```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import pandas as pd

# Try reading the CSV file with an alternative encoding
df = pd.read_csv("intern_details 20241226_RG.csv", encoding='ISO-8859-1')

# Check the first few rows of the DataFrame to ensure it loaded correctly
print(df.head())
```

```
        Id        Date  Status                              Name  \
0  2128156  12/24/2024  unread                         Vignesh K
1  2128008  12/24/2024  unread               Deekshitha Edukulla
2  2127964  12/23/2024  unread   AASHISH NIRANJAN BARATHYKANNAN
3  2127963  12/23/2024  unread                 Keerthi Shivampeta
4  2127962  12/23/2024  unread                 Keerthi Shivampeta

                   Interns Email                  Menu 23   Number
570  \
0         vikikannan8@gmail.com   AI(Artificial Intelligence
6369727236
1       deekshitha1745@gmail.com                       DevOps
9346380711
2      aashishniranjanb@gmail.com                 Data Science
6381477688
3   keerthishivampeta04@gmail.com                 Data Analyst
7032706017
4   keerthishivampeta04@gmail.com                 Data Analyst
7032706017

                   Menu 2 DegreeQualification
Branch  \
0  3 Months(NO STIPEND)         B.Tech/M.Tech
CSE(AIML)
1  3 Months(NO STIPEND)         B.Tech/M.Tech                       Data
science
2  3 Months(NO STIPEND)         B.Tech/M.Tech   ELECTRONICS AND
COMMUNICATION
3  3 Months(NO STIPEND)         B.Tech/M.Tech             CSE (DATA
SCIENCE)
4  3 Months(NO STIPEND)         B.Tech/M.Tech             CSE (DATA
SCIENCE)

     Experienceifany                              CollegeUniversity  \
0                 NaN  KPR Institute of engineering and technology
```

```
1                NaN          Institute of aeronautical engineering
2   INTERN AT CODSOFT       SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
3                 No              Malla reddy college of engineering
4                 No              Malla reddy college of engineering

              Location              City Internshipstartdate
Message
0  Work From Home(WFH)          Coimbatore             2/5/2025
NaN
1  Work From Home(WFH)  Hyderabad South East          3/12/2025
NaN
2  Work From Home(WFH)             CHENNAI            4/15/2025
NaN
3  Work From Home(WFH)           Hyderabad           2/24/2024
NaN
4  Work From Home(WFH)           Hyderabad           2/24/2024
NaN
```

Basic Statistic

```
df.describe()

           Id        Date   Status                  Name
Interns  Email  \
count    4908        4908     4908                  4908
4510
unique   4905         365        3                  2799
2533
top      ?Id   11/19/2024   unread         KULDEEP SINGH
singh.kuldeep22798@gmail.com
freq        4         139     4889                    31
51

            Menu 23   Number 570                  Menu 2
DegreeQualification  \
count              4908         4908                    4908
4908
unique               50         2690                       4
10
top      Web Development   6392033597   3 Months(NO STIPEND)
B.Tech/M.Tech
freq               1538           51                    4193
2942

     Branch Experienceifany
CollegeUniversity  \
count    4585          2674
4908
unique    881           537
2133
```

```
top         CSE              No  Trinity institute of professional
studies
freq       359             846
42

                   Location   City Internshipstartdate  \
count                  4891   4906                 4908
unique                    3   1002                  529
top      Work From Home(WFH)  Delhi           2/28/2025
freq                   4806    191                  157

                                                    Message
count                                                  1609
unique                                                  948
top      Thank you for considering my application. I lo...
freq                                                     23
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4908 entries, 0 to 4907
Data columns (total 16 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Id                 4908 non-null   object
 1   Date               4908 non-null   object
 2   Status             4908 non-null   object
 3   Name               4908 non-null   object
 4   Interns Email      4510 non-null   object
 5   Menu 23            4908 non-null   object
 6   Number 570         4908 non-null   object
 7   Menu 2             4908 non-null   object
 8   DegreeQualification 4908 non-null  object
 9   Branch             4585 non-null   object
 10  Experienceifany    2674 non-null   object
 11  CollegeUniversity  4908 non-null   object
 12  Location           4891 non-null   object
 13  City               4906 non-null   object
 14  Internshipstartdate 4908 non-null  object
 15  Message            1609 non-null   object
dtypes: object(16)
memory usage: 613.6+ KB
```

Check for duplicates-Identify and handle Duplicates.

```
df.duplicated().sum()
```

```
np.int64(3)
```

```python
df = df.drop_duplicates()  # Drop duplicates
df
```

```
          Id        Date  Status                             Name  \
0    2128156  12/24/2024  unread                        Vignesh K
1    2128008  12/24/2024  unread               Deekshitha Edukulla
2    2127964  12/23/2024  unread  AASHISH NIRANJAN BARATHYKANNAN
3    2127963  12/23/2024  unread               Keerthi Shivampeta
4    2127962  12/23/2024  unread               Keerthi Shivampeta
...      ...         ...     ...                              ...
4903     999   7/23/2023  unread                     Simran jaggi
4904     998   7/23/2023  unread                     Simran jaggi
4905     997   7/23/2023  unread                     Simran jaggi
4906     996   7/23/2023  unread                     Simran jaggi
4907     995   7/23/2023  unread                     Simran jaggi

                         Interns Email                      Menu 23
Number 570  \
0              vikikannan8@gmail.com   AI(Artificial Intelligence
6369727236
1            deekshitha1745@gmail.com                      DevOps
9346380711
2          aashishniranjanb@gmail.com                Data Science
6381477688
3       keerthishivampeta04@gmail.com                Data Analyst
7032706017
4       keerthishivampeta04@gmail.com                Data Analyst
7032706017
...                               ...                          ...
...
4903                              NaN           Digital Marketing
9899548755
4904                              NaN           Digital Marketing
9899548755
4905                              NaN           Digital Marketing
9899548755
4906                              NaN           Digital Marketing
9899548755
4907                              NaN           Digital Marketing
9899548755

                    Menu 2 DegreeQualification
Branch  \
0      3 Months(NO STIPEND)        B.Tech/M.Tech
CSE(AIML)
1      3 Months(NO STIPEND)        B.Tech/M.Tech                      Data
science
2      3 Months(NO STIPEND)        B.Tech/M.Tech  ELECTRONICS AND
COMMUNICATION
3      3 Months(NO STIPEND)        B.Tech/M.Tech             CSE (DATA
```

```
SCIENCE)
4       3 Months(NO STIPEND)       B.Tech/M.Tech              CSE (DATA
SCIENCE)
...                      ...                       ...
...
4903  3 Months(NO STIPEND)      Post-Graduation
Delhi
4904  3 Months(NO STIPEND)      Post-Graduation
Delhi
4905  3 Months(NO STIPEND)      Post-Graduation
Delhi
4906  3 Months(NO STIPEND)      Post-Graduation
Delhi
4907  3 Months(NO STIPEND)      Post-Graduation
Delhi

        Experienceifany
CollegeUniversity  \
0                    NaN  KPR Institute of engineering and technology

1                    NaN        Institute of aeronautical engineering

2     INTERN AT CODSOFT      SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

3                     No        Malla reddy college of engineering

4                     No        Malla reddy college of engineering

...                   ...                                       ...

4903                  NaN                            Iilm,lodhi road

4904                  NaN                            Iilm,lodhi road

4905                  NaN                            Iilm,lodhi road

4906                  NaN                            Iilm,lodhi road

4907                  NaN                            Iilm,lodhi road

              Location                City Internshipstartdate  \
0     Work From Home(WFH)          Coimbatore           2/5/2025
1     Work From Home(WFH)  Hyderabad South East         3/12/2025
2     Work From Home(WFH)             CHENNAI           4/15/2025
3     Work From Home(WFH)           Hyderabad           2/24/2024
4     Work From Home(WFH)           Hyderabad           2/24/2024
...                   ...                 ...                ...
4903                  NaN           New delhi          10/24/2023
4904                  NaN           New delhi          10/24/2023
```

```
4905                    NaN          New delhi          10/24/2023
4906                    NaN          New delhi          10/24/2023
4907                    NaN          New delhi          10/24/2023

                                                          Message
0                                                             NaN
1                                                             NaN
2                                                             NaN
3                                                             NaN
4                                                             NaN
...                                                           ...
4903   I want to join this internship as it will be a...
4904   I want to join this internship as it will be a...
4905                                                          NaN
4906                                                          NaN
4907                                                          NaN

[4905 rows x 16 columns]
```

Check unique values

```python
df['Location'].unique()
```

```
array(['Work From Home(WFH)', 'Select Location', 'Location', nan],
      dtype=object)
```

```python
df.isnull().sum()
```

```
Id                         0
Date                       0
Status                     0
Name                       0
Interns Email            398
Menu 23                    0
Number 570                 0
Menu 2                     0
DegreeQualification        0
Branch                   323
Experienceifany         2234
CollegeUniversity          0
Location                  17
City                       2
Internshipstartdate        0
Message                 3299
dtype: int64
```

```python
# Fill missing values in 'Interns Email' with a placeholder email
df['Interns Email'] = df['Interns
Email'].fillna('noemail@example.com')
```

```
# Check the result
print(df.head())

        Id        Date  Status                              Name  \
0  2128156  12/24/2024  unread                         Vignesh K
1  2128008  12/24/2024  unread               Deekshitha Edukulla
2  2127964  12/23/2024  unread   AASHISH NIRANJAN BARATHYKANNAN
3  2127963  12/23/2024  unread                 Keerthi Shivampeta
4  2127962  12/23/2024  unread                 Keerthi Shivampeta


                    Interns Email                 Menu 23  Number
570  \
0          vikikannan8@gmail.com  AI(Artificial Intelligence
6369727236
1        deekshitha1745@gmail.com                     DevOps
9346380711
2       aashishniranjanb@gmail.com               Data Science
6381477688
3  keerthishivampeta04@gmail.com              Data Analyst
7032706017
4  keerthishivampeta04@gmail.com              Data Analyst
7032706017


                  Menu 2 DegreeQualification
Branch  \
0  3 Months(NO STIPEND)      B.Tech/M.Tech
CSE(AIML)
1  3 Months(NO STIPEND)      B.Tech/M.Tech                     Data
science
2  3 Months(NO STIPEND)      B.Tech/M.Tech  ELECTRONICS AND
COMMUNICATION
3  3 Months(NO STIPEND)      B.Tech/M.Tech            CSE (DATA
SCIENCE)
4  3 Months(NO STIPEND)      B.Tech/M.Tech            CSE (DATA
SCIENCE)


    Experienceifany                            CollegeUniversity  \
0               NaN  KPR Institute of engineering and technology
1               NaN        Institute of aeronautical engineering
2   INTERN AT CODSOFT     SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
3                No         Malla reddy college of engineering
4                No         Malla reddy college of engineering


             Location              City Internshipstartdate
Message
0  Work From Home(WFH)         Coimbatore           2/5/2025
NaN
1  Work From Home(WFH)  Hyderabad South East        3/12/2025
NaN
2  Work From Home(WFH)            CHENNAI           4/15/2025
```

```
NaN
3  Work From Home(WFH)              Hyderabad              2/24/2024
NaN
4  Work From Home(WFH)              Hyderabad              2/24/2024
NaN
```

Some Missing Values are present in the dataset so i just manage or filled .its called a Cleaning Process.

```
df.isnull().sum()
```

```
Id                        0
Date                      0
Status                    0
Name                      0
Interns Email             0
Menu 23                   0
Number 570                0
Menu 2                    0
DegreeQualification       0
Branch                  323
Experienceifany        2234
CollegeUniversity         0
Location                 17
City                      2
Internshipstartdate       0
Message                3299
dtype: int64
```

```
df['Branch'] = df['Branch'].fillna("Not Provided")
```

```
df.isnull().sum()
```

```
Id                        0
Date                      0
Status                    0
Name                      0
Interns Email             0
Menu 23                   0
Number 570                0
Menu 2                    0
DegreeQualification       0
Branch                    0
Experienceifany        2234
CollegeUniversity         0
Location                 17
City                      2
Internshipstartdate       0
```

```
Message                    3299
dtype: int64

df['Experienceifany'] = df['Experienceifany'].fillna("No Experience")

df['Experienceifany']

0              No Experience
1              No Experience
2          INTERN AT CODSOFT
3                         No
4                         No
              ...
4903          No Experience
4904          No Experience
4905          No Experience
4906          No Experience
4907          No Experience
Name: Experienceifany, Length: 4905, dtype: object

df.isnull().sum()

Id                         0
Date                       0
Status                     0
Name                       0
Interns Email              0
Menu 23                    0
Number 570                 0
Menu 2                     0
DegreeQualification        0
Branch                     0
Experienceifany            0
CollegeUniversity          0
Location                  17
City                       2
Internshipstartdate        0
Message                 3299
dtype: int64

df['Location'] = df['Location'].fillna("Unknown")
df['Location']

0         Work From Home(WFH)
1         Work From Home(WFH)
2         Work From Home(WFH)
3         Work From Home(WFH)
4         Work From Home(WFH)
              ...
4903                  Unknown
4904                  Unknown
```

```
4905              Unknown
4906              Unknown
4907              Unknown
Name: Location, Length: 4905, dtype: object
```

```
df.isnull().sum()
```

```
Id                         0
Date                       0
Status                     0
Name                       0
Interns Email              0
Menu 23                    0
Number 570                 0
Menu 2                     0
DegreeQualification        0
Branch                     0
Experienceifany            0
CollegeUniversity          0
Location                   0
City                       2
Internshipstartdate        0
Message                 3299
dtype: int64
```

```
df['City']=df['City'].fillna("N/A")
```

```
df['City']
```

```
0              Coimbatore
1       Hyderabad South East
2                 CHENNAI
3               Hyderabad
4               Hyderabad
                ...
4903            New delhi
4904            New delhi
4905            New delhi
4906            New delhi
4907            New delhi
Name: City, Length: 4905, dtype: object
```

```
df.isnull().sum()
```

```
Id                         0
Date                       0
Status                     0
Name                       0
Interns Email              0
Menu 23                    0
Number 570                 0
```

```
Menu 2                      0
DegreeQualification         0
Branch                      0
Experienceifany             0
CollegeUniversity           0
Location                    0
City                        0
Internshipstartdate         0
Message                  3299
dtype: int64
```

```python
df['Message']=df['Message'].fillna("No Message")
```

```python
df['Message']
```

```
0                                            No Message
1                                            No Message
2                                            No Message
3                                            No Message
4                                            No Message
                           ...
4903    I want to join this internship as it will be a...
4904    I want to join this internship as it will be a...
4905                                         No Message
4906                                         No Message
4907                                         No Message
Name: Message, Length: 4905, dtype: object
```

```python
df.isnull().sum()
```

```
Id                     0
Date                   0
Status                 0
Name                   0
Interns Email          0
Menu 23                0
Number 570             0
Menu 2                 0
DegreeQualification    0
Branch                 0
Experienceifany        0
CollegeUniversity      0
Location               0
City                   0
Internshipstartdate    0
Message                0
dtype: int64
```
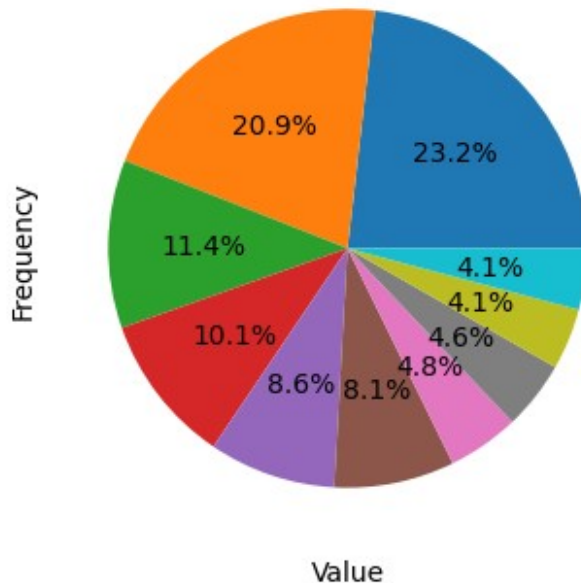
```python
df.columns
```

```
Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message'],
      dtype='object')

df.head(2)

        Id        Date  Status                      Name
Interns Email  \
0  2128156  12/24/2024  unread             Vignesh K
vikikannan8@gmail.com
1  2128008  12/24/2024  unread  Deekshitha Edukulla
deekshitha1745@gmail.com

                     Menu 23  Number 570                 Menu 2  \
0  AI(Artificial Intelligence  6369727236  3 Months(NO STIPEND)
1                      DevOps  9346380711  3 Months(NO STIPEND)

  DegreeQualification          Branch Experienceifany  \
0       B.Tech/M.Tech      CSE(AIML)    No Experience
1       B.Tech/M.Tech  Data science    No Experience

                            CollegeUniversity             Location  \
0  KPR Institute of engineering and technology  Work From Home(WFH)
1        Institute of aeronautical engineering   Work From Home(WFH)

                City Internshipstartdate     Message
0          Coimbatore            2/5/2025  No Message
1  Hyderabad South East          3/12/2025  No Message
```

Create a visual Histogram plot to finding easily to understandable data

```python
import matplotlib.pyplot as plt
plt.figure(figsize=(12,2))
df['Menu 23'].hist(bins=20)
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.title("Histogram of Column")
plt.xticks(rotation=90)
plt.show()
```

Histogram of Column
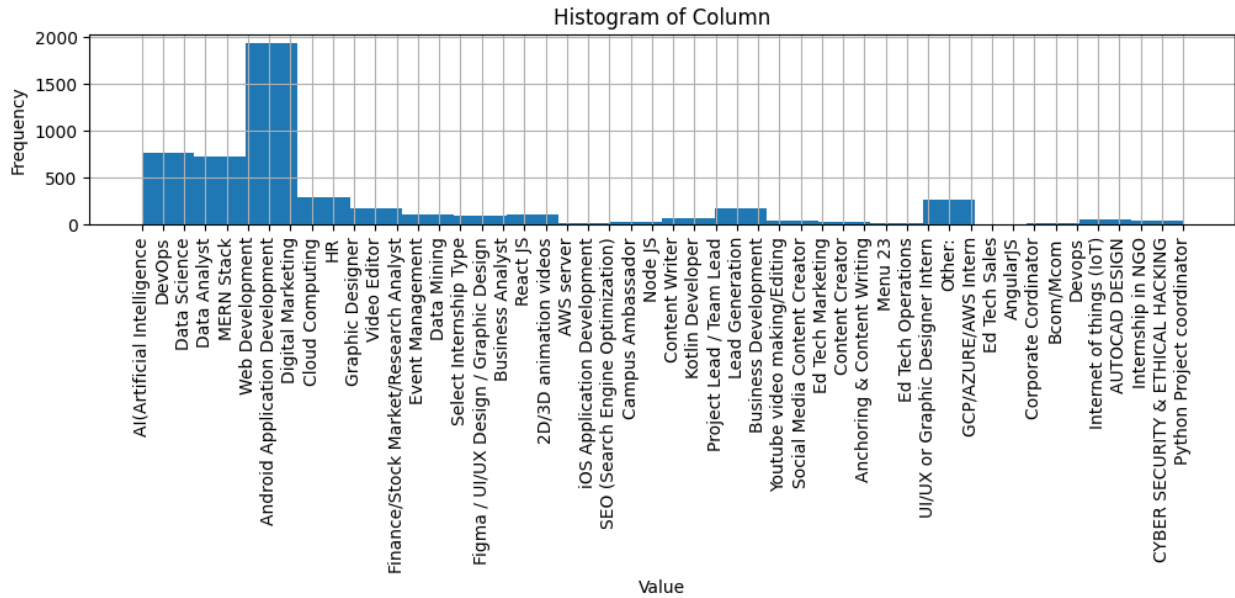
Visual BY using Count Plot to report Status Distribution

```python
plt.figure(figsize=(6, 3))
df['Status'].value_counts().plot(kind='bar', title='Status Distribution')
```
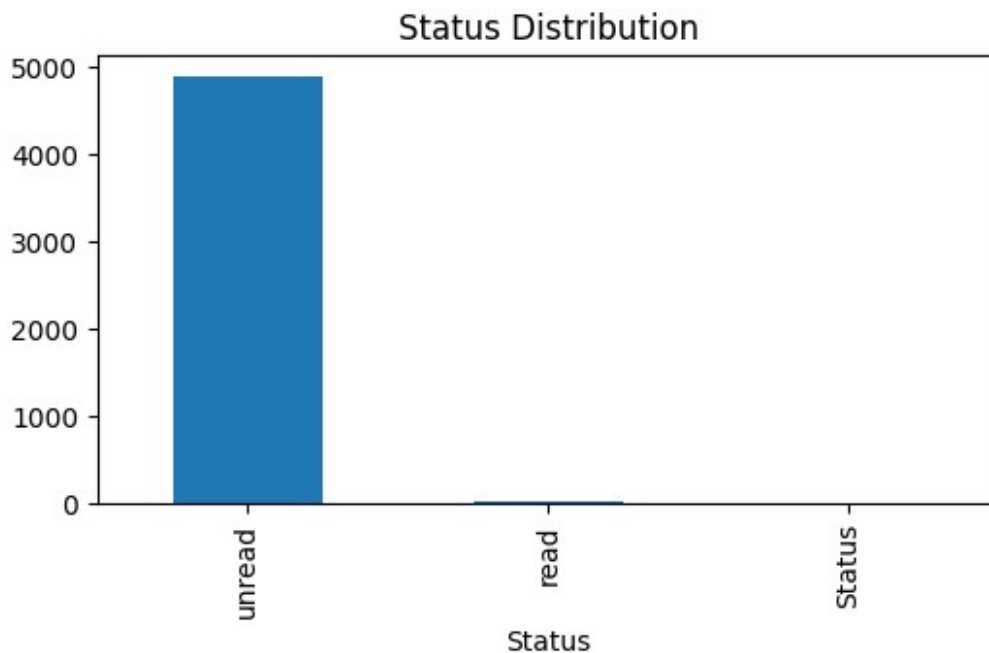
```
<Axes: title={'center': 'Status Distribution'}, xlabel='Status'>
```


Status Distribution

Visaul BY using Count Plot to report Frequency Distribution of Email Domains

```python
df['Email Domain'] = df['Interns Email'].str.split('@').str[1]

# Plot the frequency distribution of email domains
plt.figure(figsize=(6, 3))
df['Email Domain'].value_counts().head(10).plot(kind='bar',
color='skyblue', edgecolor='black')
plt.xlabel("Email Domains")
plt.ylabel("Frequency")
plt.title("Frequency Distribution of Email Domains")
plt.xticks(rotation=45)
plt.show()
```
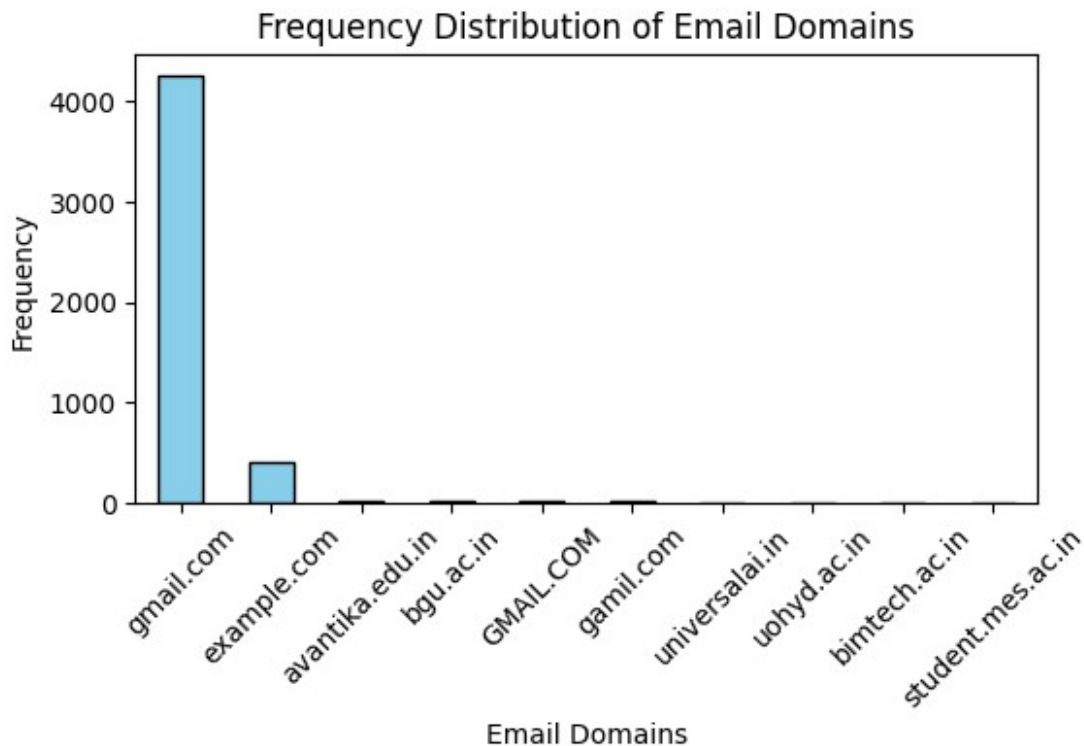

Visual BY using Count Plot to report Frequency Distribution of Menu 23

```python
menu_counts = df['Menu 23'].value_counts().head(10)

# Plot a bar chart
plt.figure(figsize=(6, 3))
plt.bar(menu_counts.index, menu_counts.values, color='skyblue',
edgecolor='black')
plt.xlabel("Menu 23 Values")
plt.ylabel("Frequency")
```

```
plt.title("Frequency Distribution of Menu 23")
plt.xticks(rotation=45)
plt.show()
```
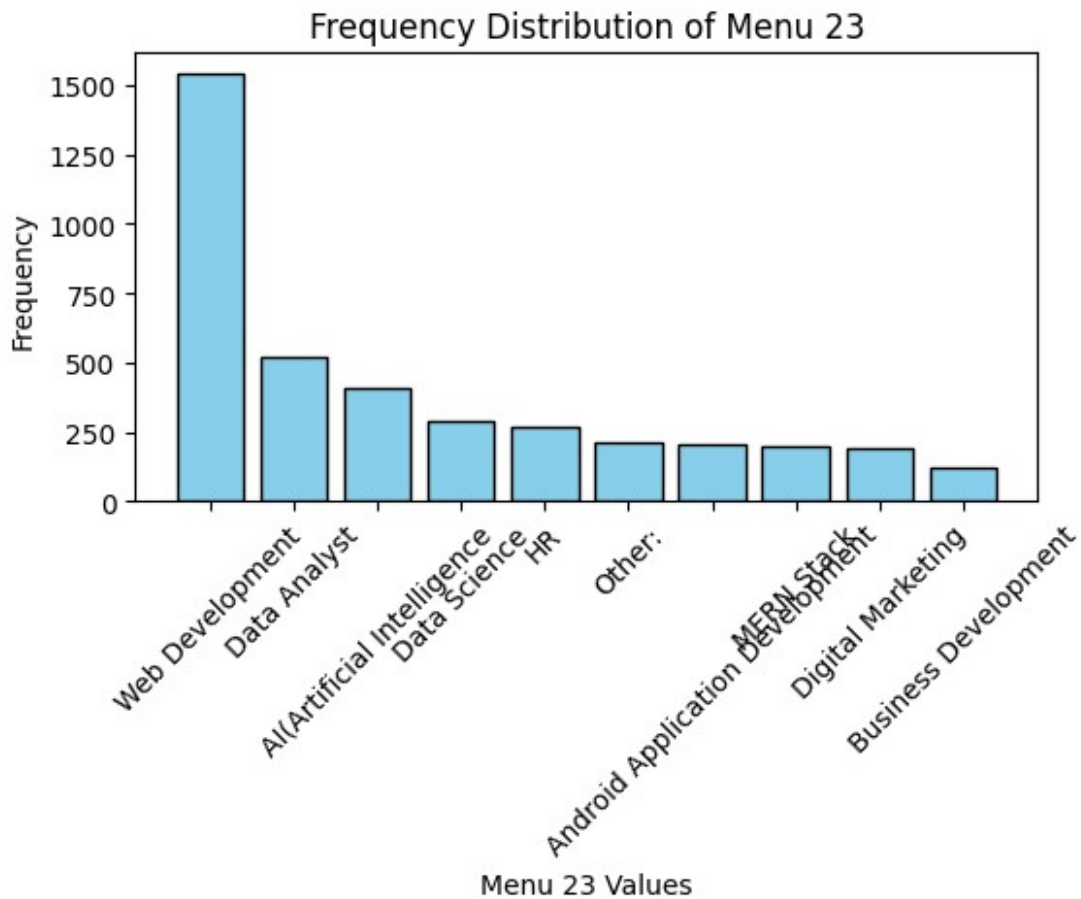

Frequency Distribution of Menu 23

```
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')
```

Visual by Pie Chart to report Histogram with Branch of Data

```
# Plot histogram of the sampled data
plt.figure(figsize=(4,4))
plt.pie(df["Branch"].value_counts().head(10),autopct='%1.1f%%')
plt.xlabel("Value",size=10)
plt.ylabel("Frequency",size=10)
plt.title("Histogram with Branch of Data",size=13)
```

```
plt.xticks(rotation=90)
plt.show()
```

Histogram with Branch of Data

Visual by Bar Plot to report Bar Plot of DegreeQualification

```
plt.figure(figsize=(6, 3))

# Plot the bar chart for DegreeQualification frequencies
df['DegreeQualification'].value_counts().plot(kind='bar',
color='skyblue')

# Add labels and title
plt.xlabel("Degree Qualification")
plt.ylabel("Frequency")
plt.title("Bar Plot of DegreeQualification")
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.show()
```
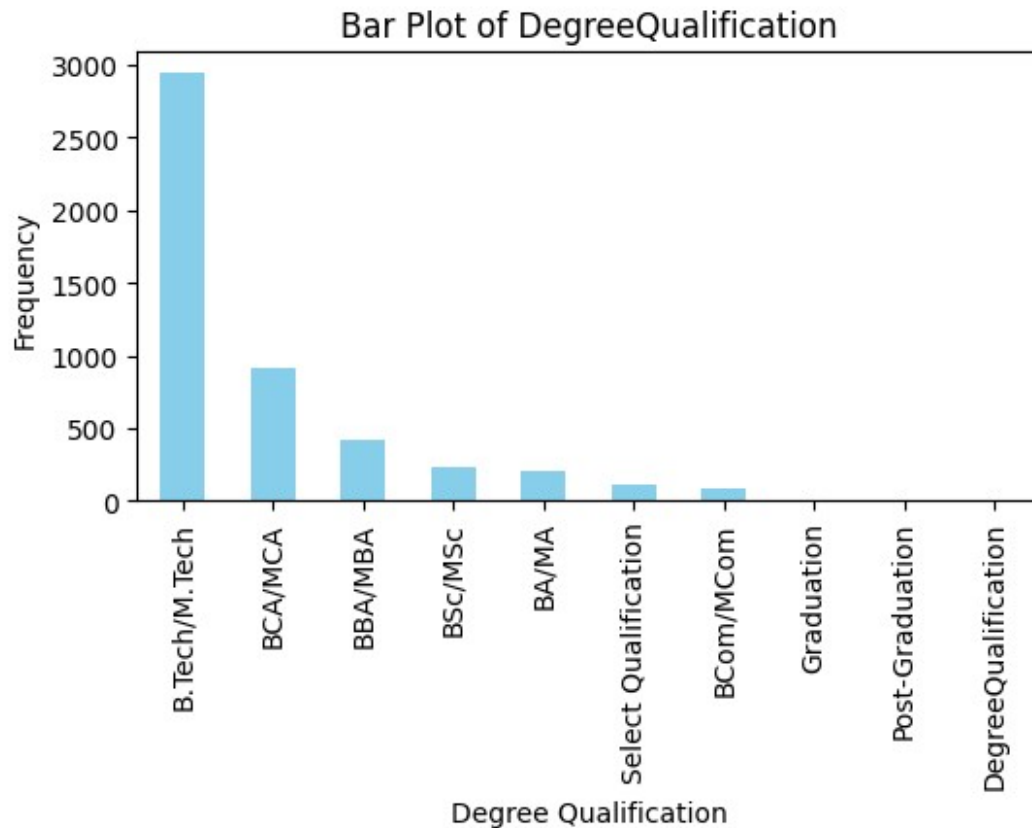
Bar Plot of DegreeQualification


```
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')

df['Location']

0         Work From Home(WFH)
1         Work From Home(WFH)
2         Work From Home(WFH)
3         Work From Home(WFH)
4         Work From Home(WFH)
               ...
4903               Unknown
4904               Unknown
4905               Unknown
4906               Unknown
4907               Unknown
Name: Location, Length: 4905, dtype: object
```

Visual by Pie Chart to report Location Distribution

```python
location_counts = df['Location'].value_counts()

# Plot a pie chart
plt.figure(figsize=(4,4))
plt.pie(
    location_counts,
    labels=location_counts.index,  # Set the labels for each location
    autopct='%1.1f%%',  # Display percentages
    startangle=90,  # Rotate for better layout
    colors=plt.cm.Paired.colors  # Color map for slices
)

plt.title("Location Distribution")  # Title of the pie chart
plt.show()
```
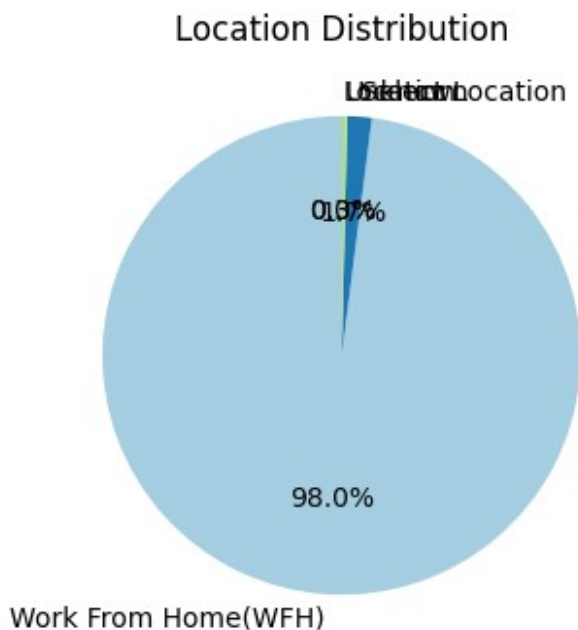

Location Distribution

```python
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')
```

```python
df['Internshipstartdate']

0        2/5/2025
1        3/12/2025
2        4/15/2025
```

```
3          2/24/2024
4          2/24/2024
            ...
4903     10/24/2023
4904     10/24/2023
4905     10/24/2023
4906     10/24/2023
4907     10/24/2023
Name: Internshipstartdate, Length: 4905, dtype: object
```

Visual BY Count Plot to report Internship Start Dates Over Time

```python
df['Internshipstartdate'] = pd.to_datetime(df['Internshipstartdate'],
errors='coerce')

# Group by date and count occurrences
date_counts = df['Internshipstartdate'].value_counts().sort_index()

# Plot the counts over time
plt.figure(figsize=(6, 3))

plt.plot(date_counts.index, date_counts.values, marker='o',
color='skyblue')
plt.xlabel("Start Date")
plt.ylabel("Number of Internships")
plt.title("Internship Start Dates Over Time")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```
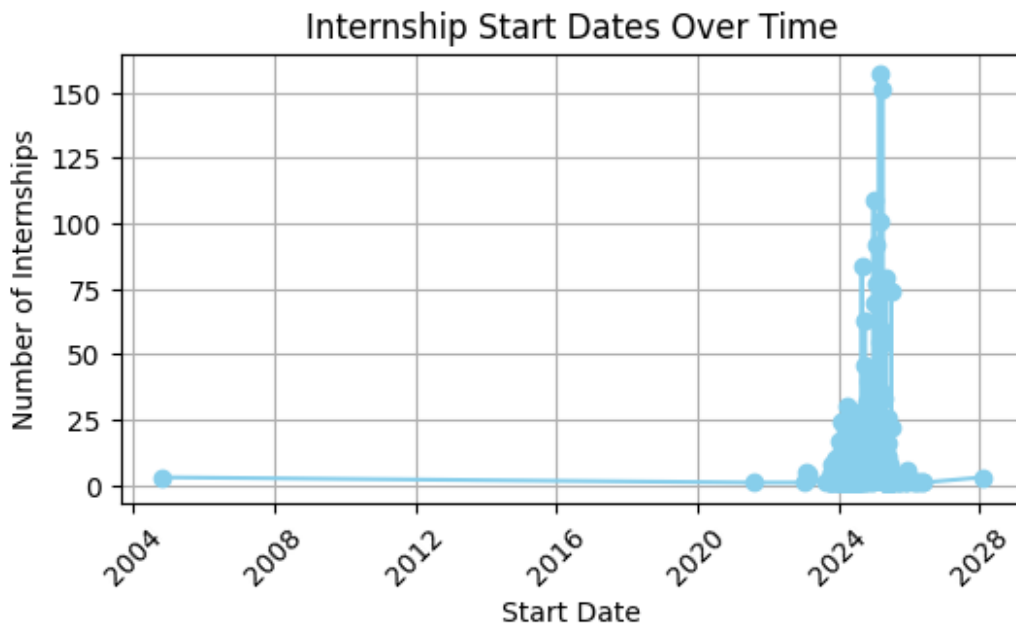
```
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')

df['Experienceifany'].value_counts()

Experienceifany
No Experience
2234
No
846
0
147
Fresher
133
no
106
                                                                    .
..
NO till student
1
Capabl web development intern
1
Capabl one month internship as full stack developer
1
1 month Web Development
1
Team lead and then HR team lead at younity India private limited
1
Name: count, Length: 538, dtype: int64
```

Visual By Pie Plot to report Distribution of Experience (Yes/No or Categories)

```python
experience_counts = df['Experienceifany'].value_counts().head(4)

# Plot a pie chart
plt.figure(figsize=(6,6))
plt.pie(
    experience_counts,
    labels=experience_counts.index,
    autopct='%1.1f%%',   # Display percentages on the pie slices
    startangle=90,       # Rotate for better layout
    colors=plt.cm.Paired.colors,  # Color map
)
```

```
plt.title("Distribution of Experience (Yes/No or Categories)")
plt.show()
```

## Distribution of Experience (Yes/No or Categories)
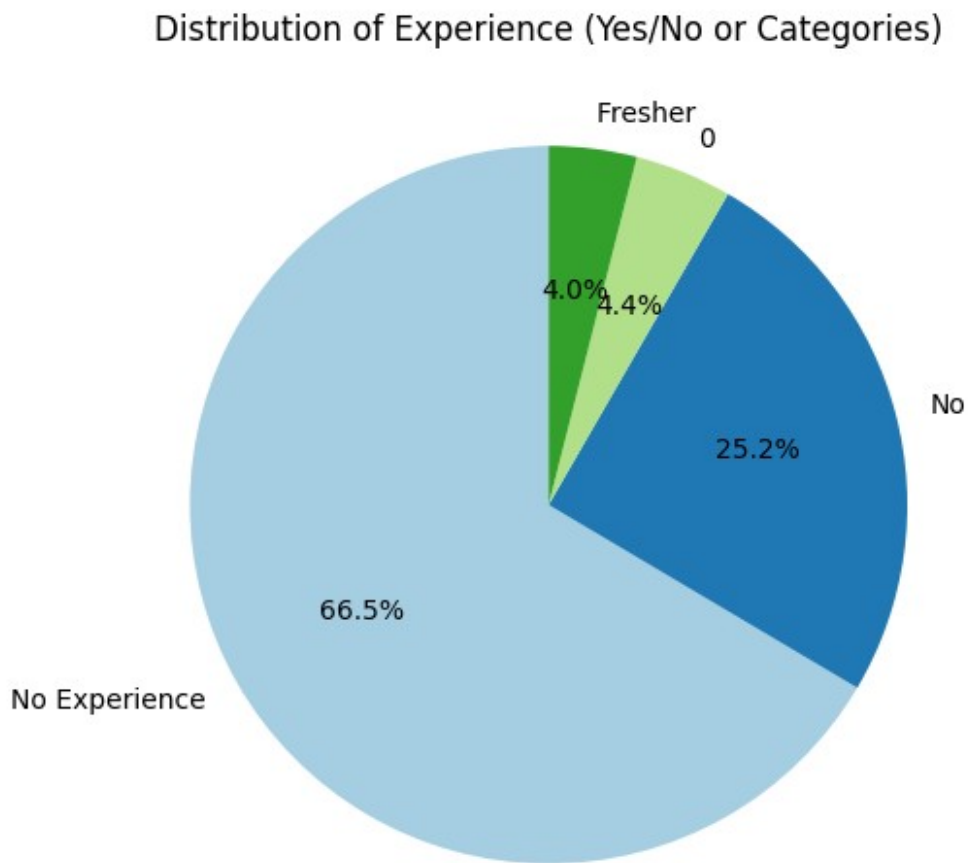

```
df['CollegeUniversity'].value_counts()

CollegeUniversity
Trinity institute of professional studies       42
Guru Gobind Singh Indraprastha University       39
Sage University                                 36
Sage university indore                          29
Sage University Indore                          27
                                                ..
Karpagam college of Engineering                  1
Finolex Academy of Management and Technology     1
North eastern hill University                    1
Integral University                              1
Kcmt                                             1
Name: count, Length: 2133, dtype: int64
```

Visual BY Pie Chart to report Proportion of Students from Different Colleges/Universities

```python
college_counts = df['CollegeUniversity'].value_counts().head(10)

# Plot a pie chart
plt.figure(figsize=(6, 6))
plt.pie(
    college_counts,
    labels=college_counts.index,
    autopct='%1.1f%%',
    startangle=90,
    colors=plt.cm.Paired.colors
)
plt.title("Proportion of Students from Different Colleges/Universities")
plt.show()
```
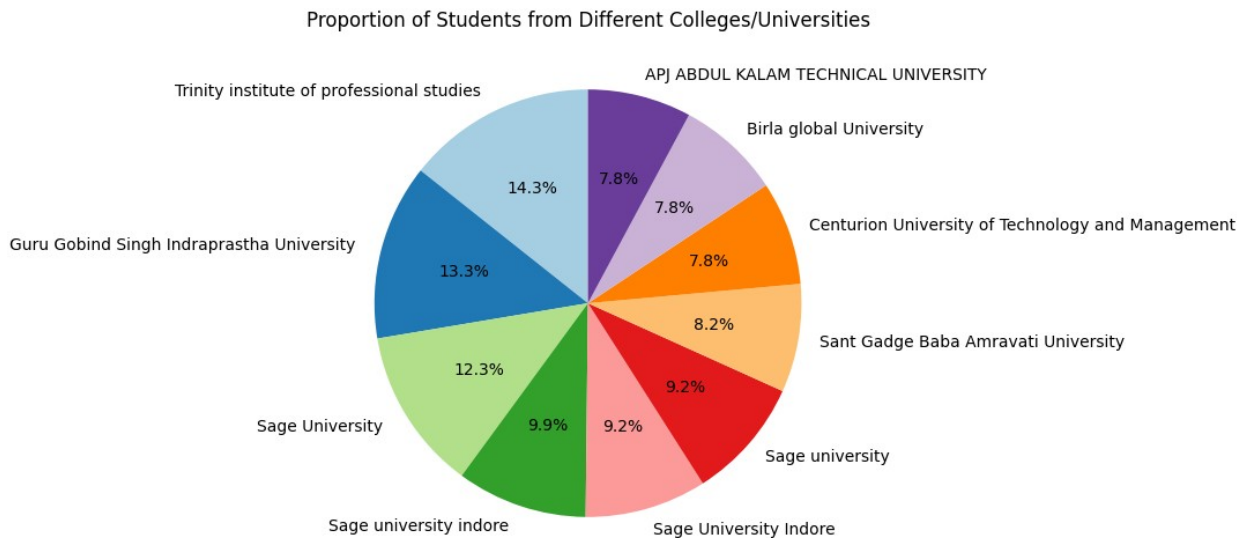

Proportion of Students from Different Colleges/Universities

```python
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')

df['Message'].value_counts()

Message
No Message
3299
Thank you for considering my application. I look forward to the
```

```
possibility of contributing to this exciting internship. Please feel
free to contact me via email or phone to discuss my application
further.
23
I have done my B.sc graduation and recently doing my MCA degree from
Sant Gadge baba Amravati university and want to tell you that I am
Comfortable with offline internship also if you are providing me.
22
Hi ,I am Naman Singh I am applying for an internship for data
Analyst , I am skilled in tableau, python(pandas , matplotlib, seaborn
and numpy ), exel, EDA, SQL (still learning)\r\n\r\n\r\nThank you \r\
nHoping for a positive result from your side.
21
want an app development internship
19

...
Hi ,Iam yogish Bakshi I am applying internship for data scientist , I
am skilled in tableau, python(pandas , matplotlib, seaborn  and
numpy ),EDA, SQL,exel, machine learning and deep learning (still
learning)\r\n\r\n\r\nThank you \r\nHopping for a positive result from
your side        1
I whoud like to gain the experience in working with company or start-
up to improve my skills to work with teams
1
Yes lam interested
1
To experience new things
1
Hello there , \r\nIm khushi gaur , I?m currently pursuing bcom hons at
trinity institute of professional studies. \r\nIvr previous done
internship in digital marketing . \r\nI m also doing a digital
marketing course right now
1
Name: count, Length: 949, dtype: int64
```

Visaul BY Histogram to report Distribution of Message Lengths

```python
# Calculate the length of each message
message_lengths = df['Message'].dropna().apply(len)

# Plot a histogram of message lengths
plt.figure(figsize=(6, 3))
plt.hist(message_lengths, bins=30, color='skyblue', edgecolor='black')
plt.xlabel("Message Length")
plt.ylabel("Frequency")
plt.title("Distribution of Message Lengths")
plt.show()
```
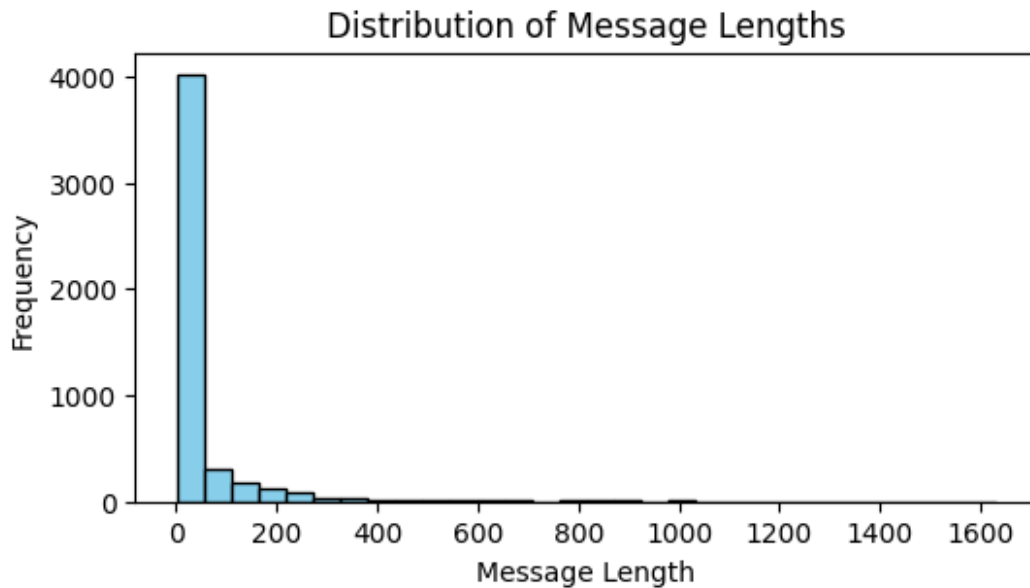
Distribution of Message Lengths


```
df.columns

Index(['Id', 'Date', 'Status', 'Name', 'Interns Email', 'Menu 23',
       'Number 570', 'Menu 2', 'DegreeQualification', 'Branch',
       'Experienceifany', 'CollegeUniversity', 'Location', 'City',
       'Internshipstartdate', 'Message', 'Email Domain'],
      dtype='object')

df['City'].value_counts()

City
Delhi             191
Hyderabad         167
Hyderabad         160
Indore            134
New Delhi         107
                 ...
Joda                1
BHILAI              1
Kanpur dehat        1
PAURI GARHWAL       1
Ayodhya             1
Name: count, Length: 1003, dtype: int64
```

Visual BY Pie chart to report Proportion of Interns from Different Cities

```
city_counts = df['City'].value_counts().head(10).dropna()

# Plot a pie chart
plt.figure(figsize=(4, 4))
plt.pie(
```

```
    city_counts,
  labels=city_counts.index,
    autopct='%1.1f%%',
  # startangle=90,
    #colors=plt.cm.Paired.colors
)
plt.title("Proportion of Interns from Different Cities")
plt.show()
```

Proportion of Interns from Different Cities