

```

# importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from scipy.stats import skew, kurtosis
from datetime import datetime

# reading the data into the dataframe
df=pd.read_csv(r"C:\Users\vamsh\Downloads\AMCAT_data.csv")
df
df1=df #i am just copying the data as a backup here,incase if i have to
use it later

#checking if there are any null values in the data
df.isnull().sum()


```

Unnamed: 0	0
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	0
ComputerProgramming	0
ElectronicsAndSemicon	0
ComputerScience	0
MechanicalEngg	0
ElectricalEngg	0

```
TelecomEngg          0  
CivilEngg           0  
conscientiousness   0  
agreeableness       0  
extraversion        0  
nueroticism         0  
openness_to_experience 0  
dtype: int64
```

## #data transformation

```
df[df.columns[0]].nunique()  
#since there is only one value i.e train i am removing it  
#since my first column is not so important to me, i like to remove it  
from the dataframe  
df=df.drop(df.columns[0],axis=1) #alternative way df=df.iloc[:,1:]  
df
```

	ID	Salary	DOJ	DOL	\
0	203097	420000.0	6/1/12 0:00	present	
1	579905	500000.0	9/1/13 0:00	present	
2	810601	325000.0	6/1/14 0:00	present	
3	267447	1100000.0	7/1/11 0:00	present	
4	343523	200000.0	3/1/14 0:00	3/1/15 0:00	
...	...	...	...	...	
3993	47916	280000.0	10/1/11 0:00	10/1/12 0:00	
3994	752781	100000.0	7/1/13 0:00	7/1/13 0:00	
3995	355888	320000.0	7/1/13 0:00	present	
3996	947111	200000.0	7/1/14 0:00	1/1/15 0:00	
3997	324966	400000.0	2/1/13 0:00	present	

DOB	Designation	JobCity	Gender
0 0:00	senior quality engineer	Bangalore	f 2/19/90
1 0:00	assistant manager	Indore	m 10/4/89
2 0:00	systems engineer	Chennai	f 8/3/92
3 0:00	senior software engineer	Gurgaon	m 12/5/89
4 0:00	get	Manesar	m 2/27/91
...	...	...	...
3993 0:00	software engineer	New Delhi	m 4/15/87
3994 0:00	technical writer	Hyderabad	f 8/27/92
3995	associate software engineer	Bangalore	m 7/3/91

0:00					
3996	software developer	Asifabadbanglore	f	3/20/92	
0:00					
3997	senior systems engineer	Chennai	f	2/26/91	
0:00					
	10percentage	10board	...		
ComputerScience	\				
0	84.30	board ofsecondary education,ap	...		
-1					
1	85.40	cbse	...		
-1					
2	85.00	cbse	...		
-1					
3	85.60	cbse	...		
-1					
4	78.00	cbse	...		
-1					
...	...	...	...	...	.
3993	52.09	cbse	...		
-1					
3994	90.00	state board	...		
-1					
3995	81.86	bse,odisha	...		
-1					
3996	78.72	state board	...		
438					
3997	70.60	cbse	...		
-1					
	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg	
conscientiousness	\				
0	-1	-1	-1	-1	-1
0.9737					
1	-1	-1	-1	-1	-1
-0.7335					
2	-1	-1	-1	-1	-1
0.2718					
3	-1	-1	-1	-1	-1
0.0464					
4	-1	-1	-1	-1	-1
-0.8810					
...	...	...	...	...	...
3993	-1	-1	-1	-1	-1
-0.1082					
3994	-1	-1	-1	-1	-1
-0.3027					

```

3995          -1          -1          -1          -1
-1.5765
3996          -1          -1          -1          -1
-0.1590
3997          -1          -1          -1          -1
-1.1128

      agreeableness  extraversion  nueroticism  openness_to_experience
0            0.8128       0.5269     1.35490      -0.4455
1            0.3789       1.2396    -0.10760      0.8637
2            1.7109       0.1637    -0.86820      0.6721
3            0.3448      -0.3440    -0.40780     -0.9194
4           -0.2793      -1.0697     0.09163     -0.1295
...
3993          0.3448       0.2366     0.64980     -0.9194
3994          0.8784       0.9322     0.77980     -0.0943
3995         -1.5273      -1.5051    -1.31840     -0.7615
3996          0.0459      -0.4511    -0.36120     -0.0943
3997         -0.2793      -0.6343     1.32553     -0.6035

```

[3998 rows x 38 columns]

`df.columns #to see the column names`

```

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'],
      dtype='object')

```

*#now i like to convert the  
ID,12graduation,CollegeID,CollegeTier,CollegeCityTier,GraduationYear  
to categorical or object*

```

df['ID']=df['ID'].astype('object')
df['12graduation']=df['12graduation'].astype('category')
df['CollegeID']=df['CollegeID'].astype('category')
df['CollegeTier']=df['CollegeTier'].astype('category')
df['CollegeCityID']=df['CollegeCityID'].astype('category')

```

```

df['CollegeCityTier']=df['CollegeCityTier'].astype('category')
df['GraduationYear']=df['GraduationYear'].astype('category')

#converting the DOB,DOJ,DOL columns to date format or extracting only
date and making them as object type
#first removing the time portion from the columns before
transformation

df['DOB'] = df['DOB'].str.split(' ').str[0]
df['DOJ'] = df['DOJ'].str.split(' ').str[0]
df['DOL']=df['DOL'].str.split(' ').str[0]

df['DOB']=pd.to_datetime(df['DOB'],format='%m/%d/%y').dt.date
df['DOJ']=pd.to_datetime(df['DOJ'],format='%m/%d/%y').dt.date

#to convert DOL column, since it is neither completely the date format
nor string format
df['DOL'] = np.where(
    df['DOL'].str.strip().str.lower() == 'present', # Check if value
    is "Present"
    pd.NaT, # If yes, set to NaT
    pd.to_datetime(df['DOL'], format='%m/%d/%y',
    errors='coerce').dt.date # Else, convert to date
)

# Convert 'DOB' to datetime format if it's not already
df['DOB'] = pd.to_datetime(df['DOB'], errors='coerce')

# Current date
current_date = pd.to_datetime(datetime.now())

# Calculate the age (difference in years)
df['Age'] = (current_date - df['DOB']).dt.days // 365

# Display the first few rows to verify the new 'Age' column
print(df[['DOB', 'Age']].head())

```

	DOB	Age
0	1990-02-19	34
1	1989-10-04	35
2	1992-08-03	32
3	1989-12-05	34
4	1991-02-27	33

```

#data description
df.shape
(3998, 39)

```

```
df.dtypes
```

```
ID          object
Salary      float64
DOJ         object
DOL         object
Designation object
JobCity     object
Gender      object
DOB         datetime64[ns]
10percentage float64
10board     object
12graduation category
12percentage float64
12board     object
CollegeID   category
CollegeTier category
Degree      object
Specialization object
collegeGPA   float64
CollegeCityID category
CollegeCityTier category
CollegeState object
GraduationYear category
English     int64
Logical     int64
Quant       int64
Domain      float64
ComputerProgramming int64
ElectronicsAndSemicon int64
ComputerScience   int64
MechanicalEngg    int64
ElectricalEngg    int64
TelecomEngg      int64
CivilEngg       int64
conscientiousness float64
agreeableness   float64
extraversion     float64
nueroticism     float64
openess_to_experience float64
Age           int64
dtype: object
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
 #   Column            Non-Null Count  Dtype  
 ---  -- 
 0   ID                3998 non-null    object 
 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    datetime64[ns]
 8   10percentage       3998 non-null    float64
 9   10board             3998 non-null    object 
 10  12graduation        3998 non-null    category
 11  12percentage       3998 non-null    float64
 12  12board             3998 non-null    object 
 13  CollegeID          3998 non-null    category
 14  CollegeTier         3998 non-null    category
 15  Degree              3998 non-null    object 
 16  Specialization      3998 non-null    object 
 17  collegeGPA          3998 non-null    float64
 18  CollegeCityID       3998 non-null    category
 19  CollegeCityTier     3998 non-null    category
 20  CollegeState         3998 non-null    object 
 21  GraduationYear       3998 non-null    category
 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-null    int64  
 28  ComputerScience      3998 non-null    int64  
 29  MechanicalEngg       3998 non-null    int64  
 30  ElectricalEngg       3998 non-null    int64  
 31  TelecomEngg          3998 non-null    int64  
 32  CivilEngg            3998 non-null    int64  
 33  conscientiousness     3998 non-null    float64
 34  agreeableness        3998 non-null    float64
 35  extraversion          3998 non-null    float64
 36  nueroticism          3998 non-null    float64
 37  openess_to_experience 3998 non-null    float64
 38  Age                  3998 non-null    int64  
 39  dtype: object
```

```

0   ID                  3998 non-null  object
1   Salary              3998 non-null  float64
2   DOJ                 3998 non-null  object
3   DOL                 2123 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  datetime64[ns]
8   10percentage        3998 non-null  float64
9   10board              3998 non-null  object
10  12graduation        3998 non-null  category
11  12percentage        3998 non-null  float64
12  12board              3998 non-null  object
13  CollegeID           3998 non-null  category
14  CollegeTier          3998 non-null  category
15  Degree               3998 non-null  object
16  Specialization       3998 non-null  object
17  collegeGPA           3998 non-null  float64
18  CollegeCityID        3998 non-null  category
19  CollegeCityTier       3998 non-null  category
20  CollegeState          3998 non-null  object
21  GraduationYear        3998 non-null  category
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-null  int64
28  ComputerScience        3998 non-null  int64
29  MechanicalEngg        3998 non-null  int64
30  ElectricalEngg        3998 non-null  int64
31  TelecomEngg           3998 non-null  int64
32  CivilEngg              3998 non-null  int64
33  conscientiousness      3998 non-null  float64
34  agreeableness          3998 non-null  float64
35  extraversion            3998 non-null  float64
36  nueroticism             3998 non-null  float64
37  openness_to_experience 3998 non-null  float64
38  Age                   3998 non-null  int64
dtypes: category(6), datetime64[ns](1), float64(10), int64(11),
object(11)
memory usage: 1.1+ MB

```

df.describe()

	Salary	DOB	10percentage	\
count	3.998000e+03	3998	3998.000000	
mean	3.076998e+05	1990-12-06 06:01:15.637819008	77.925443	
min	3.500000e+04	1977-10-30 00:00:00	43.000000	
25%	1.800000e+05	1989-11-16 06:00:00	71.680000	

50%	3.000000e+05		1991-03-07	12:00:00	79.150000
75%	3.700000e+05		1992-03-13	18:00:00	85.670000
max	4.000000e+06		1997-05-27	00:00:00	97.760000
std	2.127375e+05			NaN	9.850162
Quant	12percentage	collegeGPA	English	Logical	
count	3998.000000	3998.000000	3998.000000	3998.000000	
3998.000000					
mean	74.466366	71.486171	501.649075	501.598799	
513.378189					
min	40.000000	6.450000	180.000000	195.000000	
120.000000					
25%	66.000000	66.407500	425.000000	445.000000	
430.000000					
50%	74.400000	71.720000	500.000000	505.000000	
515.000000					
75%	82.600000	76.327500	570.000000	565.000000	
595.000000					
max	98.700000	99.930000	875.000000	795.000000	
900.000000					
std	10.999933	8.167338	104.940021	86.783297	
122.302332					
ElectricalEngg	Domain	ComputerProgramming	...	MechanicalEngg	
count	3998.000000	3998.000000	...	3998.000000	
3998.000000					
mean	0.510490	353.102801	...	22.974737	
16.478739					
min	-1.000000	-1.000000	...	-1.000000	-
1.000000					
25%	0.342315	295.000000	...	-1.000000	-
1.000000					
50%	0.622643	415.000000	...	-1.000000	-
1.000000					
75%	0.842248	495.000000	...	-1.000000	-
1.000000					
max	0.999910	840.000000	...	623.000000	
676.000000					
std	0.468671	205.355519	...	98.123311	
87.585634					
	TelecomEngg	CivilEngg	conscientiousness	agreeableness	\
count	3998.000000	3998.000000	3998.000000	3998.000000	
mean	31.851176	2.683842	-0.037831	0.146496	
min	-1.000000	-1.000000	-4.126700	-5.781600	
25%	-1.000000	-1.000000	-0.713525	-0.287100	
50%	-1.000000	-1.000000	0.046400	0.212400	
75%	-1.000000	-1.000000	0.702700	0.812800	

max	548.000000	516.000000	1.995300	1.904800
std	104.852845	36.658505	1.028666	0.941782
count	3998.000000	3998.000000	3998.000000	3998.000000
mean	0.002763	-0.169033	-0.138110	33.374437
min	-4.600900	-2.643000	-7.375700	27.000000
25%	-0.604800	-0.868200	-0.669200	32.000000
50%	0.091400	-0.234400	-0.094300	33.000000
75%	0.672000	0.526200	0.502400	34.000000
max	2.535400	3.352500	1.822400	46.000000
std	0.951471	1.007580	1.008075	1.769877

[8 rows x 22 columns]

df.head()

	ID	Salary	DOJ	DOL	Designation
0	203097	420000.0	2012-06-01	NaT	senior quality engineer
1	579905	500000.0	2013-09-01	NaT	assistant manager
2	810601	325000.0	2014-06-01	NaT	systems engineer
3	267447	1100000.0	2011-07-01	NaT	senior software engineer
4	343523	200000.0	2014-03-01	2015-03-01	get

	JobCity	Gender	DOB	10percentage	board
0	Bangalore	f	1990-02-19	84.3	ofsecondary
1	Indore	m	1989-10-04	85.4	cbse
2	Chennai	f	1992-08-03	85.0	cbse
3	Gurgaon	m	1989-12-05	85.6	cbse
4	Manesar	m	1991-02-27	78.0	cbse

	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg
0	... conscientiousness	-1	-1	-1
1	...	-1	-1	-1
2	...	-1	-1	-1

```

3 ... -1 -1 -1 -1
0.0464
4 ... -1 -1 -1 -1
-0.8810

agreeableness extraversion nueroticism openness_to_experience Age
0 0.8128 0.5269 1.35490 -0.4455 34
1 0.3789 1.2396 -0.10760 0.8637 35
2 1.7109 0.1637 -0.86820 0.6721 32
3 0.3448 -0.3440 -0.40780 -0.9194 34
4 -0.2793 -1.0697 0.09163 -0.1295 33

```

[5 rows x 39 columns]

### #univariate analysis

```

numerical_cols=df.select_dtypes(include=[np.number]).columns
numerical_cols

Index(['Salary', '10percentage', '12percentage', 'collegeGPA',
'English',
       'Logical', 'Quant', 'Domain', 'ComputerProgramming',
       'ElectronicsAndSemicon', 'ComputerScience', 'MechanicalEngg',
       'ElectricalEngg', 'TelecomEngg', 'CivilEngg',
'conscientiousness',
       'agreeableness', 'extraversion', 'nueroticism',
'openness_to_experience',
       'Age'],
      dtype='object')

categorical_cols=df.select_dtypes(include=['object','category']).columns
categorical_cols

Index(['ID', 'DOJ', 'DOL', 'Designation', 'JobCity', 'Gender',
'10board',
       '12graduation', '12board', 'CollegeID', 'CollegeTier',
'Degree',
       'Specialization', 'CollegeCityID', 'CollegeCityTier',
'CollegeState',
       'GraduationYear'],
      dtype='object')

#for numerical analysis

for col in numerical_cols:
    print()
    print(f"Analysis for {col} column:")

```

```

min_val=df[col].min()
max_val=df[col].max()
mean=df[col].mean()
median=df[col].median()
mode=df[col].mode()[0]
skewness=skew(df[col].dropna())
kurtosis_val=kurtosis(df[col].dropna())
range_val=df[col].max() - df[col].min()
iqr=np.percentile(df[col].dropna(),75) -
np.percentile(df[col].dropna(),25)
std=df[col].std()

print(f"Min: {min_val}")
print(f"Max: {max_val}")
print(f"Mean: {mean}")
print(f"Median: {median}")
print(f"Mode: {mode}")
print(f"Skew: {skewness}")
print(f"Kurtosis: {kurtosis_val}")
print(f"Range: {range_val}")
print(f"IQR: {iqr}")
print(f"STD: {std}")

```

Analysis for Salary column:

```

Min: 35000.0
Max: 4000000.0
Mean: 307699.8499249625
Median: 300000.0
Mode: 300000.0
Skew: 6.44866054831297
Kurtosis: 80.82731823061802
Range: 3965000.0
IQR: 190000.0
STD: 212737.4999568574

```

Analysis for 10percentage column:

```

Min: 43.0
Max: 97.76
Mean: 77.9254427213607
Median: 79.15
Mode: 78.0
Skew: -0.5907967422390255
Kurtosis: -0.11164680905297963
Range: 54.760000000000005
IQR: 13.989999999999995
STD: 9.850162410652587

```

Analysis for 12percentage column:

```

Min: 40.0

```

```
Max: 98.7
Mean: 74.46636568284141
Median: 74.4
Mode: 70.0
Skew: -0.032595179202258136
Kurtosis: -0.6314493016154126
Range: 58.7
IQR: 16.59999999999994
STD: 10.999933102957728
```

Analysis for collegeGPA column:

```
Min: 6.45
Max: 99.93
Mean: 71.48617058529265
Median: 71.72
Mode: 70.0
Skew: -1.2487404274031384
Kurtosis: 10.21994936069596
Range: 93.48
IQR: 9.920000000000002
STD: 8.167337651961981
```

Analysis for English column:

```
Min: 180
Max: 875
Mean: 501.64907453726863
Median: 500.0
Mode: 475
Skew: 0.1919249750111508
Kurtosis: -0.2553151870949164
Range: 695
IQR: 145.0
STD: 104.94002081603864
```

Analysis for Logical column:

```
Min: 195
Max: 795
Mean: 501.59879939969983
Median: 505.0
Mode: 495
Skew: -0.21652053612984407
Kurtosis: -0.22597989963617682
Range: 600
IQR: 120.0
STD: 86.78329691840084
```

Analysis for Quant column:

```
Min: 120
Max: 900
Mean: 513.3781890945472
```

```
Median: 515.0
Mode: 605
Skew: -0.019391755556999294
Kurtosis: -0.10384434185307434
Range: 780
IQR: 165.0
STD: 122.30233211125697
```

```
Analysis for Domain column:
Min: -1.0
Max: 0.9999104076
Mean: 0.5104896530031996
Median: 0.6226429158
Mode: -1.0
Skew: -1.9214243230710002
Kurtosis: 3.889579528417558
Range: 1.9999104076
IQR: 0.4999334224
STD: 0.46867133859379967
```

```
Analysis for ComputerProgramming column:
Min: -1
Max: 840
Mean: 353.10280140070034
Median: 415.0
Mode: -1
Skew: -0.7778136825484333
Kurtosis: -0.6670191448718974
Range: 841
IQR: 200.0
STD: 205.3555194149209
```

```
Analysis for ElectronicsAndSemicon column:
Min: -1
Max: 612
Mean: 95.32841420710355
Median: -1.0
Mode: -1
Skew: 1.1955261109357744
Kurtosis: -0.21161173593544147
Range: 613
IQR: 234.0
STD: 158.24121827263744
```

```
Analysis for ComputerScience column:
Min: -1
Max: 715
Mean: 90.7423711855928
Median: -1.0
Mode: -1
```

```
Skew: 1.5289469492574301
Kurtosis: 0.6902745985889482
Range: 716
IQR: 0.0
STD: 175.2730830755835
```

```
Analysis for MechanicalEngg column:
Min: -1
Max: 623
Mean: 22.97473736868434
Median: -1.0
Mode: -1
Skew: 4.0280514405380154
Kurtosis: 14.998679886132297
Range: 624
IQR: 0.0
STD: 98.1233114520027
```

```
Analysis for ElectricalEngg column:
Min: -1
Max: 676
Mean: 16.478739369684842
Median: -1.0
Mode: -1
Skew: 5.058508440756092
Kurtosis: 24.845590955541997
Range: 677
IQR: 0.0
STD: 87.58563441414714
```

```
Analysis for TelecomEngg column:
Min: -1
Max: 548
Mean: 31.851175587793897
Median: -1.0
Mode: -1
Skew: 3.040119450810971
Kurtosis: 7.798956685776439
Range: 549
IQR: 0.0
STD: 104.85284546338978
```

```
Analysis for CivilEngg column:
Min: -1
Max: 516
Mean: 2.68384192096048
Median: -1.0
Mode: -1
Skew: 10.311810510449162
Kurtosis: 108.90352629121763
```

```
Range: 517
IQR: 0.0
STD: 36.65850485945898
```

```
Analysis for conscientiousness column:
Min: -4.1267
Max: 1.9953
Mean: -0.03783129064532267
Median: 0.0464
Mode: 0.2718
Skew: -0.5268055945864971
Kurtosis: 0.12094197654529593
Range: 6.122
IQR: 1.416225
STD: 1.0286656043404265
```

```
Analysis for agreeableness column:
Min: -5.7816
Max: 1.9048
Mean: 0.14649582291145571
Median: 0.2124
Mode: 0.3789
Skew: -1.2044631329796518
Kurtosis: 3.385502238121073
Range: 7.6864
IQR: 1.0998999999999999
STD: 0.9417821625653949
```

```
Analysis for extraversion column:
Min: -4.6009
Max: 2.5354
Mean: 0.0027629814907453795
Median: 0.0914
Mode: 0.4711
Skew: -0.5230704373682461
Kurtosis: 0.6416632674320177
Range: 7.1363
IQR: 1.2768000000000002
STD: 0.951471442921609
```

```
Analysis for nueroticism column:
Min: -2.643
Max: 3.3525
Mean: -0.16903257128564284
Median: -0.2344
Mode: -0.4879
Skew: 0.16564750621703866
Kurtosis: -0.19279971750140756
Range: 5.9955
IQR: 1.3944
```

```
STD: 1.0075796250873494

Analysis for openness_to_experience column:
Min: -7.3757
Max: 1.8224
Mean: -0.13811023011505755
Median: -0.0943
Mode: 0.6721
Skew: -1.5063965613424706
Kurtosis: 5.779590372046927
Range: 9.1981
IQR: 1.1716
STD: 1.008074562220052
```

```
Analysis for Age column:
Min: 27
Max: 46
Mean: 33.3744372186093
Median: 33.0
Mode: 33
Skew: 0.8574624668722381
Kurtosis: 1.641010365812554
Range: 19
IQR: 2.0
STD: 1.769877459204665
```

```
#univariate visual analysis for numerical columns
```

```
for col in numerical_cols:
    plt.figure(figsize=(10,3))

    #Histogram with KDE plot
    plt.subplot(1,2,1)
    sns.histplot(df[col],kde=True,bins=30,color='skyblue')
    plt.title(f"Histogram and KDE for {col}")

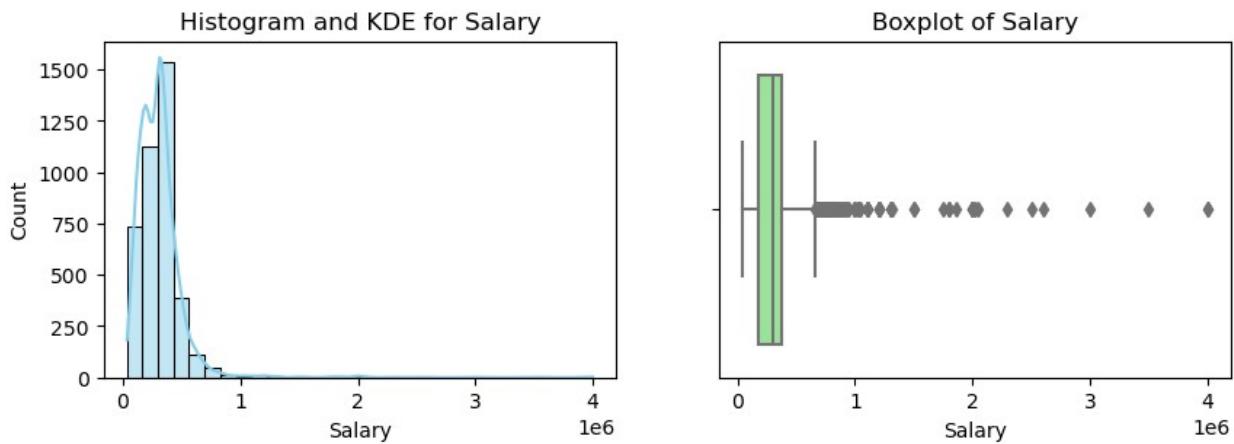
    #Box plot
    plt.subplot(1,2,2)
    sns.boxplot(x=df[col],color='lightgreen')
    plt.title(f"Boxplot of {col}")

plt.show()
```

```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
```

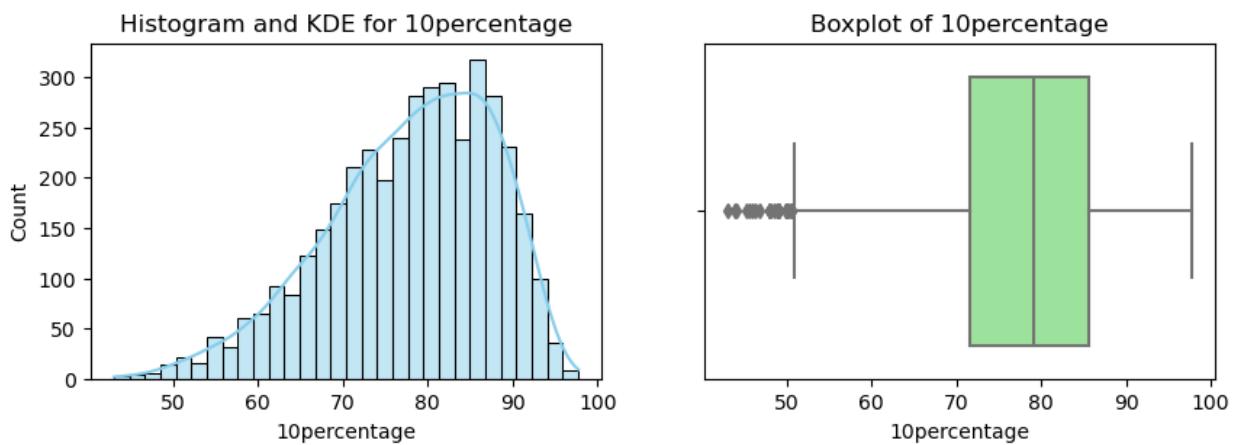
```
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



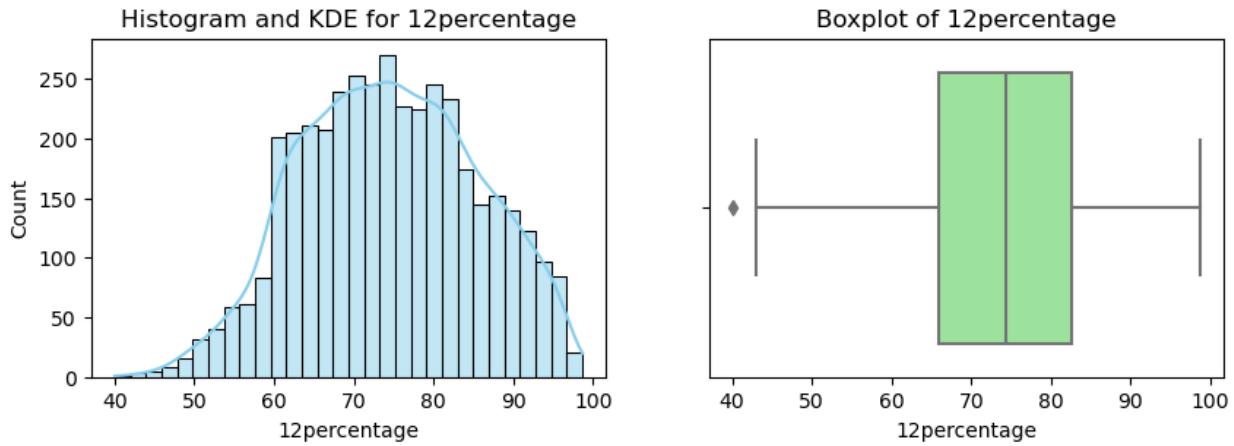
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



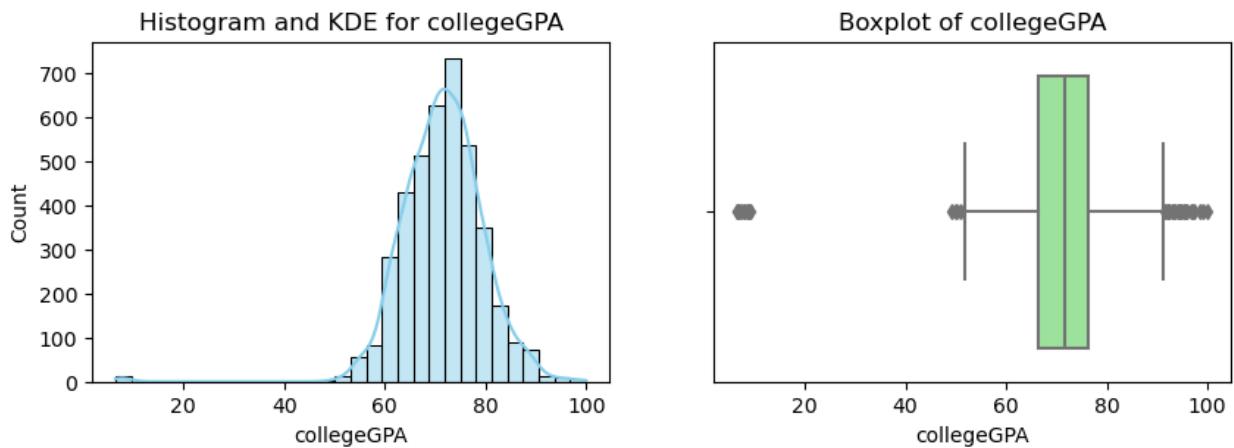
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



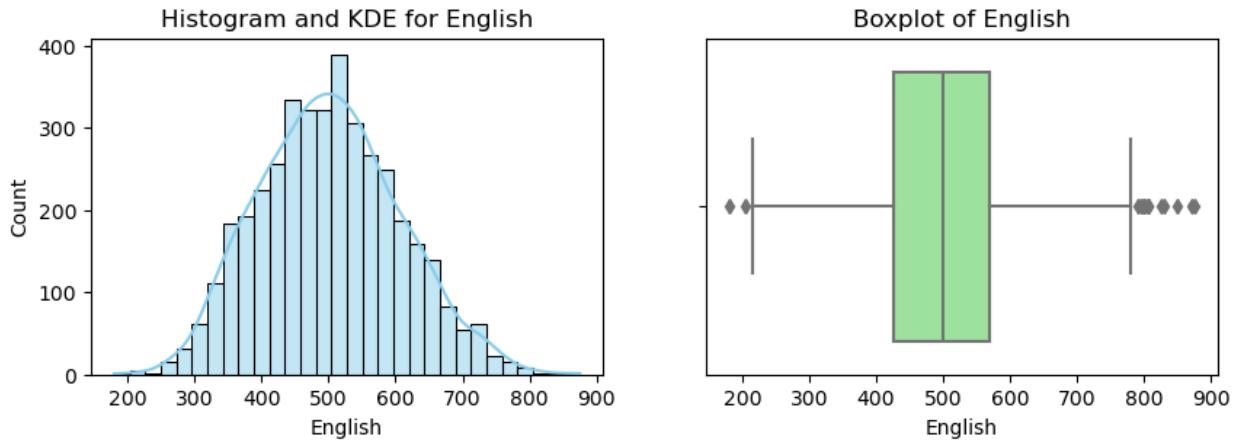
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



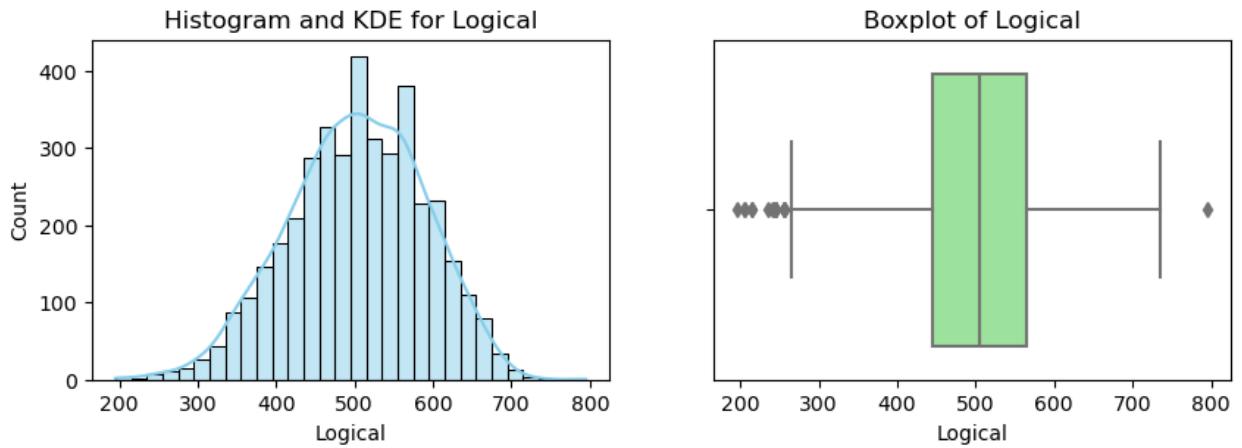
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



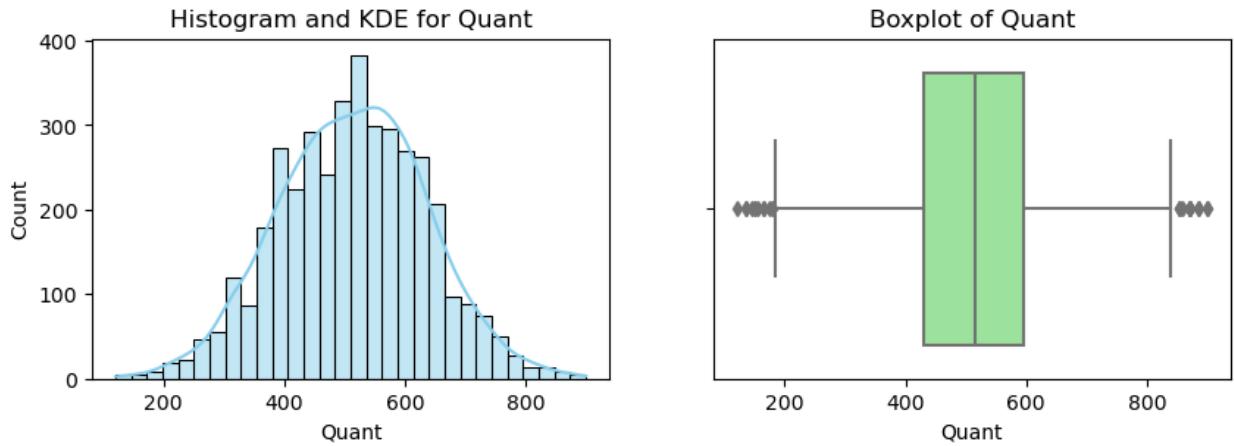
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



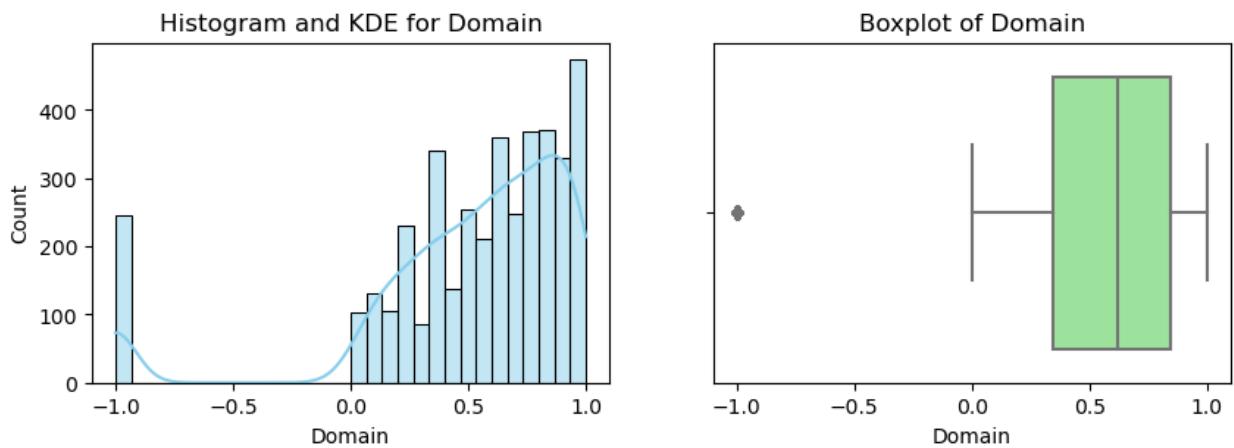
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



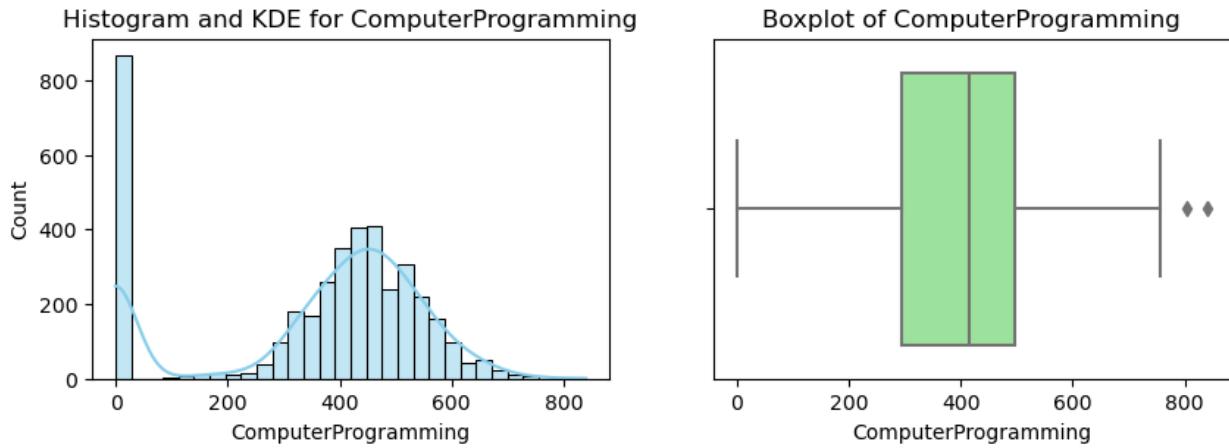
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



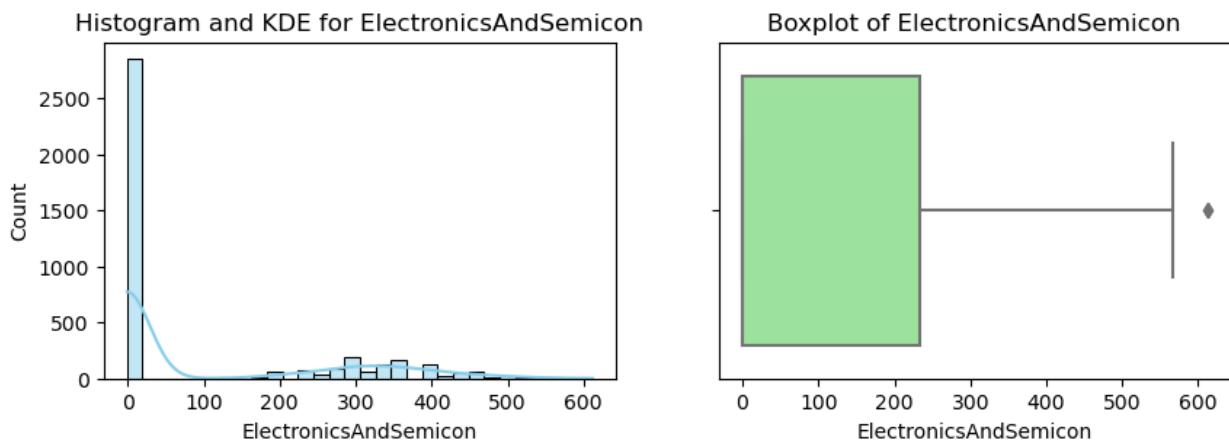
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



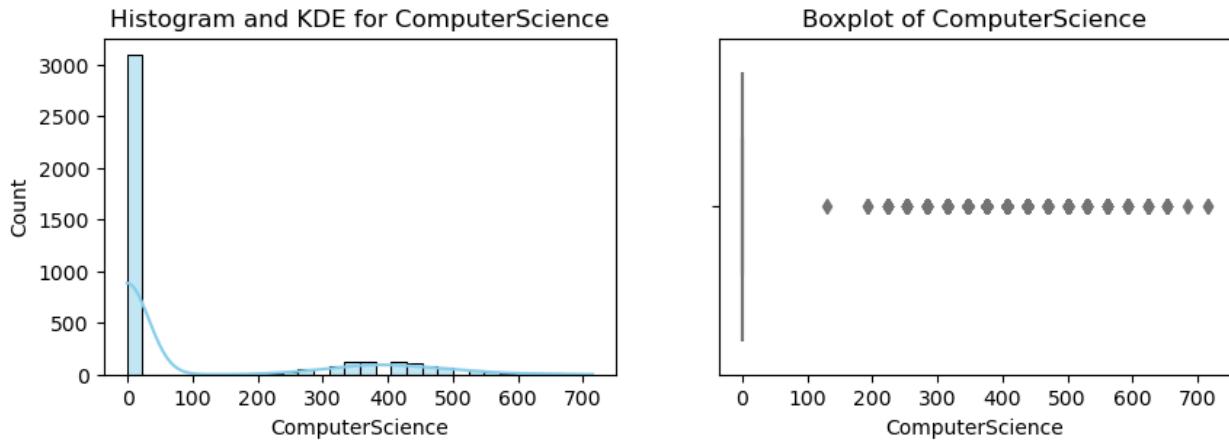
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



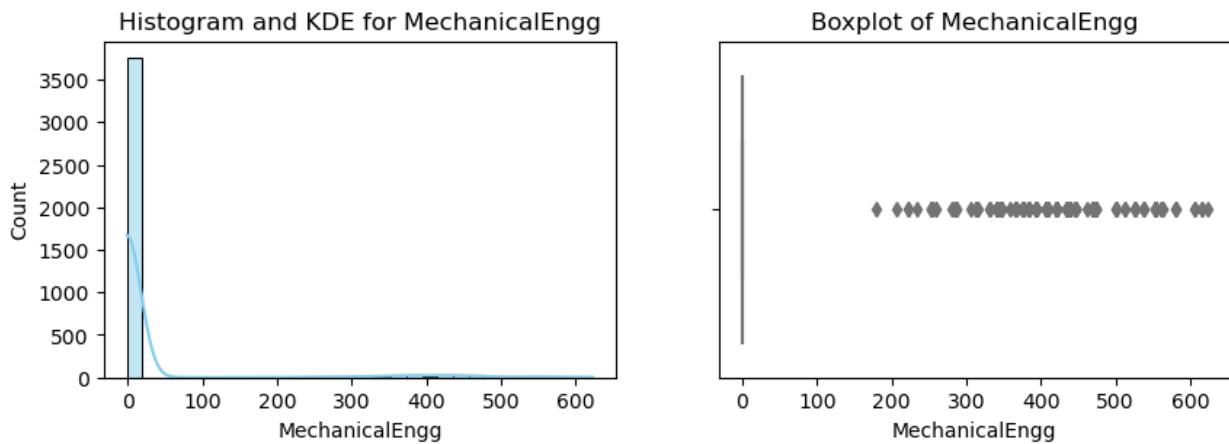
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



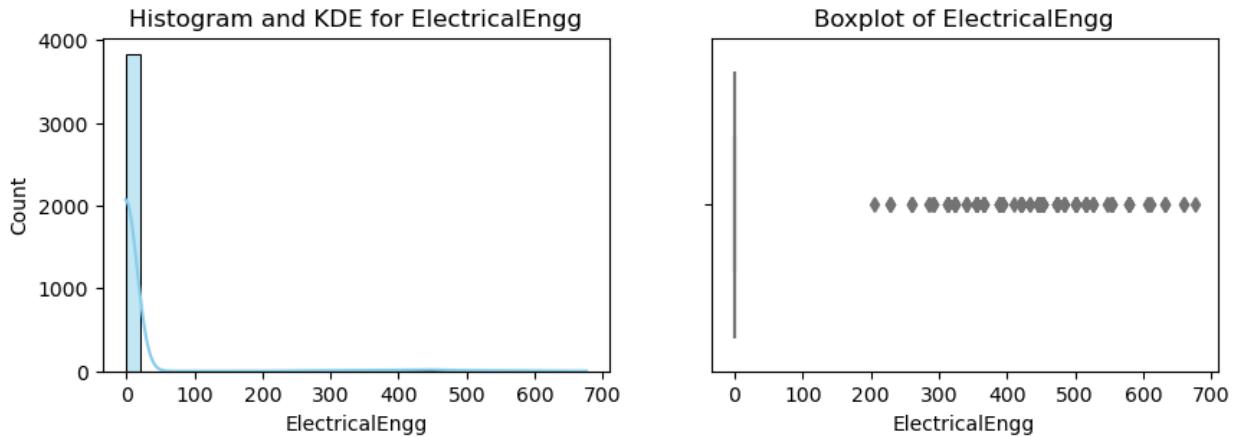
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



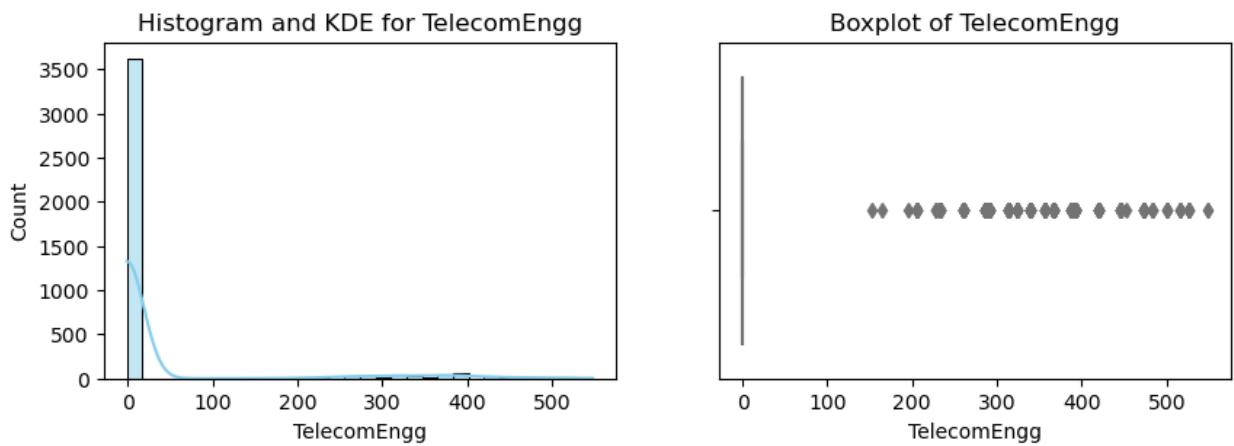
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



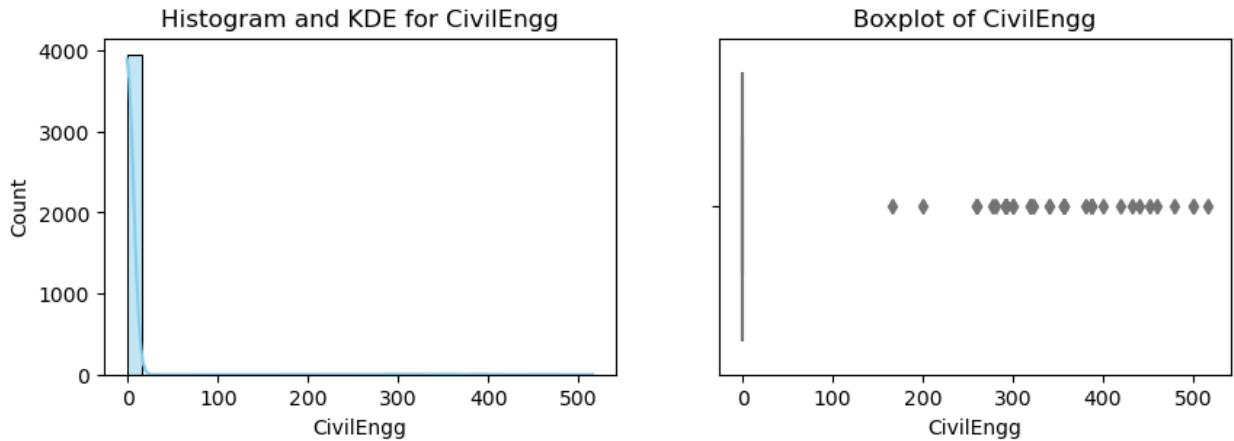
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



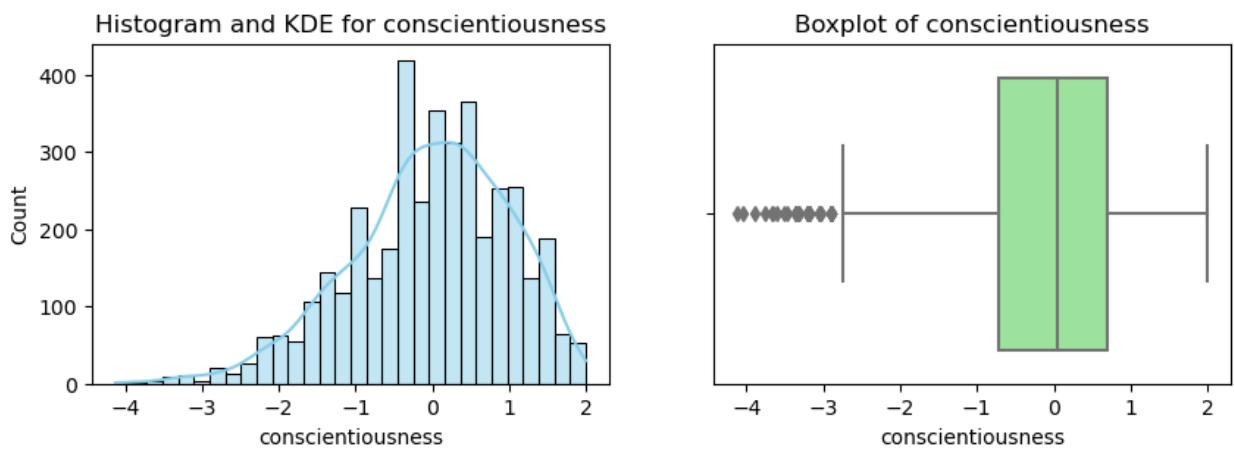
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



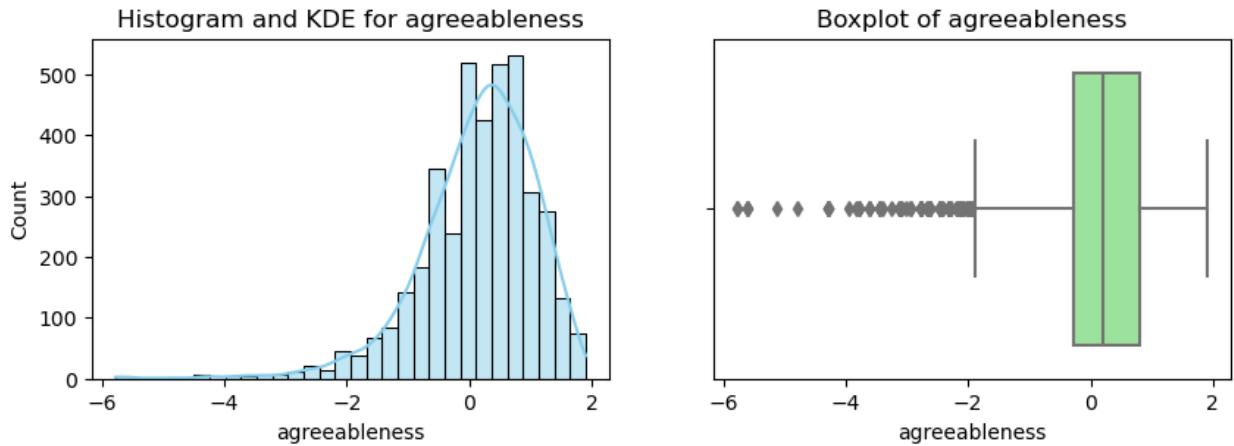
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



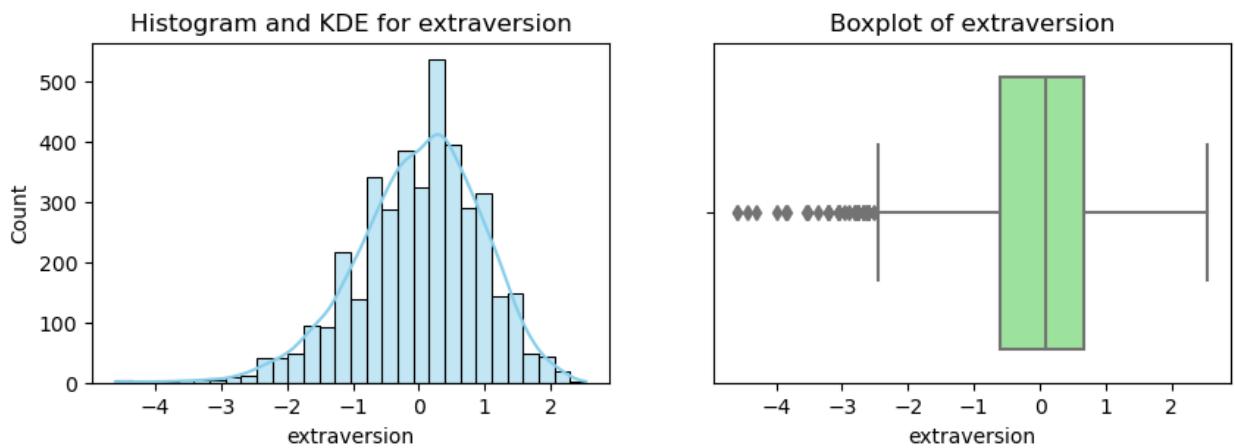
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



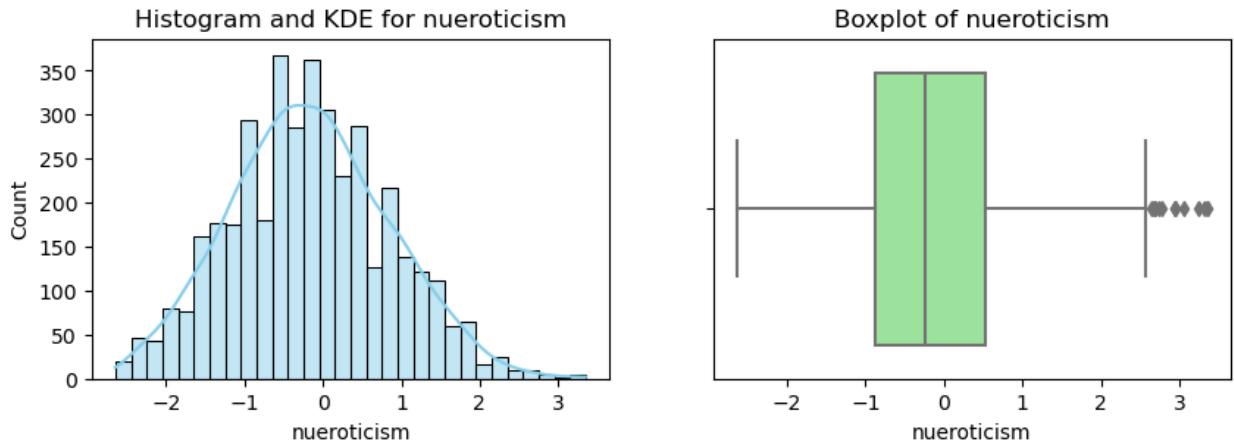
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



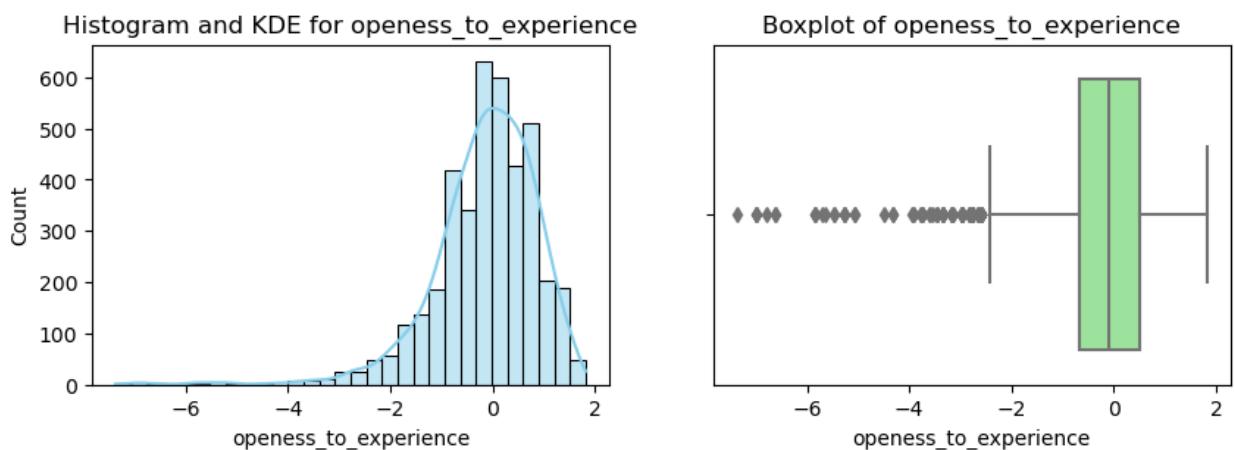
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



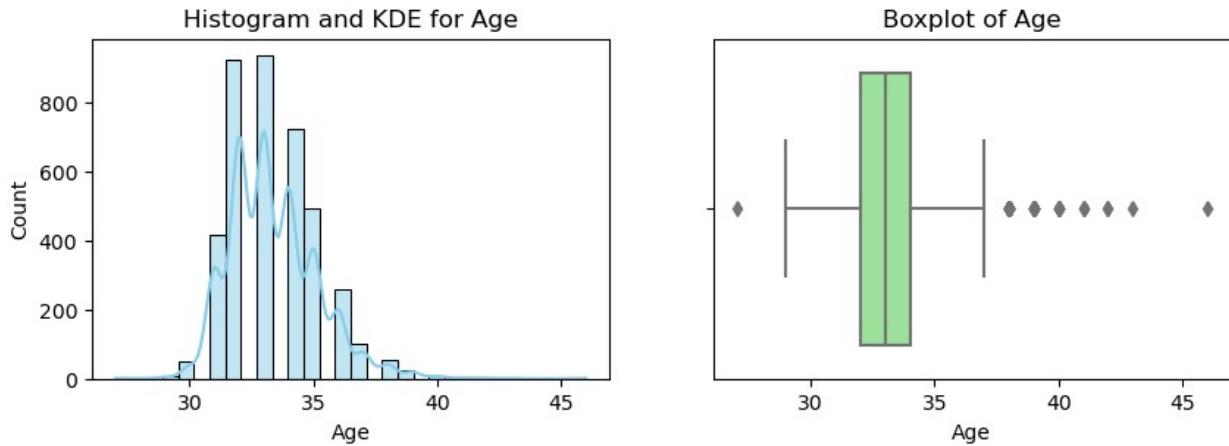
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```



```
#for categorical analysis

#categorical_cols=df.select_dtypes(include=['object','category']).columns

for col in categorical_cols:
    print()
    print(f"Analysis for {col} column:")

    unique_val=df[col].unique()
    nunique_val=df[col].nunique()
    value_counts=df[col].value_counts()

    print(f"Unique values: {unique_val}")
    print(f"No of unique values: {nunique_val}")
    print(f"value counts: \n{value_counts}")
    print()

Analysis for ID column:
Unique values: [203097 579905 810601 ... 355888 947111 324966]
No of unique values: 3998
value counts:
ID
203097      1
283847      1
1118955     1
327371      1
587792      1
...
355018      1
806160      1
102065      1
462964      1
```

```
324966      1  
Name: count, Length: 3998, dtype: int64
```

Analysis for DOJ column:

```
Unique values: [datetime.date(2012, 6, 1) datetime.date(2013, 9, 1)  
datetime.date(2014, 6, 1) datetime.date(2011, 7, 1)  
datetime.date(2014, 3, 1) datetime.date(2014, 8, 1)  
datetime.date(2014, 7, 1) datetime.date(2013, 7, 1)  
datetime.date(2011, 4, 1) datetime.date(2011, 8, 1)  
datetime.date(2013, 12, 1) datetime.date(2014, 1, 1)  
datetime.date(2013, 8, 1) datetime.date(2014, 9, 1)  
datetime.date(2010, 11, 1) datetime.date(2012, 8, 1)  
datetime.date(2013, 10, 1) datetime.date(2012, 9, 1)  
datetime.date(2011, 1, 1) datetime.date(2015, 2, 1)  
datetime.date(2014, 11, 1) datetime.date(2011, 12, 1)  
datetime.date(2014, 10, 1) datetime.date(2015, 1, 1)  
datetime.date(2013, 3, 1) datetime.date(2010, 10, 1)  
datetime.date(2013, 1, 1) datetime.date(2011, 6, 1)  
datetime.date(2014, 4, 1) datetime.date(2012, 5, 1)  
datetime.date(2012, 10, 1) datetime.date(2015, 4, 1)  
datetime.date(2012, 3, 1) datetime.date(2013, 6, 1)  
datetime.date(2009, 9, 1) datetime.date(2013, 11, 1)  
datetime.date(2010, 7, 1) datetime.date(2014, 2, 1)  
datetime.date(2015, 6, 1) datetime.date(2014, 5, 1)  
datetime.date(2014, 12, 1) datetime.date(2011, 11, 1)  
datetime.date(2015, 7, 1) datetime.date(2013, 5, 1)  
datetime.date(2011, 3, 1) datetime.date(2015, 3, 1)  
datetime.date(2012, 7, 1) datetime.date(2011, 10, 1)  
datetime.date(2010, 4, 1) datetime.date(2013, 4, 1)  
datetime.date(2010, 12, 1) datetime.date(2013, 2, 1)  
datetime.date(2011, 9, 1) datetime.date(2012, 2, 1)  
datetime.date(2012, 1, 1) datetime.date(2012, 12, 1)  
datetime.date(2010, 9, 1) datetime.date(2012, 4, 1)  
datetime.date(2012, 11, 1) datetime.date(2015, 5, 1)  
datetime.date(2010, 6, 1) datetime.date(2011, 2, 1)  
datetime.date(2010, 8, 1) datetime.date(2010, 5, 1)  
datetime.date(2011, 5, 1) datetime.date(2004, 8, 1)  
datetime.date(2008, 11, 1) datetime.date(2009, 6, 1)  
datetime.date(2010, 2, 1) datetime.date(2009, 11, 1)  
datetime.date(2010, 3, 1) datetime.date(2015, 11, 1)  
datetime.date(2006, 1, 1) datetime.date(2015, 8, 1)  
datetime.date(2010, 1, 1) datetime.date(2015, 12, 1)  
datetime.date(2007, 9, 1) datetime.date(1991, 6, 1)  
datetime.date(2007, 7, 1) datetime.date(2007, 6, 1)  
datetime.date(2007, 2, 1)]
```

No of unique values: 81

value counts:

DOJ

```
2014-07-01    199
2014-06-01    180
2014-08-01    178
2014-09-01    142
2014-01-01    142
...
2015-11-01     1
2009-11-01     1
2004-08-01     1
2009-09-01     1
2007-02-01     1
Name: count, Length: 81, dtype: int64
```

Analysis for DOL column:

```
Unique values: [NaT datetime.date(2015, 3, 1) datetime.date(2015, 5, 1)
datetime.date(2015, 7, 1) datetime.date(2015, 4, 1)
datetime.date(2014, 10, 1) datetime.date(2014, 9, 1)
datetime.date(2014, 6, 1) datetime.date(2012, 9, 1)
datetime.date(2013, 12, 1) datetime.date(2015, 6, 1)
datetime.date(2013, 10, 1) datetime.date(2015, 1, 1)
datetime.date(2014, 4, 1) datetime.date(2013, 6, 1)
datetime.date(2012, 3, 1) datetime.date(2014, 7, 1)
datetime.date(2013, 2, 1) datetime.date(2014, 1, 1)
datetime.date(2013, 4, 1) datetime.date(2012, 7, 1)
datetime.date(2014, 5, 1) datetime.date(2013, 9, 1)
datetime.date(2015, 2, 1) datetime.date(2012, 1, 1)
datetime.date(2015, 8, 1) datetime.date(2014, 8, 1)
datetime.date(2015, 12, 1) datetime.date(2014, 12, 1)
datetime.date(2012, 5, 1) datetime.date(2011, 3, 1)
datetime.date(2011, 7, 1) datetime.date(2014, 2, 1)
datetime.date(2011, 12, 1) datetime.date(2015, 10, 1)
datetime.date(2014, 11, 1) datetime.date(2014, 3, 1)
datetime.date(2011, 11, 1) datetime.date(2013, 5, 1)
datetime.date(2013, 7, 1) datetime.date(2013, 11, 1)
datetime.date(2011, 1, 1) datetime.date(2011, 5, 1)
datetime.date(2012, 2, 1) datetime.date(2012, 11, 1)
datetime.date(2012, 6, 1) datetime.date(2013, 8, 1)
datetime.date(2005, 3, 1) datetime.date(2013, 3, 1)
datetime.date(2012, 10, 1) datetime.date(2011, 2, 1)
datetime.date(2010, 2, 1) datetime.date(2013, 1, 1)
datetime.date(2011, 6, 1) datetime.date(2015, 9, 1)
datetime.date(2012, 4, 1) datetime.date(2012, 8, 1)
datetime.date(2011, 4, 1) datetime.date(2011, 10, 1)
datetime.date(2015, 11, 1) datetime.date(2012, 12, 1)
datetime.date(2011, 9, 1) datetime.date(2010, 8, 1)
datetime.date(2011, 8, 1) datetime.date(2009, 6, 1)
datetime.date(2008, 3, 1) datetime.date(2010, 10, 1)]
```

```
No of unique values: 66
value counts:
DOL
2015-04-01    573
2015-03-01    124
2015-05-01    112
2015-01-01     99
2014-04-01    82
...
2015-10-01     1
2010-02-01     1
2011-02-01     1
2005-03-01     1
2010-10-01     1
Name: count, Length: 66, dtype: int64
```

```
Analysis for Designation column:
Unique values: ['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']
```

No of unique values: 419

value counts:

Designation	
software engineer	539
software developer	265
system engineer	205
programmer analyst	139
systems engineer	118
...	
cad drafter	1
noc engineer	1
human resources intern	1
senior quality assurance engineer	1
jr. software developer	1
Name: count, Length: 419, dtype: int64	

Analysis for JobCity column:

Unique values: ['Bangalore' 'Indore' 'Chennai' 'Gurgaon' 'Manesar' 'Hyderabad' 'Banglore' 'Noida' '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' 'vsakhaptnam' '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' 'gуроан' 'Dehradun' 'Rudrapur' 'Jamshedpur' 'vizag' 'Nouda' 'Dharamshala' 'Banagalore' 'Hissar' 'Ranchi' 'BANGALORE' 'Madurai' 'Gurga' 'Chandigarh' 'Australia' 'Chennai' 'CHEYYAR' 'Mumbai' 'sonepat' 'Ghaziabad' 'Pantnagar' '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' 'Gуроан' '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' 'BAngular' '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'
'shabibabad' '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']
No of unique values: 339
value counts:
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
```

```
Analysis for Gender column:
Unique values: ['f' 'm']
No of unique values: 2
value counts:
Gender
m    3041
f     957
Name: count, dtype: int64
```

```
Analysis for 10board column:
Unique values: ['board of secondary 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' 'hbsc'  
'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 matrhrsecsschool,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' 'wbbsce'  
'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 bord' '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 academic council'
'hse,board' 'board of ssc education andhra pradesh' 'up-board'
'bse,odisha']
```

No of unique values: 275

value counts:

```
10board
cbse           1395
state board    1164
0              350
icse           281
ssc            122
...
hse,orissa     1
national public school 1
nagpur board   1
jharkhand academic council 1
bse,odisha     1
Name: count, Length: 275, dtype: int64
```

Analysis for 12graduation column:

Unique values: [2007, 2010, 2008, 2009, 2006, ..., 2002, 2001, 1998, 2013, 1999]

Length: 16

Categories (16, int64): [1995, 1998, 1999, 2001, ..., 2010, 2011, 2012, 2013]

No of unique values: 16

value counts:

```
12graduation
2009    1052
2008    935
2010    742
2007    528
2006    407
2005    160
2004    73
2011    46
2003    25
2002    14
2012    10
2001    2
1995    1
```

```
1998      1
1999      1
2013      1
Name: count, dtype: int64
```

```
Analysis for 12board column:
Unique values: ['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 vidayapeeth' '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 matrhrsecschool,neyveli,cuddalore district' 'bte up'  
'scte and vt ,orissa' 'hbsc' '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' 'aissce' '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 bord'
'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 tecnicaleducation ,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']
No of unique values: 340
value counts:
12board
cbse                      1400
state board                 1254
0                          359
icse                       129
up board                    87
...
jawahar higher secondary school      1
nagpur board                1
bsemp                      1

```

```
board of higher secondary orissa          1
boardofintermediate                      1
Name: count, Length: 340, dtype: int64

Analysis for CollegeID column:
Unique values: [1141, 5807, 64, 6920, 11368, ..., 278, 6072, 3572,
6327, 4883]
Length: 1350
Categories (1350, int64): [2, 11, 12, 13, ..., 17935, 18103, 18139,
18409]
No of unique values: 1350
value counts:
CollegeID
272      94
64       38
44       35
11759    35
47       33
...
6041     1
6072     1
11       1
6095     1
18409    1
Name: count, Length: 1350, dtype: int64

Analysis for CollegeTier column:
Unique values: [2, 1]
Categories (2, int64): [1, 2]
No of unique values: 2
value counts:
CollegeTier
2      3701
1      297
Name: count, dtype: int64

Analysis for Degree column:
Unique values: ['B.Tech/B.E.', 'MCA', 'M.Tech./M.E.', 'M.Sc. (Tech.)']
No of unique values: 4
value counts:
Degree
B.Tech/B.E.      3700
MCA              243
M.Tech./M.E.     53
M.Sc. (Tech.)    2
Name: count, dtype: int64
```

```
Analysis for Specialization column:  
Unique values: ['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']
```

No of unique values: 46

value counts:

Specialization

electronics and communication engineering	880
computer science & engineering	744
information technology	660
computer engineering	600
computer application	244
mechanical engineering	201
electronics and electrical engineering	196
electronics & telecommunications	121
electrical engineering	82
electronics & instrumentation eng	32
civil engineering	29
electronics and instrumentation engineering	27
information science engineering	27
instrumentation and control engineering	20
electronics engineering	19
biotechnology	15
other	13

```
industrial & production engineering          10
applied electronics and instrumentation       9
chemical engineering                         9
computer science and technology              6
telecommunication engineering                6
mechanical and automation                   5
automobile/automotive engineering            5
instrumentation engineering                  4
mechatronics                                4
aeronautical engineering                     3
electronics and computer engineering         3
electrical and power engineering             2
biomedical engineering                       2
information & communication technology      2
industrial engineering                      2
computer science                            2
metallurgical engineering                   2
power systems and automation                 1
control and instrumentation engineering      1
mechanical & production engineering          1
embedded systems technology                  1
polymer technology                           1
computer and communication engineering       1
information science                          1
internal combustion engine                   1
computer networking                         1
ceramic engineering                         1
electronics                                 1
industrial & management engineering         1
Name: count, dtype: int64
```

```
Analysis for CollegeCityID column:
Unique values: [1141, 5807, 64, 6920, 11368, ..., 278, 6072, 3572,
6327, 4883]
Length: 1350
Categories (1350, int64): [2, 11, 12, 13, ..., 17935, 18103, 18139,
18409]
No of unique values: 1350
value counts:
CollegeCityID
272      94
64       38
44       35
11759    35
47       33
...
6041     1
6072     1
```

```
11      1  
6095    1  
18409   1  
Name: count, Length: 1350, dtype: int64
```

```
Analysis for CollegeCityTier column:  
Unique values: [0, 1]  
Categories (2, int64): [0, 1]  
No of unique values: 2  
value counts:  
CollegeCityTier  
0    2797  
1    1201  
Name: count, dtype: int64
```

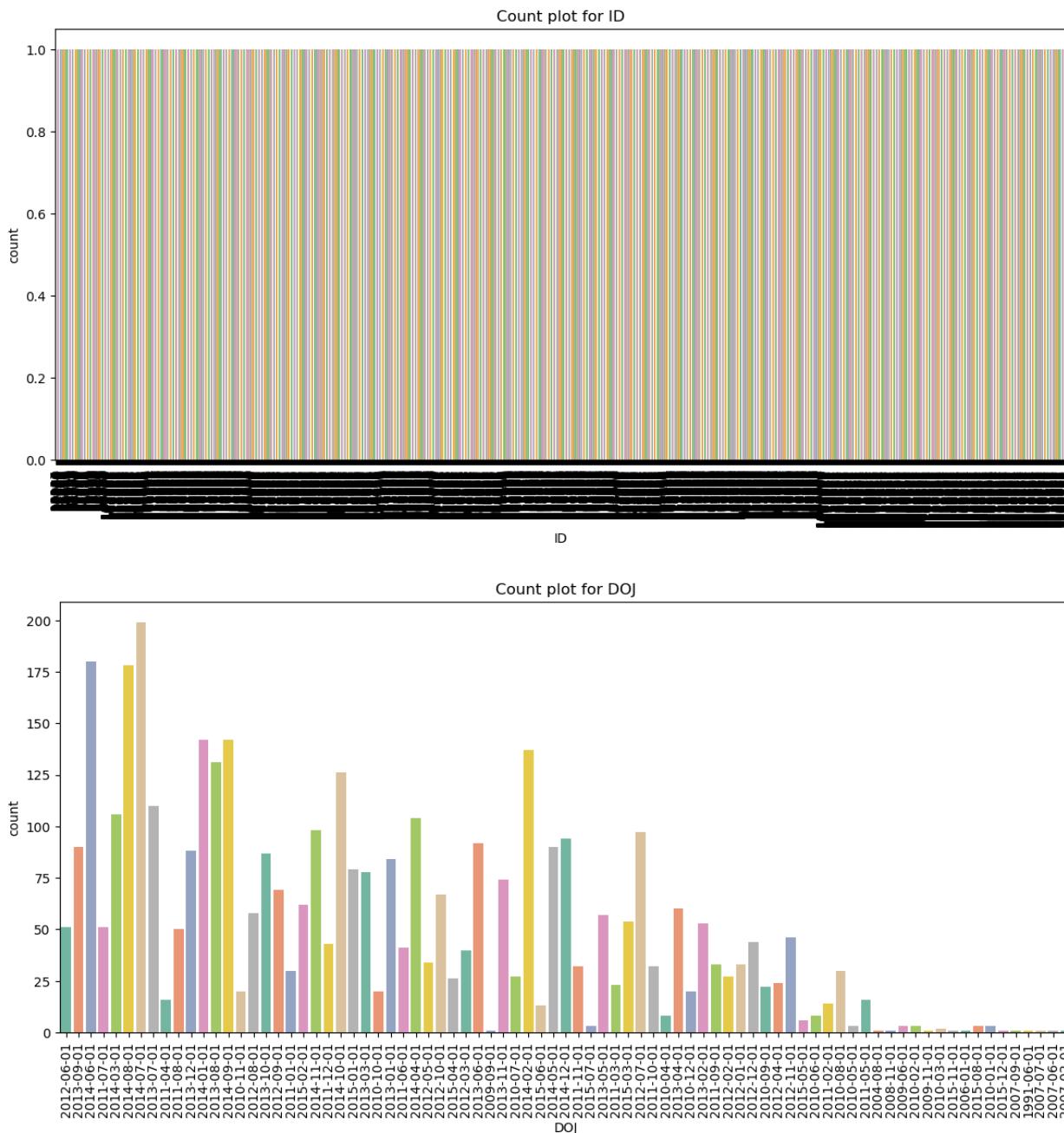
```
Analysis for CollegeState column:  
Unique values: ['Andhra Pradesh' 'Madhya Pradesh' 'Uttar Pradesh'  
'Delhi' 'Karnataka'  
'Tamil Nadu' 'West Bengal' 'Maharashtra' 'Haryana' 'Telangana'  
'Orissa'  
'Punjab' 'Kerala' 'Gujarat' 'Rajasthan' 'Chhattisgarh' 'Uttarakhand'  
'Jammu and Kashmir' 'Jharkhand' 'Himachal Pradesh' 'Bihar' 'Assam'  
'Goa'  
'Sikkim' 'Union Territory' 'Meghalaya']  
No of unique values: 26  
value counts:  
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
```

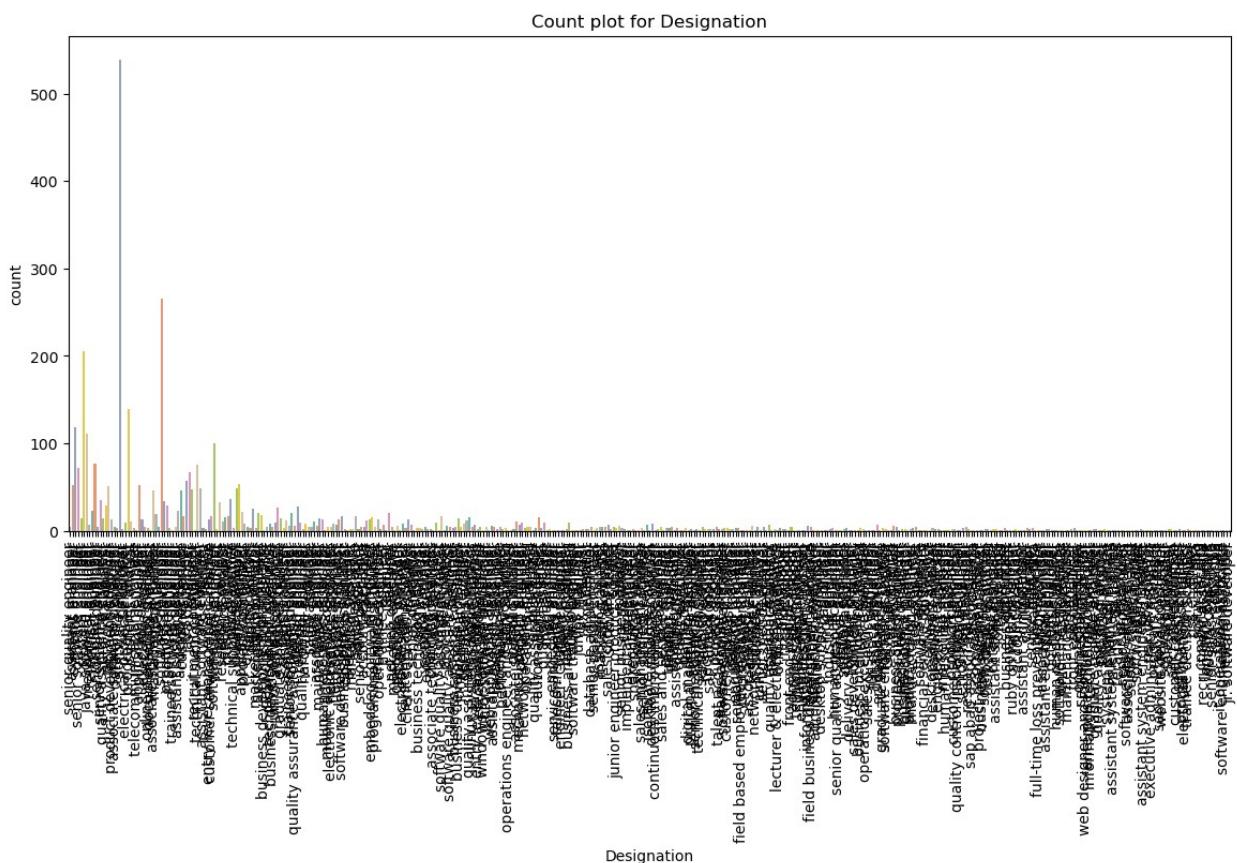
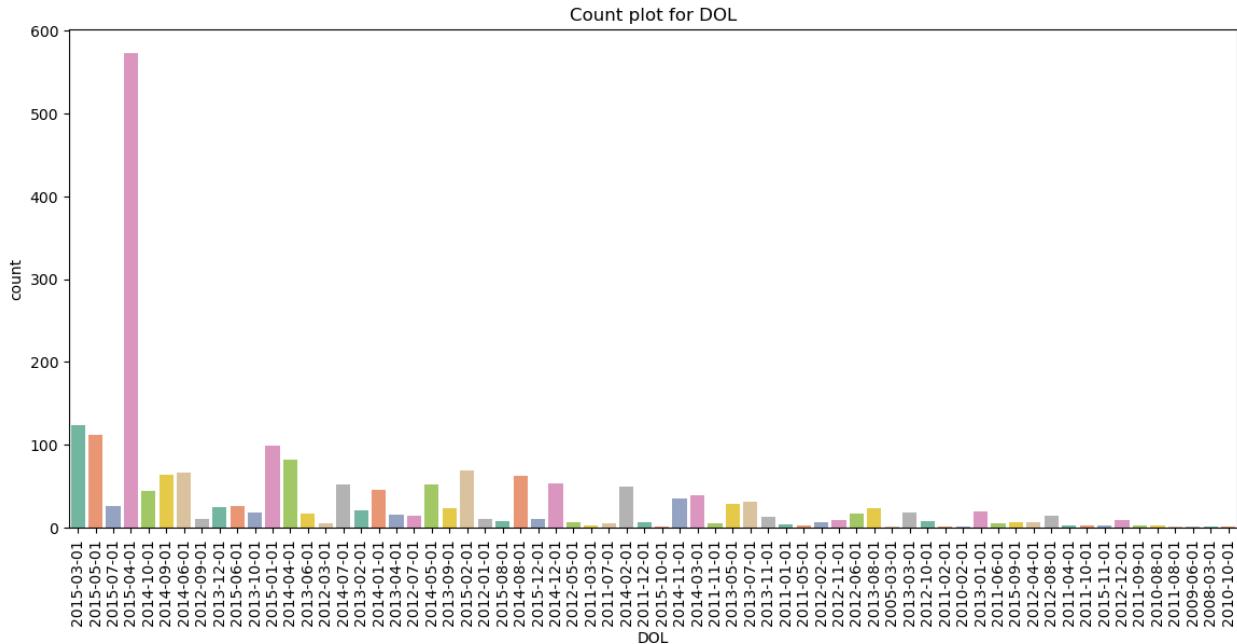
```
Assam          5
Union Territory 5
Sikkim         3
Meghalaya      2
Goa            1
Name: count, dtype: int64
```

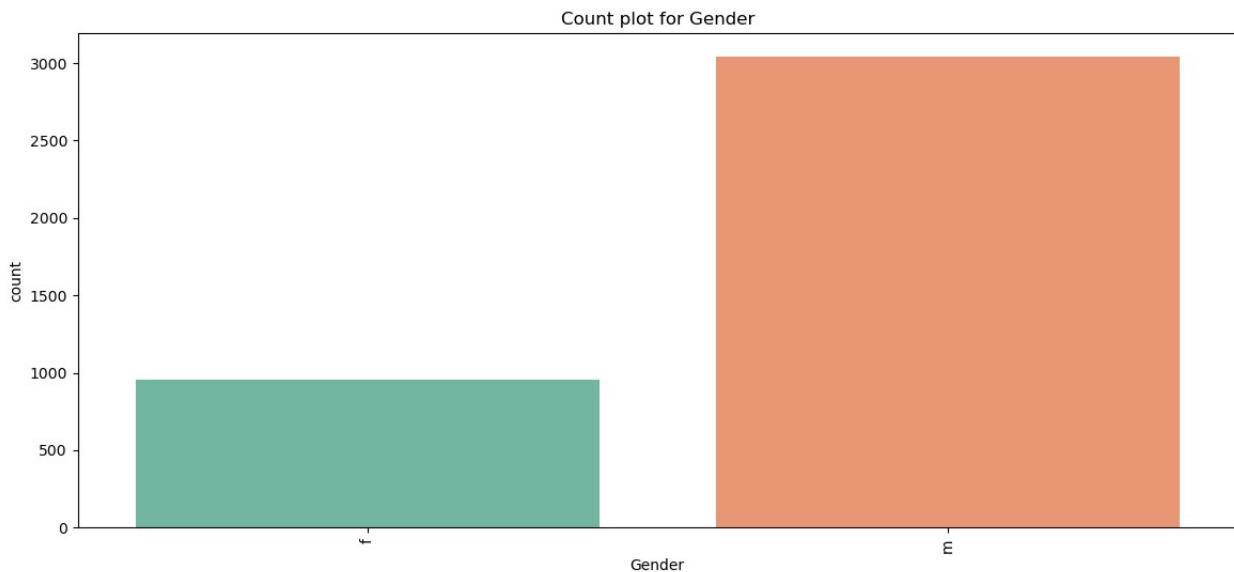
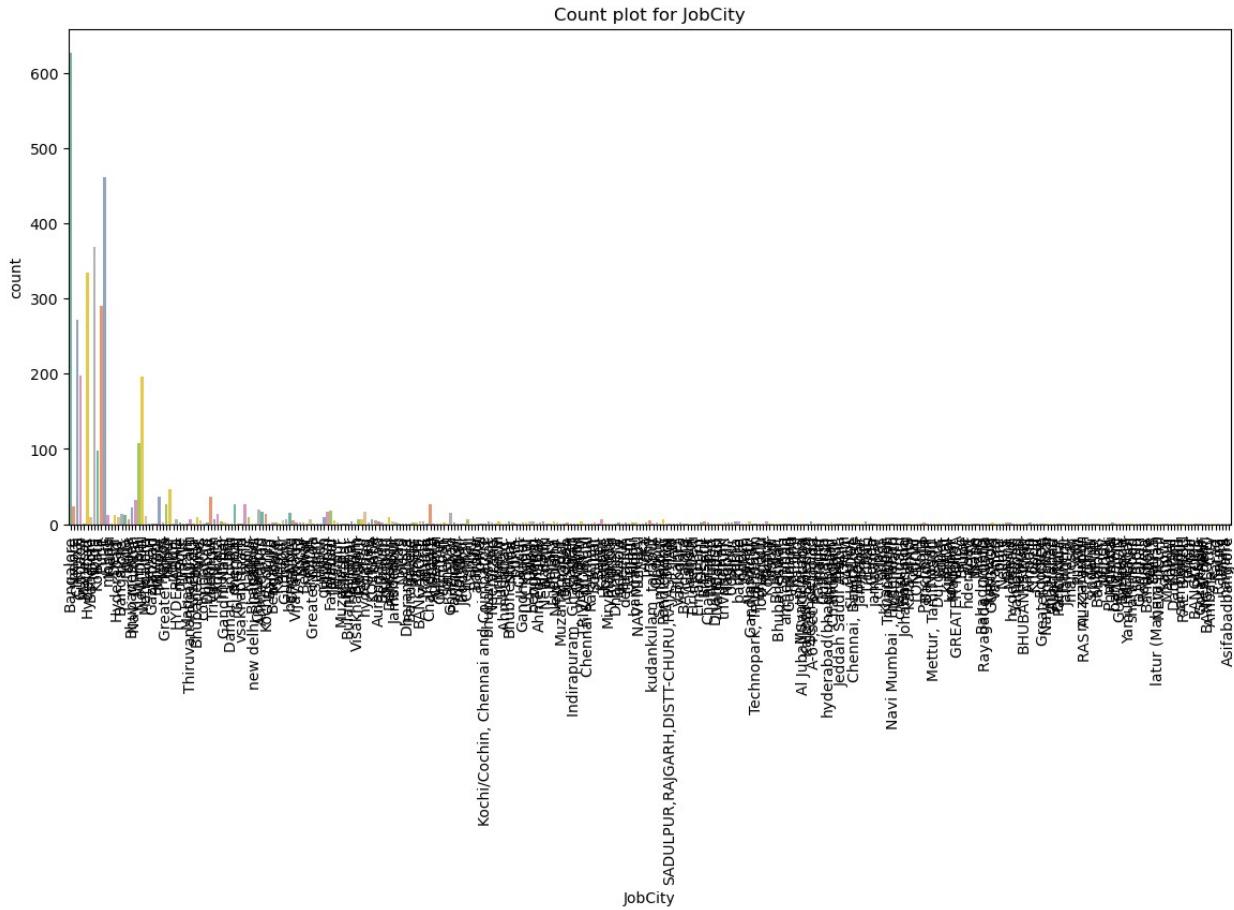
```
Analysis for GraduationYear column:
Unique values: [2011, 2012, 2014, 2016, 2013, ..., 2015, 2009, 2017,
0, 2007]
Length: 11
Categories (11, int64): [0, 2007, 2009, 2010, ..., 2014, 2015, 2016,
2017]
No of unique values: 11
value counts:
GraduationYear
2013    1181
2014    1036
2012     847
2011     507
2010     292
2015      94
2009      24
2017       8
2016       7
0          1
2007       1
Name: count, dtype: int64
```

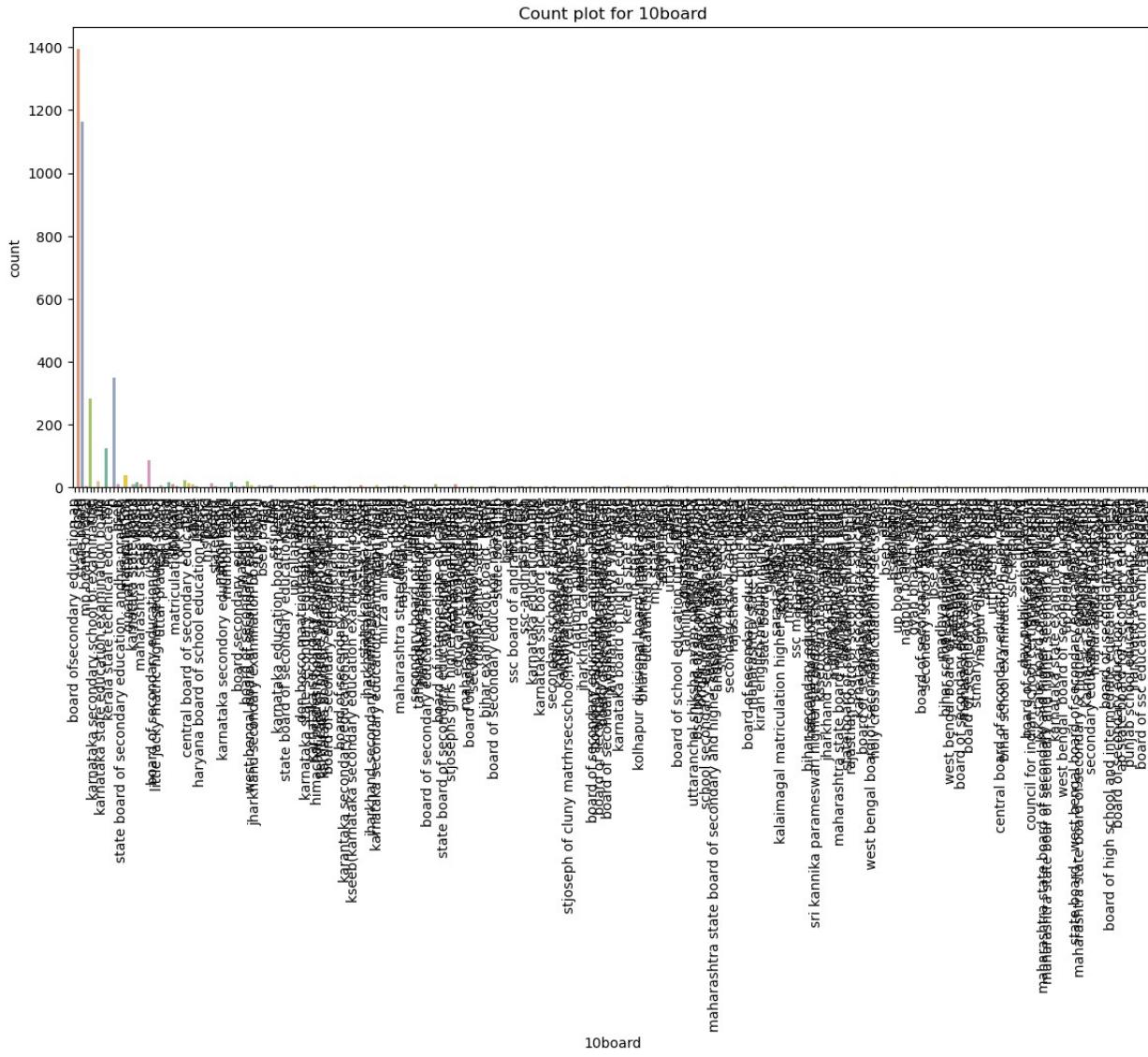
```
#univariate visual analysis for categorical columns
for col in categorical_cols:
    plt.figure(figsize=(14,6))

    #countplot
    sns.countplot(x=df[col], palette='Set2')
    plt.title(f"Count plot for {col}")
    plt.xticks(rotation=90)
    plt.show()
```



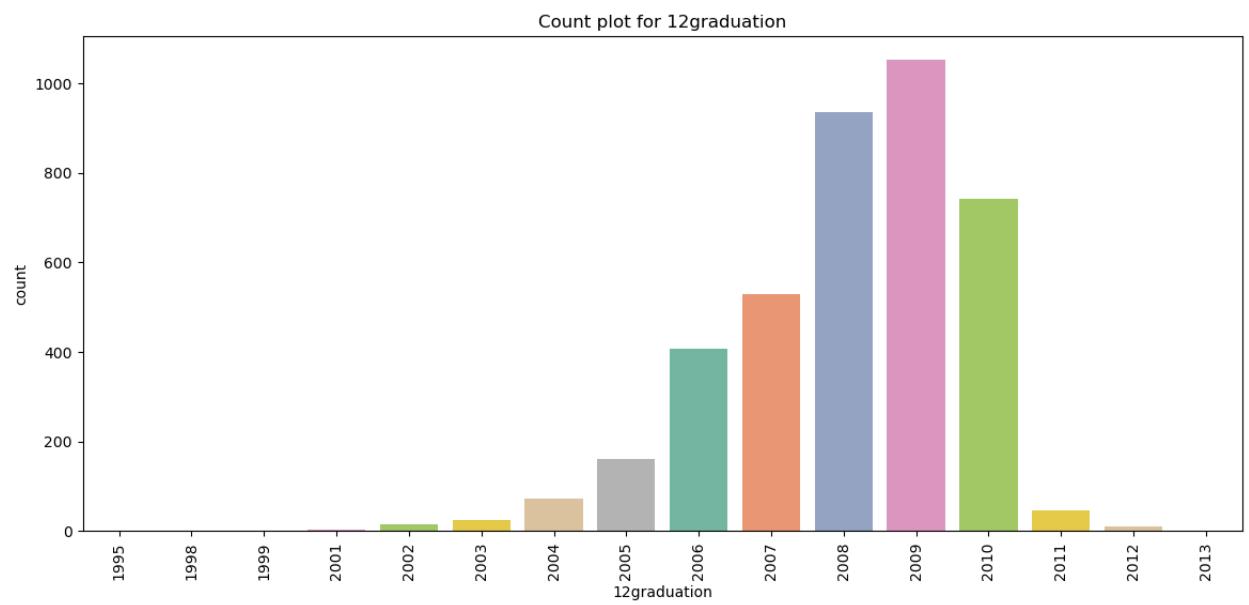


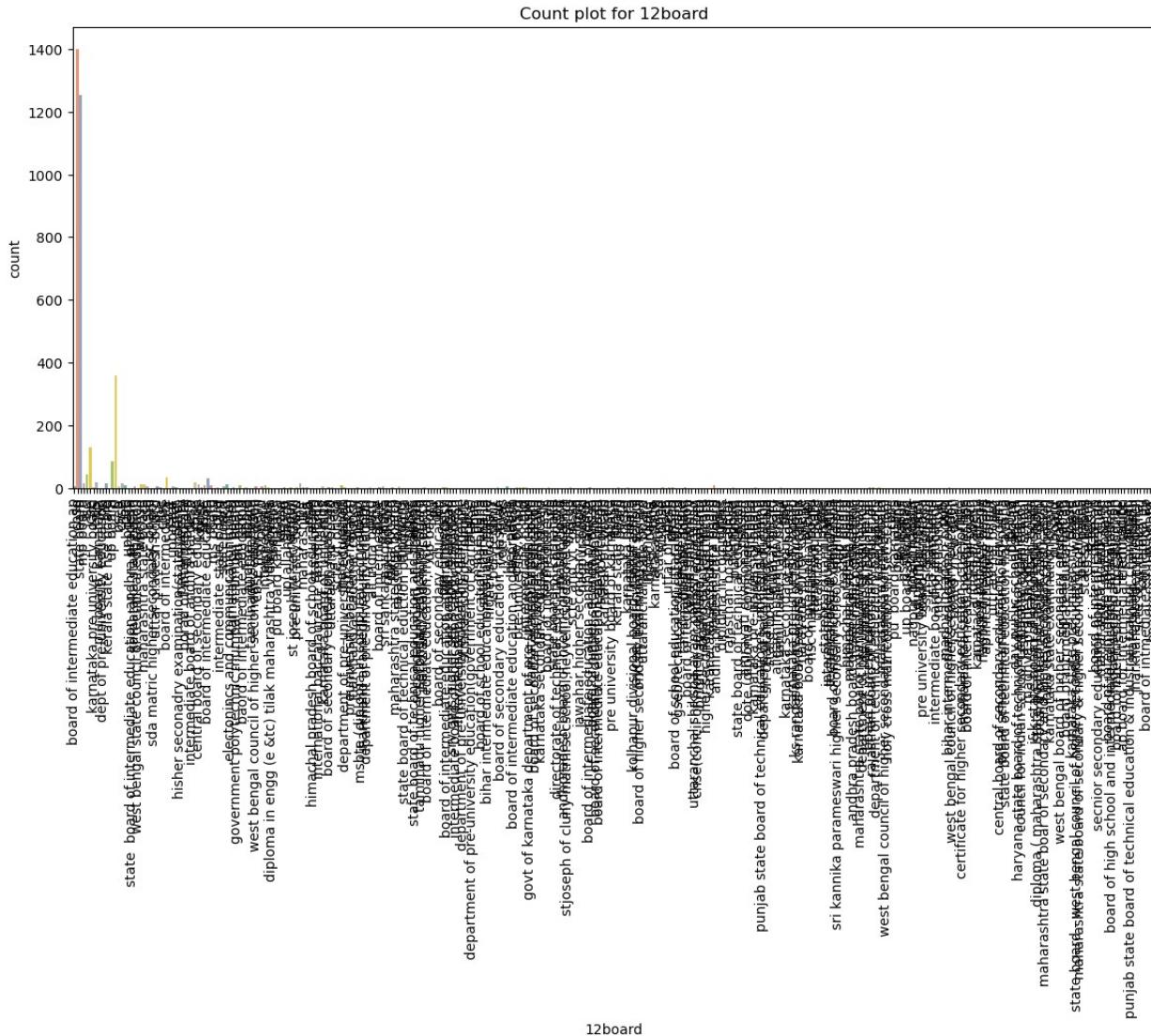




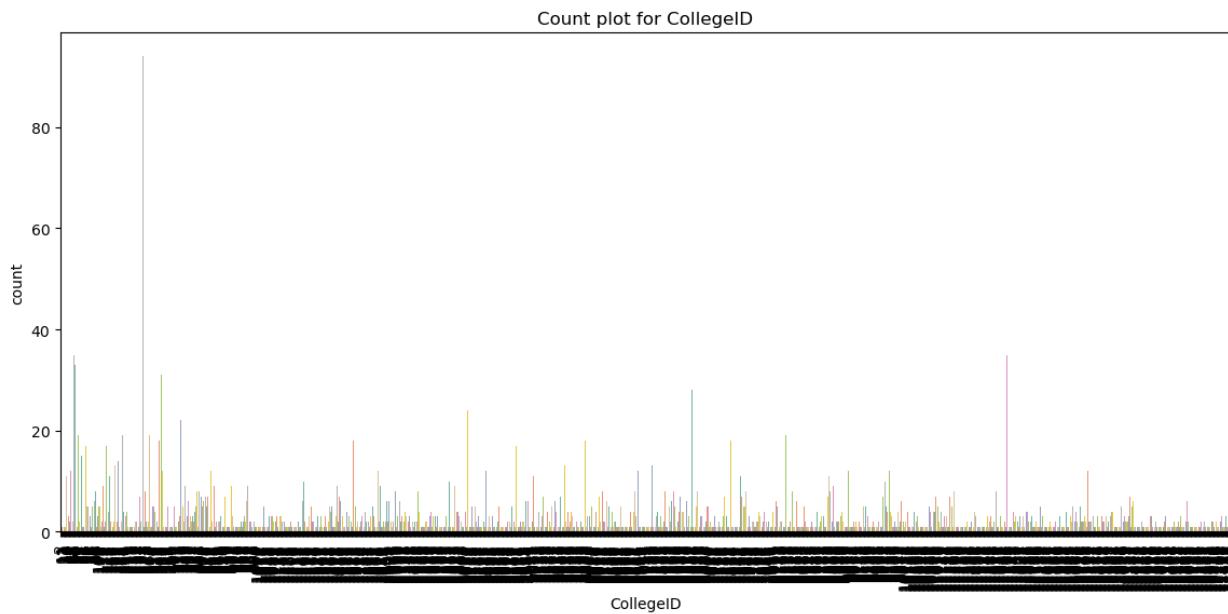
```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:
FutureWarning: The default of observed=False is deprecated and will be
changed to True in a future version of pandas. Pass observed=False to
retain current behavior or observed=True to adopt the future default
and silence this warning.
```

```
grouped_vals = vals.groupby(grouper)
```



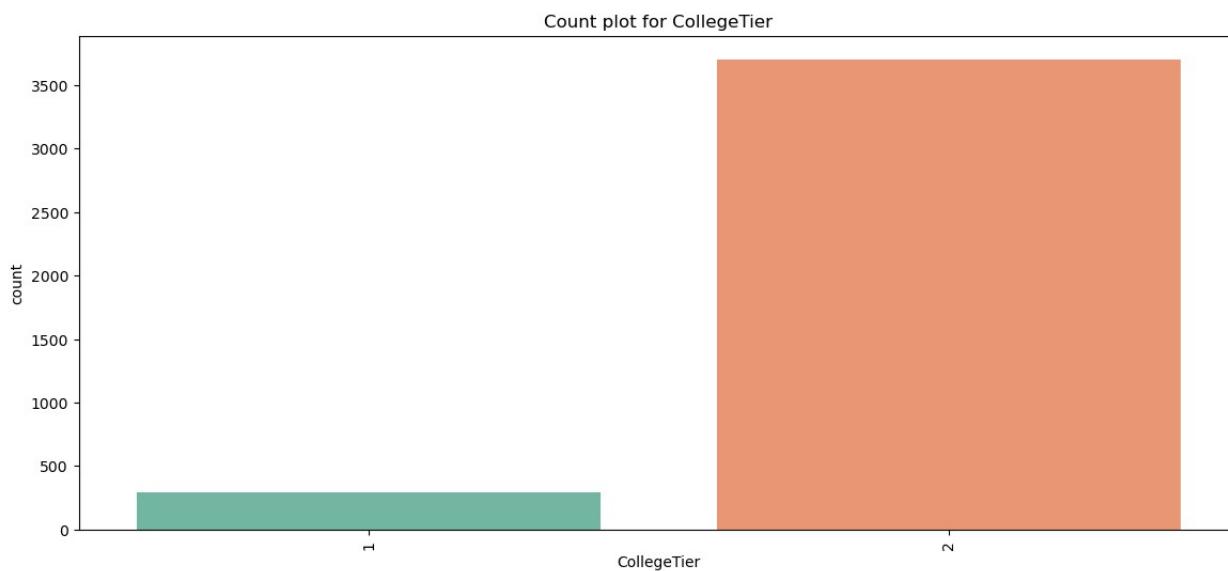


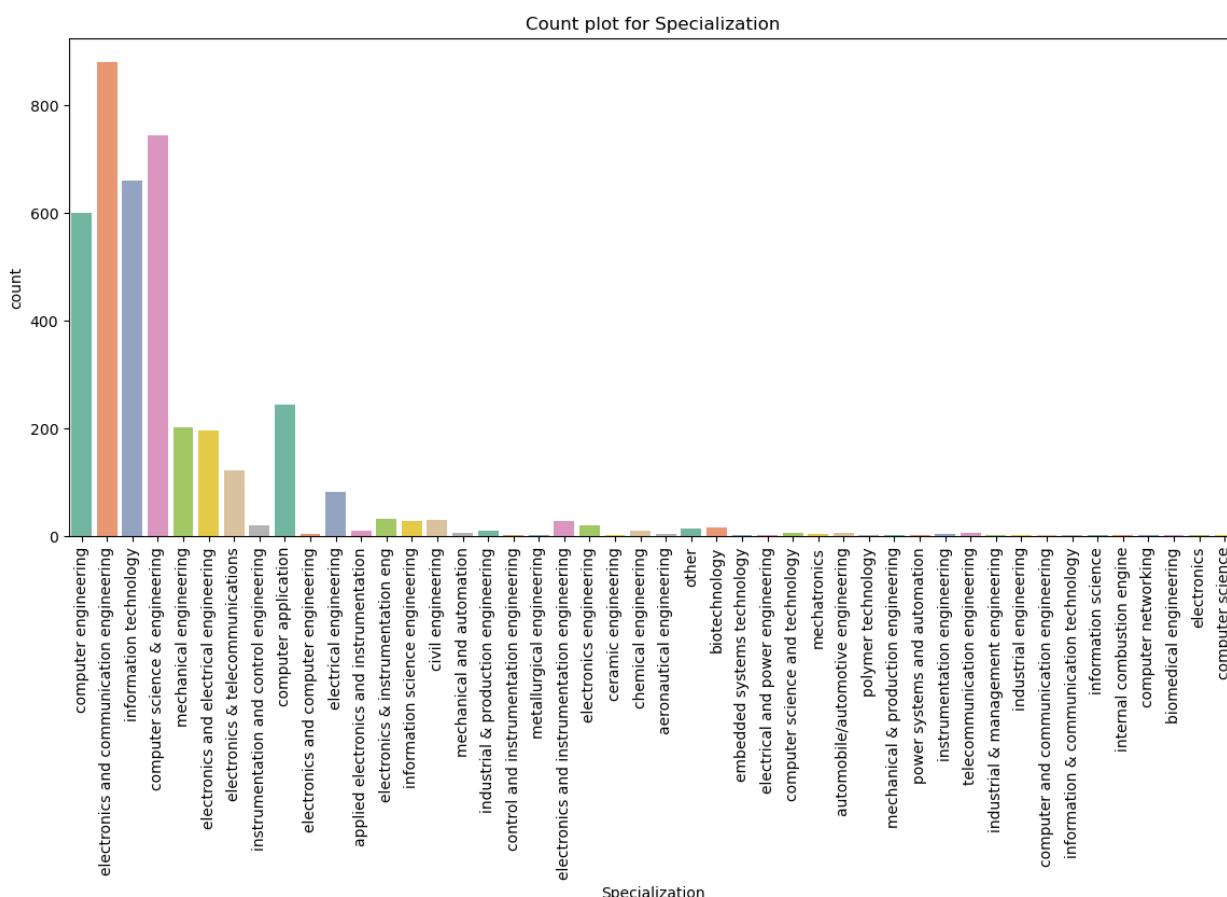
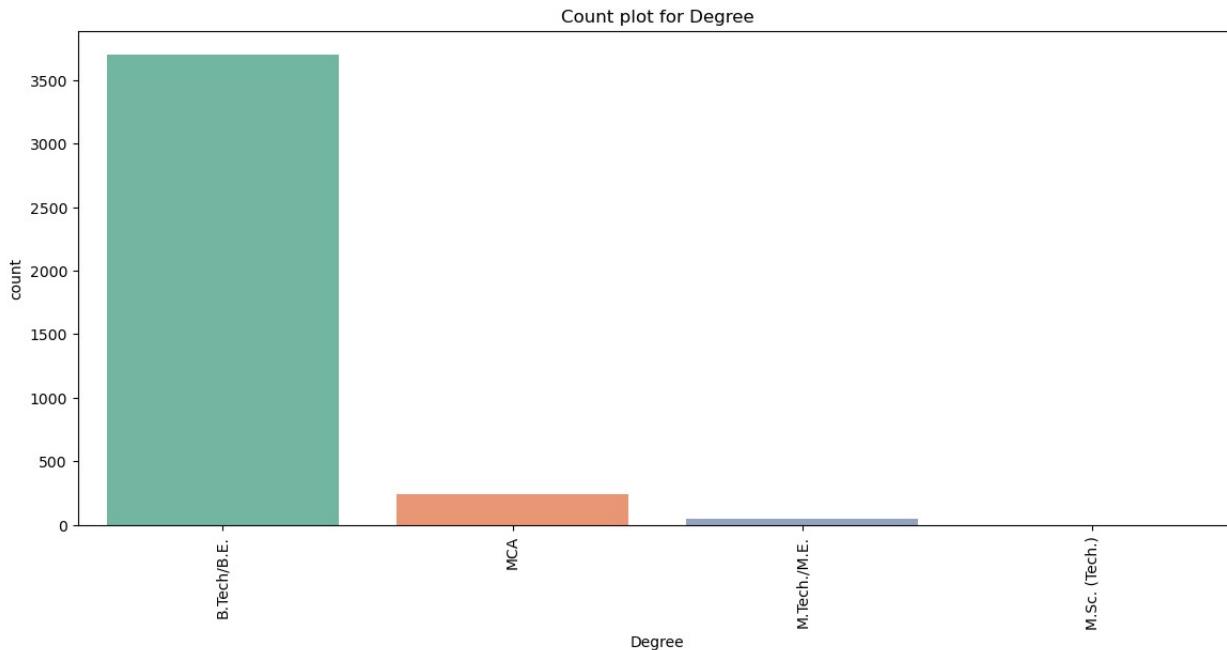
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.



```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:  
FutureWarning: The default of observed=False is deprecated and will be  
changed to True in a future version of pandas. Pass observed=False to  
retain current behavior or observed=True to adopt the future default  
and silence this warning.
```

```
grouped_vals = vals.groupby(grouper)
```

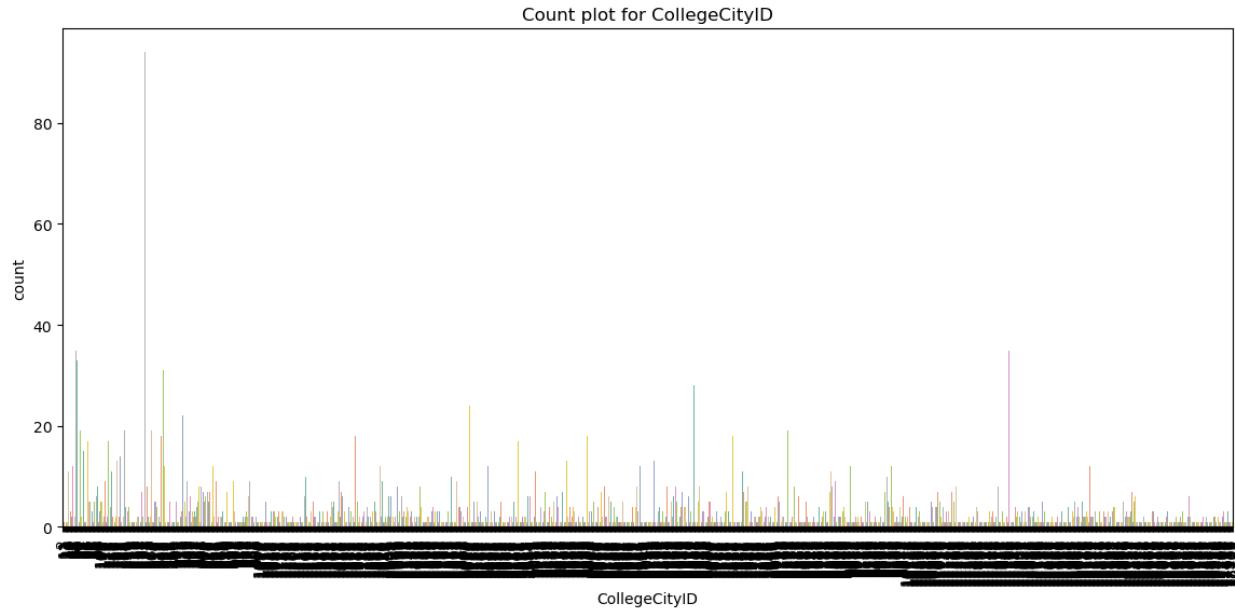




```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:
FutureWarning: The default of observed=False is deprecated and will be
```

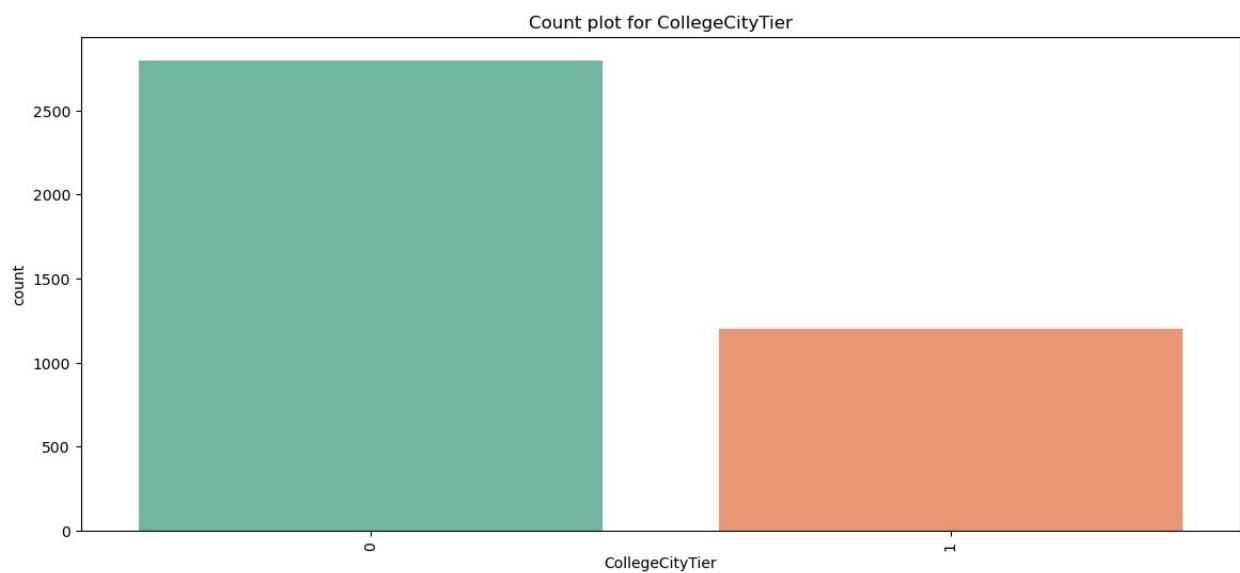
```
changed to True in a future version of pandas. Pass observed=False to  
retain current behavior or observed=True to adopt the future default  
and silence this warning.
```

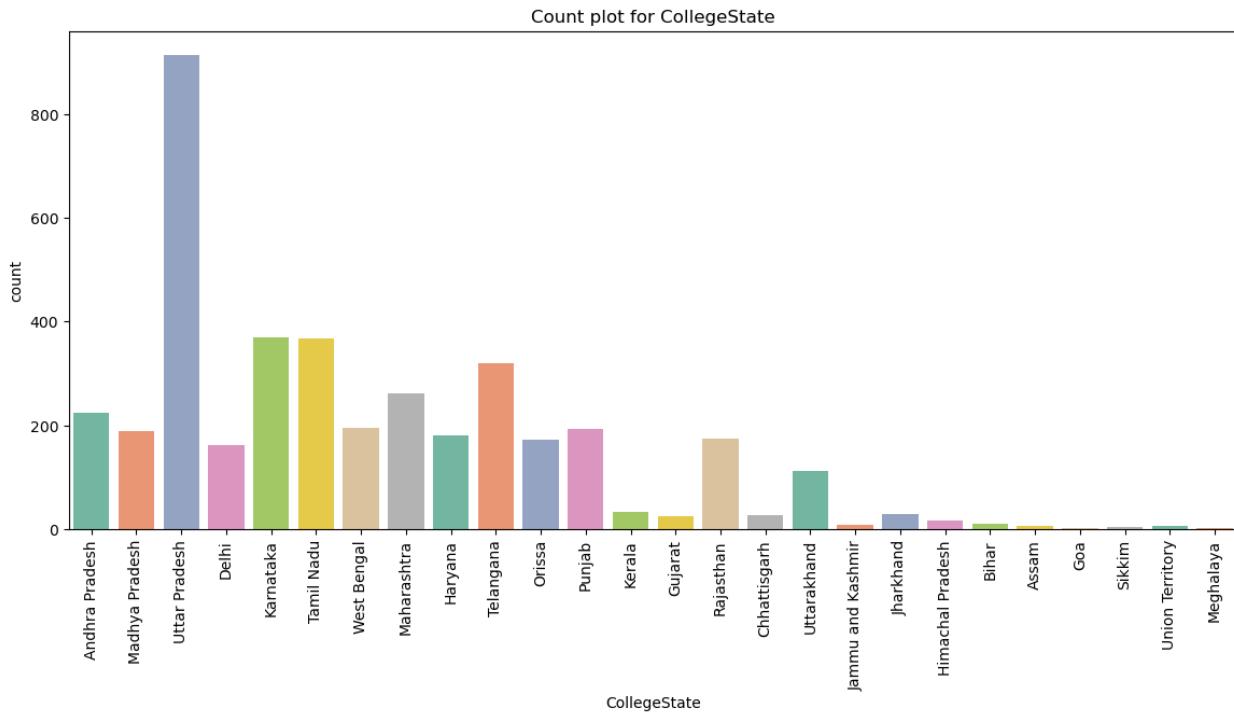
```
grouped_vals = vals.groupby(grouper)
```



```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:  
FutureWarning: The default of observed=False is deprecated and will be  
changed to True in a future version of pandas. Pass observed=False to  
retain current behavior or observed=True to adopt the future default  
and silence this warning.
```

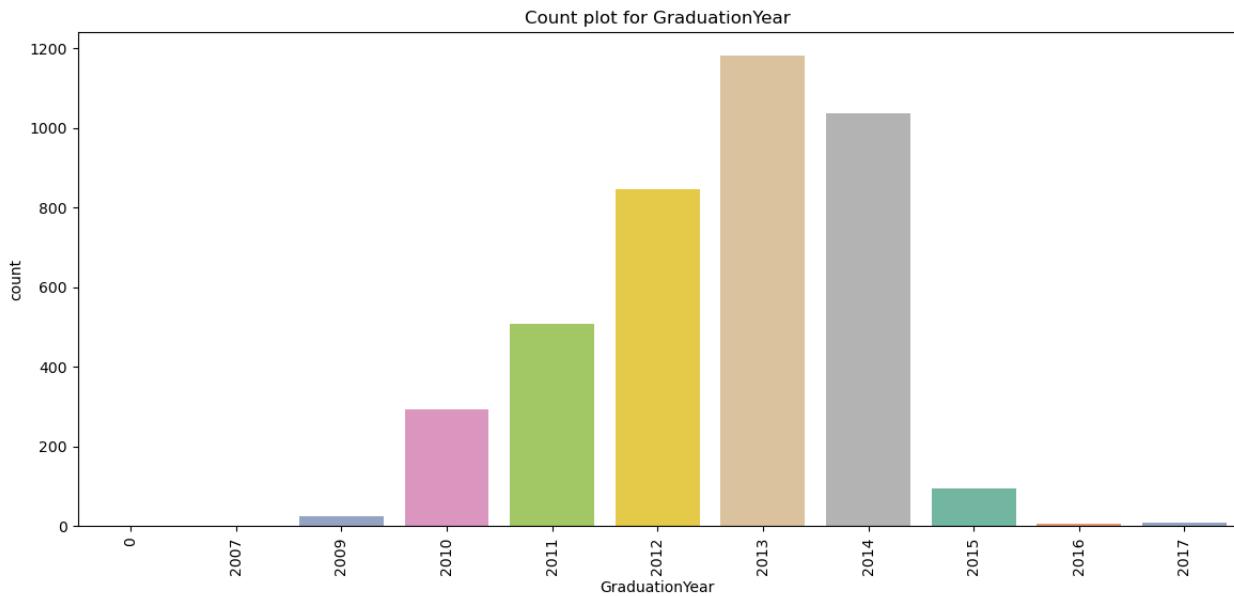
```
grouped_vals = vals.groupby(grouper)
```





```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:
FutureWarning: The default of observed=False is deprecated and will be
changed to True in a future version of pandas. Pass observed=False to
retain current behavior or observed=True to adopt the future default
and silence this warning.
```

```
grouped_vals = vals.groupby(grouper)
```



```

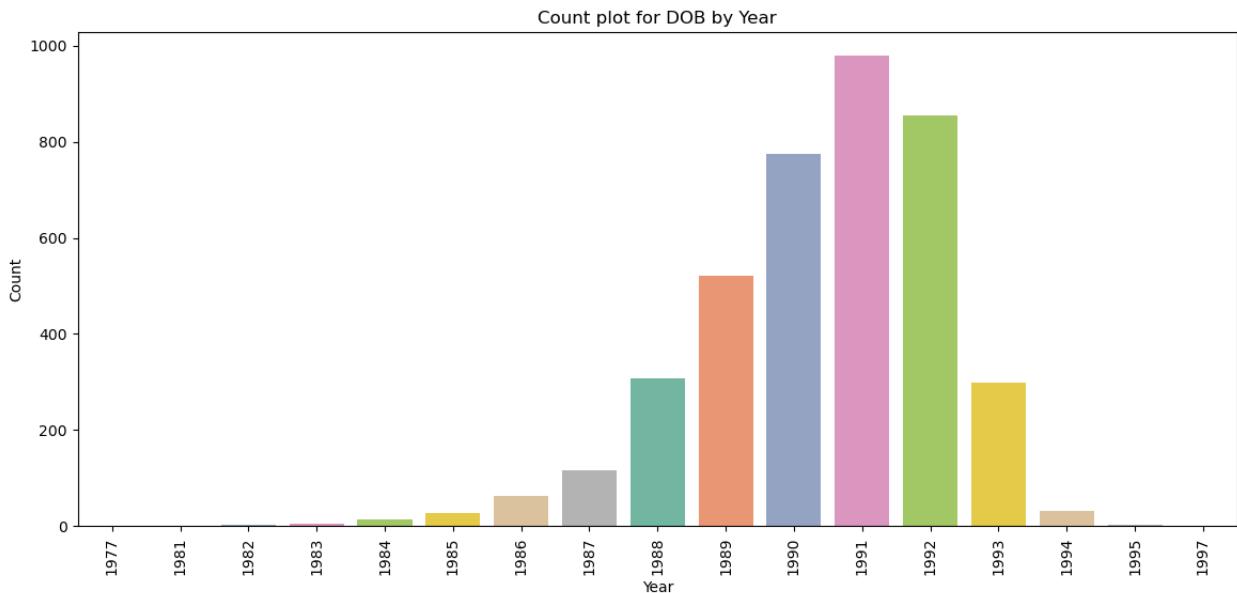
# Attempt to convert the DOB, DOJ, DOL columns to datetime
df['DOB'] = pd.to_datetime(df['DOB'], errors='coerce')
df['DOJ'] = pd.to_datetime(df['DOJ'], errors='coerce')
df['DOL'] = pd.to_datetime(df['DOL'], errors='coerce')

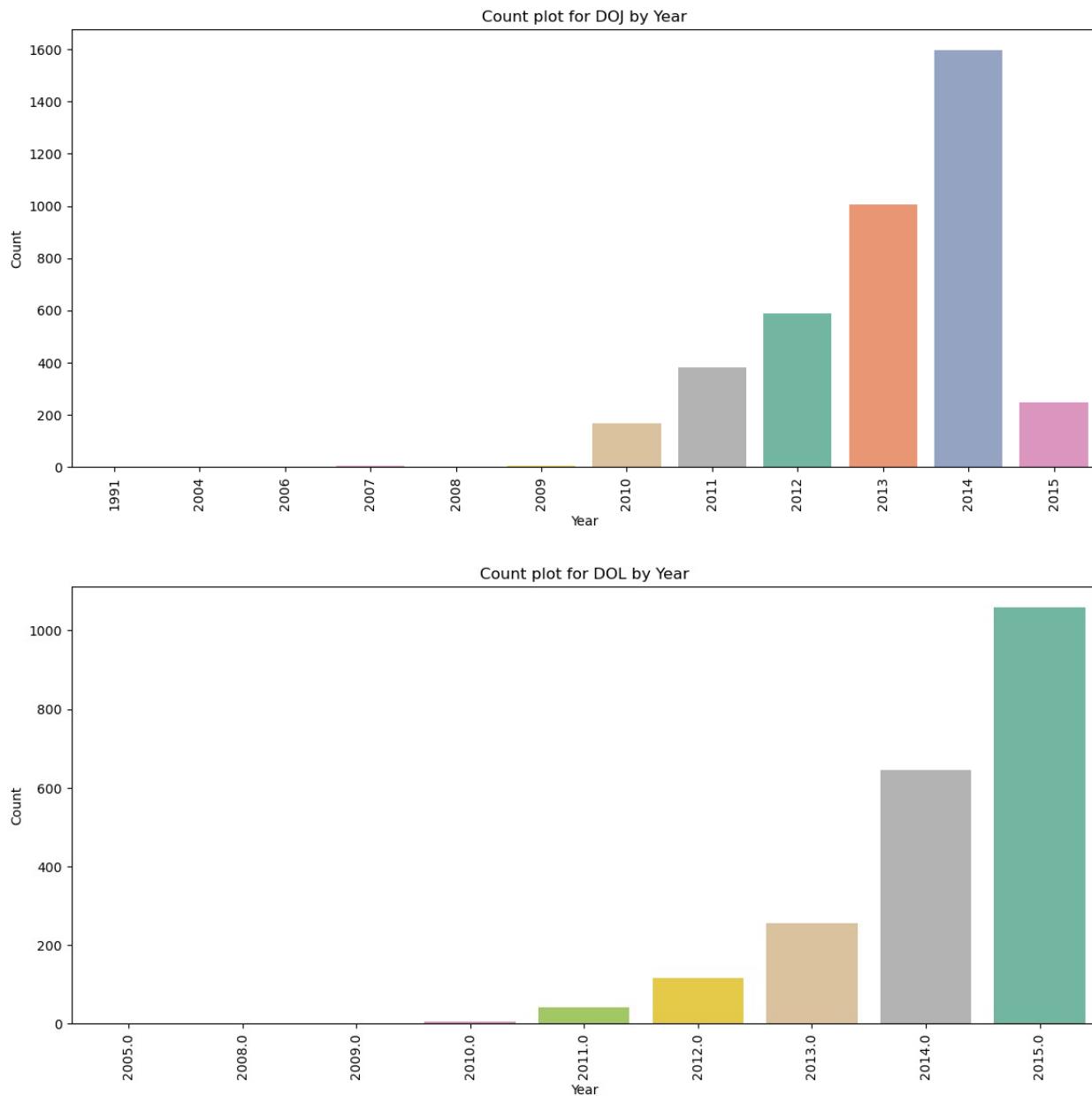
# List of columns to plot
date_columns = ['DOB', 'DOJ', 'DOL']

# Plotting the count plot for the year-based columns without creating
# new columns
for col in date_columns:
    plt.figure(figsize=(14, 6))

    # Extract year and plot
    year_counts = df[col].dt.year.value_counts().sort_index()
    sns.barplot(x=year_counts.index, y=year_counts.values,
    palette='Set2')
    plt.title(f"Count plot for {col} by Year")
    plt.xlabel('Year')
    plt.ylabel('Count')
    plt.xticks(rotation=90)
    plt.show()

```





```
# Find unique values and count in '10board' column
unique_boards = df['10board'].unique()
unique_boards_count = df['10board'].nunique()

print("Unique Boards:", unique_boards)
print("Count of Unique Boards:", unique_boards_count)

Unique Boards: ['board of secondary 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 prades'  
'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' '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 board' '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 prades' 'up-board'
'bse,odisha']

```

Count of Unique Boards: 275

*#since there are 275 unique boards and basically there are 3 types of boards, CBSE, ICSE and various state boards, but for simplification purpose i am considering all those state boards as state board itself*

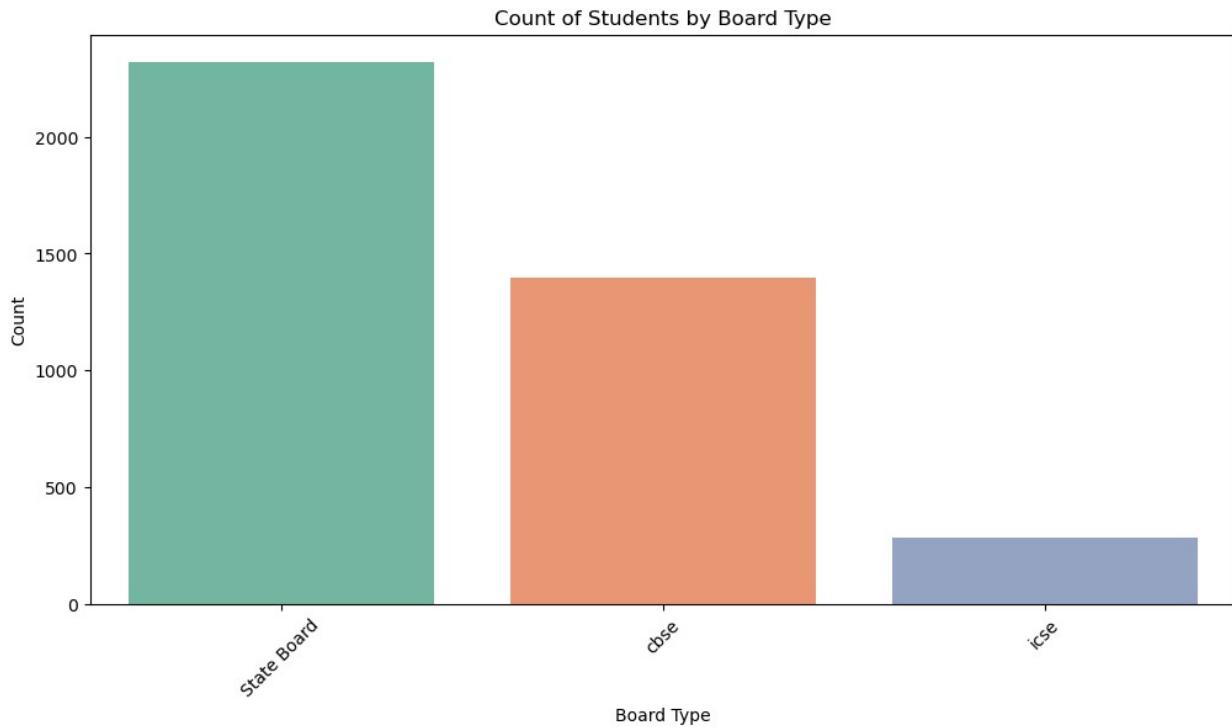
```

# Create a new column for board categorization
df['Board_Type'] = np.where(df['10board'].isin(['cbse', 'icse']),
df['10board'], 'State Board')

# Count plot for Board_Type
plt.figure(figsize=(12, 6))
sns.countplot(x='Board_Type', data=df, palette='Set2')
plt.title('Count of Students by Board Type')
plt.xlabel('Board Type')
plt.ylabel('Count')

```

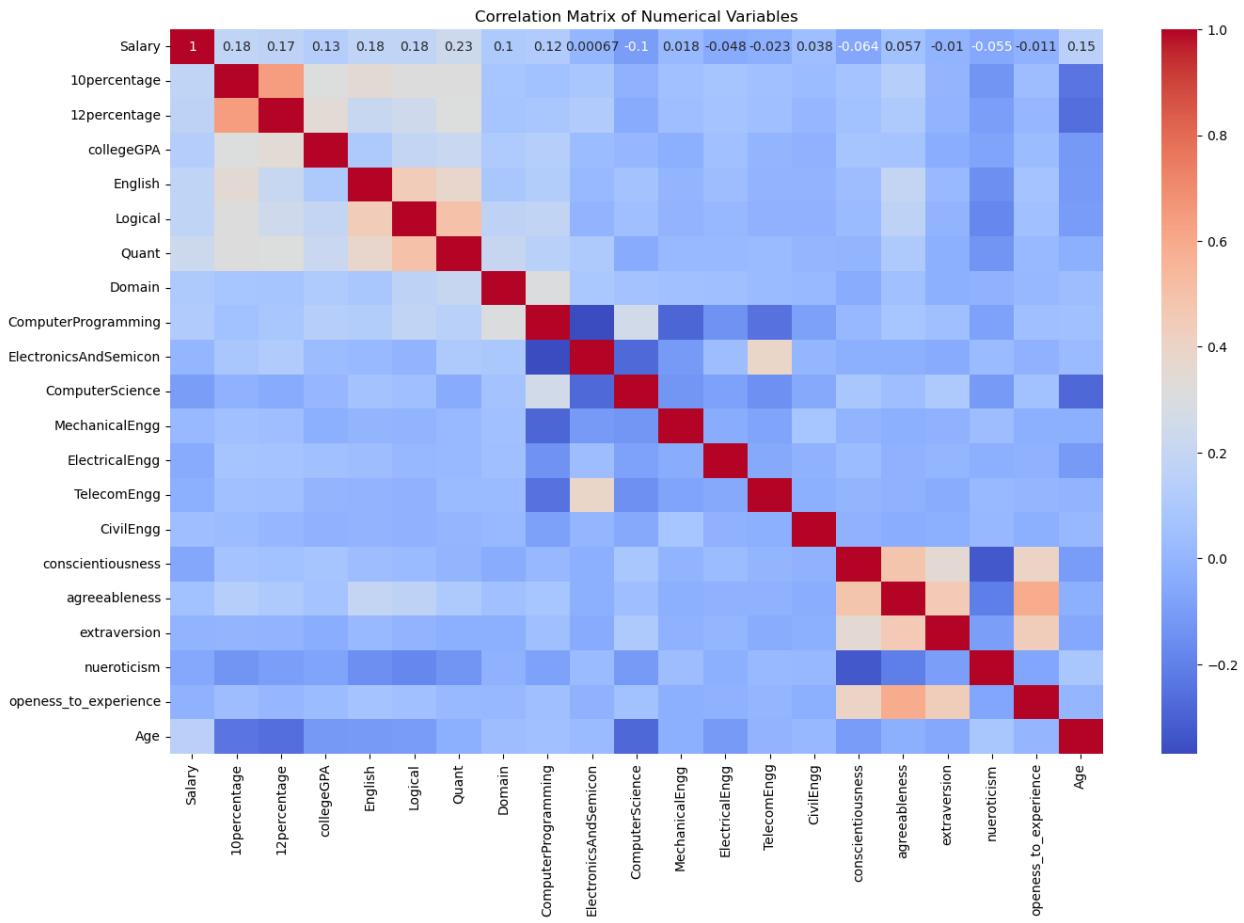
```
plt.xticks(rotation=45)
plt.show()
```



```
#Bivariate Analysis --To find relation between the variables or
columns
#since there are around approximate 40 columns it is not worth
plotting all the plots for each and every pair of columns, so as
already it is given that the Target variable is salary, we find
relation of salary with some key columns

#Heat Map

# Correlation matrix for numerical variables
correlation_matrix = df[numerical_cols].corr()
plt.figure(figsize=(16, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix of Numerical Variables')
plt.show()
```



```
#pairplot of interested columns
# Selecting numerical columns of interest
selected_cols = ['Salary', '10percentage', '12percentage',
'collegeGPA', 'English', 'Logical', 'Quant', 'Age']
# Ensure all these columns are numerical
df_selected = df[selected_cols].dropna()

# Creating a pairplot
sns.pairplot(df_selected, diag_kind='kde', kind='scatter',
palette='coolwarm')

# Show the plot
plt.suptitle('Pairplot of Selected Numerical Variables', y=1.02) # Adjusting title position
plt.show()

C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\axisgrid.py:1507:
UserWarning: Ignoring `palette` because no `hue` variable has been
assigned.
func(x=vector, **plot_kwargs)
```

```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.  
    with pd.option_context('mode.use_inf_as_na', True):  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\axisgrid.py:1507:  
UserWarning: Ignoring `palette` because no `hue` variable has been  
assigned.  
    func(x=vector, **plot_kwargs)  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.  
    with pd.option_context('mode.use_inf_as_na', True):  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\axisgrid.py:1507:  
UserWarning: Ignoring `palette` because no `hue` variable has been  
assigned.  
    func(x=vector, **plot_kwargs)  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.  
    with pd.option_context('mode.use_inf_as_na', True):  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\axisgrid.py:1507:  
UserWarning: Ignoring `palette` because no `hue` variable has been  
assigned.  
    func(x=vector, **plot_kwargs)  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.  
    with pd.option_context('mode.use_inf_as_na', True):  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\axisgrid.py:1507:  
UserWarning: Ignoring `palette` because no `hue` variable has been  
assigned.  
    func(x=vector, **plot_kwargs)  
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:  
FutureWarning: use_inf_as_na option is deprecated and will be removed  
in a future version. Convert inf values to NaN before operating  
instead.
```

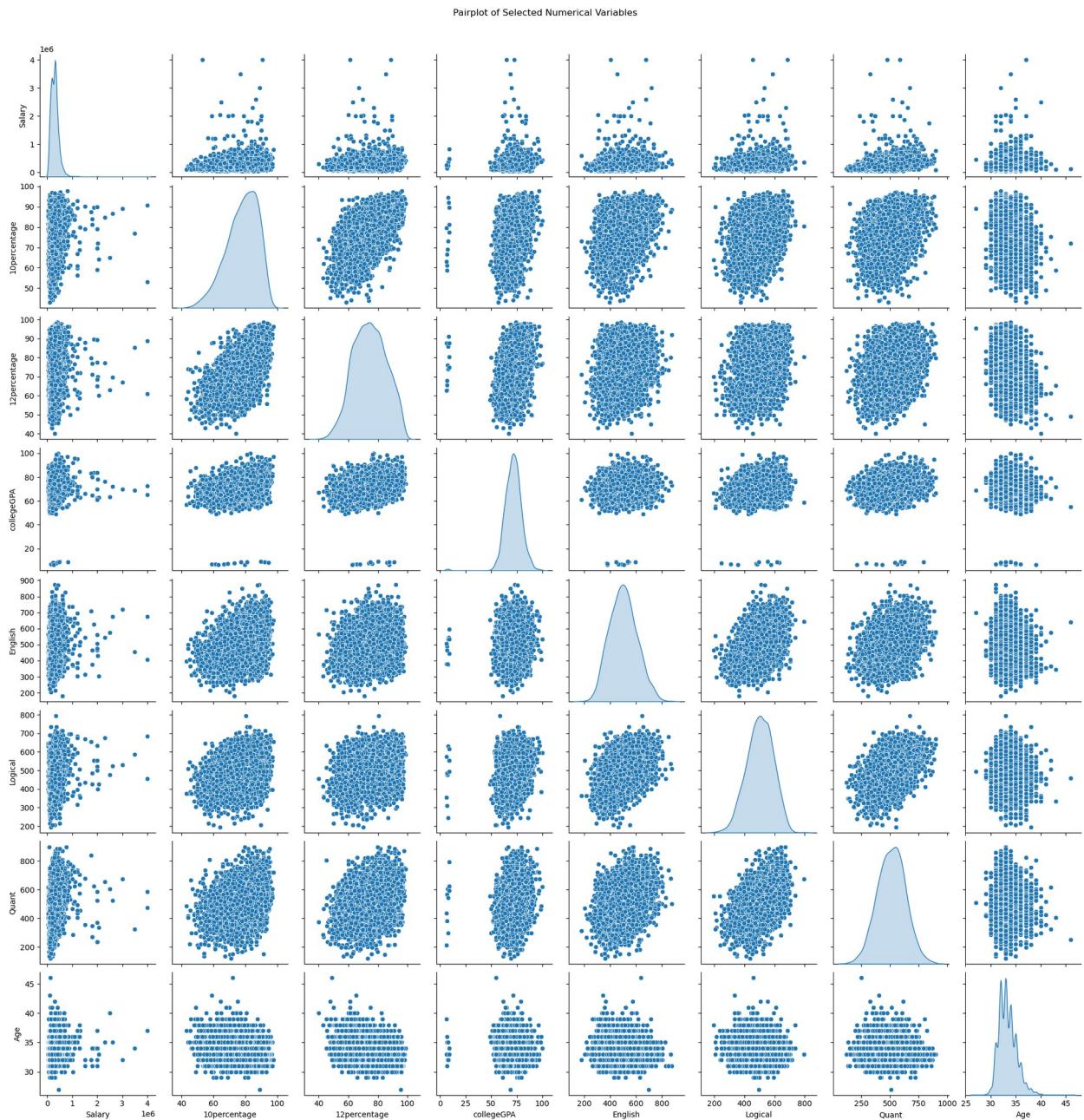




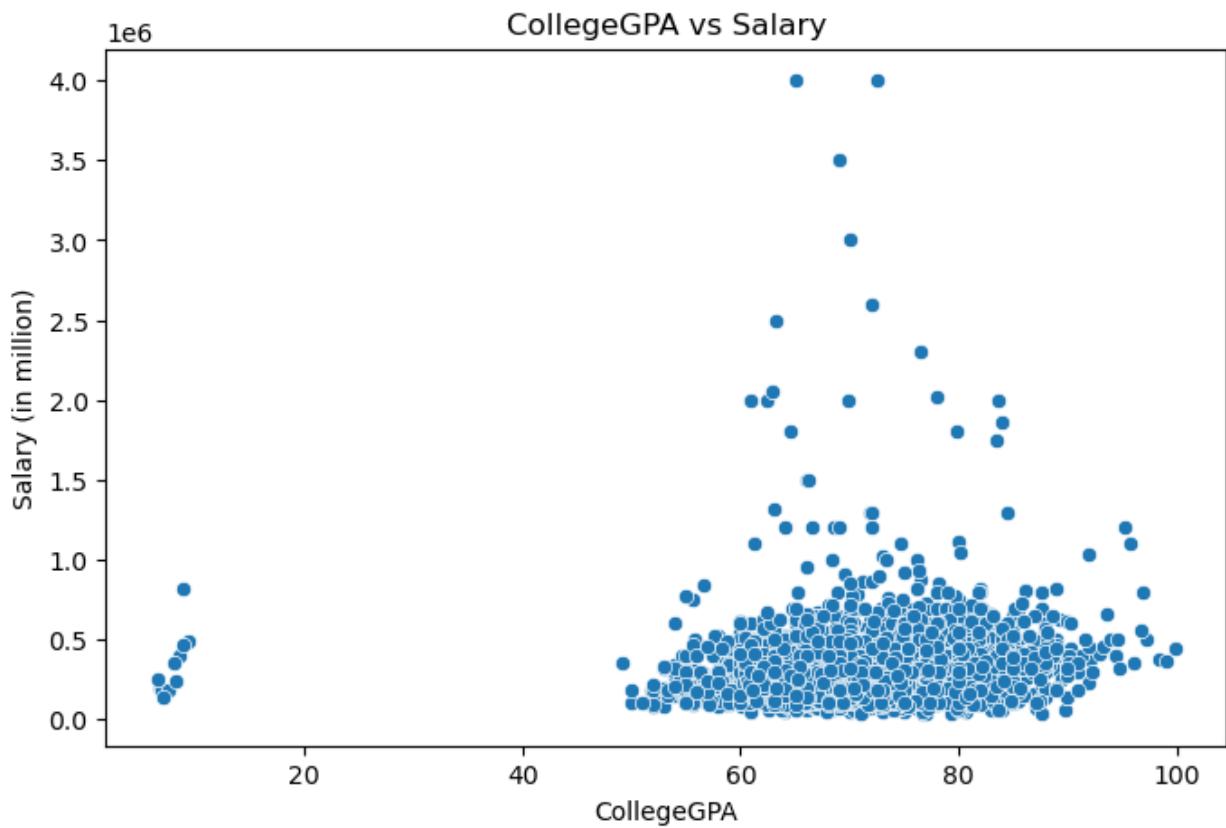




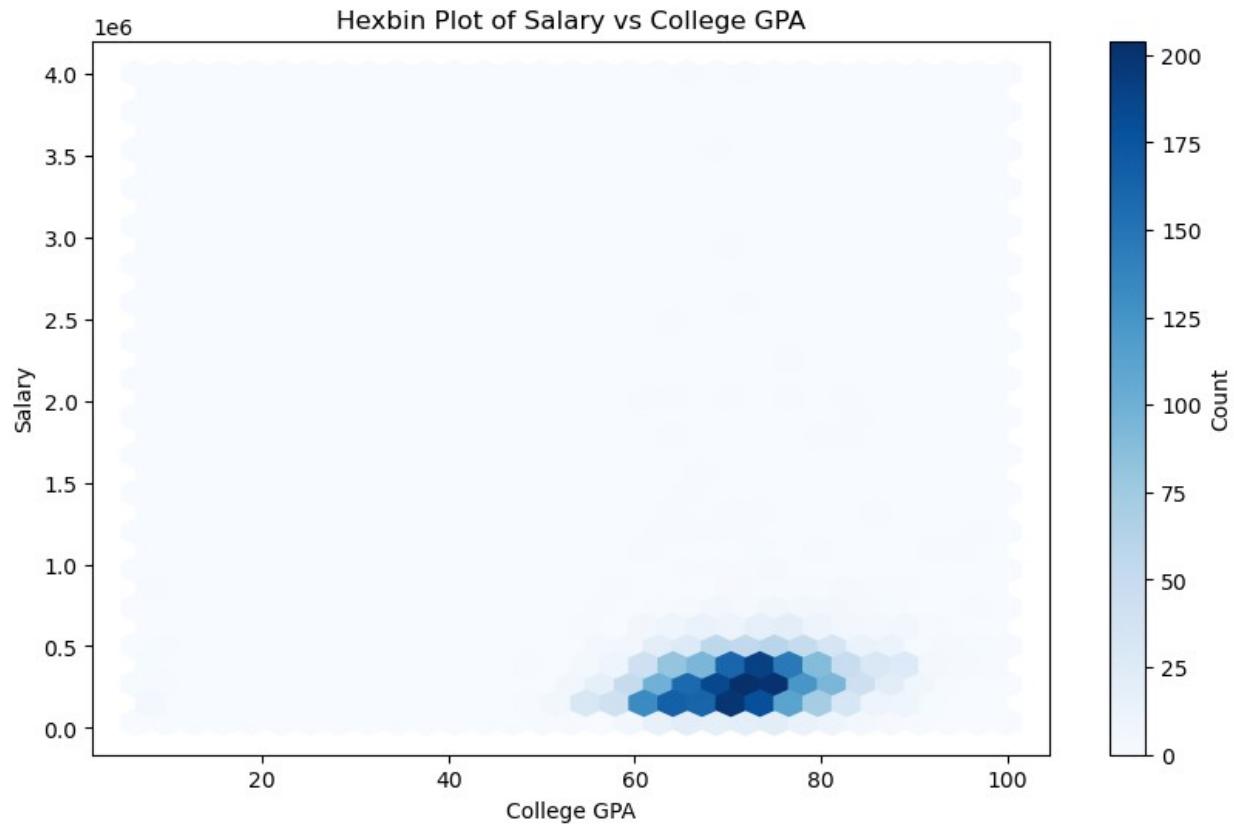




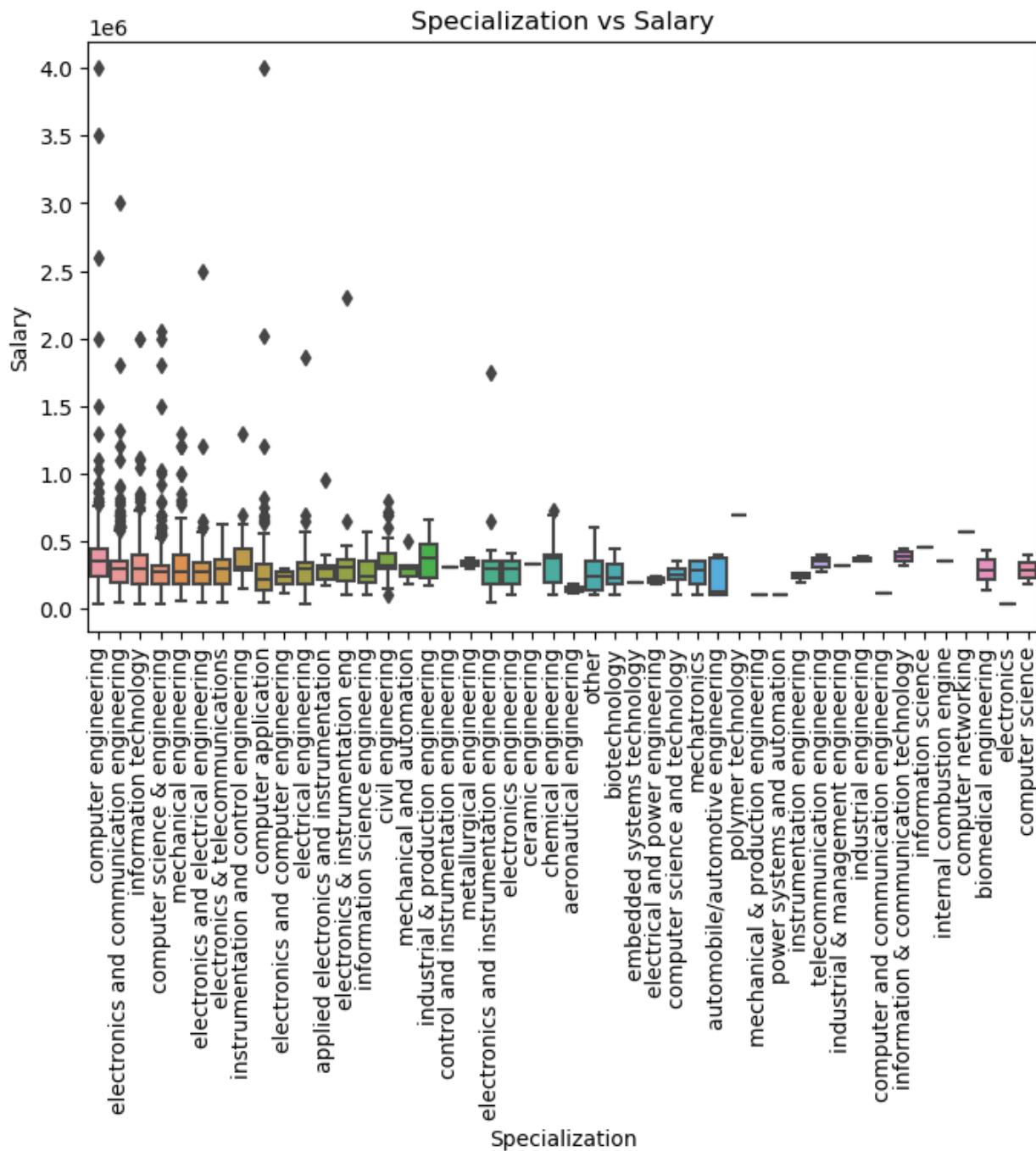
```
#scatterplot of salary vs CollegeGPA
plt.figure(figsize=(8,5))
# plt.scatter(df['collegeGPA'],df['Salary'])
sns.scatterplot(x='collegeGPA', y='Salary', data=df)
plt.xlabel('CollegeGPA')
plt.ylabel('Salary (in million)')
plt.title('CollegeGPA vs Salary')
plt.show()
```



```
#it is same as above but just the plotting is different
plt.figure(figsize=(10, 6))
plt.hexbin(df['collegeGPA'], df['Salary'], gridsize=30, cmap='Blues')
plt.colorbar(label='Count')
plt.title('Hexbin Plot of Salary vs College GPA')
plt.xlabel('College GPA')
plt.ylabel('Salary')
plt.show()
```



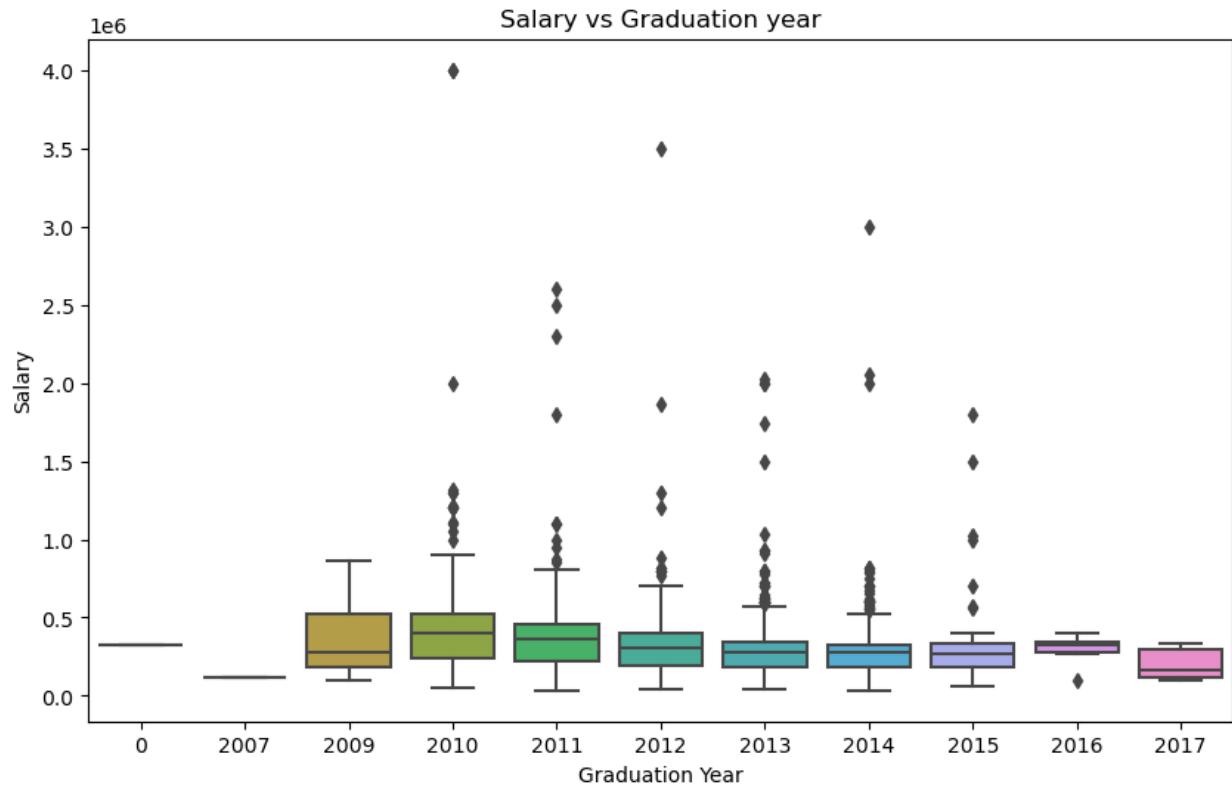
```
#Box plot of Salary by Specialization
plt.figure(figsize=(8,5))
sns.boxplot(x='Specialization',y='Salary',data=df)
plt.xticks(rotation=90)
plt.title('Specialization vs Salary')
plt.show()
```



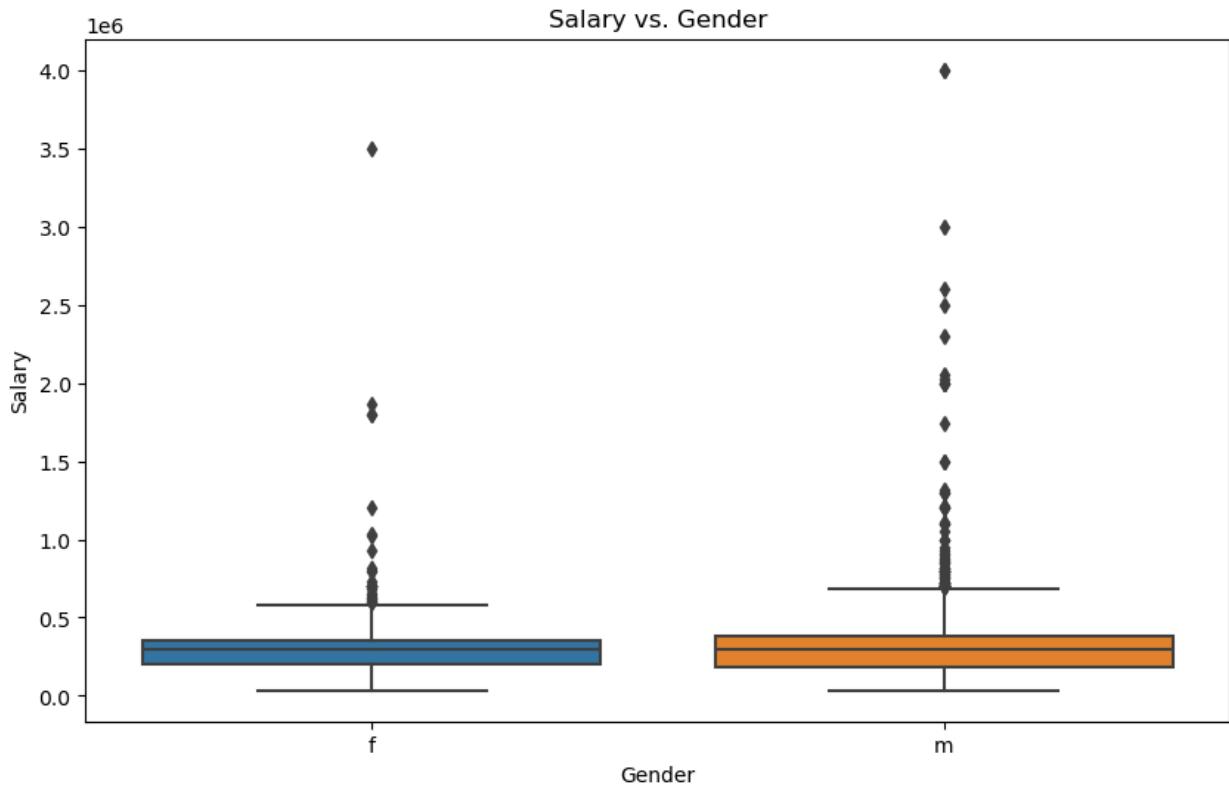
```
#salary vs Graduation year
plt.figure(figsize=(10,6))
sns.boxplot(x='GraduationYear',y='Salary',data=df)
plt.title("Salary vs Graduation year")
plt.xlabel('Graduation Year')
plt.ylabel('Salary')
plt.show()
```

```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:  
FutureWarning: The default of observed=False is deprecated and will be  
changed to True in a future version of pandas. Pass observed=False to  
retain current behavior or observed=True to adopt the future default  
and silence this warning.
```

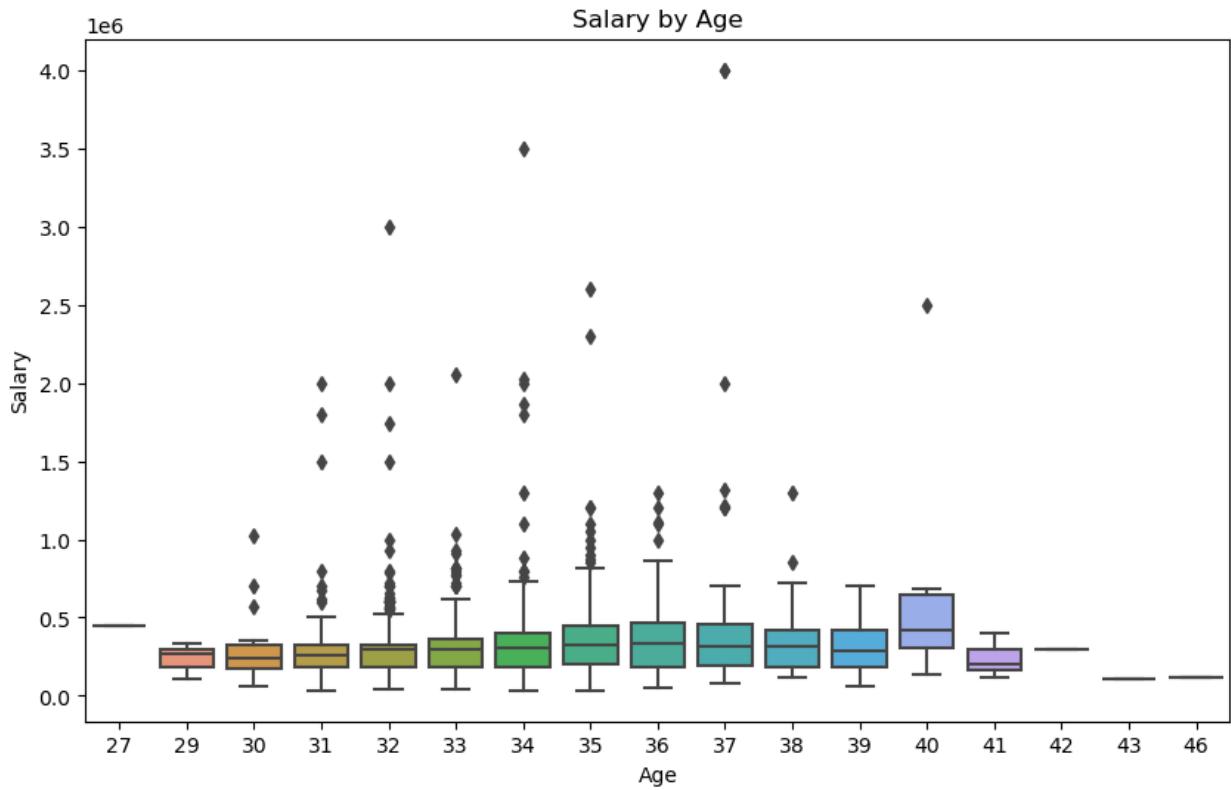
```
grouped_vals = vals.groupby(grouper)
```



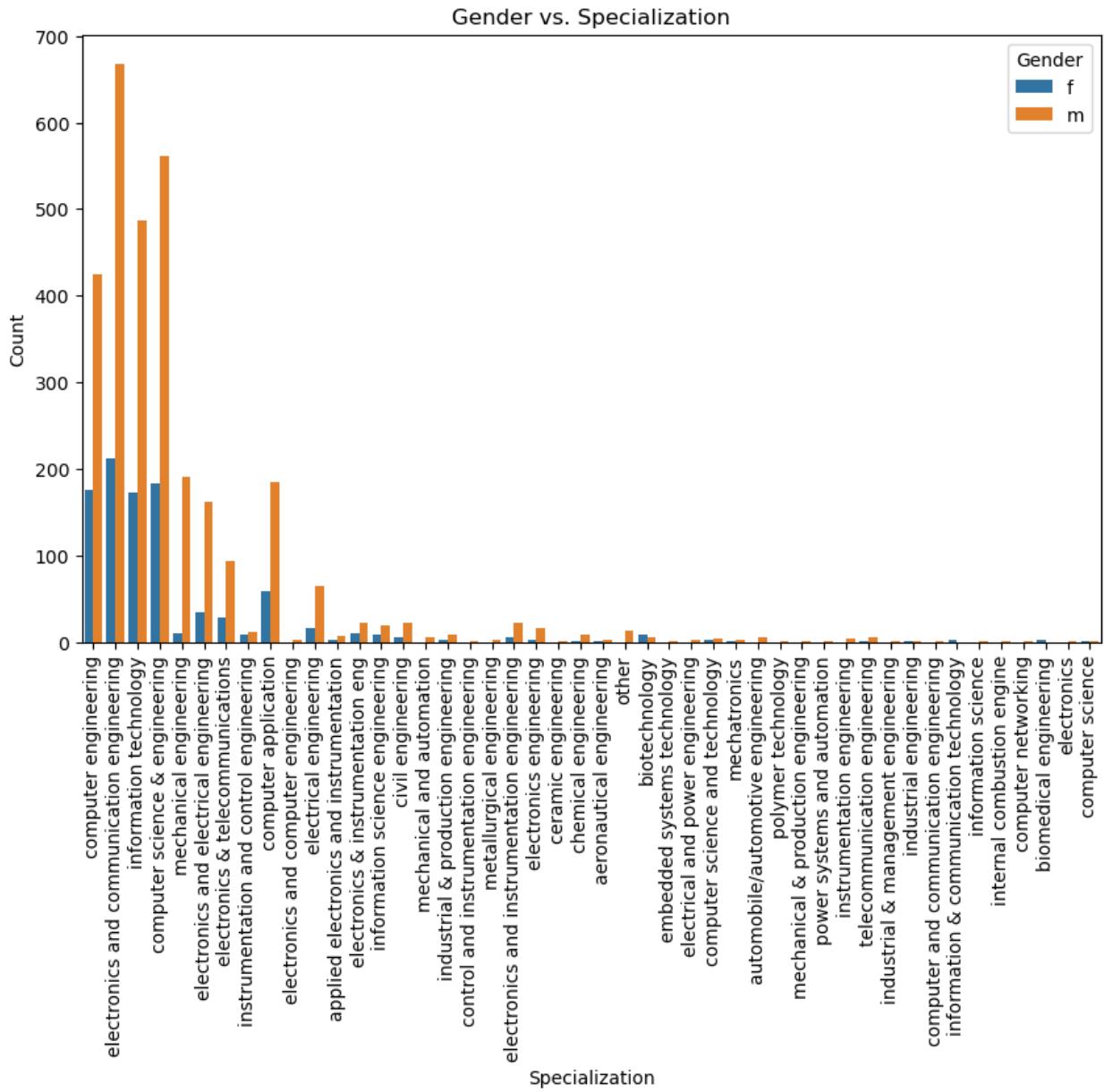
```
#Salary vs Gender  
plt.figure(figsize=(10, 6))  
sns.boxplot(x='Gender', y='Salary', data=df)  
plt.title('Salary vs. Gender')  
plt.xlabel('Gender')  
plt.ylabel('Salary')  
plt.show()
```



```
#salary vs Age
plt.figure(figsize=(10, 6))
sns.boxplot(x='Age', y='Salary', data=df)
plt.title('Salary by Age')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```



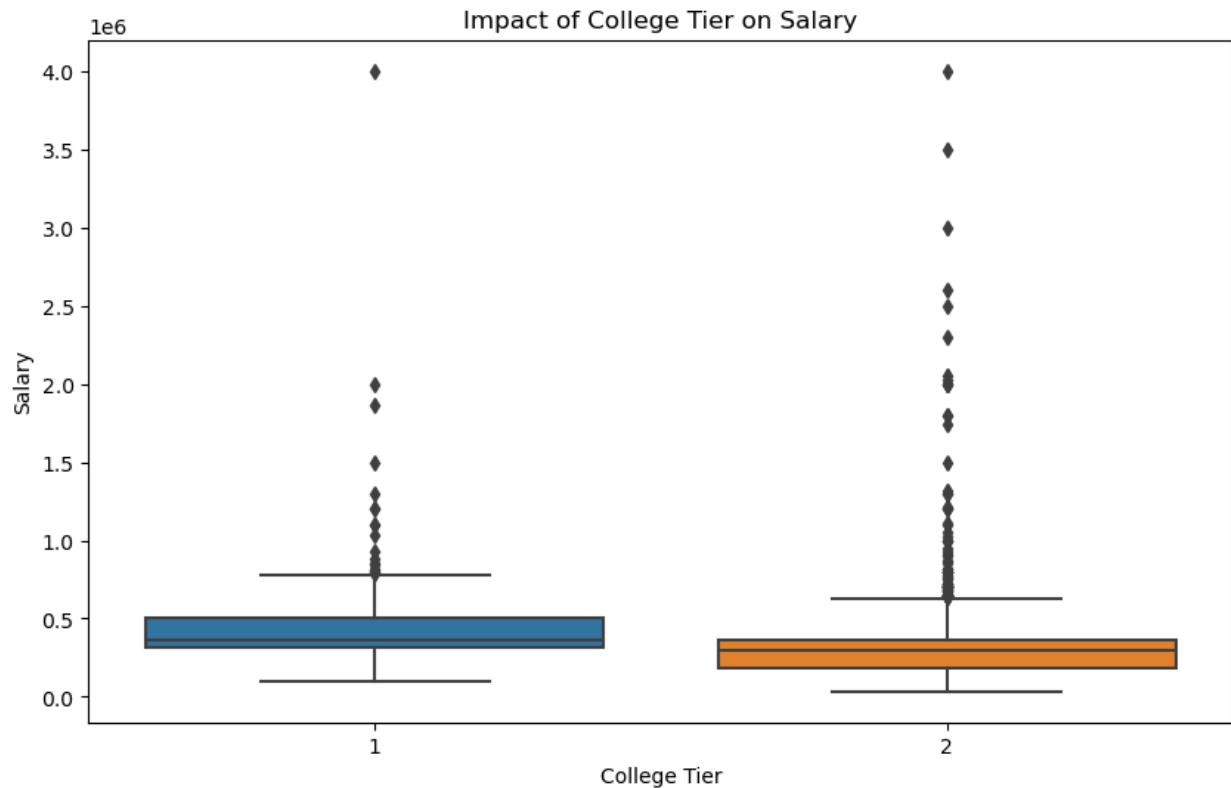
```
#Gender vs Specialization
plt.figure(figsize=(10, 6))
sns.countplot(x='Specialization', hue='Gender', data=df)
plt.title('Gender vs. Specialization')
plt.xlabel('Specialization')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```



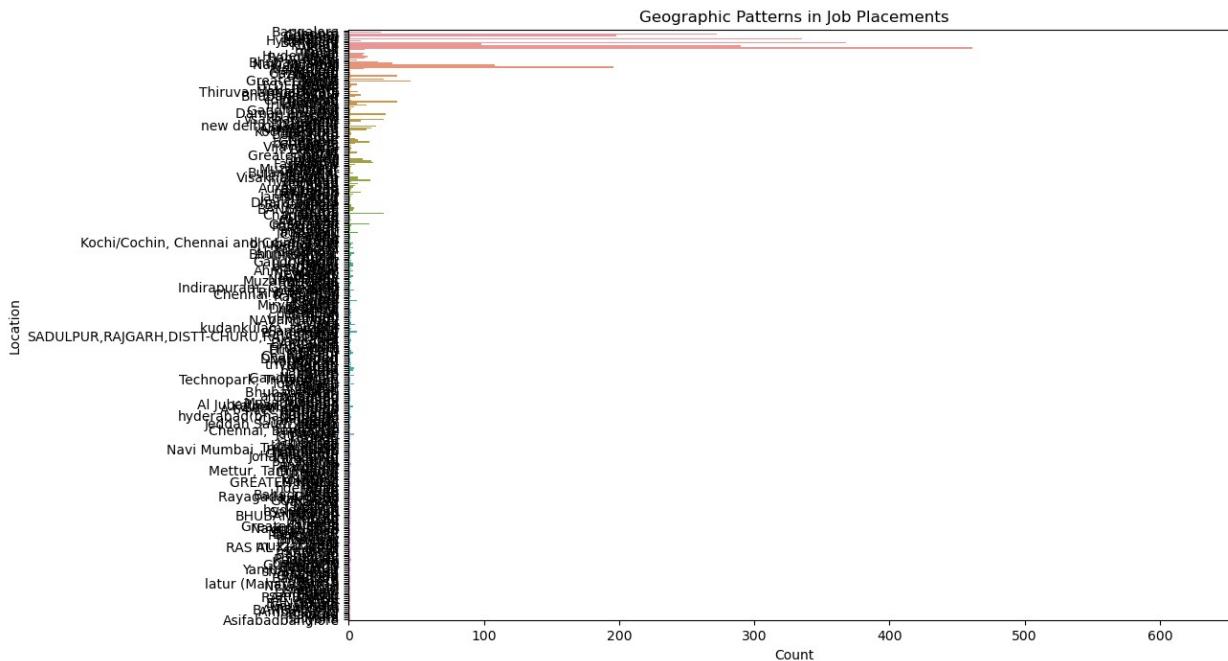
```
#impact of college tier on salary
plt.figure(figsize=(10, 6))
sns.boxplot(x='CollegeTier', y='Salary', data=df)
plt.title('Impact of College Tier on Salary')
plt.xlabel('College Tier')
plt.ylabel('Salary')
plt.show()
```

C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\categorical.py:641:  
FutureWarning: The default of observed=False is deprecated and will be  
changed to True in a future version of pandas. Pass observed=False to

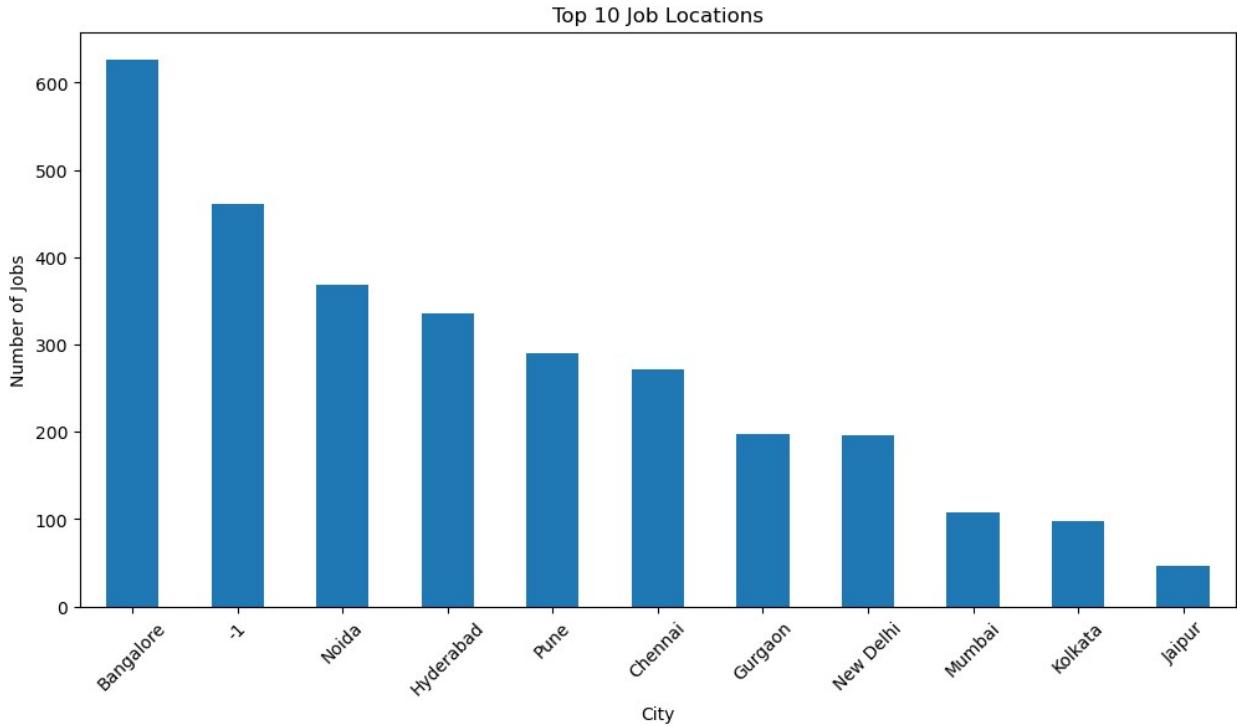
```
retain current behavior or observed=True to adopt the future default  
and silence this warning.  
grouped_vals = vals.groupby(grouper)
```



```
plt.figure(figsize=(12, 8))  
sns.countplot(y='JobCity', data=df)  
plt.title('Geographic Patterns in Job Placements')  
plt.xlabel('Count')  
plt.ylabel('Location')  
plt.show()
```



```
#since there are lots of values lets try to plot only top few locations
# Geographic patterns in job placements
job_locations = df['JobCity'].value_counts().head(11)
plt.figure(figsize=(12, 6))
job_locations.plot(kind='bar')
plt.title('Top 10 Job Locations')
plt.xlabel('City')
plt.ylabel('Number of Jobs')
plt.xticks(rotation=45)
plt.show()
```

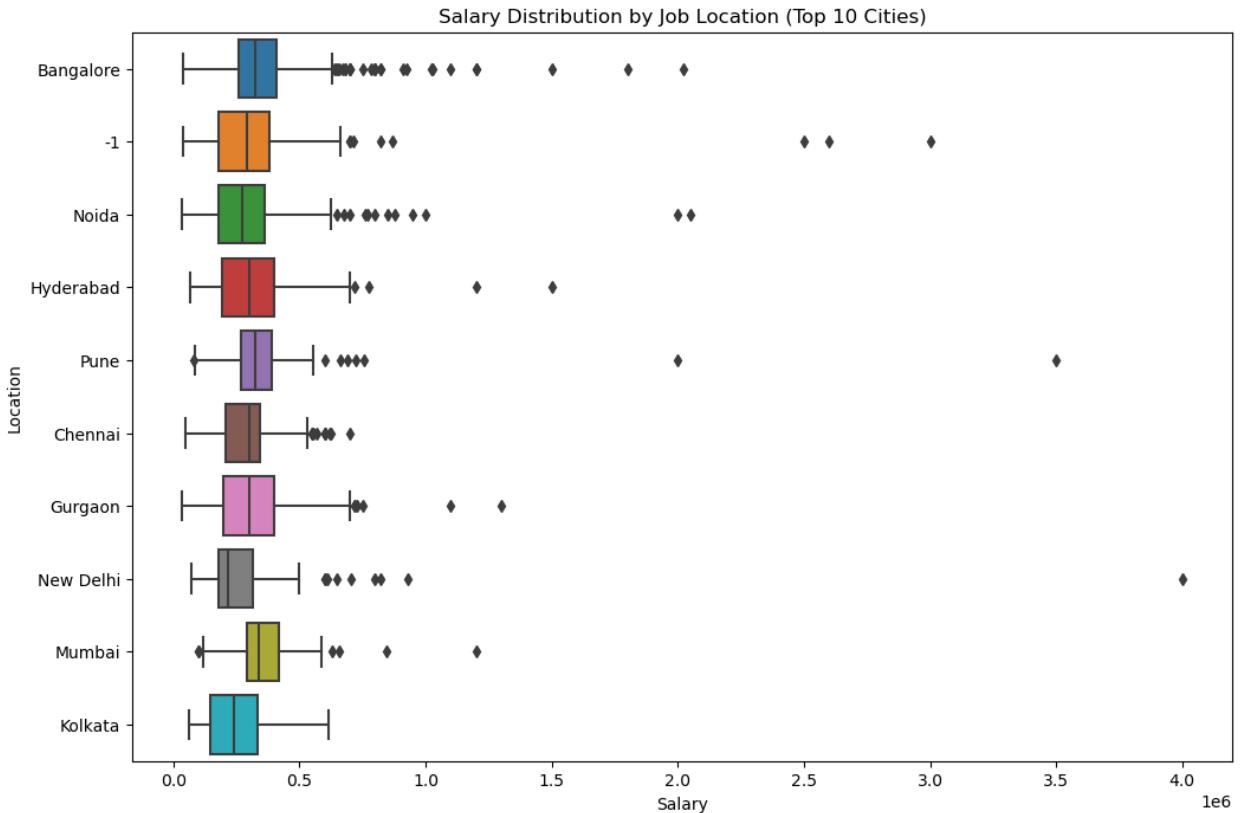


```
#job location vs salary
plt.figure(figsize=(12, 8))

# Get the top 10 job locations by count
top_10_cities = df['JobCity'].value_counts().nlargest(10).index

# Filter the DataFrame to include only rows where JobCity is in the top 10
top_10_df = df[df['JobCity'].isin(top_10_cities)]

# Plot the boxplot for salary distribution in the top 10 job locations
sns.boxplot(y='JobCity', x='Salary', data=top_10_df,
order=top_10_cities)
plt.title('Salary Distribution by Job Location (Top 10 Cities)')
plt.xlabel('Salary')
plt.ylabel('Location')
plt.show()
```



```

# Salary trends by graduation year

df['GraduationYear'].replace("0", np.nan, inplace=True)

# Drop rows where GraduationYear is NaN or invalid
df['GraduationYear'] = pd.to_datetime(df['GraduationYear'],
format='%Y', errors='coerce')

# Group by GraduationYear and calculate average salary
yearly_salary = df.groupby(df['GraduationYear'].dt.year)
['Salary'].mean()

# Plot the average salary trend
plt.figure(figsize=(12, 6))
yearly_salary.plot(kind='line', marker='o')

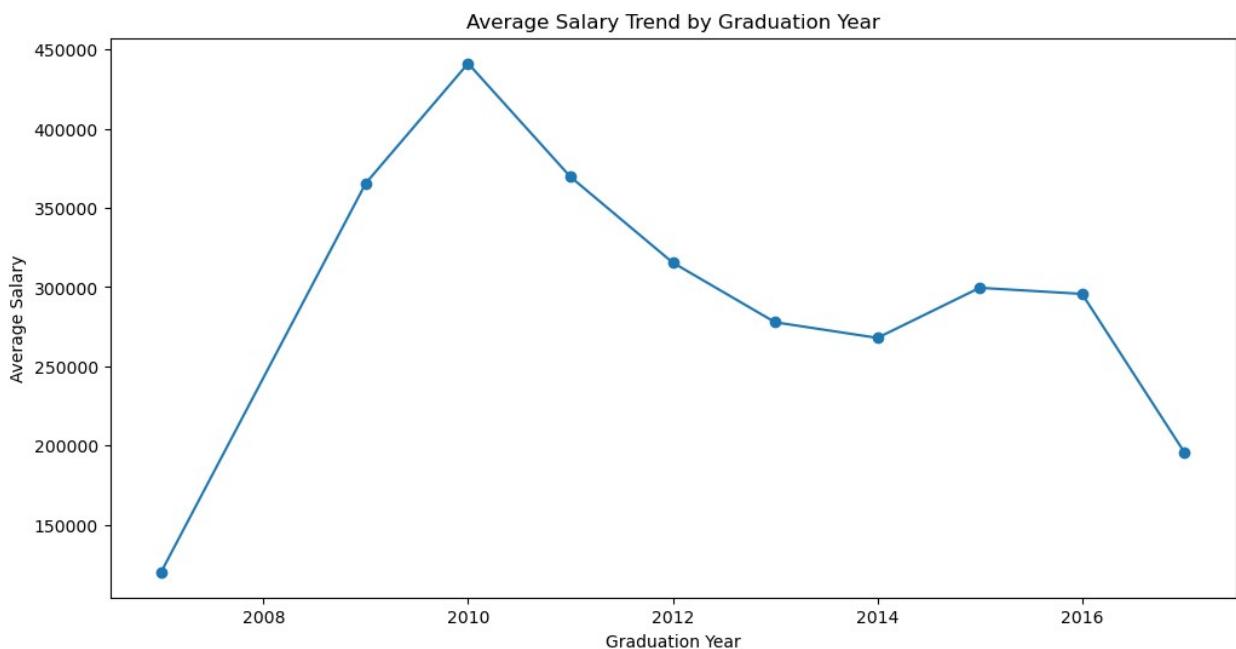
plt.title('Average Salary Trend by Graduation Year')
plt.xlabel('Graduation Year')
plt.ylabel('Average Salary')
plt.show()

```

```
C:\Users\vamsh\AppData\Local\Temp\ipykernel_4332\701605348.py:3:  
FutureWarning: A value is trying to be set on a copy of a DataFrame or  
Series through chained assignment using an inplace method.  
The behavior will change in pandas 3.0. This inplace method will never  
work because the intermediate object on which we are setting values  
always behaves as a copy.
```

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

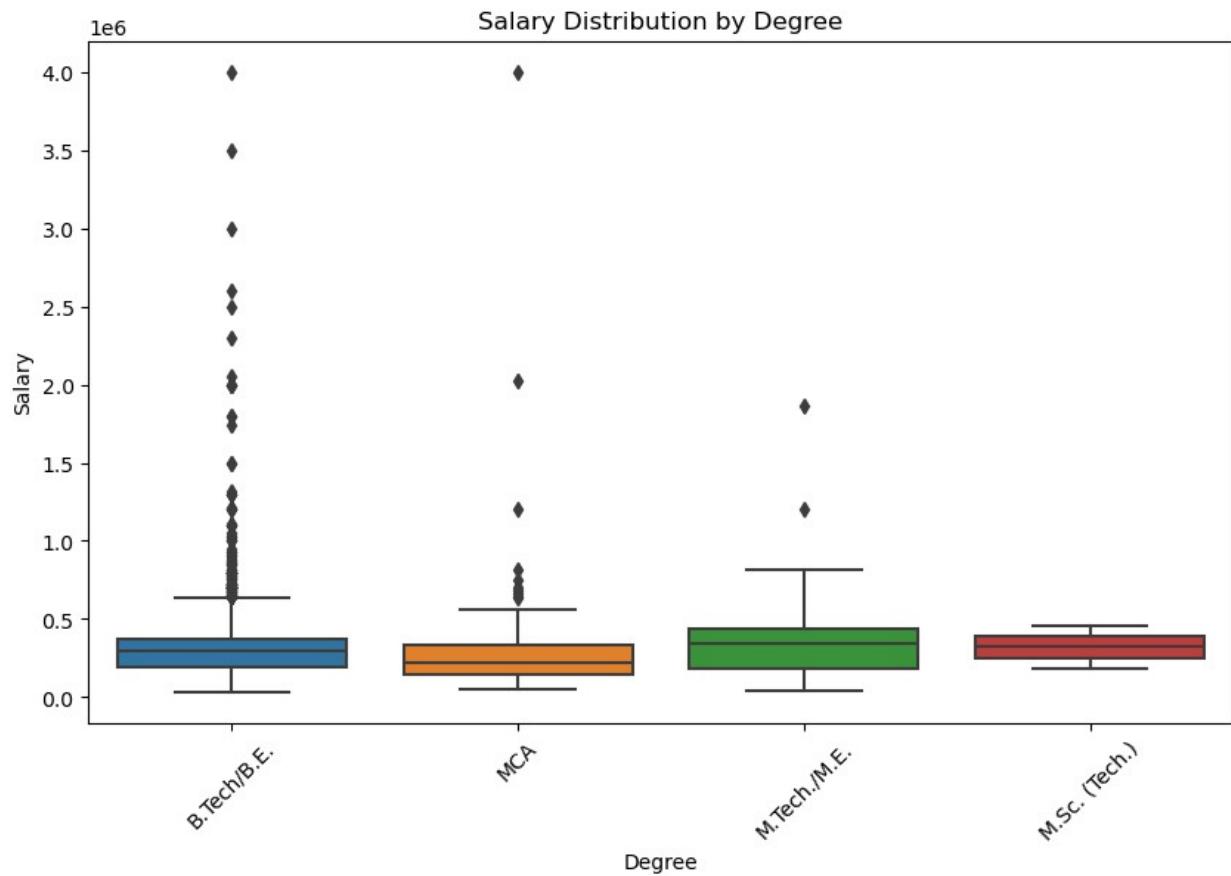
```
df['GraduationYear'].replace("0", np.nan, inplace=True)
```



```
# Groupby to calculate the average and median salary for each degree type  
degree_salary_stats = df.groupby('Degree')['Salary'].agg(['mean', 'median']).reset_index()  
  
print(degree_salary_stats)
```

	Degree	mean	median
0	B.Tech/B.E.	308622.702703	300000.0
1	M.Sc. (Tech.)	320000.000000	320000.0
2	M.Tech./M.E.	366132.075472	340000.0
3	MCA	280802.469136	220000.0

```
# Assuming 'Degree' and 'Salary' are the columns in your DataFrame
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Degree', y='Salary')
plt.title('Salary Distribution by Degree')
plt.xticks(rotation=45)
plt.show()
```



```
#Research question1
# Trying to understand Salary for Computer Science graduates in
specific roles

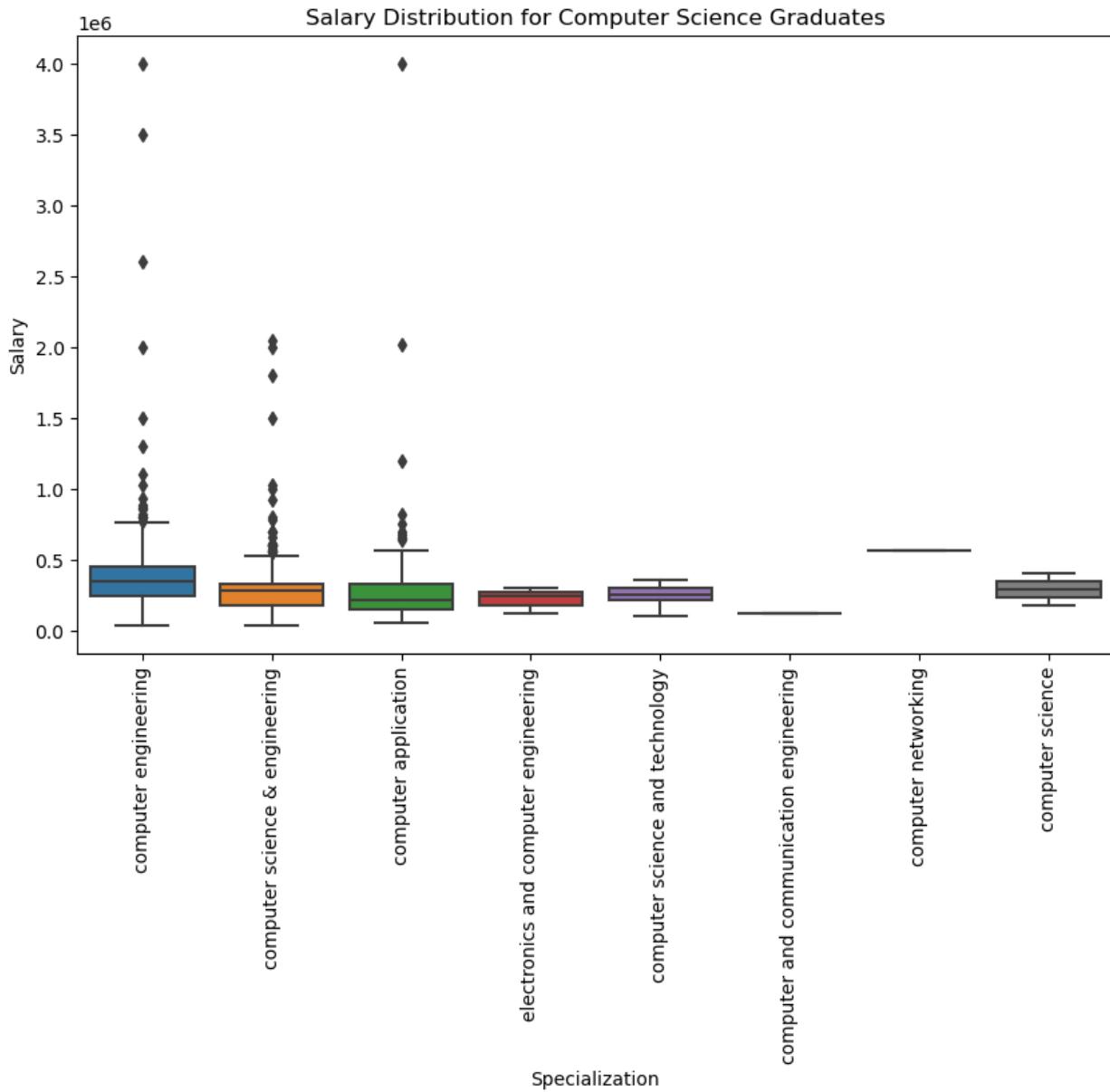
# Filter data for Computer Science graduates
cs_grads = df[df['Specialization'].str.contains('computer', na=False)]
```

```
# Calculate average and median salary
cs_mean_salary = cs_grads['Salary'].mean()
cs_median_salary = cs_grads['Salary'].median()

print(f"Average Salary for Computer Science Graduates:
{cs_mean_salary}")
print(f"Median Salary for Computer Science Graduates:
{cs_median_salary}")

# Box plot to visualize salary distribution
plt.figure(figsize=(10, 6))
sns.boxplot(x='Specialization', y='Salary', data=cs_grads)
plt.title('Salary Distribution for Computer Science Graduates')
plt.xticks(rotation=90)
plt.show()
```

```
Average Salary for Computer Science Graduates: 313985.0093691443
Median Salary for Computer Science Graduates: 300000.0
```



```
#Research question1: Now the actual research question 1 part;
# Convert Salary to lakhs (assuming salary is annual and in INR)
df1['Salary_Lakhs'] = df1['Salary'] / 100000

# 1. Analyzing the salary claim for fresh CS graduates
cs_jobs = ['programmer analyst', 'software engineer', 'system
engineer', 'associate software engineer']
fresh_cs_grads = df1[(df1['Specialization'].str.contains('computer',
case=False, na=False)) &
                     (df1['Designation'].str.lower().isin(cs_jobs)) &
                     (df1['GraduationYear'] >= 2013)] # Assuming 2013
```

*and later are fresh graduates*

```
avg_salary = fresh_cs_grads['Salary_Lakhs'].mean()
median_salary = fresh_cs_grads['Salary_Lakhs'].median()

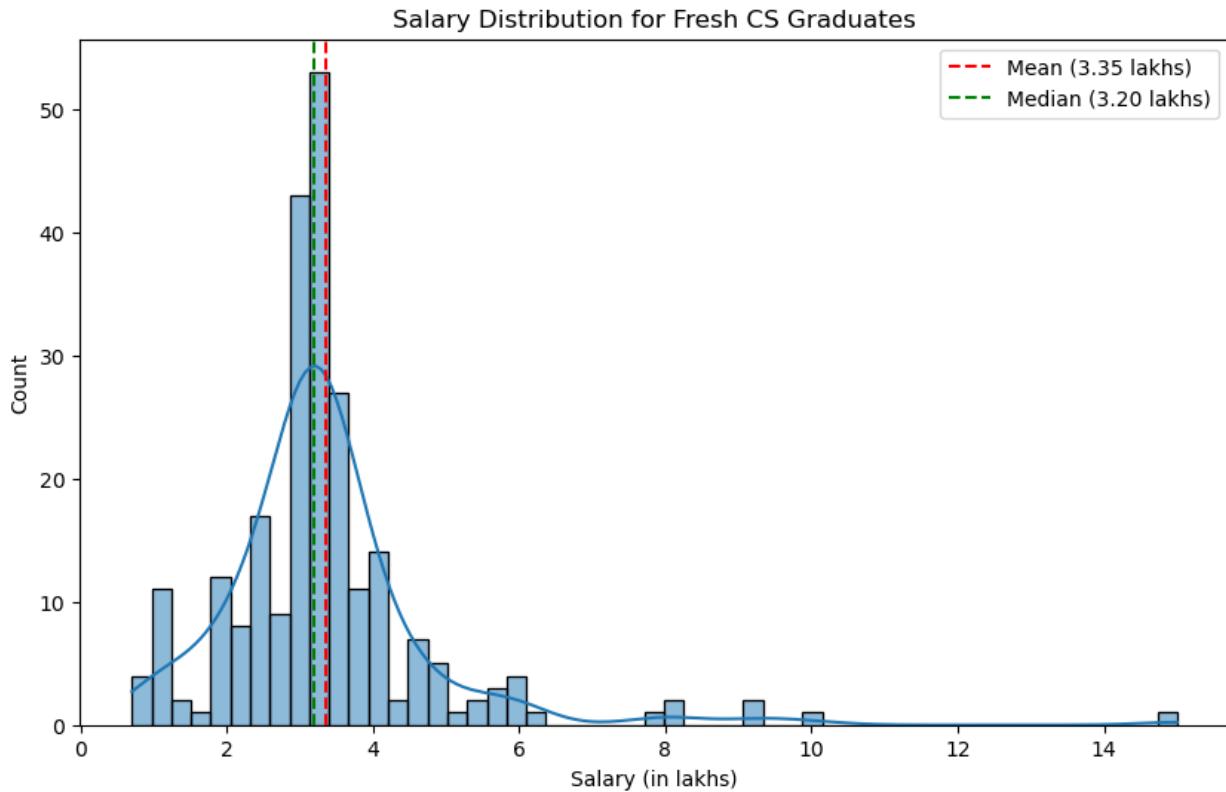
print(f"Average salary for fresh CS graduates: {avg_salary:.2f} lakhs")
print(f"Median salary for fresh CS graduates: {median_salary:.2f} lakhs")

# Visualization of salary distribution
plt.figure(figsize=(10, 6))
sns.histplot(fresh_cs_grads['Salary_Lakhs'], kde=True)
plt.title('Salary Distribution for Fresh CS Graduates')
plt.xlabel('Salary (in lakhs)')
plt.ylabel('Count')
plt.axvline(avg_salary, color='r', linestyle='--', label=f'Mean ({avg_salary:.2f} lakhs)')
plt.axvline(median_salary, color='g', linestyle='--', label=f'Median ({median_salary:.2f} lakhs)')
plt.legend()
plt.show()

Average salary for fresh CS graduates: 3.35 lakhs
Median salary for fresh CS graduates: 3.20 lakhs

C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.

with pd.option_context('mode.use_inf_as_na', True):
```



```
# 2. Relationship between gender and specialization
gender_spec = pd.crosstab(df['Gender'], df['Specialization'])
chi2, p_value, dof, expected = stats.chi2_contingency(gender_spec)

print(f"\nChi-square test results:")
print(f"Chi-square statistic: {chi2:.2f}")
print(f"p-value: {p_value:.4f}")

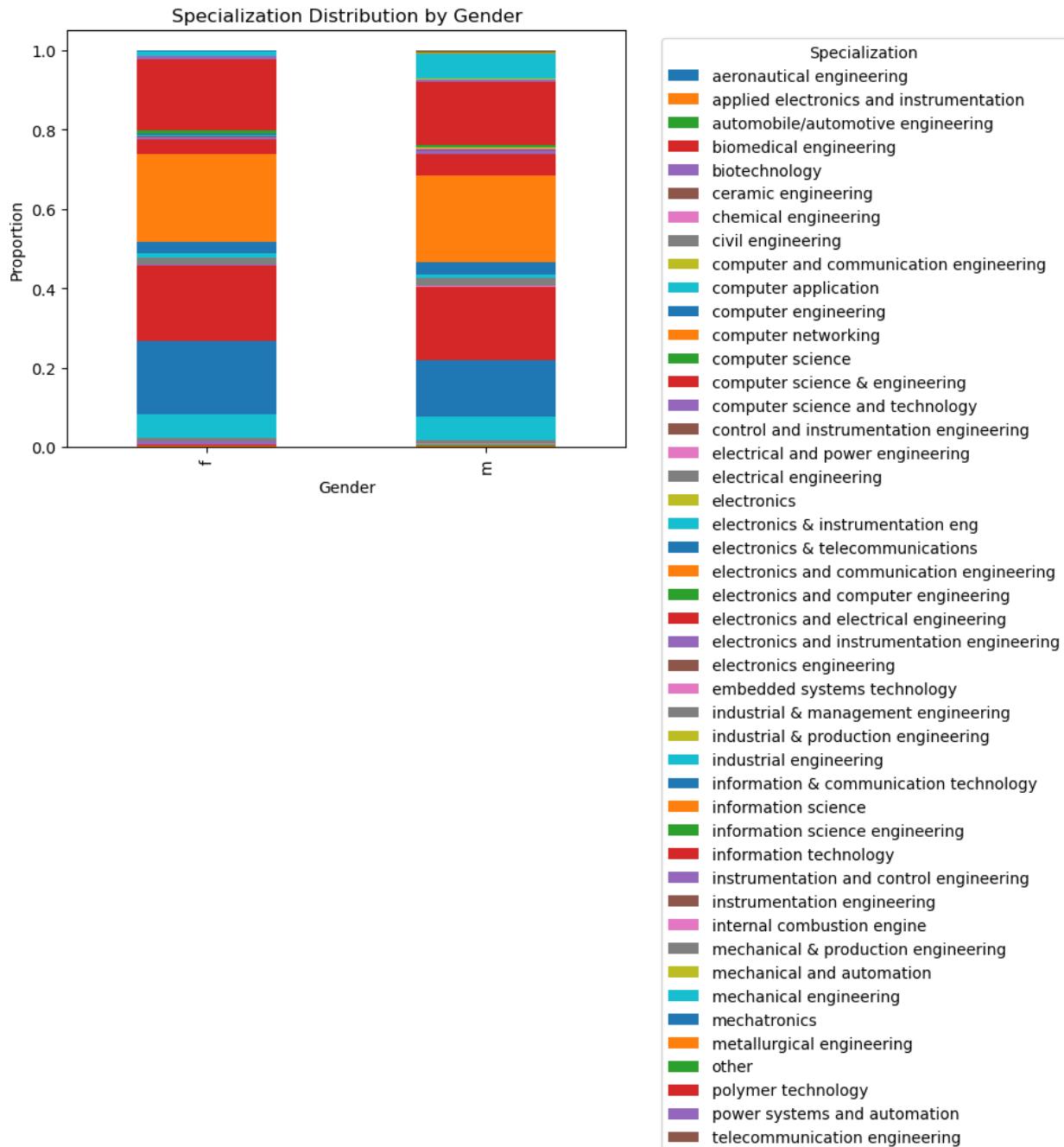
# Visualization of gender and specialization relationship
plt.figure(figsize=(12, 6))
gender_spec_norm = gender_spec.div(gender_spec.sum(axis=1), axis=0)
gender_spec_norm.plot(kind='bar', stacked=True)
plt.title('Specialization Distribution by Gender')
plt.xlabel('Gender')
plt.ylabel('Proportion')
plt.legend(title='Specialization', bbox_to_anchor=(1.05, 1),
loc='upper left')
```

```
plt.tight_layout()
plt.show()

Chi-square test results:
Chi-square statistic: 104.47
p-value: 0.0000

C:\Users\vamsh\AppData\Local\Temp\ipykernel_4332\2265005776.py:17:
UserWarning: Tight layout not applied. The bottom and top margins
cannot be made large enough to accommodate all axes decorations.
plt.tight_layout()

<Figure size 1200x600 with 0 Axes>
```



#Research question2

#since the above stacked bar plot is a bit clumsy i am using grouped bar plot for clear view, although the data is same in both of them

# Create a contingency table

```
contingency_table = pd.crosstab(df['Gender'], df['Specialization'])
```

# Calculate percentages

```

percentage_table =
contingency_table.div(contingency_table.sum(axis=1), axis=0) * 100

# Get top 5 specializations
top_10 = percentage_table.mean().nlargest(10).index

# Create a new dataframe with top 5 and 'Other'
top_10_df = percentage_table[top_10].copy()
top_10_df['Other'] = 100 - top_10_df.sum(axis=1)

# Melt the dataframe for easier plotting
melted_df = top_10_df.reset_index().melt(id_vars='Gender',
var_name='Specialization', value_name='Percentage')

# Create a grouped bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x='Specialization', y='Percentage', hue='Gender',
data=melted_df)

# Customize the plot
plt.title('Distribution of Top 10 Specializations by Gender')
plt.xlabel('Specialization')
plt.ylabel('Percentage')
plt.legend(title='Gender')
plt.xticks(rotation=45, ha='right')

# Add percentage labels on the bars
for i in plt.gca().containers:
    plt.gca().bar_label(i, fmt='%.2f%%',
label_type='center', rotation=45)

plt.tight_layout()

# Save the plot
plt.savefig('gender_specialization_distribution_top5.png')

# Display some statistics
print("Top 10 Specializations:")
print(top_10_df)

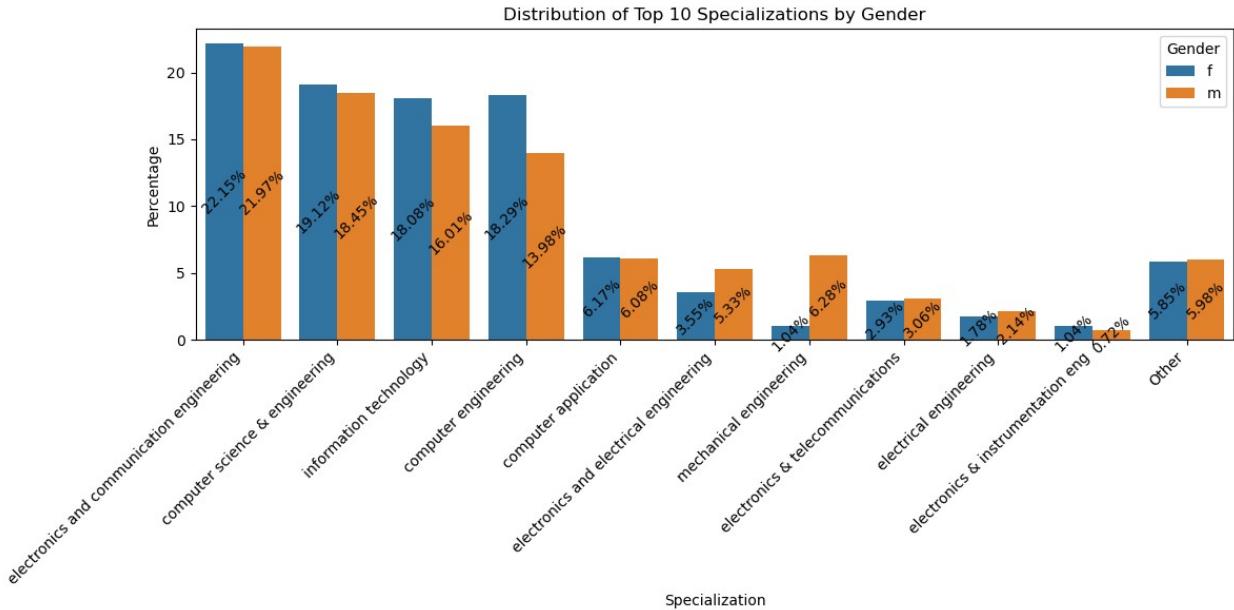
# Chi-square test for independence
from scipy.stats import chi2_contingency

chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"\nChi-square statistic: {chi2}")
print(f"p-value: {p_value}")

Top 10 Specializations:
Specialization    electronics and communication engineering \
Gender
f

```

m		21.966458
Specialization computer science & engineering information technology \ Gender		
f	19.122257	18.077325
m	18.447879	16.014469
Specialization computer engineering computer application \ Gender		
f	18.286311	6.165099
m	13.975666	6.083525
Specialization electronics and electrical engineering \ Gender		
f		3.552769
m		5.327195
Specialization mechanical engineering electronics & telecommunications \ Gender		
f	1.044932	
2.925810		
m	6.280829	
3.058205		
Specialization electrical engineering electronics & instrumentation eng \ Gender		
f	1.776385	
1.044932		
m	2.137455	
0.723446		
Specialization Other		
Gender		
f	5.851620	
m	5.984873	
Chi-square statistic: 104.46891913608455		
p-value: 1.2453868176976918e-06		

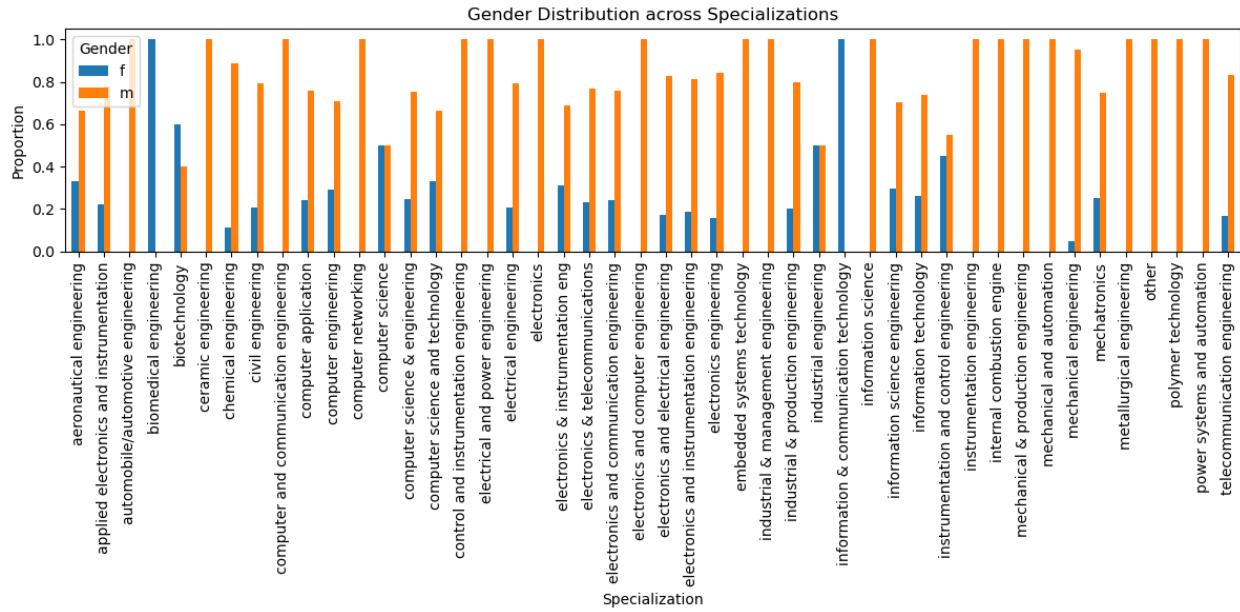


```
#here i am trying to present the gender distribution across each specialization
```

```
# Creating a grouped bar plot with specializations on the x-axis
gender_spec = pd.crosstab(df['Specialization'], df['Gender'],
normalize='index')

# Plotting the grouped bar plot
ax = gender_spec.plot(kind='bar', figsize=(12, 6))

# Set plot labels and rotate x-axis for readability
plt.title('Gender Distribution across Specializations')
plt.xlabel('Specialization')
plt.ylabel('Proportion')
plt.xticks(rotation=90) # Rotate specialization labels on x-axis
plt.tight_layout()
plt.show()
```



#some additional plottings (but already covered above, just different styles of plotting that's it.)

```
import matplotlib.pyplot as plt
import seaborn as sns

# List of numerical columns (excluding Salary)
numerical_columns = df.select_dtypes(include=['float64',
'int64']).columns.tolist()
numerical_columns.remove('Salary') # Remove Salary from the list

# Set up the plot
plt.figure(figsize=(15, 15))
for i, col in enumerate(numerical_columns, 1):
    plt.subplot(4, 4, i) # Adjust the layout based on the number of
    numerical columns
```

```
sns.scatterplot(data=df, x=col, y='Salary')
plt.title(f'Salary vs {col}')
plt.xlabel(col)
plt.ylabel('Salary')

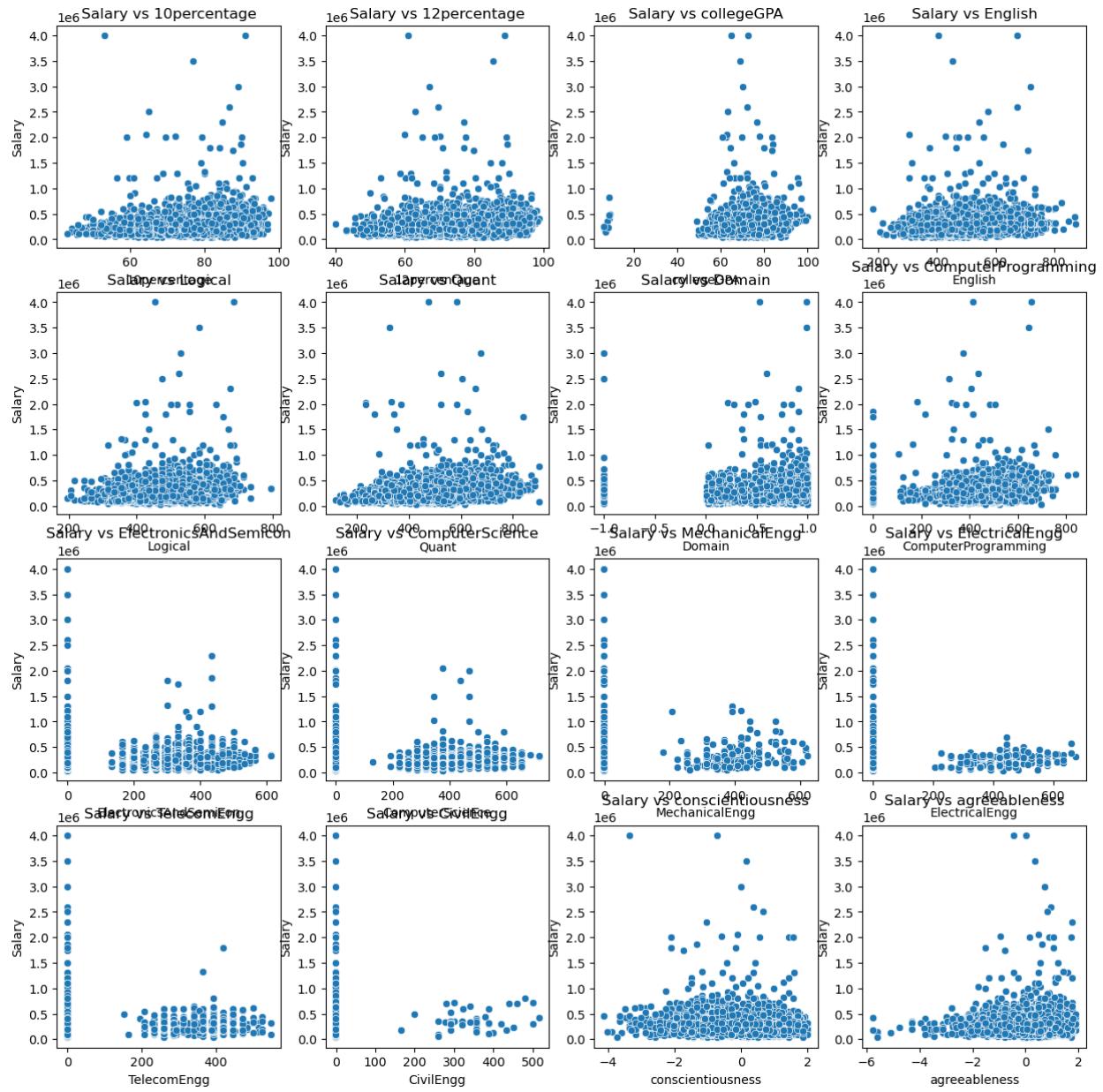
plt.tight_layout()
plt.show()

-----
-----
ValueError                                     Traceback (most recent call
last)
Cell In[59], line 11
    9 plt.figure(figsize=(15, 15))
   10 for i, col in enumerate(numerical_columns, 1):
--> 11     plt.subplot(4, 4, i) # Adjust the layout based on the
number of numerical columns
   12     sns.scatterplot(data=df, x=col, y='Salary')
   13     plt.title(f'Salary vs {col}')

File ~\anaconda3\Lib\site-packages\matplotlib\pyplot.py:1425, in
subplot(*args, **kwargs)
 1422 fig = gcf()
 1424 # First, search for an existing subplot with a matching spec.
-> 1425 key = SubplotSpec._from_subplot_args(fig, args)
 1427 for ax in fig.axes:
 1428     # If we found an Axes at the position, we can re-use it if
the user passed no
 1429     # kwargs or if the axes class and kwargs are identical.
 1430     if (ax.get_subplotspec() == key
 1431         and (kwargs == {}
 1432             or (ax._projection_init
 1433                 ==
fig._process_projection_requirements(**kwargs)))))

File ~\anaconda3\Lib\site-packages\matplotlib\gridspec.py:599, in
SubplotSpec._from_subplot_args(figure, args)
 597 else:
 598     if not isinstance(num, Integral) or num < 1 or num >
rows*cols:
--> 599         raise ValueError(
 600             f"num must be an integer with 1 <= num <=
{rows*cols}, "
 601             f"not {num!r}")
 602         )
 603     i = j = num
 604 return gs[i-1:j]

ValueError: num must be an integer with 1 <= num <= 16, not 17
```

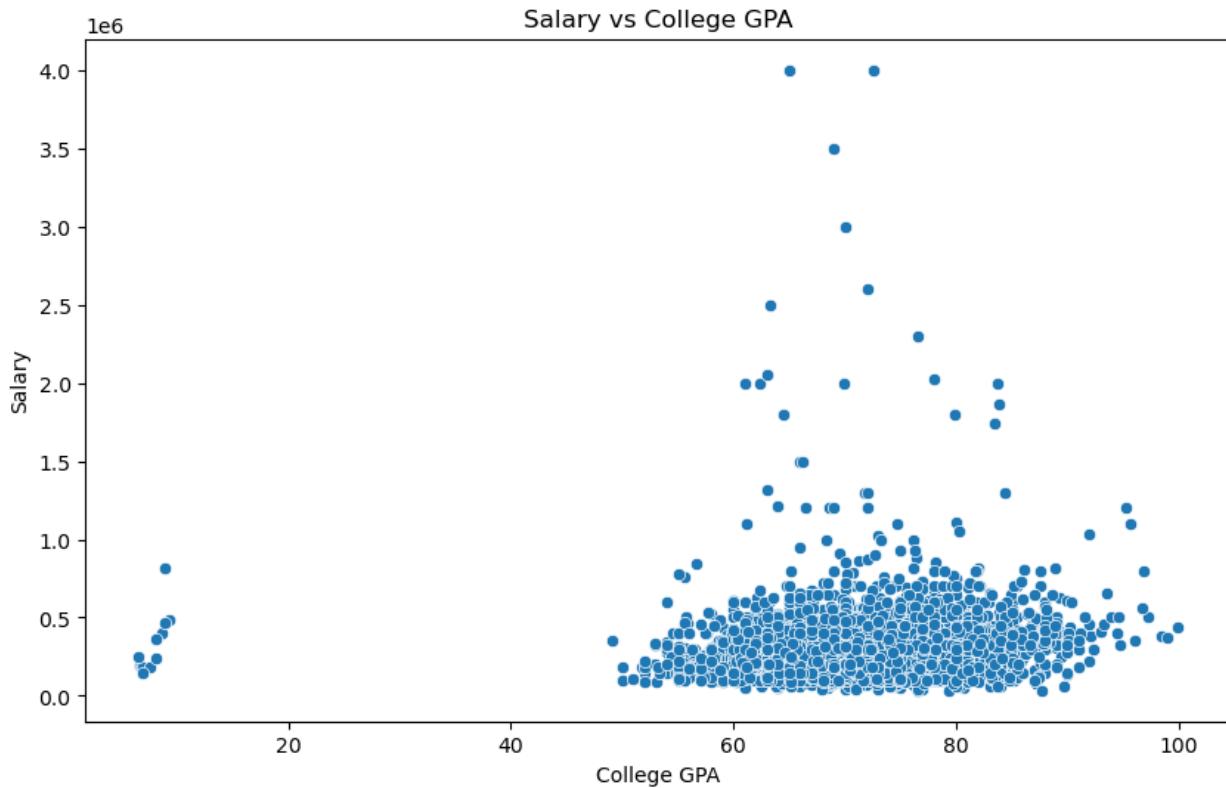


```

import seaborn as sns
import matplotlib.pyplot as plt

# Scatter plot for Salary vs College GPA
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='collegeGPA', y='Salary')
plt.title('Salary vs College GPA')
plt.xlabel('College GPA')
plt.ylabel('Salary')
plt.show()

```



```

plt.figure(figsize=(10, 6))
sns.swarmplot(x='Specialization', y='Salary', data=df)
plt.title('Salary Distribution by Specialization')
plt.xlabel('Specialization')
plt.ylabel('Salary')
plt.xticks(rotation=90)
plt.show()

C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
    with pd.option_context('mode.use_inf_as_na', True):
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
    with pd.option_context('mode.use_inf_as_na', True):
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1075:
FutureWarning: When grouping with a length-1 list-like, you will need
to pass a length-1 tuple to get_group in a future version of pandas.
Pass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 94.7% of the points cannot be

```

```
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 97.3% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 96.2% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 96.6% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 90.5% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 91.8% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 86.8% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 40.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 81.7% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 33.3% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
```

```
categorical.py:3544: UserWarning: 62.5% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 66.7% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 62.1% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 20.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 30.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 48.1% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 73.7% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 38.5% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 46.7% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 16.7% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
```

```
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 25.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 60.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 94.3% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 97.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 96.1% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 96.5% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 89.1% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 86.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 35.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 91.4% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
```

```
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 80.5% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 68.4% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 22.2% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\vamsh\anaconda3\Lib\site-packages\seaborn\
categorical.py:3544: UserWarning: 50.0% of the points cannot be
placed; you may want to decrease the size of the markers or use
stripplot.
    warnings.warn(msg, UserWarning)
```

