



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON
Exploratory Data Analysis on AMCAT

About me

- **Education:** B.Tech. in Computer Engineering from D Y Patil University, Pune.
Expected Graduation: June 2025
- **Interest in Data Science:** Passion for using data-driven methods to understand and improve modern solutions.
- **Worked as Data Science Project Intern at CodersCave** implemented Data Science Project-Based Intern at CodersCave. Analyzing data to extract insights for impactful projects.
- **LinkedIN URL :** <https://www.linkedin.com/in/rohitjakkam/>
- **Github URL :** <https://github.com/Rohitjakkam>

Agenda

- Business Problem and Use case domain understanding(If Required)
- Objective of the Project
- Web Scraping – Details (Websites, Processor you followed)
- Summary of the Data
- Exploratory Data Analysis:
 - a. *Data Cleaning Steps*
 - b. *Data Manipulation Steps*
 - c. *Univariate Analysis Steps*
 - d. *Bivariate Analysis Steps*
- Key Business Question
- Conclusion (Key finding overall)
- Q&A Slide
- Your Experience/Challenges working on Web Scraping – Data Analysis Project.

Objective

Specific Goals:

1. Identify key salary influencers (education, skills, location, personality).
2. Analyze links between academic background and professional success.
3. Examine skill impact on job outcomes (salary, designation).
4. Investigate gender and location-based salary differences.
5. Study personality traits' effect on professional outcomes.
6. Develop predictive models for salary forecasting.

Summary of the Data

Dataset Overview:

The dataset contains information on individuals' personal, academic, and professional attributes.

Data Description:

1. Unique ID
2. Salary
3. Job details (designation, city, date of joining/leaving)
4. Demographics (gender, date of birth)
5. Education (10th/12th grade percentages, boards, college details)
6. Skills (English, logical reasoning, quantitative, programming)
7. Personality traits (conscientiousness, agreeableness, extraversion)
8. Domain expertise and engineering fields (computer science, mechanical, electrical)

Analysis Goals:

1. Understand salary drivers
2. Identify trends and patterns
3. Prepare data for predictive modeling

Exploratory Data Analysis

a. Data Cleaning Steps:

1. Date Conversion: Convert object-type dates to datetime using pandas' pd.to_datetime.
2. Null Value Check: Identify and address missing values. If over 70% of values are missing, consider removing the column.
3. Duplicate Removal: Eliminate duplicate rows to prevent data redundancy.
4. Outlier Detection: Identify and handle outliers using box plots. Remove or transform outliers impacting analysis.

b. Data Manipulation Steps:

1. Handle Missing Data: Fill or drop missing values using methods like imputation (mean, median, mode) or algorithms that handle missing data.
2. Remove Duplicates: Eliminate duplicate rows.
3. Handle Outliers: Detect and address outliers to prevent skewed analysis.

✓ Rohit Keshav Jakkam - IN9240103

```
1 # Importing Required Libraries
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 import warnings
7 warnings.filterwarnings('ignore')
```

```
1 # Loading Data
2 data=pd.read_csv("amcat.csv")
3 data.head()
```

		Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	10b t ofseco education
0	train	203097	4200000.0	6/1/12 0:00	present		senior quality engineer	Bangalore	f	2/19/90 0:00	84.3	t ofseco education
1	train	579905	5000000.0	9/1/13 0:00	present		assistant manager	Indore	m	10/4/89 0:00	85.4	
2	train	810601	3250000.0	6/1/14 0:00	present		systems engineer	Chennai	f	8/3/92 0:00	85.0	
3	train	267447	1100000.0	7/1/11 0:00	present		senior software engineer	Gurgaon	m	12/5/89 0:00	85.6	
4	train	343523	2000000.0	3/1/14 0:00	3/1/15 0:00		get	Manesar	m	2/27/91 0:00	78.0	

```
1 data.shape
```

```
(3998, 39)
```

```
1 # Information about data like nullvalue count, Datatype
2 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Unnamed: 0        3998 non-null   object 
 1   ID               3998 non-null   int64  
 2   Salary            3998 non-null   float64
 3   DOJ              3998 non-null   object 
 4   DOL              3998 non-null   object 
 5   Designation       3998 non-null   object 
 6   JobCity           3998 non-null   object 
 7   Gender            3998 non-null   object 
 8   DOB              3998 non-null   object
```

```

9 10percentage      3998 non-null   float64
10 10board          3998 non-null   object
11 12graduation     3998 non-null   int64
12 12percentage     3998 non-null   float64
13 12board          3998 non-null   object
14 CollegeID         3998 non-null   int64
15 CollegeTier       3998 non-null   int64
16 Degree            3998 non-null   object
17 Specialization    3998 non-null   object
18 collegeGPA        3998 non-null   float64
19 CollegeCityID     3998 non-null   int64
20 CollegeCityTier   3998 non-null   int64
21 CollegeState       3998 non-null   object
22 GraduationYear    3998 non-null   int64
23 English           3998 non-null   int64
24 Logical           3998 non-null   int64
25 Quant              3998 non-null   int64
26 Domain             3998 non-null   float64
27 ComputerProgramming 3998 non-null   int64
28 ElectronicsAndSemicon 3998 non-null   int64
29 ComputerScience    3998 non-null   int64
30 MechanicalEngg    3998 non-null   int64
31 ElectricalEngg    3998 non-null   int64
32 TelecomEngg       3998 non-null   int64
33 CivilEngg          3998 non-null   int64
34 conscientiousness  3998 non-null   float64
35 agreeableness      3998 non-null   float64
36 extraversion        3998 non-null   float64
37 nueroticism         3998 non-null   float64
38 openness_to_experience 3998 non-null   float64
dtypes: float64(10), int64(17), object(12)
memory usage: 1.2+ MB

```

```

1 # Describing Statistical analysis
2 data.describe()

```

	ID	Salary	10percentage	12graduation	12percentage	CollegeID	CollegeTier	co
count	3.998000e+03	3.998000e+03	3998.000000	3998.000000	3998.000000	3998.000000	3998.000000	3998.000000
mean	6.637945e+05	3.076998e+05	77.925443	2008.087544	74.466366	5156.851426	1.925713	7
std	3.632182e+05	2.127375e+05	9.850162	1.653599	10.999933	4802.261482	0.262270	0.262270
min	1.124400e+04	3.500000e+04	43.000000	1995.000000	40.000000	2.000000	1.000000	1.000000
25%	3.342842e+05	1.800000e+05	71.680000	2007.000000	66.000000	494.000000	2.000000	€
50%	6.396000e+05	3.000000e+05	79.150000	2008.000000	74.400000	3879.000000	2.000000	7
75%	9.904800e+05	3.700000e+05	85.670000	2009.000000	82.600000	8818.000000	2.000000	7
max	1.298275e+06	4.000000e+06	97.760000	2013.000000	98.700000	18409.000000	2.000000	9

```

1 # Read data in depth
2 for col in data.columns:
3   print('-----{}-----'.format(col))
4   print(data[col].unique())

```

```

-----'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]

```

Dropping irrelevant feature

```
1 data.drop('Unnamed: 0', axis=1, inplace=True)
```

"Date of Joining" Changing Datatype from Object to DateTime

```
1 data['DOJ'] = pd.to_datetime(data['DOJ'])
```

```
1 data['DOL'].replace(to_replace='present', value= pd.Timestamp.now(), inplace=True)
```

"Date of Leaving" Changing Datatype from Object to DateTime

```
1 data['DOL'] = pd.to_datetime(data['DOL'])
```

```
1 from dateutil.relativedelta import relativedelta
```

```
2
```

```
3 data['Tenure'] = data.apply(lambda row: relativedelta(row['DOL'], row['DOJ']).years * 12 + relativedelta(r
```

```
4 data['Tenure'].unique()
```

```
→ array([148, 133, 124, 159, 12, 9, 135, 123, 48, 38, 130, 129, 134,
       121, 13, 5, 22, 16, 18, 8, 145, 33, 4, 2, -7, 154,
       132, 120, 117, 127, 139, 54, 122, 35, 21, 26, 20, 116, 114,
       119, 151, 7, 43, 131, 6, 24, 3, 32, 112, 125, 10, 155,
       111, 137, -4, 14, 44, 163, 1, 136, 115, 147, 42, 174, 30,
       15, 144, 138, 126, 167, 45, -9, 19, 25, 17, 27, 140, 157,
       152, 37, 11, 153, 146, 118, 31, 150, -3, 141, 0, 39, 128,
       113, 34, 149, 40, 41, 57, 28, 166, 23, 164, 142, 158, 165,
       143, 171, 170, 36, 29, -2, 161, 156, 50, 52, 169, 160, 172,
       -8, 53, 162, 62, -5, 49, 59, 175, 107, 47, 56, 168, 225,
       110, -6, -10, 106, 46, 51, 72, 400, -1, 207, 55])
```

"Date of birth" Changing Datatype from Object to DateTime

```
1 data['DOB']=pd.to_datetime(data['DOB'])
```

```
2 data['DOB'].dtype
```

```
→ dtype('M8[ns]')
```

```
1 from datetime import datetime, date
```

```
2 from dateutil.relativedelta import relativedelta
```

```
3 current_date = datetime.now()
```

```
4
```

```
5 # Calculate age in years using relativedelta
```

```
6 data['Age'] = data['DOB'].apply(lambda dob: relativedelta(current_date, dob).years)
```

```
7 print(data['Age'])
```

```
→ 0      34
  1      34
  2      32
  3      34
  4      33
  ..
3993    37
3994    32
3995    33
3996    32
3997    33
Name: Age, Length: 3998, dtype: int64
```

```
1 data['Age'].unique()
```

```
→ array([34, 32, 33, 31, 35, 37, 36, 38, 30, 46, 39, 29, 27, 40, 41, 43, 42])
```

```
1 data['10board'].unique()
```

'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',
 'msbsbhse,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', 'cbse', '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',

```

1 XBoard={'board ofsecondary education,ap':'STATE', 'cbse':'CBSE', 'state board':'STATE',
2     'mp board bhopal':'STATE', 'icse':'ICSE',
3     'karnataka secondary school of examination':'STATE', 'up':'STATE',
4     'karnataka state education examination board':'STATE', 'ssc':'STATE',
5     'kerala state technical education':'STATE', '0':'Other', 'bseb':'STATE',
6     'state board of secondary education, andhra prades': 'STATE',

```

7 'matriculation':'Other', 'gujarat state board':'STATE', 'karnataka state board':'STATE',
8 'wbbse':'STATE', 'maharashtra state board':'STATE', 'icse board':'ICSE', 'up board':'STATE',
9 'board of secondary education(bse) orissa':'STATE',
10 'little jacky matric higher secondary school':'Other',
11 'uttar pradesh board':'STATE', 'bsc,orissa':'STATE', 'mp board':'STATE', 'upboard':'STATE',
12 'matriculation board':'Other', 'j & k bord':'STATE', 'rbse':'STATE',
13 'central board of secondary education':'CBSE', 'pseb':'STATE', 'jkbose':'STATE',
14 'haryana board of school education,(hbse)':'STATE', 'metric':'Other', 'ms board':'STATE',
15 'kseeb':'STATE', 'stateboard':'STATE', 'maticulation':'Other',
16 'karnataka secondary education board':'STATE', 'mumbai board':'STATE', 'sslc':'STATE',
17 'kseb':'STATE', 'board secondary education':'STATE', 'matric board':'Other',
18 'board of secondary education':'STATE',
19 'west bengal board of secondary education':'STATE',
20 'jharkhand secondary examination board,ranchi':'STATE', 'u p board':'STATE',
21 'bseb,patna':'STATE', 'hsc':'STATE', 'bse':'STATE', 'sss pune':'STATE',
22 'karnataka education board (keeb)':'STATE', 'kerala':'STATE',
23 'state board of secondary education(ssc)':'STATE', 'gsheb':'STATE',
24 'up(allahabad)':'STATE', 'nagpur':'STATE', 'don bosco maatriculation school':'ICSE',
25 'karnataka state secondary education board':'STATE', 'maharashtra':'STATE',
26 'karnataka secondary education board':'STATE',
27 'himachal pradesh board of school education':'STATE',
28 'certificate of middle years program of ib':'Other',
29 'karnataka board of secondary education':'STATE',
30 'board of secondary education rajasthan':'STATE', 'uttarakhand board':'STATE',
31 'ua':'STATE', 'board of secendary education orissa':'STATE',
32 'karantaka secondary education and examination borad':'STATE', 'hbse':'STATE',
33 'kseeb(karnataka secondary education examination board)':'STATE',
34 'cbse[gulf zone]':'CBSE', 'hbse':'STATE', 'state(karnataka board)':'STATE',
35 'jharkhand accademic council':'STATE',
36 'jharkhand secondary examination board (ranchi)':'STATE',
37 'karnataka secondary education examination board':'STATE', 'delhi board':'STATE',
38 'mirza ahmed ali baig':'STATE', 'jseb':'STATE', 'bse, odisha':'STATE', 'bihar board':'STATE',
39 'maharashtra state(latur board)':'STATE', 'rajasthan board':'STATE', 'mpboard':'STATE',
40 'upbhsie':'STATE', 'secondary board of rajasthan':'STATE',
41 'tamilnadu matriculation board':'Other', 'jharkhand secondary board':'STATE',
42 'board of secondary education,anhara pradesh':'STATE', 'up baord':'STATE',
43 'state':'STATE', 'board of intermediate education':'Other',
44 'state board of secondary education,anhra pradesh':'STATE',
45 'up board , allahabad':'STATE',
46 'stjosephs girls higher sec school,dindigul':'Other', 'maharashtra board':'STATE',
47 'education board of kerala':'STATE', 'board of ssc':'STATE',
48 'maharashtra state board pune':'STATE',
49 'board of school education harayana':'STATE',
50 'secondary school cerfificate':'STATE', 'maharashtra sate board':'STATE', 'ksseb':'STATE',
51 'bihar examination board, patna':'STATE', 'latur':'STATE',
52 'board of secondary education, rajasthan':'STATE', 'state borad hp':'STATE',
53 'cluny':'CBSE', 'bsepatna':'STATE', 'up borad':'STATE', 'ssc board of andrapradesh':'STATE',
54 'matric':'Other', 'bse,orissa':'STATE', 'ssc-andhra prades': 'STATE', 'mp':'STATE',
55 'karnataka education board':'STATE', 'mhsbse':'STATE',
56 'karnataka sslc board bangalore':'STATE', 'karnataka':'STATE', 'u p':'STATE',
57 'secondary school of education':'STATE', 'state board of karnataka':'STATE',
58 'karnataka secondary board':'STATE', 'andhra pradesh board ssc':'STATE',
59 'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district':'CBSE',
60 'hse,orissa':'STATE', 'national public school':'ICSE', 'nagpur board':'STATE',
61 'jharkhand academic council':'STATE', 'bsemp':'STATE',
62 'board of secondary education, andhra pradesh':'STATE',
63 'board of secondary education orissa':'STATE',
64 'board of secondary education,rajasthan(rbse)':'STATE',
65 'board of secondary education,ap':'STATE',
66 'board of secondary education,anhra prades': 'STATE',
67 'jawahar navodaya vidyalaya':'CBSE', 'aisse':'CBSE',

68 'karnataka board of higher education':'STATE', 'bihar':'STATE',
69 'kerala state board':'STATE', 'cicse':'ICSE', 'tn state board':'STATE',
70 'kolhapur divisional board, maharashtra':'STATE',
71 'bharathi matriculation school':'Other', 'uttaranchal state board':'STATE',
72 'wbbse':'STATE', 'mp state board':'STATE', 'seba(assam)':'STATE', 'anglo indian':'Other', 'gseb':'',
73 'uttar pradesh':'STATE', 'ghseb':'STATE', 'board of school education uttarakhand':'STATE',
74 'msbshse,pune':'STATE', 'tamilnadu state board':'STATE', 'kerala university':'STATE',
75 'uttaranchal shiksha avam pariksha parishad':'STATE',
76 'bse(board of secondary education)':'STATE',
77 'bright way college, (up board)':'STATE',
78 'school secondary education, andhra pradesh':'STATE',
79 'secondary state certificate':'STATE',
80 'maharashtra state board of secondary and higher secondary education,pune':'STATE',
81 'andhra pradesh state board':'STATE', 'stmary higher secondary':'CBSE', 'cgbse':'STATE',
82 'secondary school certificate':'STATE', 'rajasthan board ajmer':'STATE', 'mpbse':'STATE',
83 'pune board':'STATE', 'cbse ':'CBSE', 'board of secondary education,orissa':'STATE',
84 'maharashtra state board,pune':'STATE', 'up bord':'STATE',
85 'kiran english medium high school':'Other', 'state board (jac, ranchi)':'STATE',
86 'gujarat board':'STATE', 'state board ':'STATE', 'sarada high scchool':'Other',
87 'kalaimagal matriculation higher secondary school':'Other',
88 'karnataka board':'STATE', 'maharastra board':'STATE', 'sslc board':'STATE',
89 'ssc maharashtra board':'STATE', 'tamil nadu state':'STATE', 'uttrakhand board':'STATE',
90 'bihar secondary education board,patna':'STATE',
91 'haryana board of school education':'STATE',
92 'sri kannika parameswari highier secondary school, udumalpet':'STATE',
93 'ksseb(karnataka state board)':'STATE', 'nashik board':'STATE',
94 'jharkhand secondary education board':'STATE', 'himachal pradesh board':'STATE',
95 'maharashtra sataste board':'STATE',
96 'maharashtra state board mumbai divisional board':'STATE',
97 'dav public school,hehal':'CBSE',
98 'state board of secondary education, ap':'STATE',
99 'rajasthan board of secondary education':'STATE', 'hsce':'STATE',
100 'karnataka secondary education':'STATE',
101 'board of secondary education,odisha':'STATE', 'maharashtra nasik board':'STATE',
102 'west bengal board of secondary examination (wbbse)':'STATE',
103 'holy cross matriculation hr sec school':'Other', 'cbsc':'CBSE', 'apssc':'STATE',
104 'bseb patna':'STATE', 'kolhapur':'STATE', 'bseb, patna':'STATE', 'up board allahabad':'STATE',
105 'biharboard':'STATE', 'nagpur board,nagpur':'STATE', 'pune':'STATE', 'gyan bharati school':'CBSE',
106 'rbse,ajmer':'STATE', 'board of secondaray education':'STATE',
107 'secondary school education':'STATE', 'state bord':'STATE', 'jbse,jharkhand':'STATE',
108 'hse':'STATE', 'madhya pradesh board':'STATE', 'bihar school examination board':'STATE',
109 'west bengal board of secondary eucation':'STATE', 'state boardmp board ':'STATE',
110 'icse board , new delhi':'ICSE',
111 'board of secondary education (bse) orissa':'STATE',
112 'maharashtra state board for ssc':'STATE',
113 'board of secondary school education':'STATE', 'latur board':'STATE',
114 'stmary's convent inter college':'CBSE', 'nagpur divisional board':'STATE',
115 'ap state board':'STATE', 'cgbse raipur':'STATE', 'uttranchal board':'STATE', 'ksbe':'STATE',
116 'central board of secondary education, new delhi':'CBSE',
117 'bihar school examination board patna':'CBSE', 'cbse board':'CBSE',
118 'sslc,karnataka':'STATE', 'mp-bse':'STATE', 'up bound':'STATE', 'dav public school sec 14':'CBSE',
119 'board of school education haryana':'STATE',
120 'council for indian school certificate examination':'Other',
121 'aurangabad board':'STATE', 'j&k state board of school education':'STATE',
122 'maharashtra state board of secondary and higher secondary education':'STATE',
123 'maharashtra state boar of secondary and higher secondary education':'STATE',
124 'ssc regular':'STATE', 'karnataka state examination board':'STATE', 'nasik':'STATE',
125 'west bengal board of secondary education':'STATE', 'up board,allahabad':'STATE',
126 'bseb ,patna':'STATE',
127 'state board - west bengal board of secondary education : wbbse':'STATE',
128 'maharashtra state board of secondary & higher secondary education':'STATE',

```

129      'delhi public school':'CBSE', 'karnataka secondary eduction':'STATE',
130      'secondary education board of rajasthan':'STATE',
131      'maharashtra board, pune':'STATE', 'rbse (state board)':'STATE', 'apsche':'STATE',
132      'board of secondary education':'STATE',
133      'board of high school and intermediate education uttarpradesh':'STATE',
134      'kea':'STATE', 'board of secondary education - andhra pradesh':'STATE',
135      'ap state board for secondary education':'STATE', 'seba':'STATE',
136      'punjab school education board, mohali':'STATE',
137      'jharkhand acedemic council':'STATE', 'hse,board':'STATE',
138      'board of ssc education andhra pradesh':'STATE', 'up-board':'STATE', 'bse,odisha':'STATE'}

```

```
1 data['XBoard'] = data['10board'].replace(XBoard)
```

```
1 data['XBoard'].unique()
```

```
→ array(['STATE', 'CBSE', 'ICSE', 'Other'], dtype=object)
```

```
1 data.drop('10board',axis=1,inplace=True)
```

```
1 pd.set_option('display.max_columns',None)
2 data[data['GraduationYear']==0]
```

	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	12graduation	12
2664	794381	325000.0	2014-07-01	2015-07-01	systems engineer	Chandigarh	m	1991-12-17	89.0	2010	

```
1 data['GraduationYear'].replace(to_replace = 0, value=2014,inplace = True)
2 data['GraduationYear'].unique()
```

```
→ array([2011, 2012, 2014, 2016, 2013, 2010, 2015, 2009, 2017, 2007])
```

```

1 Domains={'computer engineering':'CS','electronics and communication engineering':'EE',
2 'information technology':'IT','computer science & engineering':'CS',
3 'mechanical engineering':'ME','electronics and electrical engineering':'EE',
4 'electronics & telecommunications':'EE','instrumentation and control engineering':'EE','computer applicat:
5 'electronics and computer engineering':'EE','electrical engineering':'EE',
6 'applied electronics and instrumentation':'EE','electronics & instrumentation eng':'EE','information scie
7 'civil engineering':'CE','mechanical and automation':'ME','industrial & production engineering':'Other',
8 'control and instrumentation engineering':'EE','metallurgical engineering':'Other',
9 'electronics and instrumentation engineering':'EE', 'electronics engineering':'EE',
10 'ceramic engineering': 'Other', 'chemical engineering': 'Chem', 'aeronautical engineering': 'AE',
11 'other': 'Other', 'biotechnology': 'Other', 'embedded systems technology': 'EE',
12 'electrical and power engineering': 'EE', 'computer science and technology': 'CS',
13 'mechatronics': 'ME', 'automobile/automotive engineering': 'ME', 'polymer technology': 'Other',
14 'mechanical & production engineering': 'ME', 'power systems and automation': 'EE',
15 'instrumentation engineering': 'EE' , 'telecommunication engineering': 'IT',
16 'industrial & management engineering': 'ME', 'industrial engineering': 'ME',
17 'computer and communication engineering': 'CS',
18 'information & communication technology': 'IT', 'information science': 'CS',
19 'internal combustion engine': 'Other', 'computer networking': 'CS',
20 'biomedical engineering': 'Other', 'electronics': 'EE', 'computer science': 'CS'}

```

```
1 data['Specialization'].replace(Domains,inplace=True)
2 data['Specialization'].unique()
```

```
array(['CS', 'EE', 'IT', 'ME', 'CE', 'Other', 'Chem', 'AE'], dtype=object)
```

```
1 data[data['Domain'] == -1]
```

	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	12gradl
5	1027655	300000.0	2014-06-01	2024-10-03 20:09:22.038981	system engineer	Hyderabad	m	1992-07-02	89.92	
14	1279958	300000.0	2013-07-01	2024-10-03 20:09:22.038981	java software engineer	Bangalore	m	1992-07-02	81.20	
16	1087625	325000.0	2014-09-01	2024-10-03 20:09:22.038981	product development engineer	Noida	m	1992-06-10	85.80	
22	995191	190000.0	2013-10-01	2015-04-01 00:00:00.000000	developer	Hyderabad	m	1992-01-23	92.50	
35	942679	350000.0	2014-03-01	2024-10-03 20:09:22.038981	online marketing manager	mohali	f	1990-09-15	85.00	
...
3942	1067015	280000.0	2013-07-01	2024-10-03 20:09:22.038981	software engineer	-1	m	1990-03-17	78.00	
3958	1039020	310000.0	2014-06-01	2024-10-03 20:09:22.038981	software engineer	Hyderabad	m	1992-09-09	79.90	
3960	1037712	240000.0	2013-08-01	2014-01-01 00:00:00.000000	web developer	Kolkata	m	1992-07-10	77.00	
3969	1233826	330000.0	2015-06-01	2024-10-03 20:09:22.038981	technical engineer	pune	m	1993-01-24	76.00	
3978	1274658	320000.0	2013-08-01	2014-01-01 00:00:00.000000	software engineer	Hyderabad	f	1991-04-24	89.70	

246 rows × 40 columns

```
1 len(list(data['JobCity'].unique()))
```

→ 339

```
1 data['JobCity'].replace(-1,'Remote/Others',inplace=True)
```

```

12 'Mysore', 'Nagari', 'Nagpur', 'Nalagarh', 'Nanded', 'Nashik', 'Navi Mumbai', 'Hyderabad', 'Neemrana', 'NCR', 'Nellore',
13 'Panchkula', 'Panchnagar', 'Patiala', 'Patna', 'Phagwara', 'Pilani', 'Pondicherry', 'Pune', 'RAE BARELI', 'RAS AL KHAI
14 'Raigarh', 'Raipur', 'Rajasthan', 'Rajkot', 'Rajpura', 'Ranchi', 'Ratnagiri', 'Rayagada', 'Odisha', 'Rewari', 'Rohtak',
15 'Rourkela', 'Rudrapur', 'SHAHDOL', 'Sahibabad', 'Salem', 'Sambalpur', 'Secunderabad', 'Shahdol', 'Shimla', 'Siliguri',
16 'Surat', 'Trivandrum', 'Thane', 'Thiruvananthapuram', 'Tirunelveli', 'Tirupati', 'Tornagallu', 'Trichur', 'Trichy',
17 'Udaipur', 'Una', 'Unnao', 'Vadodara', 'Vandavasi', 'Varanasi', 'Vellore', 'Vijayawada', 'Visakhapatnam', 'Vizag',

```

```
1 !pip install fuzzywuzzy
```

→ Requirement already satisfied: fuzzywuzzy in /usr/local/lib/python3.10/dist-packages (0.18.0)

```

1 from fuzzywuzzy import process
2
3 def correct_spelling_errors(target_word, choices, threshold=80):
4     match, score = process.extractOne(target_word, choices)
5     if score >= threshold:
6         return match
7     else:
8         return target_word

```

```
1 data['JobCities'] = data['JobCity'].apply(lambda x : correct_spelling_errors(str(x),choice))
```

```
1 len(list(data['JobCities'].unique()))
```

→ 200

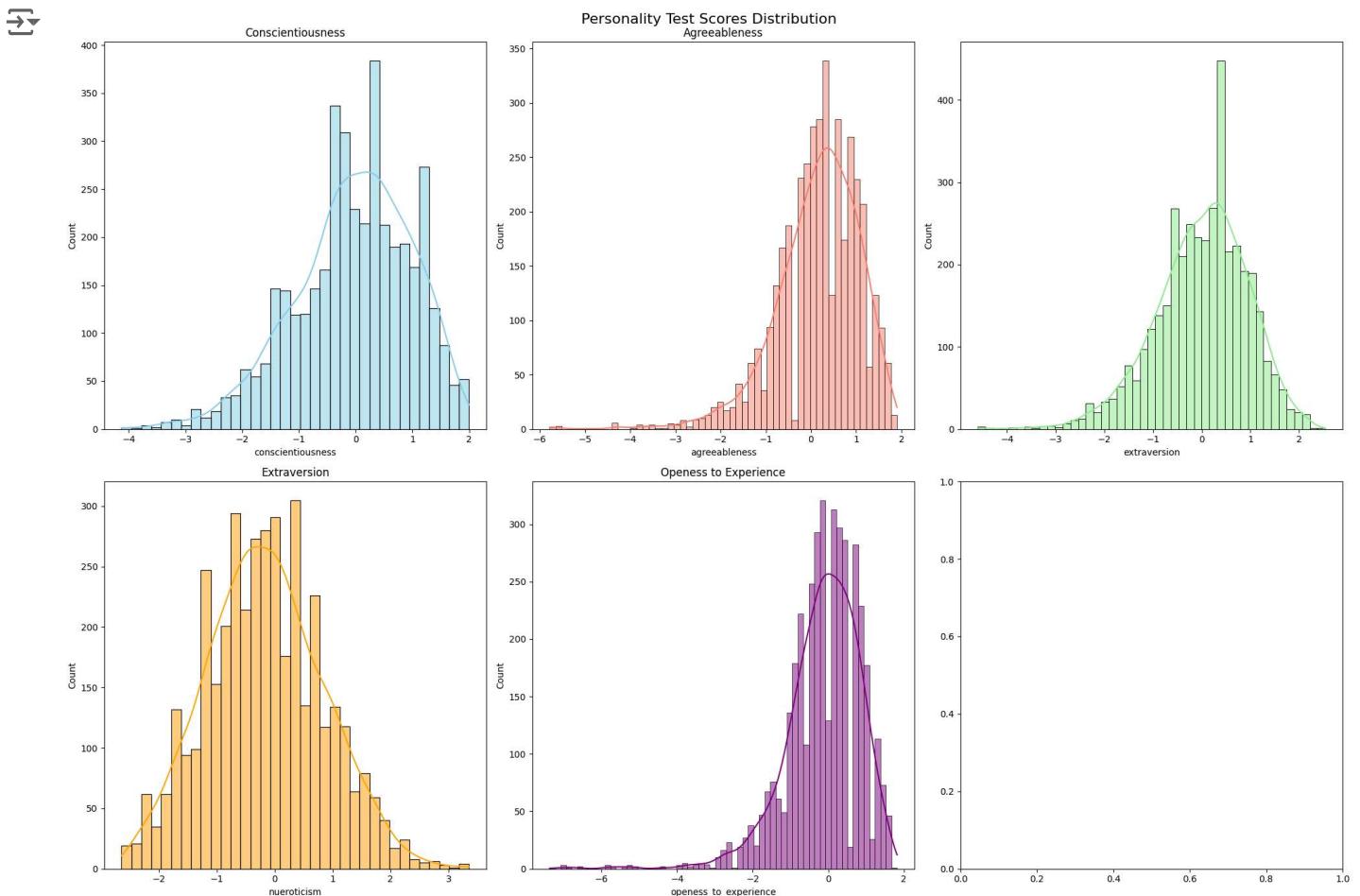
```
1 data['JobCities'].unique()
```

→ array(['Bangalore', 'Indore', 'Chennai', 'Gurgaon', 'Manesar',
 'Hyderabad', 'Noida', 'Kolkata', 'Pune', '-1', 'Mohali', 'Jhansi',
 'Delhi', 'Bhubaneshwar', 'Mumbai', 'Mangalore', 'Rewari',
 'Ghaziabad', 'Bhiwadi', 'Mysore', 'Rajkot', 'Greater Noida',
 'Jaipur', 'Thane', 'Maharajganj', 'Thiruvananthapuram',
 'Panchkula', 'Coimbatore', 'Dhanbad', 'Lucknow', 'Trivandrum',
 'Gandhi Nagar', 'Una', 'Daman and Diu', 'AM', 'Nagpur',
 'Bhagalpur', 'Ahmedabad', 'Kochi', 'Bankura', 'Bengaluru',
 'Kanpur', 'Vijayawada', 'Beawar', 'Alwar', 'Siliguri', 'Raipur',
 'Bhopal', 'Faridabad', 'Jodhpur', 'Udaipur', 'Muzaffarpur',
 'Bulandshahar', 'Haridwar', 'Raigarh', 'Visakhapatnam', 'Jabalpur',
 'Unnao', 'Aurangabad', 'Belgium', 'Dehradun', 'Rudrapur',
 'Jamshedpur', 'Vizag', 'Dharamshala', 'Hissar', 'Ranchi',
 'Madurai', 'Gurga', 'Chandigarh', 'Australia', 'CHEYYAR',
 'Sonipat', 'Pantnagar', 'Jagdalpur', 'Angul', 'Baroda', 'Ariyalur',
 'Jowai', 'Kochi/Cochin, Chennai and Coimbatore', 'Neemrana',
 'Tirupati', 'Calicut', 'Dubai', 'bengaluru', 'Ahmednagar',
 'Nashik', 'Bellary', 'Ludhiana', 'Muzaffarnagar', 'Gagret',
 'Indirapuram', 'Ghaziabad', 'Gwalior', 'Chennai & Mumbai',
 'Rajasthan', 'Bareli', 'Hospete', 'Miryalaguda', 'Dharuhera',
 'Meerut', 'Ganjam', 'Hubli', 'NCR', 'Agra', 'Trichy',
 'kudankulam ,tarapur', 'Ongole', 'Sambalpur', 'Pondicherry',
 'Bundi', 'Bikaner', 'Vadodara', 'Bathinda', 'Asansol',
 'Tirunelveli', 'Ernakulam', 'Bilaspur', 'Chandrapur', 'Nanded',
 'Dharmapuri', 'Vandavasi', 'Rohtak', 'PATNA', 'Salem', 'Nasikcity',
 'Bharuch', 'Tornagallu', 'Jaspur', 'Burdwan', 'Shimla',
 'Gajiabaad', 'Jammu', 'SHAHDOL', 'Muvattupuzha', 'Saudi Arabia',
 'Kalmar, Sweden', 'Secunderabad', 'Ratnagiri', 'Jhajjar',
 'Gulbarga', 'Nalagarh', 'Jeddah Saudi Arabia',
 'Chennai, Bangalore', 'Jamnagar', 'Gonda', 'orissa', 'kharagpur',
 'Navi Mumbai , Hyderabad', 'Joshimath', 'Johannesburg', 'Karnal',
 'London', 'Kota', 'Baddi HP', 'Nagari', 'Mettur, Tamil Nadu ',

```
'Durgapur', 'Surat', 'Kurnool', 'Kolhapur', 'Bhilai',
'Bahadurgarh', 'Rayagada, Odisha', 'Kakinada', 'Varanasi', 'punr',
'Nellore', 'Sahibabad', 'Howrah', 'Trichur', 'Ambala', 'Khopoli',
'Kerala', 'Roorkee', 'Allahabad', 'Jalandhar', 'Vapi', 'Pilani',
'RAS AL KHAIMAH', 'Bihar', 'singaruli', 'pondy', 'Phagwara',
'Baripada', 'Yamuna Nagar', 'Guwahati', 'Rourkela', 'Vellore',
'Dausa', 'Latur (Maharashtra )', 'Mainpuri', 'Dammam', 'Haldia',
'RAE BARELI', 'Patiala', 'Gorakhpur', 'new dehli', 'Karad',
'Rajpura', 'Haryana'], dtype=object)
```

❖ Univariate Analysis

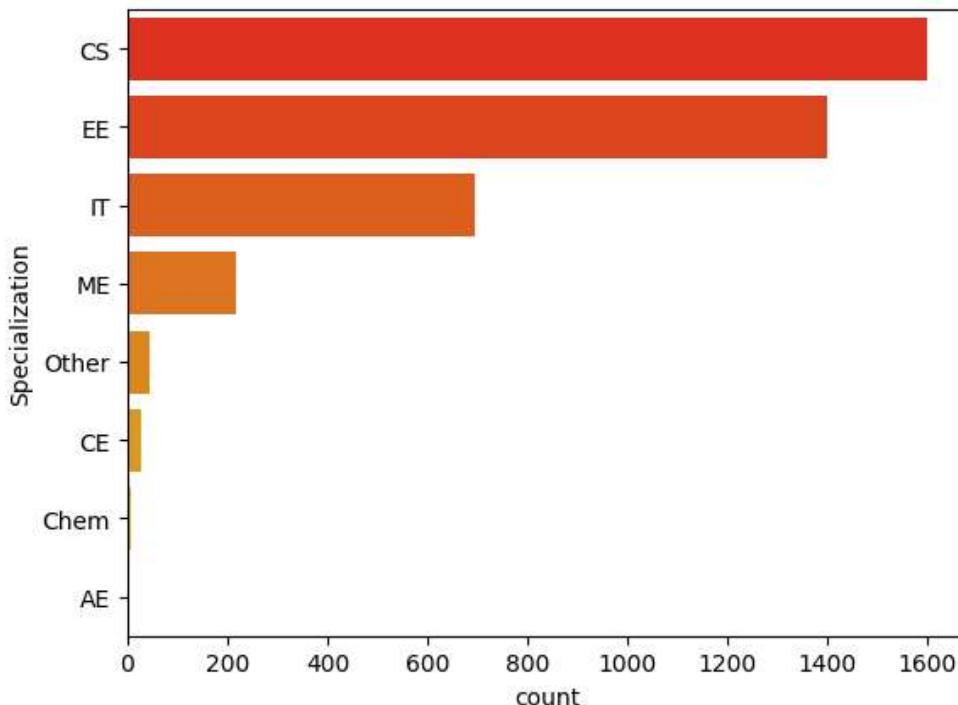
```
1 fig, axes = plt.subplots(2, 3, figsize=(20, 14))
2
3 fig.suptitle('Personality Test Scores Distribution', fontsize=16)
4
5 sns.histplot(ax=axes[0, 0], x=data['conscientiousness'], kde=True, color='skyblue')
6 axes[0, 0].set_title('Conscientiousness')
7
8 sns.histplot(ax=axes[0, 1], x=data['agreeableness'], kde=True, color='salmon')
9 axes[0, 1].set_title('Agreeableness')
10
11 sns.histplot(ax=axes[0, 2], x=data['extraversion'], kde=True, color='lightgreen')
12 axes[1, 0].set_title('Extraversion')
13
14 sns.histplot(ax=axes[1, 0], x=data['nueroticism'], kde=True, color='orange')
15 axes[1, 1].set_title('Nueroticism')
16
17 sns.histplot(ax=axes[1, 1],x=data['openess_to_experience'], kde=True, color='purple') # Added to the plot
18 axes[1,1].set_title('Openess to Experience') # Added to the plot
19
20
21 plt.tight_layout()
22 plt.show()
```



Following shows that Candidates have similar kind of Performance. But neuroticism have a bell curve equally distributed plot rest all are slightly right-skewed.

```
1 sns.countplot(data[ 'Specialization'],order=data[ 'Specialization'].value_counts().index,palette='autumn')
```

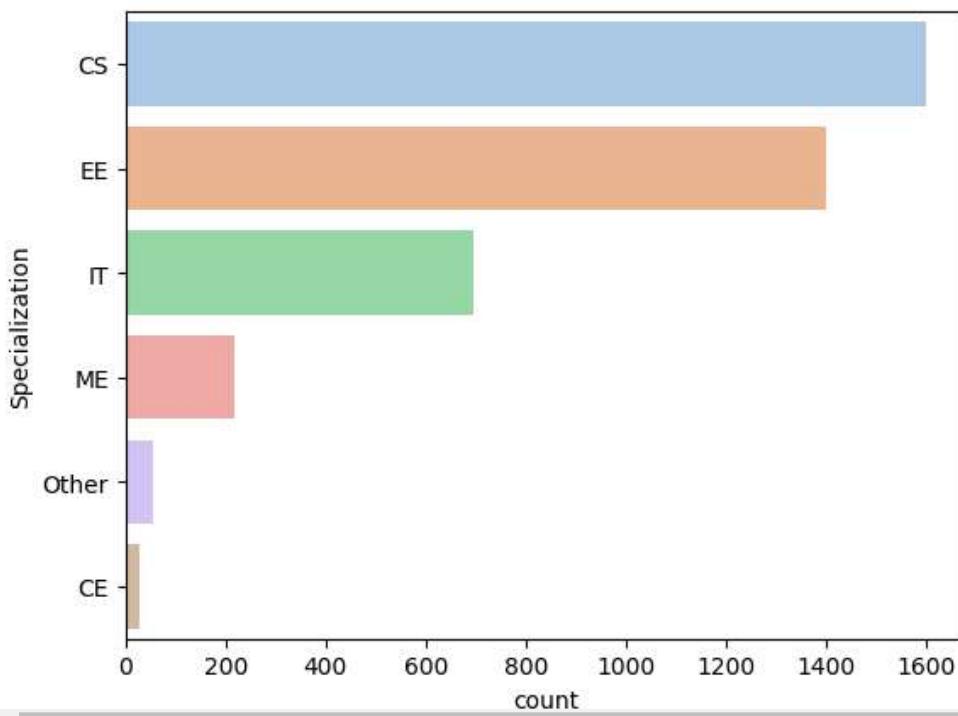
```
↳ <Axes: xlabel='count', ylabel='Specialization'>
```



Following top 5 Specialization shows the popularity of specialization.

```
1 Domain_replace = {'Chem': 'Other', 'AE': 'Other'}  
2 data['Specialization'].replace(Domain_replace,inplace=True)  
3 sns.countplot(data['Specialization'],order=data['Specialization'].value_counts().index,palette='pastel')
```

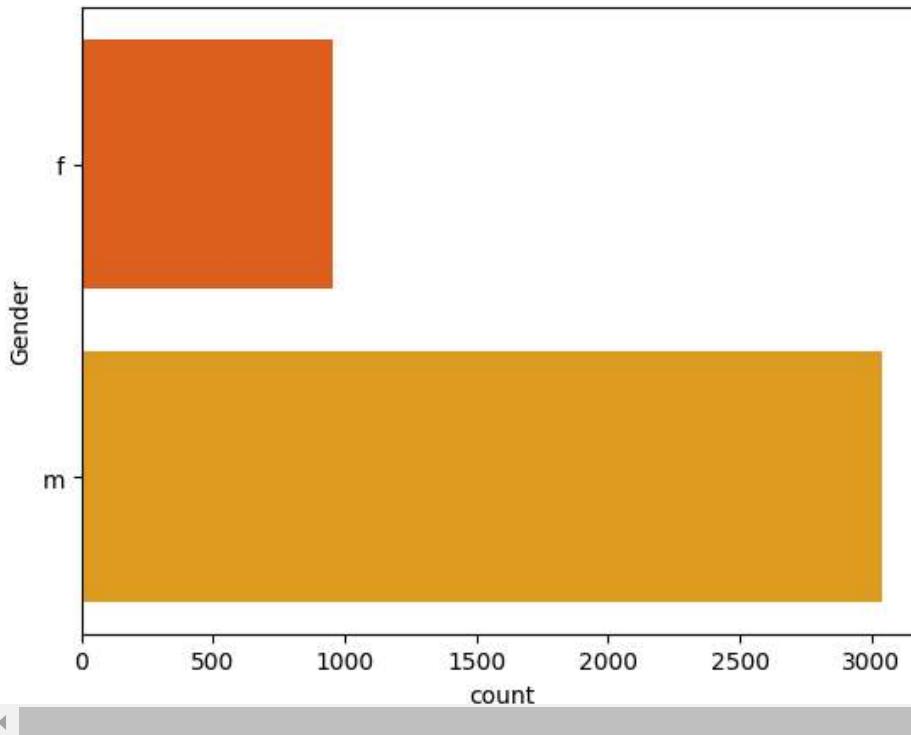
```
↳ <Axes: xlabel='count', ylabel='Specialization'>
```



Following top 5 Specialization shows the popularity of specialization.

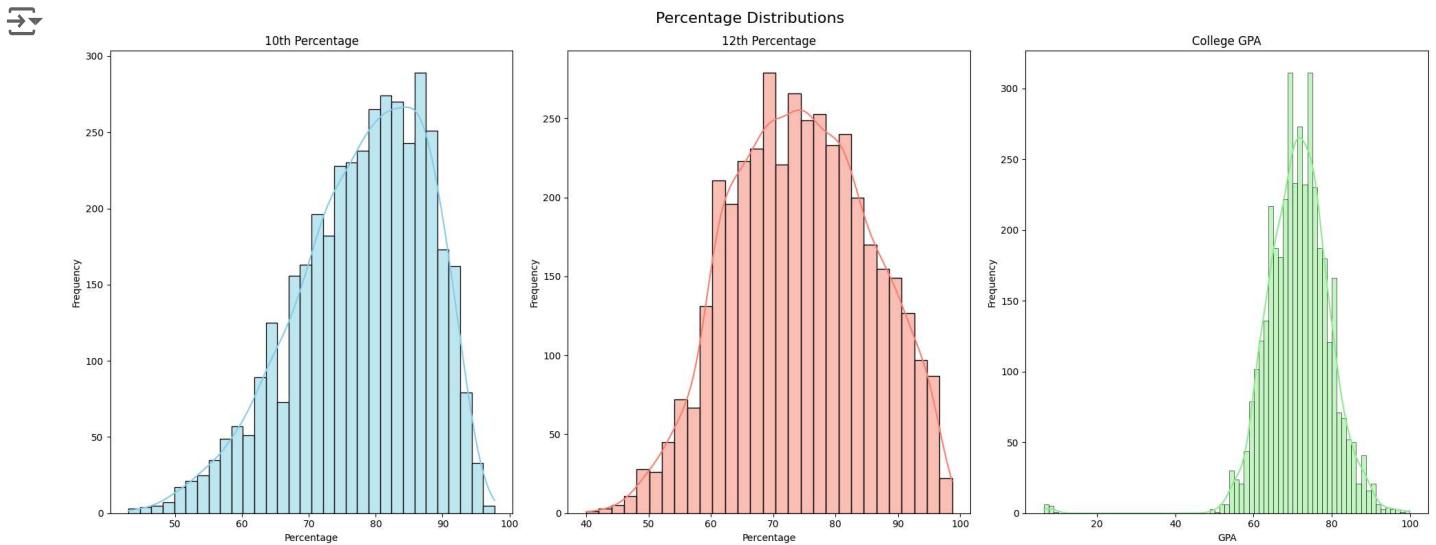
```
1 sns.countplot(data['Gender'], palette='autumn')
```

↳ <Axes: xlabel='count', ylabel='Gender'>



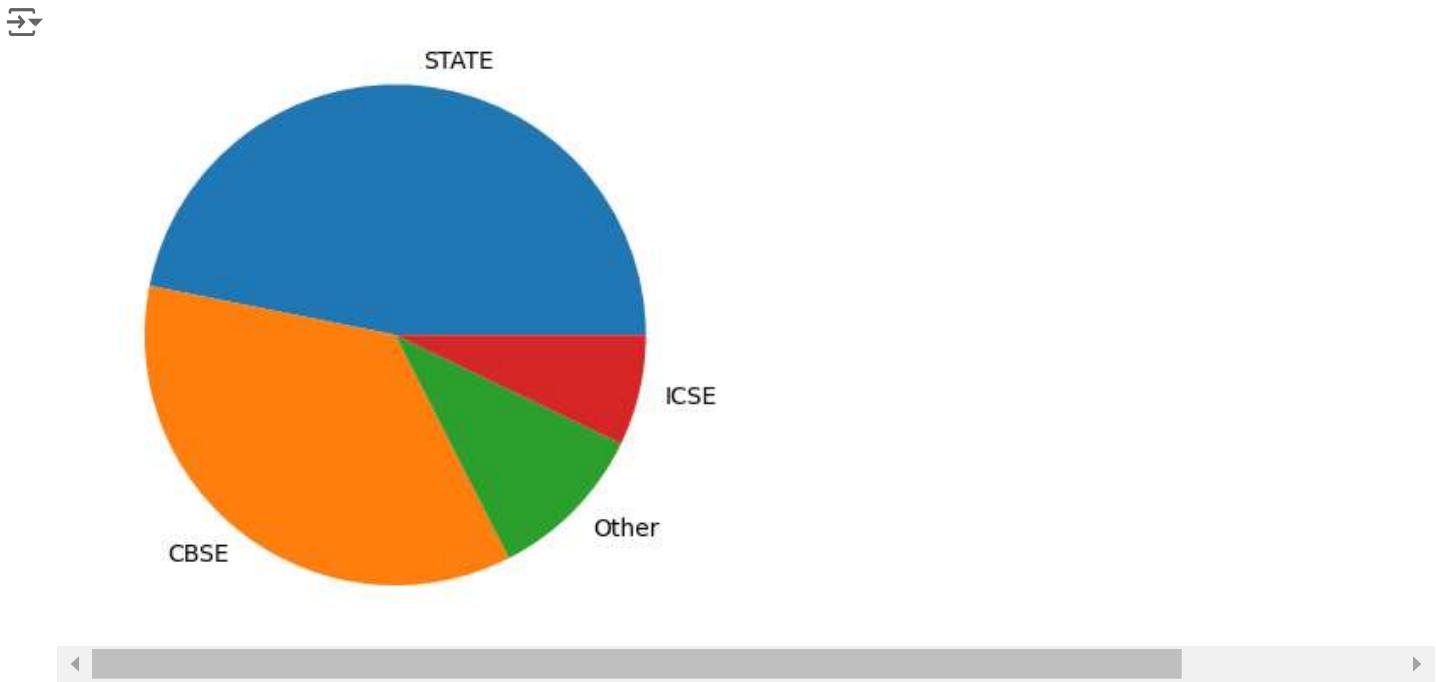
AMCAT has majority candidated with Gender Male.

```
1 tenth = data['10percentage']
2 twelve = data['12percentage']
3 clg_gpa = data['collegeGPA']
4
5 fig, axes = plt.subplots(1, 3, figsize=(20, 8))
6
7 fig.suptitle('Percentage Distributions', fontsize=16)
8
9 sns.histplot(ax=axes[0], x=tenth, kde=True, color='skyblue')
10 axes[0].set_title('10th Percentage')
11 axes[0].set_xlabel('Percentage')
12 axes[0].set_ylabel('Frequency')
13
14 sns.histplot(ax=axes[1], x=twelve, kde=True, color='salmon')
15 axes[1].set_title('12th Percentage')
16 axes[1].set_xlabel('Percentage')
17 axes[1].set_ylabel('Frequency')
18
19 sns.histplot(ax=axes[2], x=clg_gpa, kde=True, color='lightgreen')
20 axes[2].set_title('College GPA')
21 axes[2].set_xlabel('GPA')
22 axes[2].set_ylabel('Frequency')
23
24 plt.tight_layout()
25 plt.show()
```



10 , 12 results have relationship of slight growth. But not with College GPA 20% of them have failed.

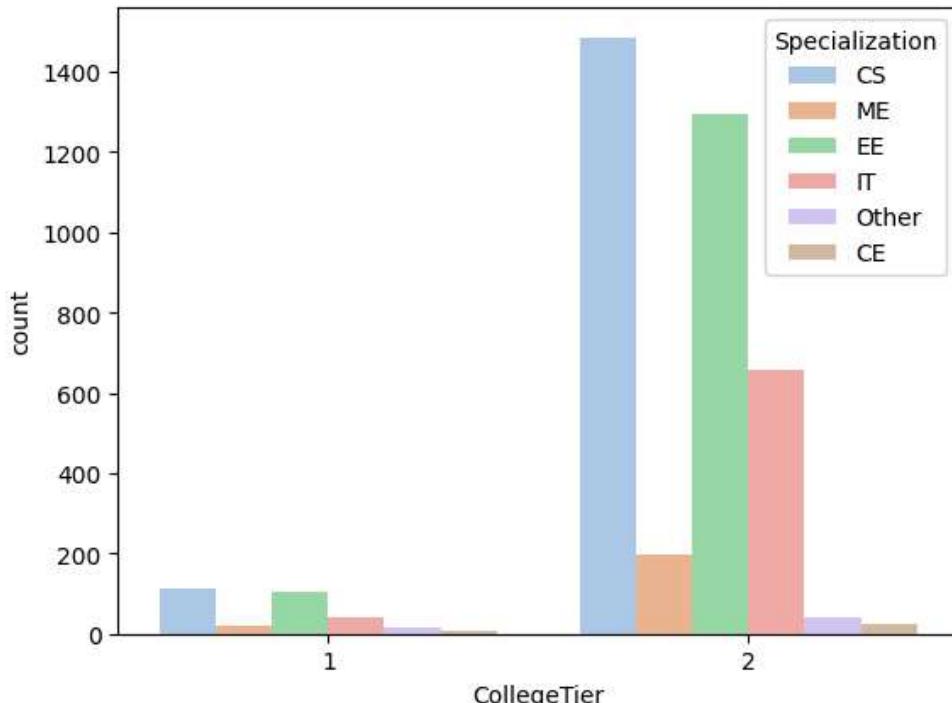
```
1 df=pd.DataFrame(data['XBoard'].value_counts())
2 plt.pie(df['count'],labels=df.index)
3 plt.show()
```



Majority of AMCAT Candidates are from State Board or CBSE Board.

```
1 sns.countplot(x='CollegeTier', hue='Specialization', data=data, palette='pastel')
```

→ <Axes: xlabel='CollegeTier', ylabel='count'>



Both College Tier 1 and 2 have similar kind of distribution but Tier 1 Colleges have Less amount of students were else totally opposite in tier 2 lots of students and huge number of colleges.

```
1 popular_role = list(pd.DataFrame(data['Designation'].value_counts()).head(15).index)
```

```
1 popular_role
```

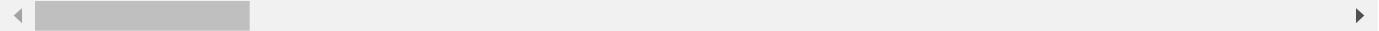
→ ['software engineer',
 'software developer',
 'system engineer',
 'programmer analyst',
 'systems engineer',
 'java software engineer',
 'software test engineer',
 'project engineer',
 'technical support engineer',
 'senior software engineer',
 'java developer',
 'test engineer',
 'web developer',
 'application developer',
 'assistant manager']

```
1 temp = data[data['Designation'].isin(popular_role)]  

2 temp
```

	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	
1	579905	500000.0	2013-09-01	2024-10-03 20:09:22.038981	assistant manager	Indore	m	1989-10-04	85.40	
2	810601	325000.0	2014-06-01	2024-10-03 20:09:22.038981	systems engineer	Chennai	f	1992-08-03	85.00	
3	267447	1100000.0	2011-07-01	2024-10-03 20:09:22.038981	senior software engineer	Gurgaon	m	1989-12-05	85.60	
5	1027655	300000.0	2014-06-01	2024-10-03 20:09:22.038981	system engineer	Hyderabad	m	1992-07-02	89.92	
6	947847	300000.0	2014-08-01	2015-05-01 00:00:00.000000	java software engineer	Banglore	m	1993-02-01	86.08	
...
3985	358393	430000.0	2013-03-01	2024-10-03 20:09:22.038981	system engineer	Gurgaon	f	1990-05-14	90.40	
3986	51897	225000.0	2010-01-01	2010-10-01 00:00:00.000000	software test engineer	Bangalore	f	1986-02-28	84.93	
3989	1204604	300000.0	2014-09-01	2024-10-03 20:09:22.038981	software engineer	Bangalore	m	1991-11-23	74.88	
3993	47916	280000.0	2011-10-01	2012-10-01 00:00:00.000000	software engineer	New Delhi	m	1987-04-15	52.09	
3996	947111	200000.0	2014-07-01	2015-01-01 00:00:00.000000	software developer	Asifabadbanglore	f	1992-03-20	78.72	

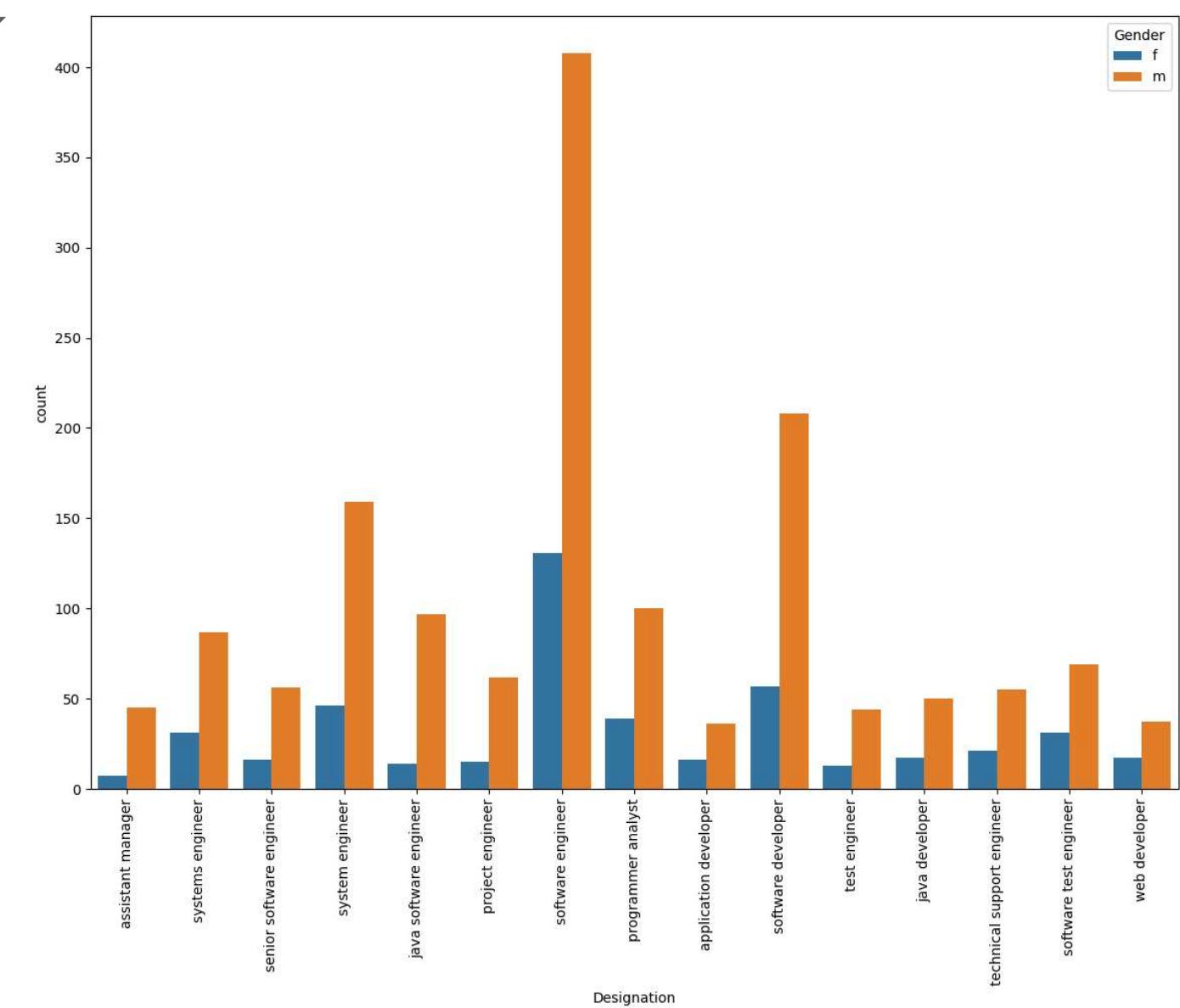
1984 rows × 41 columns



```

1 plt.figure(figsize=(14,10))
2 sns.countplot(x=data.loc[data['Designation'].isin(popular_role),'Designation'],hue=data['Gender'])
3 plt.xticks(rotation=90)
4 plt.show()

```

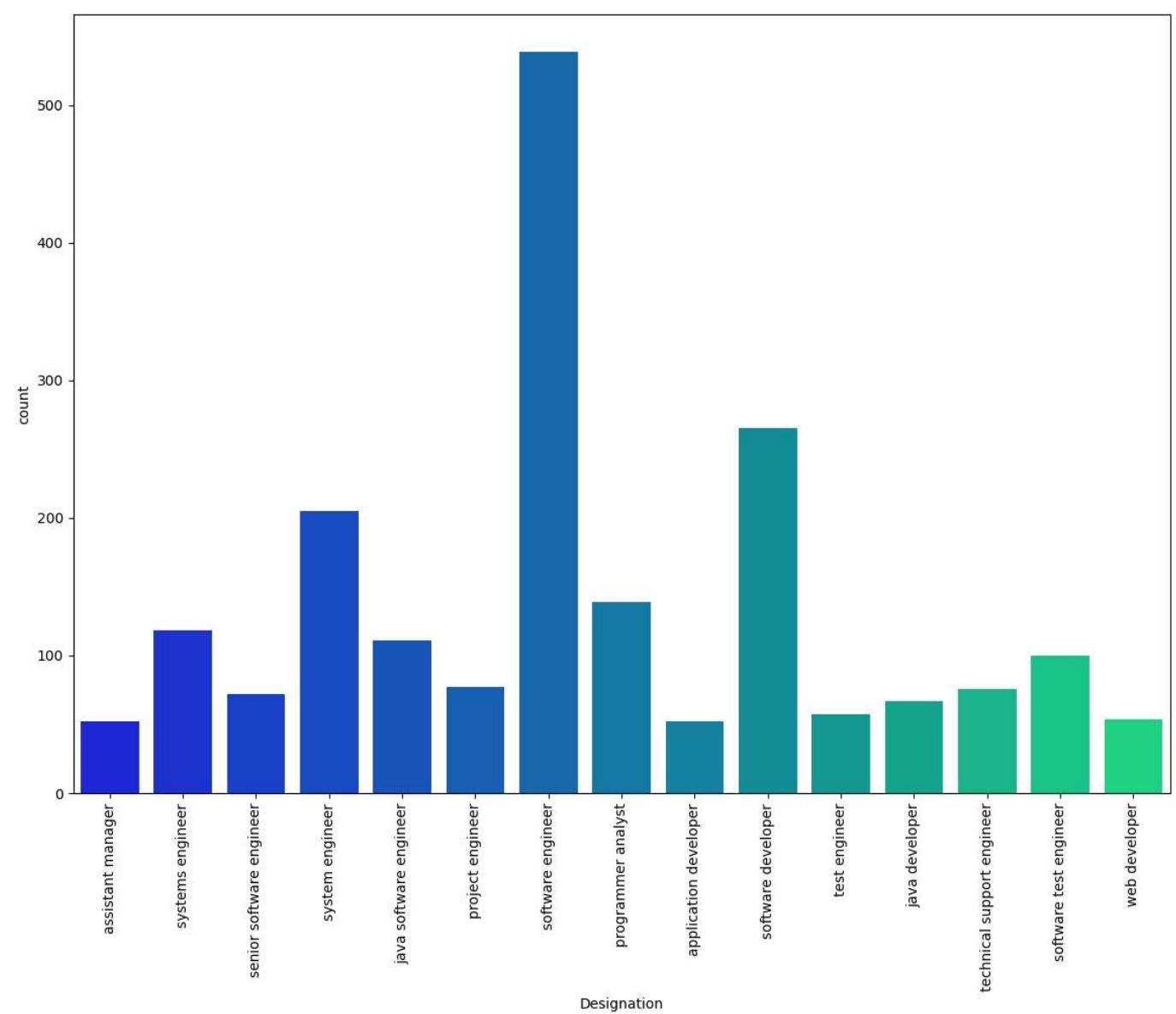


With in Male & Female top 3 Designation are Software Engineer, Software Developer, System Engineer.

```

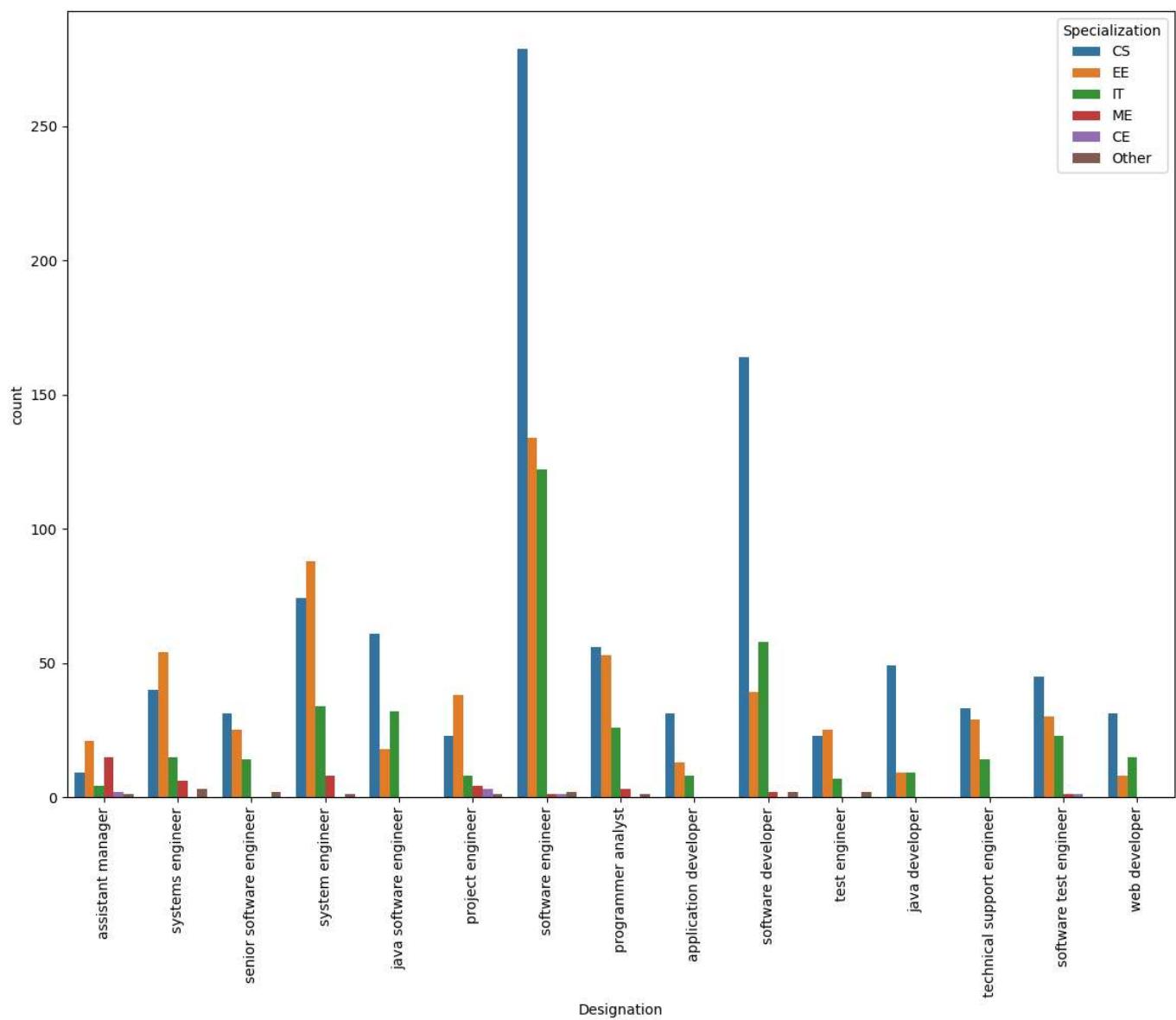
1 plt.figure(figsize=(14,10))
2 sns.countplot(x=data.loc[data['Designation'].isin(popular_role), 'Designation'], palette='winter')
3 plt.xticks(rotation=90)
4 plt.show()

```



Top 3 Designation are Software Engineer, Software Developer, System Engineer.

```
1 plt.figure(figsize=(14,10))
2 sns.countplot(x=data.loc[data['Designation'].isin(popular_role), 'Designation'], hue=data['Specialization'])
3 plt.xticks(rotation=90)
4 plt.show()
```



Most popular Designation have similar Specialization but among them popular Specialization is CS, EE, IT.

```
1 popular_city = list(pd.DataFrame(data['JobCities'].value_counts()).head(15).index)
2 popular_city
```

```
[ Bangalore',
 '-1',
 'Noida',
 'Hyderabad',
 'Pune',
 'Chennai',
 'Delhi',
 'Gurgaon',
 'Mumbai',
 'Kolkata',
 'Jaipur',
 'Lucknow',
 'Mysore',
```

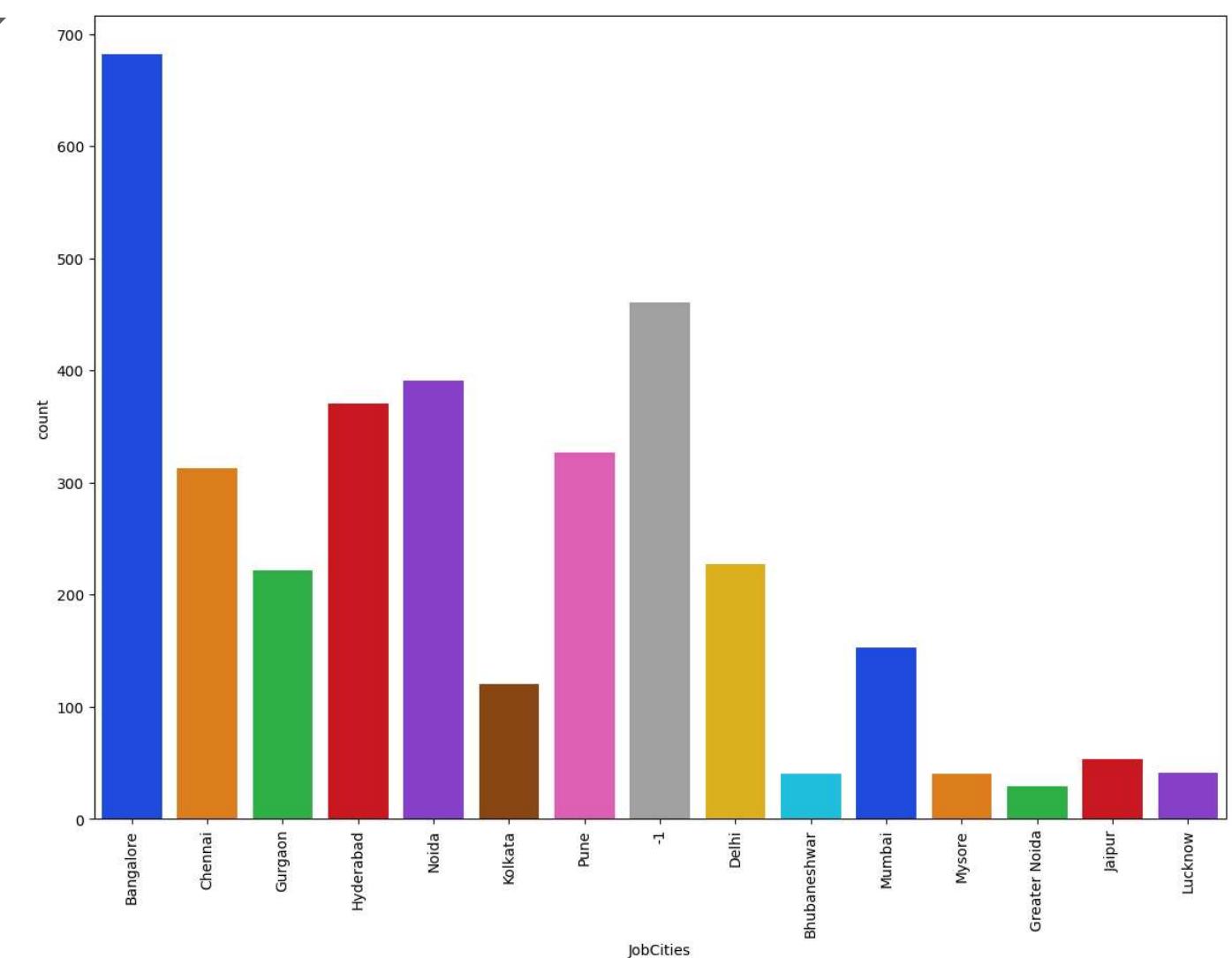
```
'Bhubaneshwar',
'Greater Noida']
```

```
1 temp_1 = data[data['JobCities'].isin(popular_city)]
2 temp_1
```

	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage :
0	203097	420000.0	2012-06-01	2024-10-03 20:09:22.038981	senior quality engineer	Bangalore	f	1990-02-19	84.30
2	810601	325000.0	2014-06-01	2024-10-03 20:09:22.038981	systems engineer	Chennai	f	1992-08-03	85.00
3	267447	1100000.0	2011-07-01	2024-10-03 20:09:22.038981	senior software engineer	Gurgaon	m	1989-12-05	85.60
5	1027655	300000.0	2014-06-01	2024-10-03 20:09:22.038981	system engineer	Hyderabad	m	1992-07-02	89.92
6	947847	300000.0	2014-08-01	2015-05-01 00:00:00.000000	java software engineer	Banglore	m	1993-02-01	86.08
...
3993	47916	280000.0	2011-10-01	2012-10-01 00:00:00.000000	software engineer	New Delhi	m	1987-04-15	52.09
3994	752781	100000.0	2013-07-01	2013-07-01 00:00:00.000000	technical writer	Hyderabad	f	1992-08-27	90.00
3995	355888	320000.0	2013-07-01	2024-10-03 20:09:22.038981	associate software engineer	Bangalore	m	1991-07-03	81.86
3996	947111	200000.0	2014-07-01	2015-01-01 00:00:00.000000	software developer	Asifabadbanglore	f	1992-03-20	78.72
3997	324966	400000.0	2013-02-01	2024-10-03 20:09:22.038981	senior systems engineer	Chennai	f	1991-02-26	70.60

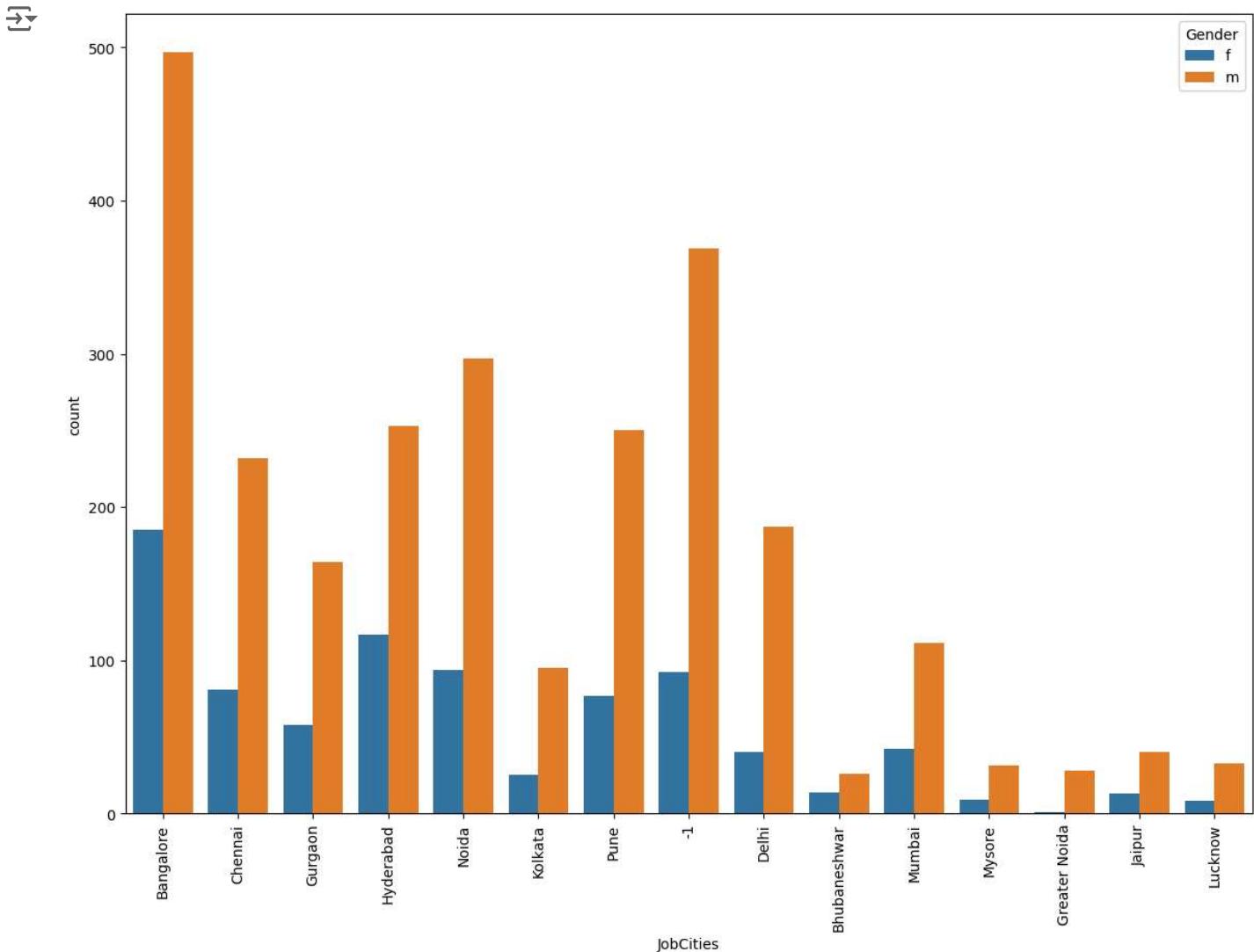
3469 rows × 41 columns

```
1 plt.figure(figsize=(14,10))
2 sns.countplot(x=data.loc[data['JobCities'].isin(popular_city),'JobCities'], palette='bright')
3 plt.xticks(rotation=90)
4 plt.show()
```



Most popular knowned cities for job are Bangalore, Noida, Hyderabad, Chennai.

```
1 plt.figure(figsize=(14,10))
2 sns.countplot(x=data.loc[data['JobCities'].isin(popular_city),'JobCities'],hue=data['Gender'])
3 plt.xticks(rotation=90)
4 plt.show()
```



No job city has different distribution and relationship between male and female.

```

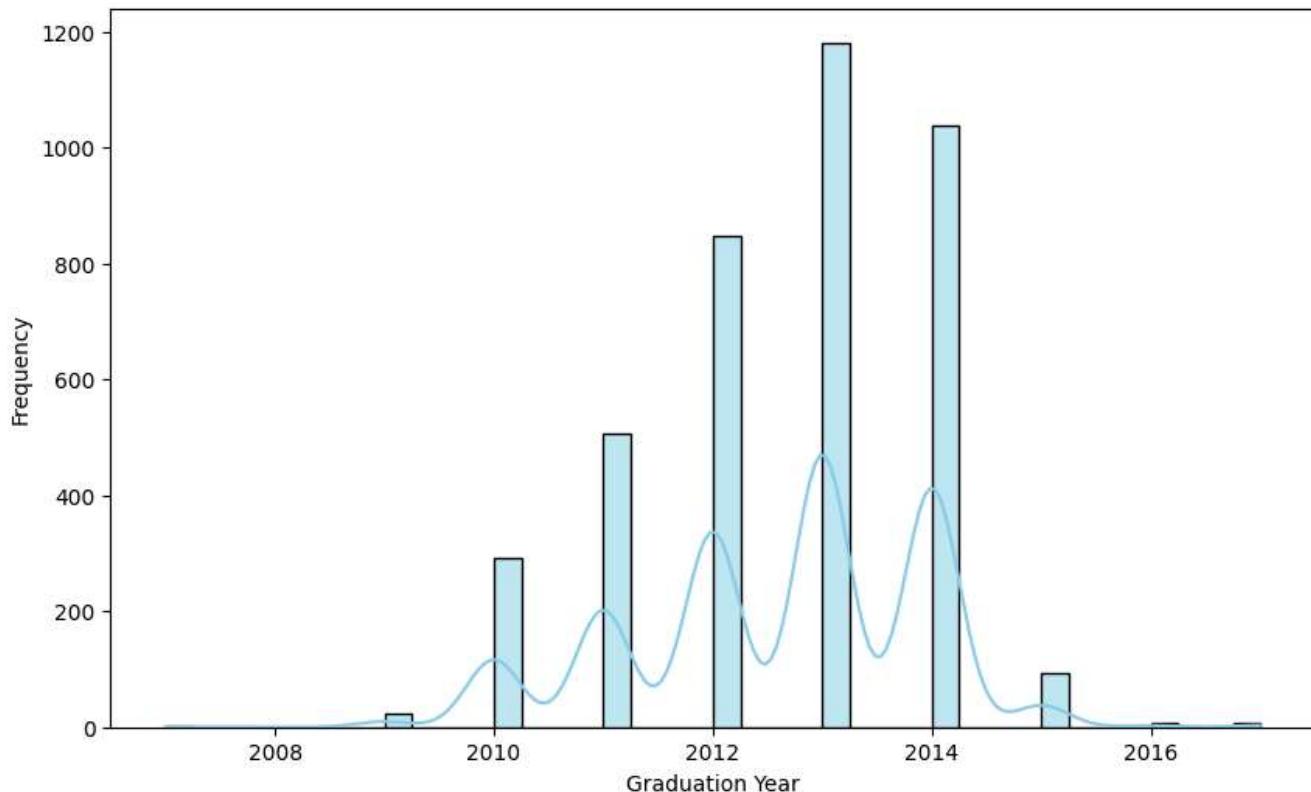
1 # Univariate analysis for 'GraduationYear'
2 plt.figure(figsize=(10, 6))
3 sns.histplot(data['GraduationYear'], kde=True, color='skyblue')
4 plt.title('Distribution of Graduation Year')
5 plt.xlabel('Graduation Year')
6 plt.ylabel('Frequency')
7 plt.show()
8
9 # Univariate analysis for 'collegeGPA'
10 plt.figure(figsize=(10, 6))
11 sns.boxplot(y=data['collegeGPA'], color='salmon')
12 plt.title('Distribution of College GPA')
13 plt.ylabel('College GPA')
14 plt.show()
15
16 # Univariate analysis for '10percentage'
17 plt.figure(figsize=(10, 6))
18 sns.violinplot(y=data['10percentage'], color='lightgreen')
19 plt.title('Distribution of 10th Percentage')
20 plt.ylabel('10th Percentage')

```

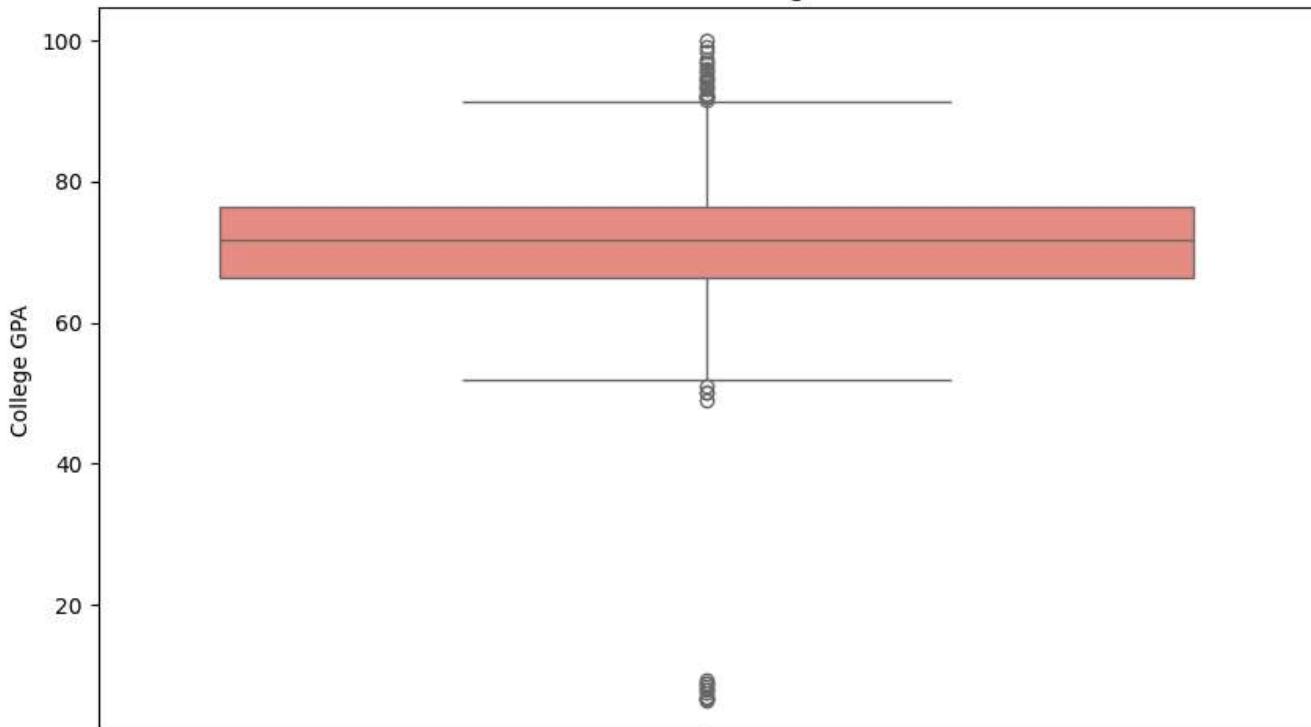
```
20 plt.ylabel('Total Percentage')
21 plt.show()
22
23 # Univariate analysis for '12percentage'
24 plt.figure(figsize=(10, 6))
25 sns.histplot(data['12percentage'], kde=True, bins=20, color='orange')
26 plt.title('Distribution of 12th Percentage')
27 plt.xlabel('12th Percentage')
28 plt.ylabel('Frequency')
29 plt.show()
30
```



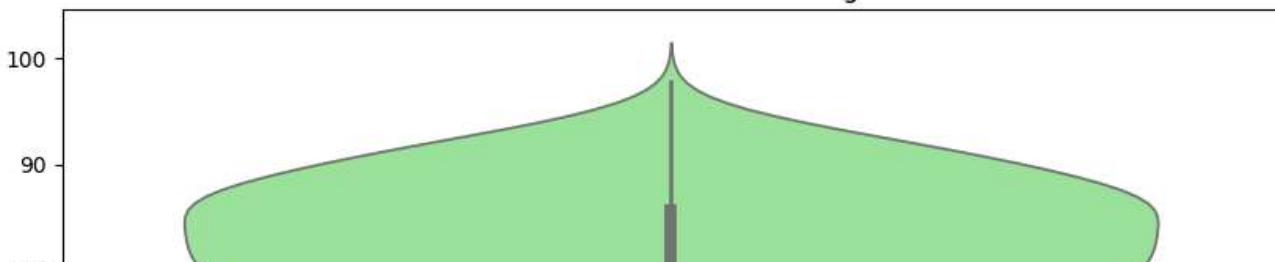
Distribution of Graduation Year

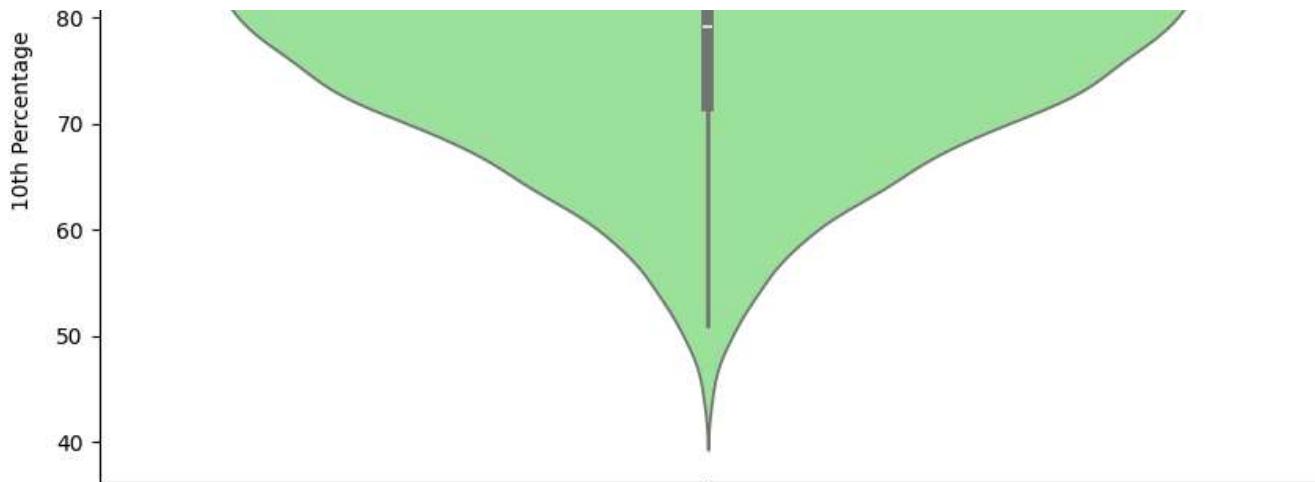


Distribution of College GPA

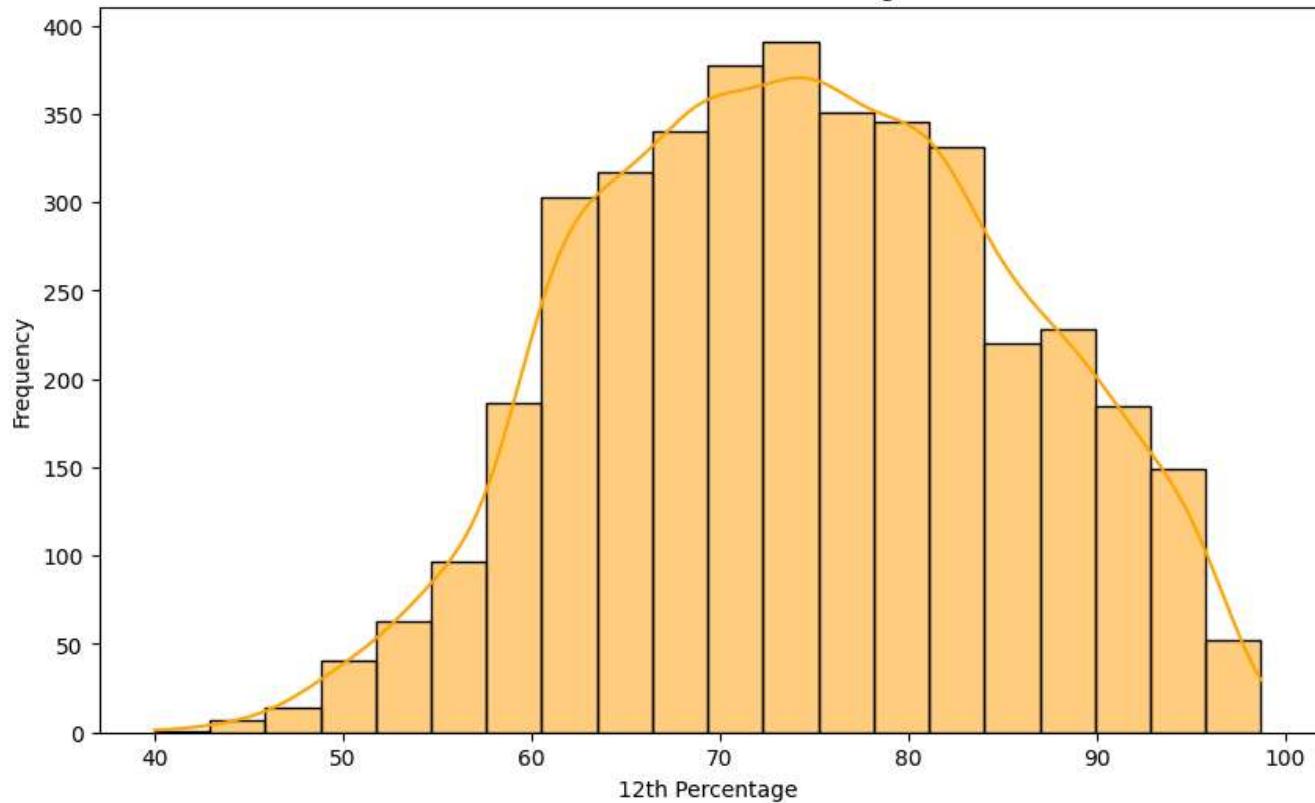


Distribution of 10th Percentage





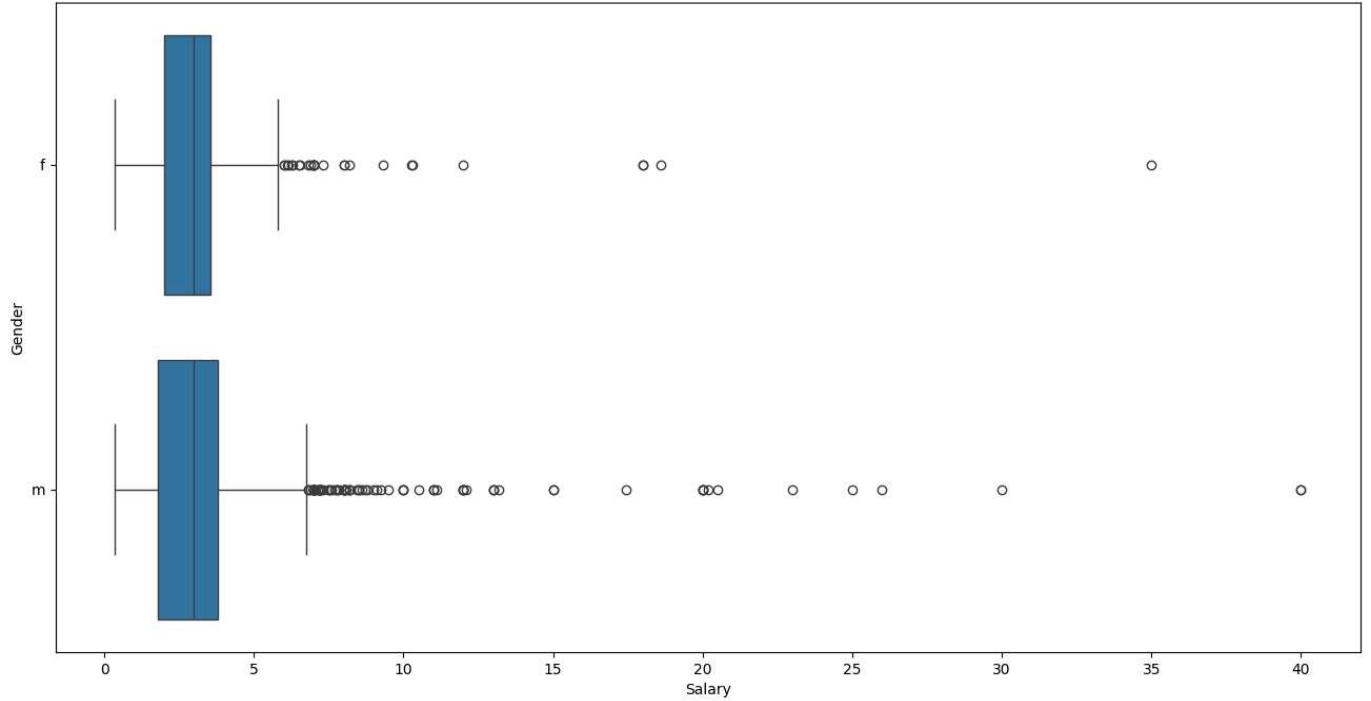
Distribution of 12th Percentage



❖ Bivariate analysis

```
1 plt.figure(figsize=(16,8))
2 sns.boxplot(x=data['Salary']/100000,y=data['Gender'])
```

↳ <Axes: xlabel='Salary', ylabel='Gender'>



As per the boxplot Male and Female has no difference on Salary.

```
1 temp_data = pd.DataFrame(data[data['Tenure']>=0])
2 temp_data.shape
```

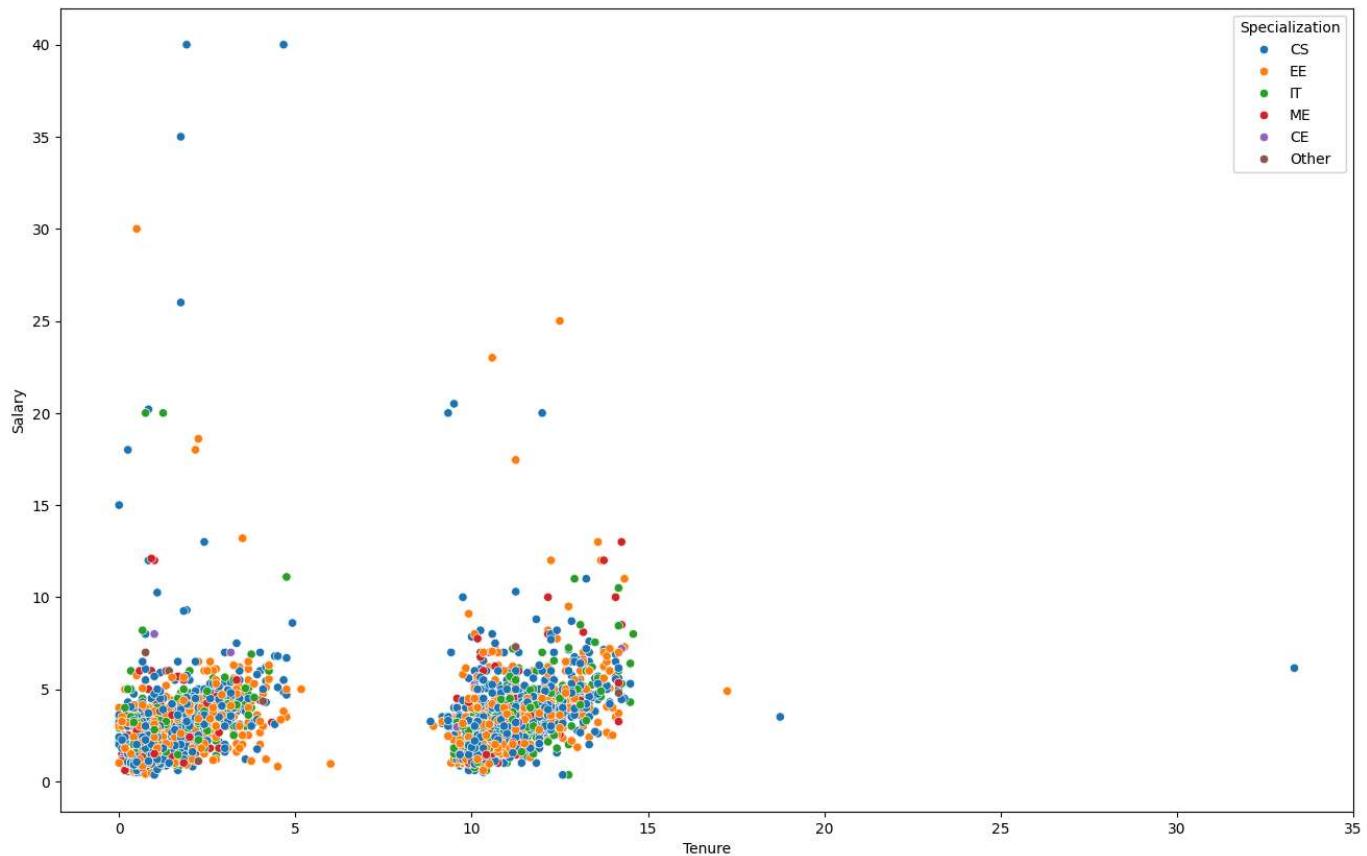
↳ (3958, 41)

```
1 temp_data.head()
```

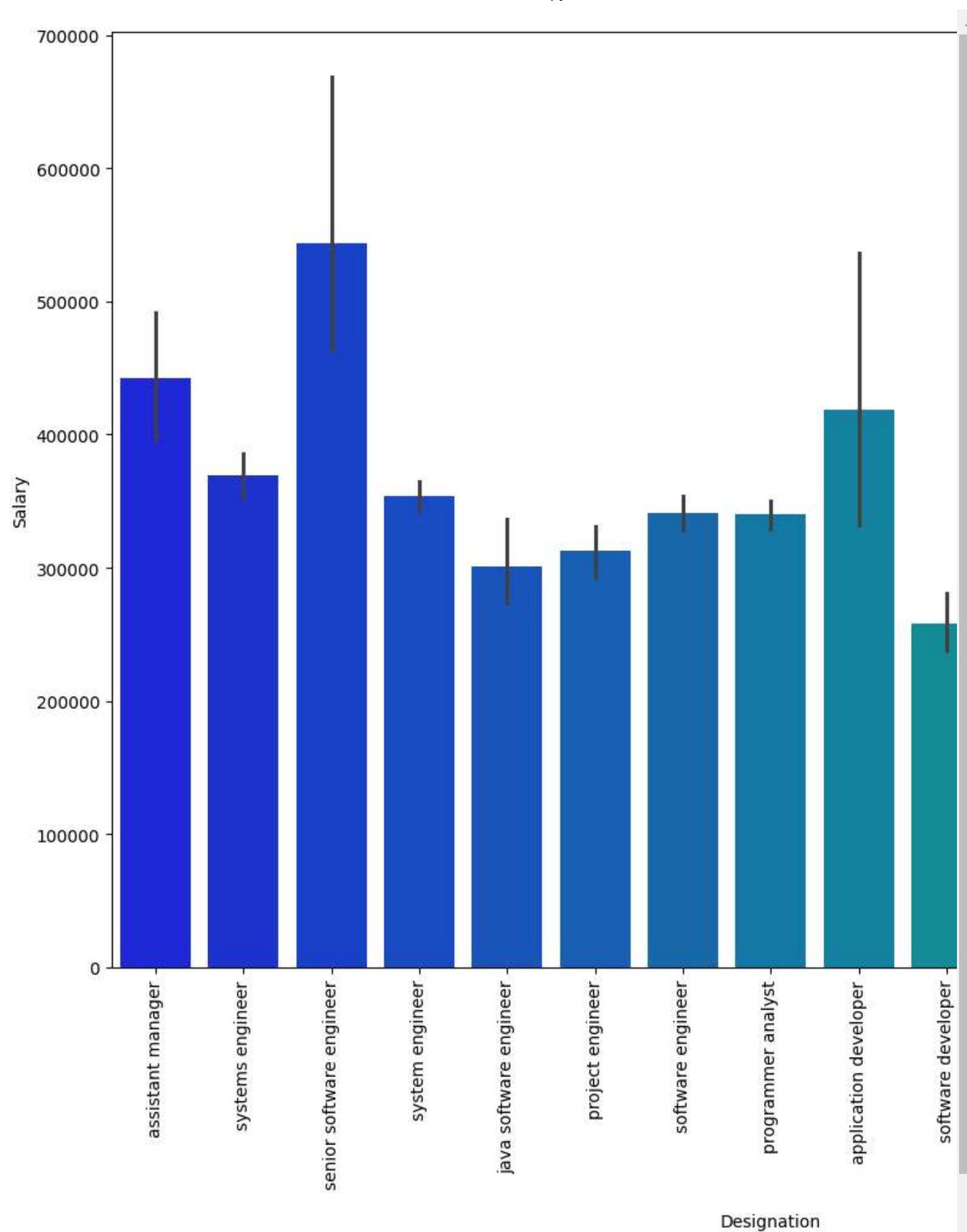
	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	12graduation
0	203097	420000.0	2012-06-01	2024-10-03 20:09:22.038981	senior quality engineer	Bangalore	f	1990-02-19	84.3	201
1	579905	500000.0	2013-09-01	2024-10-03 20:09:22.038981	assistant manager	Indore	m	1989-10-04	85.4	201
2	810601	325000.0	2014-06-01	2024-10-03 20:09:22.038981	systems engineer	Chennai	f	1992-08-03	85.0	201
3	267447	1100000.0	2011-07-01	2024-10-03 20:09:22.038981	senior software engineer	Gurgaon	m	1989-12-05	85.6	201
4	343523	200000.0	2014-03-01	2015-03-01 00:00:00.000000	get	Manesar	m	1991-02-27	78.0	201

```
1 plt.figure(figsize=(16,10))
2 sns.scatterplot(x=temp_data['Tenure']/12,y=temp_data['Salary']/100000,hue=temp_data['Specialization'])
```

↳ <Axes: xlabel='Tenure', ylabel='Salary'>

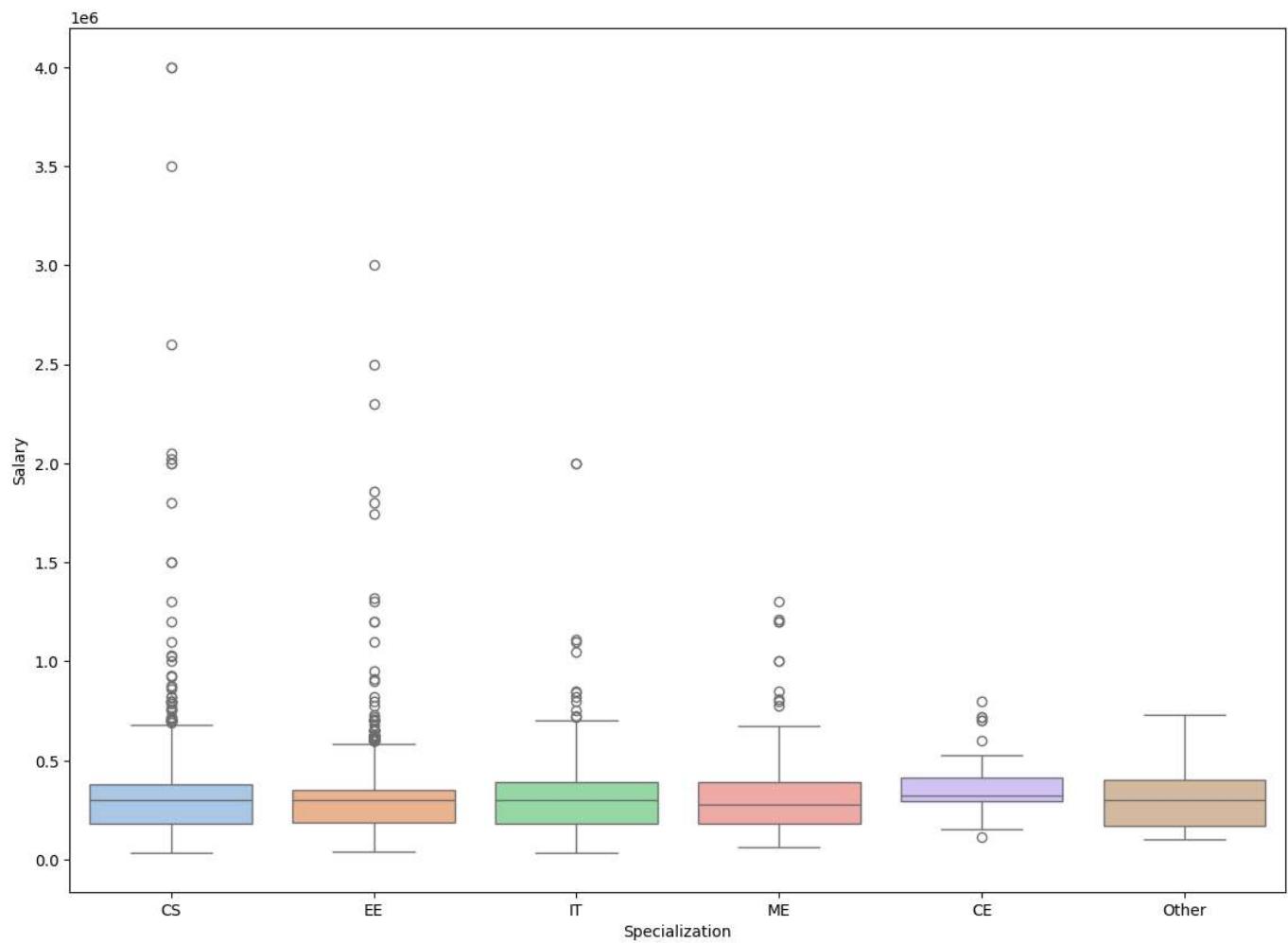


```
1 plt.figure(figsize=(14,10))
2 sns.barplot(x=data.loc[data['Designation'].isin(popular_role),'Designation'],y=data['Salary'],palette='winter')
3 plt.xticks(rotation=90)
4 plt.show()
```



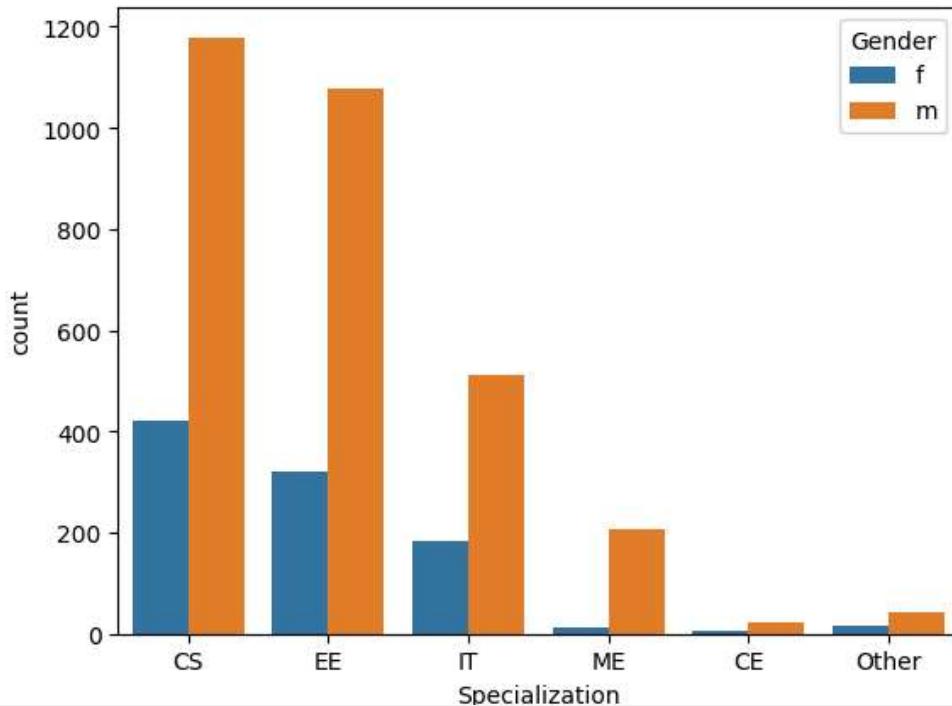
```
1 plt.figure(figsize=(14,10))
2 sns.boxplot(x=data['Specialization'],y=data['Salary'],palette='pastel')
```

→ <Axes: xlabel='Specialization', ylabel='Salary'>



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

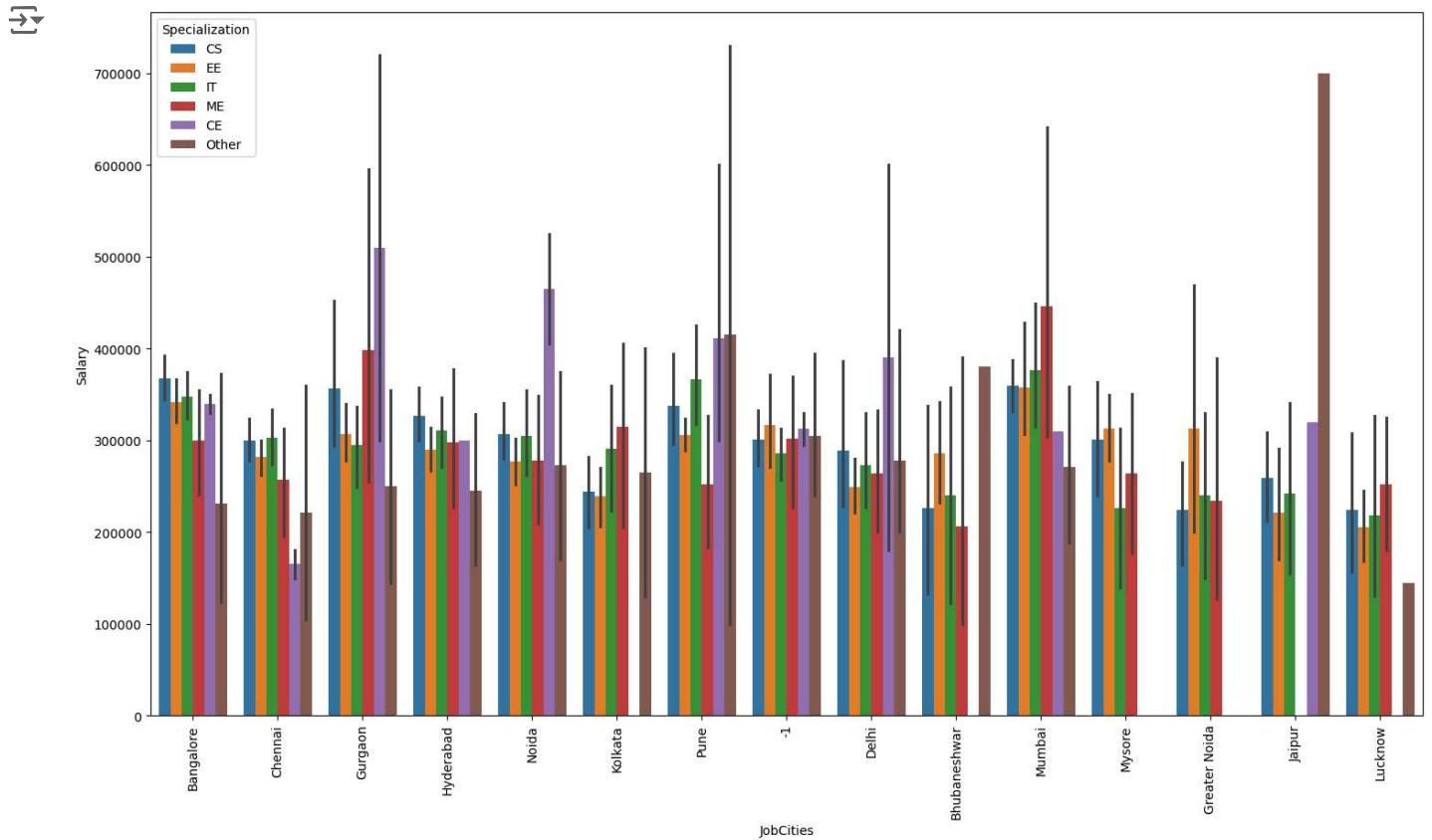
<Axes: xlabel='Specialization', ylabel='count'>



```

1 plt.figure(figsize=(18,10))
2 sns.barplot(x=data.loc[data['JobCities'].isin(popular_city),'JobCities'],y=data['Salary'],hue=data['Specialization'],
3 3 plt.xticks(rotation=90)
4 plt.show()

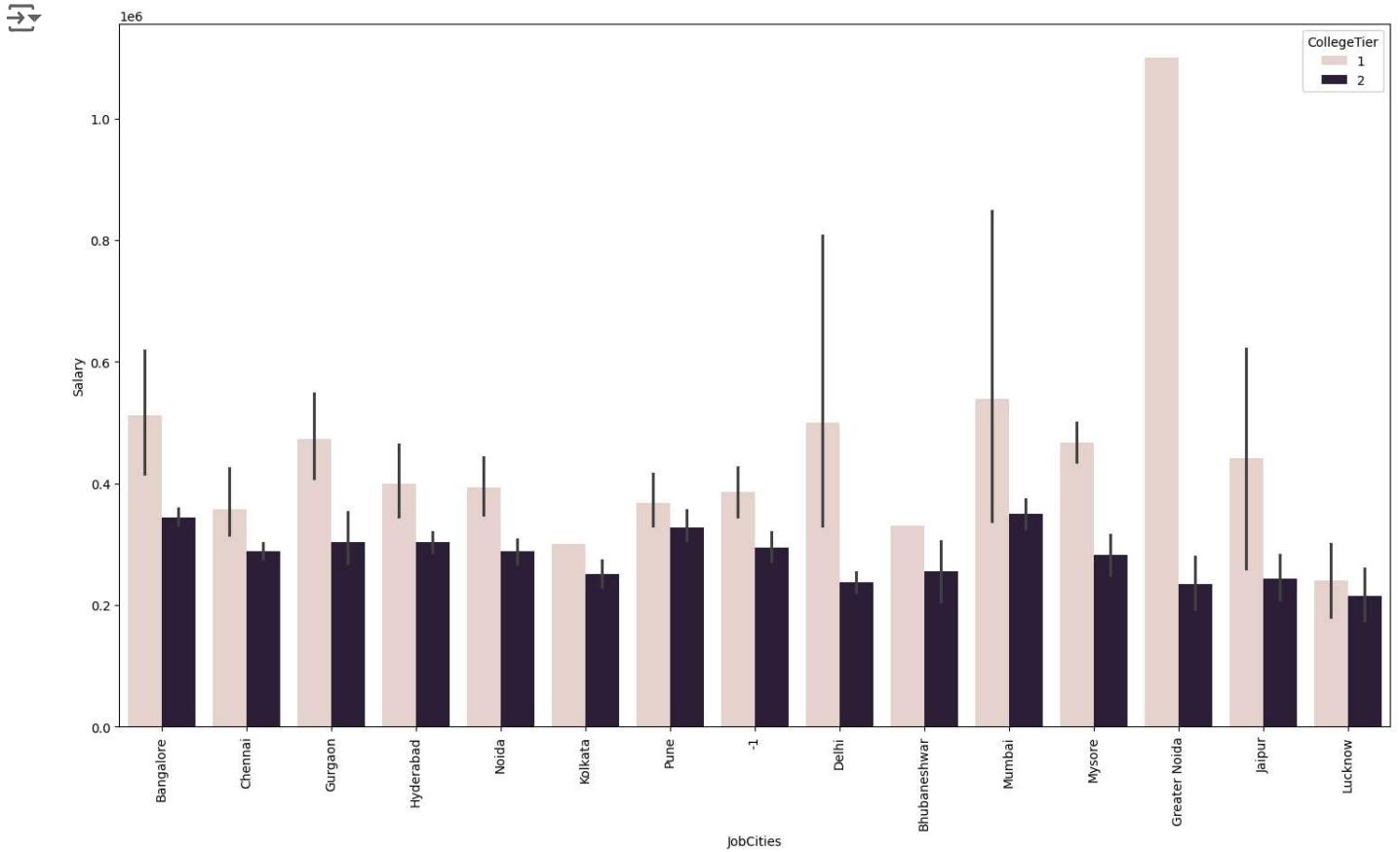
```



```

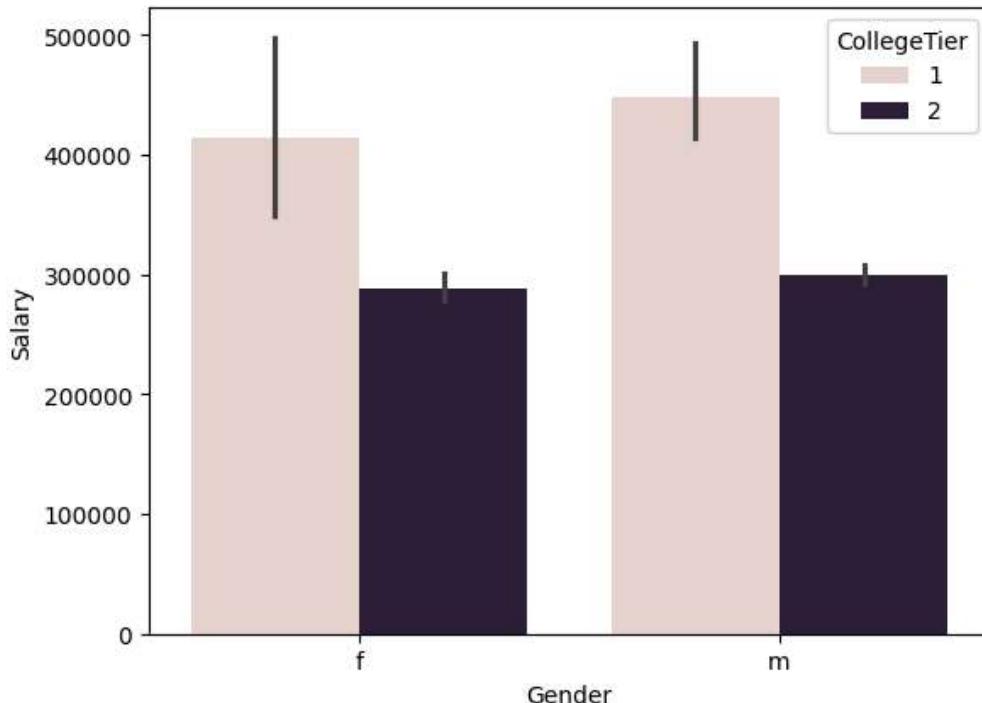
1 plt.figure(figsize=(18,10))
2 sns.barplot(x=data.loc[data['JobCities'].isin(popular_city),'JobCities'],y=data['Salary'],hue=data['CollegeTier'])
3 plt.xticks(rotation=90)
4 plt.show()

```

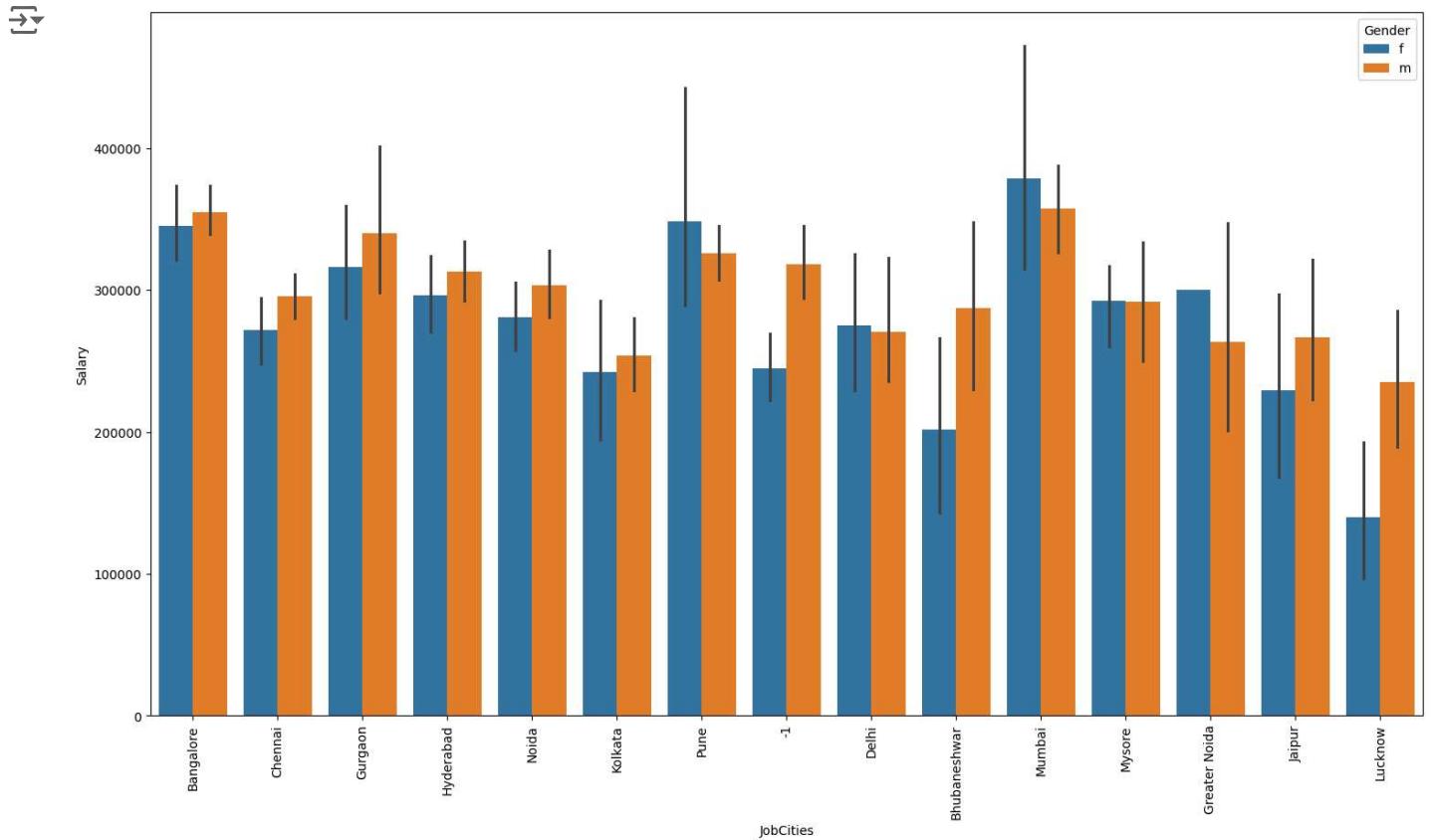


```
1 sns.barplot(x=data['Gender'],y=data['Salary'],hue=data['CollegeTier'])
```

<Axes: xlabel='Gender', ylabel='Salary'>



```
1 plt.figure(figsize=(18,10))
2 sns.barplot(x=data.loc[data['JobCities'].isin(popular_city),'JobCities'],y=data['Salary'],hue=data['Gender']
3 plt.xticks(rotation=90)
4 plt.show()
```



**THANK
YOU**

