ANALYST: SORNAPUDI PUJITHA

EMAIL: sornapudipujitha@gmail.com

LINKEDIN: PUJITHA SORNAPUDI

INTRODUCTION:

This project explores the cost of education across different countries and education levels. It compares tuition fees, living expenses, and financial aid availability globally. The study highlights affordability challenges faced by international students. Insights aim to support informed decision-making for students and policymakers.

ABOUT DATASET:

This Cost of International Education dataset compiles detailed financial information for students pursuing higher education abroad. It covers multiple countries, cities, and universities around the world, capturing the full tuition and living expenses spectrum alongside key ancillary costs. With standardized fields such as tuition in USD, living-cost indices, rent, visa fees, insurance, and up-to-date exchange rates, it enables comparative analysis across programs, degree levels, and geographies. Whether you're a prospective international student mapping out budgets, an educational consultant advising on affordability, or a researcher studying global education economics, this dataset offers a comprehensive foundation for data-driven insights.

DESCRIPTION :

Country: ISO Country name where the university is located (e.g., Germany, Australia)

City: City in which the institution sits(e.g., Munich, Melbourne)

University: Specific course or major (e.g., MAster of Computer Science , MBA)

Level: Degree level of the program : Undergraduate , Master's, PhD or other certifications.

Duration_Years: Length of the program in years (e.g.., 2 for a typical master's)

Tuition_USD: Total program tuition cost, converted into U.S. dollar for ease of comparison.

Living_Cost_Index: A normalized index(often based on global indices) reflecting relative day-t0-day living expenses(food,transport,utilities)

Rent_USD: Average monthly student accomendation rent in U.S.dollars.

Visa_Fee_USD: One-time visa application fee payable by international students , in U.S dollars.

Insurance_USD: Annual health or student insurance cost in U.S. dollars, as required by many host countries.

Exchange_Rate: Local currancy units per U.S. dollar at the time of data collection- vital for currency conversion and trend analysis if rates fluctuate.

```
In [74]:
```

Import Libraries

```
In [75]:   import numpy as np
           import pandas as pd
           import matplotlib.pyplot as plt
           import seaborn as sns
```

Load Dataset

```
In [76]:   Df=pd.read_excel('/content/Internation Cost Education Dataset.csv.xlsx')
```

```
In [77]:   Df.head()
```

Out[77]:

	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_US
0	USA	Cambridge	Harvard University	Computer Science	Master	2.0	55400	83.5	2200	16
1	UK	London	Imperial College London	Data Science	Master	1.0	41200	75.8	1800	48
2	Canada	Toronto	University of Toronto	Business Analytics	Master	2.0	38500	72.5	1600	23
3	Australia	Melbourne	University of Melbourne	Engineering	Master	2.0	42000	71.2	1400	45
4	Germany	Munich	Technical University of Munich	Mechanical Engineering	Master	2.0	500	70.5	1100	7

Exploratory Data Analysis

In [78]: `Df.head(5)`

Out[78]:

	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_US
0	USA	Cambridge	Harvard University	Computer Science	Master	2.0	55400	83.5	2200	16
1	UK	London	Imperial College London	Data Science	Master	1.0	41200	75.8	1800	48
2	Canada	Toronto	University of Toronto	Business Analytics	Master	2.0	38500	72.5	1600	23
3	Australia	Melbourne	University of Melbourne	Engineering	Master	2.0	42000	71.2	1400	45
4	Germany	Munich	Technical University of Munich	Mechanical Engineering	Master	2.0	500	70.5	1100	7

In [79]: `Df.tail(5)`

Out[79]:

	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_
902	France	Strasbourg	University of Strasbourg	Data Analytics	Master	2.0	4000	70.2	1000	
903	Malaysia	Nilai	USIM	Computer Science	Bachelor	3.0	6800	50.5	400	
904	Saudi Arabia	Al-Ahsa	King Faisal University	Information Systems	Master	2.0	4200	64.2	600	
905	USA	Seattle	University of Washington	Software Development	PhD	5.0	50000	77.8	2000	
906	UK	Nottingham	University of Nottingham	Data Engineering	Master	2.0	34000	61.2	800	

In [80]: `Df.shape`

Out[80]: (907, 12)

In [81]: `Df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 907 entries, 0 to 906
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Country           907 non-null    object
 1   City              907 non-null    object
 2   University        907 non-null    object
 3   Program           907 non-null    object
 4   Level             907 non-null    object
 5   Duration_Years    907 non-null    float64
 6   Tuition_USD       907 non-null    int64
 7   Living_Cost_Index 907 non-null    float64
 8   Rent_USD          907 non-null    int64
 9   Visa_Fee_USD      907 non-null    int64
 10  Insurance_USD     907 non-null    int64
 11  Exchange_Rate     907 non-null    float64
dtypes: float64(3), int64(4), object(5)
memory usage: 85.2+ KB
```

In [82]: `Df.describe(include='number')`

Out[82]:

	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_USD	Insurance_USD	Exchange_Rate
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000
mean	2.836825	16705.016538	64.437486	969.206174	211.396913	700.077178	623.000695
std	0.945449	16582.385275	14.056333	517.154752	143.435740	320.374875	3801.746134
min	1.000000	0.000000	27.800000	150.000000	40.000000	200.000000	0.150000
25%	2.000000	2850.000000	56.300000	545.000000	100.000000	450.000000	0.920000
50%	3.000000	7500.000000	67.500000	900.000000	160.000000	650.000000	1.350000
75%	4.000000	31100.000000	72.200000	1300.000000	240.000000	800.000000	7.150000
max	5.000000	62000.000000	122.400000	2500.000000	490.000000	1500.000000	42150.000000

In [83]: `Df.describe(include='object')`

Out[83]:

	Country	City	University	Program	Level
count	907	907	907	907	907
unique	71	556	622	92	3
top	UK	Singapore	University of Washington	Computer Science	Master
freq	93	18	6	312	451

In [84]: `Df.isnull().sum()`

Out[84]:

	0
Country	0
City	0
University	0
Program	0
Level	0
Duration_Years	0
Tuition_USD	0
Living_Cost_Index	0
Rent_USD	0
Visa_Fee_USD	0
Insurance_USD	0
Exchange_Rate	0

dtype: int64

```python
In [85]:  Tuition_mean=Df[Df['Tuition_USD']!=0]['Tuition_USD'].mean()
          Df['Tuition_USD']=Df['Tuition_USD'].replace(0,Tuition_mean)
```

Basic Level Analysis

```python
In [86]:  # 1. How many countries are present in the dataset ?
          Df['Country'].unique()
```

```
Out[86]:  array(['USA', 'UK', 'Canada', 'Australia', 'Germany', 'Japan',
                 'Netherlands', 'Singapore', 'France', 'Switzerland', 'Sweden',
                 'Denmark', 'China', 'South Korea', 'Ireland', 'New Zealand',
                 'Austria', 'Belgium', 'Hong Kong', 'Portugal', 'Israel', 'Taiwan',
                 'Czech Republic', 'India', 'Poland', 'Malaysia', 'Spain', 'Italy',
                 'Finland', 'Norway', 'Brazil', 'Turkey', 'Russia', 'Mexico',
                 'Greece', 'Thailand', 'UAE', 'South Africa', 'Egypt', 'Argentina',
                 'Indonesia', 'Saudi Arabia', 'Nigeria', 'Vietnam', 'Hungary',
                 'Iceland', 'Colombia', 'Romania', 'Luxembourg', 'Tunisia',
                 'Cyprus', 'Croatia', 'Dominican Republic', 'Morocco', 'Peru',
                 'Ecuador', 'Lebanon', 'Bahrain', 'Uruguay', 'Bulgaria', 'Ghana',
                 'Algeria', 'Panama', 'Bangladesh', 'Kuwait', 'Ukraine', 'Slovenia',
                 'Serbia', 'Iran', 'Uzbekistan', 'El Salvador'], dtype=object)
```

```python
In [87]:  Df['Country'].nunique()
```

```
Out[87]:  71
```

```python
In [88]:  #2.Which country has the highest cost of education?
          highest=Df.loc[Df['Tuition_USD'].idxmax()]
          print(f"Country with highest cost of education: {highest['Country']} (${highest['Tuition_USD']}")
```

```
          Country with highest cost of education: USA ($62000.0
```

```python
In [89]:  Countries=Df.groupby('Country')['Tuition_USD'].max()
          Top_5_Tuition=Countries.sort_values(ascending=False).head(7)
          print(Top_5_Tuition)
```

```
            Country
            USA              62000.0
            Australia        45800.0
            Canada           45200.0
            UK               45000.0
            Singapore        42000.0
            New Zealand      32000.0
            UAE              29500.0
            Name: Tuition_USD, dtype: float64
```

In [90]:
```python
Countries=Df.sort_values(by='Tuition_USD',ascending=False).drop_duplicates('Country').head(7)
x=Countries['Country']
y=Countries['Tuition_USD']
plt.figure(figsize=(10,5))
plt.bar(x,y,color='brown')
plt.gca().set_facecolor('#EFDFD8')
plt.title('top_7_Countries with highest cost of Educations',fontsize=12,fontweight='bold',color='black')
plt.xlabel('Countries',fontweight='bold')
plt.ylabel('Tuition_Cost',fontweight='bold')
plt.xticks(rotation=45)
plt.show()
```

top_7_Countries with highest cost of Educations
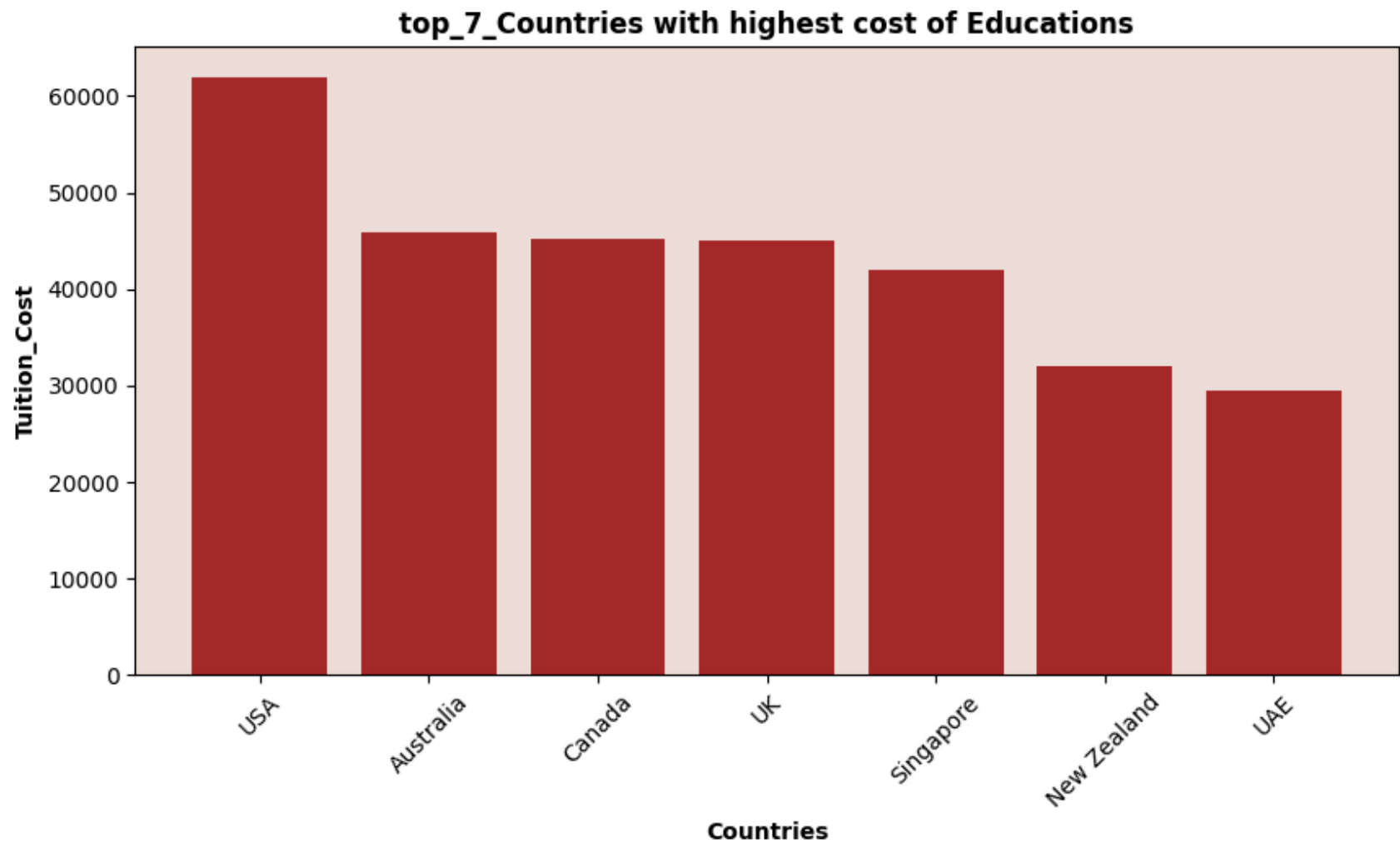

```
In [91]:  #3.What is the average education cost across all countries?
          Average_Education_Cost=Df['Tuition_USD'].mean()
          print(f'Average_Education_Cost:',Average_Education_Cost)
```

Average_Education_Cost: 18845.087064676616

```
In [92]:  #4.Which countries have an education cost below the global average?
          filter_df=Df[Df['Tuition_USD']>0]
          global_Average=filter_df['Tuition_USD'].mean()
          below_avg_df=filter_df[filter_df['Tuition_USD']<global_Average]
```

```python
countries_below_average=below_avg_df['Country'].unique()
countries_below_average
```

Out[92]: array(['Germany', 'Japan', 'Netherlands', 'France', 'Switzerland',
               'China', 'South Korea', 'Austria', 'Belgium', 'Portugal', 'Israel',
               'Taiwan', 'Czech Republic', 'Singapore', 'India', 'Poland',
               'Malaysia', 'Spain', 'Italy', 'Turkey', 'Russia', 'Mexico',
               'Greece', 'Thailand', 'South Africa', 'Indonesia', 'Saudi Arabia',
               'Nigeria', 'Vietnam', 'Denmark', 'UK', 'Ireland', 'Hungary',
               'Egypt', 'Brazil', 'Colombia', 'Romania', 'Luxembourg', 'Tunisia',
               'Cyprus', 'Croatia', 'Dominican Republic', 'Morocco', 'Peru',
               'Ecuador', 'Bahrain', 'Bulgaria', 'Ghana', 'Algeria', 'Panama',
               'Lebanon', 'Uruguay', 'Bangladesh', 'Kuwait', 'Ukraine',
               'Slovenia', 'Serbia', 'Iran', 'Uzbekistan', 'El Salvador'],
              dtype=object)

In [93]:
```python
#5. Which Coutry has the highest rent?
Highest_Rent_Cost=Df.loc[Df['Rent_USD'].idxmax()]
Highest_Rent_Cost
```

Out[93]:

	131
Country	USA
City	New York
University	Columbia University
Program	Data Science
Level	Master
Duration_Years	2.0
Tuition_USD	52800.0
Living_Cost_Index	100.0
Rent_USD	2500
Visa_Fee_USD	160
Insurance_USD	1500
Exchange_Rate	1.0

dtype: object

In [74]:
```python
top=Df.groupby('Country')[['Country','Rent_USD']].max().sort_values(by='Rent_USD',ascending=False).head(10)
plt.figure(figsize=(10,5))
plt.bar(" (" + top['Country'] + ")", top['Rent_USD'], color='brown')
plt.gca().set_facecolor('#EFDFD8')
plt.title('Top_10_Countries with Highest Rent_Cost',fontsize=12,fontweight='bold',color='black')
plt.xlabel('Countries',fontweight='bold',color='black')
plt.ylabel('Rent_Cost',fontweight='bold',color='black')
plt.xticks(rotation=90)
plt.show()
```

## Top_10_Countries with Highest Rent_Cost
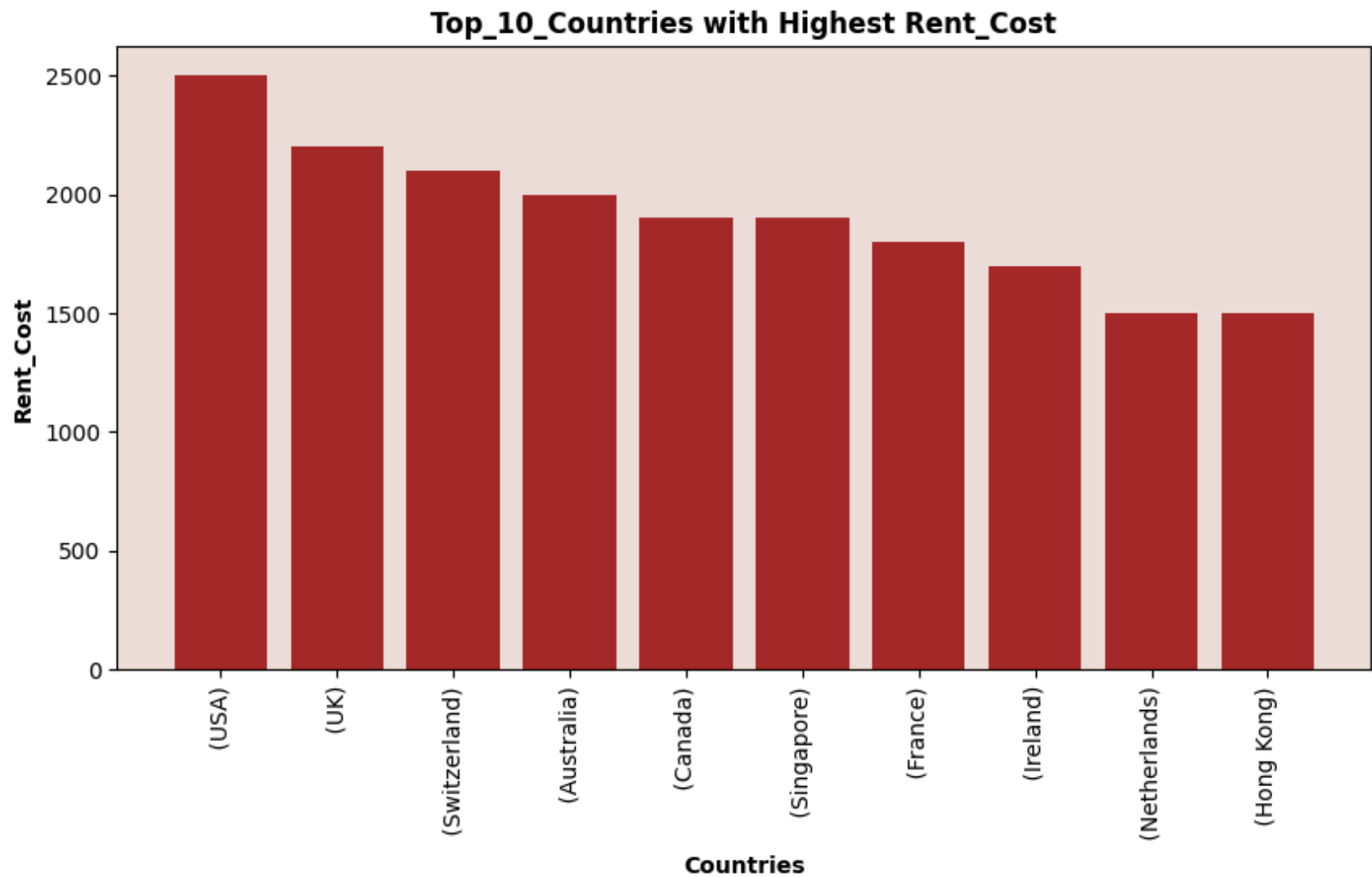

```
In [94]:   #5.List the top 10 countries with the lowest education cost.
           min_tuition_per_country=Df.groupby('Country')['Tuition_USD'].min()
           top10_lowest=min_tuition_per_country.sort_values(ascending=True).head(10)
           top10_lowest
```

Out[94]:

	Tuition_USD
Country	
Luxembourg	400.0
Germany	450.0
Mexico	800.0
Algeria	900.0
Turkey	1000.0
Tunisia	1200.0
Switzerland	1380.0
Austria	1500.0
India	1600.0
Morocco	1700.0

dtype: float64

In [95]:
```python
plt.figure(figsize=(10,5))
plt.bar(top10_lowest.index,top10_lowest.values,color='brown')
plt.title('Top_10_Countries with Lowest Tuition_Cost',  fontsize=12,fontweight='bold',color='black')
plt.xlabel('Countries',fontweight='bold',color='black')
plt.ylabel('Tuition_Cost',fontweight='bold',color='black')
plt.xticks(rotation=90)
plt.show()
```

Top_10_Countries with Lowest Tuition_Cost
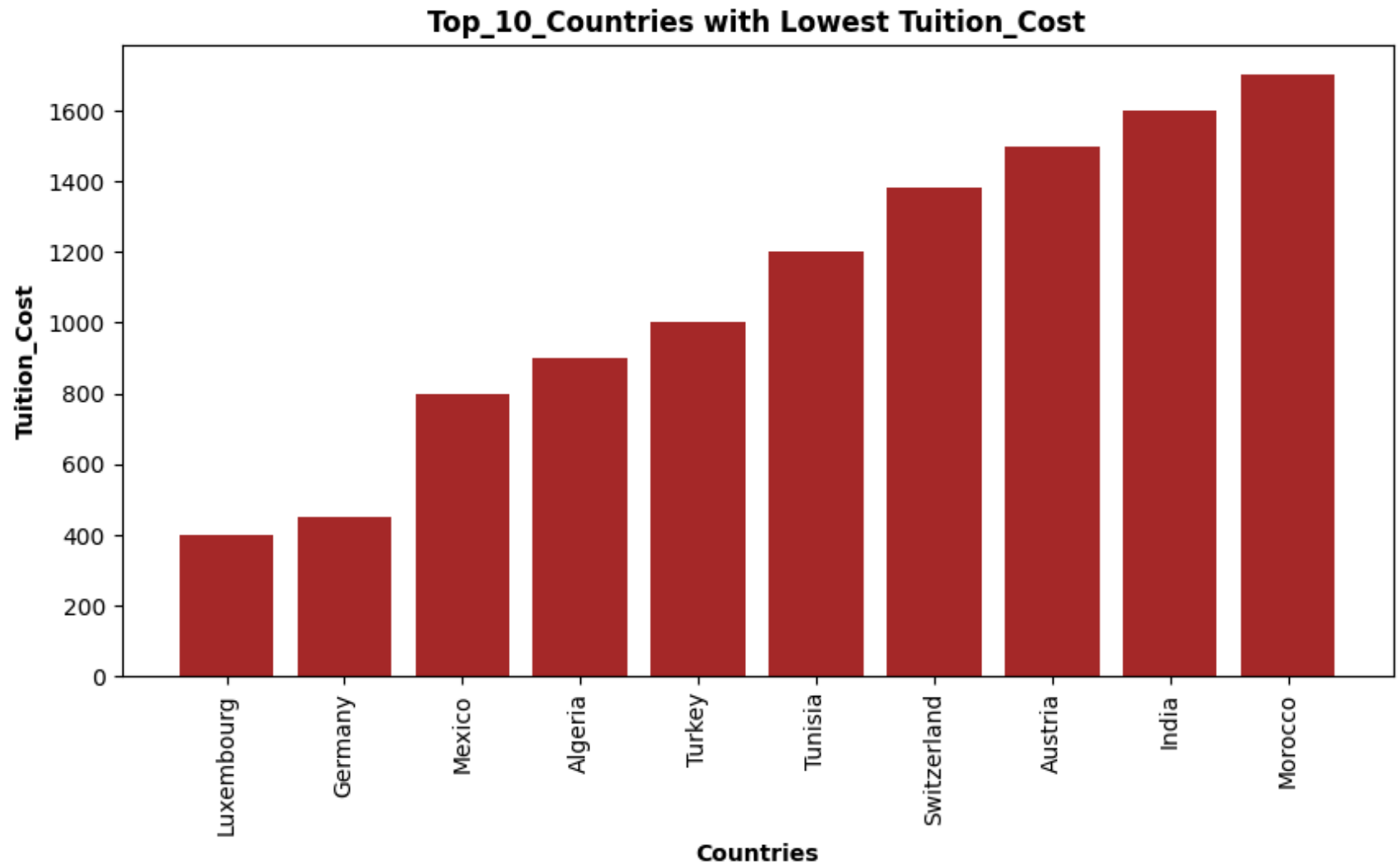

```
In [96]:  #6.How many countries have an education cost between $5,000 and $10,000?
          Countries=Df[(Df['Tuition_USD']>=5000) & (Df['Tuition_USD']<=10000)]
          Countries.head(10)
```

Out[96]:

	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_
5	Japan	Tokyo	University of Tokyo	Information Science	Master	2.0	8900.0	76.4	1300	
12	China	Beijing	Tsinghua University	Computer Engineering	Master	2.5	8900.0	52.3	800	
13	South Korea	Seoul	Seoul National University	Digital Media	Master	2.0	7200.0	68.7	900	
15	South Korea	Busan	Pusan National University	Business	Bachelor	4.0	5900.0	62.4	700	
27	Taiwan	Taipei	National Taiwan University	Electronics Engineering	Bachelor	4.0	5200.0	62.8	700	
28	Czech Republic	Prague	Charles University	Medicine	Bachelor	3.0	9800.0	54.6	650	
36	Singapore	Singapore	NTU Singapore	Data Science	PhD	4.0	8900.0	81.5	1600	
48	South Korea	Daejeon	KAIST	Computer Science	PhD	4.0	6900.0	65.4	700	
60	Japan	Osaka	Osaka University	Engineering	Bachelor	4.0	6800.0	71.2	900	
61	Japan	Kyoto	Kyoto University	Computer Science	Master	2.0	8500.0	71.8	900	

In [97]:
```python
#7. How many unique university present in the dataset?
Number_University=Df['University'].nunique()
Name_University=Df['University'].unique()
print(f'Number_University:',Number_University)
print(Name_University[:10])
```

```
Number_University: 622
['Harvard University' 'Imperial College London' 'University of Toronto'
 'University of Melbourne' 'Technical University of Munich'
 'University of Tokyo' 'University of Amsterdam'
 'National University of Singapore' 'Sorbonne University' 'ETH Zurich']
```

In [98]:
```python
#8.What is The total Visa Fee Across all entries?
Total_Visa=Df['Visa_Fee_USD'].sum()
print(f'Total_Visa:',Total_Visa)
```

```
Total_Visa: 191737
```

In [99]:
```python
#9.How many Entries Have insurance costs above $1,000?
Df[Df['Insurance_USD']>1000][['Country','Insurance_USD']]
```

Out[99]:

	Country	Insurance_USD
0	USA	1500
9	Switzerland	1200
19	USA	1500
20	USA	1500
21	USA	1500
...	...	...
877	USA	1500
884	USA	1500
891	USA	1500
898	USA	1500
905	USA	1500

98 rows × 2 columns

In [100…
```python
#10.Different type of program and there Duration_year
Df.groupby(['Program','Duration_Years'])[['Program','Duration_Years']].value_counts()
```

Out[100...

| | | count |
Program	Duration_Years	
AI Engineering	2.0	1
Aerospace Engineering	2.0	1
	4.0	1
Applied Mathematics	2.0	1
Architecture	3.0	1
...	...	...
Software Systems	4.0	1
Sustainable Energy	2.0	1
Sustainable Technology	2.0	1
Systems Engineering	2.0	1
Technology & Innovation	2.0	1

160 rows × 1 columns

dtype: int64

In [101...

```python
# different Program course available in Bachelor Degree
Filter_educations3=Df[(Df['Level']=='Bachelor')]
Result=Filter_educations3.groupby(['Level','Duration_Years'])[['Level','Duration_Years','Program']].value_counts()
Result
```

Out[101...

			count
Level	**Duration_Years**	**Program**	
Bachelor	**3.0**	**Computer Science**	83
		Computer Engineering	9
		Data Science	9
		Information Systems	8
		Data Analytics	5
		Artificial Intelligence	4
		Physics	4
		Software Engineering	4
		Software Development	3
		Digital Innovation	2
		Economics	2
		Environmental Science	2
		Information Technology	2
		Marine Biology	2
		Mathematics	2
		Mechanical Engineering	2
		Architecture	1
		Biomedical Sciences	1
		Business	1
		Business Informatics	1
		Computer Systems	1

			count
Level	**Duration_Years**	**Program**	
		Computing	1
		Computing Systems	1
		Design	1
		Digital Business	1
		Digital Systems	1
		Energy Engineering	1
		Engineering	1
		International Business	1
		Mechatronics	1
		Medicine	1
		Robotics	1
		Social Sciences	1
	4.0	Computer Science	67
		Computer Engineering	17
		Software Engineering	12
		Information Systems	8
		Information Technology	5
		Software Development	4
		Data Analytics	3
		Data Science	3
		Artificial Intelligence	2

			count
Level	Duration_Years	Program	
		Aerospace Engineering	1
		Architecture	1
		Business	1
		Chemical Engineering	1
		Computer Science & Design	1
		Data Engineering	1
		Electronics Engineering	1
		Engineering	1
		Finance	1
		Industrial Engineering	1
		Information Science	1
		Mathematics	1
		Political Science	1
		Psychology	1
		Software Systems	1
	5.0	Software Engineering	1

dtype: int64

In [102…
```python
# different programs courses available in Master level
Filter_Education2=Df[(Df['Level']=='Master')]
result=Filter_Education2.groupby(['Level','Duration_Years'])[['Level','Duration_Years','Program']].value_counts()
result
```

Out[102...

| | | | count |
Level	Duration_Years	Program	
Master	1.0	Computer Science	6
		Data Science	4
		Data Analytics	3
		Artificial Intelligence	1
		Cybersecurity	1
	...	...	...
	3.0	Computer Engineering	2
		Computer Science	2
		Data Science	2
		Information Technology	1
		Software Engineering	1

76 rows × 1 columns

dtype: int64

In [103...

```python
#different programs courses available in PHD level
Filter_Education1=Df[(Df['Level']=='PhD')]
result=Filter_Education1.groupby(['Level','Duration_Years'])[['Level','Duration_Years','Program']].value_counts()
result
```

Out[103...

Level	Duration_Years	Program	count
PhD	3.0	Artificial Intelligence	2
		Molecular Biology	1
	4.0	Computer Science	42
		Artificial Intelligence	14
		Data Science	13
		Software Engineering	9
		Data Analytics	8
		Computer Engineering	7
		Information Systems	5
		Information Technology	5
		Robotics	4
		Software Development	4
		Data Engineering	3
		Physics	3
		Chemistry	2
		Mathematics	2
		Biomedical Sciences	1
		Biotechnology	1
		Climate Science	1
		Computing Science	1
		Cybersecurity	1

	count
Level	**Duration_Years**	**Program**	
		Digital Systems & AI	1
		Electrical Engineering	1
		Electronics	1
		Electronics Engineering	1
		Environmental Computing	1
		Forestry Sciences	1
		Industrial Engineering	1
		Marine Technology	1
		Mechanical Engineering	1
		Mechatronics	1
		Neuroscience	1
		Pharmaceutical Sciences	1
		Quantum Computing	1
		Renewable Energy	1
	5.0	Computer Science	6
		Robotics	2
		Artificial Intelligence	1
		Chemistry	1
		Computer Engineering	1
		Electrical Engineering	1
		Engineering	1

			count
Level	**Duration_Years**	**Program**	
		Machine Learning	1
		Physics	1
		Software Development	1

dtype: int64

In [104...
```python
#11.highest Rate of Visa Country?
Df.loc[Df['Visa_Fee_USD'].idxmax()]
```

Out[104...

	234
Country	UK
City	Cambridge
University	University of Cambridge
Program	Computer Science
Level	Bachelor
Duration_Years	3.0
Tuition_USD	38500.0
Living_Cost_Index	78.5
Rent_USD	1800
Visa_Fee_USD	490
Insurance_USD	800
Exchange_Rate	0.79

dtype: object

```
In [105…   #what are the most popular program subject ?
           Program=Df['Program'].value_counts().head(15)
           Program
```

Out[105…

	count
Program	
Computer Science	312
Data Science	82
Computer Engineering	71
Software Engineering	57
Artificial Intelligence	53
Data Analytics	50
Information Systems	40
Information Technology	24
Software Development	22
Data Engineering	19
Robotics	14
Physics	11
Mathematics	10
Engineering	9
Mechanical Engineering	6

dtype: int64

```
In [106…   x=Program.index
           y=Program.values
           plt.figure(figsize=(10,5))
```

```python
plt.bar(x,y,color='brown')
plt.gca().set_facecolor('#EFDFD8')
plt.title('Top_15_FamousProgram',fontsize=12,fontweight='bold',color='black')
plt.xlabel('Program',fontweight='bold')
plt.ylabel('Count of program',fontweight='bold')
plt.xticks(rotation=90)
plt.show()
```

Top_15_FamousProgram
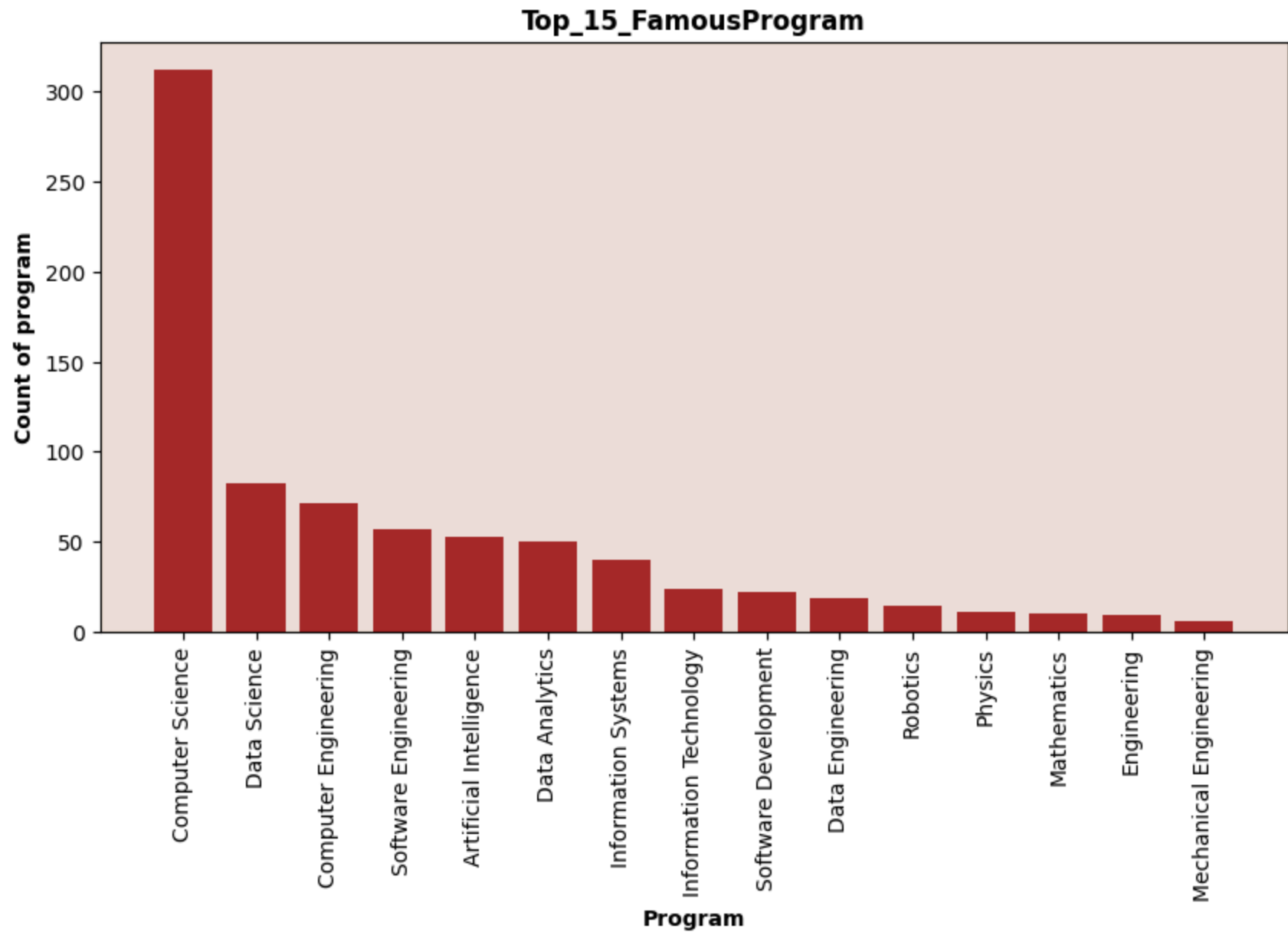

```
#Which country do most people prefer to pursue their higher education in?"
Countries=Df['Country'].value_counts().head(10)
Countries
```

Out[107…

	count
Country	
UK	93
Australia	86
USA	78
Canada	76
Germany	33
France	27
South Korea	23
Netherlands	21
Switzerland	20
Singapore	18

dtype: int64

In [108…

```python
x=Countries.index
y=Countries.values
plt.figure(figsize=(10,5))
plt.bar(x,y,color='brown')
plt.gca().set_facecolor('#EFDFD8')
plt.title('Top_10_Countries most people prefer to pursue their Higher education',fontweight='bold',color='black')
plt.xlabel('Country',fontweight='bold')
plt.ylabel('Student preference Count',fontweight='bold')
plt.xticks(rotation=90)
plt.show()
```

Top_10_Countries most people prefer to pursue their Higher education
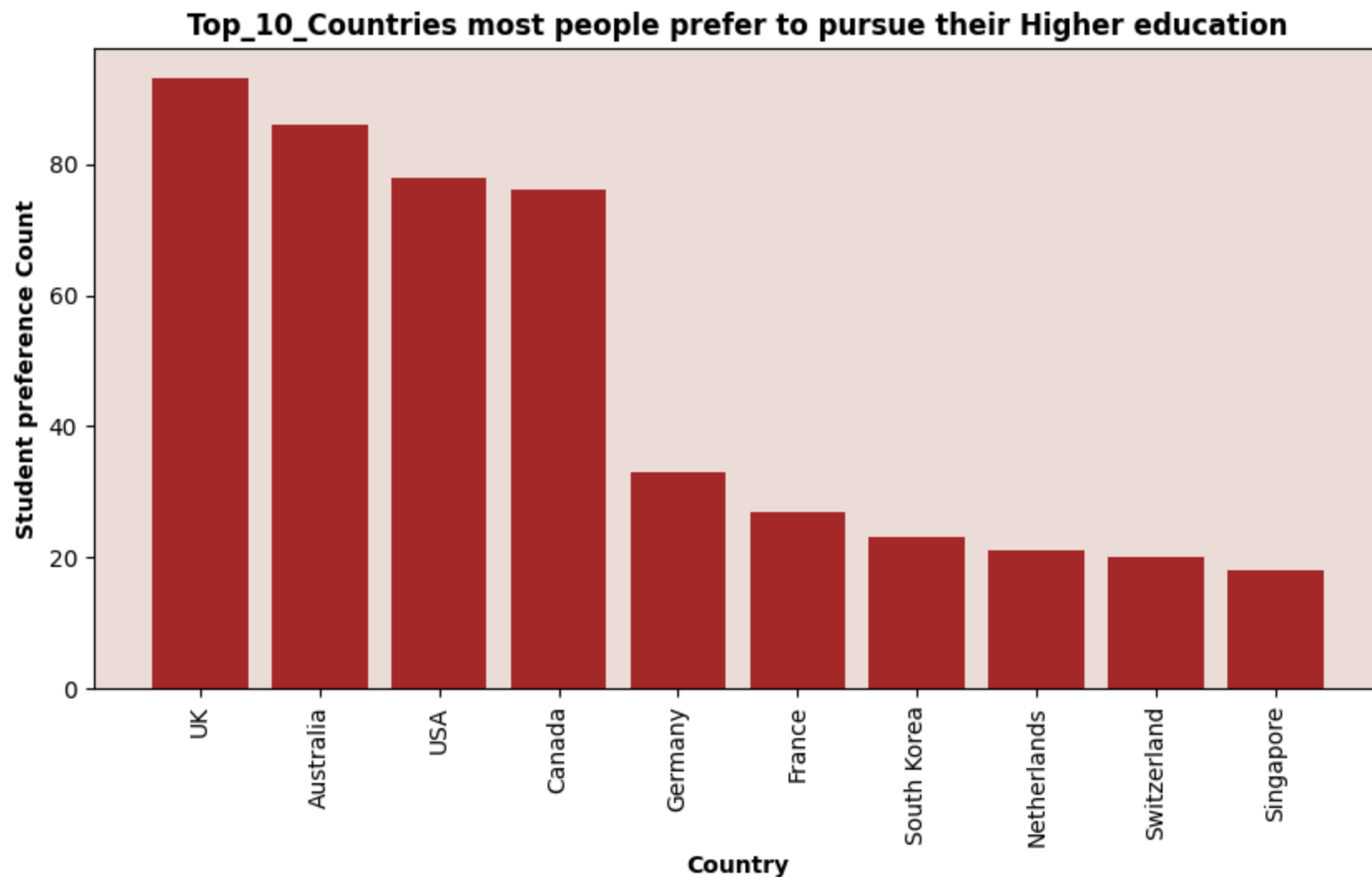
Intermediate Analysis

```
In [109…  #1.Which country has the highest average total cost?
          Df['Total_Cost']=Df['Tuition_USD']+Df['Living_Cost_Index']+Df['Rent_USD']+Df['Visa_Fee_USD']+Df['Insurance_USD']
          Result=Df.groupby('Country')['Total_Cost'].mean().idxmax()
          Result
```

Out[109...   'USA'

In [110...
```python
#2. Group Countries into cost brackets (Low,Medium,High) and show the count in each
def Cost_brackets(Total_Cost):
    if Total_Cost <= 15000:
        return 'Low'
    elif Total_Cost <= 40000:
        return 'Medium'
    else:
        return 'High'
Df['Cost_Bracket'] = Df['Total_Cost'].apply(Cost_brackets)
brackets_Counts = Df['Cost_Bracket'].value_counts()
print(brackets_Counts)
```

```
Cost_Bracket
Low        401
Medium     377
High       129
Name: count, dtype: int64
```

Apply() function is used to apply a function to each element (or row/column) of a pandas Series or Dataframe.

In [111...
```python
#3.what percentage of countries fall above and below a specific cost thereshold (e.g 10,000)
threshold=10000
total_countries=len(Df)
below_threshold=len(Df[Df['Total_Cost']<=threshold])
above_threshold=len(Df[Df['Total_Cost']>threshold])
Below_Percentage=(below_threshold/total_countries)*100
Above_Percentage=(above_threshold/total_countries)*100
print(f"Percentage of countries below or equal to ${threshold}: {Below_Percentage:.2f}%")
print(f"Percentage of countries above ${threshold}: {Above_Percentage:.2f}%")
```

```
Percentage of countries below or equal to $10000: 41.35%
Percentage of countries above $10000: 58.65%
```

In [112...
```python
#4.what is correlation between tuition_fee and rent?
Correlation=Df['Tuition_USD'].corr(Df['Rent_USD'])
print(f"Correlation between tuition_USD & Rent_USD: {Correlation:.2f}")
```

```
Correlation between tuition_USD & Rent_USD: 0.74
```

```python
plt.figure(figsize=(8, 6))
sns.regplot(x='Tuition_USD', y='Rent_USD', data=Df, color='brown', marker='o')
plt.title('Correlation between Tuition Fee and Rent')
plt.xlabel('Tuition Fee (USD)')
plt.ylabel('Rent (USD)')
plt.grid(True)
plt.tight_layout()
plt.show()
```

Correlation between Tuition Fee and Rent
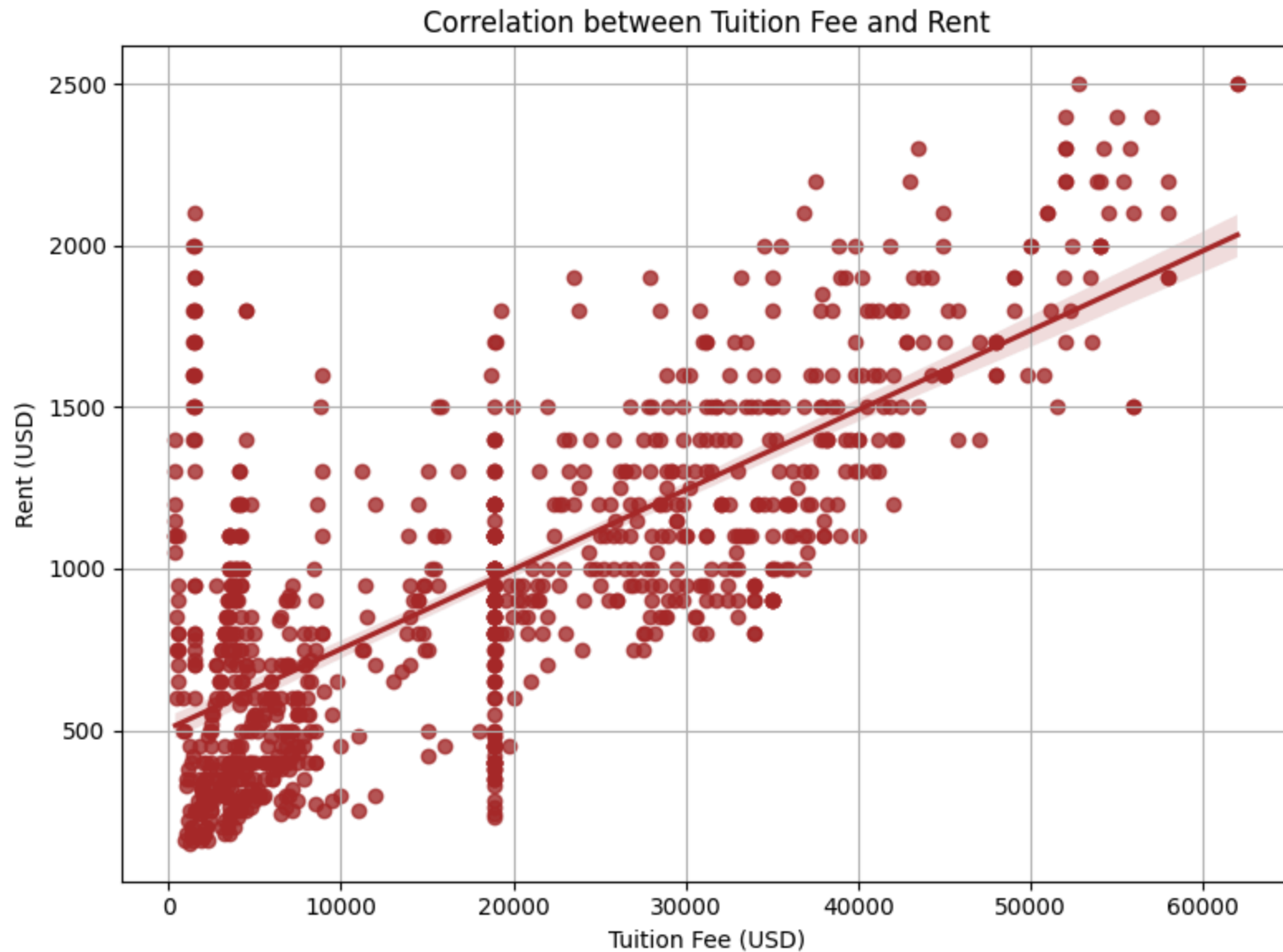
This chart reveals a clear positive relationship: higher tuition fees are often associated with higher living costs, highlighting how studying abroad can be a financial double impact.

```
In [114…   #5.which 5 Countries have the highest average living+rent cost combined?
           Df['Living+Rent']=Df['Living_Cost_Index']+Df['Rent_USD']
```

```python
top_5_Countries=Df.groupby('Country')['Living+Rent'].mean().sort_values(ascending=False).head(5)
round(top_5_Countries,2)
```

Out[114…

	Living+Rent
Country	
USA	1936.77
Switzerland	1787.82
Hong Kong	1578.60
Singapore	1476.47
Australia	1429.87

dtype: float64

In [115…
```python
#6.Calculate the percentage contribution of each country's cost to the global total.
Total_Cost_Mean=Df['Total_Cost'].sum()
Df['Cost_Percentage']=(Df['Total_Cost']/Total_Cost_Mean)*100
Df['Cost_Percentage']=Df['Cost_Percentage'].round(2)
```

In [116…
```python
Df[['Country','Cost_Percentage']]
```

Out[116...

	Country	Cost_Percentage
0	USA	0.31
1	UK	0.24
2	Canada	0.22
3	Australia	0.24
4	Germany	0.01
...	...	...
902	France	0.03
903	Malaysia	0.04
904	Saudi Arabia	0.03
905	USA	0.28
906	UK	0.19

907 rows × 2 columns

In [117...

```python
#7.what is the median total cost by country
Median=Df.groupby('Country')['Total_Cost'].median()
Median
```

Out[117…]

	Total_Cost
Country	
Algeria	1636.400000
Argentina	19524.237065
Australia	36667.050000
Austria	3331.200000
Bahrain	8871.500000
...	...
USA	53737.800000
Ukraine	4313.200000
Uruguay	19661.587065
Uzbekistan	4169.800000
Vietnam	4063.600000

71 rows × 1 columns

dtype: float64

In [118…]
```python
#8.Find how many countries have tuition fees above the dataset's average tuition
Average_Tuition=Df['Tuition_USD'].mean()
Above_Average=Df[Df['Tuition_USD']>Average_Tuition]
Countries_above_Avg=Above_Average['Country'].nunique()
CountryName_Above_avg=Above_Average['Country'].unique()
print(f"Number of countries with tuition above average:{Countries_above_Avg}")
print(f"Name of Countries with tuition Above average:\n{CountryName_Above_avg}")
```

```
Number of countries with tuition above average:11
Name of Countries with tuition Above average:
['USA' 'UK' 'Canada' 'Australia' 'Singapore' 'Ireland' 'New Zealand'
 'Hong Kong' 'UAE' 'Egypt' 'Lebanon']
```

```
In [119…    Df.head(5)
```

Out[119…

	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_US
0	USA	Cambridge	Harvard University	Computer Science	Master	2.0	55400.0	83.5	2200	16
1	UK	London	Imperial College London	Data Science	Master	1.0	41200.0	75.8	1800	48
2	Canada	Toronto	University of Toronto	Business Analytics	Master	2.0	38500.0	72.5	1600	23
3	Australia	Melbourne	University of Melbourne	Engineering	Master	2.0	42000.0	71.2	1400	45
4	Germany	Munich	Technical University of Munich	Mechanical Engineering	Master	2.0	500.0	70.5	1100	7

Advance Level Analysis

```
In [120…    #1.Which country has the highest average total cost(tuition+rent+insurance+visa fees) for a master's Program?
           Average=Df.groupby('Country')['Total_Cost'].mean()
           highest_Average_Country_wise=Average.sort_values(ascending=False).head(10)
           round(highest_Average_Country_wise,2)
```

Out[120…

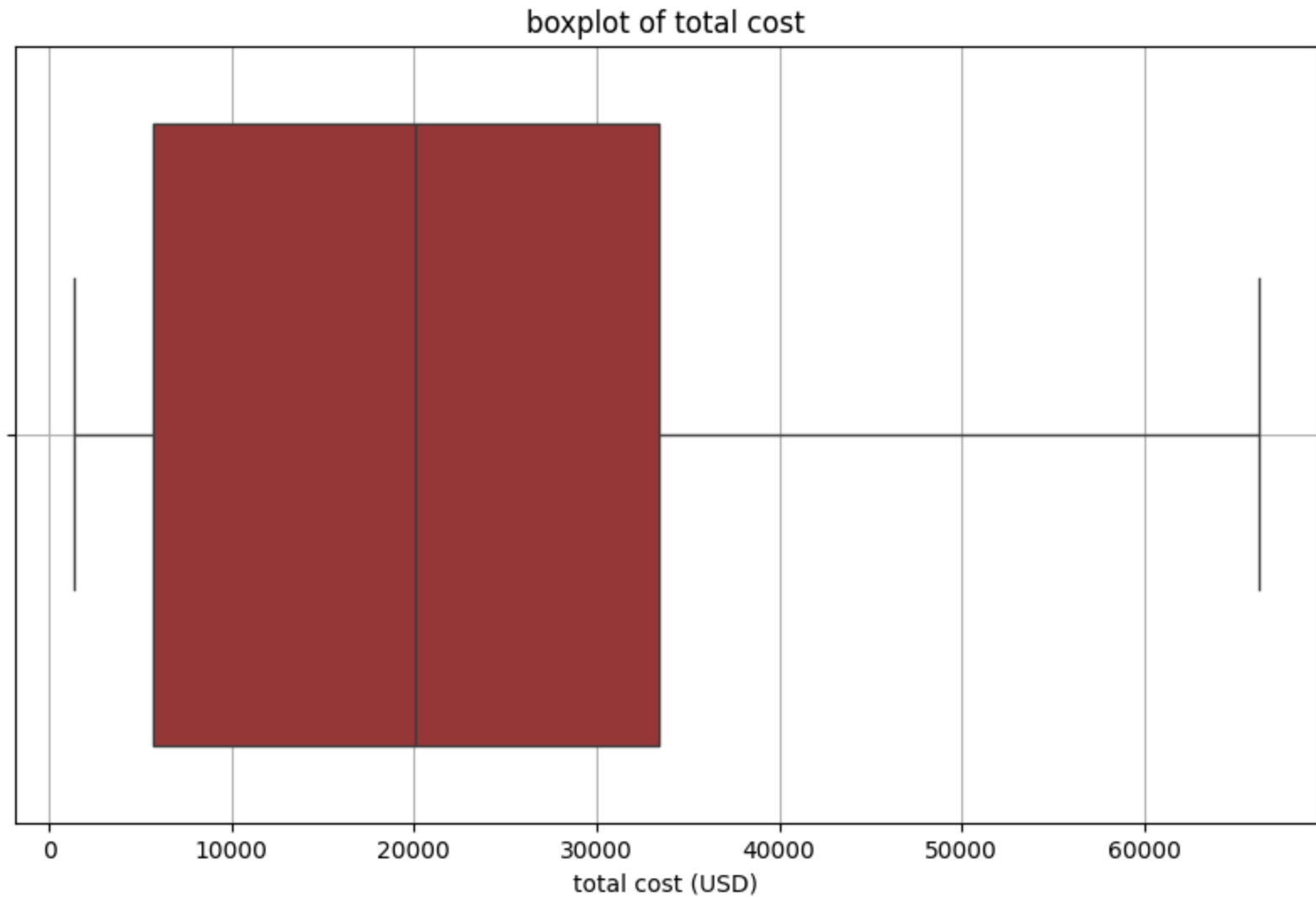|	Total_Cost
Country	
USA	51339.08
Australia	37140.34
Singapore	36538.69
UK	32916.35
Canada	31605.19
New Zealand	31332.31
UAE	26668.58
Hong Kong	24298.60
Iceland	20983.50
Norway	20893.32

dtype: float64

In [121…

```python
#2. Are there any countries where the total cost is a significant outlier compared to the global average?
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
Q1=Df['Total_Cost'].quantile(0.25)
Q2=Df['Total_Cost'].quantile(0.75)
IQR=Q2-Q1
print(f"Q1 value:{Q1}")
print(f"Q2 value:{Q2}")
print(f"InterQuartile_Range:{IQR}")
Lower_bound=Q1-1.5*IQR
Upper_bound=Q2+1.5*IQR
print(f"Lower_bound:{Lower_bound}")
print(f"Upper_bound:{Upper_bound}")
```

```
iqr_outlier=Df[(Df['Total_Cost']<Lower_bound)|(Df['Total_Cost']>Upper_bound)]
print(iqr_outlier[['Country','University','Total_Cost']])
```

```
Q1 value:5708.5
Q2 value:33408.35
InterQuartile_Range:27699.85
Lower_bound:-35841.274999999994
Upper_bound:74958.125
Empty DataFrame
Columns: [Country, University, Total_Cost]
Index: []
```

In [127…
```
plt.figure(figsize=(10,6))
sns.boxplot(x=Df['Total_Cost'],color='brown')
plt.title("boxplot of total cost")
plt.xlabel("total cost (USD)")
plt.grid(True)
plt.show()
```

boxplot of total cost

total cost (USD)

In [123…
```
#3.How do average total education costs differ across regions (e.g., Europe, Asia, North America),
# and which region offers the best balance between cost and living standards?
Mean_Country=Df.groupby('Country')[['Total_Cost','Living_Cost_Index']].mean()
Mean_Country['Balance_Score']=Mean_Country['Living_Cost_Index']/Mean_Country['Total_Cost']
Mean_Country=Mean_Country.sort_values(by='Balance_Score',ascending=False)
print(Mean_Country)
```

```
            Total_Cost  Living_Cost_Index  Balance_Score
Country
Luxembourg   2564.614286          78.900000       0.030765
Algeria      1617.680000          35.680000       0.022056
Switzerland  4556.620000          97.820000       0.021468
Austria      3336.622222          71.066667       0.021299
Turkey       2307.766667          46.655556       0.020217
...                  ...                ...            ...
UK          32916.346237          70.593548       0.002145
Argentina   19561.370398          39.616667       0.002025
Canada      31605.186842          62.950000       0.001992
Australia   37140.338372          68.245349       0.001837
USA         51339.078205          77.796154       0.001515

[71 rows x 3 columns]
```

```python
import seaborn as sns
import matplotlib.pyplot as plt
sns.pairplot(Mean_Country[['Total_Cost','Living_Cost_Index','Balance_Score']],kind='scatter',diag_kind='kde')
plt.suptitle('Pairplot of Education Cost,Living Index, and Balance Score',y=1.02)
plt.show()
```

Pairplot of Education Cost, Living Index, and Balance Score
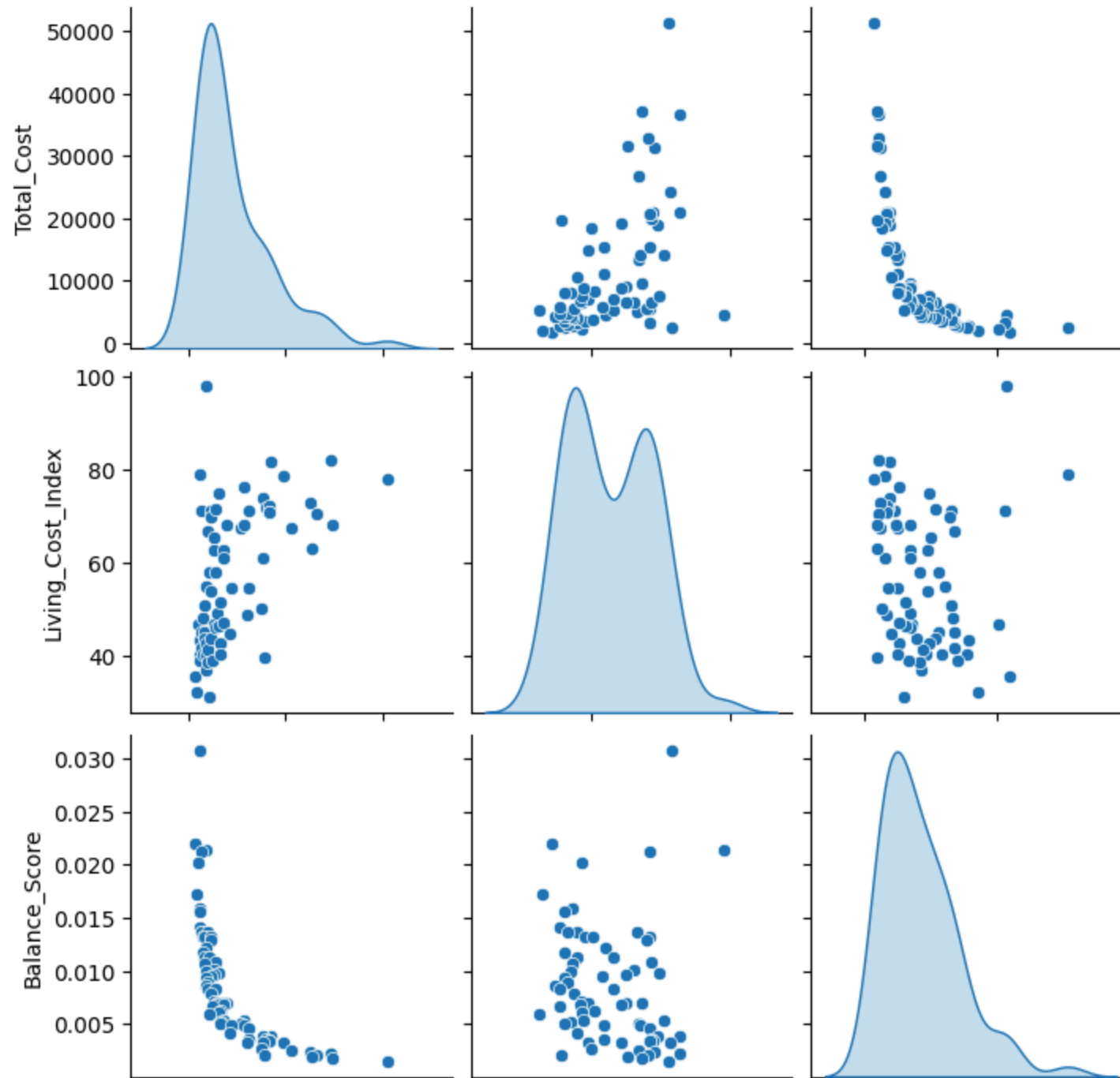
	Total_Cost	Living_Cost_Index	Balance_Score
	0　25000　50000	50　　100	0.00　0.02　0.04

In [125...   `Df.head(5)`

Out[125...

|	Country	City	University	Program	Level	Duration_Years	Tuition_USD	Living_Cost_Index	Rent_USD	Visa_Fee_US
0	USA	Cambridge	Harvard University	Computer Science	Master	2.0	55400.0	83.5	2200	16
1	UK	London	Imperial College London	Data Science	Master	1.0	41200.0	75.8	1800	48
2	Canada	Toronto	University of Toronto	Business Analytics	Master	2.0	38500.0	72.5	1600	23
3	Australia	Melbourne	University of Melbourne	Engineering	Master	2.0	42000.0	71.2	1400	45
4	Germany	Munich	Technical University of Munich	Mechanical Engineering	Master	2.0	500.0	70.5	1100	7

In [71]:
```python
cols_to_check = ['Tuition_USD', 'Living_Cost_Index', 'Rent_USD', 'Visa_Fee_USD', 'Insurance_USD', 'Exchange_Rate']
Df[cols_to_check] = Df[cols_to_check].apply(pd.to_numeric, errors='coerce')
df_clean = Df[cols_to_check].dropna()
sns.pairplot(df_clean)
plt.suptitle("Multivariate Pair Plot", y=1.02)
plt.show()
```

Multivariate Pair Plot
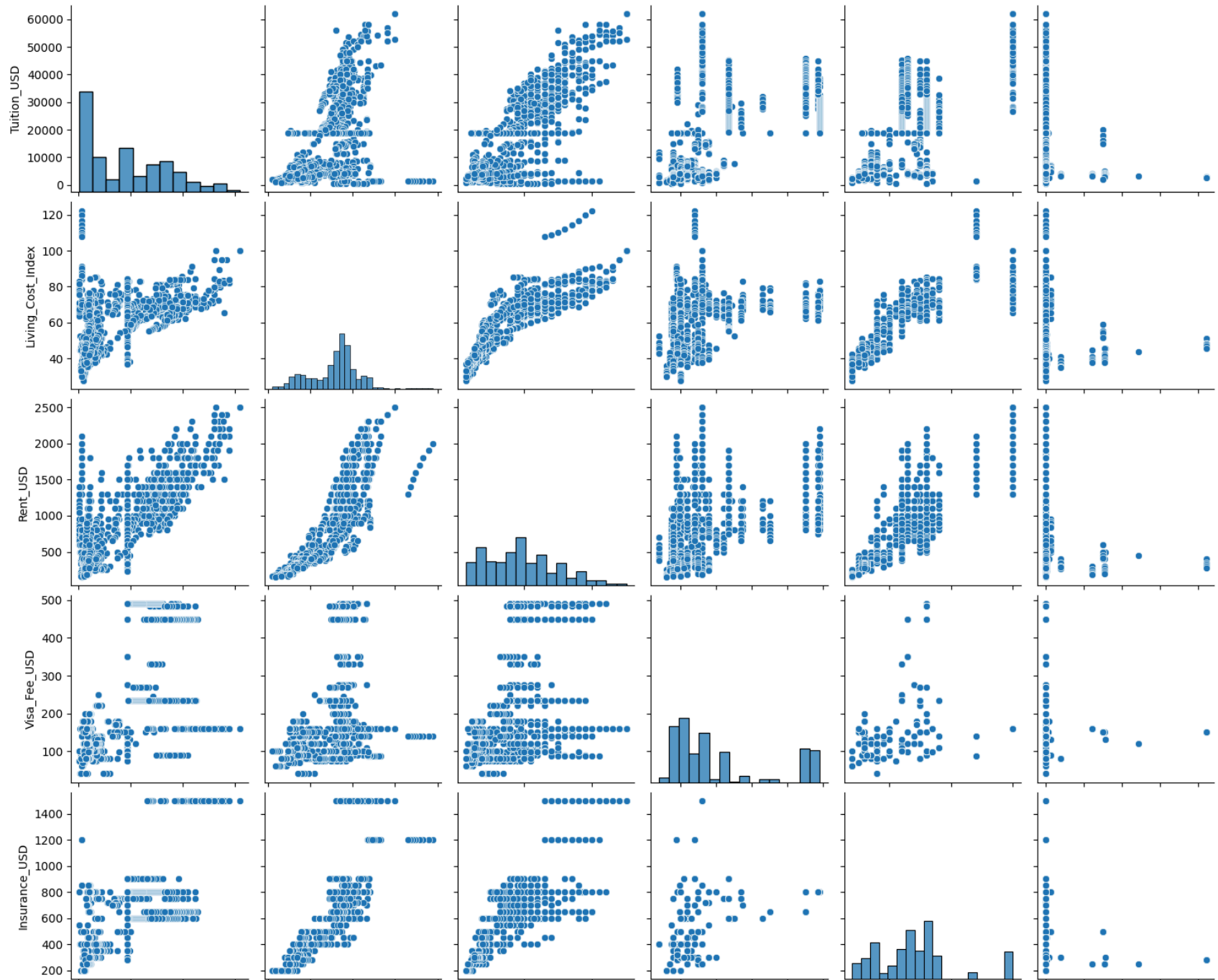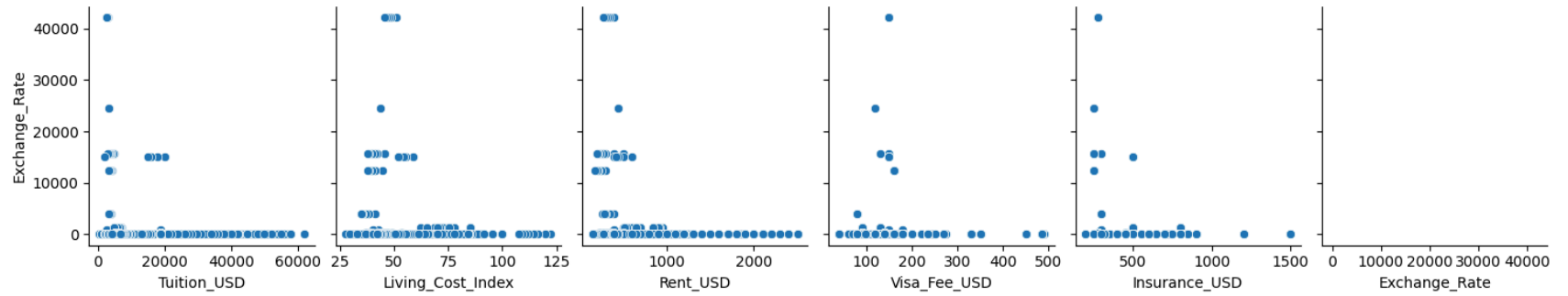
```
In [72]:  plt.figure(figsize=(8,5))
          sns.scatterplot(data=Df, x='Rent_USD', y='Tuition_USD', hue='Country', alpha=0.7, legend=False)
          plt.title('Rent vs Tuition Cost')
          plt.xlabel('Rent (USD)')
          plt.ylabel('Tuition (USD)')
          plt.show()
```

Rent vs Tuition Cost
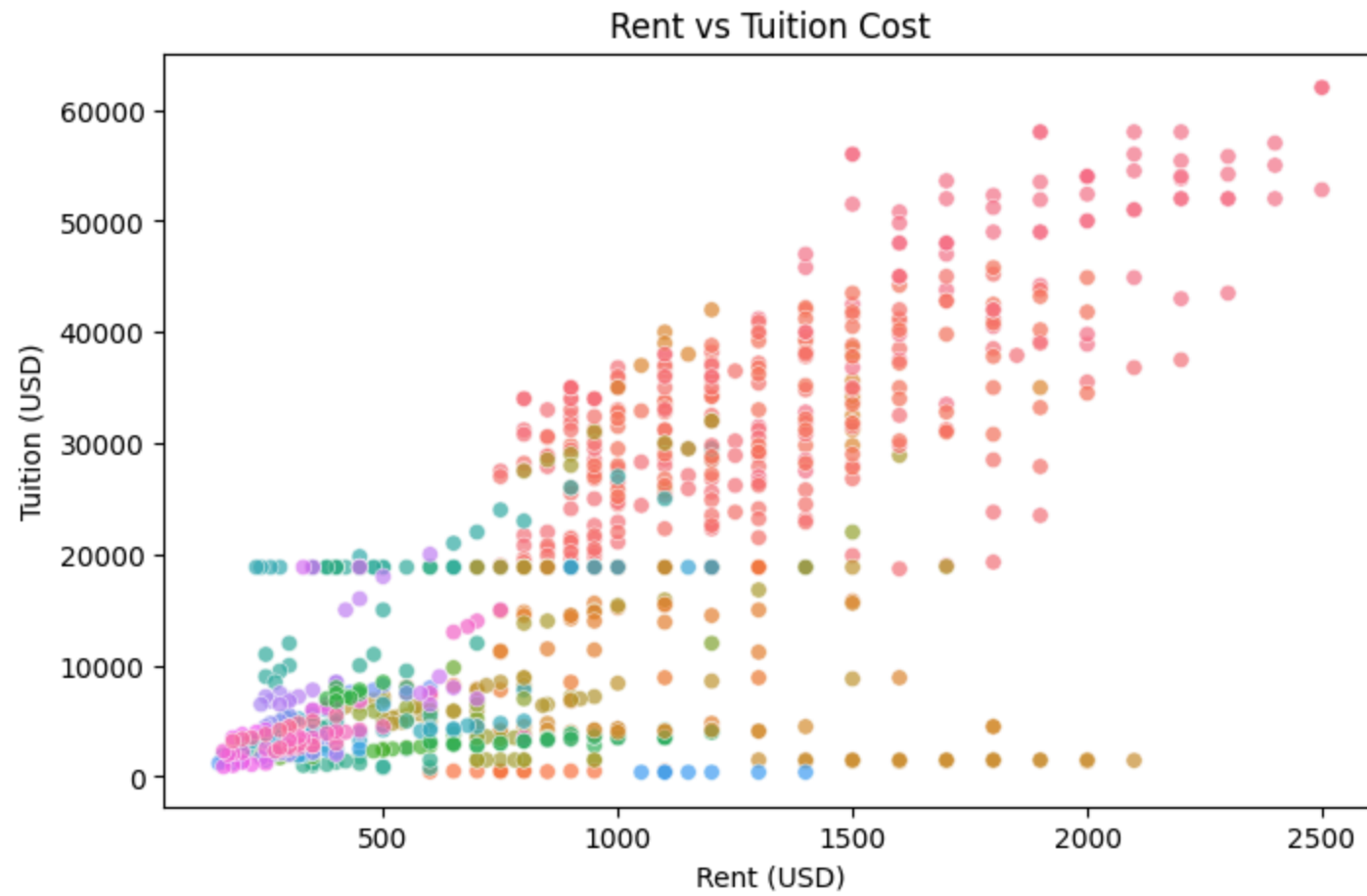

```
In [73]:  top_countries = Df.groupby('Country')['Tuition_USD'].mean().sort_values(ascending=False).head(10)
          plt.figure(figsize=(10,6))
          sns.lineplot(x=top_countries.index, y=top_countries.values, marker='o')
          plt.xticks(rotation=45)
          plt.title('Avg Tuition by Top 10 Countries')
          plt.ylabel('Average Tuition (USD)')
          plt.xlabel('Country')
          plt.grid(True)
          plt.tight_layout()
          plt.show()
```
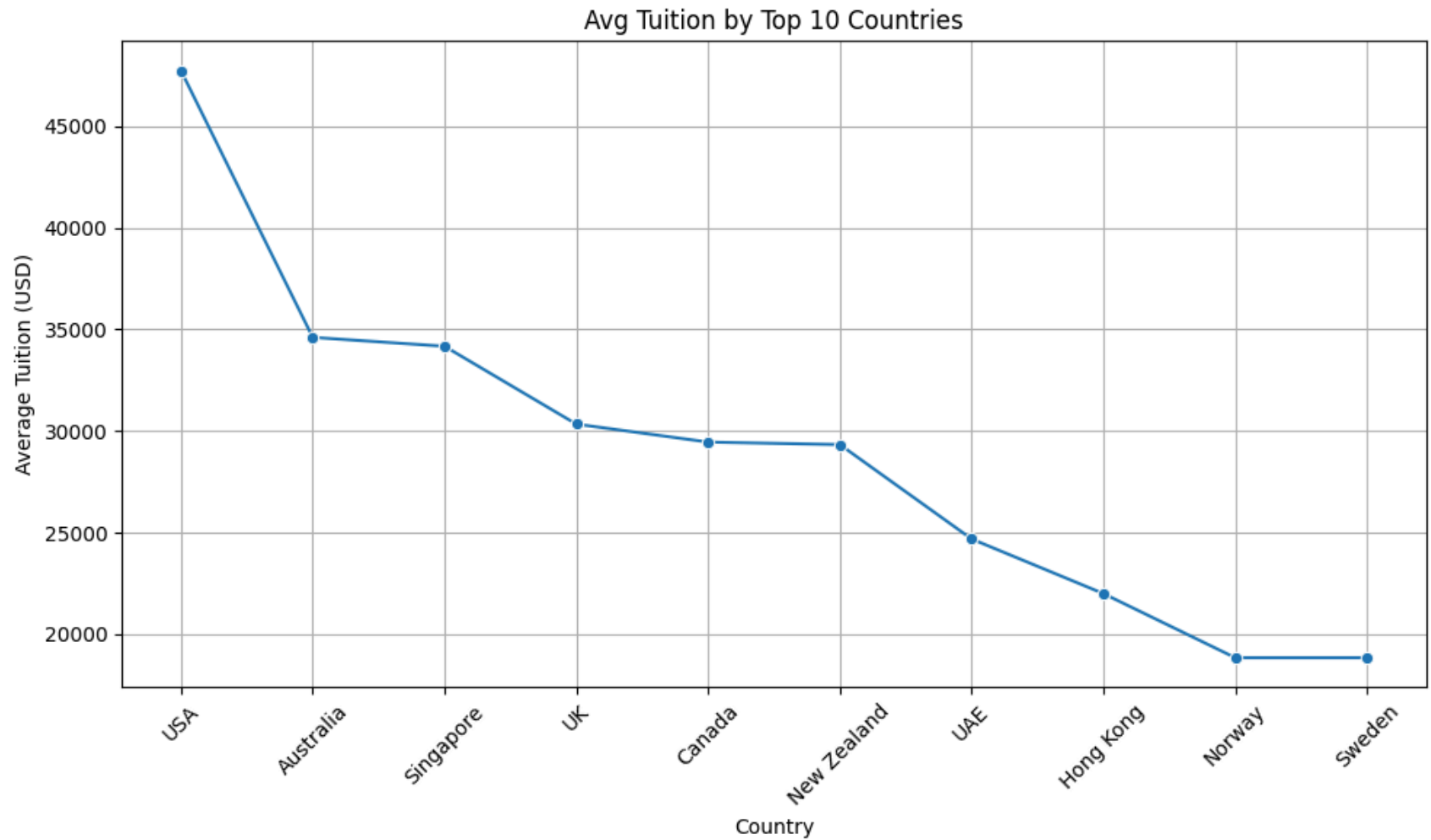
INSIGHT :

1. USA has the highest Tuition Fees

- The USA ranks highest in education cost, with tuition fees going up to $62,000 in top universitites.

2. Top 7 Most Expensive Countrries

- Countries like Australia , Canada, Uk, Singapore, New Zealand, and UAE also rank high in tuition cost, following the USA

3. Germany offers Extremely Low Education Costs

- Some countries like Germany offer tuition as low as $500 , making it a cost-effective destination for international students.

4 .Average Tuition Cost Across All Countries

- The global average tuition fee in your dataset is around $18,845 USD.

5. UK is the most preferred Destination

- The UK has the highest number of entries in the dataset, showing its a top choice for international students.

6. Over 50 Countries Have Tuition Below Global Average

- Many countries like Malaysia, Poland, India, and Turkey offer programs well below the global tuition average.

7. Top 5 Most Popular Programs

- Computer Science, Data Science, Software Engineering, Artificial Intelligence, Information Systems.

8 . Rent Costs are highest in the USA

- Cities like New York(USA) have the highest rent, up to $2,500/month, significantly contributing to total education costs.

9. Visa and Insurance Fees also vary widely

- The UK has the highest visa fees up to dollar 490 and the USA often has high insurance costs $1,500.

10. Tuition and rent Have Strong Correlation

- A Correlation score of 0.74 between Tuition and Rent suggests that expensive universities are alsp located in high-rent areas.