Importing Libraries

```
import pandas as pd # Importing the pandas library for data manipulation
import numpy as np  # Importing the numpy library for numerical operations
import matplotlib.pyplot as plt # Importing the matplotlib library for data visualization
import seaborn as sns # Importing the seaborn library for enhanced data visualization
from scipy.stats import norm
```

Importing the Dataset.

```
df = pd.read_csv('FT_Global_Business_School_MBA_Ranking_2023[1].csv')

df.head() #Getting the data for the first 5 rows.
```

```
   Rank                            School Name  Female faculty (%)  \
0     1             Columbia Business School                  26
1     2                              Insead                   22
2     3                 Iese Business School                  33
3     4               Harvard Business School                 33
4     4  Stanford Graduate School of Business                 26

   International mobility rank Salary today (US$) **  \
0                           49               228,425
1                            9               202,568
2                           19               182,278
3                           58               235,177
4                           56               253,435

   International students (%) Weighted salary (US$) International
board (%)  \
0                          57               226,359
39
1                          95               198,363
89
2                          88               181,270
90
3                          38               235,019
0 ‡
4                          46               248,669
25

   Value for money rank Audit year *  ...  Careers service rank  \
0                    80        2,019  ...                    23
1                    15        2,020  ...                    43
2                    92        2,022  ...                     5
3                    89        2,018  ...                    20
```

```
4                     72        2,022  ...                      11

   Overall satisfaction **  Carbon footprint rank  Rank in 2022  \
0                     9.51                      53           2.0
1                     9.31                      53           3.0
2                     9.61                       6          10.0
3                     9.69                      14           3.0
4                     9.98                      53           6.0

  Faculty with doctorates (%)  Female students (%)  \
0                          96                   44
1                          99                   36
2                         100                   37
3                          88                   46
4                          89                   44

  International course experience rank  Women on board (%)  Rank in
2021  \
0                                   48                  19
NaN
1                                    7                  50
1.0
2                                    5                  29
4.0
3                                   57                 0 ‡
NaN
4                                   27                  31
NaN

   Three-year average rank
0                      NaN
1                      2.0
2                      6.0
3                      NaN
4                      NaN

[5 rows x 28 columns]
```

Data Cleaning

```
df.shape         # Checking the shape of the DataFrame
```

```
(100, 28)
```

```
df.dtypes       # Checking the data types of each column
```

```
Rank                                  int64
School Name                          object
Female faculty (%)                    int64
International mobility rank           int64
```

```
Salary today (US$) **                      object
International students (%)                  int64
Weighted salary (US$)                      object
International board (%)                    object
Value for money rank                        int64
Audit year *                               object
Career progress rank                        int64
FT research rank                            int64
Aims achieved (%)                           int64
Sector diversity rank                       int64
Location, by primary campus                object
Alumni network rank                         int64
International faculty (%)                  object
ESG and net zero teaching rank              int64
Careers service rank                        int64
Overall satisfaction **                   float64
Carbon footprint rank                       int64
Rank in 2022                              float64
Faculty with doctorates (%)                 int64
Female students (%)                         int64
International course experience rank        int64
Women on board (%)                         object
Rank in 2021                              float64
Three-year average rank                   float64
dtype: object
```

```
df.describe()            # Display summary statistics of the DataFrame
```

```
              Rank  Female faculty (%)  International mobility rank  \
count  100.000000          100.000000                   100.000000
mean    50.310000           31.270000                    50.500000
std     29.019219            6.884113                    29.011492
min      1.000000           13.000000                     1.000000
25%     25.750000           26.750000                    25.750000
50%     50.500000           32.000000                    50.500000
75%     75.250000           36.000000                    75.250000
max    100.000000           49.000000                   100.000000

       International students (%)  Value for money rank  Career
progress rank  \
count                  100.000000            100.000000
100.000000
mean                    59.800000             50.500000
50.500000
std                     30.361459             29.011492
29.011492
min                      0.000000              1.000000
1.000000
25%                     39.750000             25.750000
25.750000
```

```
50%                      53.000000             50.500000
50.500000
75%                      93.250000             75.250000
75.250000
max                     100.000000            100.000000
100.000000

       FT research rank  Aims achieved (%)  Sector diversity rank  \
count        100.000000         100.000000             100.000000
mean          50.360000          87.410000              50.430000
std           28.944224           3.156187              28.914452
min            1.000000          80.000000               1.000000
25%           25.750000          85.000000              25.750000
50%           50.500000          88.000000              50.500000
75%           75.250000          90.000000              75.250000
max          100.000000          93.000000              97.000000

       Alumni network rank  ESG and net zero teaching rank  \
count           100.000000                      100.000000
mean             50.500000                       50.500000
std              29.011492                       29.011492
min               1.000000                        1.000000
25%              25.750000                       25.750000
50%              50.500000                       50.500000
75%              75.250000                       75.250000
max             100.000000                      100.000000

       Careers service rank  Overall satisfaction **  Carbon footprint
rank  \
count            100.000000               100.000000
100.000000
mean              50.500000                 9.016300
46.260000
std               29.011492                 0.457849
27.910253
min                1.000000                 7.610000
1.000000
25%               25.750000                 8.780000
17.000000
50%               50.500000                 9.090000
50.500000
75%               75.250000                 9.310000
75.000000
max              100.000000                 9.980000
89.000000

       Rank in 2022  Faculty with doctorates (%)  Female students (%)
\
count     83.000000                    100.00000           100.000000
```

```
mean      46.614458                    94.02000            39.260000

std       28.338903                     7.22367             9.560673

min        2.000000                    70.00000            13.000000

25%       22.500000                    90.00000            34.000000

50%       45.000000                    96.00000            39.000000

75%       70.500000                   100.00000            45.000000

max       99.000000                   100.00000            63.000000
```

```
       International course experience rank  Rank in 2021  \
count                            100.000000     81.000000
mean                              46.380000     44.123457
std                               23.967057     27.588214
min                                1.000000      1.000000
25%                               25.750000     21.000000
50%                               50.500000     42.000000
75%                               72.000000     64.000000
max                               72.000000    100.000000

       Three-year average rank
count                76.000000
mean                 46.157895
std                  25.562500
min                   2.000000
25%                  25.000000
50%                  44.500000
75%                  65.000000
max                  91.000000
```

```
df.isnull().sum()                          # Count the number of missing
values in each column of the DataFrame
```

```
Rank                                    0
School Name                             0
Female faculty (%)                      0
International mobility rank             0
Salary today (US$) **                   0
International students (%)              0
Weighted salary (US$)                   0
International board (%)                 0
Value for money rank                    0
Audit year *                            1
Career progress rank                    0
FT research rank                        0
Aims achieved (%)                       0
```

```
Sector diversity rank                    0
Location, by primary campus              0
Alumni network rank                      0
International faculty (%)                0
ESG and net zero teaching rank           0
Careers service rank                     0
Overall satisfaction **                  0
Carbon footprint rank                    0
Rank in 2022                            17
Faculty with doctorates (%)              0
Female students (%)                      0
International course experience rank     0
Women on board (%)                       0
Rank in 2021                            19
Three-year average rank                 24
dtype: int64
```

```
mode_value = df['Rank in 2022'].mode()[0]          # Calculate the
mode value of 'Rank in 2022' and fill missing values with it
df['Rank in 2022'].fillna(mode_value, inplace=True)

mode_value = df['Audit year *'].mode()[0]          # Calculate the
mode value of 'audit year ' and fill missing values with it
df['Audit year *'].fillna(mode_value, inplace=True)

median_value = df['Rank in 2021'].median()        # Calculate the
median value of 'Rank in 2021' and fill missing values with it
df['Rank in 2021'].fillna(median_value, inplace=True)

median_value = df['Three-year average rank'].median()   # Calculate
the median value of 'Three year average rank' and fill missing values
with it
df['Three-year average rank'].fillna(median_value, inplace=True)

df.isnull().sum()         #Again Counting the number of missing values
in each column of the DataFrame foe verification
```

```
Rank                                    0
School Name                             0
Female faculty (%)                      0
International mobility rank             0
Salary today (US$) **                   0
International students (%)              0
Weighted salary (US$)                   0
International board (%)                 0
Value for money rank                    0
Audit year *                            0
Career progress rank                    0
FT research rank                        0
Aims achieved (%)                       0
Sector diversity rank                   0
```

```
Location, by primary campus                0
Alumni network rank                        0
International faculty (%)                  0
ESG and net zero teaching rank             0
Careers service rank                       0
Overall satisfaction **                    0
Carbon footprint rank                      0
Rank in 2022                               0
Faculty with doctorates (%)                0
Female students (%)                        0
International course experience rank       0
Women on board (%)                         0
Rank in 2021                               0
Three-year average rank                    0
dtype: int64
```

```
print(df.describe())         # Display summary statistics for numeric
columns
```

```
              Rank  Female faculty (%)  International mobility rank  \
count  100.000000          100.000000                   100.000000
mean    50.310000           31.270000                    50.500000
std     29.019219            6.884113                    29.011492
min      1.000000           13.000000                     1.000000
25%     25.750000           26.750000                    25.750000
50%     50.500000           32.000000                    50.500000
75%     75.250000           36.000000                    75.250000
max    100.000000           49.000000                   100.000000

       International students (%)  Value for money rank  Career
progress rank  \
count                  100.000000            100.000000
100.000000
mean                    59.800000             50.500000
50.500000
std                     30.361459             29.011492
29.011492
min                      0.000000              1.000000
1.000000
25%                     39.750000             25.750000
25.750000
50%                     53.000000             50.500000
50.500000
75%                     93.250000             75.250000
75.250000
max                    100.000000            100.000000
100.000000

       FT research rank  Aims achieved (%)  Sector diversity rank  \
count        100.000000         100.000000             100.000000
```

```
mean          50.360000          87.410000              50.430000
std           28.944224           3.156187              28.914452
min            1.000000          80.000000               1.000000
25%           25.750000          85.000000              25.750000
50%           50.500000          88.000000              50.500000
75%           75.250000          90.000000              75.250000
max          100.000000          93.000000              97.000000

       Alumni network rank  ESG and net zero teaching rank  \
count           100.000000                      100.000000
mean             50.500000                       50.500000
std              29.011492                       29.011492
min               1.000000                        1.000000
25%              25.750000                       25.750000
50%              50.500000                       50.500000
75%              75.250000                       75.250000
max             100.000000                      100.000000

       Careers service rank  Overall satisfaction **  Carbon footprint
rank  \
count            100.000000               100.000000
100.000000
mean              50.500000                 9.016300
46.260000
std               29.011492                 0.457849
27.910253
min                1.000000                 7.610000
1.000000
25%               25.750000                 8.780000
17.000000
50%               50.500000                 9.090000
50.500000
75%               75.250000                 9.310000
75.000000
max              100.000000                 9.980000
89.000000

       Rank in 2022  Faculty with doctorates (%)  Female students (%)
\
count    100.000000                    100.00000           100.000000

mean      39.200000                     94.02000            39.260000

std       30.599069                      7.22367             9.560673

min        2.000000                     70.00000            13.000000

25%        9.750000                     90.00000            34.000000

50%       34.000000                     96.00000            39.000000
```

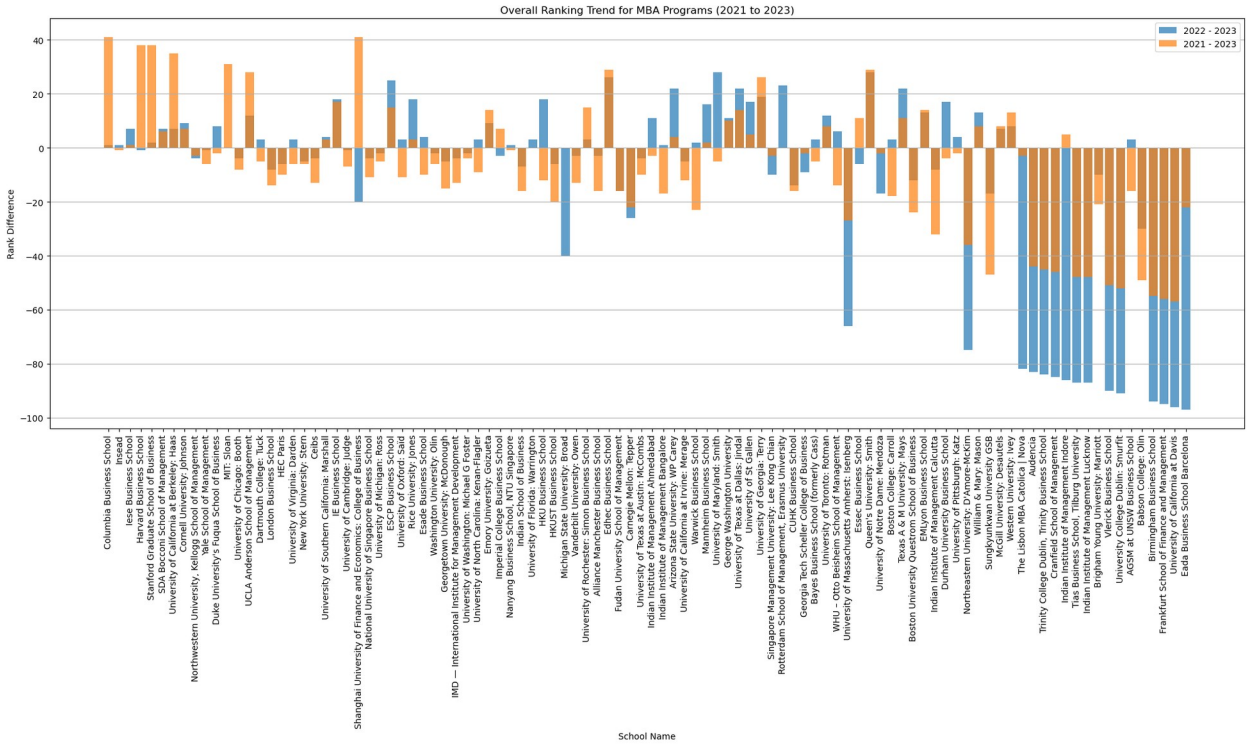
```
75%       65.000000                      100.00000             45.000000

max       99.000000                      100.00000             63.000000


       International course experience rank  Rank in 2021  \
count                             100.000000    100.000000
mean                               46.380000     43.720000
std                                23.967057     24.814089
min                                 1.000000      1.000000
25%                                25.750000     25.750000
50%                                50.500000     42.000000
75%                                72.000000     60.250000
max                                72.000000    100.000000

       Three-year average rank
count                100.000000
mean                  45.760000
std                   22.260678
min                    2.000000
25%                   32.500000
50%                   44.500000
75%                   63.000000
max                   91.000000
```

Analyzing and Visualizing the questions in Dataset

1) What is the overall ranking trend for MBA programs globally in 2023 compared to the previous years?

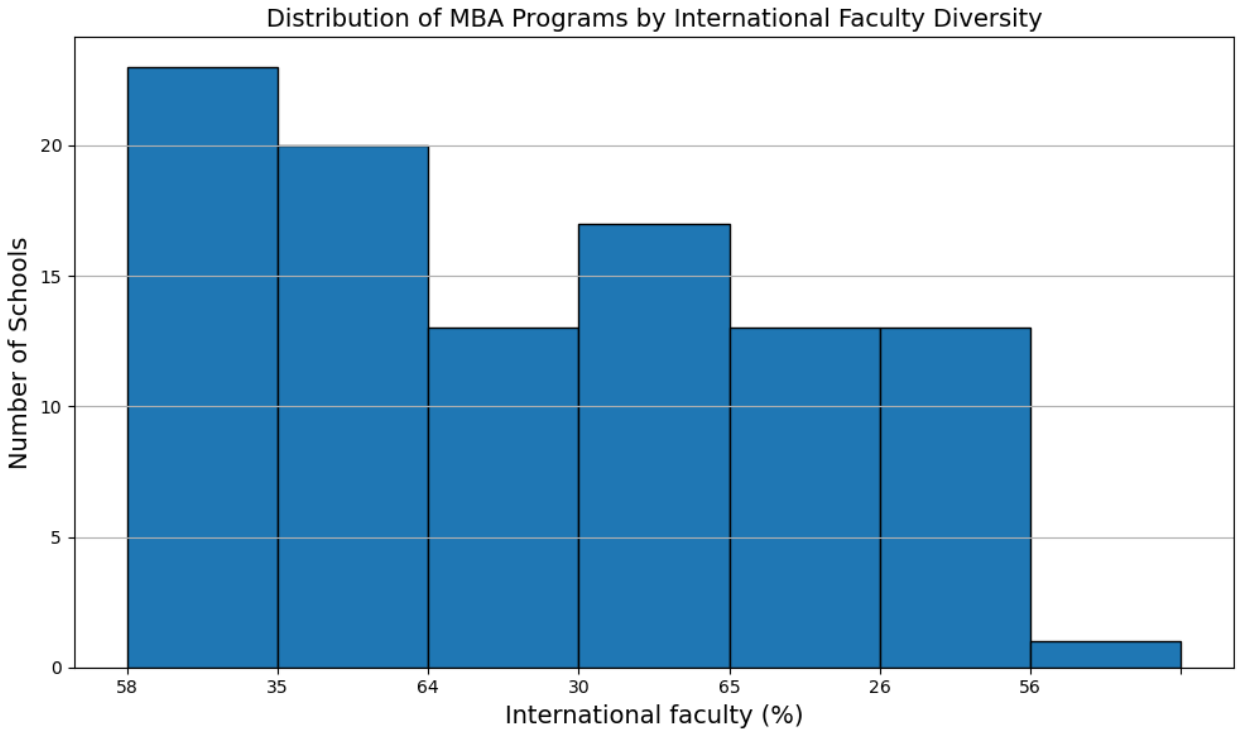
```
df['Rank Difference 2022'] = df['Rank in 2022'] - df['Rank']  # Calculate the differences in ranks between 2023 and 2022, and between 2023 and 2021
df['Rank Difference 2021'] = df['Rank in 2021'] - df['Rank']

# Creating a bar plot to visualize the overall ranking trend for each year
plt.figure(figsize=(20, 12))
plt.bar(df['School Name'], df['Rank Difference 2022'], label='2022 - 2023', alpha=0.7)
plt.bar(df['School Name'], df['Rank Difference 2021'], label='2021 - 2023', alpha=0.7)
plt.xlabel('School Name')
plt.ylabel('Rank Difference')
plt.title('Overall Ranking Trend for MBA Programs (2021 to 2023)')
plt.xticks(rotation=90)
plt.legend()
plt.grid(axis='y')
```

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



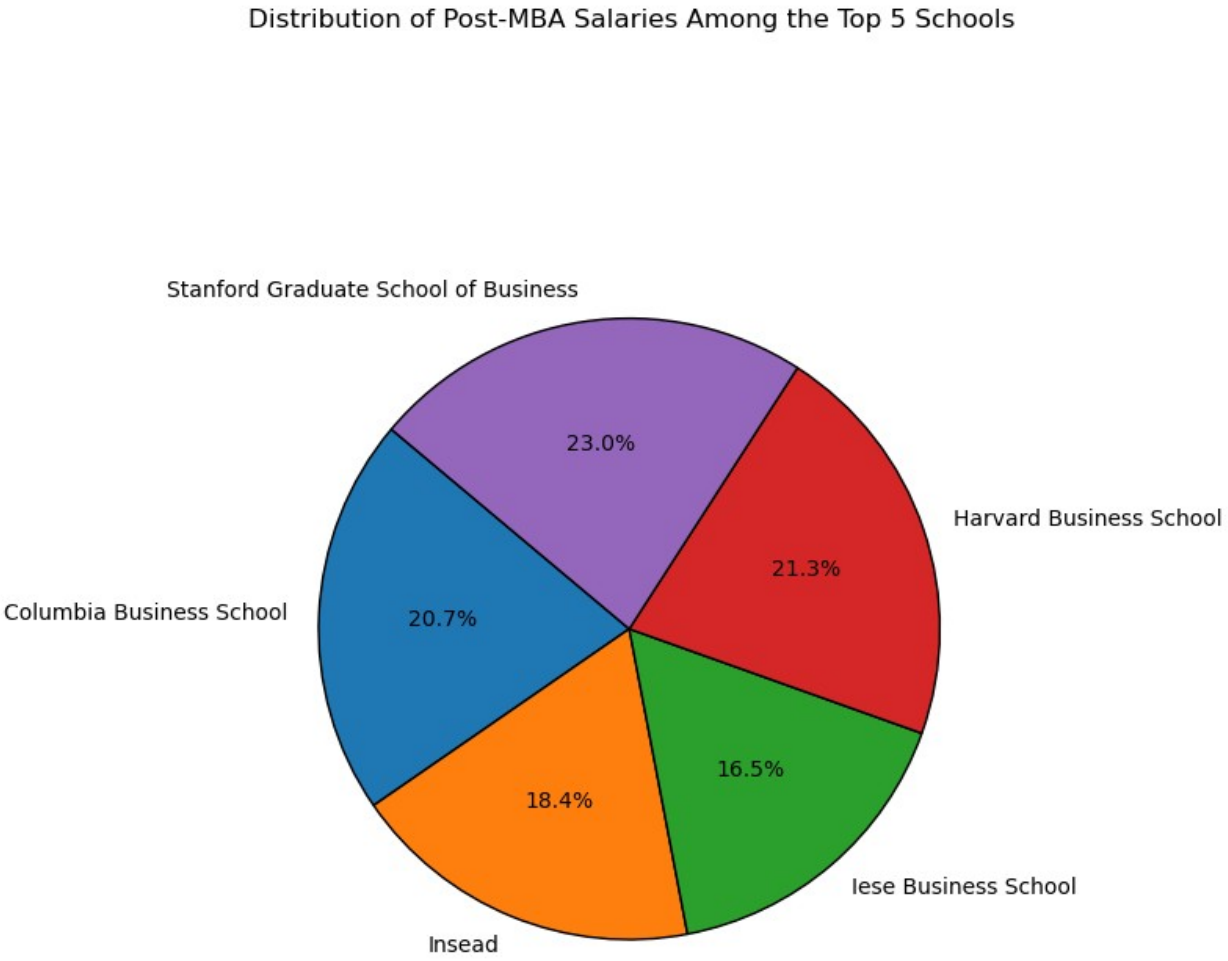
2) How do MBA programs rank in terms of international faculty diversity?

```
# Create a histogram for international faculty diversity
plt.figure(figsize=(10, 6))
bins = [0, 10, 20, 30, 40, 50, 60, 70]
plt.hist(df['International faculty (%)'], bins = bins ,
edgecolor='black')
plt.xlabel('International faculty (%)' , fontsize=14)
plt.ylabel('Number of Schools' , fontsize=14)
plt.title('Distribution of MBA Programs by International Faculty
Diversity' , fontsize=14)
plt.xticks(bins)
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```



3) What is the distribution of post-MBA salaries among the top-ranked schools?

```
# Creating a piechart for the top universities for post-MBA salaries
top_n = 5
top_ranked_schools = df.sort_values(by='Rank').head(top_n)
post_mba_salaries = top_ranked_schools['Salary today (US$)
**'].str.replace(',', '', regex=True).astype(float)
# Removing commas from 'Salary today (US$)' values and converting them
to float
labels = top_ranked_schools['School Name']

plt.figure(figsize=(8, 8))
plt.pie(post_mba_salaries, labels=labels, autopct='%1.1f%%',
startangle=140, wedgeprops={"edgecolor": "black"})
plt.title(f'Distribution of Post-MBA Salaries Among the Top {top_n}
Schools')
plt.axis('equal')  # Equal aspect ratio ensures that the pie chart is
circular.
plt.tight_layout()
plt.show()
```

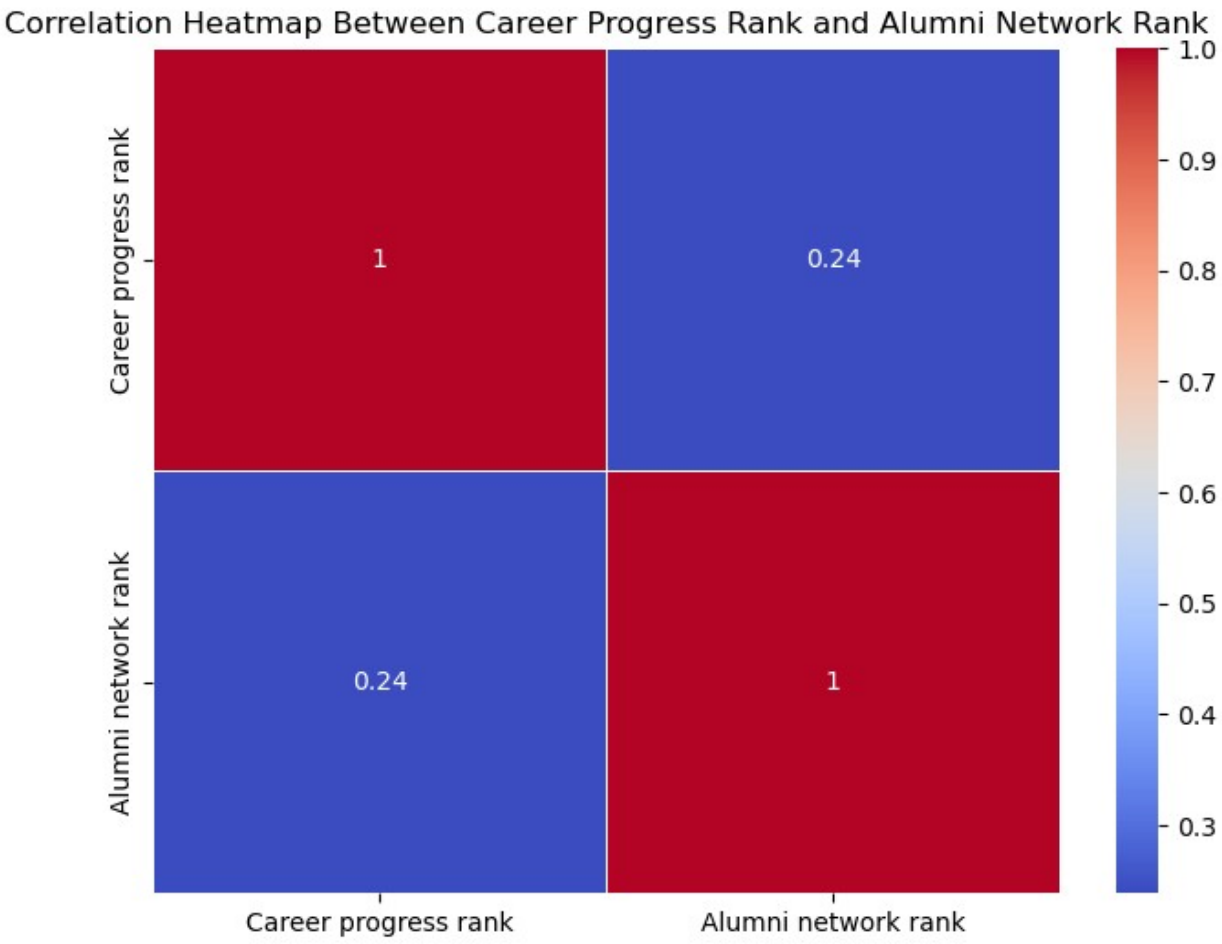


4) Are there any correlations between the career progress rank and alumni network rank?

```
# Calculating the correlation matrix
correlation_matrix = df[['Career progress rank', 'Alumni network
rank']].corr()
print(correlation_matrix)
```

```
                      Career progress rank  Alumni network rank
Career progress rank              1.000000             0.238512
Alumni network rank               0.238512             1.000000
```

```
# Creating a correlation heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title('Correlation Heatmap Between Career Progress Rank and Alumni
Network Rank')
plt.show()
```



5) Which MBA programs demonstrate the highest value for money based on the dataset?

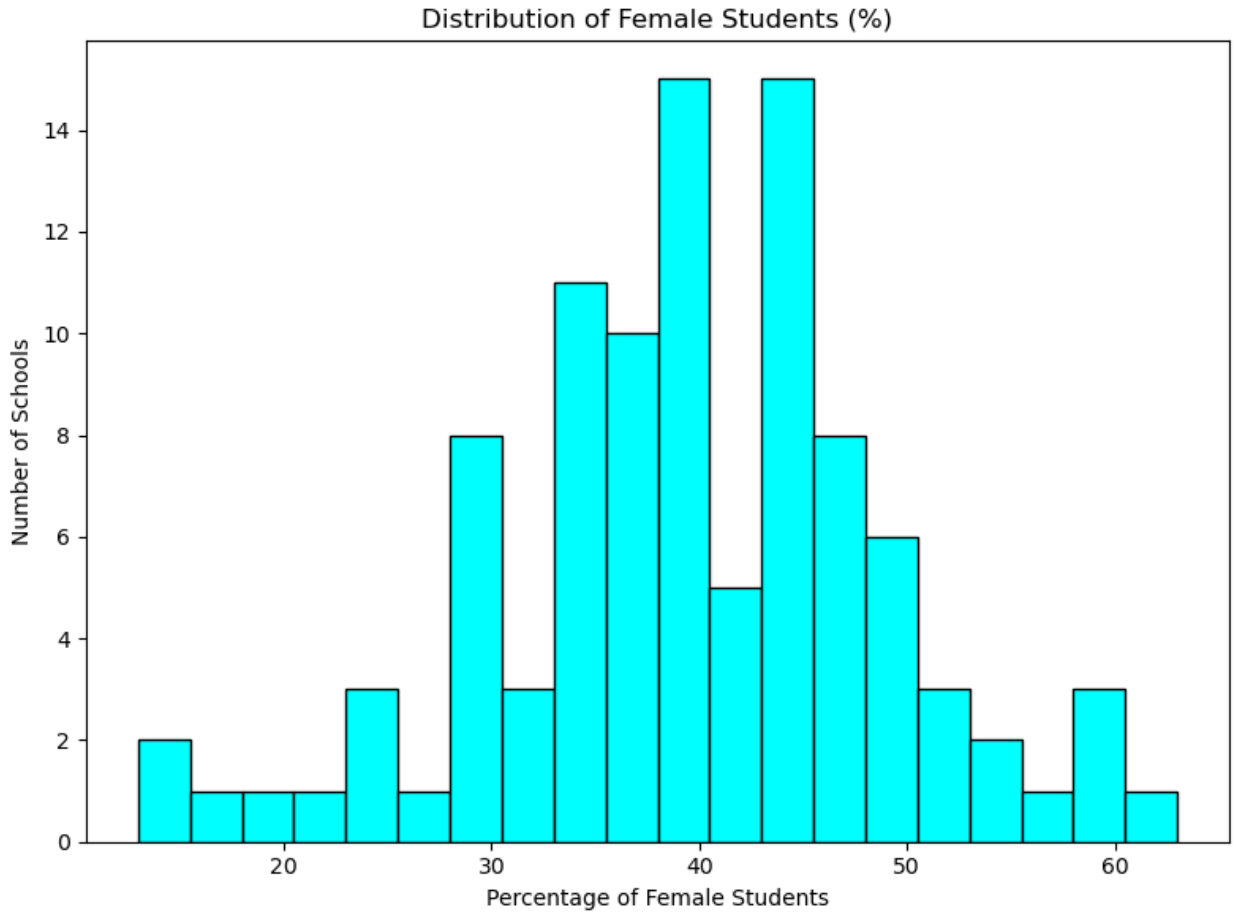
```
# Ensuring that 'Salary today (US$)' columns are numeric
df['Salary today (US$)'] = pd.to_numeric(df['Salary today (US$)
**'].str.replace(',', '', regex=True), errors='coerce')
# Calculating the value-for-money score
df['Value for Money Score'] = df['Salary today (US$) **']

# Sort the programs by the value-for-money score in descending order
df.sort_values(by='Value for Money Score', ascending=False,
```

```
inplace=True)

# Display the top programs with the highest value for money
top_programs = df[['School Name', 'Value for Money Score']].head(10)
print(top_programs)
```

```
                                          School Name Value for Money
Score
89           Tias Business School, Tilburg University
94,786
85                                            Audencia
79,457
4               Stanford Graduate School of Business
253,435
3                            Harvard Business School
235,177
0                           Columbia Business School
228,425
12                        University of Chicago: Booth
216,295
6          University of California at Berkeley: Haas
213,321
8   Northwestern University, Kellogg School of Man...
211,718
11                                          MIT: Sloan
207,100
1                                             Insead
202,568
```

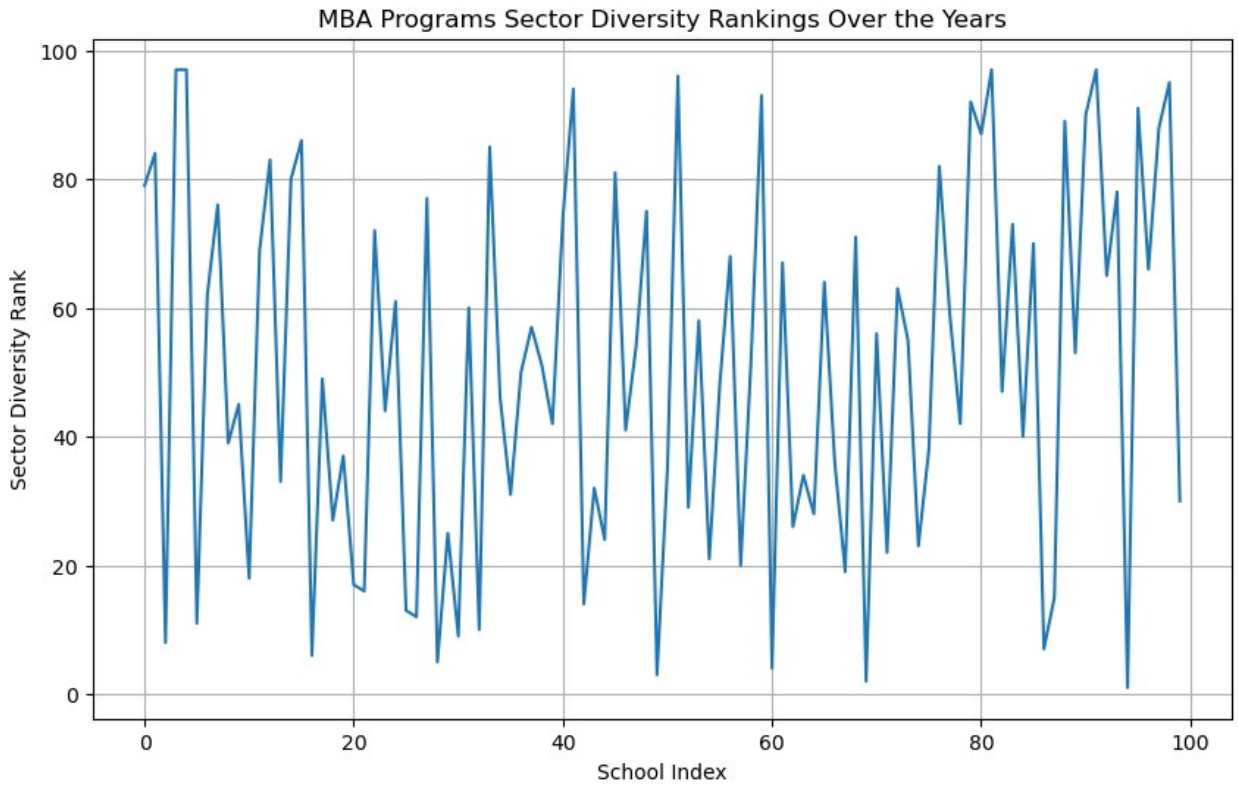
6) How does the percentage of female students vary across different MBA programs?

```
# Creating a histogram chart to visualize the percentage of female
students
plt.figure(figsize=(8, 6))
plt.hist(df['Female students (%)'], bins=20, color='cyan',
edgecolor='black')
plt.title('Distribution of Female Students (%)')
plt.xlabel('Percentage of Female Students')
plt.ylabel('Number of Schools')
plt.tight_layout()
plt.show()
```



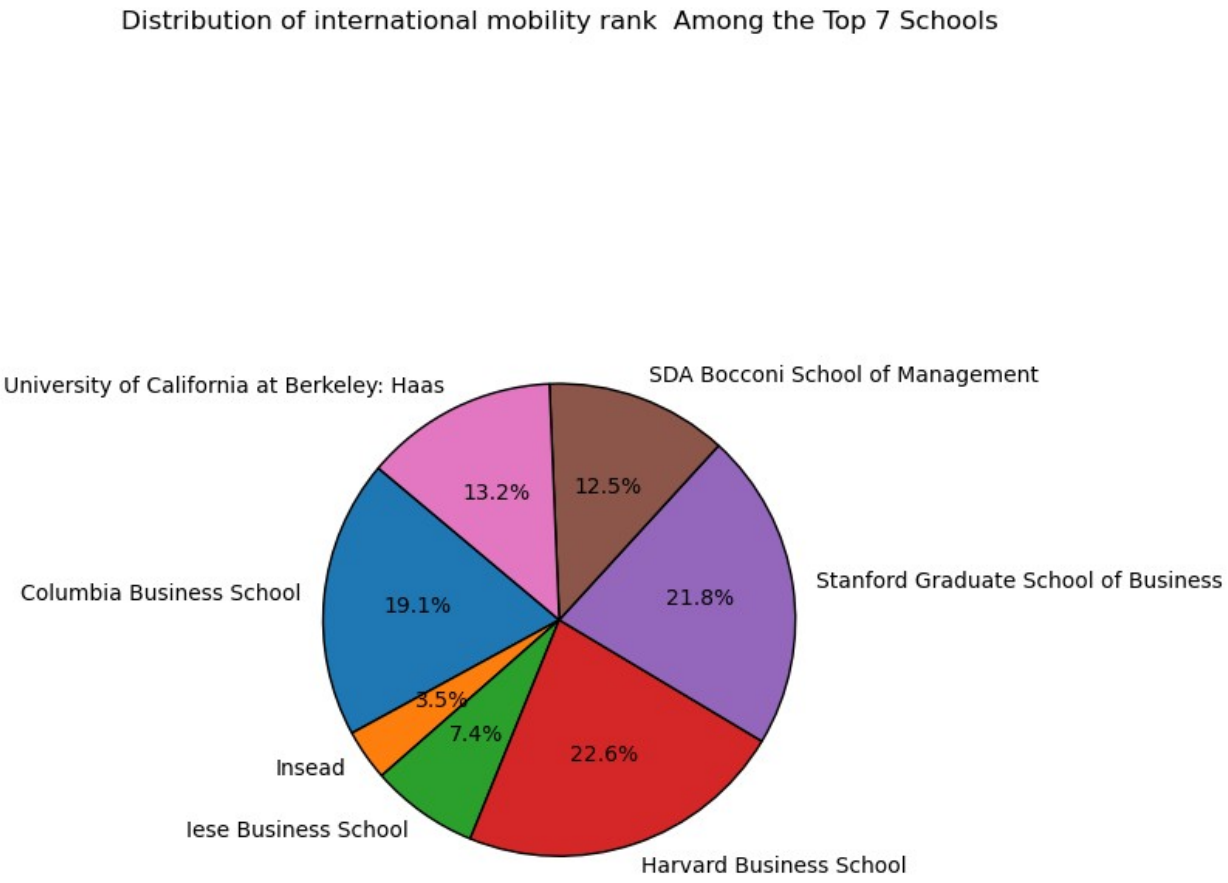
7) Can we identify any trends in MBA programs' sector diversity rankings over the years?

```
# Creating a line plot to visualize sector diversity rankings over the years for each school
plt.figure(figsize=(10, 6))
sector_diversity_rank = df['Sector diversity rank']
plt.plot(sector_diversity_rank)
plt.title('MBA Programs Sector Diversity Rankings Over the Years')
plt.xlabel('School Index')
plt.ylabel('Sector Diversity Rank')
plt.grid(True)
plt.show()
```



8) What are the top-ranked MBA programs for international mobility in 2023?

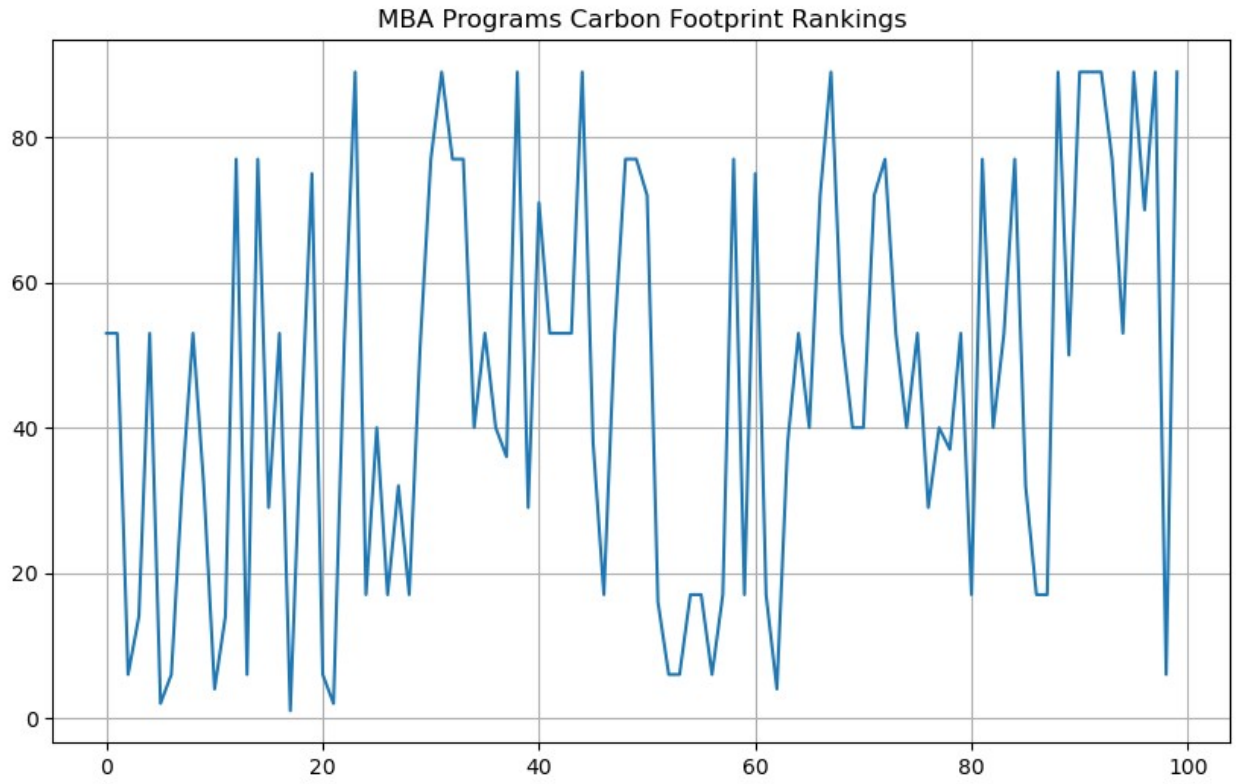
```
# Creating a piechart for the top universities for international
mobility rank
top_n = 7
top_ranked_schools = df.sort_values(by='Rank').head(top_n)
internatinal_mobility_rank = top_ranked_schools['International
mobility rank']
labels = top_ranked_schools['School Name']
plt.figure(figsize=(8,8))
plt.pie(internatinal_mobility_rank, labels=labels, autopct='%1.1f%%',
startangle=140, wedgeprops={"edgecolor": "black"})
plt.title(f'Distribution of international mobility rank  Among the Top
{top_n} Schools')
plt.axis('equal')  # Equal aspect ratio ensures that the pie chart is
circular.
plt.tight_layout()
plt.show()
```

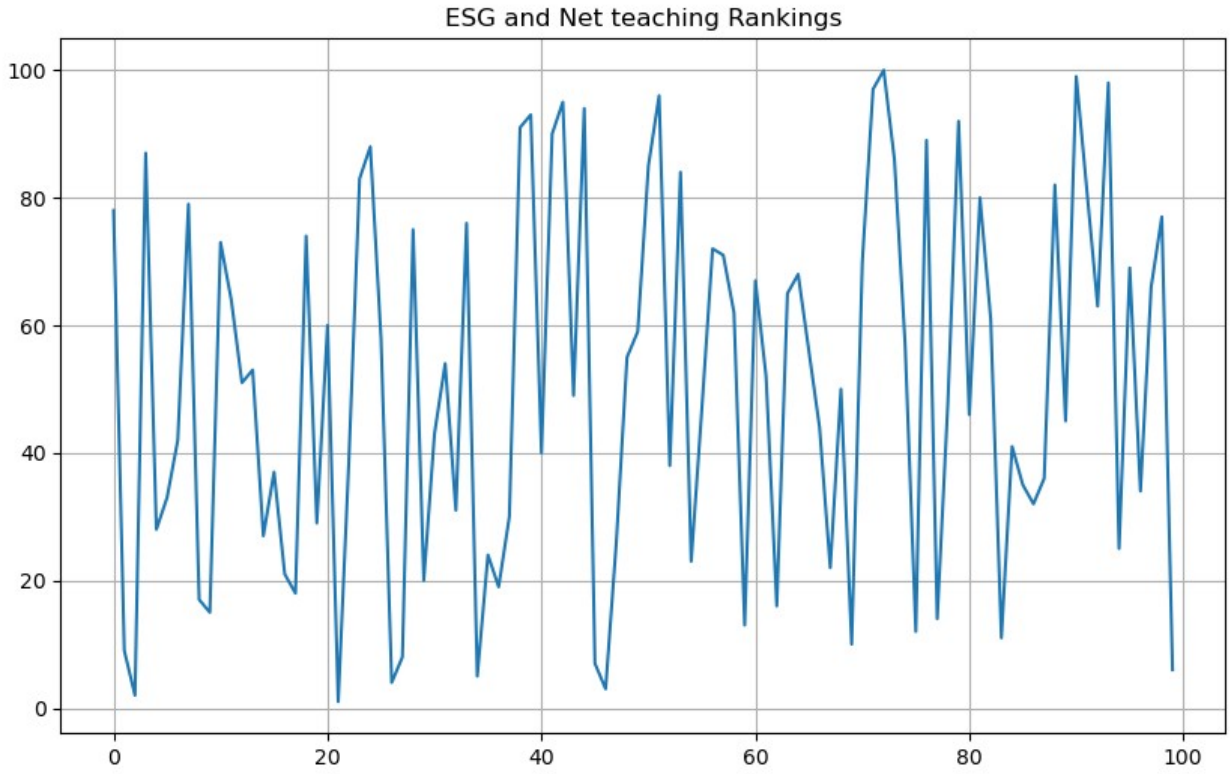


9) How do MBA programs perform in terms of carbon footprint rankings and ESG teaching?

```
# Creating a line plot to visualize sector diversity rankings for each school
plt.figure(figsize=(10, 6))
carbon_footprint_rank = df['Carbon footprint rank']
plt.plot(carbon_footprint_rank)
plt.title('MBA Programs Carbon Footprint Rankings')
plt.grid(True)
plt.show()
```

```
#creating a line plot to visualzie ESG and net zero teaching rank  for each school
plt.figure(figsize=(10,6))
ESG_Teaching_rank = df['ESG and net zero teaching rank']
plt.plot(ESG_Teaching_rank)
plt.title('ESG and Net teaching Rankings')
plt.grid(True)
plt.show()
```



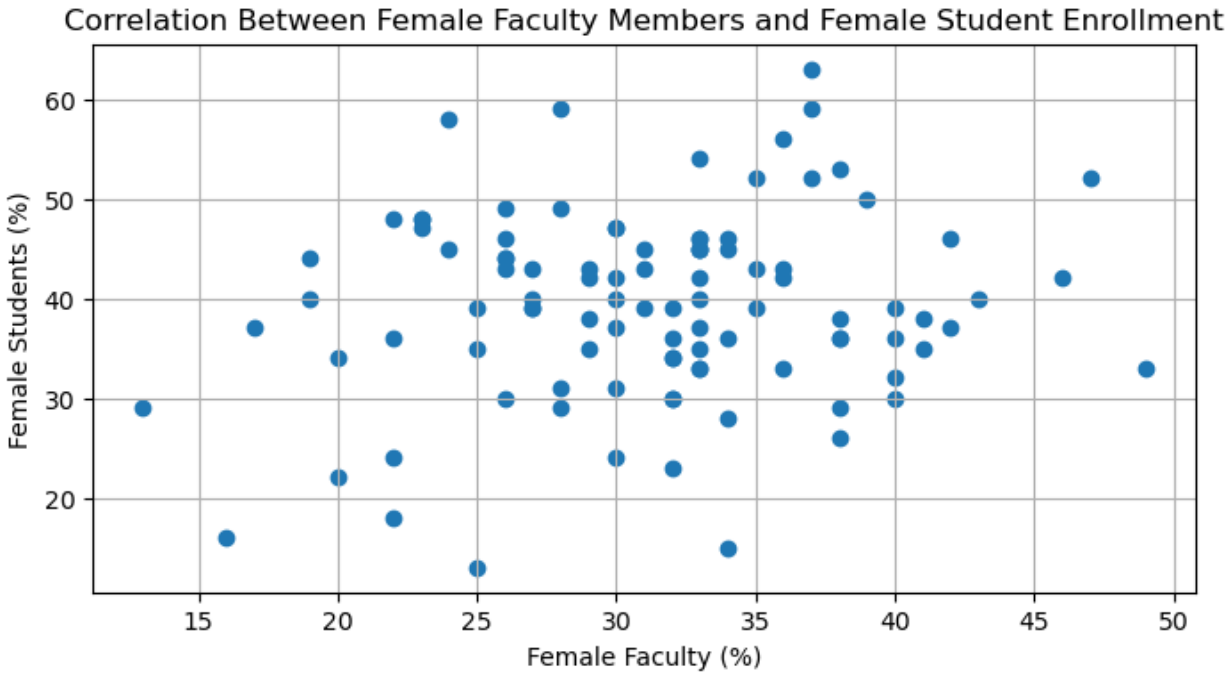


10) Is there a correlation between the international course experience rank and the overall satisfaction of students?

```
# Creating a scatter plot to show if there is any correalation bwtween
the two of them
plt.figure(figsize=(6, 4))
x = df['International course experience rank']
y = df['Overall satisfaction **']
plt.scatter(x,y)
plt.xlabel('International course experience rank')
plt.ylabel('Overall satisfaction **')
plt.title('Scatter plot of international course experience rank vs.
overall satisfaction')
plt.show()
```

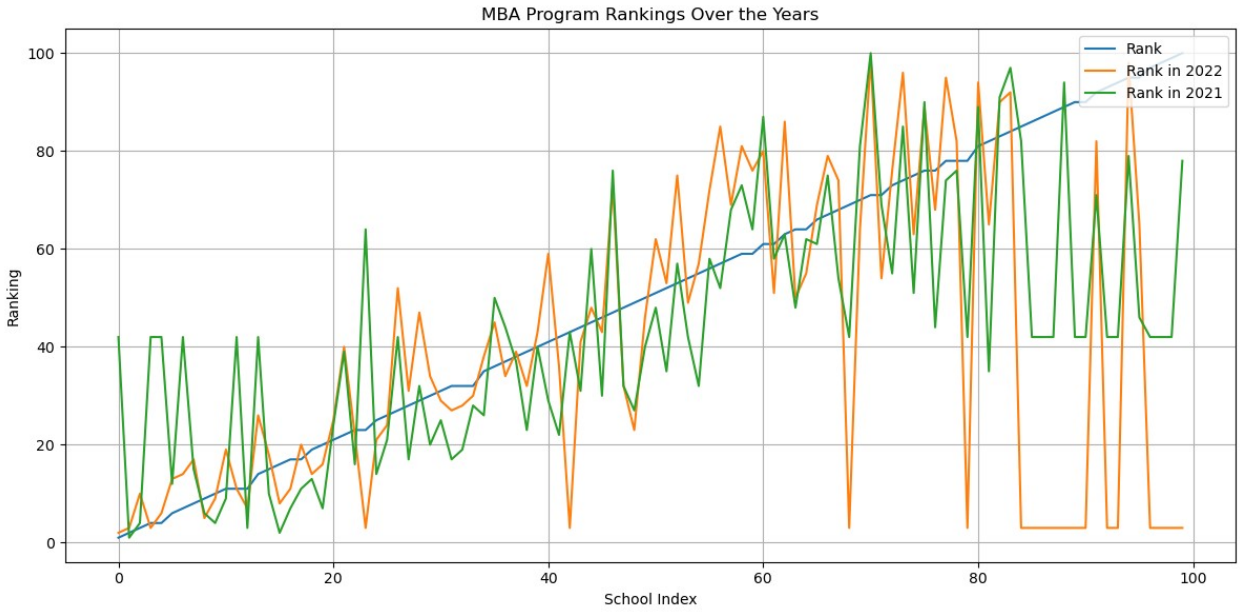
11) How does the presence of female faculty members correlate with female student enrollment?

```
# Creating a scatter plot to correlate female faculty with female
students
plt.figure(figsize=(8, 4))
x = df['Female faculty (%)']
y = df['Female students (%)']
plt.scatter(x,y)
plt.title('Correlation Between Female Faculty Members and Female
Student Enrollment')
plt.xlabel('Female Faculty (%)')
plt.ylabel('Female Students (%)')
plt.grid(True)
```



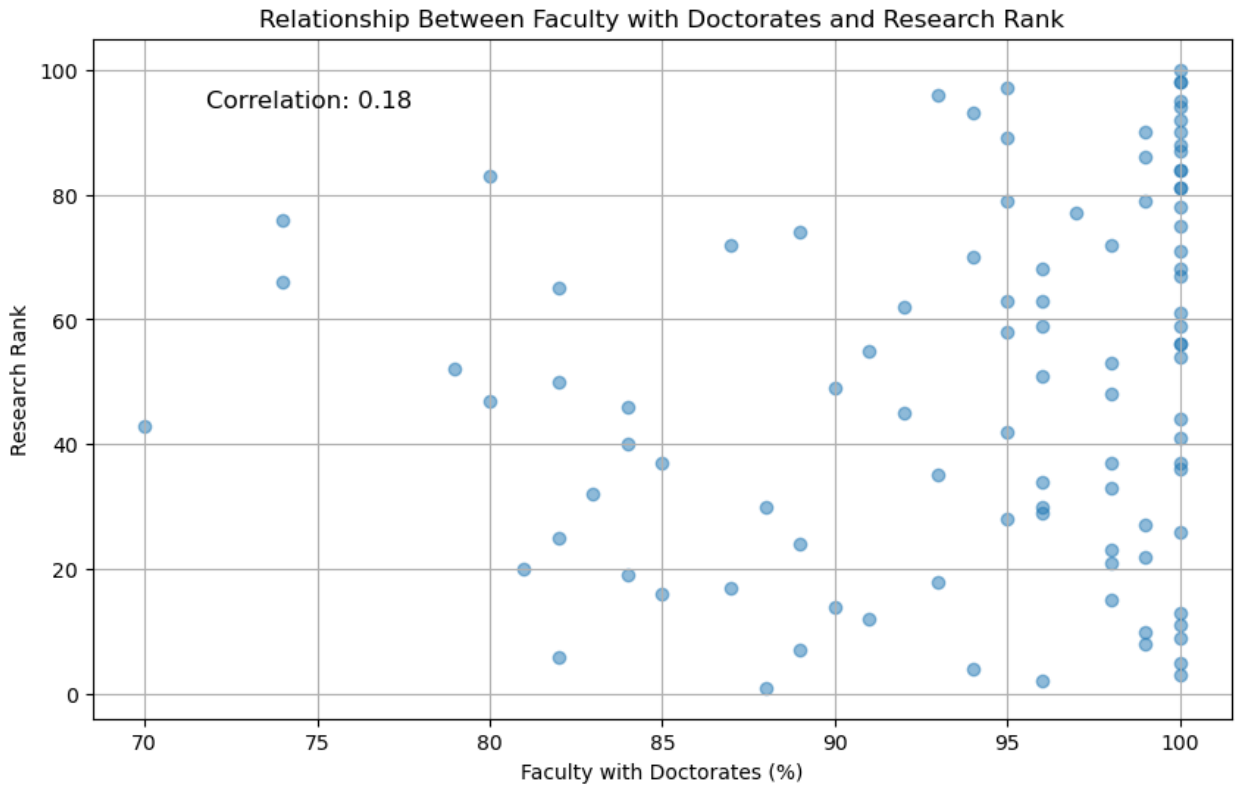
12) Are there any patterns in the rankings of MBA programs?

```
# Selecting relevant columns for ranking trends
rank_columns = ['Rank', 'Rank in 2022', 'Rank in 2021']

# Ploting the rankings for each school over the years
df[rank_columns].plot(figsize=(12, 6))
plt.title('MBA Program Rankings Over the Years')
plt.xlabel('School Index')
plt.ylabel('Ranking')
plt.legend(loc='upper right', labels=rank_columns)
plt.grid(True)
plt.tight_layout()
plt.show()
```



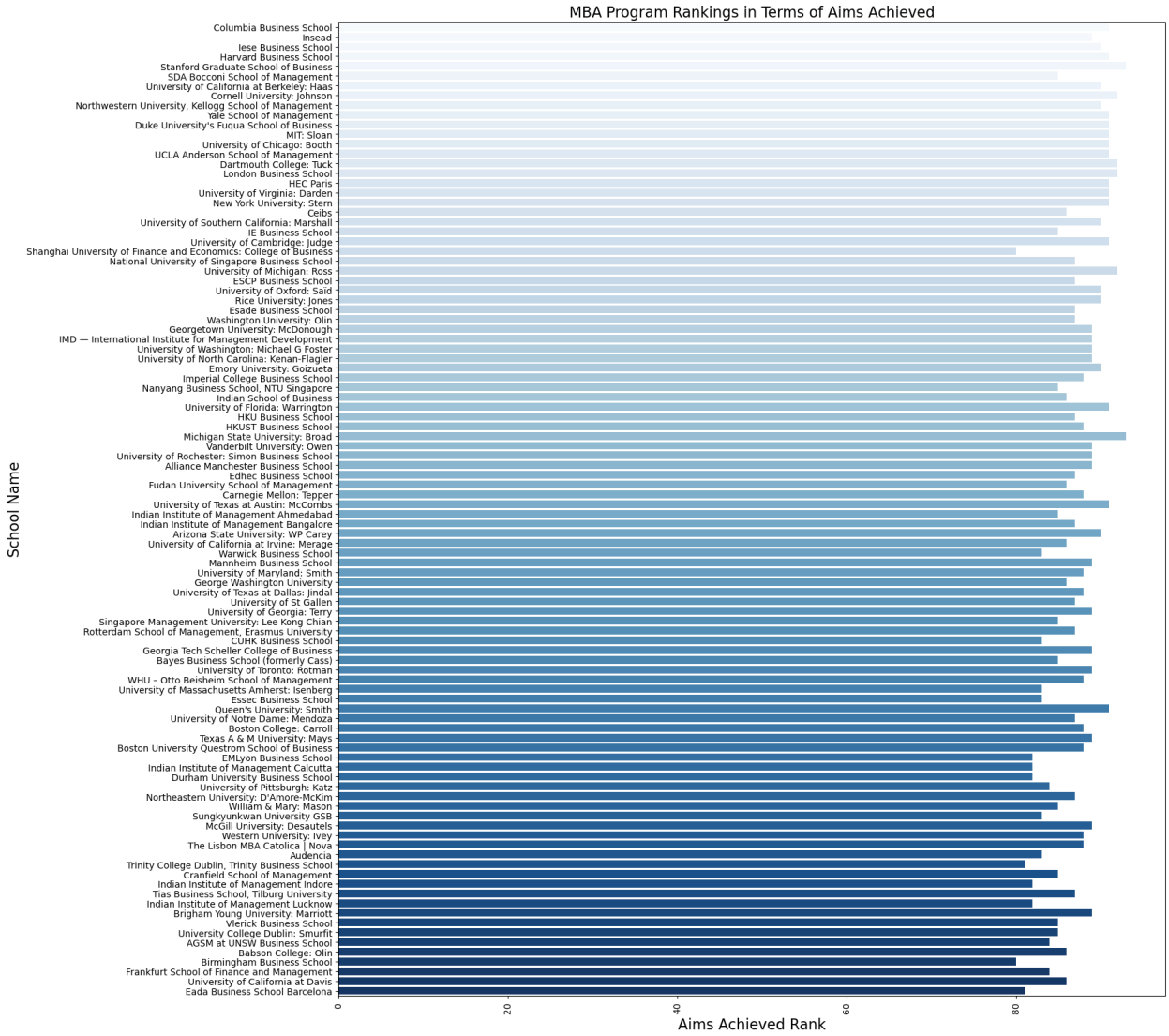
13) What is the relationship between the faculty with doctorates and the research rank of MBA programs

```
# Calculating the correlation coefficient
correlation_coefficient = df['Faculty with doctorates
(%)'].corr(df['FT research rank'])

# Creating a scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df['Faculty with doctorates (%)'], df['FT research rank'],
alpha=0.5)
plt.title('Relationship Between Faculty with Doctorates and Research
Rank')
plt.xlabel('Faculty with Doctorates (%)')
plt.ylabel('Research Rank')
plt.grid(True)
# Displaying the correlation coefficient on the plot
plt.annotate(f'Correlation: {correlation_coefficient:.2f}', xy=(0.1,
0.9), xycoords='axes fraction', fontsize=12)
plt.show()
```



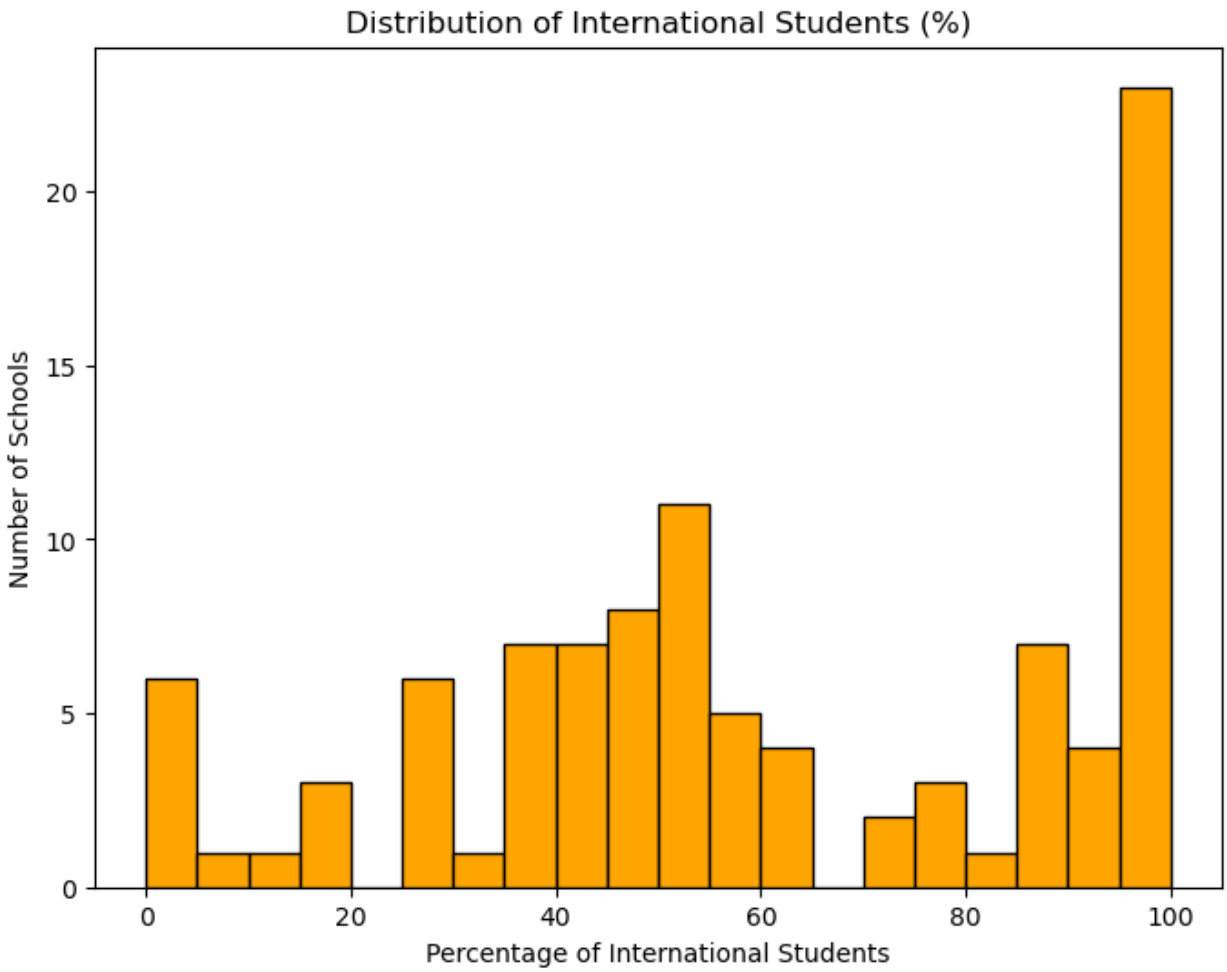
14) How do MBA programs rank in terms of aims achieved, and are there any common themes among the top performers?

```
# Identify common themes among top performers
top_performers = df[df['Aims achieved
(%)'].notna()].sort_values(by='Aims achieved (%)').head(10)
# Rank analysis (bar plot)
plt.figure(figsize=(18, 16))
sns.barplot(x='Aims achieved (%)', y='School Name', data=df,
palette='Blues')
plt.title('MBA Program Rankings in Terms of Aims Achieved', fontsize =
16)
plt.xlabel('Aims Achieved Rank', fontsize = 16)
plt.ylabel('School Name', fontsize = 16)
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



15) Can we identify a distribution in the number of international students across the schools?

```
# Plotting histogram for 'International students (%)'
plt.figure(figsize=(8,6))
plt.hist(df['International students (%)'], bins=20, color='orange',
edgecolor='black')
plt.title('Distribution of International Students (%)')
plt.xlabel('Percentage of International Students')
plt.ylabel('Number of Schools')
plt.show()
```



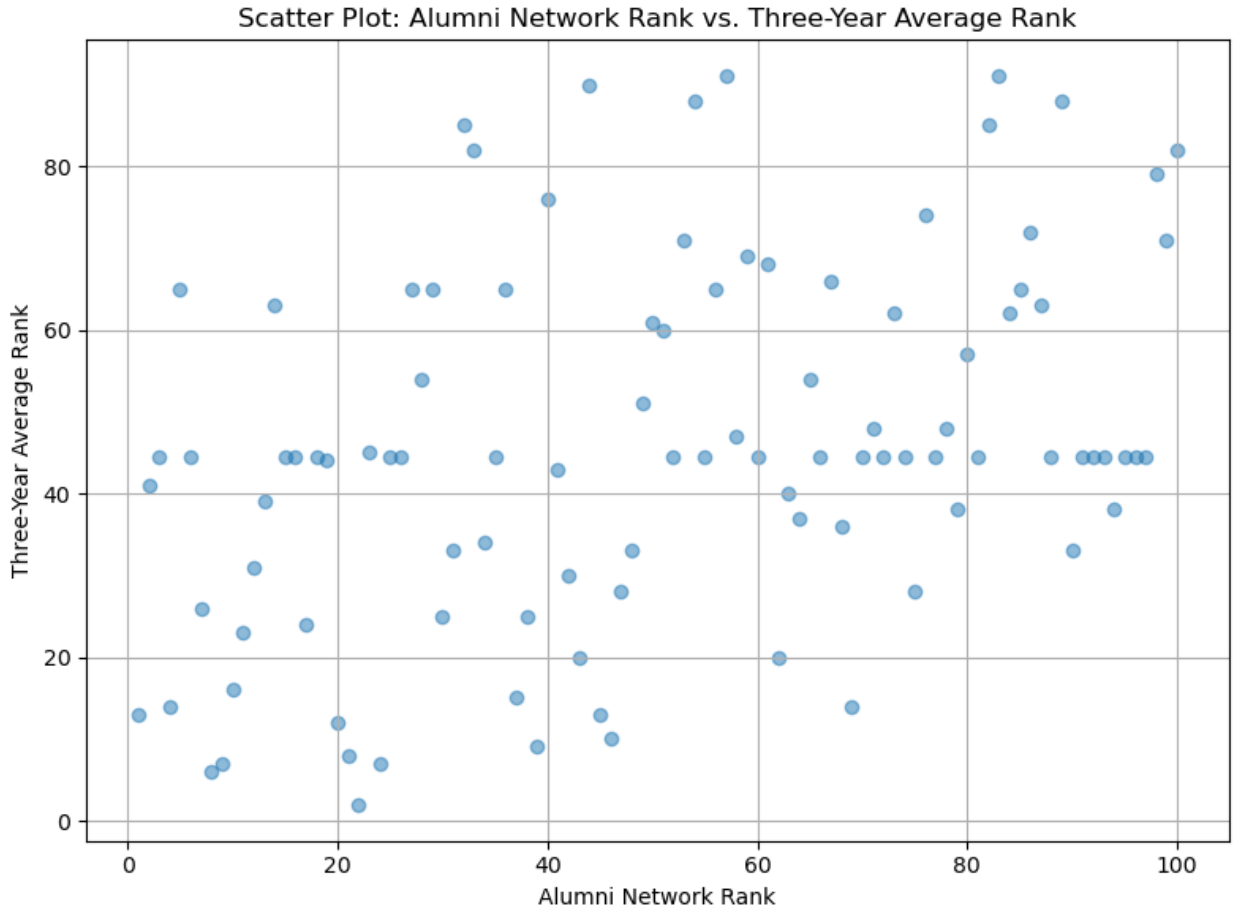
16) What is the correlation between the alumni network rank and the three-year average rank?

```
correlation = df['Alumni network rank'].corr(df['Three-year average rank'])

print(f"Correlation between Alumni Network Rank and Three-Year Average Rank: {correlation:.2f}")
```

```
Correlation between Alumni Network Rank and Three-Year Average Rank: 0.40
```

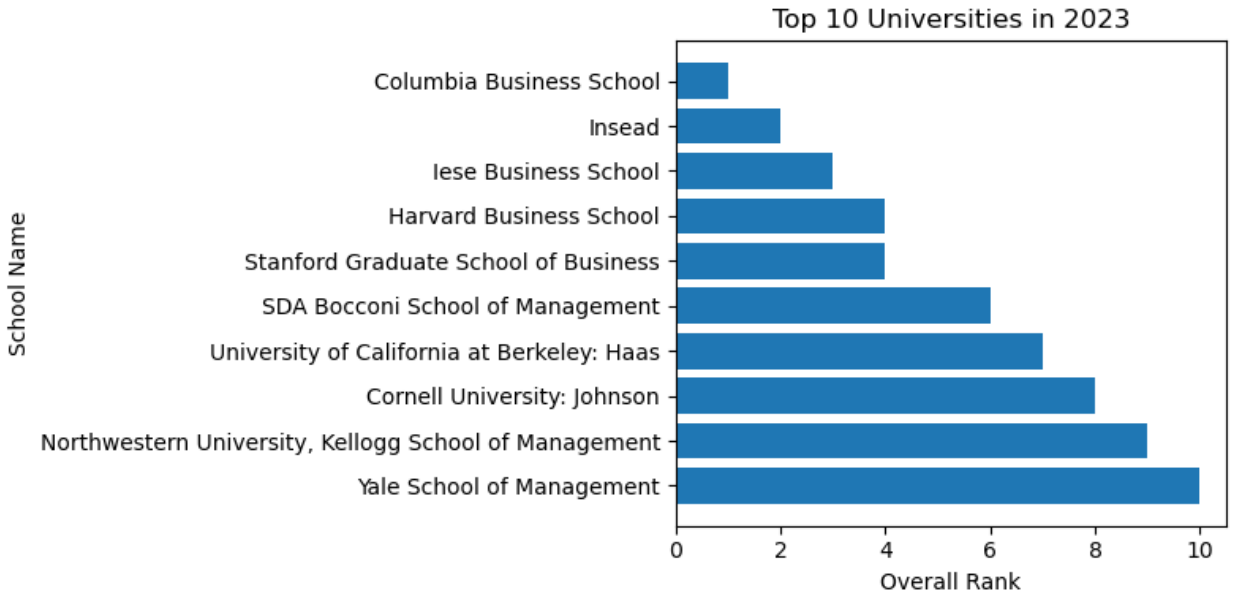
```
plt.figure(figsize=(8, 6))
plt.scatter(df['Alumni network rank'], df['Three-year average rank'], alpha=0.5)
plt.title('Scatter Plot: Alumni Network Rank vs. Three-Year Average Rank')
plt.xlabel('Alumni Network Rank')
plt.ylabel('Three-Year Average Rank')
```

```
plt.grid(True)
plt.tight_layout()
plt.show()
```



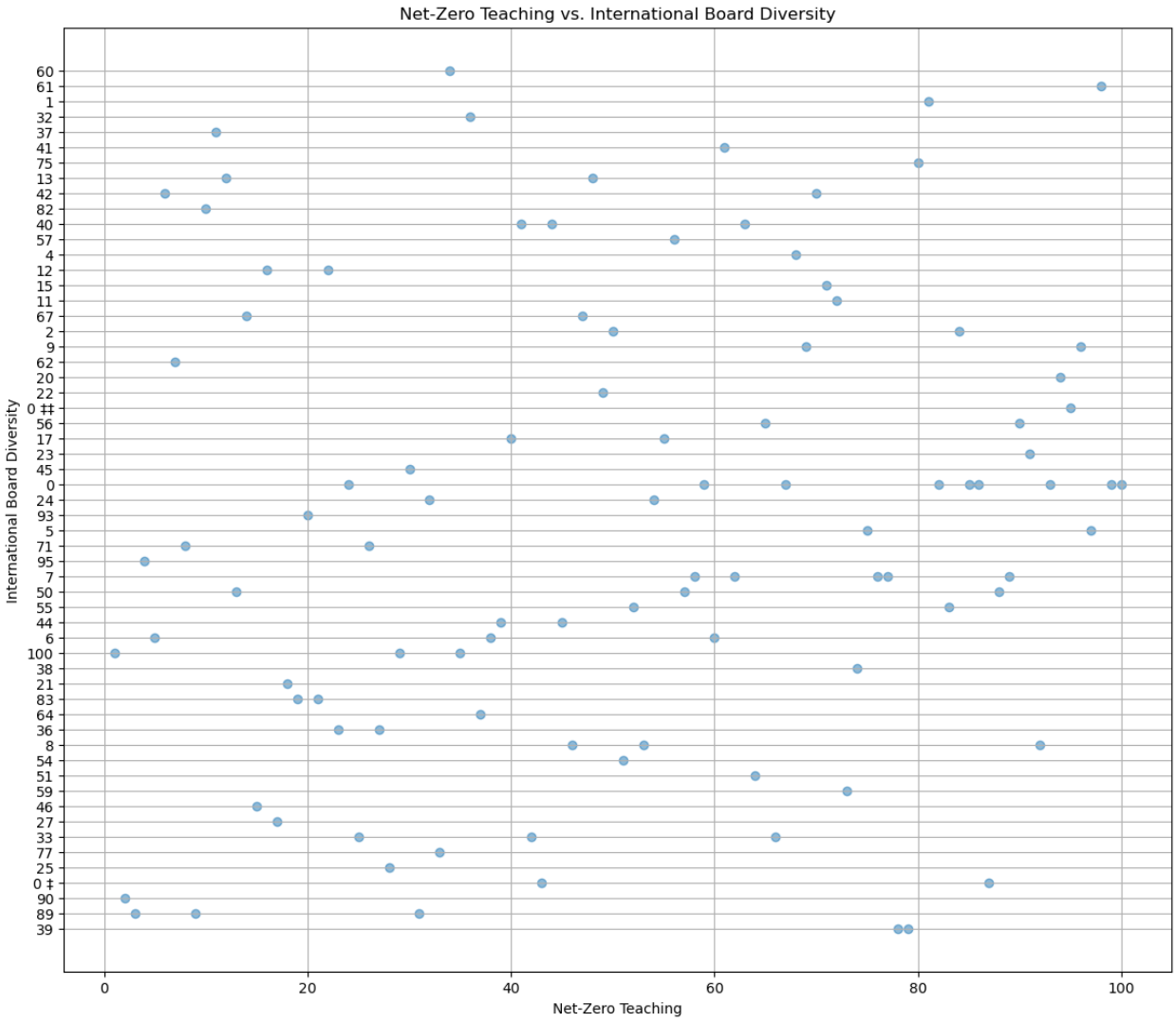
17) which are the dominating schools in the top rankings in 2023?

```
df_sorted = df.sort_values(by='Rank').head(10)
# Create a bar plot to visualize the top 10 universities
plt.figure(figsize=(8, 4))
plt.barh(df_sorted['School Name'], df_sorted['Rank'])
plt.title('Top 10 Universities in 2023')
plt.xlabel('Overall Rank')
plt.ylabel('School Name')
plt.gca().invert_yaxis()  # Invert the y-axis to show the highest-ranked at the top
plt.tight_layout()
plt.show()
```



18) How do MBA programs fare in terms of net-zero teaching and international board diversity?

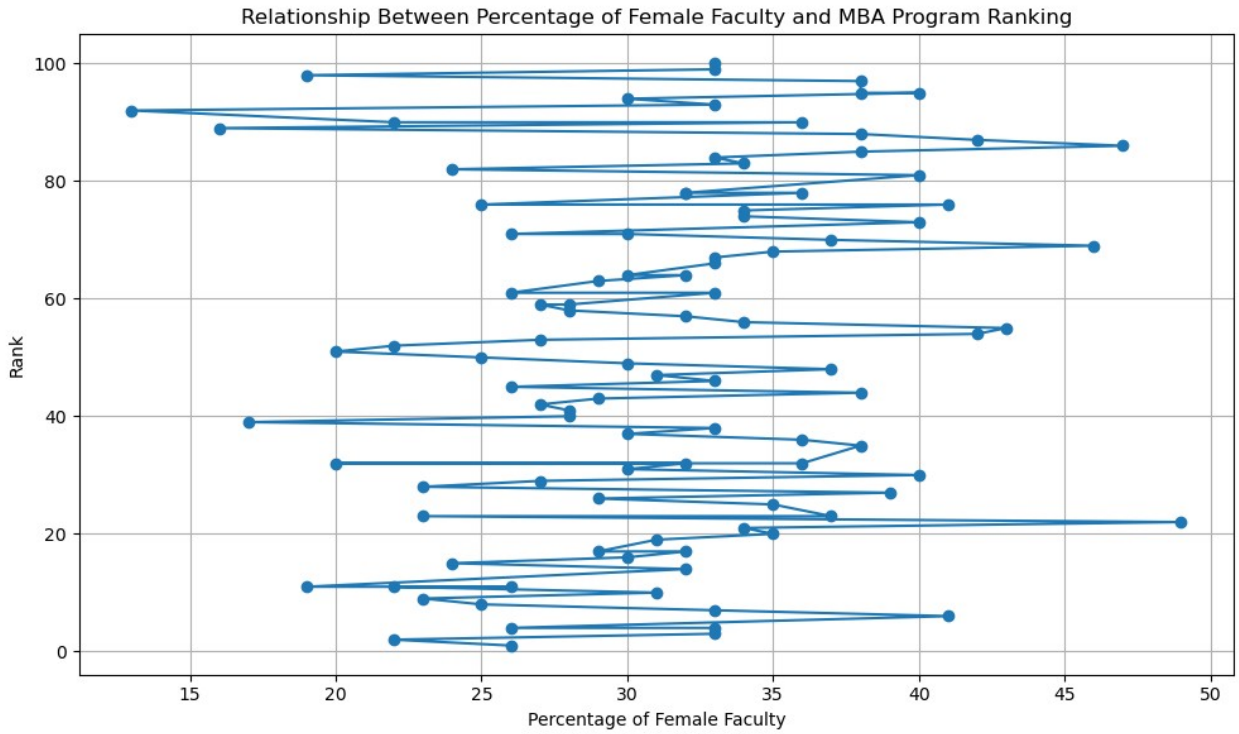
```
import matplotlib.pyplot as plt


net_zero_teaching = df['ESG and net zero teaching rank']
international_board_diversity = df['International board (%)']

plt.figure(figsize=(14, 12))
plt.scatter(net_zero_teaching, international_board_diversity,
alpha=0.5)
plt.title('Net-Zero Teaching vs. International Board Diversity')
plt.xlabel('Net-Zero Teaching')
plt.ylabel('International Board Diversity')
plt.grid(True)
plt.show()
```



19) Is there a relationship between the percentage of female faculty and the ranking of MBA programs?

```
# Specify the columns for the line plot
x_column = 'Female faculty (%)'
y_column = 'Rank'
# Creating a line plot with data points
plt.figure(figsize=(10,6))
plt.plot(df[x_column], df[y_column], marker='o', linestyle='-')
plt.title('Relationship Between Percentage of Female Faculty and MBA
Program Ranking')
plt.xlabel('Percentage of Female Faculty')
plt.ylabel('Rank')
plt.grid(True)
plt.tight_layout()
```

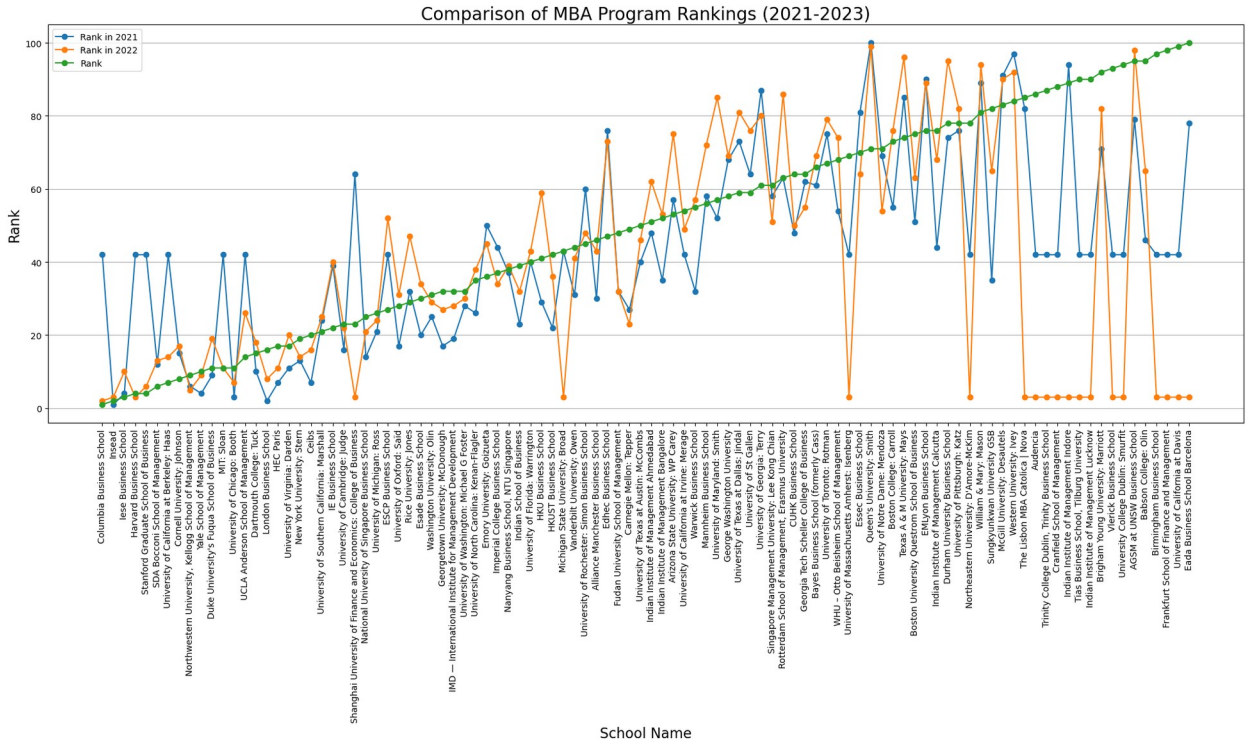
```
plt.show()
```



20) What insights can be gained from comparing the rankings in 2021, 2022, and 2023, and are there any consistent trends or outliers?

```
# Specifing the columns for the line plot
years = ['Rank in 2021', 'Rank in 2022', 'Rank']

# Creating a line plot to compare rankings over the years
plt.figure(figsize=(20, 12))
for year in years:
    plt.plot(df['School Name'], df[year], marker='o', label=year)

plt.title('Comparison of MBA Program Rankings (2021-2023)', fontsize =
20)
plt.xlabel('School Name', fontsize=16)
plt.ylabel('Rank', fontsize=16)
plt.xticks(rotation=90)  # Rotating x-axis labels for readability
plt.legend()
plt.grid(True, axis='y')
```

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

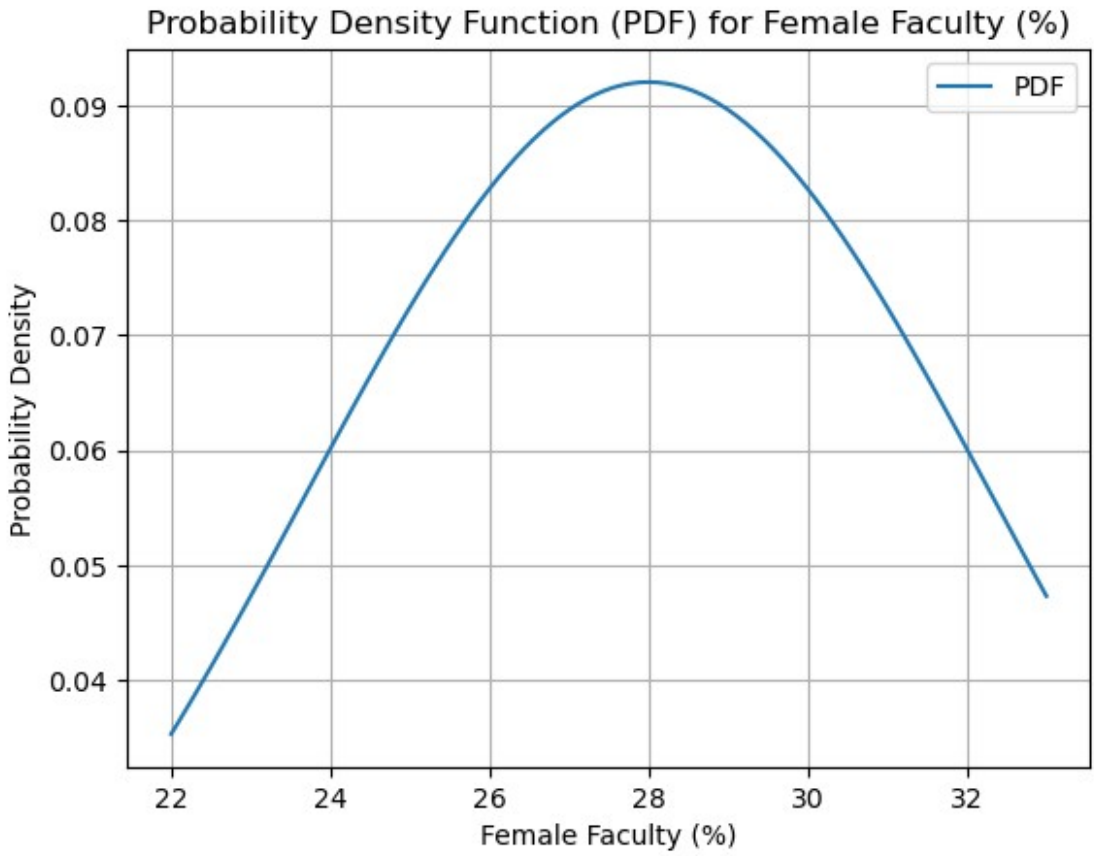


Distribution

probability density function (PDF)

```
# Sample data for the "Female faculty (%)" column
female_faculty_data = [26, 22, 33, 33, 26]

# Calculating the mean and standard deviation of the data
mean = np.mean(female_faculty_data)
std_dev = np.std(female_faculty_data)

# Generating a range of values for the x-axis using NumPy
x = np.linspace(min(female_faculty_data), max(female_faculty_data),
100)

# Calculating the probability density function (PDF) for a normal
distribution
pdf = norm.pdf(x, mean, std_dev)

# Creating a plot to visualize the PDF
plt.plot(x, pdf, label='PDF')
```

```
plt.xlabel("Female Faculty (%)")
plt.ylabel("Probability Density")
plt.title("Probability Density Function (PDF) for Female Faculty (%)")
plt.legend()
plt.grid(True)
plt.show()
```



Hypothesis Testing

T-Test

```
from scipy.stats import ttest_1samp

salaries = [228425,202568,182278,235177,253435]

salaries = pd.Series(salaries)

mean = salaries.mean()
print('Mean salary:', mean)

# Performing one sample t-test
tstat, pval = ttest_1samp(salaries, popmean=200000)
```

```python
print('t-statistic:', tstat)
print('p-value:', pval)

if pval < 0.05:
    print('Reject null hypothesis')
else:
    print('Fail to reject null hypothesis')
```

```
Mean salary: 220376.6
t-statistic: 1.6248497153861678
p-value: 0.1795194563593551
Fail to reject null hypothesis
```