D

November 13, 2023

1 Assignment [CSE445]

1.1 Data : Investigating the career trajectory of NSU Alumni (Responses)

1.1.1 Tasks:

1. Conduct the cleaning on the data if there is any:
    1. Empty cells
    2. Data in wrong format
    3. Wrong data
    4. Duplicates
2. Organize the data into a JSON file and then read this JSON data.
3. Do you think scaling is necessary in this case? Justify
4. Develop necessary hypothesis that might support your understanding on the developed datasets
5. Calculate the relationship between each column in your data set. The find out the most valuable features that will support you hypothesis and you will do your further analysis and data mining task. For this use seaborn python library to visualize the correlation matrix

[1]:
```
# %pip install nbconvert
```

[3]:
```
# !jupyter nbconvert --to pdf D.ipynb
```

[3]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
```

[4]:
```
data = pd.read_csv("data.csv")
data.head()
```

[4]:
```
             Timestamp                Email Address  Gender  \
0                  NaN                          NaN     NaN
1  8/24/2023 23:03:30      dewanaminul12@gmail.com    Male
2   8/25/2023 3:21:54      tanzilshahria@gmail.com    Male
```

```
3   8/25/2023 3:27:49      progga.tasnim@gmail.com  Female
4  8/26/2023 11:45:09  saif.ahmed02@northsouth.edu    Male

   Cumulative Grade Point Average(CGPA)  \
0                                   NaN
1                          3 to < 3.5
2                            3.5 to 4
3                            3.5 to 4
4                            3.5 to 4

  Total year now since your graduation    \
0                                   NaN
1                                     1
2                                     4
3                                     3
4                                     8

                       Fields of research interest  Number of publications  \
0                                                NaN                     NaN
1  Internet of Things(IoT) and Embedded Systems, …                   1 - 5
2  Machine Learning, Artificial Intelligence, Nat…            More than 10
3                          Human-Computer interaction                   1 - 5
4  Machine Learning, Data Science/Data mining, Ar…                    6-10

  Were you a teaching assistant (TA) or research assistant(RA) during your time
in NSU?  \
0                                                NaN
1                                                 No
2                                                Yes
3                                                Yes
4                                                Yes

  Did you participate in any extra curricular activities while you were studying
in NSU?  \
0                                                NaN
1                                                Yes
2                                                Yes
3                                                Yes
4                                                Yes

  Were you interested in coding?  \
0                            NaN
1                            Yes
2                            Yes
3                            Yes
4                            Yes
```

```
  Did you participate in any competitive programming?   \
0                                                  NaN
1                                                  Yes
2                                                  Yes
3                                                  Yes
4                                                   No

                     Programming languages you know  \
0                                                  NaN
1                                       C, C++, HTML
2                               C, Java, Python, C++, R
3  C, Python, PHP, C++, HTML, CSS, JavaScript, R,…
4                                    C, Python, R, SQL

              Frameworks you know  \
0                             NaN
1                  Not interested
2                         Laravel
3  React, Django, Laravel, Node.js
4                          Django

  Did you start working as a software engineer after graduation?  \
0                                                  NaN
1                                                   No
2                                                   No
3                                                  Yes
4                                                   No

  What was the name of the company you worked for right after your graduation?
if your answer is not write N/A.  \
0                                                  NaN
1                  Venturous Professional Institute
2                          AST International Limited
3              Banglalink (Internship), Therap BD Ltd
4                              North South University

  What was your starting salary range?  \
0                                   NaN
1                        BDT 15K - 25K
2                          > 25K - 40K
3                                > 40K
4                        BDT 15K - 25K

  Did you have any Start-ups after graduation?  \
0                                          NaN
1                                           No
2                                           No
```

```
3                                                 No
4                                                 No

  Which country did you go to for higher studies?   \
0                                                NaN
1                  United States of America (USA)
2                  United States of America (USA)
3                  United States of America (USA)
4                                        Germany

  What was the name of the University you completed your higher studies from? if
answer is no write N/A.  \
0                                                  NaN
1                        Florida Atlantic University
2                  University of Wisconsin-Milwaukee
3                             Marquette University
4      University of Jena (Abbe School of Photonics)

  Where did you start working after completing your higher studies?
0                                                  NaN
1                                             Industry
2                                                  NaN
3                                                  NaN
4                              Educational institute
```

[5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 65 entries, 0 to 64
Data columns (total 20 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Timestamp
64 non-null     object
 1   Email Address
64 non-null     object
 2   Gender
64 non-null     object
 3    Cumulative Grade Point Average(CGPA)
64 non-null     object
 4   Total year now since your graduation
64 non-null     object
 5   Fields of research interest
62 non-null     object
 6   Number of publications
38 non-null     object
```

```
 7   Were you a teaching assistant (TA) or research assistant(RA) during your
time in NSU?                          64 non-null     object
 8   Did you participate in any extra curricular activities while you were
studying in NSU?                        64 non-null     object
 9   Were you interested in coding?
64 non-null     object
 10  Did you participate in any competitive programming?
64 non-null     object
 11  Programming languages you know
64 non-null     object
 12  Frameworks you know
64 non-null     object
 13  Did you start working as a software engineer after graduation?
64 non-null     object
 14  What was the name of the company you worked for right after your
graduation? if your answer is not write N/A.  39 non-null     object
 15  What was your starting salary range?
42 non-null     object
 16  Did you have any Start-ups after graduation?
64 non-null     object
 17  Which country did you go to for higher studies?
63 non-null     object
 18  What was the name of the University you completed your higher studies from?
if answer is no write N/A.          14 non-null     object
 19  Where did you start working after completing your higher studies?
29 non-null     object
dtypes: object(20)
memory usage: 10.3+ KB
```

We can see the data has 65 rows and 20 columns

And the Data types are of object types

Thus, it tells us that the dataset contains a lot of Categorical values

[6]:
```
print(f"Rows: {data.shape[0]}")
print(f"Columns: {data.shape[1]}")
```

```
Rows: 65
Columns: 20
```

[7]:
```
data.columns
```

[7]:
```
Index(['Timestamp', 'Email Address', 'Gender',
       ' Cumulative Grade Point Average(CGPA)',
       'Total year now since your graduation  ',
       'Fields of research interest ', 'Number of publications',
       'Were you a teaching assistant (TA) or research assistant(RA) during your
 time in NSU?',
```

```
       'Did you participate in any extra curricular activities while you were
studying in NSU?',
       'Were you interested in coding?',
       'Did you participate in any competitive programming? ',
       'Programming languages you know', 'Frameworks you know',
       'Did you start working as a software engineer after graduation?',
       'What was the name of the company you worked for right after your
graduation? if your answer is not write N/A.',
       'What was your starting salary range?',
       'Did you have any Start-ups after graduation?',
       'Which country did you go to for higher studies? ',
       'What was the name of the University you completed your higher studies
from? if answer is no write N/A.',
       'Where did you start working after completing your higher studies?'],
      dtype='object')
```

1.2 Renaming the columns and Removing the Whitespaces

[8]:
```
data.rename(columns={' Cumulative Grade Point Average(CGPA)': 'Cumulative Grade␣
 ↪Point Average(CGPA)'}, inplace=True)
data.rename(columns={'Total year now since your graduation  ': 'Total year now␣
 ↪since your graduation'}, inplace=True)
data.rename(columns={'Did you participate in any competitive programming? ':␣
 ↪'Did you participate in any competitive programming?'}, inplace=True)
data.rename(columns={'Which country did you go to for higher studies? ': 'Which␣
 ↪country did you go to for higher studies?'}, inplace=True)
data.rename(columns={'Fields of research interest ': 'Fields of research␣
 ↪interest'}, inplace=True)
```

2 Data Cleaning

2.0.1 We can even see that the data contains a lot of null values

3 Null Value Handling!

[9]:
```
data.isnull().sum()
```

[9]:
```
Timestamp
1
Email Address
1
Gender
1
Cumulative Grade Point Average(CGPA)
1
Total year now since your graduation
```

```
1
Fields of research interest
3
Number of publications
27
Were you a teaching assistant (TA) or research assistant(RA) during your time in
NSU?                              1
Did you participate in any extra curricular activities while you were studying
in NSU?                            1
Were you interested in coding?
1
Did you participate in any competitive programming?
1
Programming languages you know
1
Frameworks you know
1
Did you start working as a software engineer after graduation?
1
What was the name of the company you worked for right after your graduation? if
your answer is not write N/A.    26
What was your starting salary range?
23
Did you have any Start-ups after graduation?
1
Which country did you go to for higher studies?
2
What was the name of the University you completed your higher studies from? if
answer is no write N/A.           51
Where did you start working after completing your higher studies?
36
dtype: int64
```

3.0.1 The first row of the data contains all null value

3.0.2 Thus we shall deal with the null values now, By dropping the first row and repplacing the Null Values

[10]:
```
data.drop(0, inplace=True)
```

3.1 'Number of publications' Column Cleaning :

We can see it has nan value which doesn't represent anything in python so we replace it with 0

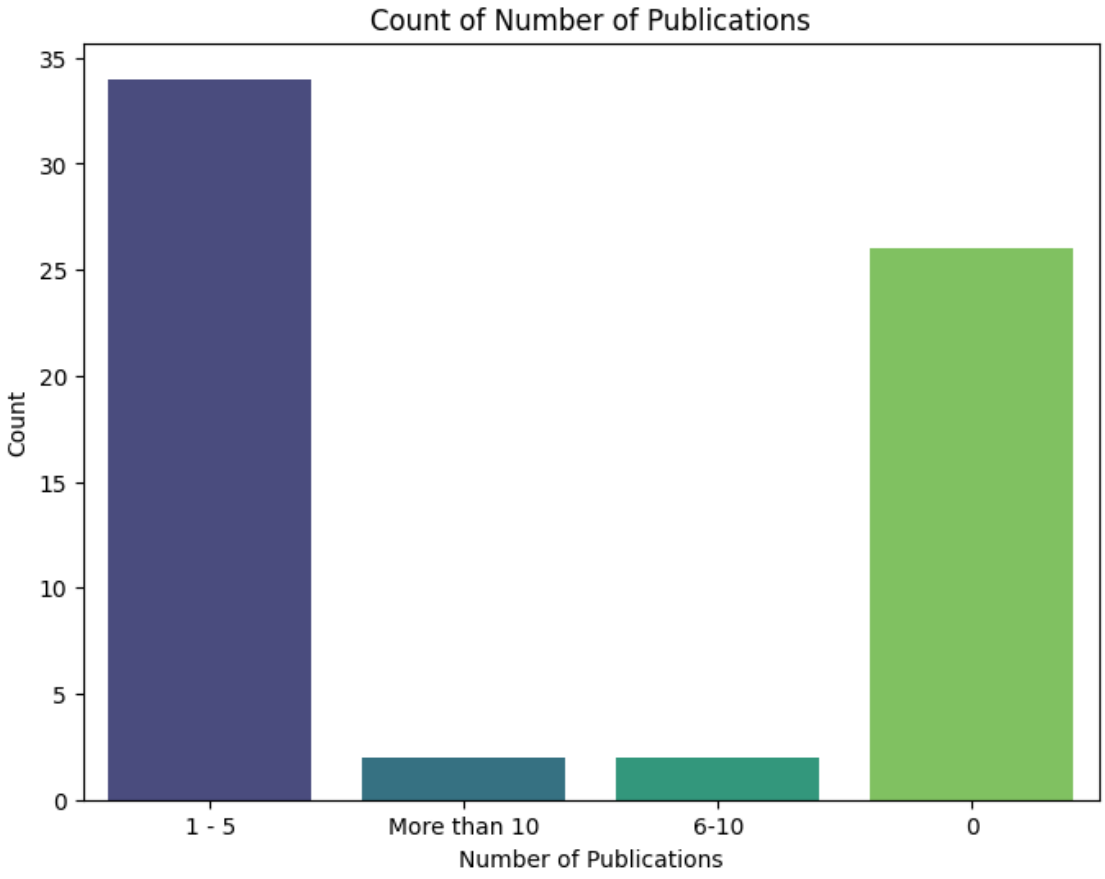
[11]:
```
data['Number of publications'].unique()
```

[11]:
```
array(['1 - 5', 'More than 10', '6-10', nan], dtype=object)
```

[12]:
```
data['Number of publications'].fillna('0', inplace=True)
data['Number of publications'].unique()
```

[12]:
```
array(['1 - 5', 'More than 10', '6-10', '0'], dtype=object)
```

[13]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data=data, x='Number of publications', palette='viridis')

plt.title('Count of Number of Publications')
plt.xlabel('Number of Publications')
plt.ylabel('Count')

plt.show()
```



[14]:
```
data['Number of publications'].unique()
```

[14]:
```
array(['1 - 5', 'More than 10', '6-10', '0'], dtype=object)
```

3.2 Now we will deal with 'Fields of research interest' column

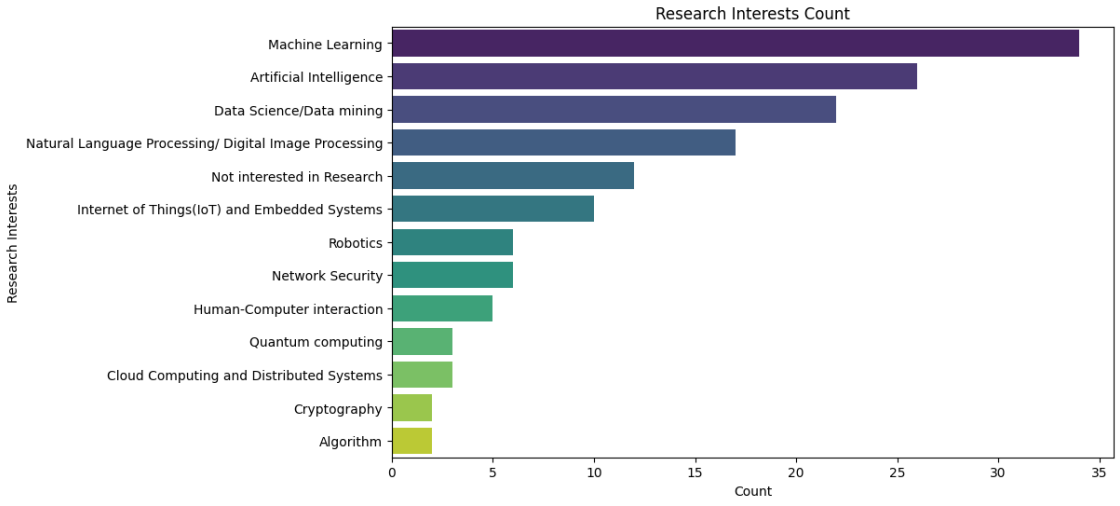
This column has 2 Null Values, First We will deal with these

As this data has a lot of unique values we are going to conduct One-Hot Encoding in the feature engineering

[15]:
```
print(data[data['Fields of research interest'].isnull()].index)
data['Fields of research interest'].fillna('Not interested in Research',␣
 ↪inplace=True)
print(data[data['Fields of research interest'].isnull()].index)
```

```
Index([10, 11], dtype='int64')
Index([], dtype='int64')
```

[16]:
```
interests = data['Fields of research interest'].str.split(', ').explode()

plt.figure(figsize=(10, 6))
sns.countplot(y=interests, order=interests.value_counts().index,␣
 ↪palette='viridis')
plt.title('Research Interests Count')
plt.xlabel('Count')
plt.ylabel('Research Interests')

plt.show()
```



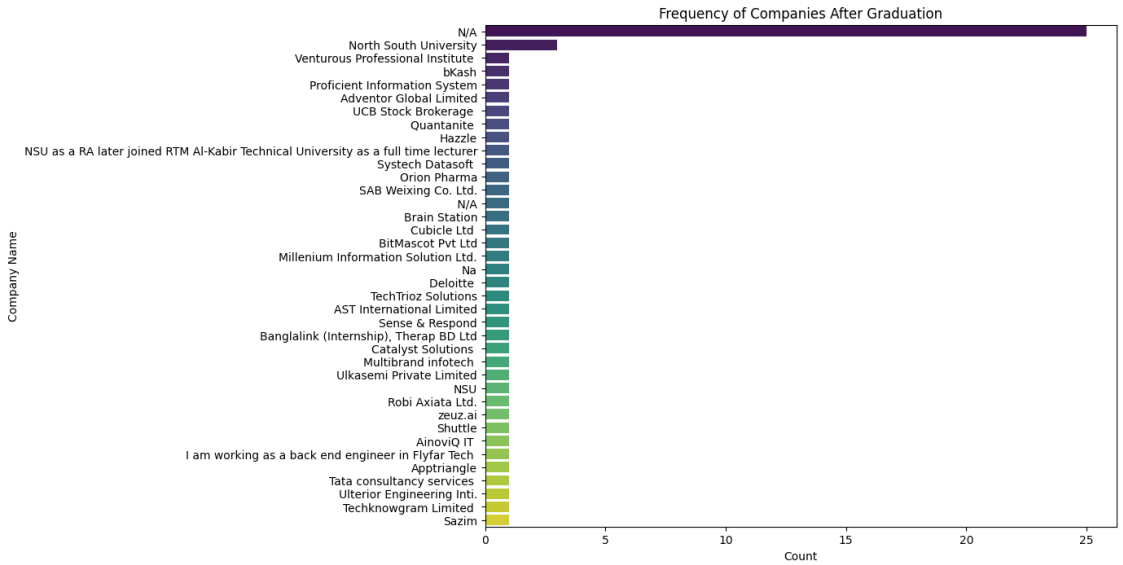
3.2.1 Now dealing with 'What was the name of the company you worked for right after your graduation? if your answer is not write N/A.' column

But we will drop this column because it has a lot of different values and the column is not actually needed in our prediction model

[17]:
```
print(data['What was the name of the company you worked for right after your␣
 ↪graduation? if your answer is not write N/A.'].isnull().sum())
data['What was the name of the company you worked for right after your␣
 ↪graduation? if your answer is not write N/A.'].fillna('N/A', inplace=True)
print(data['What was the name of the company you worked for right after your␣
 ↪graduation? if your answer is not write N/A.'].isnull().sum())
```

```
25
0
```

[18]:
```
plt.figure(figsize=(10, 8))
sns.countplot(y='What was the name of the company you worked for right after␣
 ↪your graduation? if your answer is not write N/A.', data=data,␣
 ↪palette='viridis',
              order=data['What was the name of the company you worked for right␣
 ↪after your graduation? if your answer is not write N/A.'].value_counts().
 ↪index)
plt.title('Frequency of Companies After Graduation')
plt.xlabel('Count')
plt.ylabel('Company Name')

plt.show()
```

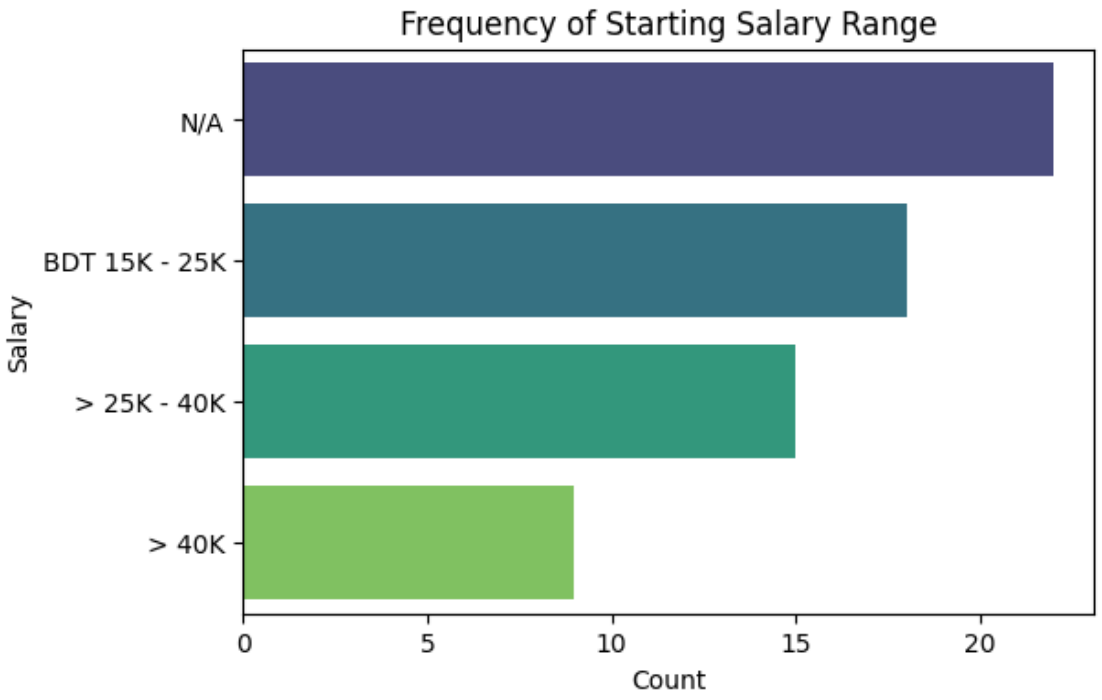


3.3 Now 'What was your starting salary range?' Column

[19]:
```
print(data['What was your starting salary range?'].unique())
data['What was your starting salary range?'].fillna('N/A', inplace=True)
print(data['What was your starting salary range?'].unique())
```

```
['BDT 15K - 25K' '> 25K - 40K' '> 40K' nan]
['BDT 15K - 25K' '> 25K - 40K' '> 40K' 'N/A']
```

[20]:
```
plt.figure(figsize=(6, 4))
sns.countplot(y='What was your starting salary range?', data=data,␣
 ↪palette='viridis',
              order=data['What was your starting salary range?'].value_counts().
 ↪index)
plt.title('Frequency of Starting Salary Range')
plt.xlabel('Count')
plt.ylabel('Salary')

plt.show()
```



3.3.1 Now Dealing with 'What was the name of the University you completed your higher studies from? if answer is no write N/A.'

1. First we are going to remove the null values

2. Then We are going to replace the Na with N/A

3. Then we are going to visualize it

[21]:
```
print("Null Values : ",data['What was the name of the University you completed␣
 ↪your higher studies from? if answer is no write N/A.'].isnull().sum())

print("\n\nWhat was the name of the University you completed your higher␣
 ↪studies from? if answer is no write N/A.\n\n",data['What was the name of the␣
 ↪University you completed your higher studies from? if answer is no write N/A.
 ↪'].unique())
```

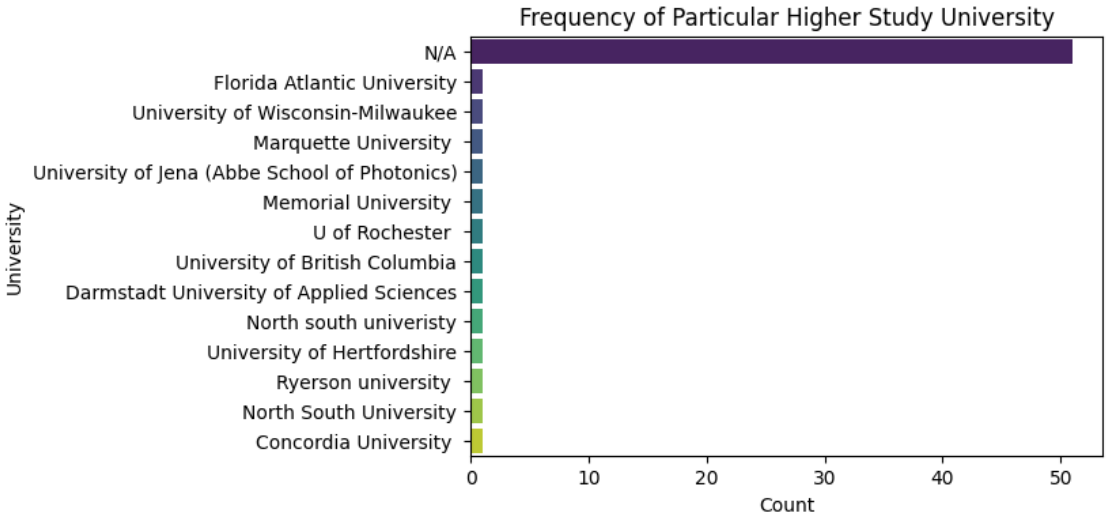
```
Null Values :  50


What was the name of the University you completed your higher studies from? if
answer is no write N/A.

 ['Florida Atlantic University' 'University of Wisconsin-Milwaukee'
 'Marquette University ' 'University of Jena (Abbe School of Photonics)'
 nan 'Memorial University ' 'U of Rochester '
 'University of British Columbia'
 'Darmstadt University of Applied Sciences' 'North south univeristy'
 'University of Hertfordshire' 'Ryerson university '
 'North South University' 'Concordia University ' 'Na']
```

[22]:
```
data['What was the name of the University you completed your higher studies␣
 ↪from? if answer is no write N/A.'].fillna('N/A', inplace=True)
data['What was the name of the University you completed your higher studies␣
 ↪from? if answer is no write N/A.'].replace('Na', 'N/A', inplace=True)
data['What was the name of the University you completed your higher studies␣
 ↪from? if answer is no write N/A.'].unique()
```

[22]:
```
array(['Florida Atlantic University', 'University of Wisconsin-Milwaukee',
       'Marquette University ',
       'University of Jena (Abbe School of Photonics)', 'N/A',
       'Memorial University ', 'U of Rochester ',
       'University of British Columbia',
       'Darmstadt University of Applied Sciences',
       'North south univeristy', 'University of Hertfordshire',
       'Ryerson university ', 'North South University',
       'Concordia University '], dtype=object)
```

[23]:
```
plt.figure(figsize=(6, 4))
sns.countplot(y='What was the name of the University you completed your higher␣
 ↪studies from? if answer is no write N/A.', data=data, palette='viridis',
              order=data['What was the name of the University you completed␣
 ↪your higher studies from? if answer is no write N/A.'].value_counts().index)
```

```
plt.title('Frequency of Particular Higher Study University')
plt.xlabel('Count')
plt.ylabel('University')

plt.show()
```



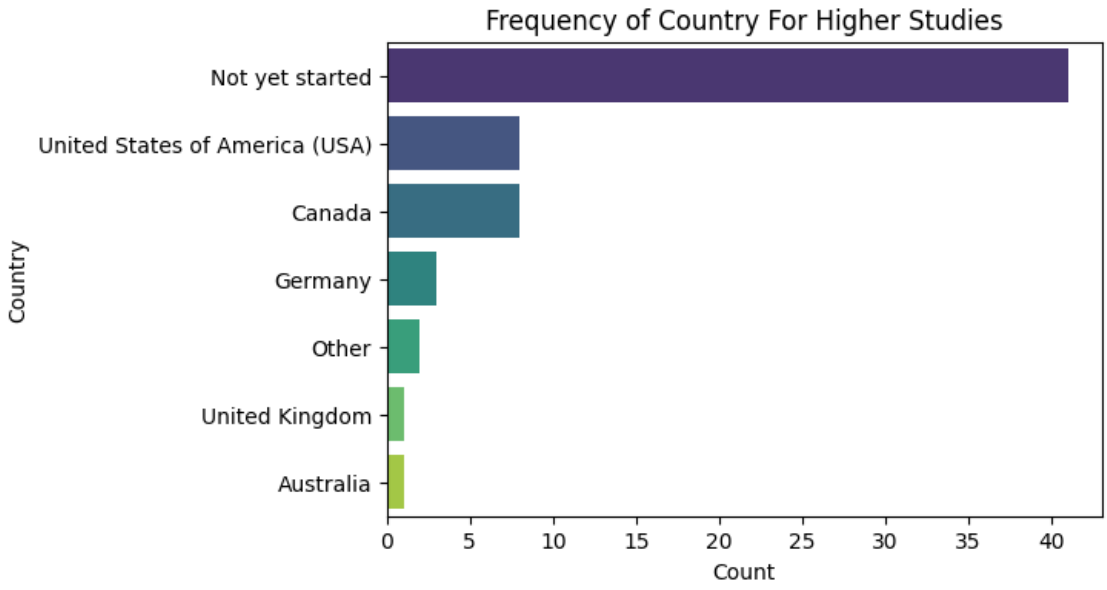
3.4 Now for the 'Which country did you go to for higher studies?' Column

1. We will just replace the 1 nan value with 'Not yet started'

2. Then we will visualize it

[24]:
```
print(data['Which country did you go to for higher studies?'].unique())

data['Which country did you go to for higher studies?'].replace('N/A', 'Not yet␣
 ↪started', inplace=True)
data['Which country did you go to for higher studies?'].fillna("Not yet␣
 ↪started",inplace=True)

plt.figure(figsize=(6, 4))
sns.countplot(y='Which country did you go to for higher studies?', data=data,␣
 ↪palette='viridis',
              order=data['Which country did you go to for higher studies?'].
 ↪value_counts().index)
plt.title('Frequency of Country For Higher Studies')
plt.xlabel('Count')
plt.ylabel('Country')

plt.show()
```

```
['United States of America (USA)' 'Germany' 'Not yet started' 'Canada'
 'United Kingdom' 'Australia' 'Other' nan]
```



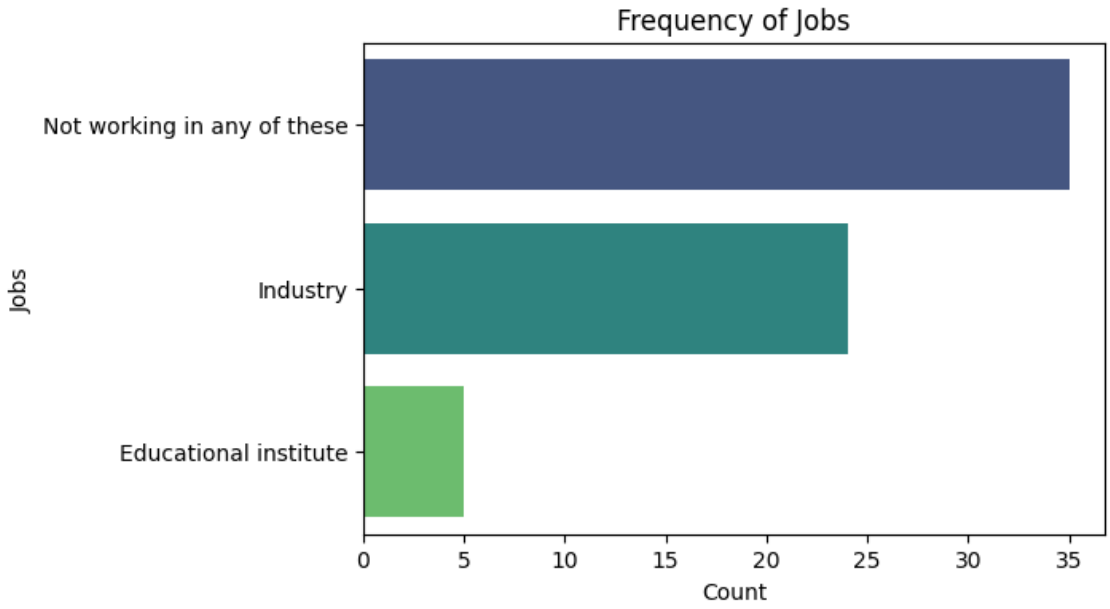
3.5 Now for the 'Where did you start working after completing your higher studies?' Column

3.5.1 1. We will just replace the nan value with 'Not Working'

3.5.2 2. Then We will visualize it

[25]:
```
print(data['Where did you start working after completing your higher studies?'].
 ↪unique())

data['Where did you start working after completing your higher studies?'].
 ↪fillna('Not working in any of these', inplace=True)

plt.figure(figsize=(6, 4))
sns.countplot(y='Where did you start working after completing your higher␣
 ↪studies?', data=data, palette='viridis',
              order=data['Where did you start working after completing your␣
 ↪higher studies?'].value_counts().index)
plt.title('Frequency of Jobs')
plt.xlabel('Count')
plt.ylabel('Jobs')

plt.show()
```

```
['Industry' nan 'Educational institute']
```



[26]:
```
data.isnull().sum()
```

[26]:
```
Timestamp
0
Email Address
0
Gender
0
Cumulative Grade Point Average(CGPA)
0
Total year now since your graduation
0
Fields of research interest
0
Number of publications
0
Were you a teaching assistant (TA) or research assistant(RA) during your time in
NSU?                                0
Did you participate in any extra curricular activities while you were studying
in NSU?                              0
Were you interested in coding?
0
Did you participate in any competitive programming?
0
Programming languages you know
0
```

```
Frameworks you know
0
Did you start working as a software engineer after graduation?
0
What was the name of the company you worked for right after your graduation? if
your answer is not write N/A.    0
What was your starting salary range?
0
Did you have any Start-ups after graduation?
0
Which country did you go to for higher studies?
0
What was the name of the University you completed your higher studies from? if
answer is no write N/A.           0
Where did you start working after completing your higher studies?
0
dtype: int64
```

3.6 We have succesfully removed all the Null Values/Wrong Data/Duplicate Data

3.6.1 Now we will drop the Timestamp and Email Address from our dataset because We do not need these columns

[27]:
```
data.drop("Timestamp", axis=1, inplace=True)
data.drop("Email Address", axis=1, inplace=True)
```

4 Column Investigation for Wrong Data

4.1 We will go Column by Column

4.1.1 Starting with the column : Total year now since your graduation

We can see we have a lot of garbage inputs and we are gonna manipulate the data with proper values according to those garbage wrong values

[28]:
```
data['Total year now since your graduation'].unique()
```

[28]:
```
array(['1', '4', '3', '8', '0', '2', '6', '5', 'one', '1.5', '3.5',
       'almost 5 years', '4.5 years', '1 year', '2.5 years', '8 Months',
       '0.5 year ', '5 years ', 'aprox 1 year', '0.5', '1 year ',
       '2 months', '7 months. ', '1 Year', '2 Month', 'less than 1 year',
       '0.2'], dtype=object)
```

[29]:
```
data['Total year now since your graduation'].replace('8 Months', 'less than 1␣
 ↪year', inplace=True)
data['Total year now since your graduation'].replace('0.5 year ', 'less than 1␣
 ↪year', inplace=True)
```

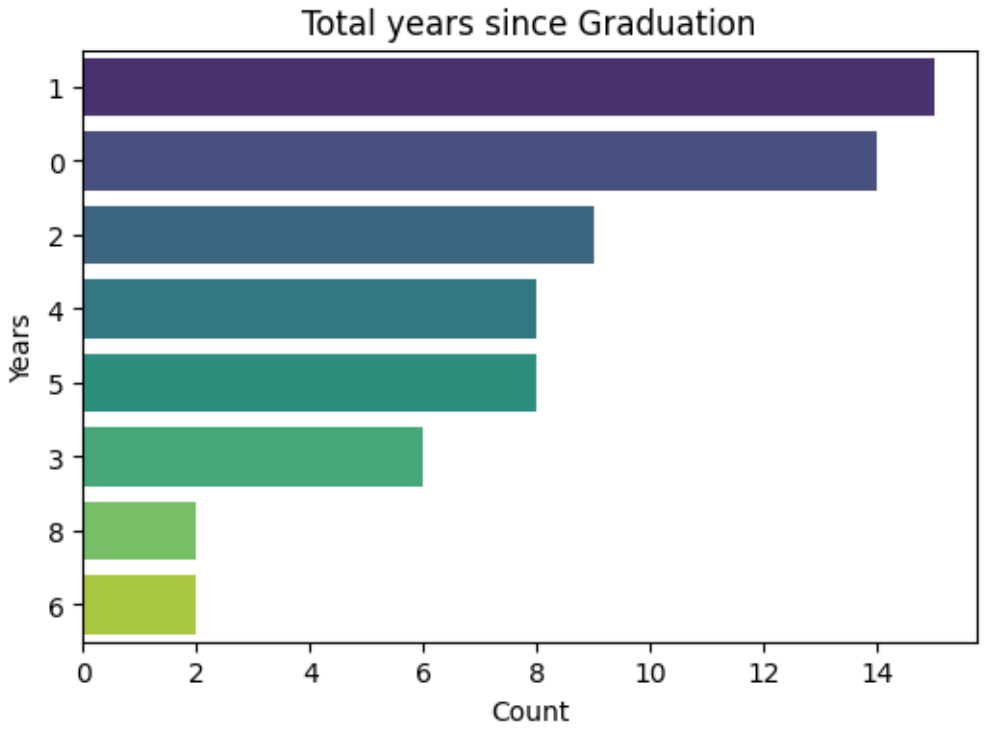
```
data['Total year now since your graduation'].replace('0.5', 'less than 1 year',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('2 months', 'less than 1␣
 ↪year', inplace=True)
data['Total year now since your graduation'].replace('7 months. ', 'less than 1␣
 ↪year', inplace=True)
data['Total year now since your graduation'].replace('2 Month', 'less than 1␣
 ↪year', inplace=True)
data['Total year now since your graduation'].replace('0.2', 'less than 1 year',␣
 ↪inplace=True)

data['Total year now since your graduation'].replace('one', '1', inplace=True)
data['Total year now since your graduation'].replace('1 year', '1',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('1 year ', '',␣
 ↪inplace=True)


data['Total year now since your graduation'].replace('less than 1 year', '0',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('almost 5 years', '5',␣
 ↪inplace=True)


data['Total year now since your graduation'].replace('4.5 years', '4',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('2.5 years', '2',␣
 ↪inplace=True)

data['Total year now since your graduation'].replace('5 years ', '5',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('aprox 1 year', '1',␣
 ↪inplace=True)
data['Total year now since your graduation'].replace('1 Year', '1',␣
 ↪inplace=True)

data['Total year now since your graduation'].replace('1.5', '1', inplace=True)
data['Total year now since your graduation'].replace('3.5', '3', inplace=True)
data['Total year now since your graduation'].replace('', '0', inplace=True)
```

[30]:
```
data['Total year now since your graduation'].isnull().sum()
```

[30]:
```
0
```

[31]:
```
plt.figure(figsize=(6, 4))
```

```
sns.countplot(y='Total year now since your graduation', data=data,␣
 ↪palette='viridis',
             order=data['Total year now since your graduation'].value_counts().
 ↪index)
plt.title('Total years since Graduation')
plt.xlabel('Count')
plt.ylabel('Years')

plt.show()
```



[32]:
```
data.columns
```

[32]:
```
Index(['Gender', 'Cumulative Grade Point Average(CGPA)',
       'Total year now since your graduation', 'Fields of research interest',
       'Number of publications',
       'Were you a teaching assistant (TA) or research assistant(RA) during your
time in NSU?',
       'Did you participate in any extra curricular activities while you were
studying in NSU?',
       'Were you interested in coding?',
       'Did you participate in any competitive programming?',
       'Programming languages you know', 'Frameworks you know',
       'Did you start working as a software engineer after graduation?',
```

```
       'What was the name of the company you worked for right after your
graduation? if your answer is not write N/A.',
       'What was your starting salary range?',
       'Did you have any Start-ups after graduation?',
       'Which country did you go to for higher studies?',
       'What was the name of the University you completed your higher studies
from? if answer is no write N/A.',
       'Where did you start working after completing your higher studies?'],
      dtype='object')
```

4.2 Question 4. Organize the data into a JSON file and then read this JSON data.

[33]:
```
data.to_json('DatasetJson.json')
```

[34]:
```
read_data = pd.read_json('DatasetJson.json')
read_data.head(5)
```

[34]:
```
   Gender Cumulative Grade Point Average(CGPA)  \
1    Male                           3 to < 3.5
2    Male                             3.5 to 4
3  Female                             3.5 to 4
4    Male                             3.5 to 4
5  Female                             3.5 to 4

   Total year now since your graduation  \
1                                     1
2                                     4
3                                     3
4                                     8
5                                     0

                        Fields of research interest Number of publications  \
1  Internet of Things(IoT) and Embedded Systems, …                  1 - 5
2  Machine Learning, Artificial Intelligence, Nat…           More than 10
3                        Human-Computer interaction                 1 - 5
4  Machine Learning, Data Science/Data mining, Ar…                 6-10
5  Machine Learning, Data Science/Data mining, Ar…                 1 - 5

  Were you a teaching assistant (TA) or research assistant(RA) during your time
in NSU?  \
1                                                 No
2                                                Yes
3                                                Yes
4                                                Yes
5                                                 No
```

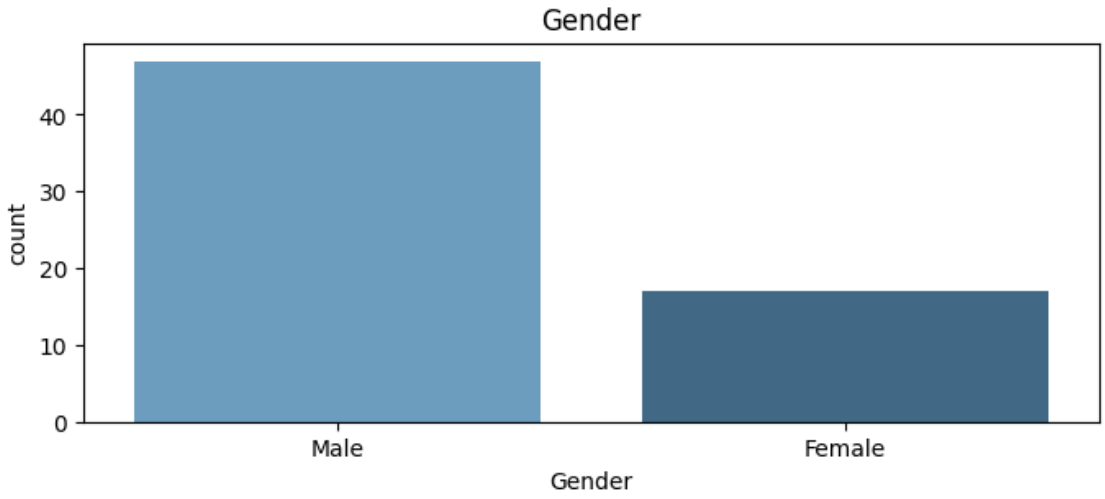
```
  Did you participate in any extra curricular activities while you were studying
in NSU?  \
1                                                  Yes
2                                                  Yes
3                                                  Yes
4                                                  Yes
5                                                   No

  Were you interested in coding?  \
1                            Yes
2                            Yes
3                            Yes
4                            Yes
5                             No

  Did you participate in any competitive programming?  \
1                                                  Yes
2                                                  Yes
3                                                  Yes
4                                                   No
5                                                   No

                     Programming languages you know  \
1                                      C, C++, HTML
2                             C, Java, Python, C++, R
3  C, Python, PHP, C++, HTML, CSS, JavaScript, R,…
4                                   C, Python, R, SQL
5                                   C, HTML, CSS, SQL

              Frameworks you know  \
1                   Not interested
2                          Laravel
3  React, Django, Laravel, Node.js
4                           Django
5                           jQuery

  Did you start working as a software engineer after graduation?  \
1                                                   No
2                                                   No
3                                                  Yes
4                                                   No
5                                                   No

  What was the name of the company you worked for right after your graduation?
if your answer is not write N/A.  \
1                  Venturous Professional Institute
2                          AST International Limited
```

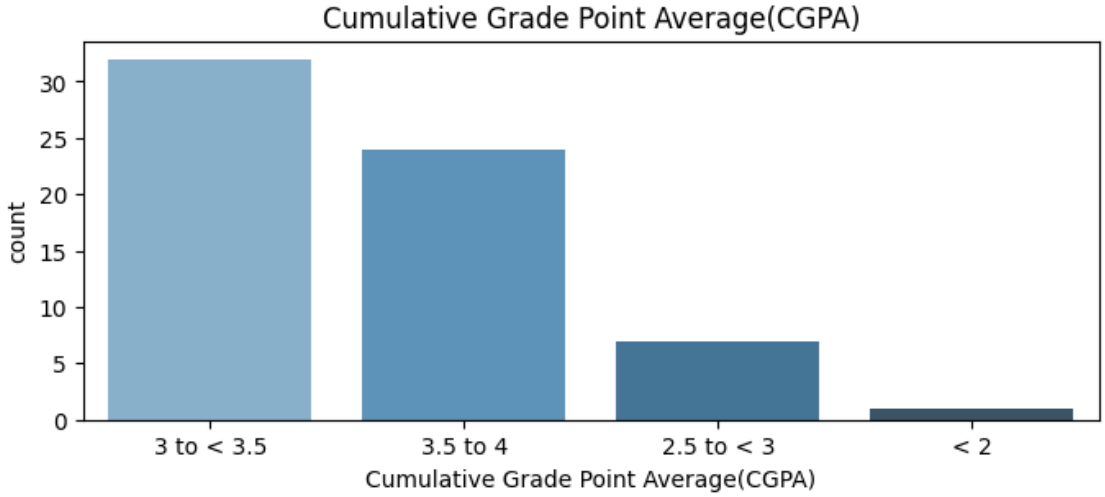
```
3             Banglalink (Internship), Therap BD Ltd
4                               North South University
5                                  Catalyst Solutions

  What was your starting salary range?  \
1                          BDT 15K - 25K
2                            > 25K - 40K
3                                 > 40K
4                          BDT 15K - 25K
5                          BDT 15K - 25K

  Did you have any Start-ups after graduation?  \
1                                           No
2                                           No
3                                           No
4                                           No
5                                           No

  Which country did you go to for higher studies?  \
1                  United States of America (USA)
2                  United States of America (USA)
3                  United States of America (USA)
4                                        Germany
5                                 Not yet started

  What was the name of the University you completed your higher studies from? if
answer is no write N/A.  \
1                          Florida Atlantic University
2                  University of Wisconsin-Milwaukee
3                               Marquette University
4      University of Jena (Abbe School of Photonics)
5                                              N/A

  Where did you start working after completing your higher studies?
1                                           Industry
2                        Not working in any of these
3                        Not working in any of these
4                              Educational institute
5                        Not working in any of these
```

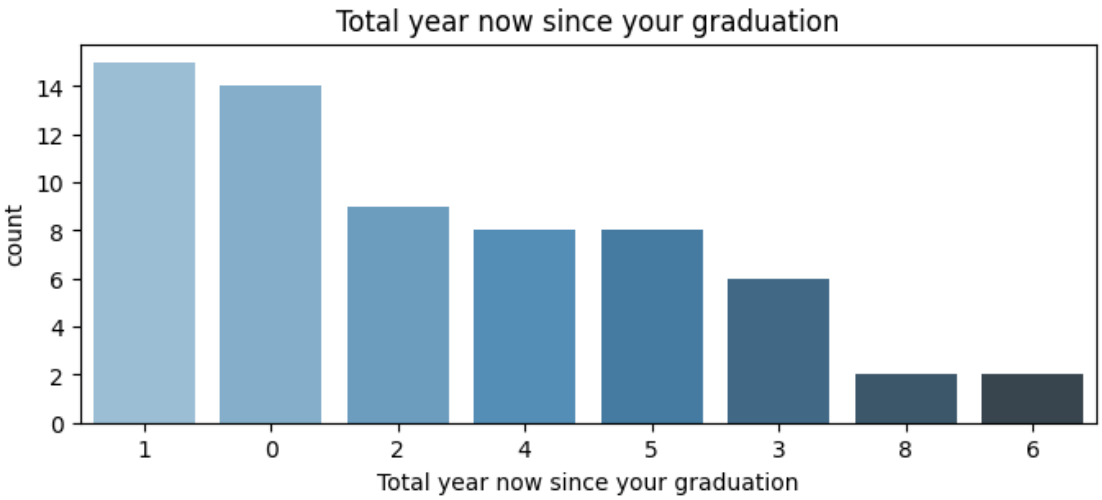
4.3 Visualize the Dataset [An Overview of the distribution of Data]

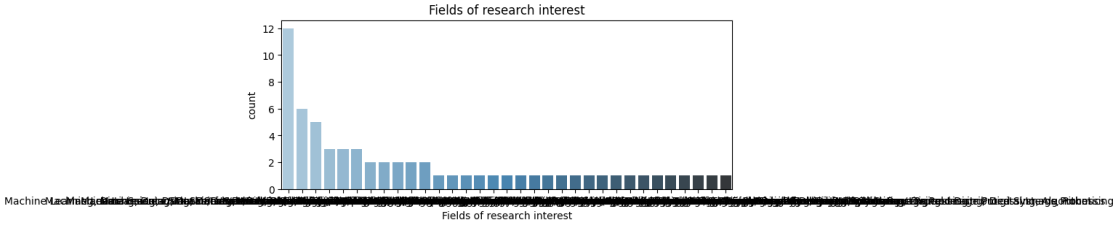
[35]:
```
for i in data.columns:
    plt.figure(figsize=(8, 3))
    sns.barplot(x=data[i].value_counts().index, y=data[i].value_counts(),␣
 ↪palette="Blues_d")
    plt.title(i)
```

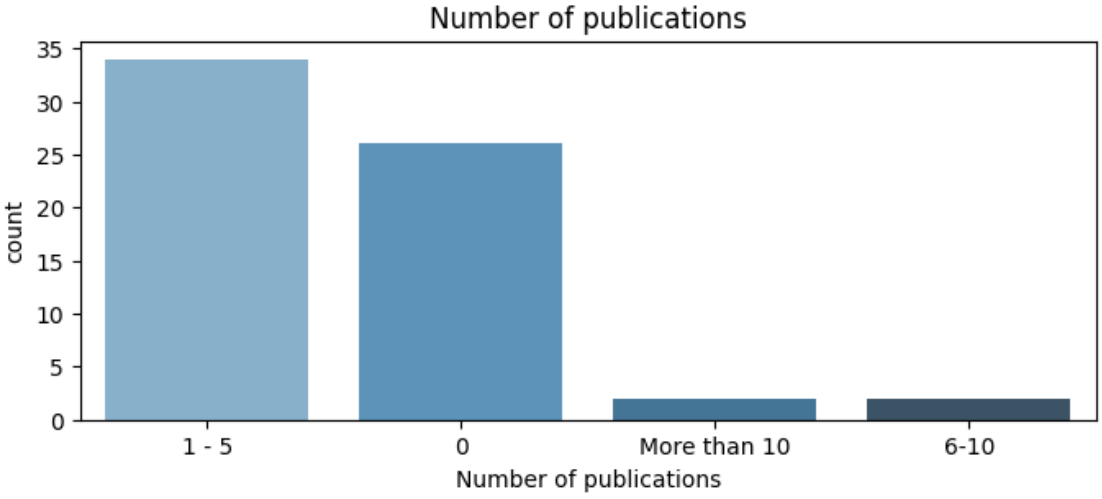
```
    plt.show()
```

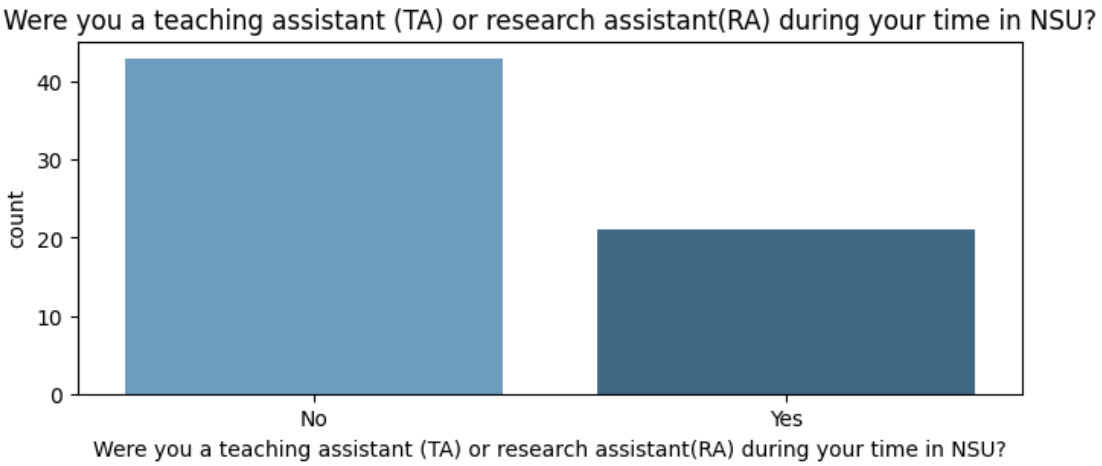

Gender
count
40
30
20
10
0
Male
Female
Gender

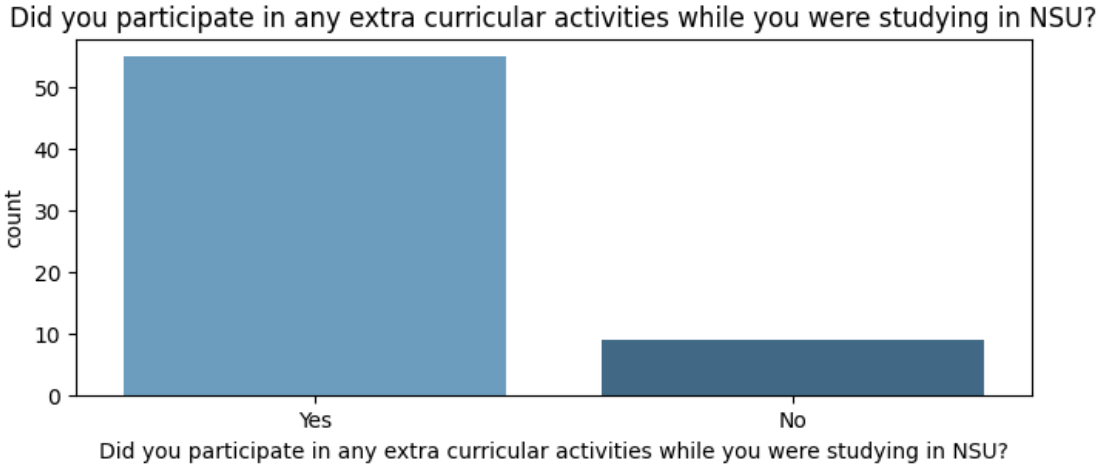

Cumulative Grade Point Average(CGPA)
count
30
25
20
15
10
5
0
3 to < 3.5
3.5 to 4
2.5 to < 3
< 2
Cumulative Grade Point Average(CGPA)

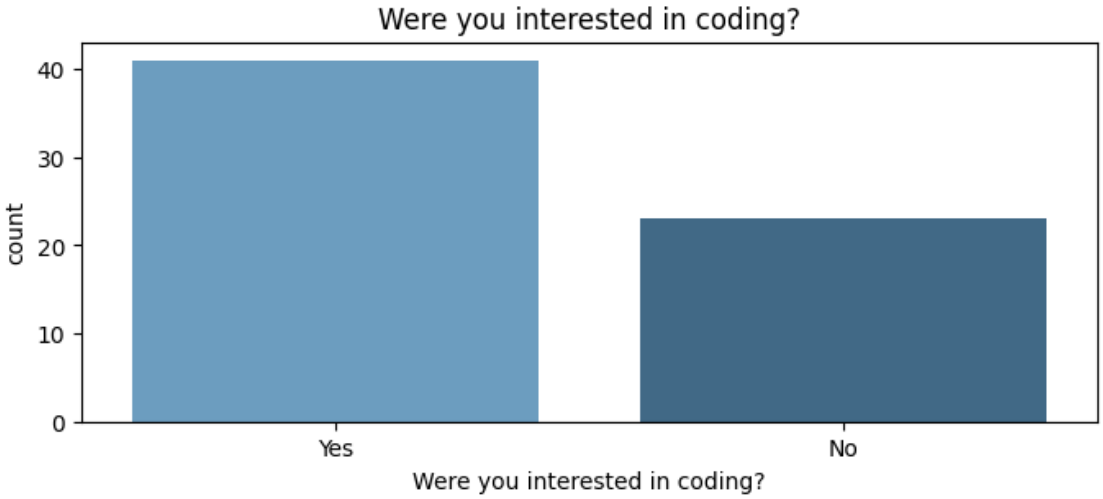

Total year now since your graduation
count
0
2
4
6
8
10
12
14
1
0
2
4
5
3
8
6
Total year now since your graduation

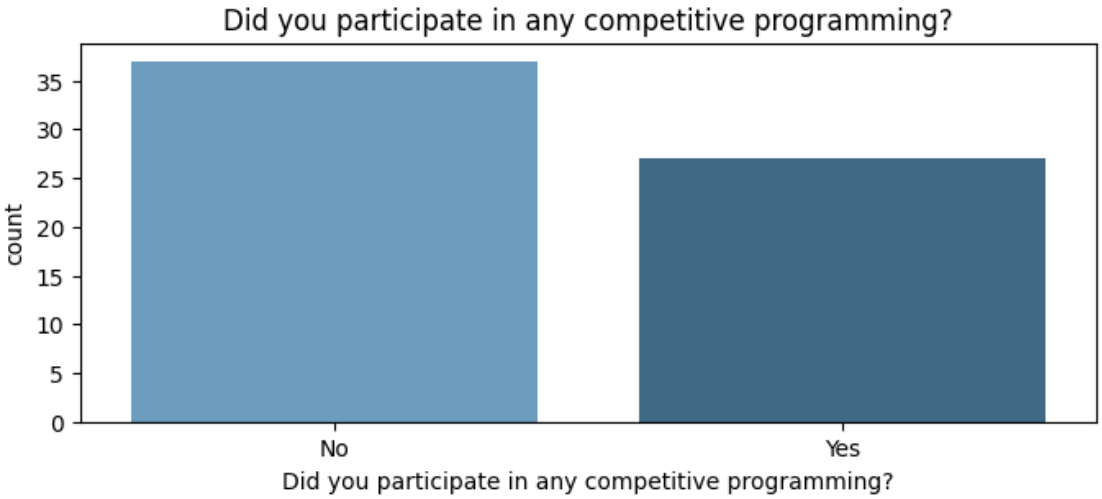

Fields of research interest
count
0
2
4
6
8
10
12
Fields of research interest

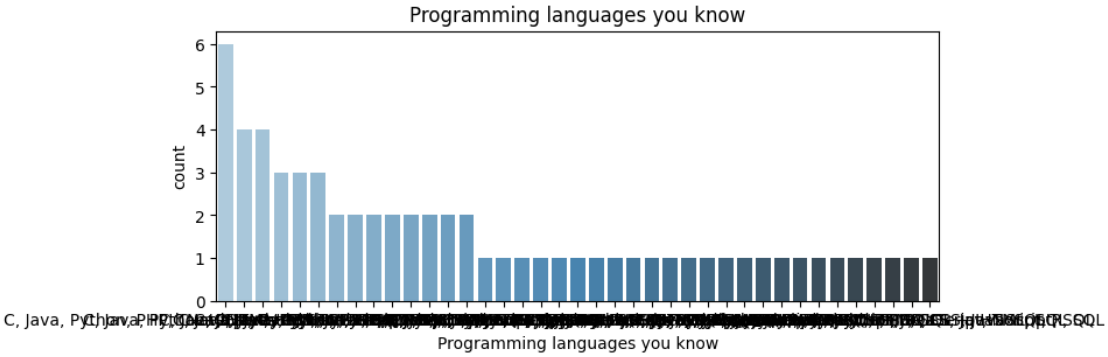

Number of publications
count
0
5
10
15
20
25
30
35
1 - 5
0
More than 10
6-10
Number of publications

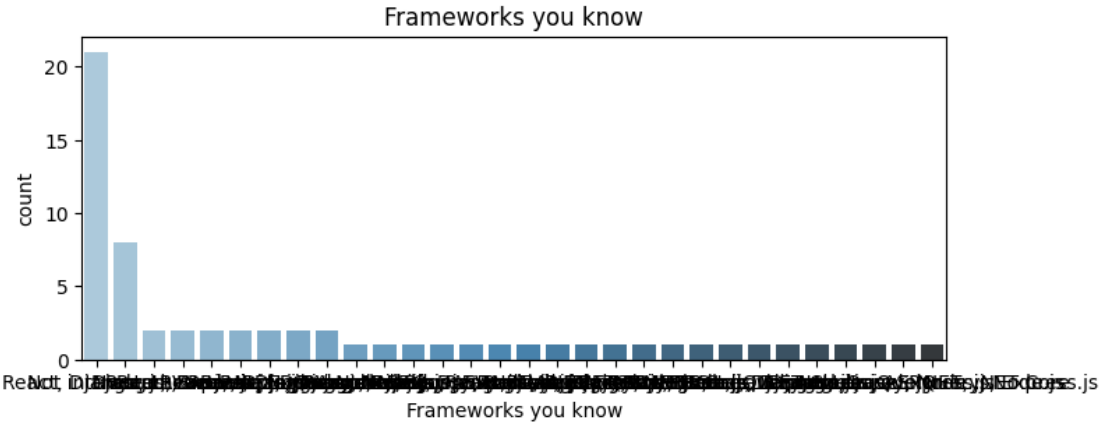

Were you a teaching assistant (TA) or research assistant(RA) during your time in NSU?
count
0
10
20
30
40
No
Yes
Were you a teaching assistant (TA) or research assistant(RA) during your time in NSU?

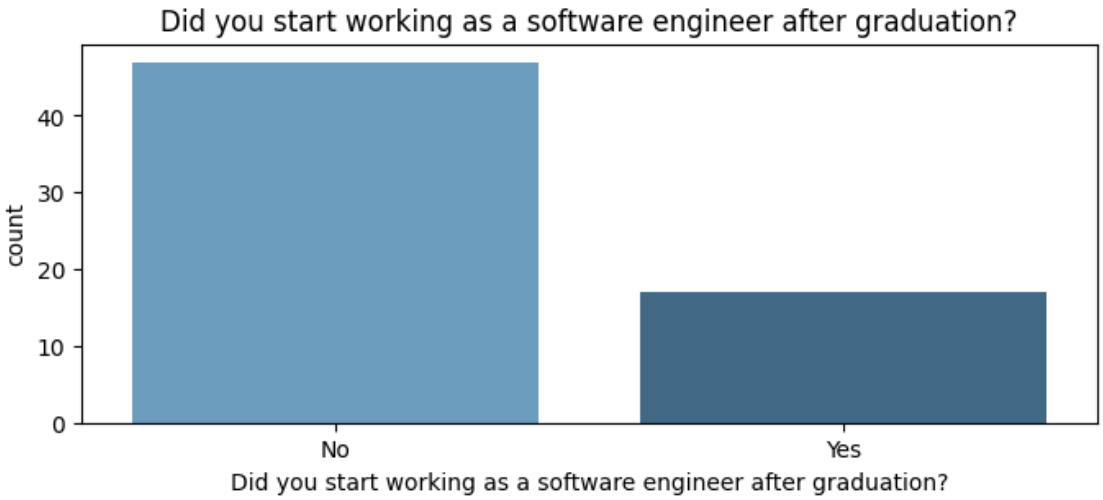

Did you participate in any extra curricular activities while you were studying in NSU?
count
0
10
20
30
40
50
Yes
No
Did you participate in any extra curricular activities while you were studying in NSU?

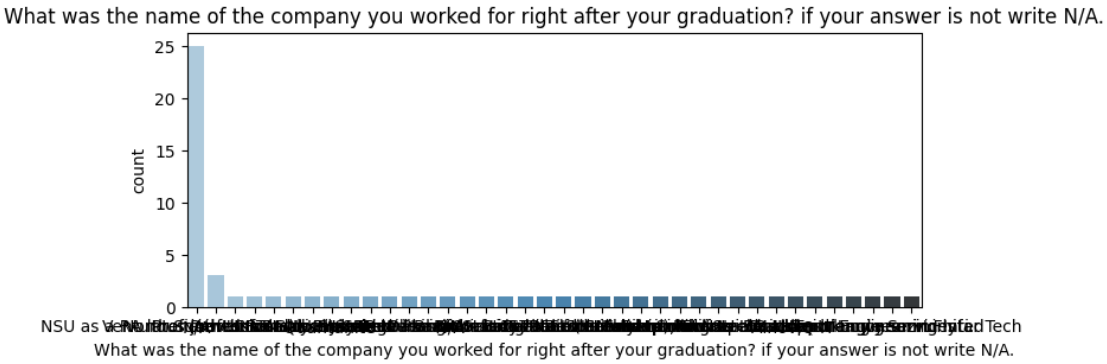

Were you interested in coding?
count
0
10
20
30
40
Yes
No
Were you interested in coding?

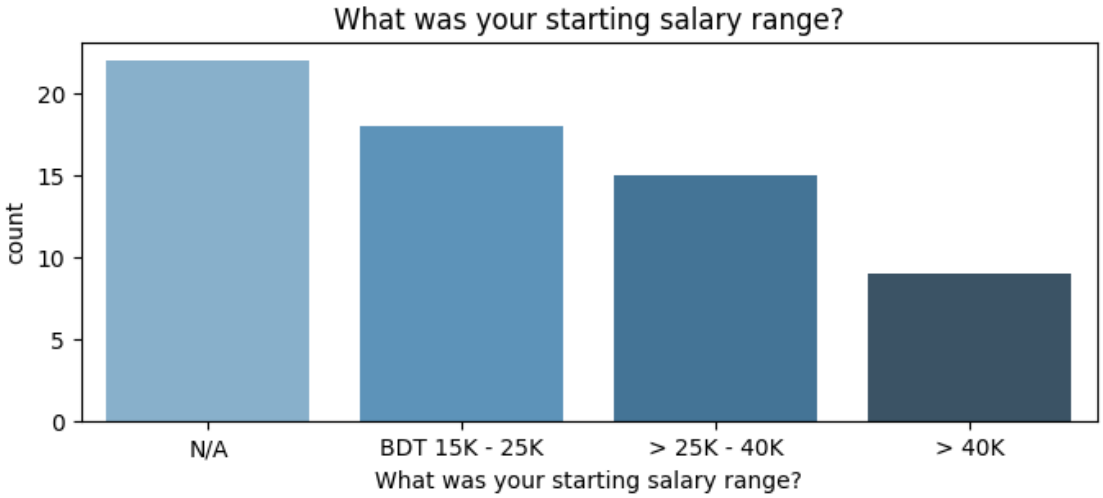

Did you participate in any competitive programming?
count
35
30
25
20
15
10
5
0
No
Yes
Did you participate in any competitive programming?

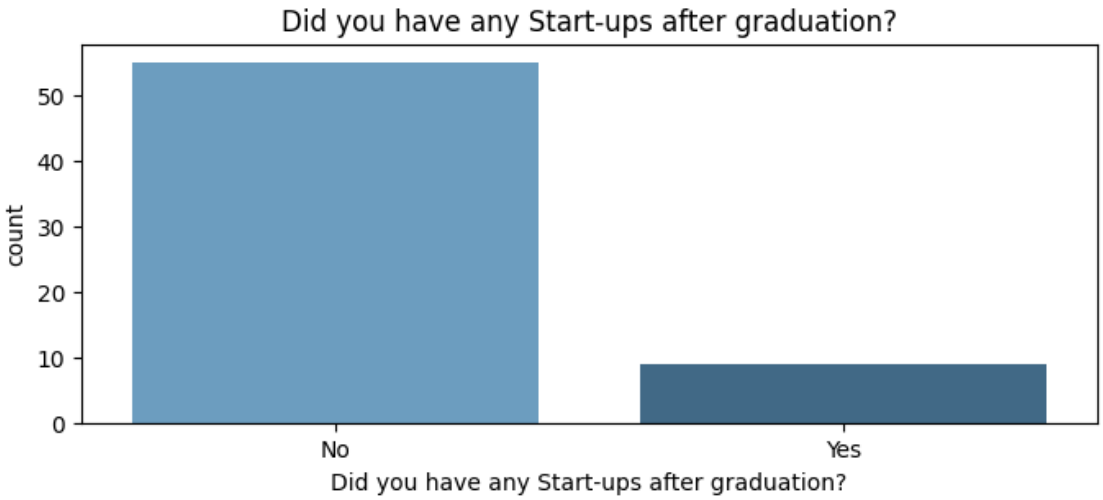

Programming languages you know
count
6
5
4
3
2
1
0
C, Java, Python
Programming languages you know

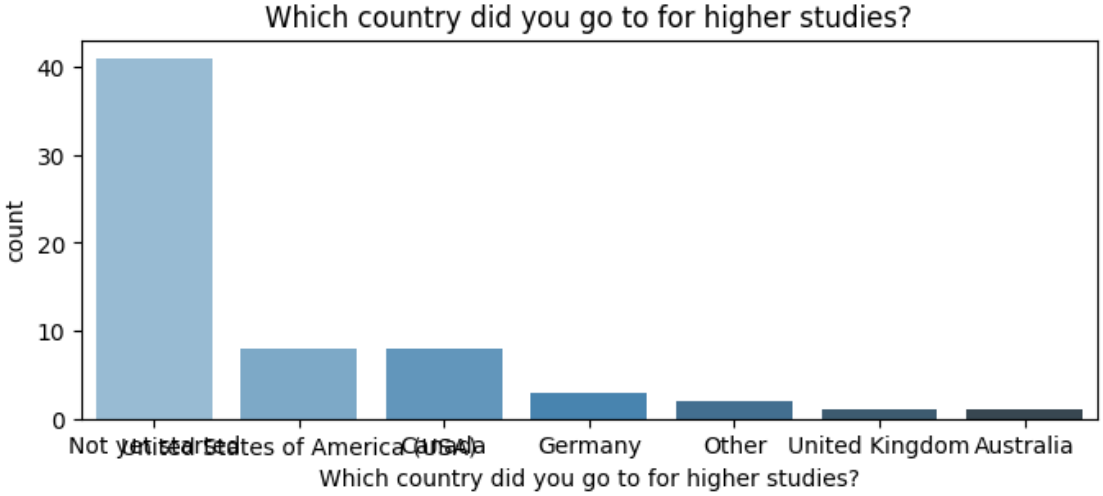

Frameworks you know
count
20
15
10
5
0
React
Frameworks you know


Did you start working as a software engineer after graduation?
count
40
30
20
10
0
No
Yes
Did you start working as a software engineer after graduation?

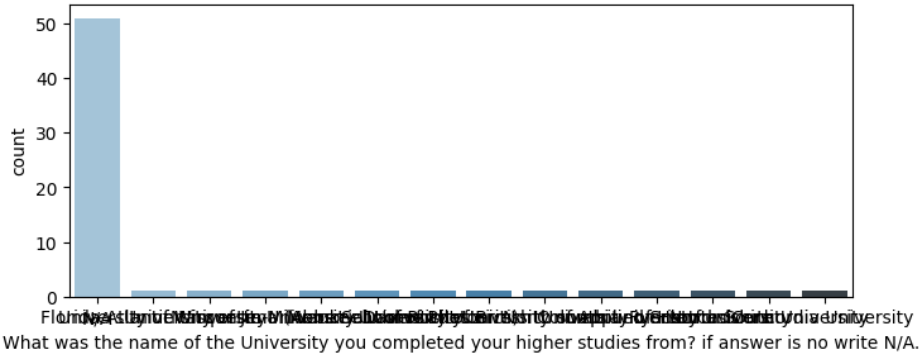

What was the name of the company you worked for right after your graduation? if your answer is not write N/A.
count
25
20
15
10
5
0
What was the name of the company you worked for right after your graduation? if your answer is not write N/A.


What was your starting salary range?
count
20
15
10
5
0
N/A
BDT 15K - 25K
> 25K - 40K
> 40K
What was your starting salary range?





What was the name of the University you completed your higher studies from? if answer is no write N/A.



5 Question 5 : Do you think scaling is necessary in this case? Justify

5.1 Answer: Yes Scaling is indeed needed because our dataset has only Categorical values, So we need to convert them to Numerical Values

6 Scaling/Encoding [Categorical Values –> Numerical Values]

We can see our dataset has a lot of categorical values, We need to convert them to Numerical values in order to show the Correlation betweet the features # So Scaling the data is really needed ## We will do 4 types of Scaling/Encoding 1. Binary Label Encoding {For features that has yes/no or male/female values} 2. Ordinal Encoding {For Features that has values that maintains an order} 3. One Hot Encoding {For Features that has Multiple Choice options} 4. Nominal Encoding {For Features that maybe the Dependable Feature or The Target Value}

6.1 We will Do Binary Label Encoding for these columns

Because These columns have either yes or no Values

[36]:
```
print(f"Gender Column Values : {data['Gender'].unique()}\n")
print(f"Were you a teaching assistant (TA) or research assistant(RA) during␣
 ↪your time in NSU? Column Values : {data['Were you a teaching assistant (TA)␣
 ↪or research assistant(RA) during your time in NSU?'].unique()}\n")
print(f"Did you participate in any extra curricular activities while you were␣
 ↪studying in NSU? Column Values : {data['Did you participate in any extra␣
 ↪curricular activities while you were studying in NSU?'].unique()}\n")
```

```
print(f"Were you interested in coding? Column Values : {data['Were you␣
 ↪interested in coding?'].unique()}\n")
print(f"Did you participate in any competitive programming? Column Values :␣
 ↪{data['Did you participate in any competitive programming?'].unique()}\n")
print(f"Did you start working as a software engineer after graduation? Column␣
 ↪Values : {data['Did you start working as a software engineer after␣
 ↪graduation?'].unique()}\n")
print(f"Did you have any Start-ups after graduation? Column Values : {data['Did␣
 ↪you have any Start-ups after graduation?'].unique()}\n")
```

```
Gender Column Values : ['Male' 'Female']

Were you a teaching assistant (TA) or research assistant(RA) during your time in
NSU? Column Values : ['No' 'Yes']

Did you participate in any extra curricular activities while you were studying
in NSU? Column Values : ['Yes' 'No']

Were you interested in coding? Column Values : ['Yes' 'No']

Did you participate in any competitive programming? Column Values : ['Yes' 'No']

Did you start working as a software engineer after graduation? Column Values :
['No' 'Yes']

Did you have any Start-ups after graduation? Column Values : ['No' 'Yes']
```

[37]:
```
data.head(5)
```

[37]:

	Gender	Cumulative Grade Point Average(CGPA)	Total year now since your graduation	Fields of research interest	Number of publications
1	Male	3 to < 3.5	1		
2	Male	3.5 to 4	4		
3	Female	3.5 to 4	3		
4	Male	3.5 to 4	8		
5	Female	3.5 to 4	0		
1				Internet of Things(IoT) and Embedded Systems, …	1 - 5
2				Machine Learning, Artificial Intelligence, Nat…	More than 10

```
3                          Human-Computer interaction                  1 - 5
4  Machine Learning, Data Science/Data mining, Ar…                    6-10
5  Machine Learning, Data Science/Data mining, Ar…                   1 - 5

  Were you a teaching assistant (TA) or research assistant(RA) during your time
in NSU?  \
1                                                    No
2                                                   Yes
3                                                   Yes
4                                                   Yes
5                                                    No

  Did you participate in any extra curricular activities while you were studying
in NSU?  \
1                                                   Yes
2                                                   Yes
3                                                   Yes
4                                                   Yes
5                                                    No

  Were you interested in coding?  \
1                             Yes
2                             Yes
3                             Yes
4                             Yes
5                              No

  Did you participate in any competitive programming?  \
1                                                   Yes
2                                                   Yes
3                                                   Yes
4                                                    No
5                                                    No

                     Programming languages you know  \
1                                       C, C++, HTML
2                             C, Java, Python, C++, R
3  C, Python, PHP, C++, HTML, CSS, JavaScript, R,…
4                                   C, Python, R, SQL
5                                   C, HTML, CSS, SQL

              Frameworks you know  \
1                   Not interested
2                          Laravel
3  React, Django, Laravel, Node.js
4                           Django
5                           jQuery
```

```
  Did you start working as a software engineer after graduation?  \
1                                                        No
2                                                        No
3                                                       Yes
4                                                        No
5                                                        No

  What was the name of the company you worked for right after your graduation?
if your answer is not write N/A.  \
1                  Venturous Professional Institute
2                          AST International Limited
3             Banglalink (Internship), Therap BD Ltd
4                              North South University
5                                 Catalyst Solutions

  What was your starting salary range?  \
1                         BDT 15K - 25K
2                           > 25K - 40K
3                                > 40K
4                         BDT 15K - 25K
5                         BDT 15K - 25K

  Did you have any Start-ups after graduation?  \
1                                          No
2                                          No
3                                          No
4                                          No
5                                          No

  Which country did you go to for higher studies?  \
1                  United States of America (USA)
2                  United States of America (USA)
3                  United States of America (USA)
4                                        Germany
5                                 Not yet started

  What was the name of the University you completed your higher studies from? if
answer is no write N/A.  \
1                        Florida Atlantic University
2                  University of Wisconsin-Milwaukee
3                              Marquette University
4      University of Jena (Abbe School of Photonics)
5                                              N/A

  Where did you start working after completing your higher studies?
1                                          Industry
```

```
2                         Not working in any of these
3                         Not working in any of these
4                               Educational institute
5                         Not working in any of these
```

[38]:
```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
columns_to_encode = [
    'Gender',
    'Were you a teaching assistant (TA) or research assistant(RA) during your␣
 ↪time in NSU?',
    'Did you participate in any extra curricular activities while you were␣
 ↪studying in NSU?',
    'Were you interested in coding?',
    'Did you participate in any competitive programming?',
    'Did you start working as a software engineer after graduation?',
    'Did you have any Start-ups after graduation?',

]
encoded_binary = data[columns_to_encode]
for column in columns_to_encode:
    encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
encoded_binary
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1078086232.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_binary[column] = label_encoder.fit_transform(encoded_binary[column])
```

[38]:

```
    Gender  \
1        1
2        1
3        0
4        1
5        0
..     …
60       1
61       1
62       1
63       1
64       1

    Were you a teaching assistant (TA) or research assistant(RA) during your
```

```
time in NSU?  \
1                                                     0
2                                                     1
3                                                     1
4                                                     1
5                                                     0
..                                                  …
60                                                    0
61                                                    0
62                                                    1
63                                                    1
64                                                    0

    Did you participate in any extra curricular activities while you were
studying in NSU?  \
1                                                     1
2                                                     1
3                                                     1
4                                                     1
5                                                     0
..                                                  …
60                                                    0
61                                                    1
62                                                    0
63                                                    1
64                                                    1

    Were you interested in coding?  \
1                                   1
2                                   1
3                                   1
4                                   1
5                                   0
..                                …
60                                  0
61                                  1
62                                  1
63                                  1
64                                  1

    Did you participate in any competitive programming?  \
1                                                     1
2                                                     1
3                                                     1
4                                                     0
5                                                     0
..                                                  …
```

```
60                                                    0
61                                                    1
62                                                    1
63                                                    0
64                                                    0

    Did you start working as a software engineer after graduation?  \
1                                                     0
2                                                     0
3                                                     1
4                                                     0
5                                                     0
..                                                  ...
60                                                    0
61                                                    1
62                                                    1
63                                                    1
64                                                    0

    Did you have any Start-ups after graduation?
1                                                   0
2                                                   0
3                                                   0
4                                                   0
5                                                   0
..                                                ...
60                                                  0
61                                                  0
62                                                  0
63                                                  0
64                                                  0

[64 rows x 7 columns]
```

[ ]:

6.2 We will do Ordinal Encoding for these columns

Because the data of these columns maintains a sort of order

[39]:
```
print(f"Cumulative Grade Point Average(CGPA) column Values : {data['Cumulative␣
 ↪Grade Point Average(CGPA)'].unique()}\n")
print(f"Number of publications column Values : {data['Number of publications'].
 ↪unique()}\n")
print(f"What was your starting salary range? Column Values : {data['What was␣
 ↪your starting salary range?'].unique()}\n")
```

```
Cumulative Grade Point Average(CGPA) column Values : ['3 to < 3.5' '3.5 to 4'
```

```
'2.5 to < 3' '< 2']

Number of publications column Values : ['1 - 5' 'More than 10' '6-10' '0']

What was your starting salary range? Column Values : ['BDT 15K - 25K' '> 25K -
40K' '> 40K' 'N/A']
```

[40]:
```
columns_to_encode = [
    'Cumulative Grade Point Average(CGPA)',
    'Number of publications',
    'What was your starting salary range?'
]
encoded_ordinal = data[columns_to_encode]
encoded_ordinal.head(5)
```

[40]:
```
  Cumulative Grade Point Average(CGPA) Number of publications  \
1                           3 to < 3.5                   1 - 5
2                             3.5 to 4           More than 10
3                             3.5 to 4                   1 - 5
4                             3.5 to 4                    6-10
5                             3.5 to 4                   1 - 5

  What was your starting salary range?
1                        BDT 15K - 25K
2                          > 25K - 40K
3                                > 40K
4                        BDT 15K - 25K
5                        BDT 15K - 25K
```

[41]:
```
data['Number of publications'].unique()
```

[41]:
```
array(['1 - 5', 'More than 10', '6-10', '0'], dtype=object)
```

[42]:
```
cgpa_rank_mapping = {
    '3.5 to 4': 3,
    '3 to < 3.5': 2,
    '2.5 to < 3': 1,
    '< 2': 0
}
encoded_ordinal['Cumulative Grade Point Average(CGPA)'] =␣
 ↪encoded_ordinal['Cumulative Grade Point Average(CGPA)'].
 ↪map(cgpa_rank_mapping)
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1535697869.py:7:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_ordinal['Cumulative Grade Point Average(CGPA)'] =
encoded_ordinal['Cumulative Grade Point Average(CGPA)'].map(cgpa_rank_mapping)
```

[43]:
```
publication_mapping = {
    'More than 10': 3,
    '6-10': 2,
    '1 - 5': 1,
    '0': 0
}
encoded_ordinal['Number of publications'] = encoded_ordinal['Number of␣
 ↪publications'].map(publication_mapping)
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\2042818071.py:7:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_ordinal['Number of publications'] = encoded_ordinal['Number of
publications'].map(publication_mapping)
```

[44]:
```
salary_mapping = {
    '> 40K': 3,
    '> 25K - 40K': 2,
    'BDT 15K - 25K': 1,
    'N/A': 0
}
encoded_ordinal['What was your starting salary range?'] = encoded_ordinal['What␣
 ↪was your starting salary range?'].map(salary_mapping)
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\1304109159.py:7:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_ordinal['What was your starting salary range?'] =
encoded_ordinal['What was your starting salary range?'].map(salary_mapping)
```

[45]:
```
encoded_ordinal.head(5)
```

[45]:

	Cumulative Grade Point Average(CGPA)	Number of publications	\
1	2	1	
2	3	3	

```
3                                   3                                   1
4                                   3                                   2
5                                   3                                   1

   What was your starting salary range?
1                                   1
2                                   2
3                                   3
4                                   1
5                                   1
```

[46]:
```
encoded_ordinal.isnull().sum()
```

[46]:
```
Cumulative Grade Point Average(CGPA)    0
Number of publications                  0
What was your starting salary range?    0
dtype: int64
```

[47]:
```
encoded_ordinal['Number of publications'].unique()
```

[47]:
```
array([1, 3, 2, 0], dtype=int64)
```

6.2.1 For 'Programming languages you know', 'Frameworks you know' and 'Fields of research interest' We will do One Hot Encoding

[48]:
```
print(f"Programming languages you know column Values : {data['Programming␣
 ↪languages you know'].unique()}\n")
print(f"Frameworks you know column Values : {data['Frameworks you know'].
 ↪unique()}\n")
print(f"Fields of research interest column Values : {data['Fields of research␣
 ↪interest'].unique()}\n")
```

```
Programming languages you know column Values : ['C, C++, HTML' 'C, Java, Python,
C++, R'
 'C, Python, PHP, C++, HTML, CSS, JavaScript, R, SQL, NoSQL'
 'C, Python, R, SQL' 'C, HTML, CSS, SQL' 'C, Java, Python'
 'C, Java, C++, HTML, CSS, JavaScript, SQL' 'C, Python, HTML'
 'C, Python, HTML, CSS, SQL' 'Python' 'C' 'C, Python'
 'C, Python, HTML, CSS, JavaScript, SQL, NoSQL' 'C, Python, SQL'
 'C, Java, Python, PHP, C++, HTML, SQL'
 'C, Java, Python, C++, HTML, CSS, JavaScript, R, SQL' 'C, Python, C++'
 'C, Java, Python, PHP, C++, HTML, CSS, JavaScript, SQL, NoSQL'
 'C, Java, Python, PHP, HTML, CSS, JavaScript, NoSQL'
 'C, Java, Python, PHP, C++, HTML, CSS, JavaScript, SQL'
 'C, Java, Python, PHP, HTML, CSS, SQL'
 'C, Java, Python, C++, JavaScript, SQL' 'C, Java, Python, C++, SQL' 'SQL'
 'Not interested' 'C, Java, C++, HTML, CSS, SQL'
 'C, Python, HTML, CSS, JavaScript' 'Java, Python, HTML, JavaScript'
```

```
 'C, Java, C++, HTML, CSS, JavaScript'
 'C, Java, Python, PHP, HTML, CSS, JavaScript, SQL'
 'C, Java, Python, C++, SQL, NoSQL'
 'C, Java, Python, PHP, C++, HTML, CSS, SQL'
 'C, Python, PHP, C++, HTML, CSS, JavaScript, SQL, NoSQL'
 'C, Java, Python, C++, HTML, CSS, JavaScript, SQL' 'C++, HTML, SQL'
 'C, Java, Python, C++, HTML, SQL' 'Java, Python' 'C, Python, HTML, SQL'
 'Java, Python, HTML, CSS, JavaScript, SQL']

Frameworks you know column Values : ['Not interested' 'Laravel' 'React, Django,
Laravel, Node.js' 'Django'
 'jQuery' 'Node.js, Express.js' 'React, Django, Laravel, jQuery, Node.js'
 'React, Node.js, Express.js' 'Django, Laravel'
 'Angular, React, Vue.js, Django, jQuery, Node.js, Express.js'
 'Angular, React, jQuery, Node.js, Express.js' 'React, Laravel'
 'React, jQuery, Node.js' 'React, Node.js, .NET Core' '.NET Core'
 'React, Django, Laravel, jQuery, Node.js, Express.js'
 'Angular, React, Django, Laravel, Node.js' 'ASP.NET'
 'React, Django, ASP.NET, .NET Core' 'Vue.js, Django, Node.js'
 'React, Vue.js, Django, Laravel, jQuery, Node.js, Express.js'
 'React, Node.js' 'React, Django, Node.js' 'Django, Node.js'
 'Ruby on Rails, React, Node.js'
 'React, Vue.js, Django, Node.js, Express.js, .NET Core'
 'Angular, Ruby on Rails, Vue.js, ASP.NET'
 'Angular, Django, Laravel, jQuery, .NET Core' 'Vue.js'
 'Angular, ASP.NET, .NET Core']

Fields of research interest column Values : ['Internet of Things(IoT) and
Embedded Systems, Robotics'
 'Machine Learning, Artificial Intelligence, Natural Language Processing/
Digital Image Processing, Robotics'
 'Human-Computer interaction'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Natural
Language Processing/ Digital Image Processing'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence'
 'Machine Learning, Cryptography'
 'Internet of Things(IoT) and Embedded Systems'
 'Not interested in Research'
 'Machine Learning, Quantum computing, Natural Language Processing/ Digital
Image Processing'
 'Machine Learning'
 'Machine Learning, Data Science/Data mining, Internet of Things(IoT) and
Embedded Systems, Natural Language Processing/ Digital Image Processing,
Algorithm'
 'Machine Learning, Artificial Intelligence'
 'Machine Learning, Artificial Intelligence, Internet of Things(IoT) and
Embedded Systems'
 'Machine Learning, Artificial Intelligence, Internet of Things(IoT) and
```

```
Embedded Systems, Robotics'
 'Data Science/Data mining, Network Security'
 'Machine Learning, Artificial Intelligence, Natural Language Processing/
Digital Image Processing'
 'Machine Learning, Natural Language Processing/ Digital Image Processing'
 'Data Science/Data mining' 'Network Security'
 'Data Science/Data mining, Artificial Intelligence, Internet of Things(IoT) and
Embedded Systems'
 'Machine Learning, Data Science/Data mining, Quantum computing, Artificial
Intelligence, Natural Language Processing/ Digital Image Processing, Algorithm,
Cloud Computing and Distributed Systems, Robotics'
 'Artificial Intelligence'
 'Machine Learning, Artificial Intelligence, Natural Language Processing/
Digital Image Processing, Cloud Computing and Distributed Systems'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Network
Security'
 'Quantum computing, Artificial Intelligence'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Human-
Computer interaction, Natural Language Processing/ Digital Image Processing,
Robotics'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Network
Security, Natural Language Processing/ Digital Image Processing'
 'Machine Learning, Artificial Intelligence, Cryptography'
 'Machine Learning, Data Science/Data mining'
 'Data Science/Data mining, Network Security, Cloud Computing and Distributed
Systems'
 'Natural Language Processing/ Digital Image Processing'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Internet
of Things(IoT) and Embedded Systems, Natural Language Processing/ Digital Image
Processing'
 'Machine Learning, Data Science/Data mining, Artificial Intelligence, Human-
Computer interaction, Natural Language Processing/ Digital Image Processing']
```

6.3 Performing One Hot Encoding

It takes the multiple choice answers from the user and based on the answers we will create new columns for each of the options ticked and mark as yes or no(1 or 0) for those options/columns

[49]:
```
df_PL_encoded = data['Programming languages you know'].str.get_dummies(', ')
print(f"Programming languages you know dataframe Columns: {df_PL_encoded.
  ↪columns}\nColumn Length : {len(df_PL_encoded.columns)}\n")
df_PL_encoded.head(15)
```

```
Programming languages you know dataframe Columns: Index(['C', 'C++', 'CSS',
'HTML', 'Java', 'JavaScript', 'NoSQL',
       'Not interested', 'PHP', 'Python', 'R', 'SQL'],
```

```
      dtype='object')
Column Length : 12
```

[49]:

	C	C++	CSS	HTML	Java	JavaScript	NoSQL	Not interested	PHP	Python	\
1	1	1	0	1	0	0	0	0	0	0	
2	1	1	0	0	1	0	0	0	0	1	
3	1	1	1	1	0	1	1	0	1	1	
4	1	0	0	0	0	0	0	0	0	1	
5	1	0	1	1	0	0	0	0	0	0	
6	1	0	0	0	1	0	0	0	0	1	
7	1	1	1	1	1	1	0	0	0	0	
8	1	0	0	1	0	0	0	0	0	1	
9	1	0	1	1	0	0	0	0	0	1	
10	0	0	0	0	0	0	0	0	0	1	
11	1	0	0	0	0	0	0	0	0	0	
12	1	0	0	0	0	0	0	0	0	1	
13	1	0	1	1	0	1	1	0	0	1	
14	1	0	0	0	0	0	0	0	0	1	
15	1	1	0	1	1	0	0	0	1	1	

	R	SQL
1	0	0
2	1	0
3	1	1
4	1	1
5	0	1
6	0	0
7	0	1
8	0	0
9	0	1
10	0	0
11	0	0
12	0	0
13	0	1
14	0	1
15	0	1

[50]:
```
df_FoR_encoded = data['Fields of research interest'].str.get_dummies(', ')
print(f"Fields of research interest dataframe Columns: {df_FoR_encoded.
  ↪columns}\n\nColumn Length : {len(df_FoR_encoded.columns)}\n")
df_FoR_encoded.head(15)
```

```
Fields of research interest dataframe Columns: Index(['Algorithm', 'Artificial
Intelligence',
       'Cloud Computing and Distributed Systems', 'Cryptography',
       'Data Science/Data mining', 'Human-Computer interaction',
       'Internet of Things(IoT) and Embedded Systems', 'Machine Learning',
```

```
       'Natural Language Processing/ Digital Image Processing',
       'Network Security', 'Not interested in Research', 'Quantum computing',
       'Robotics'],
      dtype='object')

Column Length : 13
```

[50]:

	Algorithm	Artificial Intelligence	Cloud Computing and Distributed Systems	Cryptography	Data Science/Data mining	Human-Computer interaction
1	0	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	0	0	0	1
4	0	1	0	0	1	0
5	0	1	0	0		
6	0	0	0	1		
7	0	1	0	0		
8	0	0	0	0		
9	0	0	0	0		
10	0	0	0	0		
11	0	0	0	0		
12	0	0	0	0		
13	0	0	0	0		
14	0	0	0	0		
15	0	1	0	0		

5	1	0
6	0	0
7	1	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	1	0

	Internet of Things(IoT) and Embedded Systems	Machine Learning	\
1	1	0	
2	0	1	
3	0	0	
4	0	1	
5	0	1	
6	0	1	
7	0	1	
8	1	0	
9	0	0	
10	0	0	
11	0	0	
12	0	1	
13	0	1	
14	0	0	
15	0	1	

	Natural Language Processing/ Digital Image Processing	Network Security	\
1	0	0	
2	1	0	
3	0	0	
4	1	0	
5	0	0	
6	0	0	
7	0	0	
8	0	0	
9	0	0	
10	0	0	
11	0	0	
12	1	0	
13	0	0	
14	0	0	
15	1	0	

	Not interested in Research	Quantum computing	Robotics

1	0	0	1
2	0	0	1
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	1	0	0
10	1	0	0
11	1	0	0
12	0	1	0
13	0	0	0
14	1	0	0
15	0	0	0

[51]:
```
df_FW_encoded = data['Frameworks you know'].str.get_dummies(', ')
print(f"Frameworks you know dataframe Columns: {df_FW_encoded.
  ↪columns}\n\nColumn Length : {len(df_FW_encoded.columns)}\n")
df_FW_encoded.head(15)
```

```
Frameworks you know dataframe Columns: Index(['.NET Core', 'ASP.NET', 'Angular',
'Django', 'Express.js', 'Laravel',
       'Node.js', 'Not interested', 'React', 'Ruby on Rails', 'Vue.js',
       'jQuery'],
      dtype='object')

Column Length : 12
```

[51]:

	.NET Core	ASP.NET	Angular	Django	Express.js	Laravel	Node.js	\
1	0	0	0	0	0	0	0	
2	0	0	0	0	0	1	0	
3	0	0	0	1	0	1	1	
4	0	0	0	1	0	0	0	
5	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	
7	0	0	0	0	1	0	1	
8	0	0	0	0	0	0	0	
9	0	0	0	1	0	1	1	
10	0	0	0	0	0	0	0	
11	0	0	0	0	0	0	0	
12	0	0	0	0	0	0	0	
13	0	0	0	0	1	0	1	
14	0	0	0	1	0	0	0	
15	0	0	0	1	0	0	0	

	Not interested	React	Ruby on Rails	Vue.js	jQuery

1	1	0	0	0	0
2	0	0	0	0	0
3	0	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	1
6	1	0	0	0	0
7	0	0	0	0	0
8	1	0	0	0	0
9	0	1	0	0	1
10	1	0	0	0	0
11	1	0	0	0	0
12	1	0	0	0	0
13	0	1	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0

6.4 We will be removing the columns which has a lot of different data and the columns that are not important

[52]:
```
# data['What was the name of the company you worked for right after your␣
↪graduation? if your answer is not write N/A.'].unique()
```

[53]:
```
# data['What was the name of the University you completed your higher studies␣
↪from? if answer is no write N/A.'].unique()
```

7 Nominal Encoding

[54]:
```
print(f"Where did you start working after completing your higher studies?␣
↪Column Values : {data['Where did you start working after completing your␣
↪higher studies?'].unique()}\n")
print(f"Which country did you go to for higher studies? Column Values :␣
↪{data['Which country did you go to for higher studies?'].unique()}\n")
print(f"'What was the name of the University you completed your higher studies␣
↪from? if answer is no write N/A.' Column Values : {data['What was the name␣
↪of the University you completed your higher studies from? if answer is no␣
↪write N/A.'].unique()}")
```

```
Where did you start working after completing your higher studies? Column Values
: ['Industry' 'Not working in any of these' 'Educational institute']

Which country did you go to for higher studies? Column Values : ['United States
of America (USA)' 'Germany' 'Not yet started' 'Canada'
 'United Kingdom' 'Australia' 'Other']

'What was the name of the University you completed your higher studies from? if
answer is no write N/A.' Column Values : ['Florida Atlantic University'
'University of Wisconsin-Milwaukee'
```

```
 'Marquette University ' 'University of Jena (Abbe School of Photonics)'
 'N/A' 'Memorial University ' 'U of Rochester '
 'University of British Columbia'
 'Darmstadt University of Applied Sciences' 'North south univeristy'
 'University of Hertfordshire' 'Ryerson university '
 'North South University' 'Concordia University ']
```

[55]:
```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
columns_to_encode = [
    'Where did you start working after completing your higher studies?',
    'Which country did you go to for higher studies?',
    'What was the name of the University you completed your higher studies from?
 ↪ if answer is no write N/A.'
]
encoded_Dependable = data[columns_to_encode]
for column in columns_to_encode:
    encoded_Dependable[column] = label_encoder.
 ↪fit_transform(encoded_Dependable[column])
encoded_Dependable
```

```
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\3544060623.py:10:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_Dependable[column] =
label_encoder.fit_transform(encoded_Dependable[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\3544060623.py:10:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_Dependable[column] =
label_encoder.fit_transform(encoded_Dependable[column])
C:\Users\nahia\AppData\Local\Temp\ipykernel_12048\3544060623.py:10:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  encoded_Dependable[column] =
label_encoder.fit_transform(encoded_Dependable[column])
```

[55]:
```
    Where did you start working after completing your higher studies?  \
1                                                                       1
2                                                                       2
3                                                                       2
4                                                                       0
5                                                                       2
..                                                                    …
60                                                                      1
61                                                                      1
62                                                                      0
63                                                                      2
64                                                                      2

    Which country did you go to for higher studies?  \
1                                                   6
2                                                   6
3                                                   6
4                                                   2
5                                                   3
..                                                …
60                                                  3
61                                                  3
62                                                  1
63                                                  3
64                                                  6

    What was the name of the University you completed your higher studies from?
if answer is no write N/A.
1                                                    2
2                                                   13
3                                                    3
4                                                   12
5                                                    5
..                                                 …
60                                                   5
61                                                   5
62                                                   5
63                                                   5
64                                                   5

[64 rows x 3 columns]
```

[56]:
```
data
```

[56]:
```
    Gender Cumulative Grade Point Average(CGPA)  \
1     Male                          3 to < 3.5
2     Male                            3.5 to 4
```

```
3   Female                              3.5 to 4
4     Male                              3.5 to 4
5   Female                              3.5 to 4
..     …                                   …
60    Male                              3.5 to 4
61    Male                              3.5 to 4
62    Male                              3.5 to 4
63    Male                              3.5 to 4
64    Male                            3 to < 3.5

   Total year now since your graduation  \
1                                      1
2                                      4
3                                      3
4                                      8
5                                      0
..                                   …
60                                     1
61                                     5
62                                     1
63                                     0
64                                     0

                          Fields of research interest Number of publications  \
1   Internet of Things(IoT) and Embedded Systems, …                  1 - 5
2   Machine Learning, Artificial Intelligence, Nat…           More than 10
3                           Human-Computer interaction                 1 - 5
4   Machine Learning, Data Science/Data mining, Ar…                   6-10
5   Machine Learning, Data Science/Data mining, Ar…                  1 - 5
..                                                 …                    …
60                                  Machine Learning                      0
61                          Human-Computer interaction                    0
62  Machine Learning, Data Science/Data mining, Ar…                     0
63  Machine Learning, Data Science/Data mining, Ar…                  1 - 5
64                             Artificial Intelligence                    0

   Were you a teaching assistant (TA) or research assistant(RA) during your time
in NSU?  \
1                                                   No
2                                                  Yes
3                                                  Yes
4                                                  Yes
5                                                   No
..                                                 …
60                                                  No
61                                                  No
62                                                 Yes
```

```
63                                                  Yes
64                                                   No

   Did you participate in any extra curricular activities while you were
studying in NSU?  \
1                                                   Yes
2                                                   Yes
3                                                   Yes
4                                                   Yes
5                                                    No
..                                                  …
60                                                   No
61                                                  Yes
62                                                   No
63                                                  Yes
64                                                  Yes

   Were you interested in coding?  \
1                              Yes
2                              Yes
3                              Yes
4                              Yes
5                               No
..                             …
60                              No
61                             Yes
62                             Yes
63                             Yes
64                             Yes

   Did you participate in any competitive programming?  \
1                                                   Yes
2                                                   Yes
3                                                   Yes
4                                                    No
5                                                    No
..                                                  …
60                                                   No
61                                                  Yes
62                                                  Yes
63                                                   No
64                                                   No

                       Programming languages you know  \
1                                          C, C++, HTML
2                                C, Java, Python, C++, R
3   C, Python, PHP, C++, HTML, CSS, JavaScript, R,…
```

```
4                                   C, Python, R, SQL
5                                   C, HTML, CSS, SQL
..                                                …
60                                C, Python, HTML, SQL
61                         C, Java, C++, HTML, CSS, SQL
62                            C, Java, Python, C++, SQL
63           Java, Python, HTML, CSS, JavaScript, SQL
64                                               Python

                Frameworks you know  \
1                        Not interested
2                               Laravel
3    React, Django, Laravel, Node.js
4                                Django
5                                jQuery
..                                  …
60                       Not interested
61        Angular, ASP.NET, .NET Core
62                Node.js, Express.js
63         React, Node.js, Express.js
64                               Django

   Did you start working as a software engineer after graduation?  \
1                                                          No
2                                                          No
3                                                         Yes
4                                                          No
5                                                          No
..                                                        …
60                                                         No
61                                                        Yes
62                                                        Yes
63                                                        Yes
64                                                         No

   What was the name of the company you worked for right after your graduation?
if your answer is not write N/A.  \
1                      Venturous Professional Institute
2                              AST International Limited
3                Banglalink (Internship), Therap BD Ltd
4                                  North South University
5                                     Catalyst Solutions
..                                                      …
60                                                    N/A
61                  Millenium Information Solution Ltd.
62                                                     Na
63                                                 Sazim
```

```
64                                                  N/A

   What was your starting salary range?  \
1                          BDT 15K - 25K
2                            > 25K - 40K
3                                  > 40K
4                          BDT 15K - 25K
5                          BDT 15K - 25K
..                                     …
60                         BDT 15K - 25K
61                                   N/A
62                                   N/A
63                           > 25K - 40K
64                         BDT 15K - 25K

   Did you have any Start-ups after graduation?  \
1                                              No
2                                              No
3                                              No
4                                              No
5                                              No
..                                            …
60                                             No
61                                             No
62                                             No
63                                             No
64                                             No

   Which country did you go to for higher studies?  \
1                    United States of America (USA)
2                    United States of America (USA)
3                    United States of America (USA)
4                                          Germany
5                                  Not yet started
..                                             …
60                                 Not yet started
61                                 Not yet started
62                                          Canada
63                                 Not yet started
64                   United States of America (USA)

   What was the name of the University you completed your higher studies from?
if answer is no write N/A.  \
1                           Florida Atlantic University
2                    University of Wisconsin-Milwaukee
3                                 Marquette University
4       University of Jena (Abbe School of Photonics)
```

```
5                                                   N/A
..                                                  …
60                                                  N/A
61                                                  N/A
62                                                  N/A
63                                                  N/A
64                                                  N/A

   Where did you start working after completing your higher studies?
1                                            Industry
2                        Not working in any of these
3                        Not working in any of these
4                              Educational institute
5                        Not working in any of these
..                                                  …
60                                           Industry
61                                           Industry
62                             Educational institute
63                       Not working in any of these
64                       Not working in any of these

[64 rows x 18 columns]
```

[57]:
```
print(f"'Where did you start working after completing your higher studies?'␣
 ↪Column Values : {data['Where did you start working after completing your␣
 ↪higher studies?'].unique()}\n")

work = data['Where did you start working after completing your higher studies?']
work = pd.get_dummies(work, drop_first = False).astype(int)
work.head()
```

```
'Where did you start working after completing your higher studies?' Column
Values : ['Industry' 'Not working in any of these' 'Educational institute']
```

[57]:

	Educational institute	Industry	Not working in any of these
1	0	1	0
2	0	0	1
3	0	0	1
4	1	0	0
5	0	0	1

[58]:
```
print(f"Which country did you go to for higher studies? Column Values :␣
 ↪{data['Which country did you go to for higher studies?'].unique()}\n")

higherStudies = data["Which country did you go to for higher studies?"]
higherStudies = pd.get_dummies(higherStudies, drop_first = True).astype(int)
higherStudies.head()
```

```
Which country did you go to for higher studies? Column Values : ['United States
of America (USA)' 'Germany' 'Not yet started' 'Canada'
 'United Kingdom' 'Australia' 'Other']
```

[58]:
```
   Canada  Germany  Not yet started  Other  United Kingdom  \
1       0        0                0      0               0
2       0        0                0      0               0
3       0        0                0      0               0
4       0        1                0      0               0
5       0        0                1      0               0

   United States of America (USA)
1                               1
2                               1
3                               1
4                               0
5                               0
```

[59]:
```
print(f"'What was the name of the University you completed your higher studies␣
 ↪from? if answer is no write N/A.' Column Values : {data['What was the name␣
 ↪of the University you completed your higher studies from? if answer is no␣
 ↪write N/A.'].unique()}")

higherStudiesUniversity = data['What was the name of the University you␣
 ↪completed your higher studies from? if answer is no write N/A.']
higherStudiesUniversity = pd.get_dummies(higherStudiesUniversity, drop_first =␣
 ↪True).astype(int)
higherStudiesUniversity.head()
```

```
'What was the name of the University you completed your higher studies from? if
answer is no write N/A.' Column Values : ['Florida Atlantic University'
'University of Wisconsin-Milwaukee'
 'Marquette University ' 'University of Jena (Abbe School of Photonics)'
 'N/A' 'Memorial University ' 'U of Rochester '
 'University of British Columbia'
 'Darmstadt University of Applied Sciences' 'North south univeristy'
 'University of Hertfordshire' 'Ryerson university '
 'North South University' 'Concordia University ']
```

[59]:
```
   Darmstadt University of Applied Sciences  Florida Atlantic University  \
1                                         0                            1
2                                         0                            0
3                                         0                            0
4                                         0                            0
5                                         0                            0

   Marquette University   Memorial University   N/A  North South University  \
```

```
1                       0                       0    0                       0
2                       0                       0    0                       0
3                       1                       0    0                       0
4                       0                       0    0                       0
5                       0                       0    1                       0

   North south univeristy  Ryerson university   U of Rochester   \
1                       0                       0                0
2                       0                       0                0
3                       0                       0                0
4                       0                       0                0
5                       0                       0                0

   University of British Columbia  University of Hertfordshire  \
1                               0                             0
2                               0                             0
3                               0                             0
4                               0                             0
5                               0                             0

   University of Jena (Abbe School of Photonics)  \
1                                              0
2                                              0
3                                              0
4                                              1
5                                              0

   University of Wisconsin-Milwaukee
1                                  0
2                                  1
3                                  0
4                                  0
5                                  0
```

7.1 So We end up with 6 new datasets

1. **encoded_binary {Values with yes/no, Male/Female}**
    1. 'Gender'
    2. 'Were you a teaching assistant (TA) or research assistant(RA) during your time in NSU?'
    3. 'Did you participate in any extra curricular activities while you were studying in NSU?'
    4. 'Were you interested in coding?'
    5. 'Did you participate in any competitive programming?'
    6. 'Did you start working as a software engineer after graduation?'
    7. 'Did you have any Start-ups after graduation?'
2. **encoded_ordinal {Values that maintains an order}**
    1. 'Cumulative Grade Point Average(CGPA)',

2. 'Number of publications',
3. 'What was your starting salary range?'

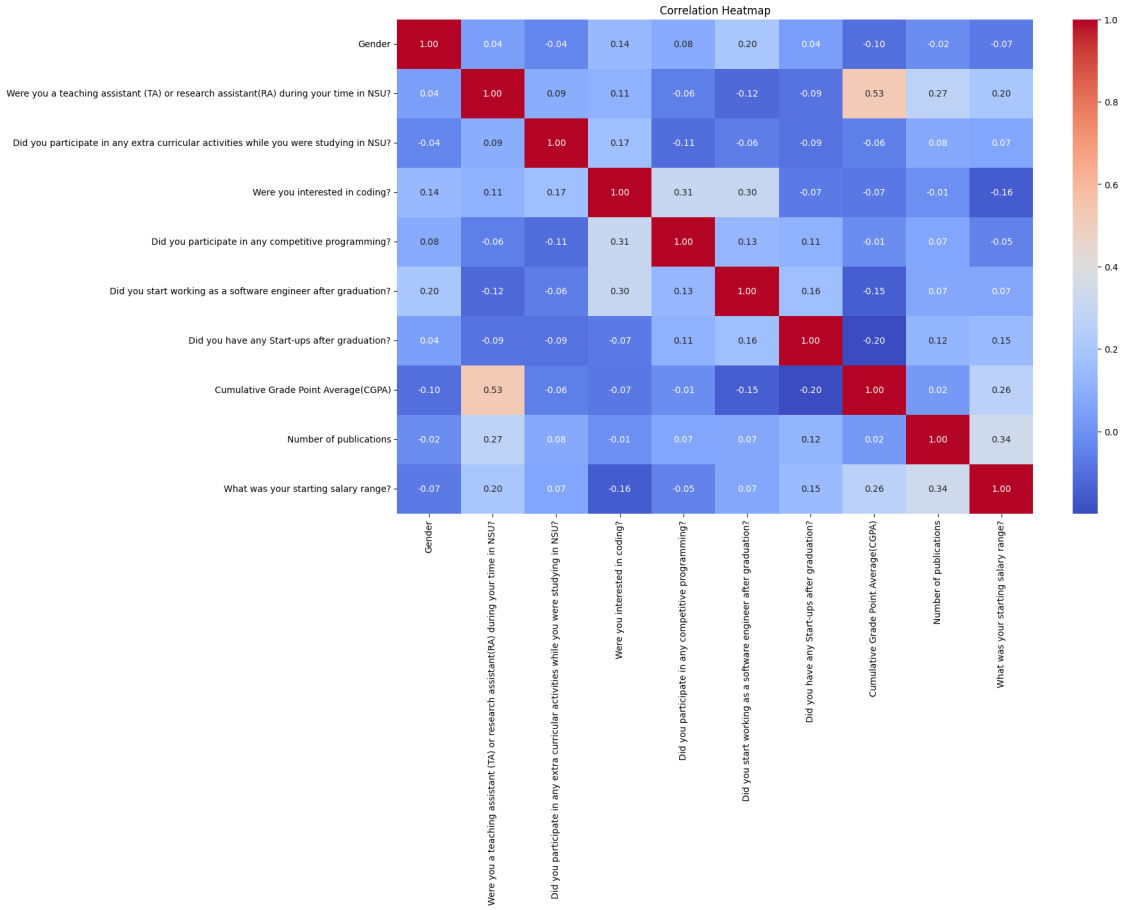
3. **df_PL_encoded {Programing Languages You know} [One Hot Encoding]**
4. **df_FoR_encoded {Fields of Research Interest} [One Hot Encoding]**
5. **df_FW_encoded {FrameWorks You know} [One Hot Encoding]**
6. **encoded_Dependable {Columns that maybe used as target variable}**
    1. 'Where did you start working after completing your higher studies?',
    2. 'Which country did you go to for higher studies?',
    3. 'What was the name of the University you completed your higher studies from? if answer is no write N/A.'

8 Question 7 : Correlation

8.1 Now we can concat each of the datasets and show the correlation

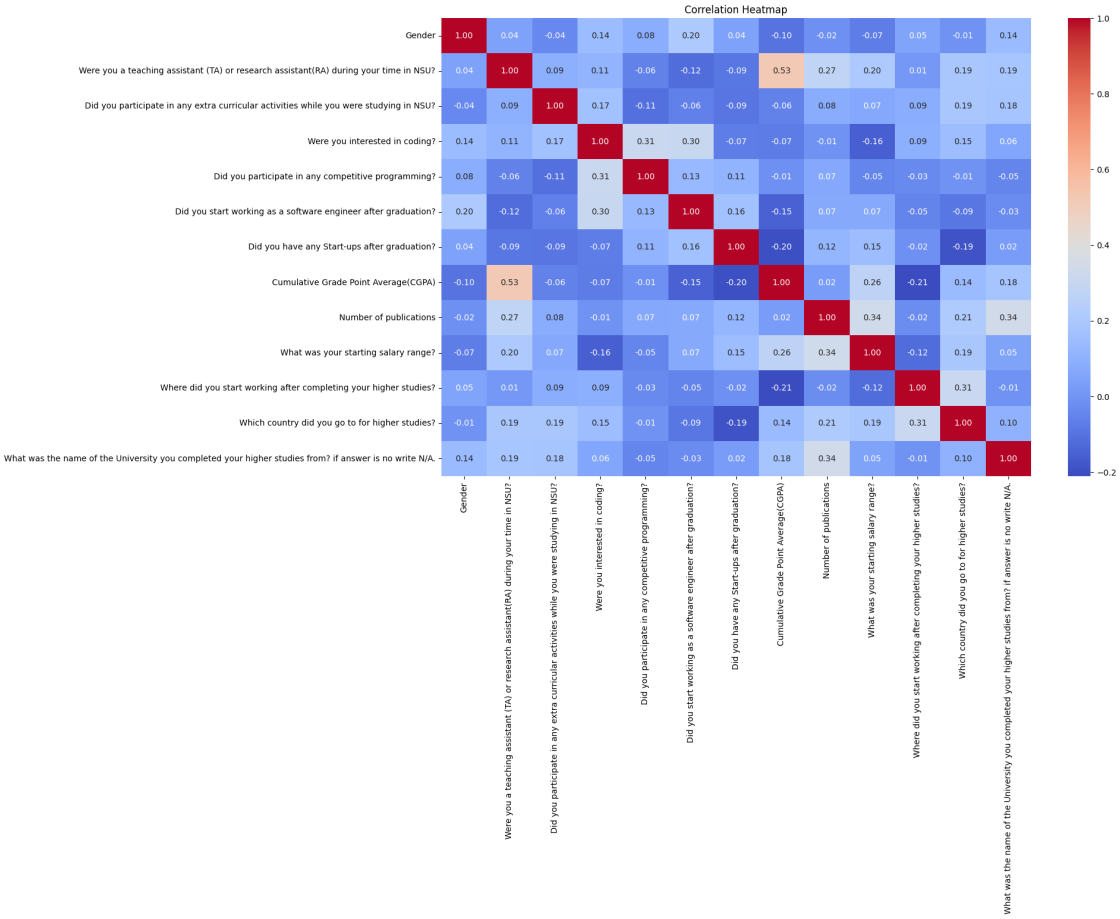
8.1.1 First we are concating the encoded_binary and encoded_ordinal Datasets

[60]:
```
firstDF = pd.concat([encoded_binary,encoded_ordinal], axis=1)
correlation_matrix = firstDF.corr(numeric_only=True)

plt.figure(figsize=(16, 10))  # Set the figure size
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



8.1.2 Now we are concating encoded_binary,encoded_ordinal and encoded_Dependable

[61]:
```
secondDF = pd.concat([encoded_binary,encoded_ordinal ,encoded_Dependable ],␣
↪axis=1)
correlation_matrix = secondDF.corr(numeric_only=True)

plt.figure(figsize=(16, 10))  # Set the figure size
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



[62]:
```
csv_file_path = 'encBin-endOrd-encDep.csv'

# Save the DataFrame to a CSV file
secondDF.to_csv(csv_file_path, index=False)
```

[63]:
```
secondDF.isnull().sum()
```

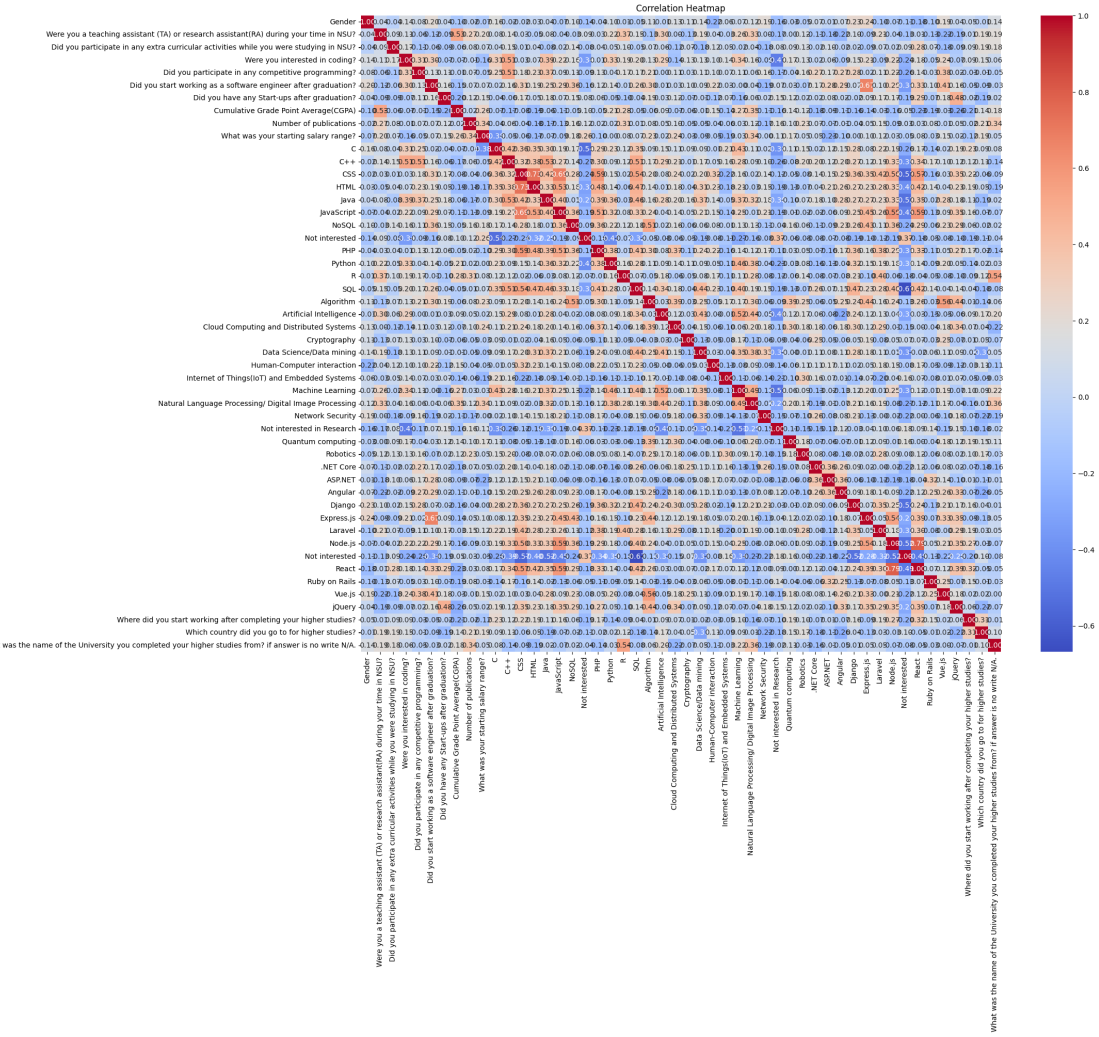
[63]:
```
Gender
0
Were you a teaching assistant (TA) or research assistant(RA) during your time in
NSU?                    0
Did you participate in any extra curricular activities while you were studying
in NSU?                    0
Were you interested in coding?
0
Did you participate in any competitive programming?
0
Did you start working as a software engineer after graduation?
```

```
0
Did you have any Start-ups after graduation?
0
Cumulative Grade Point Average(CGPA)
0
Number of publications
0
What was your starting salary range?
0
Where did you start working after completing your higher studies?
0
Which country did you go to for higher studies?
0
What was the name of the University you completed your higher studies from? if
answer is no write N/A.    0
dtype: int64
```

8.2 Now we concate all the datasets together

[64]:
```
concatenat = pd.
 ↪concat([encoded_binary,encoded_ordinal,df_PL_encoded,df_FoR_encoded,␣
 ↪df_FW_encoded, encoded_Dependable], axis=1)
# concatenat.reset_index(drop=True, inplace=True)
```

[65]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming 'concatenated_df' is your concatenated DataFrame

# Calculate the correlation matrix
correlation_matrix = concatenat.corr()

# Create a heatmap
plt.figure(figsize=(20, 16))  # Set the figure size
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```


Correlation Heatmap
Gender
Were you a teaching assistant (TA) or research assistant(RA) during your time in NSU?
Did you participate in any extra curricular activities while you were studying in NSU?
Were you interested in coding?
Did you participate in any competitive programming?
Did you start working as a software engineer after graduation?
Did you have any Start-ups after graduation?
Cumulative Grade Point Average(CGPA)
Number of publications
What was your starting salary range?
C
C++
CSS
HTML
Java
JavaScript
NoSQL
Not interested
PHP
Python
R
SQL
Algorithm
Artificial Intelligence
Cloud Computing and Distributed Systems
Cryptography
Data Science/Data mining
Human-Computer interaction
Internet of Things(IoT) and Embedded Systems
Machine Learning
Natural Language Processing/ Digital Image Processing
Network Security
Not interested in Research
Quantum computing
Robotics
.NET Core
ASP.NET
Angular
Django
Express.js
Laravel
Node.js
Not interested
React
Ruby on Rails
Vue.js
jQuery
Where did you start working after completing your higher studies?
Which country did you go to for higher studies?
What was the name of the University you completed your higher studies from? if answer is no write N/A
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

9 Question 6. Develop necessary hypothesis that might support your understanding on the developed datasets.

10 Answer:

10.0.1 1. Students interested in coding are more likely to participate in competitive programming.

10.0.2 2. Proficiency in certain programming languages is correlated with better job placement.

10.0.3 3. Knowledge of certain frameworks is associated with industry employment.

10.0.4 4. Students who pursued higher studies in the USA are more likely to work in the industry.

10.0.5 5. Gender does not affect career choices after graduation.

10.0.6 6. Students with research interests in certain fields are more likely to be interested in coding.

[ ]:

[ ]:

[ ]: