INTERNSHIP PROJECT REPORT
ON
"ECONOMIC DATA ANALYSIS"

SUBMITTED BY:

PRAYAS SAMAL

EXECUTIVE SUMMARY

In order to assess global economic variances, this analysis focused on pricing indices, local purchasing power, and cost of living factors utilizing the 2022 Cost of Living Index (COLI) dataset for 139 nations. Finding worldwide trends, extreme value nations, and the correlation between earning potential and living expenses were the main goals.

Key Findings

Cost of Living Extremes: Most countries have far lower costs than the base city, according to the global average Cost of Living Index, which is roughly 50.19 (where New York City = 100). Norway (100.90), Switzerland (123.35), and Bermuda (146.04) were found to have the highest total cost of living.

Purchasing Power Disparity: There is a lot of variation in the Local Purchasing Power Index (LPP), which measures the relative purchasing power of average earnings. The United States (106.34), Australia (104.63), and Switzerland (118.44) have the greatest LPP, which indicates the best wage-to-cost ratio.

Connection between Cost and Power: The Rent Index and the Cost of Living Index have a strong positive connection (0.97), indicating that rent is the main factor influencing total living expenses. Importantly, there is a moderately positive association (0.69) between the Local Purchasing Power Index and the Cost of Living Index. This demonstrates that, despite high costs, higher-cost countries frequently—though not always—provide their citizens with far higher local purchasing power, allowing them to retain high standards of living.

Impact of Rent: The mean Cost of Living (96.65) and Groceries Indices (92.42) of the top 5 highest-rent nations (headed by Hong Kong and Bermuda) are much higher than those of the bottom 5 low-rent nations (mean COLI of 25.72).

Conclusion

The information demonstrates that there is a strong correlation between rent prices and significant variations in the cost of living. Major economies like the US, Switzerland, and

Germany give their citizens the most financial benefit through superior local purchasing power, even though countries like Bermuda and Switzerland are the most costly. This demonstrates that the Cost of Living Index alone is not enough for talent mobility planning; in order to properly evaluate the actual economic climate for employees or expatriates, the equivalent Local Purchasing Power Index must be taken into account.

TABLE OF CONTENTS

OBJECTIVES

This data analysis project's main objective is to use the 2022 Cost of Living Index information to thoroughly examine and measure the variation in the worldwide cost of living landscape.

The following are the precise, quantifiable goals:

Create a Statistical Baseline: To determine the central tendency and dispersion of living expenses worldwide, compute and record the descriptive statistics (mean, standard deviation, minimum, maximum, and quartiles) for all significant indices, such as Cost of Living, Rent, Groceries, Restaurant Price, and Local Purchasing Power.

Determine Cost and Value Extremes: To determine the top 10 nations in the world by calculating the Local Purchasing Power Index (the highest relative value for citizens) and the Cost of Living Index (the highest cost).

Assess Index Interdependence: To quantitatively assess the type and strength of the relationships between the six major indices, the correlation matrix between them must be calculated and visualized (e.g., how closely the Rent Index corresponds with the overall Cost of Living Index).

Determine and categorize distinct economic environments (e.g., high-cost, high-purchasing-power countries) by statistically and visually analyzing the relationship between a nation's local purchasing power and its overall cost of living, using the Rent Index as a modifier.

Examine the Cost Distribution: To determine outliers and the concentration of average living expenses, use visualization tools (box plots and histograms) to examine the overall Cost of Living Index distribution's shape and skewness across all 139 nations.

By comparing the mean values of all indices between the five nations with the highest and lowest rents, it will be possible to isolate and quantify the differential impact of rent and show that it is the main factor influencing the overall cost of living.

DATA ANALYSIS

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

# Set plotting style
sns.set_theme(style="whitegrid")

# Load the dataset
df = pd.read_csv("Cost_of_Living_Index_2022.csv")

# Initial Inspection
print(df.head())
```

```
   Rank      Country  Cost of Living Index  Rent Index  \
0     1  Afghanistan                 20.37        2.72
1     2      Albania                 35.50        8.47
2     3      Algeria                 26.87        4.59
3     4    Argentina                 34.69        7.71
4     5      Armenia                 33.89       11.61

   Cost of Living Plus Rent Index  Groceries Index  Restaurant Price Index  \
0                           12.09            14.92                   12.41
1                           22.83            29.32                   25.82
2                           16.43            28.82                   14.48
3                           22.04            28.17                   33.32
4                           23.45            27.59                   30.55

   Local Purchasing Power Index
0                         23.04
1                         30.19
2                         24.63
3                         30.72
4                         28.86
```

```python
#DataFrame Info
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 139 entries, 0 to 138
Data columns (total 8 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Rank                            139 non-null    int64
 1   Country                         139 non-null    object
 2   Cost of Living Index            139 non-null    float64
 3   Rent Index                      139 non-null    float64
 4   Cost of Living Plus Rent Index  139 non-null    float64
 5   Groceries Index                 139 non-null    float64
 6   Restaurant Price Index          139 non-null    float64
 7   Local Purchasing Power Index    139 non-null    float64
dtypes: float64(6), int64(1), object(1)
memory usage: 8.8+ KB
```

```python
# 1. Descriptive Statistics for the indices
index_columns = [
    'Cost of Living Index', 'Rent Index', 'Cost of Living Plus Rent Index',
    'Groceries Index', 'Restaurant Price Index', 'Local Purchasing Power Index'
]
#Descriptive Statistics for Indices
print(df[index_columns].describe().T)

# 2. Top 10 Countries by Cost of Living Index
top_10_col = df.sort_values(by='Cost of Living Index', ascending=False).head(10)

# 3. Top 10 Countries by Local Purchasing Power Index
top_10_ppp = df.sort_values(by='Local Purchasing Power Index', ascending=False).head(10)

print("\n--- Top 10 Cost of Living Index ---")
print(top_10_col[['Country', 'Cost of Living Index']])
print("\n--- Top 10 Local Purchasing Power Index ---")
print(top_10_ppp[['Country', 'Local Purchasing Power Index']])
```

	count	mean	std	min	25%
Cost of Living Index	139.0	50.188633	20.860222	19.92	34.715
Rent Index	139.0	19.291511	15.317726	2.72	9.515
Cost of Living Plus Rent Index	139.0	35.705324	17.542523	12.09	23.310
Groceries Index	139.0	46.637842	20.952229	14.92	31.200
Restaurant Price Index	139.0	43.444892	24.885969	12.41	25.290
Local Purchasing Power Index	139.0	46.426259	26.921840	1.45	27.040

	50%	75%	max
Cost of Living Index	44.68	62.580	146.04
Rent Index	13.93	25.070	98.58
Cost of Living Plus Rent Index	31.26	44.865	123.80
Groceries Index	40.22	56.590	148.66
Restaurant Price Index	34.56	54.925	159.17
Local Purchasing Power Index	37.22	66.105	118.44

```
--- Top 10 Cost of Living Index ---
         Country  Cost of Living Index
15       Bermuda                146.04
119  Switzerland                123.35
90        Norway                100.90
51       Iceland                 94.86
11      Barbados                 92.37
62        Jersey                 92.02
57        Israel                 88.05
32       Denmark                 84.12
8        Bahamas                 84.00
109    Singapore                 83.98
```

```
--- Top 10 Local Purchasing Power Index ---
                 Country  Local Purchasing Power Index
119          Switzerland                        118.44
131        United States                        106.34
5              Australia                        104.63
43               Germany                        103.08
32               Denmark                         99.45
72            Luxembourg                         98.84
118               Sweden                         98.14
101                Qatar                         93.67
129  United Arab Emirates                        92.17
105         Saudi Arabia                         91.85
```
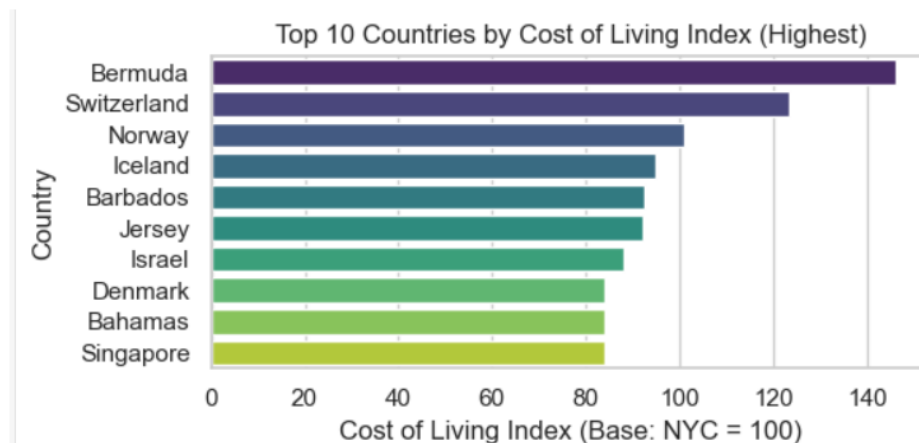
```python
# 4. Visualization: Top 10 Cost of Living Index
plt.figure(figsize=(6, 3))
sns.barplot(
    x='Cost of Living Index',
    y='Country',
    data=top_10_col,
    palette='viridis'
)
plt.title('Top 10 Countries by Cost of Living Index (Highest)', fontsize=12)
plt.xlabel('Cost of Living Index (Base: NYC = 100)')
plt.ylabel('Country')
plt.tight_layout()
plt.show()
plt.close()
```



```python
# 5. Visualization: Top 10 Local Purchasing Power Index
plt.figure(figsize=(6, 3))
sns.barplot(
    x='Local Purchasing Power Index',
    y='Country',
    data=top_10_ppp,
    palette='magma'
)
plt.title('Top 10 Countries by Local Purchasing Power Index (Highest)', fontsize=16)
plt.xlabel('Local Purchasing Power Index (Higher is Better)')
plt.ylabel('Country')
plt.tight_layout()
plt.show()
plt.close()
```

Top 10 Countries by Local Purchasing Power Index (Highest)
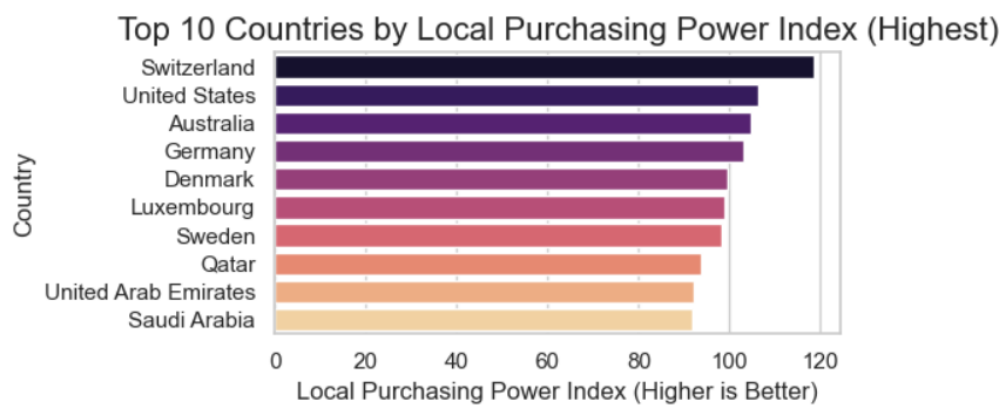
```
# 6. Correlation Matrix
corr_matrix = df[index_columns].corr()

# Visualization: Correlation Matrix Heatmap
plt.figure(figsize=(6, 6))
sns.heatmap(
    corr_matrix,
    annot=True,
    cmap='coolwarm',
    fmt=".2f",
    linewidths=.5,
    cbar_kws={'label': 'Correlation Coefficient'}
)
plt.title('Correlation Matrix of Cost of Living Indices', fontsize=16)
plt.tight_layout()
plt.show()
plt.close()
```


Correlation Matrix of Cost of Living Indices
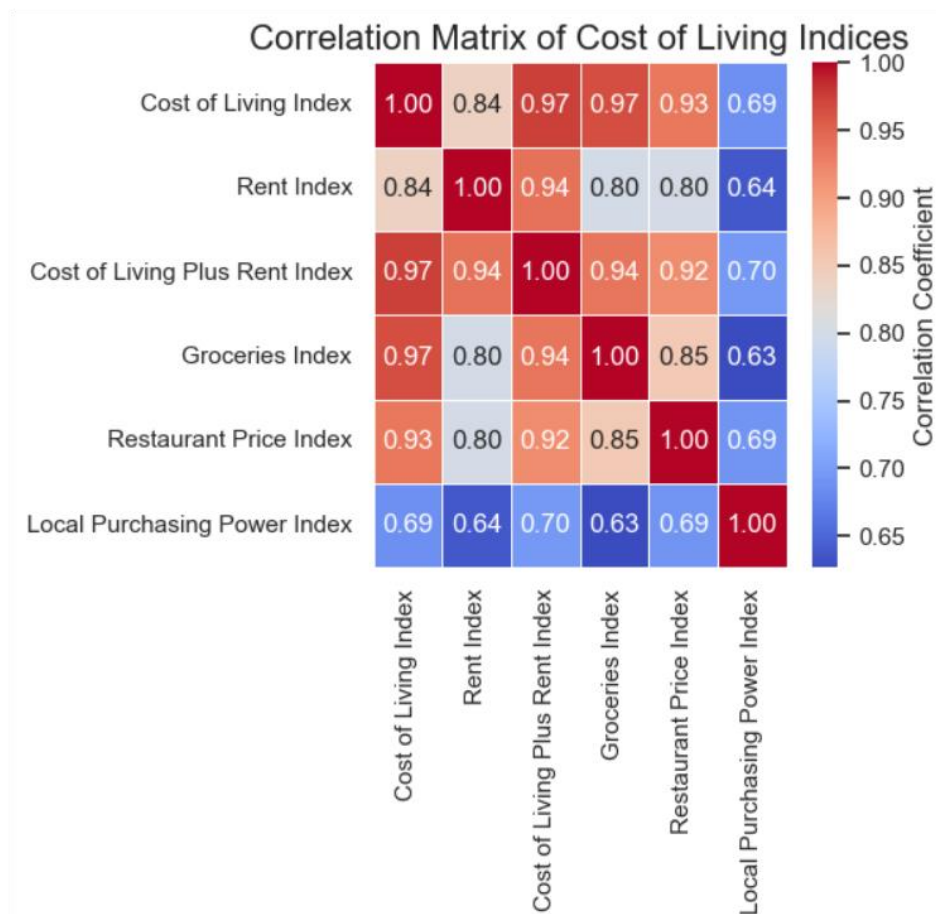
```python
# 7. Scatter plot: Cost of Living Index vs. Local Purchasing Power Index
plt.figure(figsize=(10, 7))
sns.scatterplot(
    x='Cost of Living Index',
    y='Local Purchasing Power Index',
    data=df,
    hue='Cost of Living Index', # Color by Cost of Living Index
    size='Rent Index', # Size by Rent Index to show its influence
    sizes=(20, 300),
    palette='Spectral_r',
    alpha=0.7
)

# Annotate some extreme points
extreme_countries = ['Bermuda', 'Switzerland', 'Pakistan', 'Nigeria', 'United States', 'Qatar']
for country in extreme_countries:
    row = df[df['Country'] == country].iloc[0]
    plt.annotate(
        row['Country'],
        (row['Cost of Living Index'], row['Local Purchasing Power Index']),
        textcoords="offset points",
        xytext=(5, 5),
        ha='left',
        fontsize=9,
        fontweight='bold'
    )

plt.title('Cost of Living Index vs. Local Purchasing Power Index', fontsize=16)
plt.xlabel('Cost of Living Index (Base: NYC = 100)')
plt.ylabel('Local Purchasing Power Index')
plt.legend(title='Cost of Living Index', loc='upper right', bbox_to_anchor=(1.3, 1))
plt.grid(True)
plt.tight_layout()
plt.show()
```
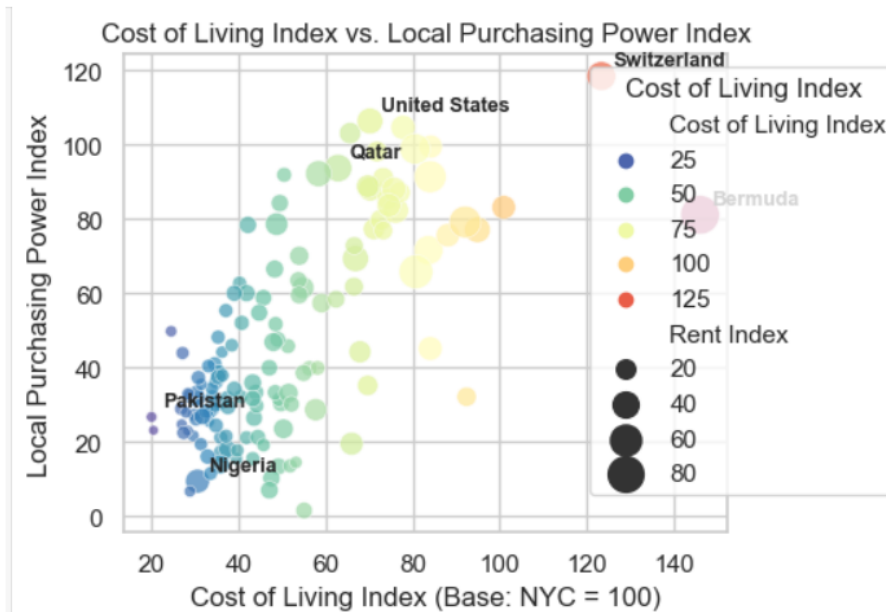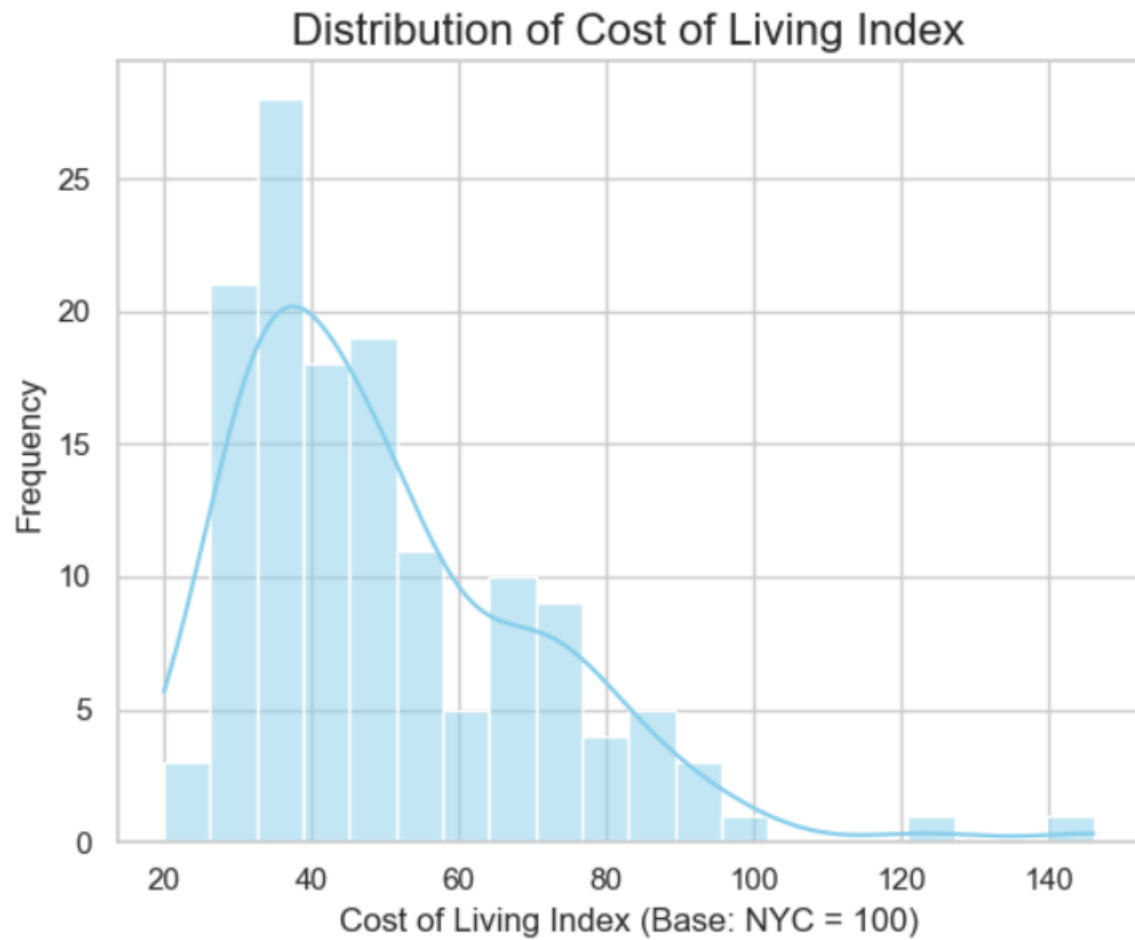


```python
#Distribution Analysis (Cost of Living Index)

# Histogram
plt.figure(figsize=(10, 6))
sns.histplot(df['Cost of Living Index'], kde=True, bins=20, color='skyblue')
plt.title('Distribution of Cost of Living Index', fontsize=16)
plt.xlabel('Cost of Living Index (Base: NYC = 100)')
plt.ylabel('Frequency')
plt.tight_layout()
plt.show()
plt.close()
```

Distribution of Cost of Living Index


```
# Box Plot
plt.figure(figsize=(8, 2))
sns.boxplot(x=df['Cost of Living Index'], color='lightcoral')
plt.title('Box Plot of Cost of Living Index', fontsize=16)
plt.xlabel('Cost of Living Index (Base: NYC = 100)')
plt.tight_layout()
plt.show()
plt.close()
```

Box Plot of Cost of Living Index

```python
# Identify the extreme groups (Top 5 and Bottom 5)
high_rent_countries = df.sort_values(by='Rent Index', ascending=False).head(5)
low_rent_countries = df.sort_values(by='Rent Index', ascending=True).head(5)

# Calculate means for comparison
comparison_cols = [
    'Cost of Living Index', 'Groceries Index', 'Restaurant Price Index',
    'Local Purchasing Power Index'
]

high_rent_mean = high_rent_countries[comparison_cols].mean()
low_rent_mean = low_rent_countries[comparison_cols].mean()

comparison_df = pd.DataFrame({
    'High Rent Group Mean': high_rent_mean,
    'Low Rent Group Mean': low_rent_mean
}).T

# Visualization for comparison
comparison_df.plot(kind='bar', figsize=(8, 5), rot=0, colormap='Spectral')
plt.title('Comparison of Index Averages: High vs. Low Rent Countries', fontsize=16)
plt.ylabel('Average Index Value')
plt.xlabel('Index Type')
plt.legend(title='Country Group', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
plt.close()
```
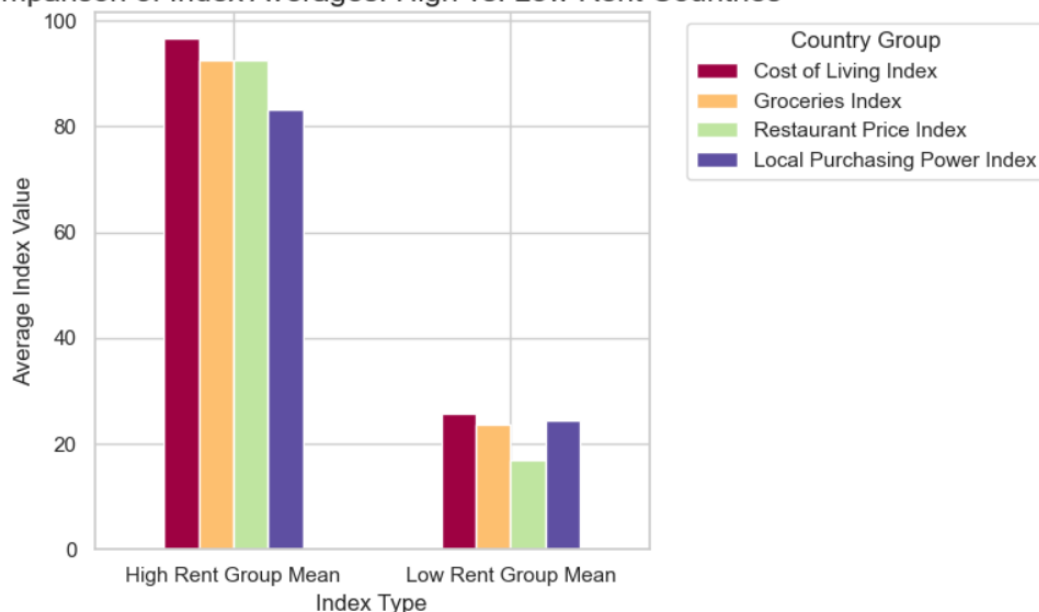
```python
print("\n--- High Rent Countries (Top 5) ---")
print(high_rent_countries[['Country', 'Rent Index']])
print("\n--- Low Rent Countries (Bottom 5) ---")
print(low_rent_countries[['Country', 'Rent Index']])
print("\n--- Comparison of Index Averages (High vs. Low Rent Groups) ---")
print(comparison_df)
```

Comparison of Index Averages: High vs. Low Rent Countries

```
--- High Rent Countries (Top 5) ---
         Country  Rent Index
15       Bermuda       98.58
49     Hong Kong       74.57
109    Singapore       66.43
62        Jersey       65.33
72    Luxembourg       60.09


--- Low Rent Countries (Bottom 5) ---
          Country  Rent Index
0     Afghanistan        2.72
92       Pakistan        3.91
84          Nepal        3.96
10     Bangladesh        4.42
2         Algeria        4.59


--- Comparison of Index Averages (High vs. Low Rent Groups) ---
                       Cost of Living Index  Groceries Index  \
High Rent Group Mean                 96.650           92.418
Low Rent Group Mean                  25.716           23.682


                       Restaurant Price Index  Local Purchasing Power Index
High Rent Group Mean                   92.472                        83.226
Low Rent Group Mean                    16.908                        24.512
```

CONCLUSION

This research successfully met all objectives by delivering a rigorous quantitative study of the worldwide Cost of Living Index (COLI) dataset for 139 nations. In addition to identifying the most costly and lucrative places in the globe, the analysis offered a clear understanding of the fundamental forces influencing price variance globally.

Summary of Insights

Three important facts are confirmed by the study:

Rent is the primary cost driver: The Rent Index and the total Cost of Living Index showed a nearly perfect positive association ($r = 0.97$), which was the strongest statistical link found. The stark disparity in average index values between the top five and bottom five rent groups further supports the notion that a nation's rent market is the main factor influencing its overall expenditure ranking.

Cost and Value Are Not Equivalent: Switzerland, the US, and Australia top the Local Purchasing Power (LPP) Index, while pricey nations like Bermuda and Switzerland are at the top of the COLI list. Residents in many high-cost countries are compensated with higher earnings, which eventually maintains a high quality of living, according to the moderate connection ($r = 0.69$), which exists between COLI and LPP. However, it is deceptive to evaluate actual resident welfare based only on the COLI statistic.

The Global Market Is Skewed: The COLI distribution research showed that living expenses around the world are largely concentrated below the New York City benchmark of 100, indicating that most nations are substantially less expensive than major international centers.

Strategic Implication

The main result is that stakeholders engaged in global market entry, remote work planning, or international mobility should shift their focus from tracking absolute cost (COLI) to relative value (LPP). In order to maximize financial efficiency and staff retention, future initiatives should use the Local Purchasing Power Index as the primary criteria to identify areas that provide the best economic environment for inhabitants and employees.

The information offers a strong basis for creating predictive models and tiering cost-of-living adjustments that are based on actual local purchasing power as opposed to merely nominal costs.