1	-1			
3995	-1	-1	-1	-1			
3996	-1	-1	-1	-1			
3997	-1	-1	-1	-1			

```
[3998 rows x 7 columns]
```

```
[77]: plt.figure(figsize=(15,5))
colors = sns.color_palette('bright',n_colors=2)
sns.FacetGrid(data, col="Gender", palette=colors).map(sns.distplot,_
    "Salary",bins=10)
plt.show()
```

<Figure size 1500x500 with 0 Axes >



- The distributions are right skewed, Male and Female have almost same distribution

2.4 Designation Column - Univariate

```
[78]: data['Designation'].unique()
```

```
[78]: 418
```

- Designation Column having More cardinality
- We can change the Designation column to low cardinality

```
[79]: desig = data['Designation'].unique()
```

```
desig.sort()
```

```
[80]: desig
```

```
[80]: array(['.net developer', '.net web developer', 'account executive',
       'account manager', 'admin assistant', 'administrative
       coordinator',
```

'administrative support', 'aircraft technician',
'android developer', 'application developer',
'application engineer', 'apprentice', 'ase', 'asp.net developer',
'assistant administrator', 'assistant electrical engineer',
'assistant engineer', 'assistant manager', 'assistant professor',
'assistant programmer', 'assistant software engineer',
'assistant store manager', 'assistant system engineer',
'assistant system engineer - trainee',
'assistant system engineer trainee', 'assistant systems
engineer',
'associate developer', 'associate engineer', 'associate manager',
'associate qa', 'associate software developer',
'associate software engg', 'associate software engineer',
'associate system engineer', 'associate technical operations',
'associate test engineer', 'asst. manager', 'automation
engineer',
'branch manager', 'bss engineer', 'business analyst',
'business analyst consultant', 'business consultant',
'business development executive', 'business development manager',
'business development managerde', 'business intelligence
analyst',
'business office manager', 'business process analyst',
'business system analyst', 'business systems analyst',
'business systems consultant', 'business technology analyst',
'c# developer', 'cad designer', 'cad drafter', 'catalog
associate',
'civil engineer', 'clerical', 'clerical assistant',
'client services associate', 'cloud engineer', 'cnc programmer',
'co faculty', 'computer faculty',
'continuous improvement engineer', 'controls engineer',
'corporate recruiter', 'customer care executive',
'customer service', 'customer service manager',
'customer service representative', 'customer support engineer',
'data analyst', 'data entry operator', 'data scientist',
'database administrator', 'database developer', 'db2 dba', 'dba',
'dcs engineer', 'delivery software engineer', 'design engineer',
'designer', 'desktop support analyst', 'desktop support
engineer',
'desktop support technician', 'developer',
'digital marketing specialist', 'documentation specialist',
'dotnet developer', 'editor', 'educator',
'electrical controls engineer', 'electrical design engineer',
'electrical designer', 'electrical engineer',
'electrical field engineer', 'electrical project engineer',
'electronic field service engineer', 'embedded engineer',

'embedded software engineer', 'engineer', 'engineer trainee',
'engineer- customer support', 'engineer-hws',
'engineering manager', 'engineering technician',
'enterprise solutions developer', 'entry level management
trainee',
'entry level sales and marketing', 'environmental engineer',
'etl developer', 'executive administrative assistant',
'executive assistant', 'executive engg', 'executive engineer',
'executive hr', 'executive recruiter', 'faculty',
'field based employee relations manager',
'field business development associate', 'field engineer',
'field service engineer', 'financial analyst',
'financial service consultant', 'firmware engineer',
'front end developer', 'front end web developer',
'full stack developer', 'full-time loss prevention associate',
'game developer', 'general manager', 'gis/cad engineer',
'graduate apprentice trainee', 'graduate engineer',
'graduate engineer trainee', 'graduate trainee engineer',
'graphic designer', 'hardware engineer', 'help desk analyst',
'help desk technician', 'hr assistant', 'hr executive',
'hr generalist', 'hr manager', 'hr recruiter', 'html developer',
'human resource assistant', 'human resources analyst',
'human resources associate', 'human resources intern',
'implementation engineer', 'industrial engineer',
'information security analyst',
'information technology specialist', 'ios developer', 'it
analyst',
'it assistant', 'it business analyst', 'it developer',
'it engineer', 'it executive', 'it operations associate',
'it recruiter', 'it support specialist', 'it technician',
'java developer', 'java software engineer', 'java trainee',
'javascript developer', 'jr. software developer',
'jr. software engineer', 'junior .net developer',
'junior engineer', 'junior engineer product support',
'junior manager', 'junior recruiter', 'junior research fellow',
'junior software developer', 'junior software engineer',
'junior system analyst', 'lead engineer', 'lecturer',
'lecturer & electrical maintenance', 'linux systems
administrator',
'logistics executive', 'maintenance engineer',
'maintenance supervisor', 'management trainee', 'manager',
'manual tester', 'manufacturing engineer', 'marketing analyst',
'marketing assistant', 'marketing coordinator',
'marketing executive', 'marketing manager',
'mechanical design engineer', 'mechanical engineer',

'mis executive', 'mobile application developer',
'network administrator', 'network engineer',
'network security engineer', 'network support engineer',
'noc engineer', 'office coordinator', 'online marketing manager',
'operation engineer', 'operation executive',
'operational excellence manager', 'operational executive',
'operations', 'operations analyst', 'operations assistant',
'operations engineer', 'operations engineer and jetty handling',
'operations executive', 'operations manager', 'oracle dba',
'performance engineer', 'phone banking officer', 'php developer',
'planning engineer', 'portfolio analyst',
'principal software engineer', 'process advisor',
'process associate', 'process control engineer',
'process engineer', 'process executive', 'product design engineer',
'product developer', 'product development engineer',
'product engineer', 'product manager', 'production engineer',
'professor', 'program analyst trainee', 'program manager',
'programmer', 'programmer analyst', 'programmer analyst trainee',
'project administrator', 'project assistant',
'project coordinator', 'project engineer',
'project management officer', 'project manager',
'python developer', 'qa analyst', 'qa engineer', 'qa trainee',
'quality analyst', 'quality associate', 'quality assurance',
'quality assurance analyst', 'quality assurance auditor',
'quality assurance automation engineer',
'quality assurance engineer', 'quality assurance test engineer',
'quality assurance tester', 'quality consultant',
'quality control engineer',
'quality control inspection technician',
'quality control inspector', 'quality controller',
'quality engineer', 'r & d', 'r&d engineer', 'recruiter',
'recruitment associate', 'recruitment coordinator',
'research analyst', 'research associate', 'research engineer',
'research scientist', 'research staff member', 'rf engineer',
'rf/dt engineer', 'risk consultant', 'risk investigator',
'ruby on rails developer', 'sales & service engineer',
'sales account manager', 'sales and service engineer',
'sales associate', 'sales coordinator',
'sales development manager', 'sales engineer', 'sales executive',
'sales management trainee', 'sales manager', 'sales support',
'sales trainer', 'salesforce developer',
'sap abap associate consultant', 'sap abap consultant',
'sap analyst', 'sap consultant', 'sap functional consultant',
'sap mm consultant', 'secretary', 'senior .net developer',

'senior business analyst', 'senior design engineer',
'senior developer', 'senior engineer', 'senior java developer',
'senior mechanical engineer', 'senior network engineer',
'senior php developer', 'senior programmer',
'senior project engineer', 'senior quality assurance engineer',
'senior quality engineer', 'senior research fellow',
'senior risk consultant', 'senior sales executive',
'senior software developer', 'senior software engineer',
'senior systems engineer', 'senior test engineer',
'senior web developer', 'seo', 'seo analyst', 'seo engineer',
'seo executive', 'seo trainee', 'service and sales engineer',
'service coordinator', 'service engineer', 'service manager',
'sharepoint developer', 'shift engineer', 'site engineer',
'site manager', 'software analyst', 'software architect',
'software designer', 'software developer',
'software development engineer', 'software devloper',
'software eng', 'software engg', 'software engineer',
'software engineer analyst', 'software engineer associate',
'software engineer trainee', 'software engineere',
'software engineering associate', 'software enginner',
'software executive', 'software programmer',
'software quality assurance analyst',
'software quality assurance tester', 'software test engineer',
'software test engineer (etl)', 'software test engineerte',
'software tester', 'software trainee', 'software trainee
engineer',
'sql dba', 'sql developer', 'sr. database engineer',
'sr. engineer', 'staffing recruiter', 'supply chain analyst',
'support engineer', 'system administrator', 'system analyst',
'system engineer', 'system engineer trainee',
'systems administrator', 'systems analyst', 'systems engineer',
'talent acquisition specialist', 'team lead', 'team leader',
'technical analyst', 'technical assistant', 'technical
consultant',
'technical engineer', 'technical lead',
'technical operations analyst', 'technical recruiter',
'technical support engineer', 'technical support executive',
'technical support specialist', 'technical writer',
'technology analyst', 'technology lead', 'telecom engineer',
'telecom support engineer', 'telecommunication engineer',
'teradata dba', 'teradata developer', 'territory sales manager',
'test engineer', 'test technician', 'testing engineer',
'trainee decision scientist', 'trainee engineer',
'trainee software developer', 'trainee software engineer',
'training specialist', 'ui developer', 'ux designer',

```
'visiting faculty', 'web application developer', 'web designer',
'web designer and joomla administrator', 'web designer and seo',
'web developer', 'web intern', 'website developer/tester',
'windows systems administrator'], dtype=object)
```

```
[81]: def feature_cleaning(input_val, input_list):
    if type(input_val) == str:
        for item in [i for i in input_list if len(i.split()) > 1]:
            if all([x in input_val for x in item.split()]):
                return item.title()

        for item in [i for i in input_list if len(i.split()) == 1]:
            if item in input_val:
                return item.title()
            if 'engineer' in input_val:
                return 'Hardware Engineer'
        try:
            matched_item = get_close_matches(input_val, input_list)[0]
            return matched_item.title()
        except:
            return 'Other'

    else:
        return np.nan
```

```
[82]: roles_list = ['software engineer', 'system engineer', 'developer', 'analyst',
    ↵'test engineer', 'dba',
    ↵'administrator', 'customer service', 'quality engineer', 'quality',
    ↵'automation engineer',
    ↵'network engineer', 'support', 'it engineer', 'manager',
    ↵'management', 'programmer',
    ↵'tester', 'qa engineer', 'design']
```

```
[83]: data['Job_Role'] = data['Designation'].apply(lambda x: feature_cleaning(x,
    ↵roles_list))
jr_sorted = data['Job_Role'].unique()
jr_sorted.sort()
jr_sorted
```

```
[83]: array(['Administrator', 'Analyst', 'Automation Engineer',
    'Customer Service', 'Dba', 'Design', 'Developer',
    'Hardware Engineer', 'It Engineer', 'Management', 'Manager',
    'Network Engineer', 'Other', 'Programmer', 'Qa Engineer',
    'Quality', 'Quality Engineer', 'Software Engineer', 'Support',
    'System Engineer', 'Test Engineer', 'Tester'], dtype=object)
```

```
[84]: data['Job_Role'] = data['Job_Role'].replace({'It Engineer':  
'Software_Engineer', 'Network Engineer': 'System Engineer', 'Dba': 'System  
Engineer',  
'Support': 'Administrator', 'Customer_Service': 'Administrator',  
'Tester': 'Test Engineer', 'Qa Engineer': 'Test Engineer',  
'Quality': 'Test Engineer',  
'Quality Engineer': 'Test Engineer',  
'Automation Engineer': 'Test Engineer',  
'Programmer': 'Developer',  
'Management': 'Manager', 'Design': 'Other'})
```

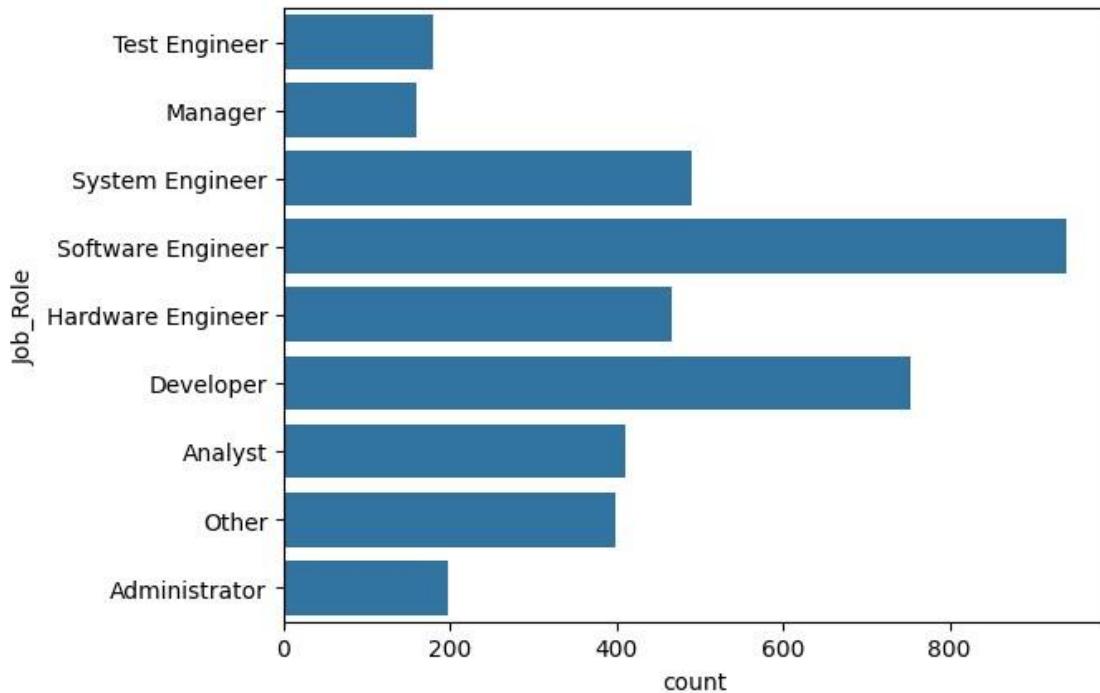
```
[85]: data['Job_Role'].value_counts(dropna=False)
```

```
[85]: Job_Role
```

```
Software Engineer 941  
Developer 753  
System Engineer 491  
Hardware Engineer 466  
Analyst 411  
Other 399  
Administrator 197  
Test Engineer 180  
Manager 160  
Name: count, dtype: int64
```

```
[86]: sns.countplot(data['Job_Role'])
```

```
[86]: <AxesSubplot: xlabel='count', ylabel='Job_Role'>
```



- By this analysis Software Engineer are more and second Highest is Developer
- From now onwards in the analysis we can use Job_Role instead Designation

2.5 JobCity Column

```
[87]: data['JobCity'].value_counts()
```

```
[87]: JobCity
```

Bengaluru	1126
Noida	390
Hyderabad	371
Pune	327
Chennai	313
...	
GANJAM	1
GAZIABAAD	1
Dharuhera	1
Miryalaguda	1
ASIFABADBANGLORE	1

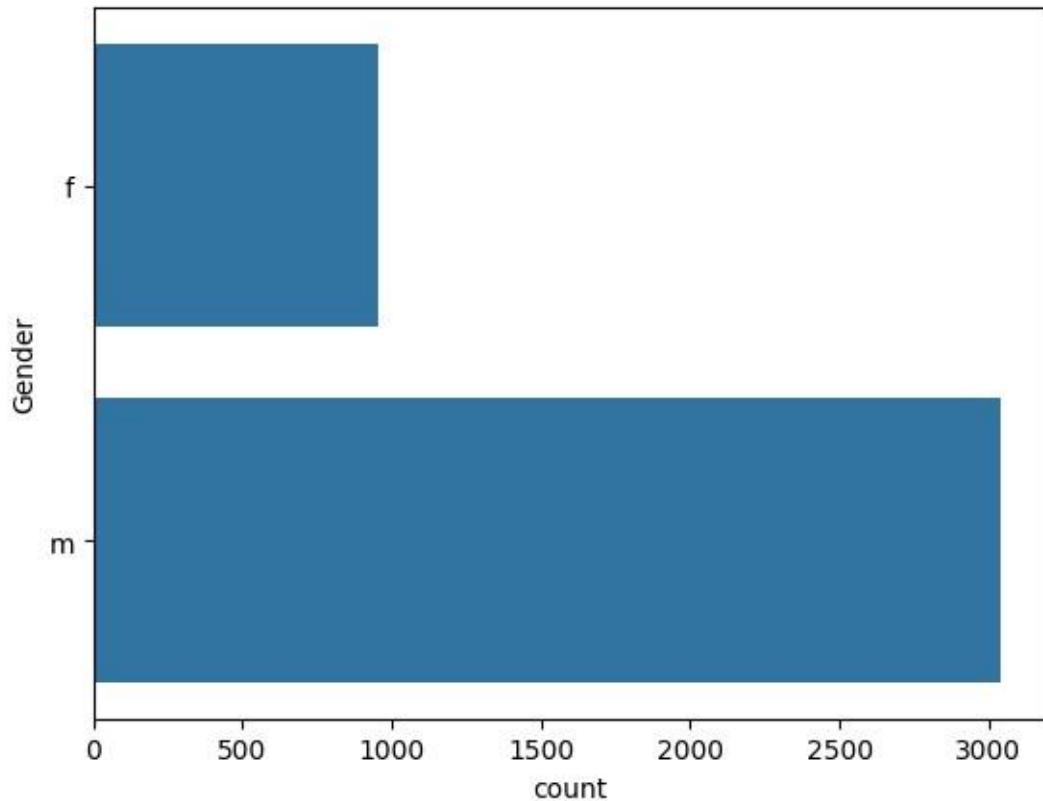
Name: count, Length: 222, dtype: int64

- **JobCity having Highest cardinality**
- In JobCity more people are working in Bengaluru and second most is Noida

2.6 Gender column

```
[88]: sns.countplot(data['Gender'])
```

```
[88]: <AxesSubplot:xlabel='count', ylabel='Gender'>
```

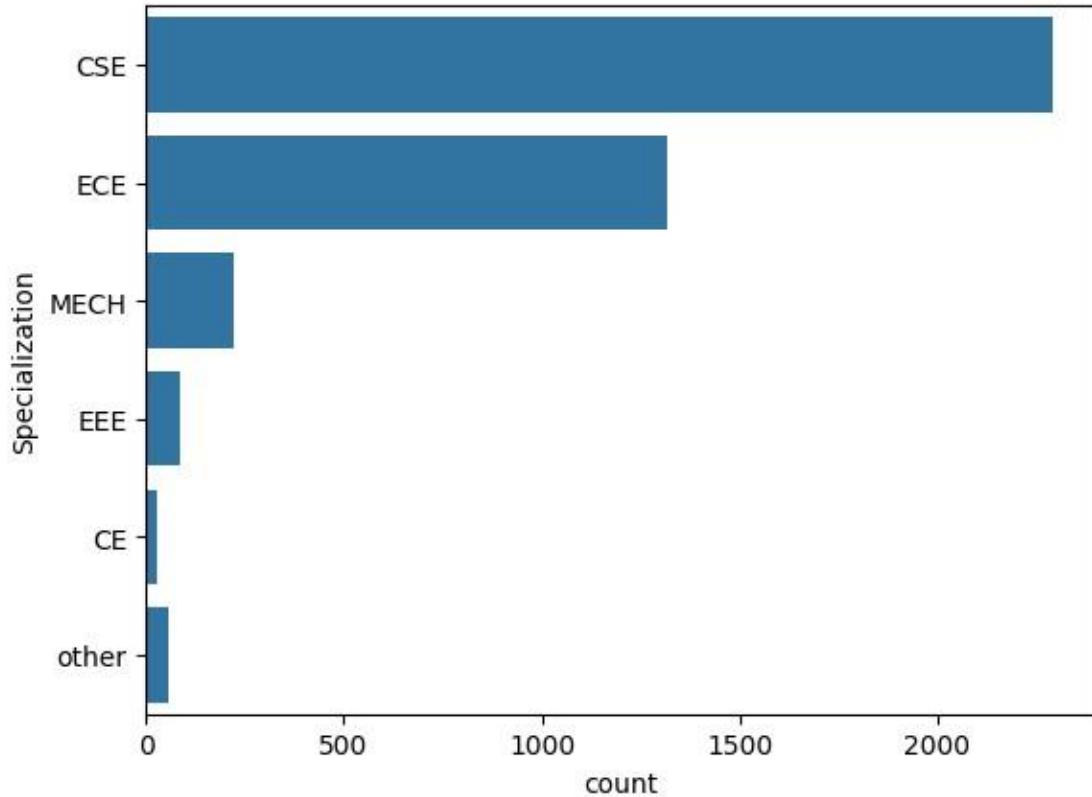


- From the Gender column - proportion of Male are more compared to Female
- The Proportion looks like $f : m = 1 : 3$

2.7 Specialization Column

```
[89]: sns.countplot(data['Specialization'])
```

```
[89]: <AxesSubplot:xlabel='count', ylabel='Specialization'>
```

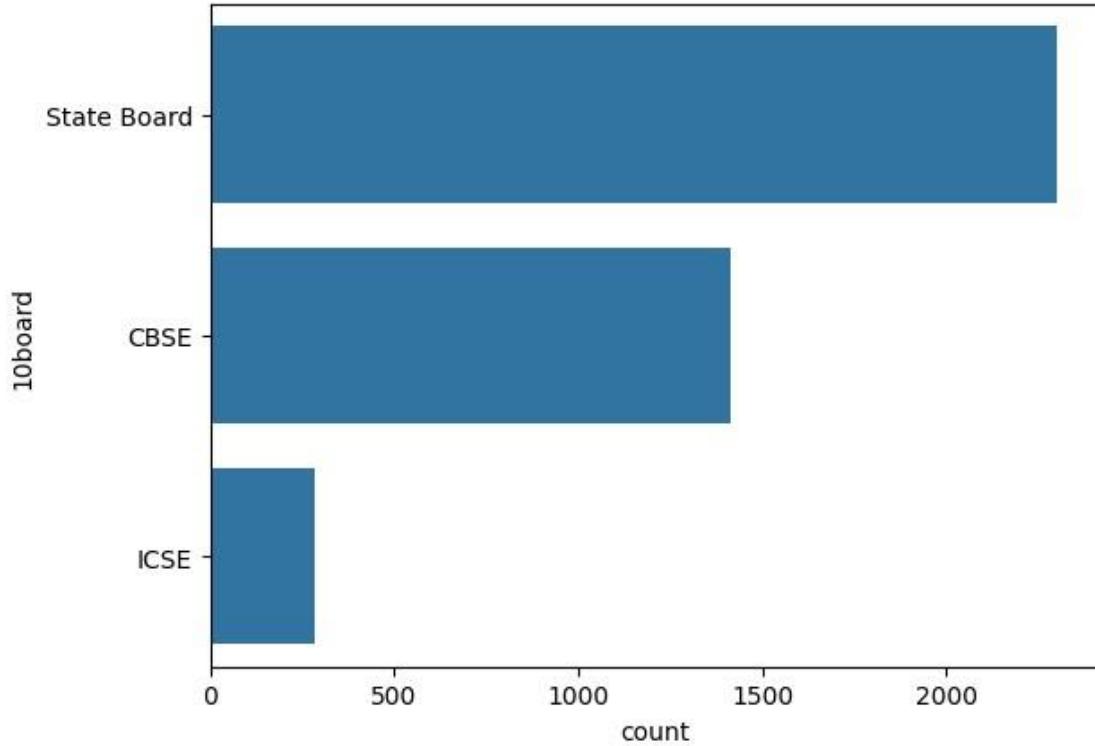


- Based on Specialization most of the people are from CSE > ECE > MECH > EEE > CE > other

2.8 10board & 12board Columns

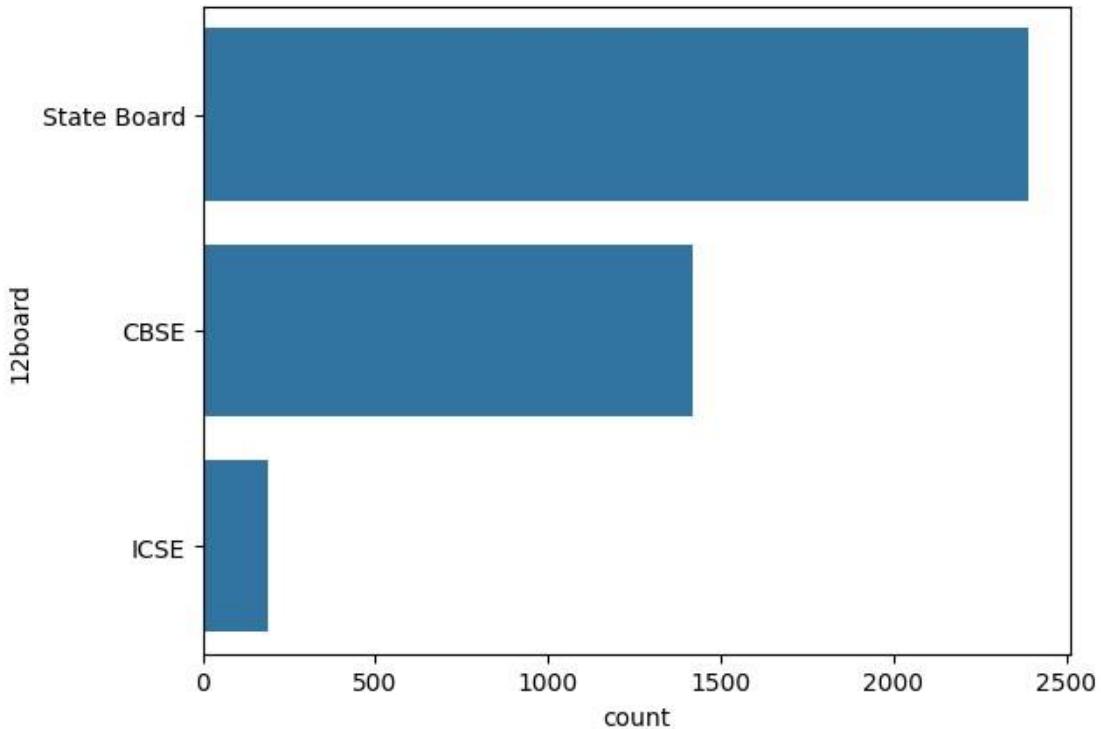
```
[90]: sns.countplot(data['10board'])
```

```
[90]: <AxesSubplot:xlabel='count', ylabel='10board'>
```



```
[91]: sns.countplot(data['12board'])
```

```
[91]: <AxesSubplot:xlabel='count', ylabel='12board'>
```



- Based on 10board & 12board most of the people are from State Board, CBSE and Followed by ICSE

2.9 GradAge Column

```
[92]: data['GradAge'].mode()
```

```
[92]: 0    21
Name: GradAge, dtype: int32
```

- People who are appearing For AMCAT exam - Most of the people are Completed their Graduation at the age of 21 Years

2.10 AMCAT - Scores

```
[93]: amcat_scores = ['English', 'Logical', 'Quant', 'Domain',
'ComputerProgramming',
'ElectronicsAndSemicon', 'ComputerScience', 'MechanicalEngg',
'ElectricalEngg', 'TelecomEngg', 'CivilEngg',
'conscientiousness',
'agreeableness', 'extraversion', 'nueroticism', 'openess_to_experience']
```

```
[94]: np.round(data[amcat_scores].describe(), 2)
```

[94]:

	English	Logical	Quant	Domain	ComputerProgramming	\
count	3998.00	3998.00	3998.00	3998.00		3998.00
mean	501.65	501.60	513.38	0.61		353.10
std	104.94	86.78	122.30	0.26		205.36
min	180.00	195.00	120.00	0.00		-1.00
25%	425.00	445.00	430.00	0.41		295.00
50%	500.00	505.00	515.00	0.62		415.00
75%	570.00	565.00	595.00	0.84		495.00
max	875.00	795.00	900.00	1.00		840.00
	ElectronicsAndSemicon		ComputerScience		MechanicalEngg	
	ElectricalEngg	\				
count		3998.00		3998.00		3998.00
mean		95.33		90.74		22.97
std		158.24		175.27		98.12
min		-1.00		-1.00		-1.00
25%		-1.00		-1.00		-1.00
50%		-1.00		-1.00		-1.00
75%		233.00		-1.00		-1.00
max		612.00		715.00		623.00
	676.00					
	TelecomEngg	CivilEngg		conscientiousness		agreeableness
	extraversion	\				
count	3998.00	3998.00		3998.00		3998.00
mean	31.85	2.68		-0.04		0.15
std	104.85	36.66		1.03		0.94
min	-1.00	-1.00		-4.13		-5.78
25%	-1.00	-1.00		-0.71		-0.29
50%	-1.00	-1.00		0.05		0.21
75%	-1.00	-1.00		0.70		0.81
max	548.00	516.00		2.00		1.90
						0.67
						2.54
	nueroticism	openess_to_experience				
count	3998.00		3998.00			
mean	-0.17		-0.14			
std	1.01		1.01			
min	-2.64		-7.38			
25%	-0.87		-0.67			
50%	-0.23		-0.09			
75%	0.53		0.50			
max	3.35		1.82			

```
[95]: data[amcat_scores].replace(-1, np.nan, inplace=True)
```

```
[96]: data[amcat_scores].describe().round()
```

```
[96]: English Logical Quant Domain ComputerProgramming \
count      3998.0      3998.0 3998.0 3998.0 3998.0 mean
      502.0 502.0 513.0 1.0    353.0

std      105.0      87.0 122.0    0.0      205.0
min      180.0     195.0 120.0    0.0      -1.0
25%     425.0     445.0 430.0    0.0      295.0
50%     500.0     505.0 515.0    1.0      415.0
75%     570.0     565.0 595.0    1.0      495.0
max     875.0     795.0 900.0    1.0      840.0

ElectronicsAndSemicon      ComputerScience      MechanicalEngg
ElectricalEngg \
count          3998.0      3998.0 3998.0 3998.0
mean          95.0        91.0   23.0   16.0
std          158.0      175.0   98.0   88.0
min         -1.0        -1.0   -1.0   -1.0
25%         -1.0        -1.0   -1.0   -1.0
50%         -1.0        -1.0   -1.0   -1.0
75%        233.0       -1.0   -1.0   -1.0
max        612.0       715.0   623.0   676.0

TelecomEngg      CivilEngg conscientiousness agreeableness
extraversion \
count          3998.0      3998.0 3998.0 3998.0 3998.0
mean          32.0        3.0    -0.0    0.0    0.0
std          105.0      37.0    1.0    1.0    1.0
min         -1.0        -1.0   -4.0   -6.0   -5.0
25%         -1.0        -1.0   -1.0   -0.0   -1.0
50%         -1.0        -1.0    0.0    0.0    0.0
75%         -1.0        -1.0    1.0    1.0    1.0
max        548.0       516.0    2.0    2.0    3.0

nueroticism
openess_to_experience
count          3998.0      3998.0
mean         -0.0        -0.0
std           1.0        1.0
min          -3.0       -7.0
25%         -1.0       -1.0
```

50%	-0.0	-0.0
75%	1.0	1.0
max	3.0	2.0

```
[97]: for col in amcat_scores:
    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12,4),  

    ↪constrained_layout=True, facecolor='lightyellow',  

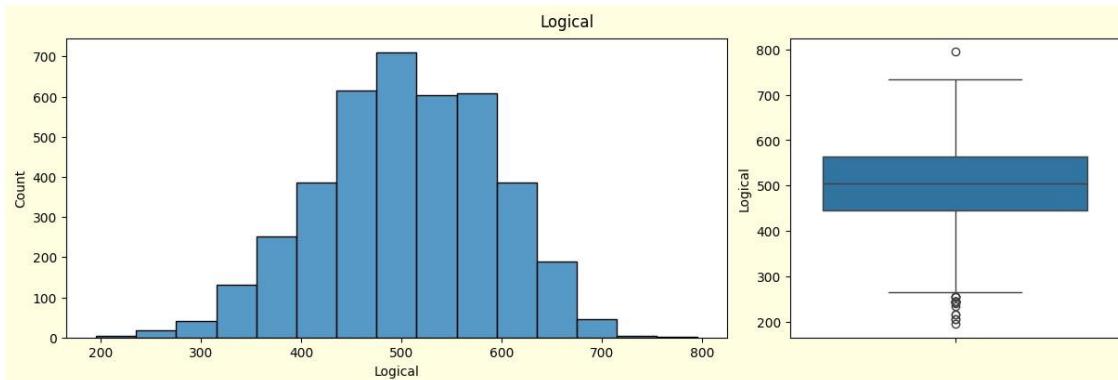
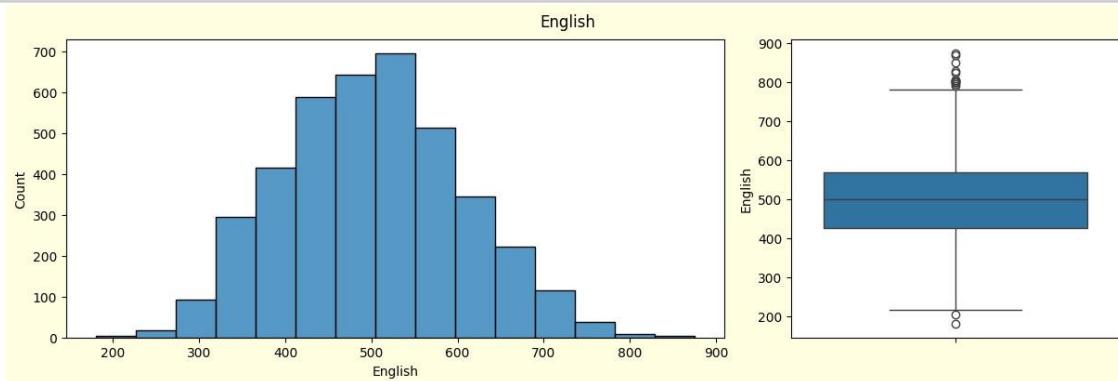
    ↪gridspec_kw={'width_ratios': 4, 2})  

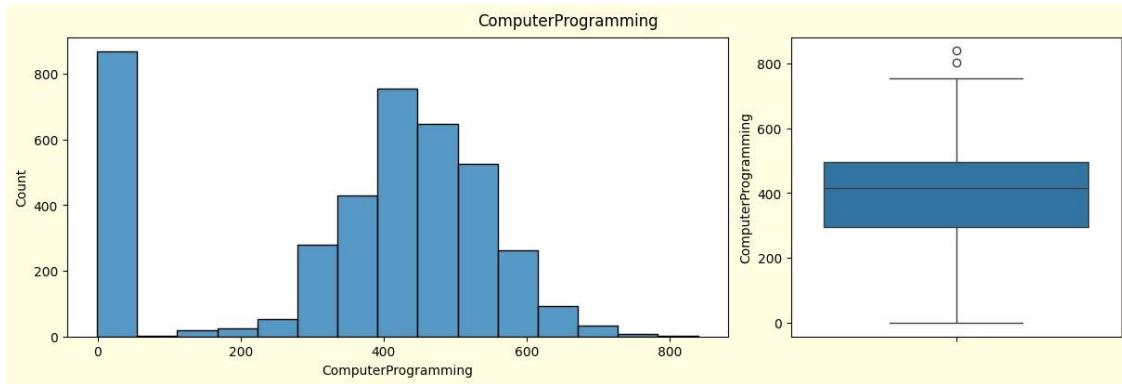
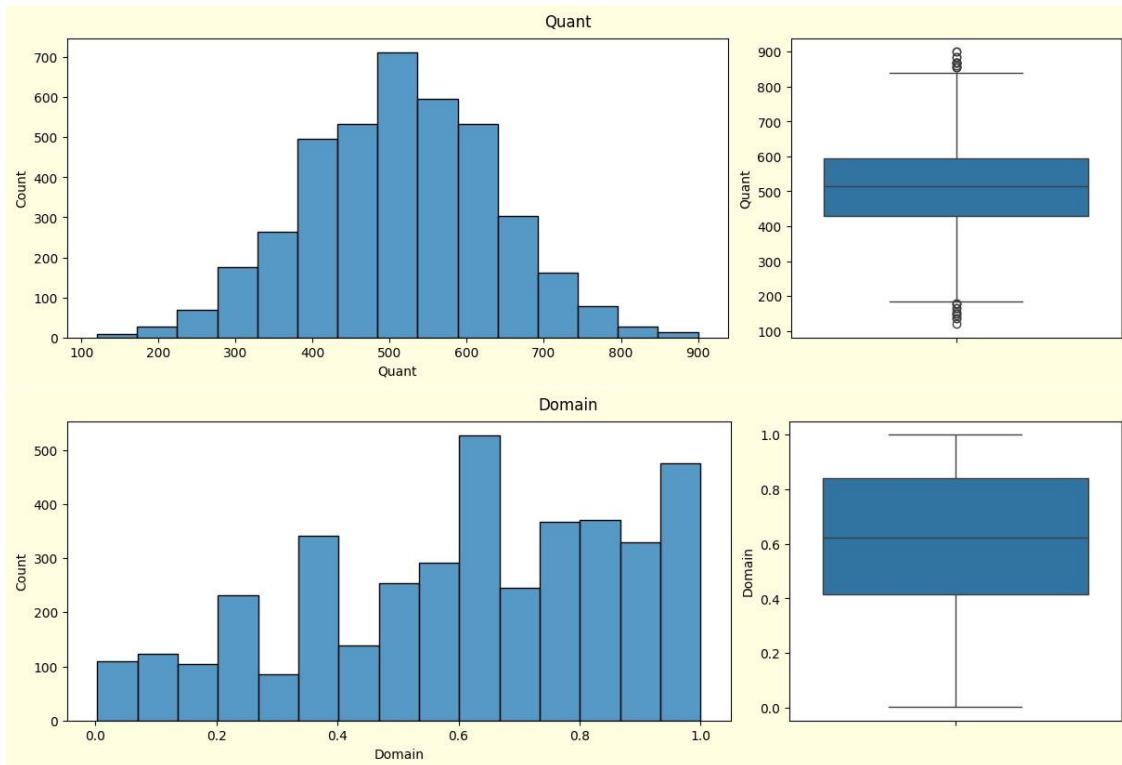
    sns.histplot(data[col], ax=ax1, bins=15)  

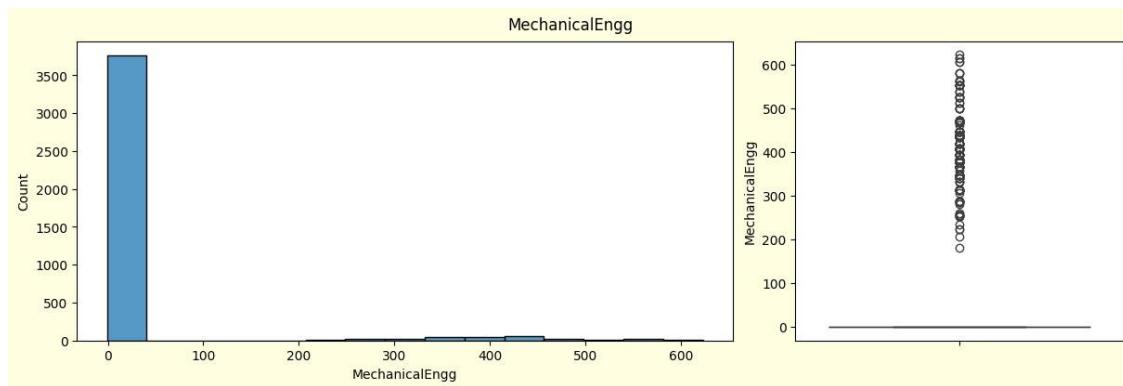
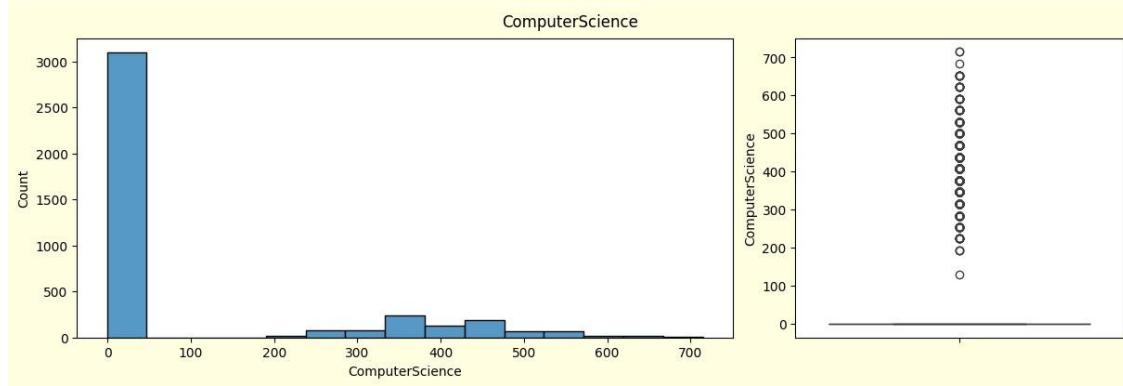
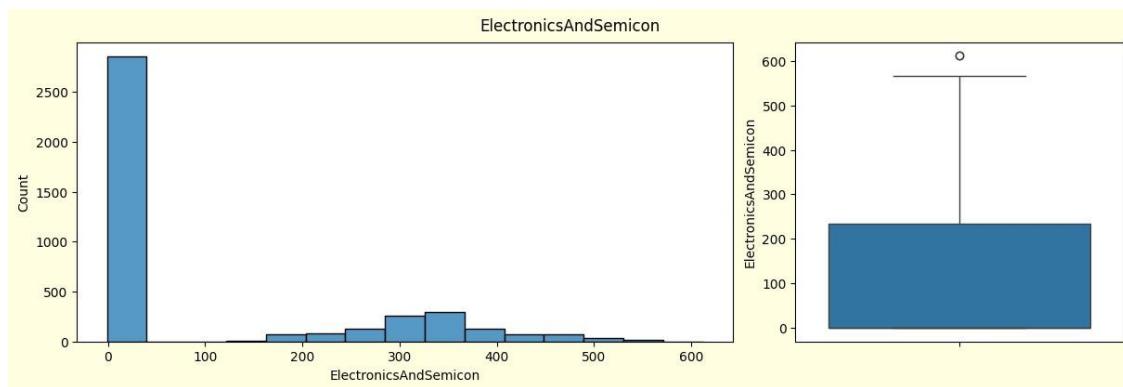
    sns.boxplot(data[col], ax=ax2)  

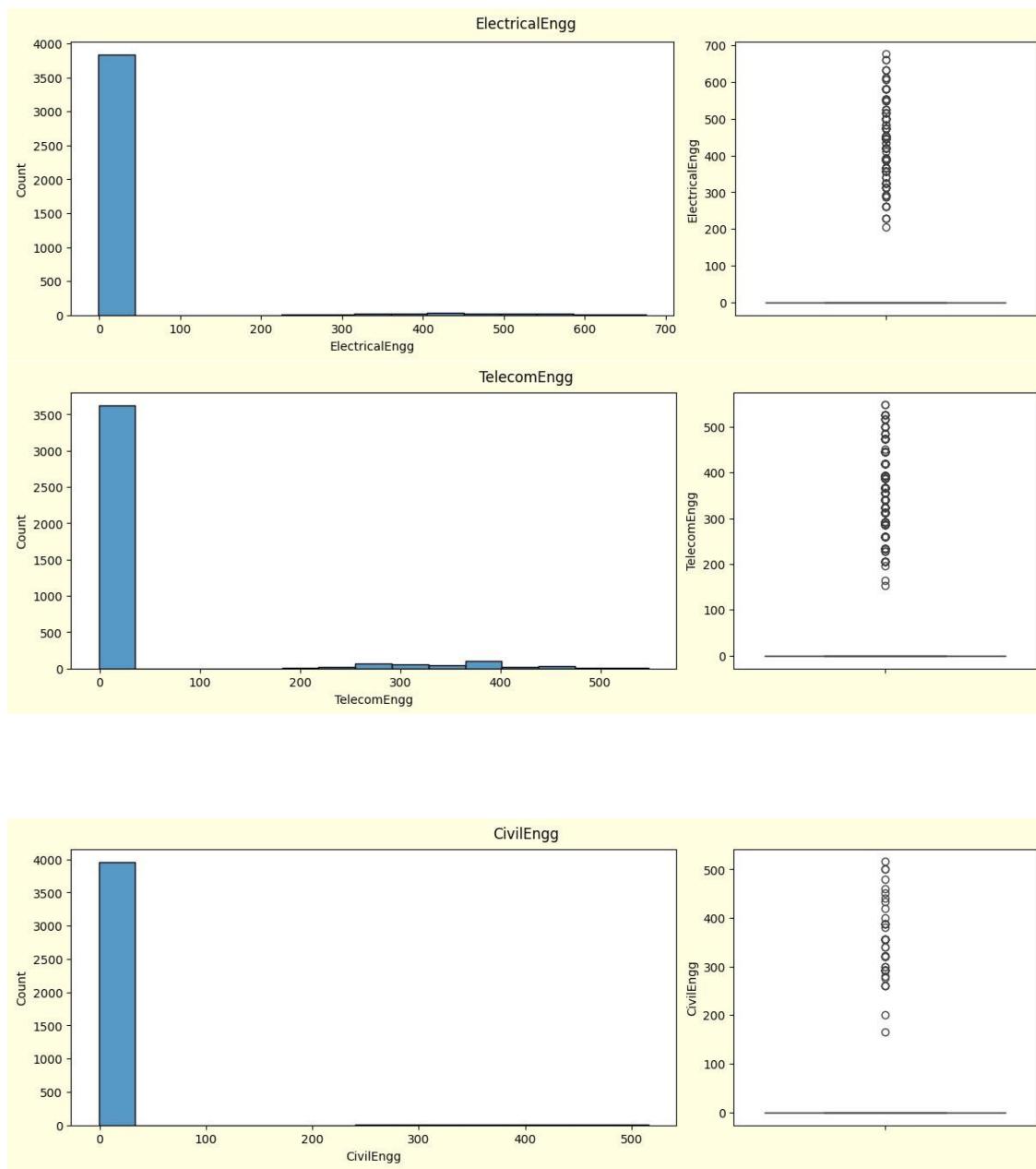
    plt.suptitle(col)  

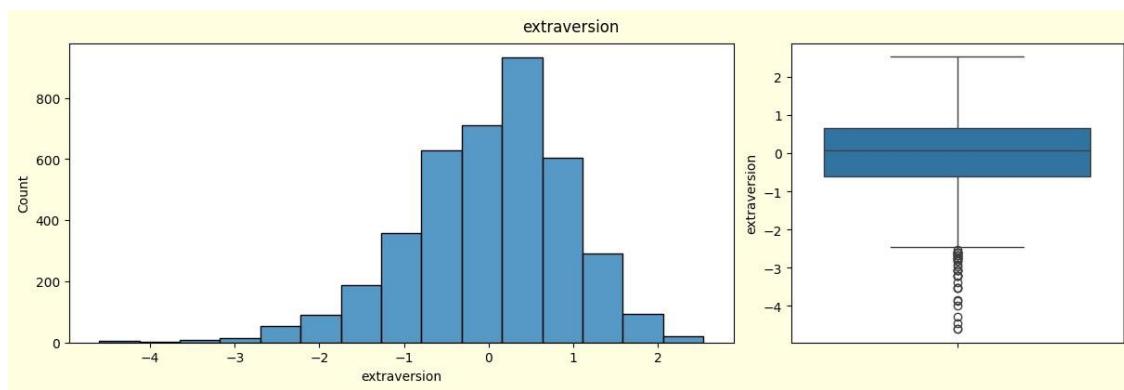
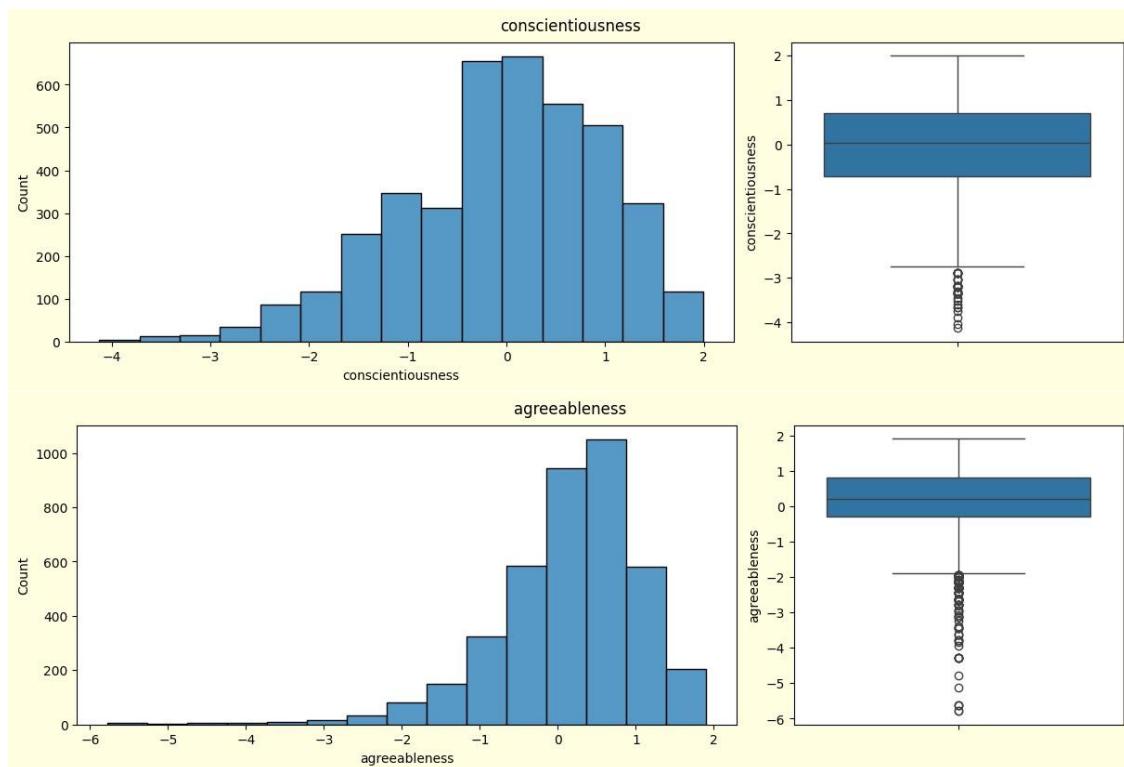
    plt.show()
```

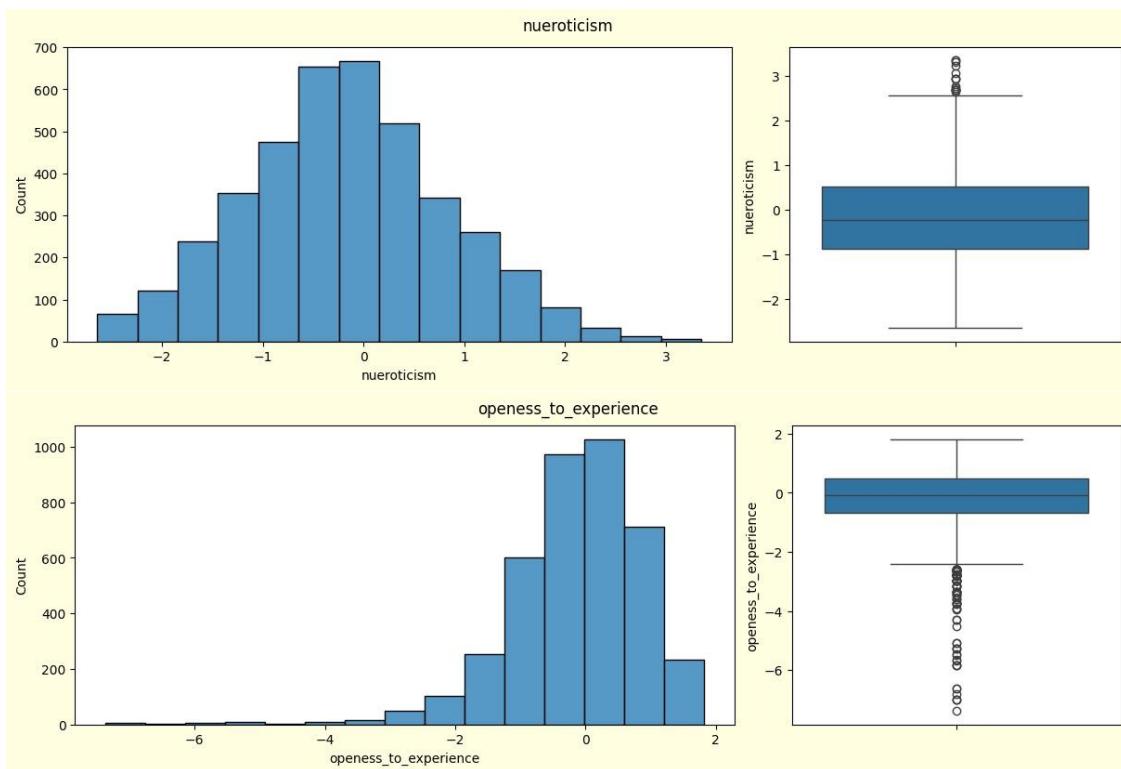








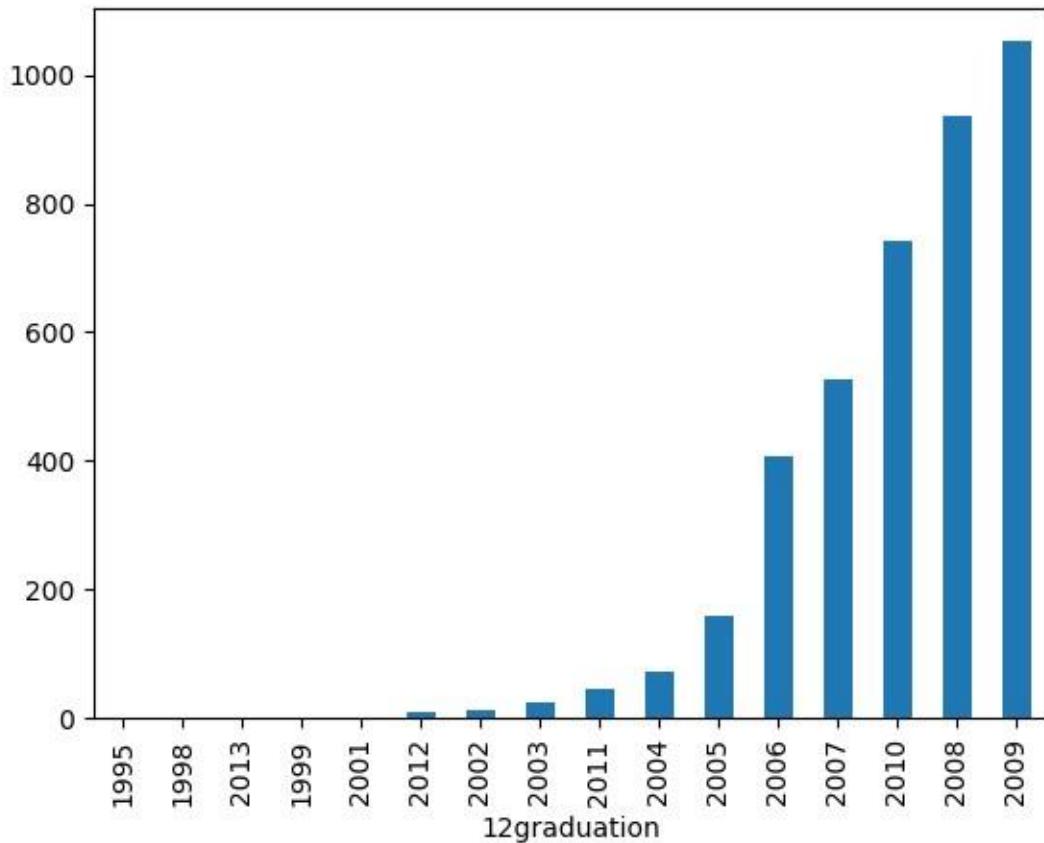




2.11 12graduation Column

```
[153]: data['12graduation'].value_counts().sort_values().plot(kind='bar')
```

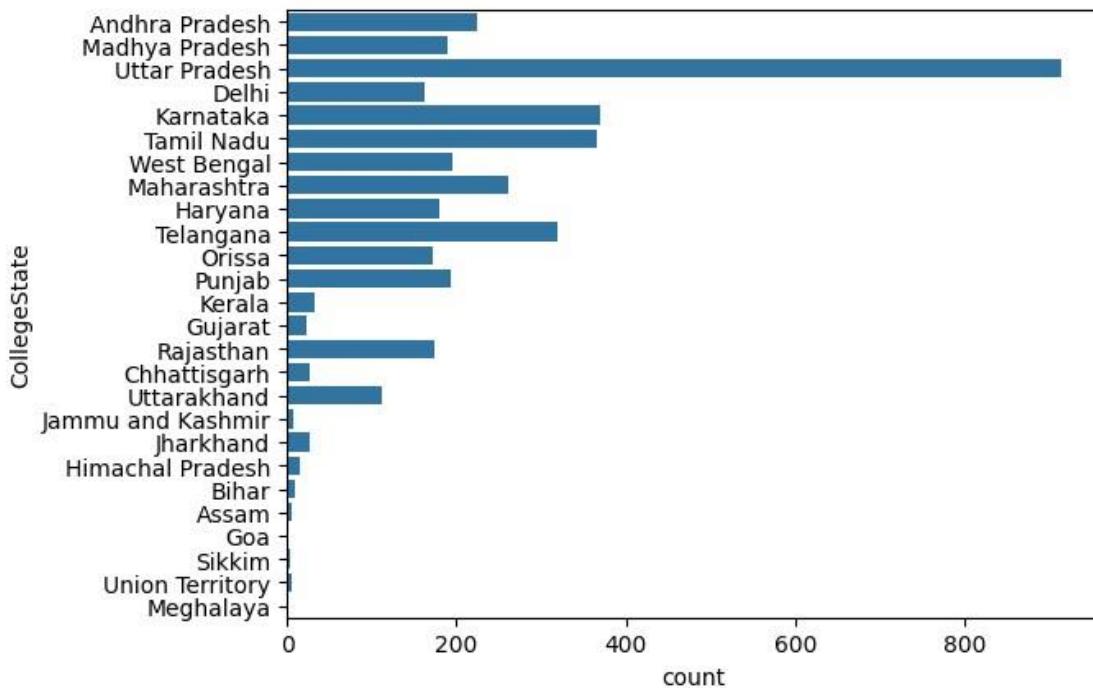
```
[153]: <AxesSubplot:xlabel='12graduation'>
```



2.12 CollegeState Column

```
[99]: sns.countplot(data['CollegeState'])
```

```
[99]: <AxesSubplot:xlabel='count', ylabel='CollegeState'>
```

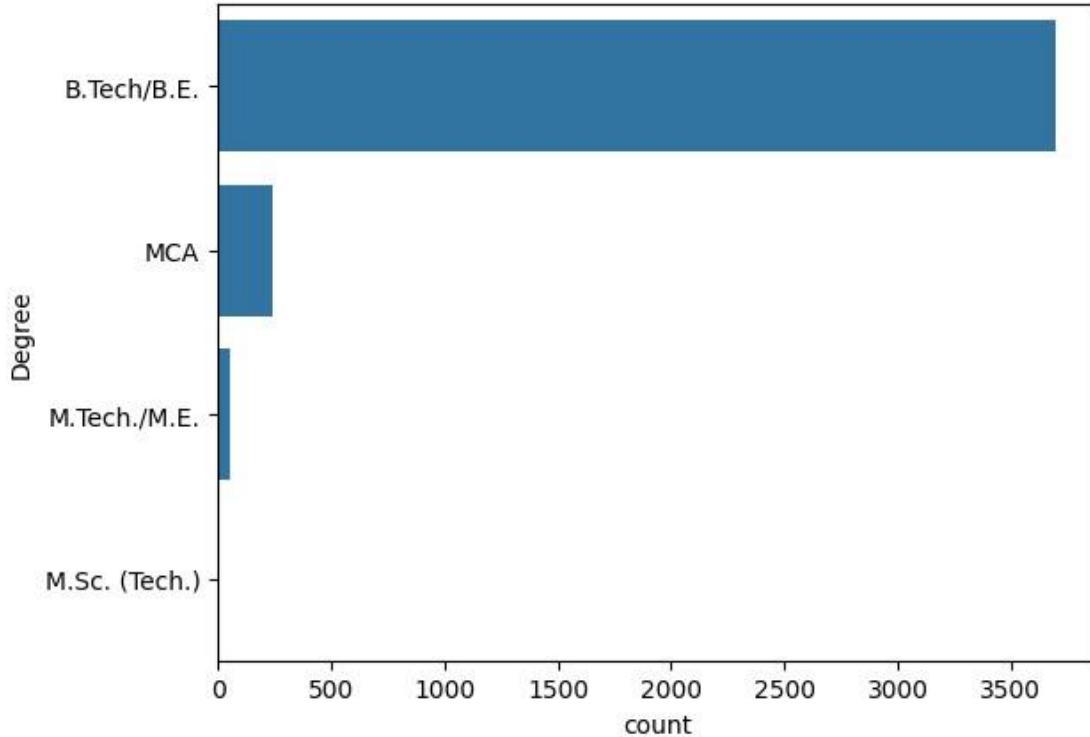


- Most of the people are from UP, next Karnataka and Tamil Nadu

2.13 Degree Column

```
[100]: sns.countplot(data['Degree'])
```

```
[100]: <AxesSubplot:xlabel='count', ylabel='Degree'>
```

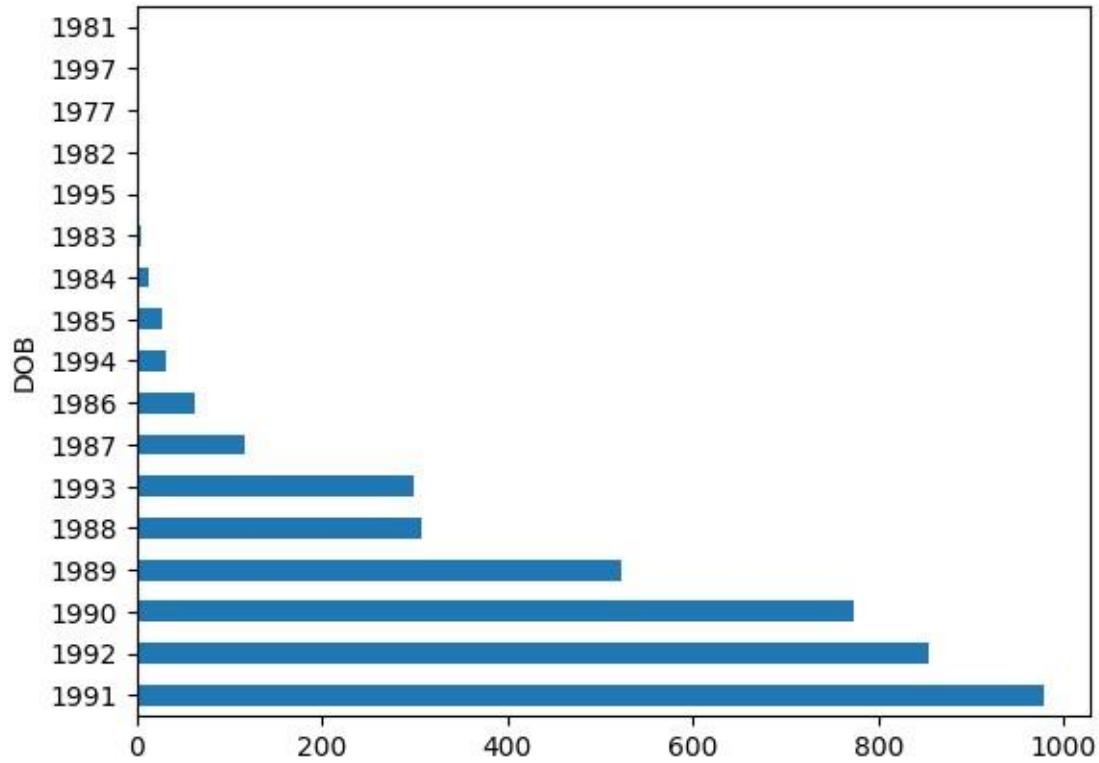


- Most of the students are from B.Tech/B.E and second most is MCA

2.14 DOB, DOJ, DOL Columns

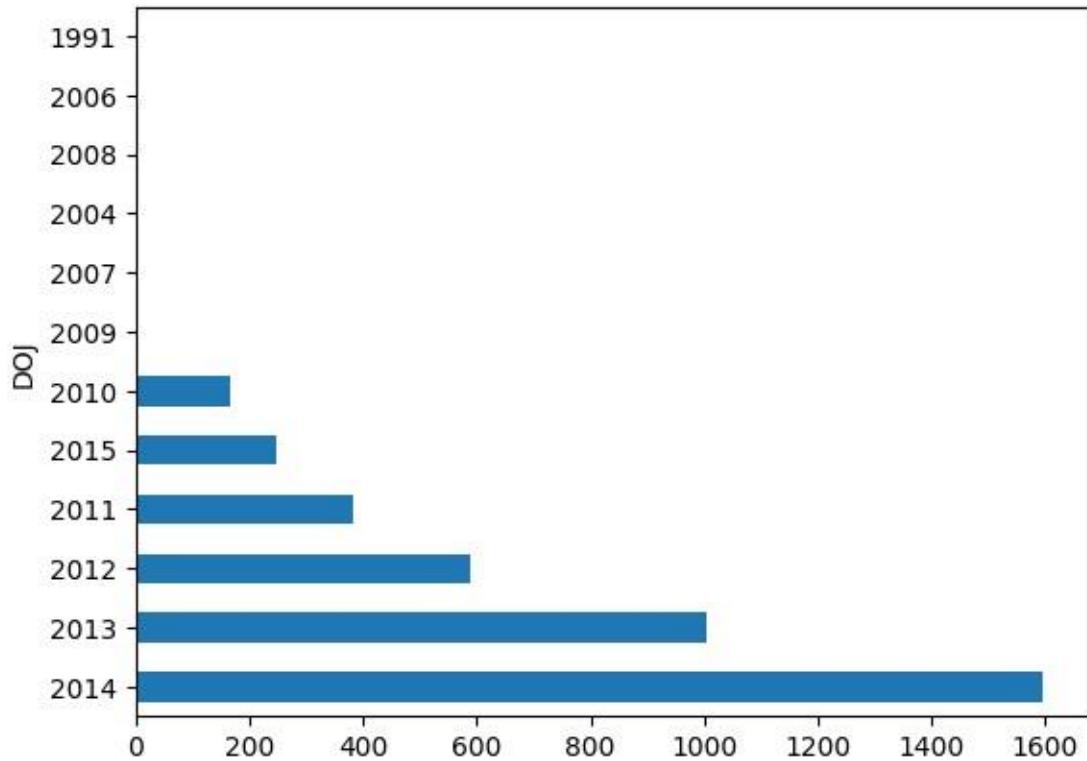
```
[101]: data['DOB'].dt.year.value_counts().plot(kind='barh')
```

```
[101]: <AxesSubplot:ylabel='DOB'>
```



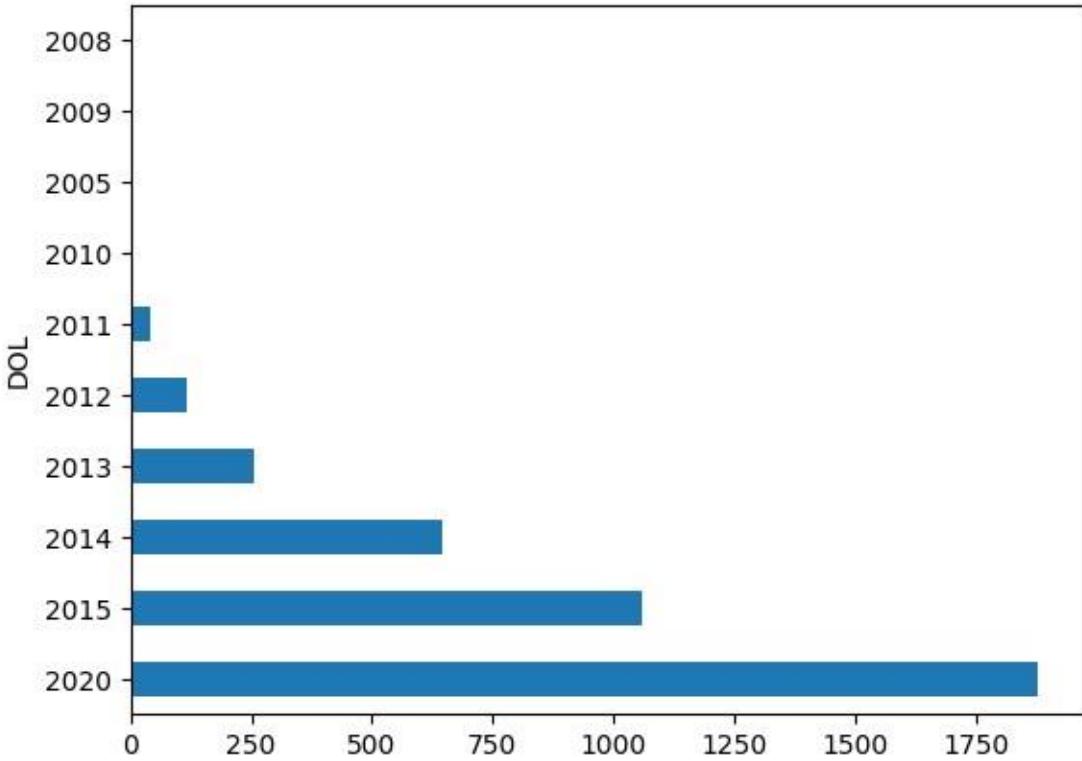
```
[102]: data['DOJ'].dt.year.value_counts().plot(kind='barh')
```

```
[102]: <AxesSubplot:ylabel='DOJ'>
```



```
[103]: data['DOL'].dt.year.value_counts().plot(kind='barh')
```

```
[103]: <AxesSubplot:ylabel='DOL'>
```



2.15 Bi Variate Analysis

[104]: data.columns

```
[104]: Index(['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',
       'openness_to_experience',
       'period', 'DOB_Year', 'GYear', '12GradAge', 'GradAge',
       'Job_Role'], dtype='object')
```

2.16 Gender vs Specialization

- cat vs cat

```
[105]: pd.crosstab(index=data['Gender'], columns=data['Specialization'])
```

```
[105]: Specialization CECSE ECE EEE MECH other
```

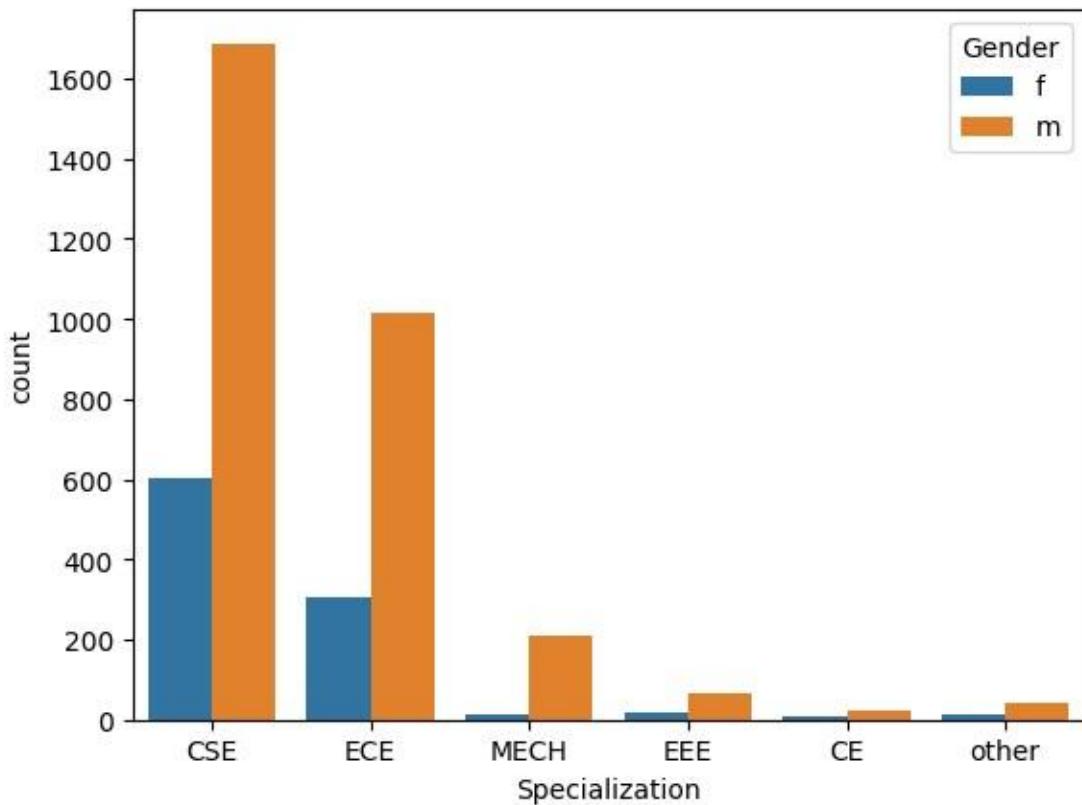
Gender

f 6 601 306 17 12 15 m 23 1688 1013 68 208

41

```
[106]: sns.countplot(data=data, x='Specialization', hue='Gender')
```

```
[106]: <AxesSubplot:xlabel='Specialization', ylabel='count'>
```



- Observation:

1. By seeing the plot most of the Male are preferring CSE as their course than other course
2. When compared to Male with Female, Male are More in every course than Female.
3. Female are more interested to opt CSE as their course as same as Male.

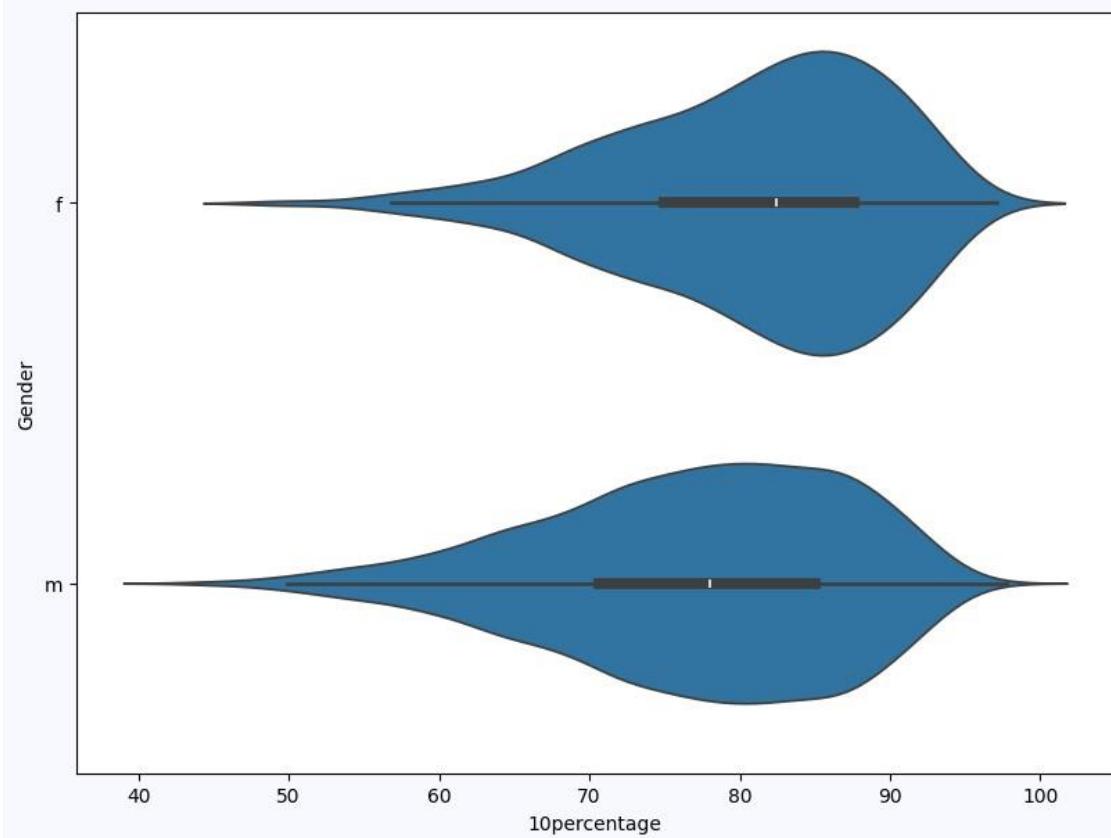
2.17 Gender vs 10percentage

```
[107]: data.groupby('Gender')['10percentage'].describe()
```

```
[107]:      count      mean       std    min    25%    50%    75%    max
Gender
f      957.0  80.932894  8.848698  48.8  75.00  82.4  87.63  97.12
m     3041.0  76.979000  9.960078  43.0  70.66  78.0  85.00  97.76
```

```
[108]: fig, ax = plt.subplots(figsize=(8, 6), constrained_layout=True, facecolor='ghostwhite')
sns.violinplot(data=data, y='Gender', x='10percentage', ax=ax)
```

```
[108]: <AxesSubplot:xlabel='10percentage', ylabel='Gender'>
```



- **Observation:**

- The two plots are right skewed, it means the students are performing in 10th exam are well.
- Comparing Male with Female, Female performance is Good (i.e; Females are performed well in the exam) their 50th percentile is more than Male

2.18 Gender vs 12percentage

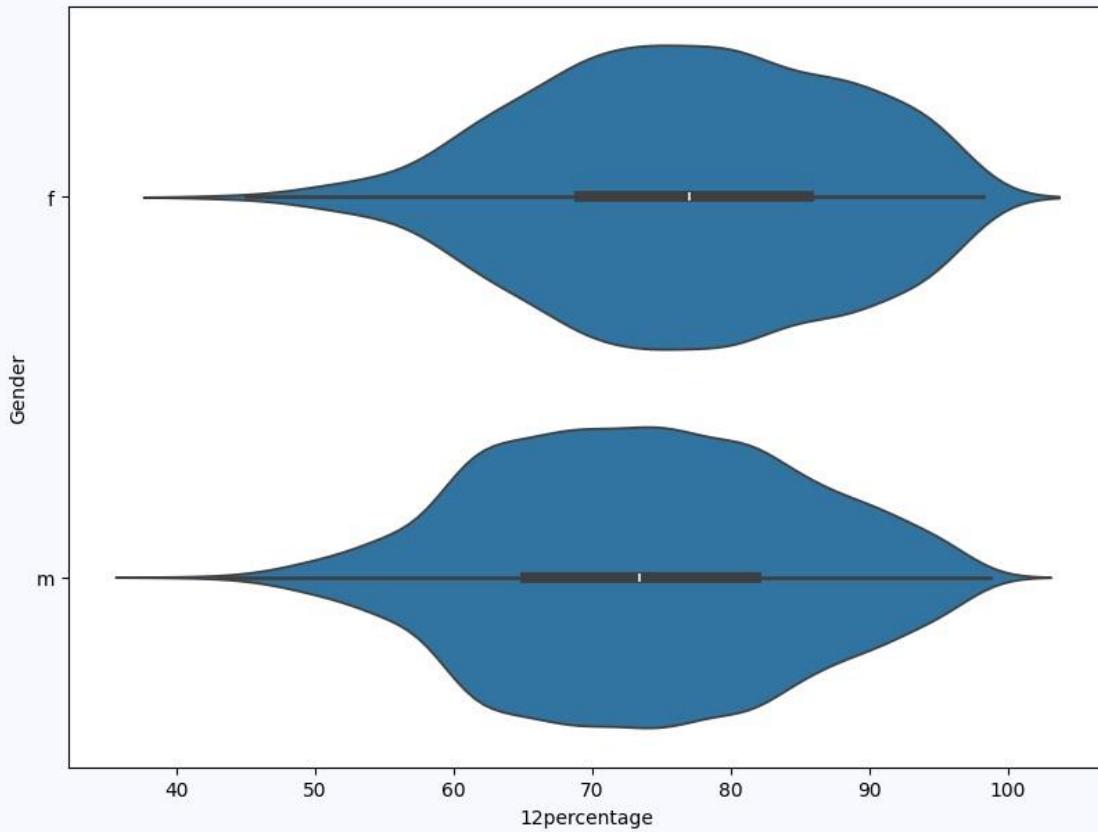
```
[109]: data.groupby('Gender')['12percentage'].describe()
```

```
[109]:      count      mean       std    min    25%    50%    75%    max
```

```
Gender
f      957.0 77.007618 10.852544 43.12 69.16 77.0 85.6 98.2
m     3041.0 73.666636 10.926044 40.00 65.20 73.4 81.8 98.7
```

```
[110]: fig, ax = plt.subplots(figsize=(8, 6), constrained_layout=True,
                           facecolor='ghostwhite')      sns.violinplot(data=data,
                           y='Gender', x='12percentage', ax=ax)
```

```
[110]: <AxesSubplot:xlabel='12percentage', ylabel='Gender'>
```



- **Observation:**

- Here also we see the same pattern as 10percentage.
- The Distribution is again Right skewed, the students perform well on 12Board exam

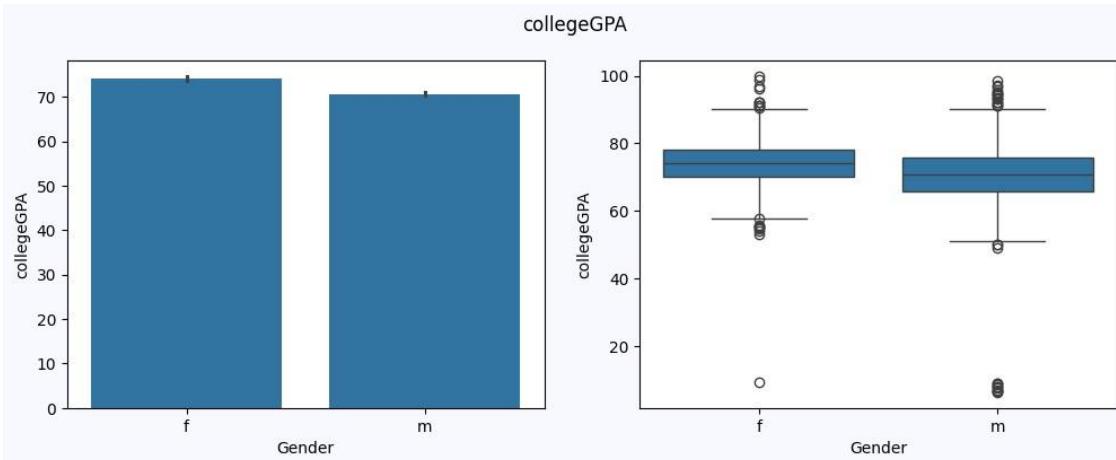
2.19 Gender vs collegeGPA

```
[111]: data.groupby('Gender')[['collegeGPA']].describe()
```

```
[111]:          count        mean         std    min     25%     50%     75%     max
Gender
```

```
f      957.0 74.048056 7.266736 9.30 70.00 74.00 78.15 99.93 m
      3041.0 70.679947 8.269352 6.45 65.63 70.66 75.70 98.40
```

```
[112]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12,4), facecolor='ghostwhite')
sns.barplot(x='Gender', y='collegeGPA', data=data, ax=ax1)
sns.boxplot(x='Gender', y='collegeGPA', data=data, ax=ax2)
plt.suptitle('collegeGPA')
plt.show()
```



• Observation:

- Male and Female performed well in the Exam.
- 25th Percentile of the Male & Female are greater than 65 GPA.
- There are some people not performed well and we consider them as outliers.
- Female have Higher GPA than Male.

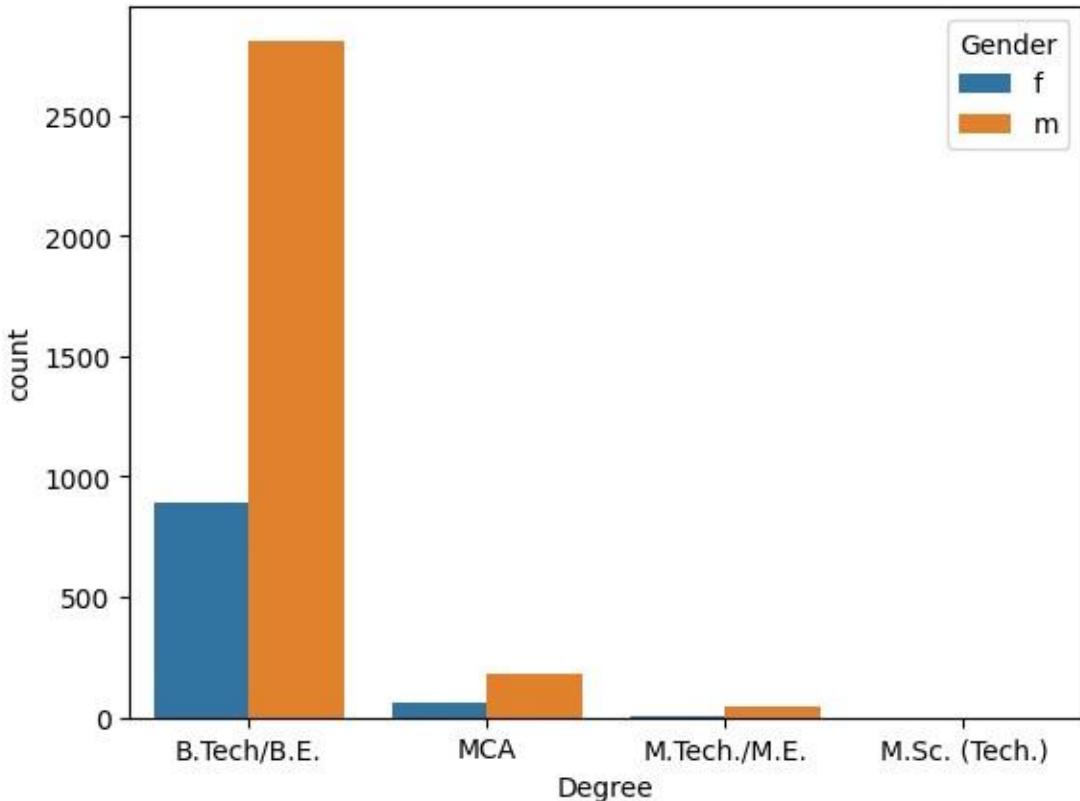
2.20 Gender vs Degree

```
[113]: pd.crosstab(index=data['Gender'], columns=data['Degree'])
```

```
[113]: Degree    B.Tech/B.E.    M.Sc. (Tech.)    M.Tech./M.E.    MCA
Gender
f        889     1       9   58    m        2811     1       44   185
```

```
[114]: sns.countplot(data=data, x='Degree', hue='Gender')
```

```
[114]: <AxesSubplot:xlabel='Degree', ylabel='count'>
```



- Observation:**

- Both Male and Female are opted their Degree course as B.Tech/B.E

2.21 Gender vs Job_Role

```
[115]: pd.crosstab(index=data['Gender'], columns=data['Job_Role'])
```

```
[115]:   Job_Role Administrator Analyst Developer Hardware Engineer Manager \
Gender
      f      55      125     185     64      29      m    142     286     568     402     131
```

Job_Role	Other	Software Engineer	System Engineer	Test Engineer
Gender				

f	112	219	119	49
m	287	722	372	131

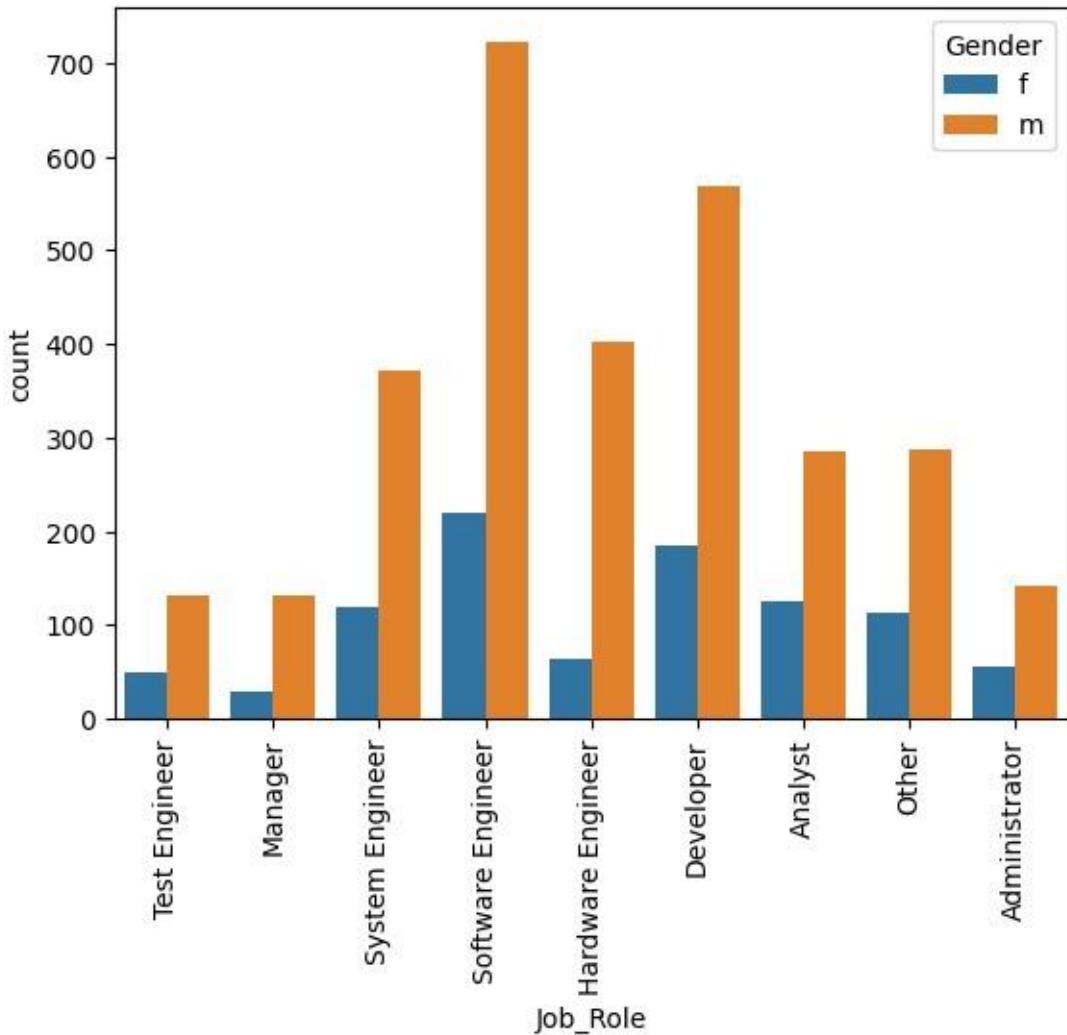
```
[116]:     sns.countplot(data=data,      x='Job_Role',      hue='Gender')
plt.xticks(rotation=90)
```

```
[116]: ([0, 1, 2, 3, 4, 5, 6, 7, 8],
[Text(0, 0, ''),
Text(0, 0, '')],
```

```

Text(0, 0, ''),
Text(0, 0, '')])

```



- **Observation:**

- Most of the people are doing job as Software Engineer, Developer and -----

2.22 Gender vs Salary

```
[117]: data.groupby('Gender')['Salary'].describe()
```

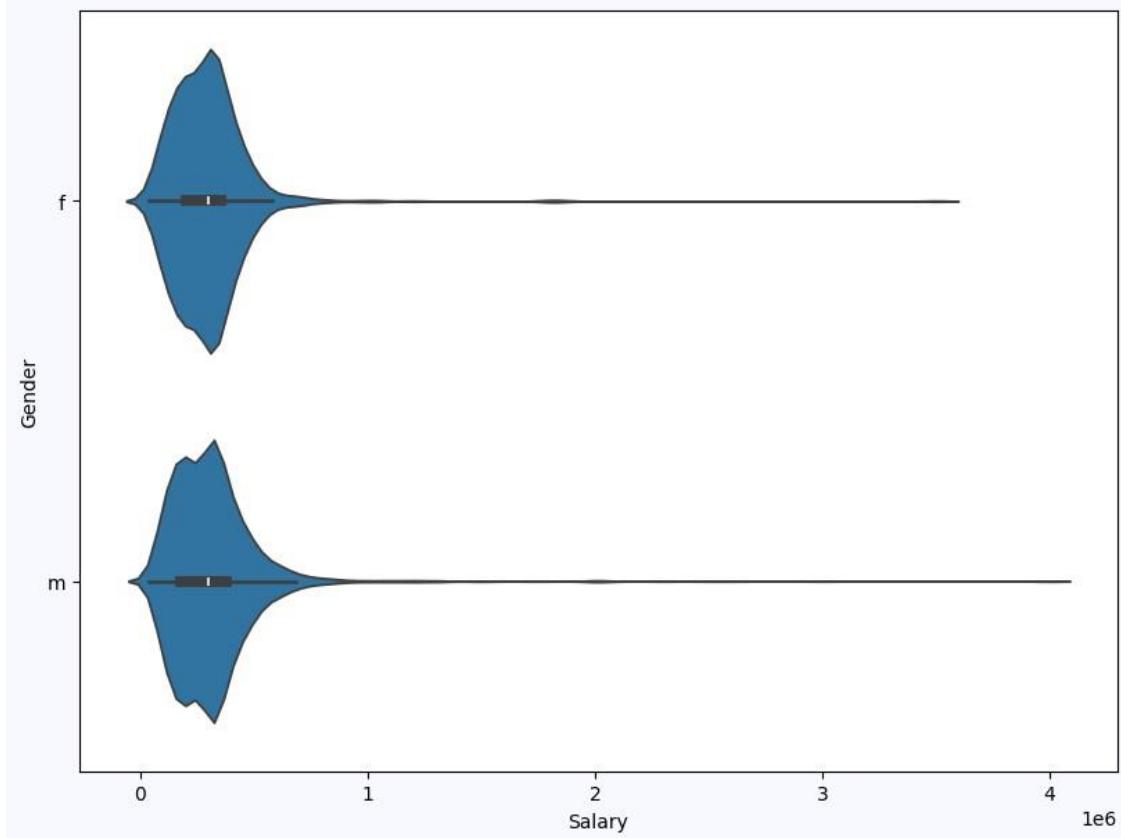
```
[117]:      count        mean         std        min      25%      50% \
Gender
```

```
f      957.0 294937.304075 191031.789503 35000.0 200000.0 300000.0
m     3041.0 311716.211772 218999.537438 35000.0 180000.0
                  300000.0
```

```
    75%          max
Gender
f      355000.0 3500000.0 m
      380000.0 4000000.0
```

```
[118]: fig, ax = plt.subplots(figsize=(8, 6), constrained_layout=True,
                           facecolor='ghostwhite')
sns.violinplot(data=data, y='Gender', x='Salary',
                 ax=ax)
```

```
[118]: <AxesSubplot:xlabel='Salary', ylabel='Gender'>
```

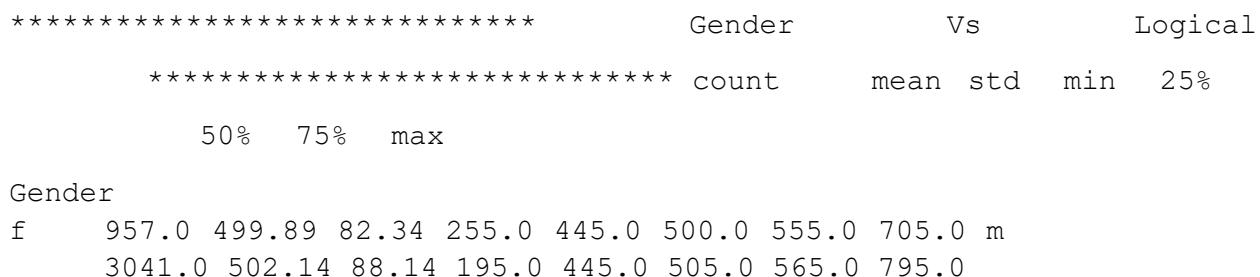
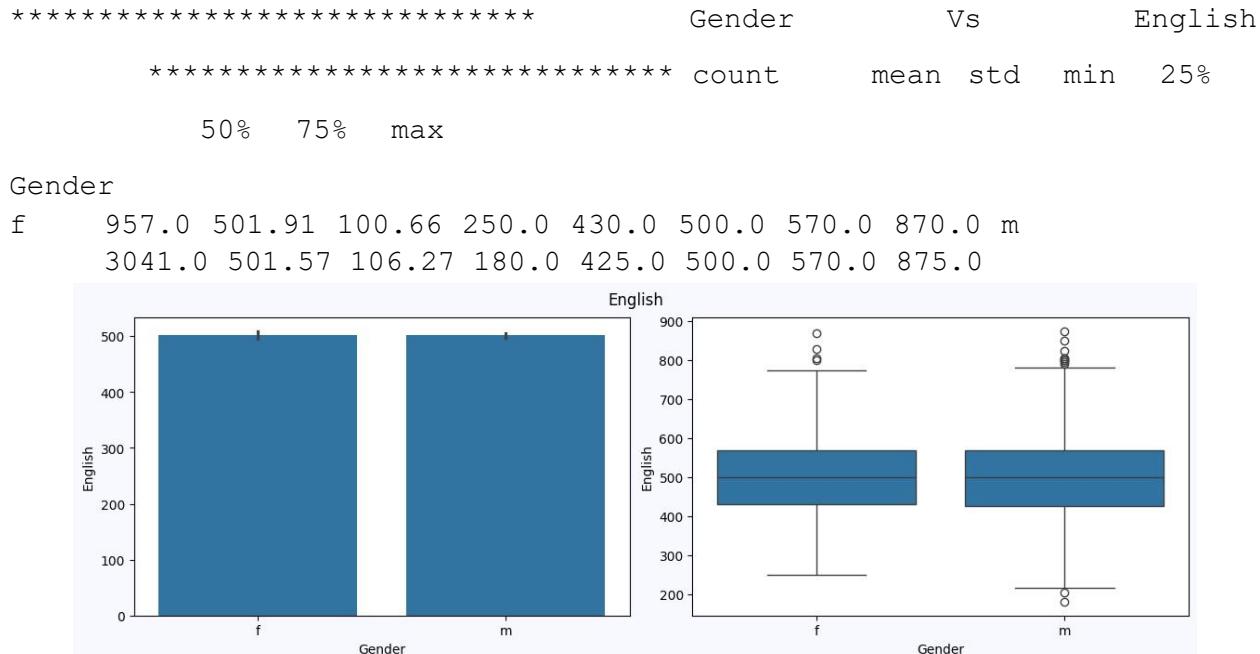


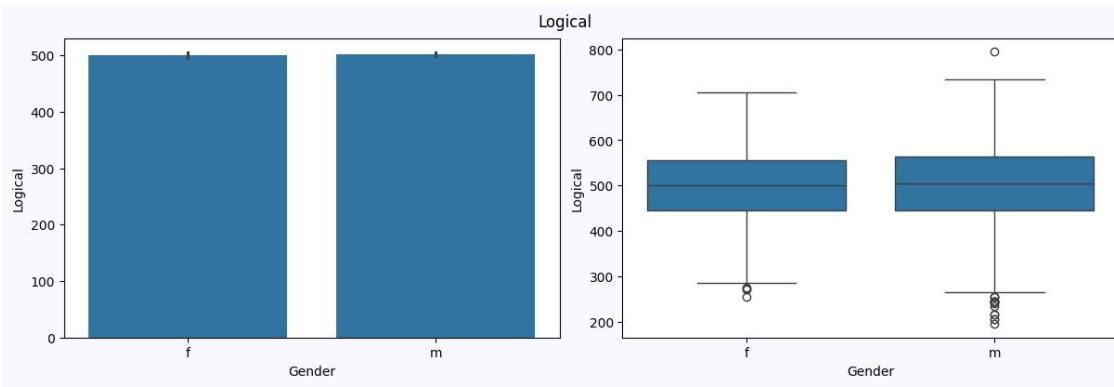
- **Observation:**

- There is a minute difference in the median salary between Female and Male, but it is hard to say whether this difference is significant or not.

2.23 Gender vs AMCAT scores

```
[119]: for col in amcat_scores: print('*'*30, 'Gender
Vs', col, '*'*30)
display(data.groupby('Gender')[col].describe().round(2))
fig, (ax1, ax2) = plt.subplots(1, 2,
figsize=(12, 4), __
facecolor='ghostwhite',
constrained_layout=True)
sns.barplot(x='Gender', y=col, data=data,
ax=ax1) sns.boxplot(x='Gender', y=col,
data=data, ax=ax2)
plt.suptitle(col)
plt.show() print()
```

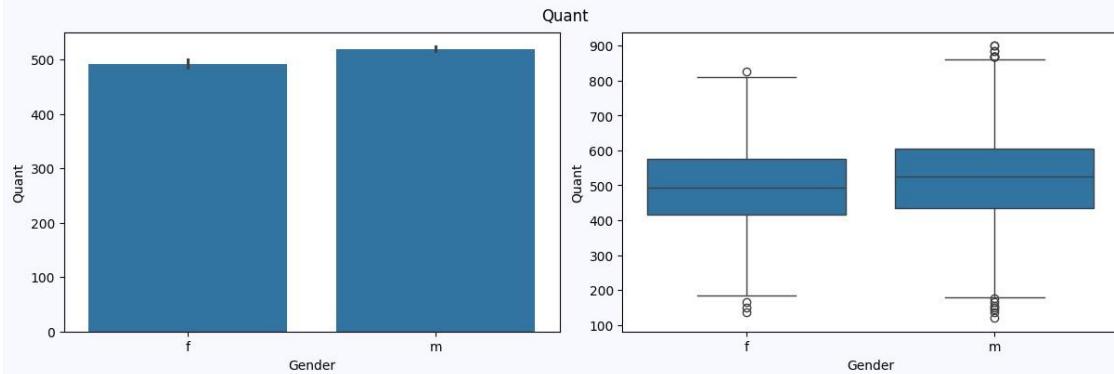




```
*****
Gender          Vs       Quant
***** count      mean   std   min   25%
50%    75%    max
```

Gender

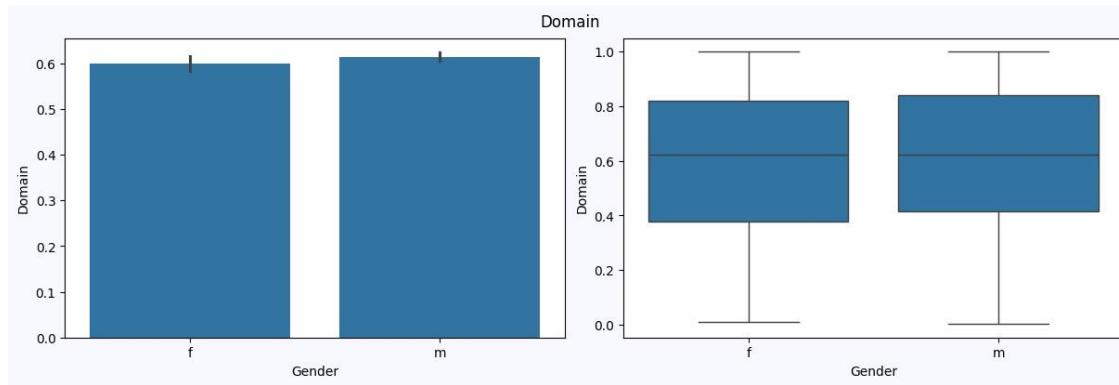
	957.0	492.36	116.25	135.0	415.0	494.0	575.0	825.0	m
f	3041.0	519.99	123.43	120.0	435.0	525.0	605.0	900.0	



```
*****
Gender          Vs       Domain
***** count   mean  std   min   25%   50%
75%    max
```

Gender

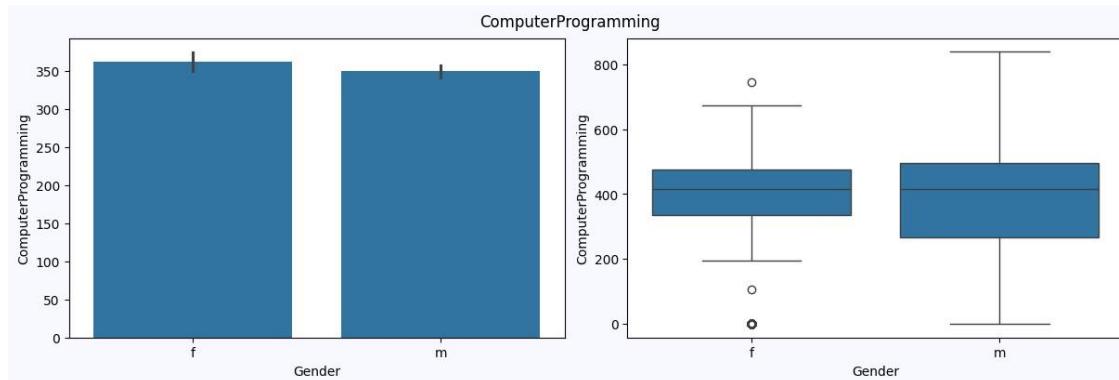
	957.0	0.60	0.26	0.01	0.38	0.62	0.82	1.0	m
f	3041.0	0.61	0.27	0.00	0.41	0.62	0.84	1.0	



***** Gender Vs *****

ComputerProgramming

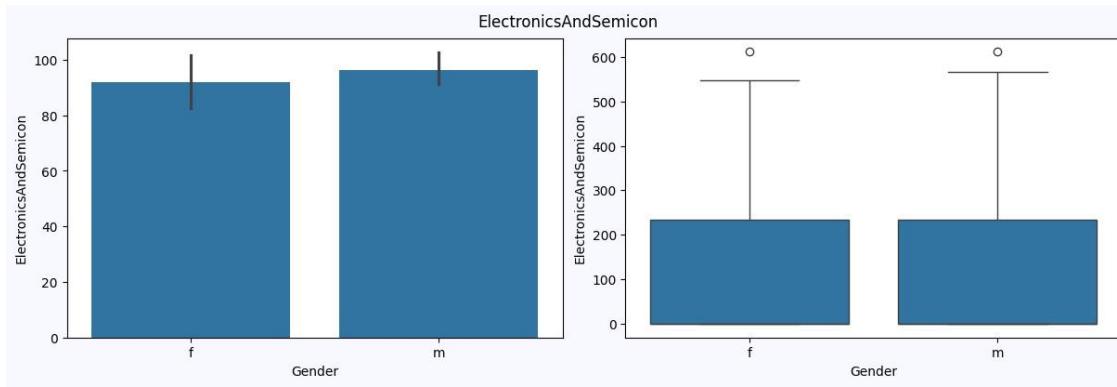
	count	mean	std	min	25%	50%	75%	max
Gender								
f	957.0	362.54	185.67	-1.0	335.0	415.0	475.0	745.0
m	3041.0	350.13	211.11	-1.0	265.0	415.0	495.0	840.0



***** Gender Vs *****

ElectronicsAndSemicon

m

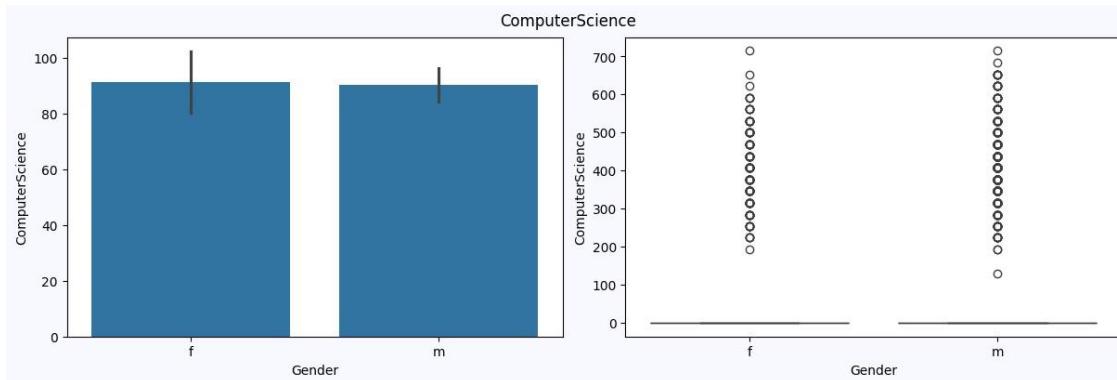


```

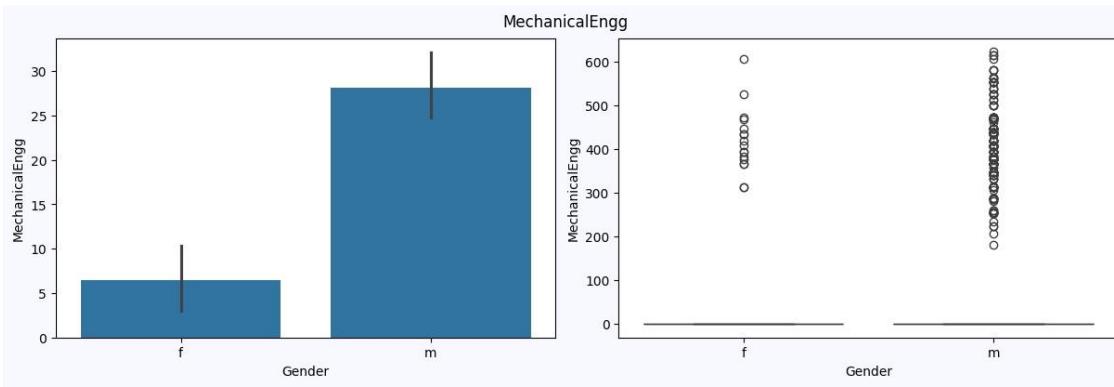
count  mean      std min 25% 50%    75%    max
Gender
f      957.0 91.79 154.41 -1.0 -1.0 -1.0 -1.0 233.0
612.0 3041.0 96.44 159.44 -1.0 -1.0 -1.0 -1.0 233.0
612.0 **** Gender Vs
***** ComputerScience
*****
```

```

count  mean      std min 25% 50% 75%    max
Gender
f      957.0 91.5 174.39 -1.0 -1.0 -1.0 -1.0 715.0
m      3041.0 90.5 175.58 -1.0 -1.0 -1.0 -1.0 715.0
```



m



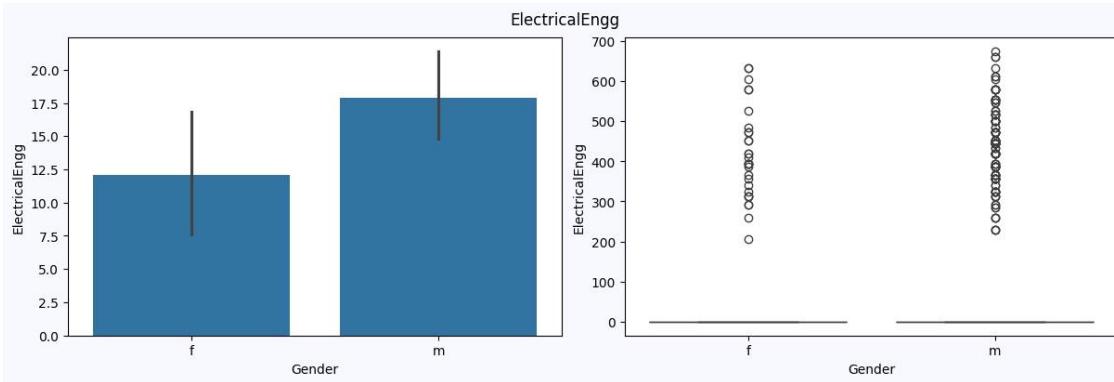
***** Gender Vs MechanicalEngg

	count	mean	std	min	25%	50%	75%	max	
Gender									
f	957.0	6.46	56.30	-1.0	-1.0	-1.0	-1.0	606.0	3041.0
	28.17	107.47	-1.0	-1.0	-1.0	-1.0	-1.0	623.0	

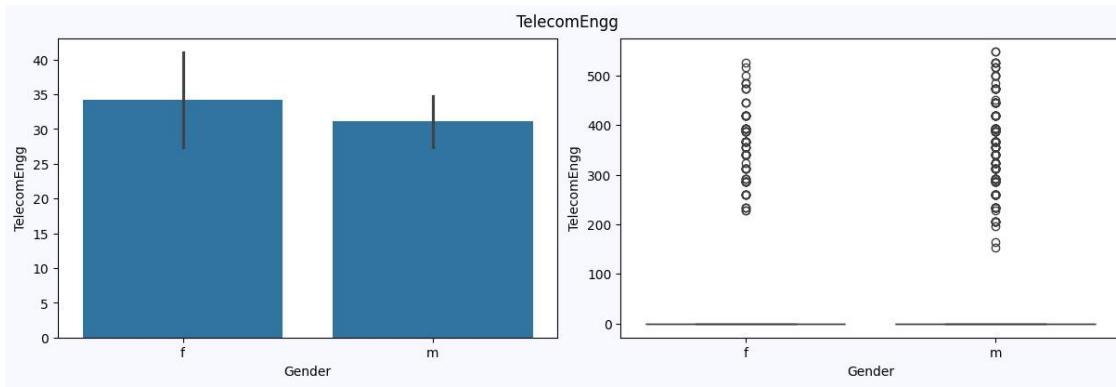
									Gender Vs

ElectricalEngg

	count	mean	std	min	25%	50%	75%	max	
Gender									
f	957.0	12.07	75.27	-1.0	-1.0	-1.0	-1.0	633.0	
m	3041.0	17.87	91.09	-1.0	-1.0	-1.0	-1.0	676.0	



m



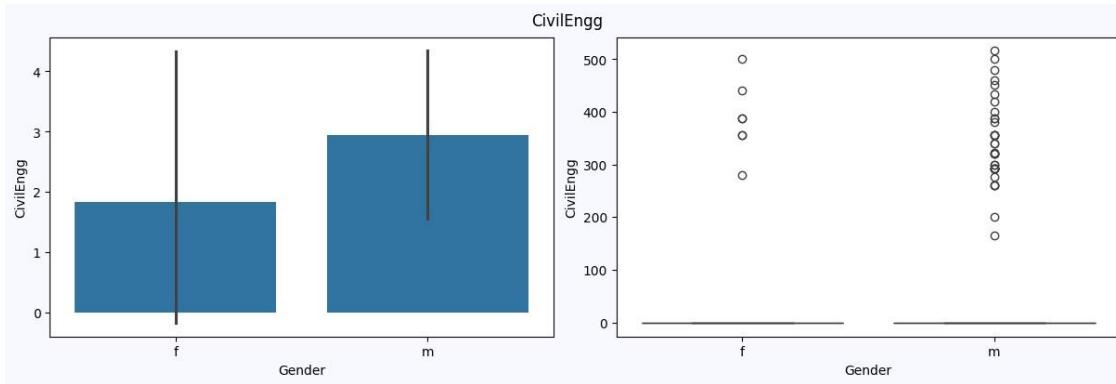
***** Gender Vs TelecomEngg

	count	mean	std	min	25%	50%	75%	max	
Gender									
f	957.0	34.24	109.11	-1.0	-1.0	-1.0	-1.0	526.0	3041.0
	31.10	103.48	-1.0	-1.0	-1.0	-1.0	-1.0	548.0	

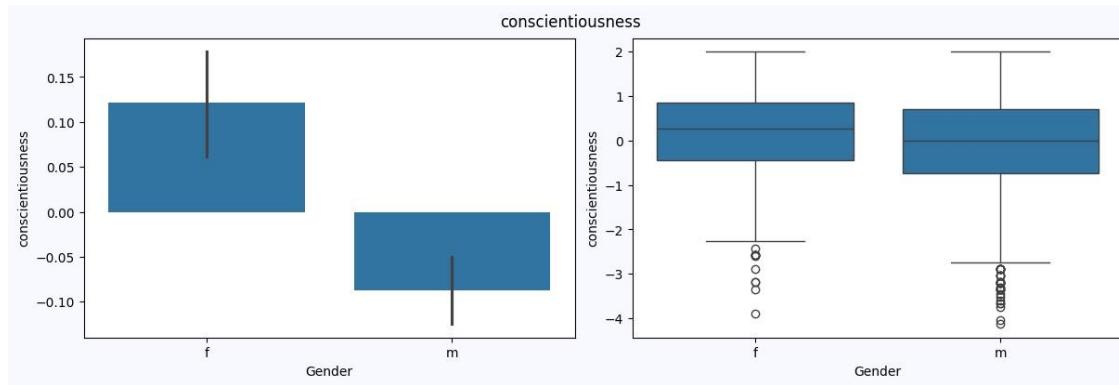
***** Gender Vs

CivilEngg

	count	mean	std	min	25%	50%	75%	max	
Gender									
f	957.0	1.84	33.52	-1.0	-1.0	-1.0	-1.0	500.0	
m	3041.0	2.95	37.59	-1.0	-1.0	-1.0	-1.0	516.0	



m



***** Gender Vs *****

conscientiousness

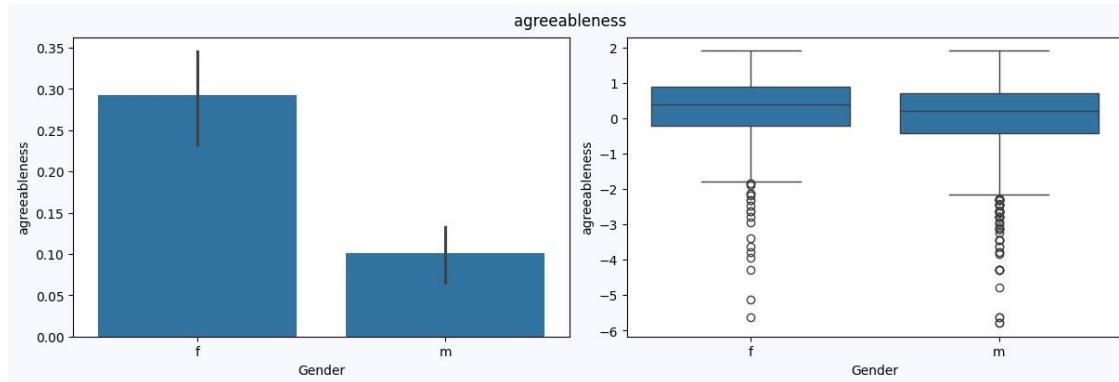
	count	mean	std	min	25%	50%	75%	max
Gender								
f	957.0	0.12	0.95	-3.89	-0.45	0.27	0.85	2.0

0.09 1.05 -4.13 -0.73 -0.02 0.70 2.0
***** Gender Vs *****

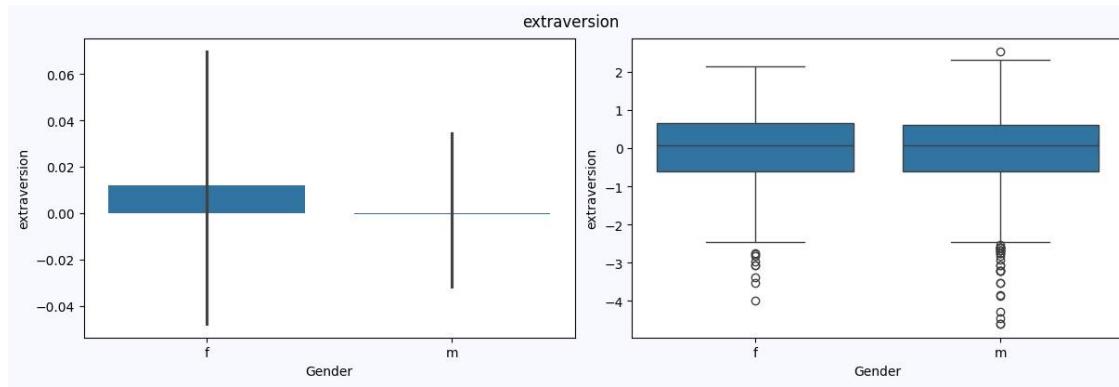
agreeableness

	count	mean	std	min	25%	50%	75%	max
Gender								
f	957.0	0.29	0.89	-5.62	-0.20	0.38	0.88	1.9

m 3041.0 0.10 0.95 -5.78 -0.44 0.21 0.71 1.9



m

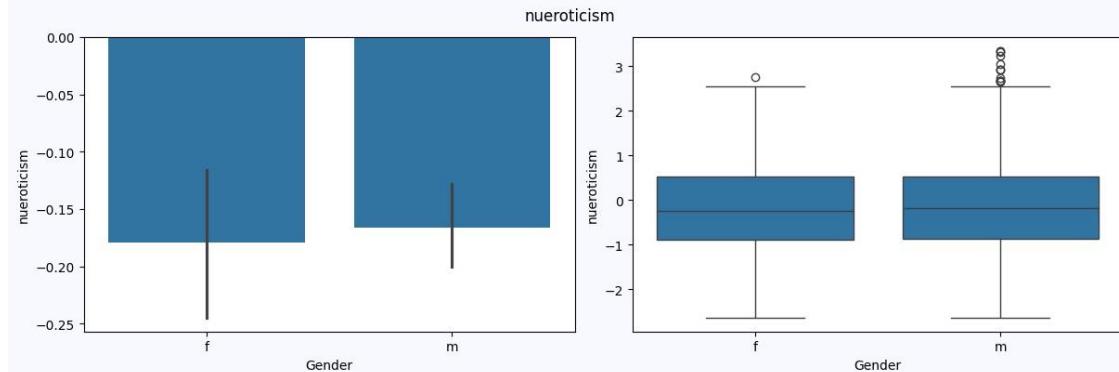


***** Gender Vs extraversion

	count	mean	std	min	25%	50%	75%	max
Gender								
f	957.0	0.01	0.94	-3.99	-0.6	0.09	0.67	2.16
m	3041.0	-0.00	0.95	-4.60	-0.6	0.09	0.62	2.54

***** Gender Vs
nuerotism

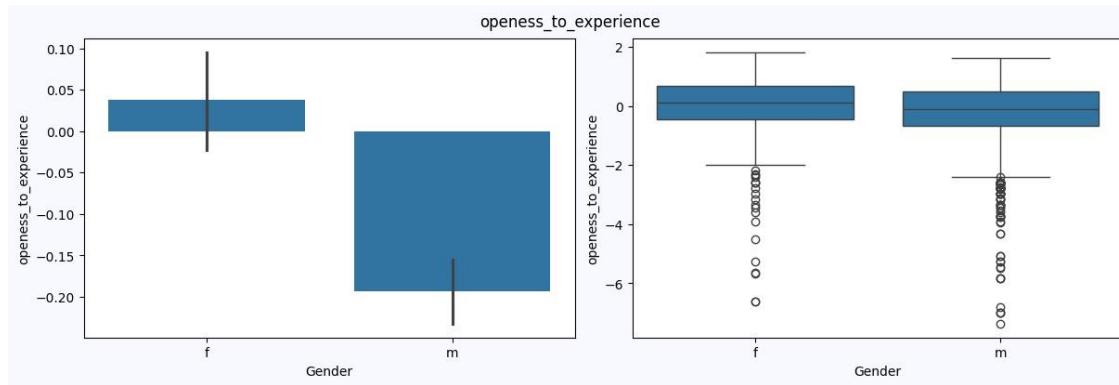
	count	mean	std	min	25%	50%	75%	max
Gender								
f	957.0	-0.18	1.01	-2.64	-0.88	-0.23	0.53	2.76
m	3041.0	-0.17	1.01	-2.64	-0.87	-0.17	0.53	3.35



m

```
*****  
openness_to_experience  
*****  
  
count mean std min 25% 50% 75% max  
Gender  
f 957.0 0.04 0.94 -6.61 -0.45 0.10 0.67 1.82 3041.0 -  
0.19 1.02 -7.38 -0.67 -0.09 0.48 1.63
```

m



- **Observation:** | Field | Gender Score Differences |
 - | English | Minor || Logical | Minor || Quantitative | Minor || Domain | Minor || Computer Programming | Minor || Electronics and Semiconductor Eng. | Minor || Computer Science | Minor || Mechanical Engineering | Minor || Telecommunication Engineering | Minor
 - | Civil Engineering | Women outperform men || Electrical Engineering | Men tend to excel

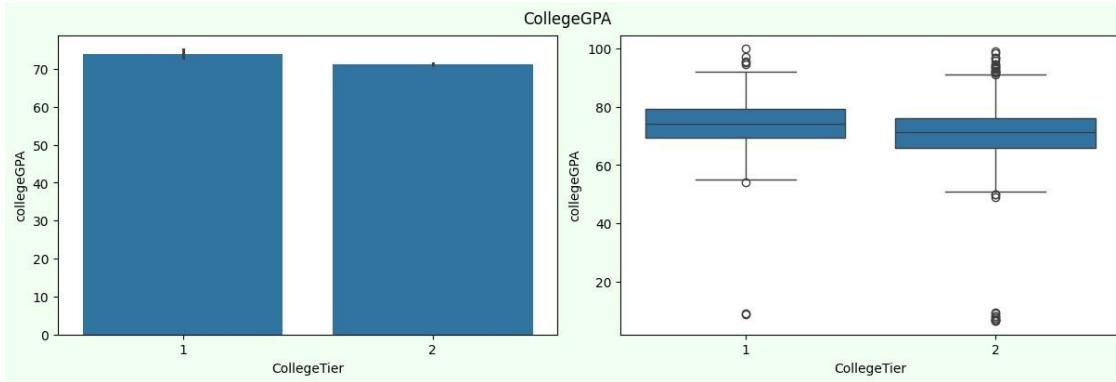
2.24 CollegeTier vs CollegeGPA

```
[120]: data.groupby('CollegeTier')[['collegeGPA']].describe()
```

```
[120]:
```

	count	mean	std	min	25%	50%	75%	max
CollegeTier								
1	297.0	73.987845	9.503475	8.58	69.40	74.30	79.29	99.93
2	3701.0	71.285415	8.018223	6.45	66.02	71.36	76.08	99.00

```
[121]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4),
constrained_layout=True,
    facecolor='honeydew') sns.barplot(x='CollegeTier',
y='collegeGPA', data=df, ax=ax1)
sns.boxplot(x='CollegeTier', y='collegeGPA',
data=df, ax=ax2) plt.suptitle('CollegeGPA')
plt.show()
```



- **Observation:**

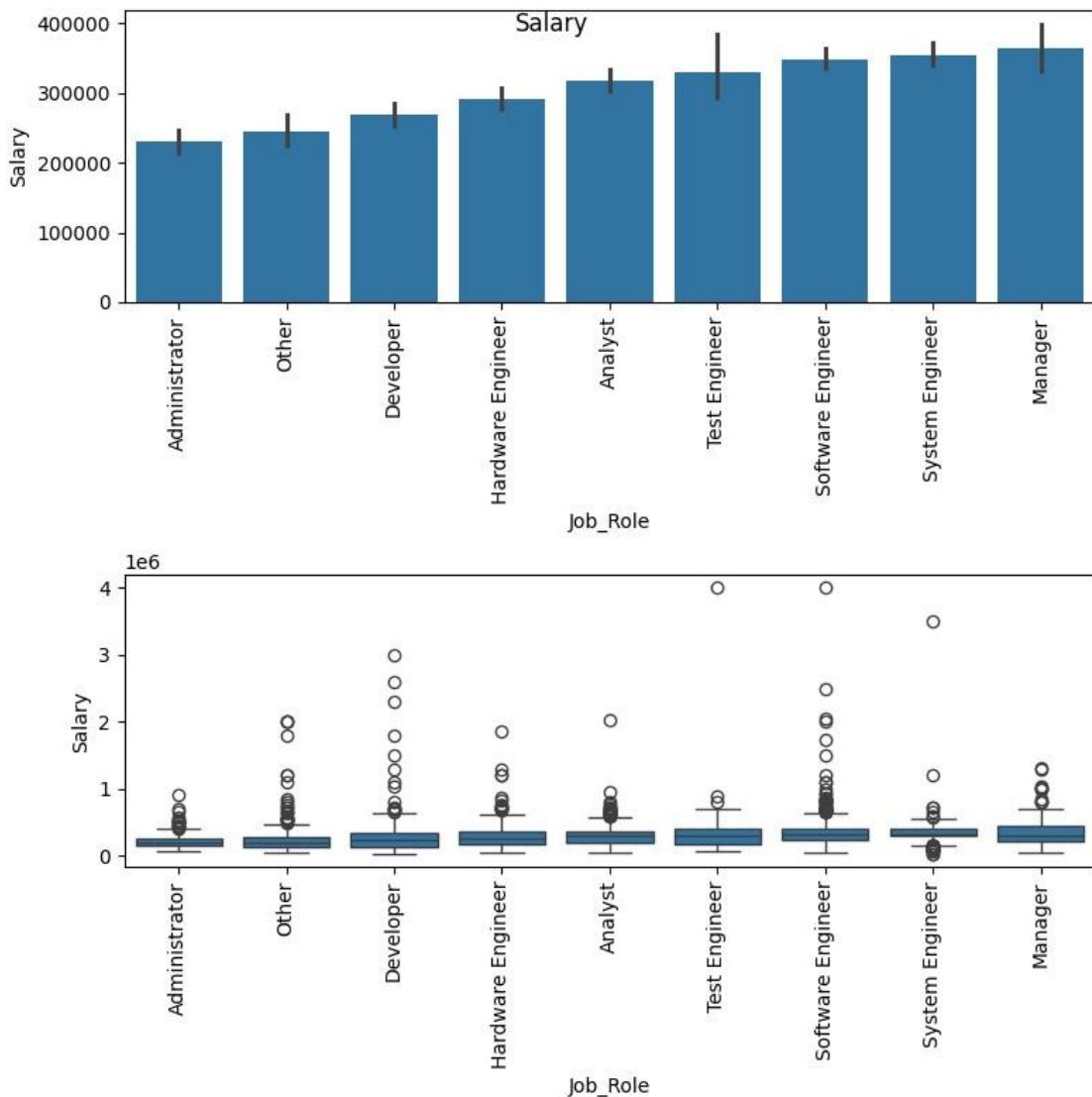
- Tier-1 college students Performance is slightly high compared to Tire-2

2.25 Salary vs Job

```
[122]: data.groupby('Job_Role')['Salary'].describe().round(2).sort_values('mean')
```

Job_Role	count	mean	std	min	25%	50%	75%	max
Administrator	197.0	230177.66	120526.46	60000.0	150000.0	200000.0	255000.0	910000.0
Other	399.0	244837.09	227884.44	40000.0	142500.0	200000.0	275000.0	2000000.0
Developer	753.0	267994.69	221494.33	35000.0	135000.0	240000.0	340000.0	3000000.0
Hardware Engineer	466.0	291083.69	172999.64	50000.0	180000.0	265000.0	355000.0	1860000.0
Analyst	411.0	317335.77	158333.61	50000.0	205000.0	305000.0	360000.0	2020000.0
Test Engineer	180.0	329022.22	314984.92	60000.0	180000.0	310000.0	400000.0	4000000.0
Software Engineer	941.0	348512.22	223980.47	50000.0	240000.0	315000.0	412500.0	3500000.0
System Engineer	491.0	353879.84	180654.33	35000.0	310000.0	330000.0	442500.0	3500000.0
Manager	160.0	364687.50	225009.22	50000.0	210000.0	312500.0	442500.0	1320000.0

```
[123]: order = data.groupby('Job_Role')['Salary'].mean().sort_values().index
[124]: fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(8,8))
sns.barplot(x='Job_Role', y='Salary', data=data, order=order, ax=ax1)
sns.boxplot(x='Job_Role', y='Salary', data=data, order=order, ax=ax2)
ax1.tick_params('x', labelrotation=90)
ax2.tick_params('x', labelrotation=90)
plt.tight_layout()
plt.suptitle('Salary')
plt.show()
```



- **Observation:**

- By the above graph Managers are Earning More than others.

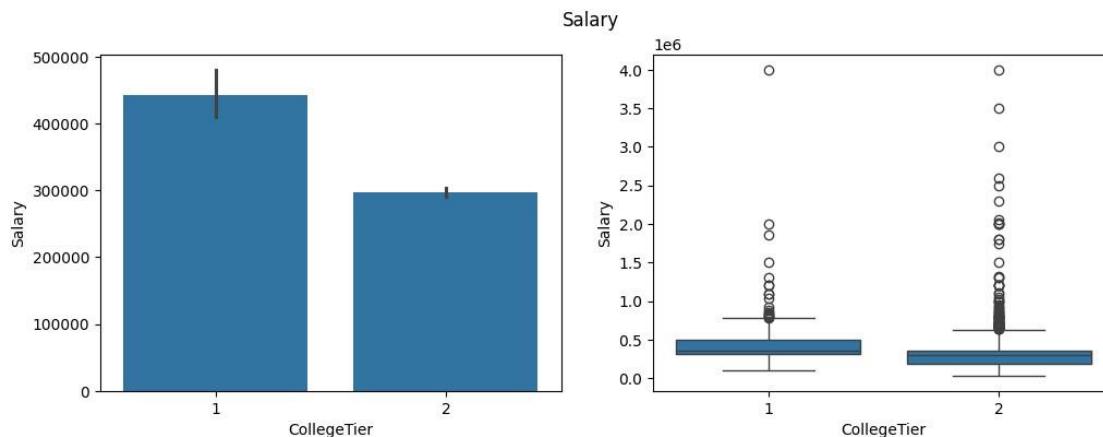
- The second Most Earner from the plot is System Engineer

2.26 Salary vs CollegeTier

```
[125]: df.groupby('CollegeTier')['Salary'].describe()
```

	count	mean	std	min	25%	50%	75%	max
CollegeTier								
1	297.0	442356.902357	312169.285374	100000.0	315000.0			
2	3701.0	296893.812483	198800.849194	35000.0	180000.0			
CollegeTier								
1		360000.0	500000.0	4000000.0				
2		300000.0	360000.0	4000000.0				

```
[126]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
sns.barplot(x='CollegeTier', 'Salary', data=data, ax=ax1)
sns.boxplot(x='CollegeTier', 'Salary', data=data, ax=ax2)
plt.suptitle('Salary')
plt.show()
```



- **Observation:**

- The people who are from Tier-1 college are Earning More as compared to Tire-2

2.27 Salary vs Specialization

```
[127]:
```

```
data.groupby('Specialization')['Salary'].describe().round(1).sort_values('mean')
```

	count	mean	std	min	25%	50%	75%	max
Specialization								
System Engineer	297.0	442356.902357	312169.285374	100000.0	315000.0			
Data Scientist	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Machine Learning Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Software Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Big Data Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Cloud Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
DevOps Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Blockchain Developer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
AI Researcher	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Product Manager	3701.0	296893.812483	198800.849194	35000.0	180000.0			
UX Designer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
UI Designer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Blockchain Developer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Machine Learning Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Big Data Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Cloud Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Software Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Machine Learning Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
System Engineer	3701.0	296893.812483	198800.849194	35000.0	180000.0			
Data Scientist	3701.0	296893.812483	198800.849194	35000.0	180000.0			

```

Specialization
EEE           85.0 289529.4 213651.1 40000.0 180000.0 265000.0
ECE          1319.0 298798.3 189421.4 40000.0 190000.0 300000.0
MECH         220.0 310836.4 206893.2 60000.0 180000.0 267500.0
CSE        2289.0 312020.5 226690.8 35000.0 180000.0 300000.0 other
               56.0 317946.4 162941.1 100000.0 180000.0 317500.0
CE            29.0 381206.9 190528.7 110000.0 295000.0 320000.0

```

75% max

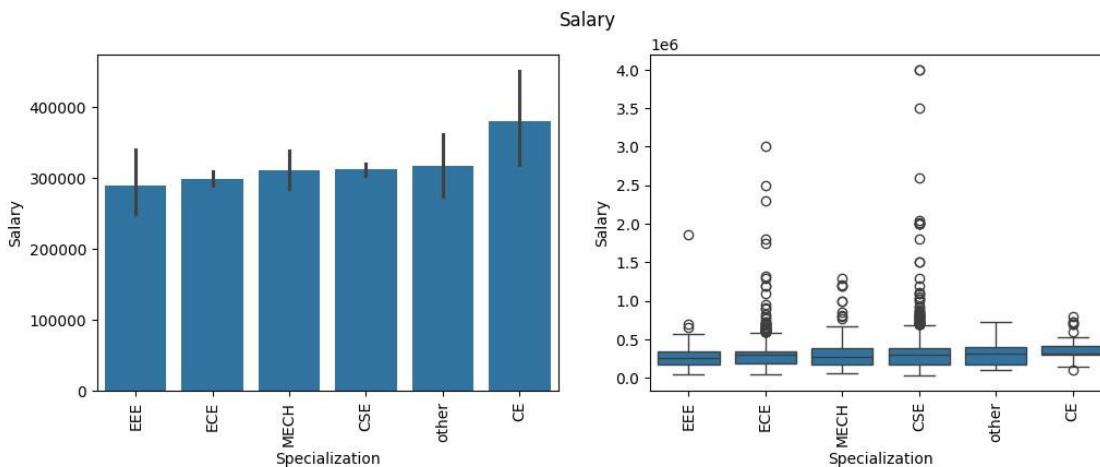
```

Specialization
EEE           340000.0 1860000.0
ECE           350000.0 3000000.0
MECH         391250.0 1300000.0 CSE
385000.0     4000000.0      other
400000.0    730000.0
CE            415000.0 800000.0

```

```
[128]: order = data.groupby('Specialization')['Salary'].mean().sort_values().index
```

```
[129]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12,4))
sns.barplot(x='Specialization', y='Salary', data=data, order=order, ax=ax1)
sns.boxplot(x='Specialization', y='Salary', data=data, order=order, ax=ax2)
ax1.tick_params('x', labelrotation=90)
ax2.tick_params('x', labelrotation=90)
plt.suptitle('Salary')
plt.show()
```



- **Observation:**

- CSE people are earning more as compared to other students

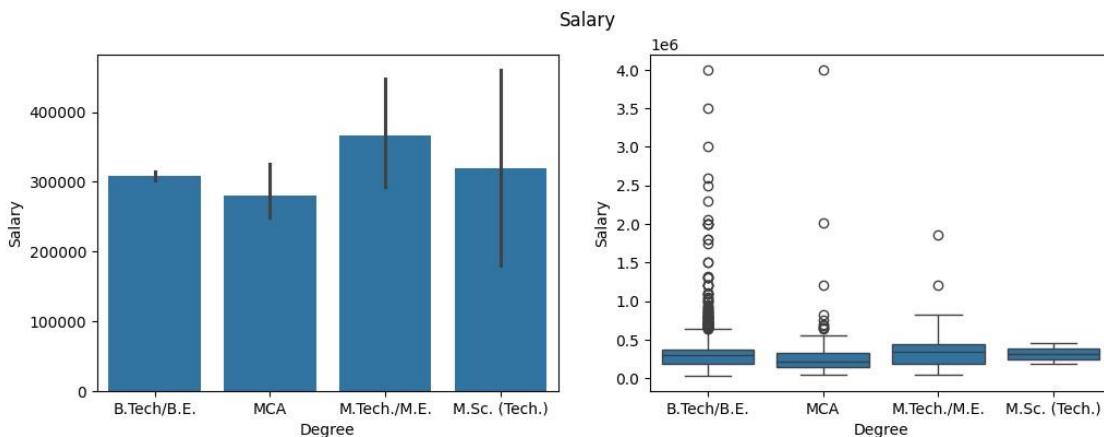
2.28 Salary vs Degree

```
[130]: data.groupby('Degree')['Salary'].describe()
```

Degree	count	mean	std	min	25%	50%	75%	max
B.Tech/B.E.	3700.0	308622.702703	203358.18609035000.0	190000.0				
M.Sc. (Tech.)	2.0	320000.000000	197989.898732	180000.0	250000.0			
M.Tech./M.E.	53.0	366132.075472	301422.335815	40000.0	180000.0			
MCA	243.0	280802.469136	306936.46554050000.0	145000.0				

Degree	300000.0	370000.0	4000000.0
B.Tech/B.E.	300000.0	370000.0	4000000.0
M.Sc. (Tech.)	320000.0	390000.0	460000.0
M.Tech./M.E.	340000.0	440000.0	1860000.0
MCA	220000.0	330000.0	4000000.0

```
[131]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12,4))
sns.barplot(x='Degree',  = 'Salary', data=data, ax=ax1)
sns.boxplot(x='Degree',  = 'Salary', data=data, ax=ax2)
plt.suptitle('Salary')
plt.show()
```



- **Observation:**

- M.Tech/M.E students are earning More than others, but B.Tech/B.E Students having more chances to earn better than M.Tech Students.

2.29 Statistical Tests

2.29.1 Research Questions

- Times of India article dated Jan 18, 2019 states that “After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate.” Test this claim with the data given to you.
- Is there a relationship between gender and specialization? (i.e. Does the preference of Specialisation depend on the Gender?)

- **Let's Verify the claim**

- **Defining hypothesis |**

Hypothesis	Description
Null Hypothesis (H0)	= 250k - 300k
Alternate Hypothesis (H1)	!= 250k - 300k

```
[146]: data['Job_Role'].unique()
```

```
[146]: array(['Test Engineer', 'Manager', 'System Engineer', 'Software Engineer',
   'Hardware Engineer', 'Developer', 'Analyst', 'Other',
   'Administrator'], dtype=object)
```

```
[147]: data['Designation'].unique()
```

```
[147]: array(['senior quality engineer', 'assistant manager', 'systems engineer',
   'senior software engineer', 'system engineer',
   'java software engineer', 'mechanical engineer',
   'electrical engineer', 'project engineer', 'senior php developer',
   'senior systems engineer', 'quality assurance engineer',
   'qa analyst', 'network engineer', 'product development engineer',
   'associate software developer', 'data entry operator',
   'software engineer', 'developer', 'electrical project engineer',
   'programmer analyst', 'systems analyst', 'ase',
   'telecommunication engineer', 'application developer',
   'ios developer', 'executive assistant', 'online marketing manager',
   'documentation specialist', 'associate software engineer',
   'management trainee', 'site manager', 'software developer',
   '.net developer', 'production engineer', 'jr. software engineer',
   'trainee software developer', 'ui developer',
   'assistant system engineer', 'android developer',
   'customer service', 'test engineer', 'java developer',
   'engineer',
   'recruitment coordinator', 'technical support engineer',
   'data analyst', 'assistant software engineer', 'faculty',
```

'entry level management trainee',
'customer service representative', 'software test engineer',
'firmware engineer', 'php developer', 'research associate',
'research analyst', 'quality engineer', 'programmer',
'technical support executive', 'business analyst', 'web
developer',
'application engineer', 'project coordinator', 'engineer
trainee',
'sap consultant', 'quality analyst', 'marketing coordinator',
'system administrator', 'senior engineer',
'business development managerde', 'network administrator',
'technical support specialist', 'business development executive',
'junior software engineer', 'asp.net developer',
'graduate engineer trainee', 'field engineer',
'assistant professor', 'trainee software engineer',
'senior software developer',
'quality assurance automation engineer', 'design engineer',
'telecom engineer', 'quality control engineer',
'hardware engineer', 'hr recruiter', 'sales associate',
'junior engineer', 'associate engineer', 'maintenance engineer',
'sales engineer', 'human resources associate',
'mobile application developer',
'electronic field service engineer', 'process associate',
'field service engineer', 'it support specialist',
'software development engineer', 'business process analyst',
'operation engineer', 'electrical designer', 'marketing
assistant',
'sales executive', 'admin assistant', 'senior java developer',
'account executive', 'oracle dba', 'rf engineer',
'embedded software engineer', 'programmer analyst trainee',
'technical engineer', 'operations executive', 'trainee engineer',
'recruiter', 'lecturer', '.net web developer',
'marketing executive', 'operations assistant', 'associate
manager',
'electrical design engineer', 'systems administrator',
'client services associate', 'it analyst', 'senior developer',
'cad designer', 'business technology analyst', 'asst. manager',
'service engineer', 'executive recruiter', 'planning engineer',
'associate technical operations', 'web designer',
'software architect', 'software quality assurance tester',
'seo trainee', 'process engineer',
'software quality assurance analyst', 'designer',
'business systems consultant', 'business development manager',
'junior research fellow', 'technical recruiter',
'operations analyst', 'quality assurance test engineer',

'linux systems administrator', 'software trainee',
'entry level sales and marketing', 'electrical field engineer',
'windows systems administrator', 'junior software developer',
'python developer', 'web application developer',
'assistant systems engineer', 'javascript developer',
'operation executive', 'performance engineer', 'technical writer',
'operations engineer and jetty handling', 'lead engineer',
'portfolio analyst', 'associate system engineer',
'mechanical design engineer', 'product engineer',
'network security engineer', 'operations manager',
'technical lead', 'operations', 'quality assurance tester',
'automation engineer', 'data scientist', 'quality associate',
'manual tester', 'sr. engineer', 'embedded engineer',
'service and sales engineer', 'telecom support engineer',
'engineer- customer support', 'cloud engineer', 'branch manager',
'business analyst consultant', 'technology lead',
'software trainee engineer', 'dcs engineer', 'junior manager',
'ux designer', 'clerical', 'hr generalist',
'database administrator', 'senior design engineer', 'seo',
'assistant engineer', 'marketing analyst', 'it executive',
'salesforce developer', 'software tester', 'sql dba',
'junior engineer product support', 'manager',
'senior business analyst', 'c# developer',
'implementation engineer', 'executive hr', 'executive engineer',
'sharepoint developer', 'system analyst',
'sales management trainee', 'senior project engineer',
'it recruiter', 'software engineer analyst',
'desktop support technician', 'continuous improvement engineer',
'process advisor', 'etl developer', 'sales and service engineer',
'project manager', 'training specialist', 'product manager',
'staffing recruiter', 'assistant programmer', 'quality controller',
'mis executive', 'game developer', 'digital marketing specialist',
'principal software engineer', 'software devloper',
'senior mechanical engineer', 'technical operations analyst',
'service coordinator', 'testing engineer', 'technical assistant',
'sap abap consultant', 'seo engineer', 'project assistant',
'talent acquisition specialist', 'sales account manager',
'software engineer trainee', 'customer service manager',
'help desk analyst', 'general manager', 'engineering manager',
'senior network engineer',
'field based employee relations manager', 'phone banking officer',

'support engineer', 'associate test engineer',
'technology analyst', 'network support engineer',
'it business analyst', 'junior system analyst',
'senior .net developer', 'secretary', 'research engineer',
'quality assurance auditor', 'process executive',
'lecturer & electrical maintenance', 'office coordinator',
'hr manager', 'html developer', 'sales support',
'front end web developer', 'administrative support',
'territory sales manager', 'project administrator',
'environmental engineer', 'web designer and seo',
'information security analyst',
'field business development associate', 'operational executive',
'administrative coordinator', 'senior risk consultant',
'desktop support engineer', 'cad drafter', 'noc engineer',
'industrial engineer', 'it engineer', 'human resources intern',
'senior quality assurance engineer', 'clerical assistant',
'software enginner', 'quality assurance',
'delivery software engineer', 'graphic designer',
'sales development manager', 'visiting faculty',
'business intelligence analyst', 'team lead',
'operational excellence manager', 'sales & service engineer',
'web intern', 'full stack developer', 'database developer',
'sr. database engineer', 'graduate apprentice trainee',
'software engineer associate', 'technical analyst',
'executive engg', 'it technician', 'business system analyst',
'process control engineer', 'technical consultant',
'business office manager', 'quality control inspector',
'product design engineer', 'manufacturing engineer',
'seo executive', 'sap analyst', 'software engineere',
'financial service consultant', 'co faculty', 'software analyst',
'desktop support analyst', 'graduate engineer',
'engineering technician', 'it assistant', 'marketing manager',
'human resource assistant', 'hr assistant', 'product developer',
'customer support engineer',
'quality control inspection technician', 'gis/cad engineer',
'senior web developer', 'sql developer', 'research staff member',
'sap abap associate consultant', 'associate qa',
'corporate recruiter', 'project management officer',
'business systems analyst', 'software programmer',
'help desk technician', 'sales manager', 'catalog associate',
'assistant store manager', 'software engg', 'it developer',
'apprentice', 'business consultant', 'controls engineer',
'ruby on rails developer', 'risk consultant', 'account manager',
'professor', 'assistant administrator', 'civil engineer',

```

'educator', 'service manager', 'teradata dba',
'full-time loss prevention associate', 'junior recruiter',
'associate developer', 'assistant electrical engineer',
'shift engineer', 'dotnet developer', 'rf/dt engineer',
'human resources analyst', 'software test engineer',
'junior .net developer', 'java trainee', 'maintenance supervisor',
'r&d engineer', 'front end developer', 'engineer-hws',
'operations engineer', 'senior research fellow',
'web designer and joomla administrator',
'enterprise solutions developer',
'information technology specialist', 'site engineer',
'graduate trainee engineer', 'quality assurance analyst',
'cnc programmer', 'financial analyst', 'system engineer trainee',
'sap mm consultant', 'assistant system engineer trainee',
'qa trainee', 'teradata developer', 'hr executive',
'senior programmer', 'software test engineer (etl)',
'associate software engg', 'supply chain analyst', 'sales trainer',
'software executive', 'team leader',
'assistant system engineer - trainee', 'seo analyst',
'risk investigator', 'executive administrative assistant',
'program manager', 'r & d', 'sap functional consultant',
'website developer/tester', 'software designer',
'sales coordinator', 'qa engineer', 'aircraft technician',
'customer care executive', 'senior test engineer',
'program analyst trainee', 'electrical controls engineer',
'trainee decision scientist', 'editor', 'bss engineer', 'dba',
'software eng', 'computer faculty', 'recruitment associate',
'logistics executive', 'quality consultant',
'senior sales executive', 'db2 dba', 'test technician',
'it operations associate', 'software engineering associate',
'research scientist', 'jr. software developer'], dtype=object)

```

```
[148]: filtered_df = data[data['Designation'].isin(['programmer analyst',
'software_',
'engineer', 'hardware engineer', 'associate engineer'])]
```

	ID	Salary	DOJ	DOL	Designation	JobCity
19	466888	325000.0	2014-09-01	2020-03-20	software engineer	Pune
20	140069	320000.0	2010-11-01	2012-01-09	software engineer	Bengaluru
21	339689	200000.0	2012-08-01	2013-01-12	software engineer	Bengaluru
24	963123	335000.0	2014-06-01	2015-01-06	programmer analyst	Hyderabad
31	1094324	340000.0	2014-08-01	2015-01-04	software engineer	Bengaluru

...
3979	212055	550000.0	2013-07-01	2014-01-04	software engineer	Bengaluru		
3981	1077872	220000.0	2014-09-01	2020-03-20	software engineer	Gurugram		
3984	305041	480000.0	2011-12-01	2020-03-20	software engineer	Gurugram		
3989	1204604	300000.0	2014-09-01	2020-03-20	software engineer	Bengaluru		
3993	47916	280000.0	2011-10-01	2012-01-10	software engineer	NEW Delhi		
	Gender	DOB	10percentage	10board	12graduation	12percentage	\	
19	f	1990-11-30	79.00	CBSE	2008	62.20		
20	f	1988-07-25	91.20	State Board	2006	84.63		
21	f	1991-08-20	75.67	State Board	2007	74.40		
24	m	1993-06-28	88.00	State Board	2010	90.00		
31	m	1992-10-23	77.20	State Board	2010	86.10		
...	
3979	m	1989-07-22	69.16	State Board	2006	65.66		
3981	m	1991-12-17	53.40	CBSE	2009	65.40		
3984	f	1990-01-18	89.80	CBSE	2007	83.80		
3989	m	1991-11-23	74.88	State Board	2010	82.55		
3993	m	1987-04-15	52.09	CBSE	2006	55.50		
	12board	CollegeID	CollegeTier	Degree	Specialization	\		
19	CBSE	7977	2	B.Tech/B.E.	CSE			
20	State Board	4437	2	B.Tech/B.E.	CSE			
21	State Board	11368	2	B.Tech/B.E.	CSE			
24	State Board	9173	2	B.Tech/B.E.	CSE			
31	State Board	7282	2	B.Tech/B.E.	CSE			
...	
3979	State Board	278	1	M.Tech./M.E.	CSE			
3981	CBSE	3879	2	B.Tech/B.E.	CSE			
3984	CBSE	5232	2	B.Tech/B.E.	CSE			
3989	State Board	3572	2	B.Tech/B.E.	CSE			
3993	CBSE	6268	2	B.Tech/B.E.	CSE			
	collegeGPA	CollegeCityID	CollegeCityTier	CollegeState				
\								
19	76.30	7977	0	Uttar Pradesh				
20	78.15	4437	0	Karnataka				
21	74.01	11368	0	Uttar Pradesh				
24	76.00	9173	1	Telangana				

31	71.00	7282	0	Tamil Nadu
...
3979	77.50	278	0	Uttar Pradesh
3981	71.20	3879	0	Madhya Pradesh
3984	70.00	5232	0	Haryana
3989	80.00	3572	0	Karnataka
3993	61.50	6268	0	Haryana
GraduationYear English Logical Quant Domain \				
19 1970-01-01 00:00:00.000002012	485	445	435	
			0.864685	
20 1970-01-01 00:00:00.000002010	455	535	445	
			0.635979	
21 1970-01-01 00:00:00.000002012	385	555	445	
			0.635979	
24 1970-01-01 00:00:00.000002014	625	555	630	
			0.356536	
31 1970-01-01 00:00:00.000002014	560	485	450	
			0.622643	
...
3979 1970-01-01 00:00:00.000002013	395	435	645	
			0.990009	
3981 1970-01-01 00:00:00.000002013	560	420	645	
			0.953900	
3984 1970-01-01 00:00:00.000002011	535	455	525	
			0.843124	
3989 1970-01-01 00:00:00.000002014	500	480	500	
			0.356536	
3993 1970-01-01 00:00:00.000002010	365	334	475	
			0.276047	
ComputerProgramming ElectronicsAndSemicon ComputerScience \				
19	525	-1	-1	
20	445	-1	-1	
21	445	-1	-1	
24	475	-1	346	
31	485	-1	407	
...

3979	655	-1	-1
3981	575	-1	530
3984	515	-1	-1
3989	465	-1	346
3993	345	-1	-1
	MechanicalEngg ElectricalEngg TelecomEngg CivilEngg \		
19 -1 -1 -1 -----			135
20 -1 -1 -1 -----			135
21 -1 -1 -1 -----			135
24 -1 -1 -1 -----			135
31 -1 -1 -1... -----			135
3979 -1 -1 -1 -----			136
3981 -1 -1 -1 -----			136
3984 -1 -1 -1 -----			136
3989 -1 -1 -1 -----			136
3993 -1 -1 -1 -----			136
	conscientiousness agreeableness extraversion nueroticism \		
19	0.8192 0.2668 -0.2714 -0.40780		
20	-0.1082 0.9688 0.5269 -0.29020		
21	-1.0355 -0.5913 -1.3599 0.06223		
24	0.4155 0.8027 0.1357 -0.99500		
31	-0.0154 1.2114 1.0859 -1.50210		
...
3979	-0.5719 0.5008 -0.4891 0.41480		
3981	0.1282 -0.2871 -0.1437 -1.12180		
3984	-0.2628 0.1888 0.3817 -0.29020		
3989	0.1282 0.0459 1.2396 1.03330		
3993	-0.1082 0.3448 0.2366 0.64980		
	openess_to_experience period DOB_Year GYear 12GradAge GradAge \		

19	-0.1295	6	1990	1970	20	20
20	0.3444	2	1988	1970	18	18
21	-1.3539	1	1991	1970	21	21
24	-0.6692	1	1993	1970	23	23
31	0.2889	1	1992	1970	22	22
...
3979	-1.2354	1	1989	1970	19	19
3981	1.4386	6	1991	1970	21	21
3984	1.6082	9	1990	1970	20	20
3989	0.6721	6	1991	1970	21	21
3993	-0.9194	1	1987	1970	17	17

Job_Role

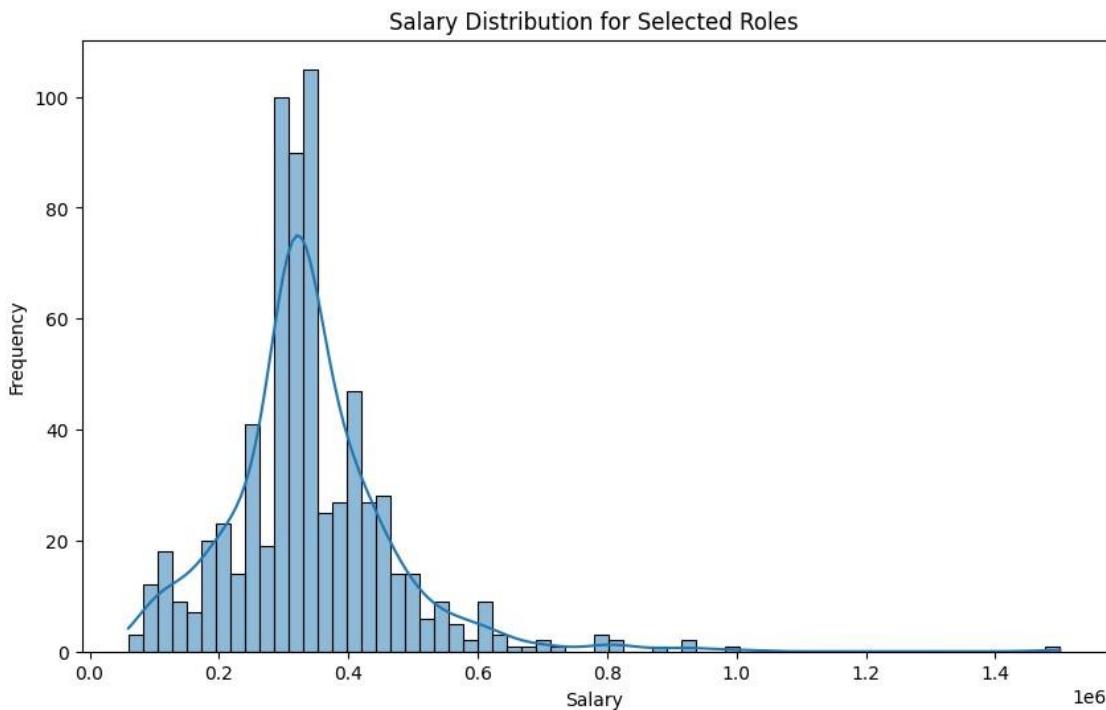
19	Software Engineer
20	Software Engineer
21	Software Engineer
24	Analyst
31	Software Engineer
...	...
3979	Software Engineer
3981	Software Engineer
3984	Software Engineer
3989	Software Engineer
3993	Software Engineer

[692 rows x 44 columns]

```
[149]: average_salary = filtered_df['Salary'].mean()
print(f'The average salary for these roles is:
{average_salary}')
```

The average salary for these roles is: 339790.4624277457

```
[150]: plt.figure(figsize=(10, 6))
sns.histplot(data=filtered_df, x="Salary", kde=True)
plt.title('Salary Distribution for Selected Roles')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.show()
```



```
[144]: max_salary = filtered_df['Salary'].max()
```

```
[152]: if max_salary >= 250000 and max_salary <= 300000: print("Fail to
        Reject Null Hypothesis (H1): The claim that Recent graduates_
        can earn between 2.5 - 3 lakhs is supported by the
        data.") else:
        print("Reject Null Hypothesis (H0): The claim that Recent
        graduates can_
        earn between 2.5 - 3 lakhs is not supported by the data.")
```

Reject Null Hypothesis (H0): The claim that Recent graduates can earn between 2.5 - 3 lakhs is not supported by the data.

- **Observation:**

- Reject Null Hypothesis (H0) - Reject Null Hypothesis (H0): The claim that Recent graduates can earn between 2.5 - 3 lakhs is not supported by the data.

2.30 Conclusions

- The data reveals a gender imbalance, indicating a need for diversity efforts in the workforce.
- Technical skills, particularly in Computer Science and Engineering, are in high demand based on the prevalence of related degrees.
- Job roles vary widely, with Software Engineer being the most common, followed by Developer.
- Educational board preferences influence policies, with a preference for State Boards, CBSE, and ICSE.
- Technical expertise is crucial, as evidenced by the prevalence of Bachelor of Technology/Engineering graduates.
- Managerial and technical positions are the highest-earning roles.
- College tier impacts earnings, with Tier-1 graduates earning more.
- Gender-based salary differences exist, though further analysis is needed for clarity.
- The claim about recent graduates' earnings in Computer Science Engineering was not supported by the data.
- There's no significant link between gender and specialization preference, challenging assumptions about their correlation.