

CS422 Data Mining Homework_1 Report

Individual Contribution			
CWID	Name	Contribution (description)	Percent Contribution
A20564206	Hanjun Tang	Pie Chart, Spearman Correlation Heatmap , Stacked Bar Chart	33.3%
A20563437	Xinhong Hu	Bar Chart , Correlation Heatmap, Pairplot	33.3%
A20563438	Zhengyang Zhu	Histogram, Box Plot and Choropleth Map, Parallel Coordinate Plot	33.3%

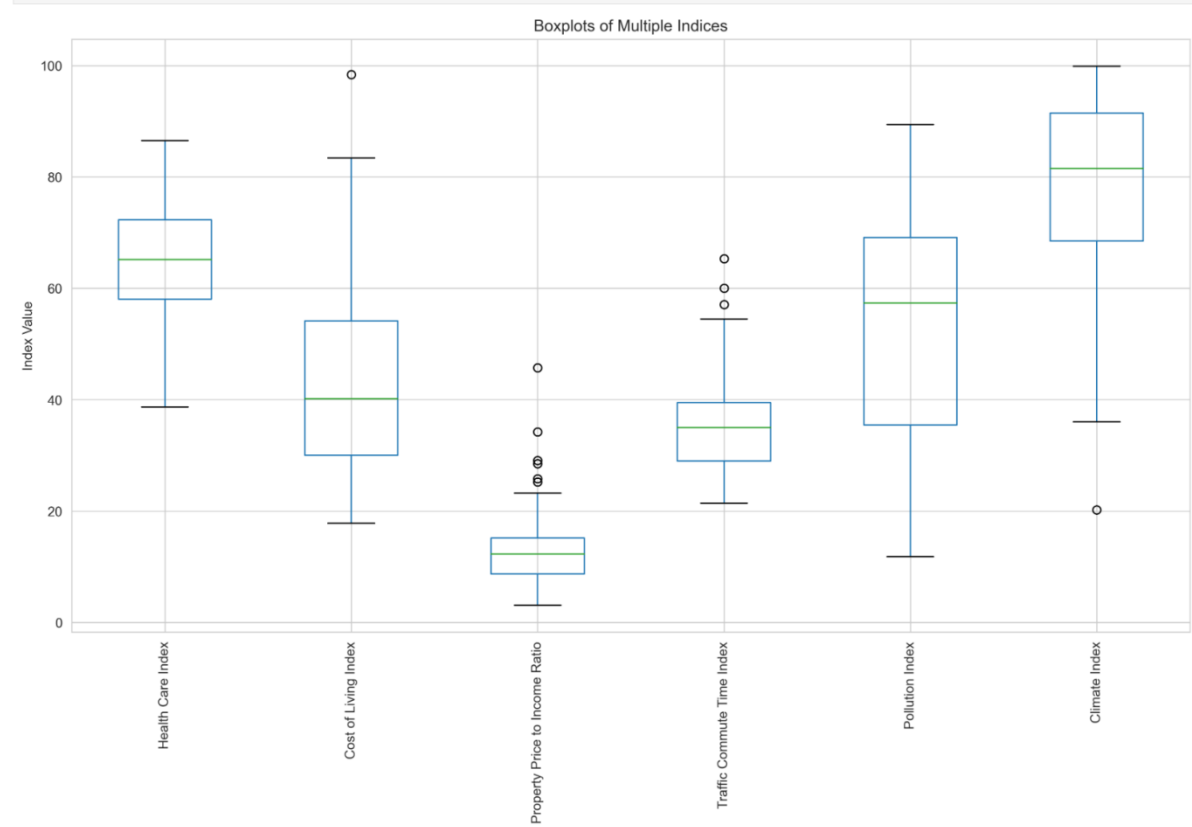
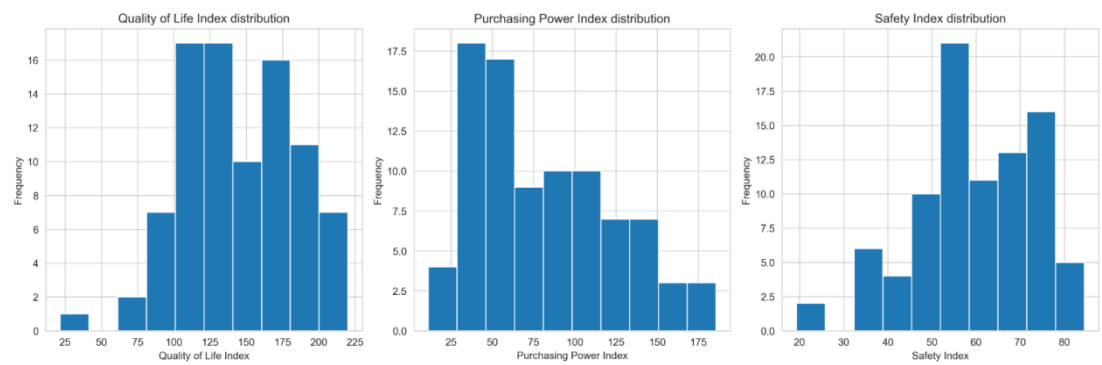
Brief description of the dataset:

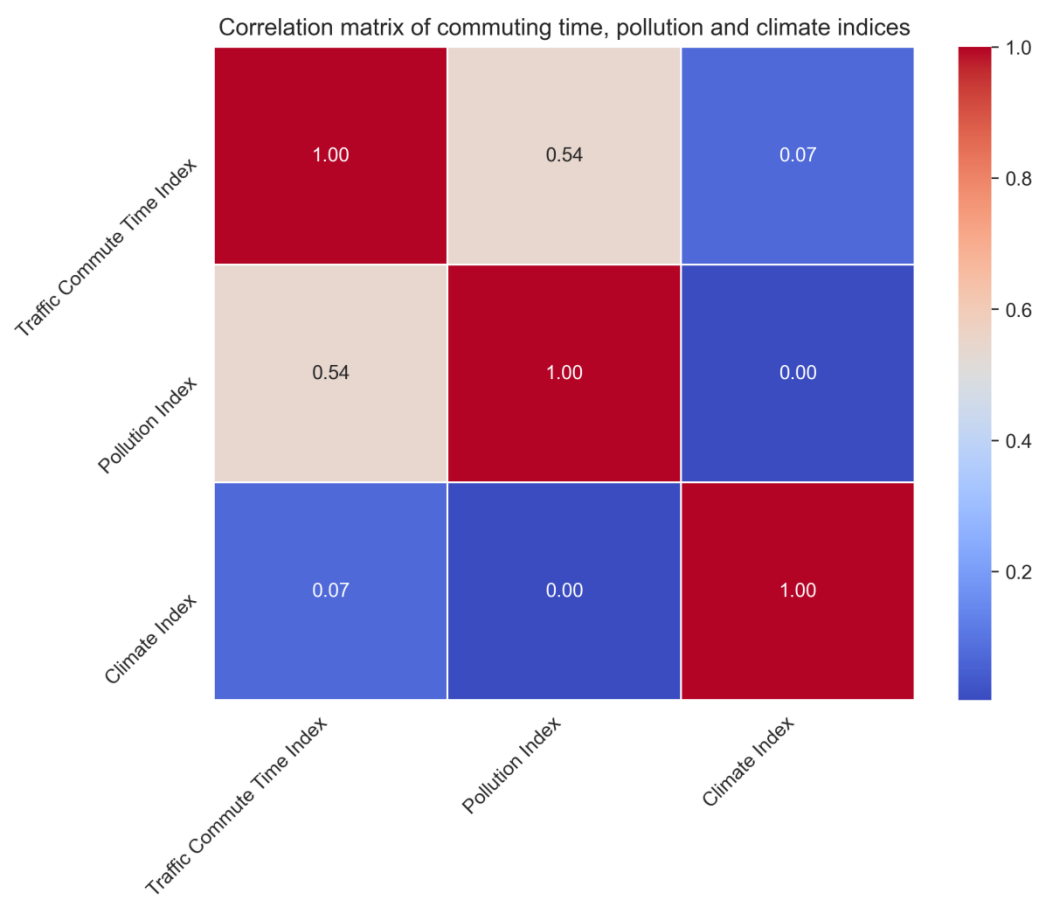
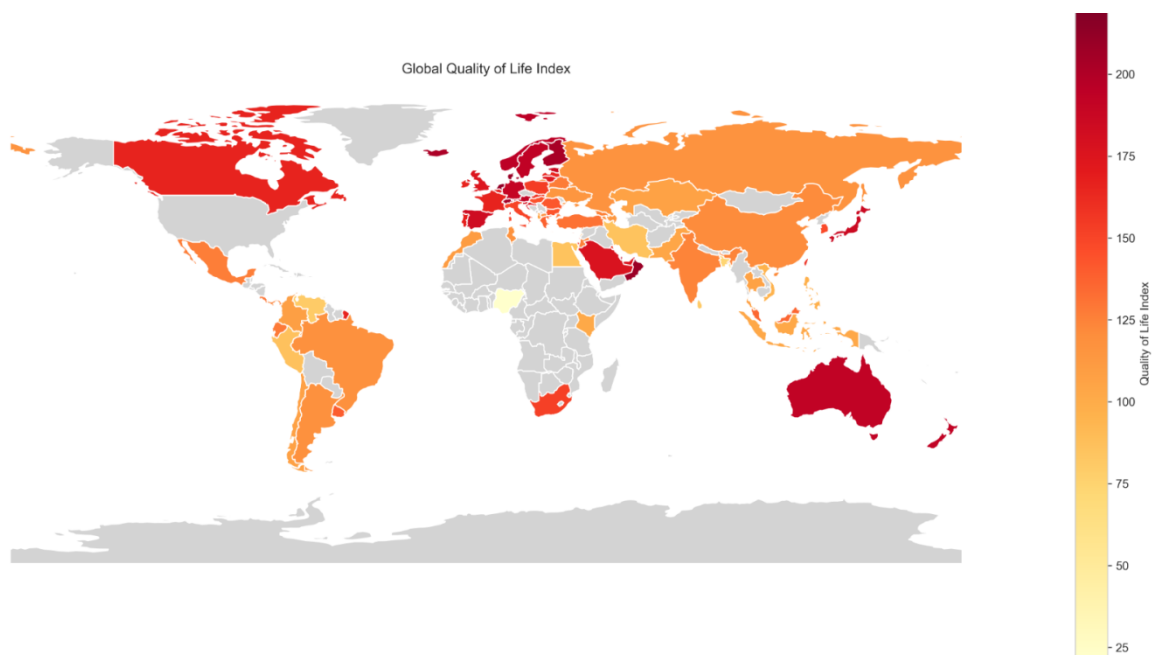
The dataset contains quality of life indices for various countries, including several key factors that contribute to overall life quality. Here's a breakdown of the columns:

- 1.Rank: The country's rank based on the overall quality of life index.
- 2.Country: The name of the country.
- 3.Quality of Life Index: A comprehensive measure that combines various factors influencing the overall quality of life in a country.
- 4.Purchasing Power Index: Represents the relative purchasing power in the country, showing how far a unit of currency goes in terms of buying goods and services.
- 5.Safety Index: Measures the perceived safety in a country, based on factors like crime rates.
- 6.Health Care Index: Indicates the quality of healthcare in the country, including accessibility and affordability.
- 7.Cost of Living Index: Shows the relative cost of living in a country, covering expenses like food, housing, and utilities.
- 8.Property Price to Income Ratio: A ratio indicating how affordable housing is in relation to income.
- 9.Traffic Commute Time Index: Measures the average commuting time in a country, highlighting traffic conditions and urban planning.
- 10.Pollution Index: Reflects the level of pollution in a country, including air and water quality.
- 11.Climate Index: Measures the general climate quality, taking into account temperature, humidity, and comfort levels.

This dataset provides a comprehensive overview of factors contributing to the quality of life in different countries, allowing comparisons across regions based on key indices.

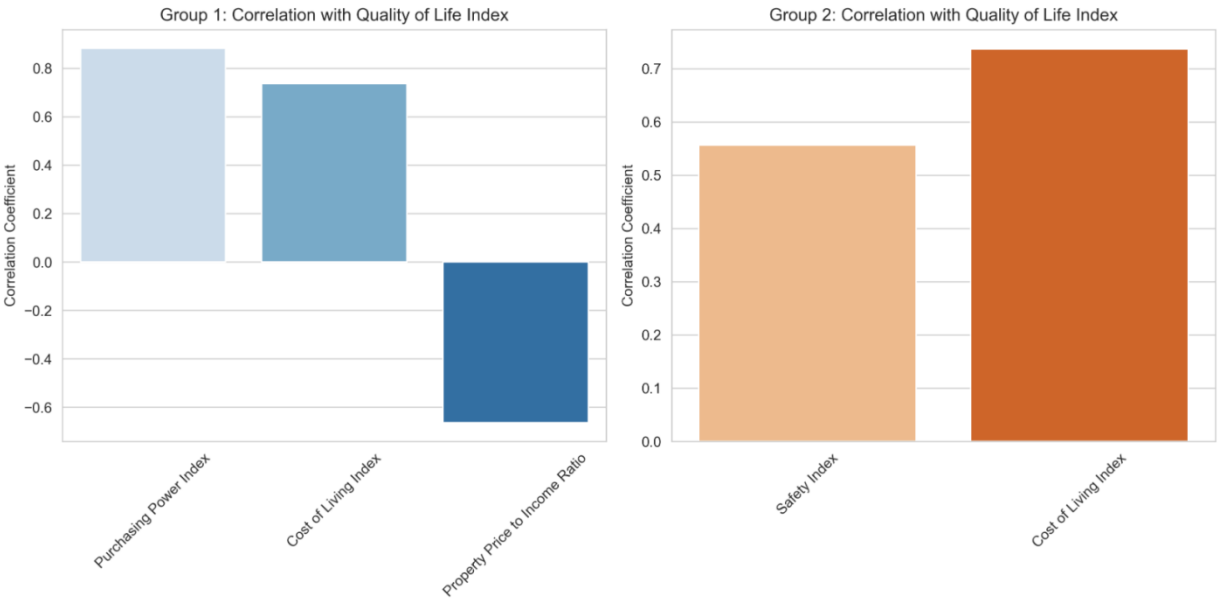
Visualization of results:





Group 1 (Purchasing Power Index, Cost of Living Index, Property Price to Income Ratio) Correlations:
Purchasing Power Index 0.882789
Cost of Living Index 0.736565
Property Price to Income Ratio -0.663569
dtype: float64

Group 2 (Safety Index, Cost of Living Index) Correlations:
Safety Index 0.556305
Cost of Living Index 0.736565
dtype: float64



Quality of Life Index Distribution

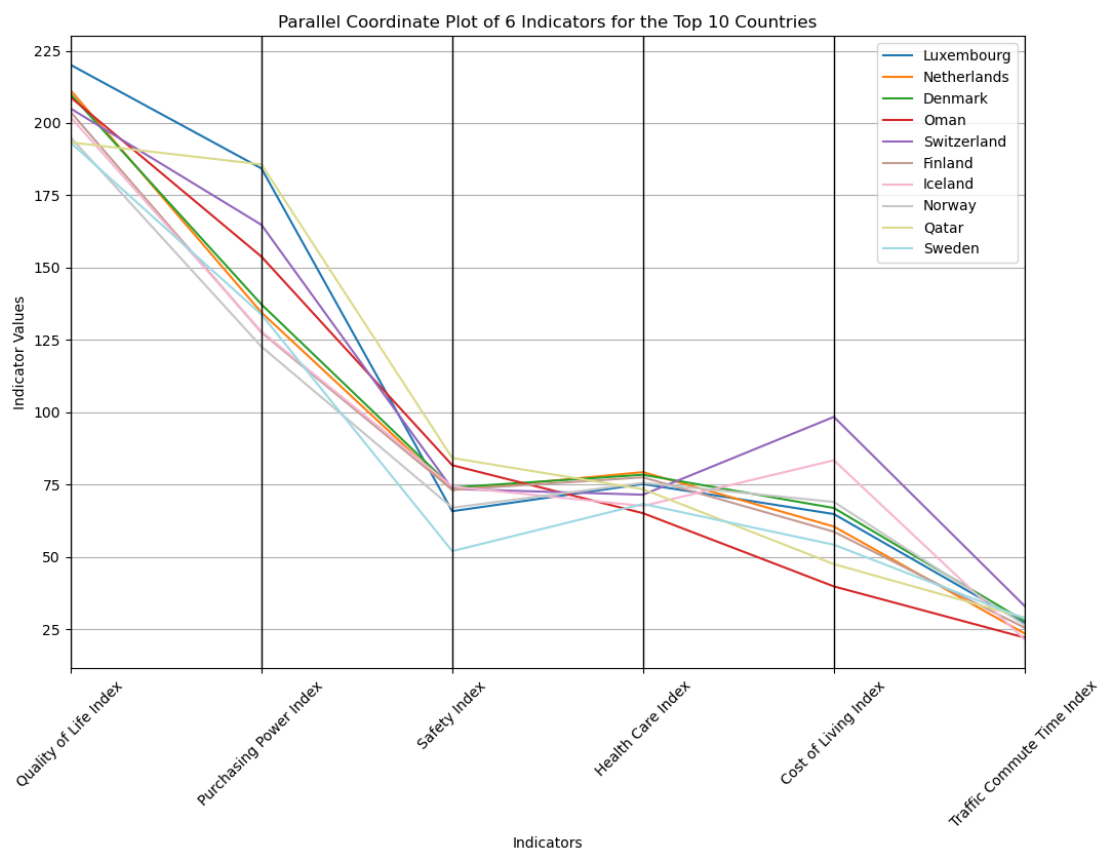
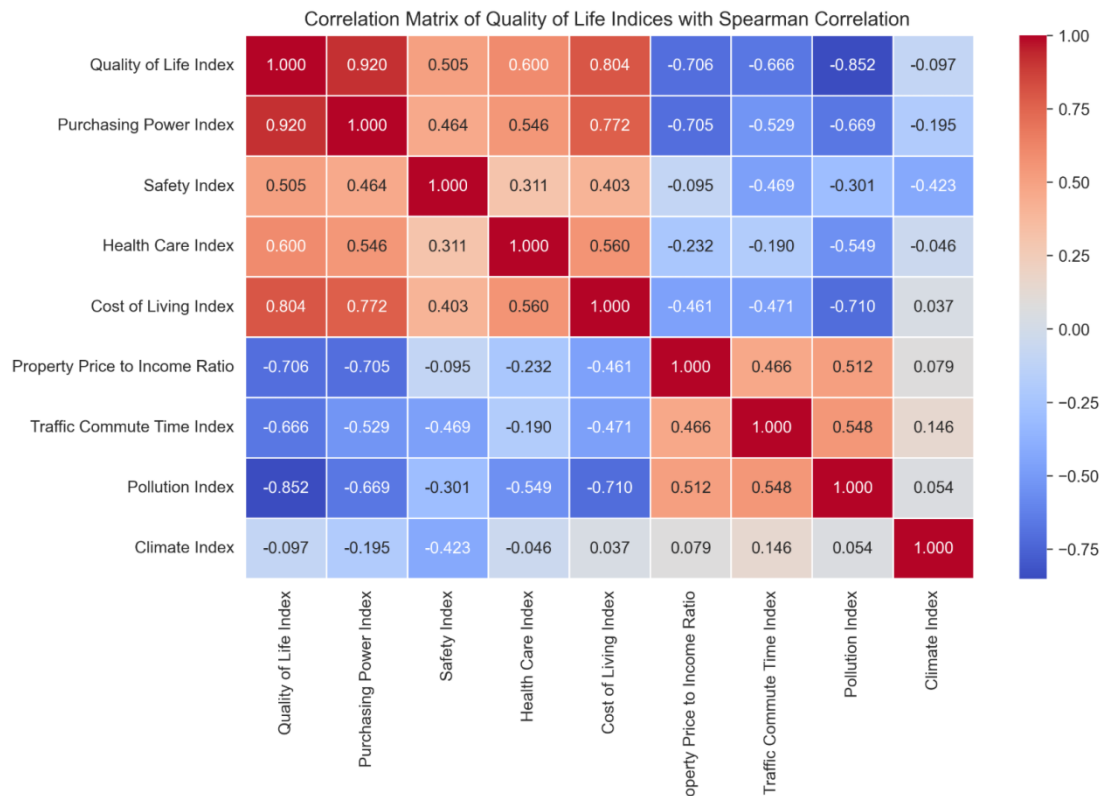


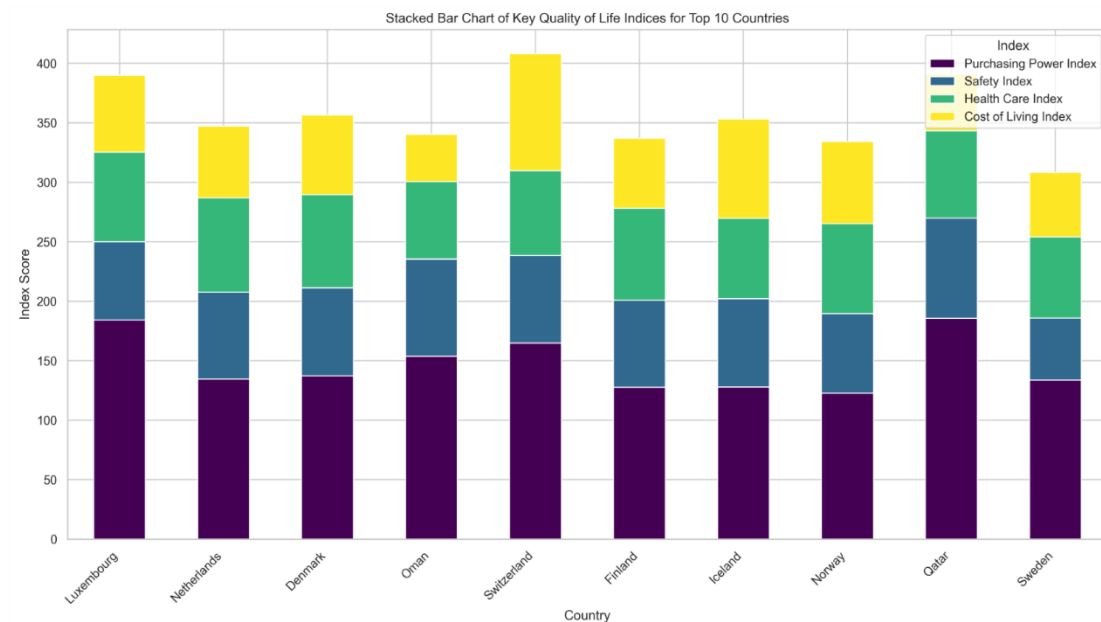
Purchasing Power Index Distribution



Pollution Index Distribution







Visualization Methods Used:

1. Histogram: Displays the distribution of indices such as Quality of Life Index, Purchasing Power Index, and Safety Index, helping to understand their frequency distribution.

2. Box Plot: Illustrates the distribution, median, and outliers of multiple indices, including Health Care Index, Cost of Living Index, Property Price to Income Ratio, Traffic Commute Time Index, Pollution Index, and Climate Index.

3. Choropleth Map: Provides a geospatial visualization of the Quality of Life Index across different countries, using color intensity to indicate index values.

4. Pairplot (Scatterplot Matrix): Shows pairwise relationships among Traffic Commute Time Index, Pollution Index, and Climate Index, allowing an assessment of potential correlations.

5. Correlation Heatmap: Visualizes the correlation between selected indices using a color-coded heatmap, helping to identify strong or weak relationships.

6. Bar Chart: Compares the correlation of different groups of indices (e.g., Purchasing Power Index, Cost of Living Index, Property Price to Income Ratio vs. Safety Index, Cost of Living Index) with the Quality of Life Index.

7. Pie Chart: Represents the proportion of countries categorized by different levels of Quality of Life Index, Purchasing Power Index, and Pollution Index.

8. Spearman Correlation Heatmap: Some of the data in the dataset are discrete, which will have a greater impact on the Pearson correlation coefficient, and the Spearman correlation coefficient can be used to reflect the correlation more accurately.

9. Parallel Coordinate Plot: use multiple parallel axes to represent variables, connect data points with lines, may use color - coding, and scale/position axes to show relationships between variables for different entities.

10. Stacked Bar Chart: use matplotlib and pandas and seaborn .stacked bar charts can show the composition of each country's scores on different quality of life indices, making it easy to compare the scores of different countries on each index.

Libraries Used:

Pandas,
matplotlib,
seaborn,
geopandas,
numpy,plotly.express
parallel_coordinates

Summary of results:

1.Histogram:

(1)Quality of Life Index: The quality - of - life indices of most countries are concentrated in the range of 75 - 200. Among them, the number of countries in the 100 - 125 range is the largest, approaching 17; followed by the 125 - 150 range, with approximately 16 countries. The number of countries at both ends (25 - 50 and 175 - 225) is relatively small, about 7 each.

(2)Purchasing Power Index: The distribution is relatively scattered, and the data mainly concentrate in the range of 25 - 150. There are relatively more countries in the 37.5 - 75 and 75 - 100 ranges, about 9 and 10 respectively. The number of countries with an index less than 25 and greater than 150 is small, only about 3 each.

(3)Safety Index: The safety indices of most countries are concentrated in the range of 30 - 70. The number of countries in the 40 - 50 range is the largest, about 20; followed by the 50 - 60 range, with approximately 16 countries. The number of countries with a safety index less than 30 and greater than 70 is small, about 2 and 5 respectively.

2.Box Plot:

In this multi-index box plot, the median of the Climate Index is the highest, indicating that the climate conditions in most countries are relatively good and the differences are small. The median of the Property Price to Income Ratio is the lowest, with a large degree of dispersion and many outliers, showing significant differences among different countries, which are influenced by factors such as economic models and policies. The overall values of the Health Care Index and the Pollution Index are high. The long whisker of the Health Care Index reflects large differences in the levels of some countries, while the Pollution Index indicates that pollution is widespread globally. The Cost of Living Index is at a medium level, with fluctuations in different countries affected by various factors. The wide box of the Traffic Commute Time Index shows obvious differences in the commuting times of residents in different countries.

3.Choropleth Map:

This "Global Quality of Life Index" heatmap intuitively displays the quality - of - life indices of different countries through color intensity. In the dark red areas, such as some European countries, Australia, and Canada, the quality - of - life index exceeds 175, indicating that these countries excel in economic development, social welfare, infrastructure construction, etc., and can provide residents with a high - quality life. The areas ranging from orange to light orange, including some countries in Asia and South America, have an index between 75 and 175, suggesting that the quality of life in these countries is at a medium level. They have made certain progress in various aspects, but there is still room for improvement compared with countries with high indices. The light yellow areas are relatively few, and the gray areas represent missing data, indicating that the quality - of - life indices of most of the surveyed countries are above a certain level. For countries with missing data, it is difficult to measure their quality of life due to statistical difficulties or other factors.

4 and 5 (Pairplot, .Correlation Heatmap):

The correlation between Traffic Commute Time Index and Pollution Index is 0.54, showing a moderate positive correlation.

The correlation between Traffic Commute Time Index and Climate Index is 0.07, indicating a very weak correlation.

The correlation between Pollution Index and Climate Index is 0.00, showing no correlation.

6.Bar Chart:

The left bar chart indicates a strong positive correlation between both the Purchasing Power Index and Cost of Living Index with the Quality of Life Index, while the Property Price to Income Ratio shows a strong negative correlation.

On the right, the Safety Index is moderately positively correlated with the Quality of Life Index, and the Cost of Living Index again demonstrates a strong positive correlation.

In conclusion, a higher purchasing power and lower relative property prices are associated with a better quality of life, whereas higher costs of living and safety concerns have a positive, though less strong, relationship with quality of life.

7.Pie Chart:

Quality of Life Index Distribution

High (≥ 180): 20.5% of countries fall into this category, indicating excellent quality of life.

Upper-Middle (150-179): 25% of countries are in this bracket, signifying a high but not the highest quality of life.

Middle (120-149): Another 25% of countries are categorized here, representing a moderate level of quality of life.

Lower-Middle (90-119): 21.6% of countries are in this range, indicating a slightly below moderate quality of life.

Low (< 90): 7.95% of countries are in this category, reflecting a lower quality of life.

Purchasing Power Index Distribution

Low (< 60): 40.9% of countries are in this category, indicating lower purchasing power.

Lower-Middle (60-89): 21.6% of countries fall into this range, showing slightly higher purchasing power than the low category.

Middle (90-119): 15.9% of countries are in this bracket, representing a moderate level of purchasing power.

Upper-Middle (120-149): 14.8% of countries are in this category, indicating higher purchasing power, though not the highest.

High (≥ 150): 6.82% of countries are in this top category, reflecting the highest purchasing power.

Pollution Index Distribution

Upper-Middle (50-69): 35.2% of countries are in this range, indicating a moderate level of pollution.

High (≥ 70): 23.9% of countries fall into this category, showing higher pollution levels.

Middle (30-49): 21.6% of countries are in this bracket, representing a lower level of pollution compared to the upper-middle and high categories.

Low (< 30): 19.3% of countries are in this category, reflecting the lowest pollution levels.

8. Spearman Correlation Heatmap:

Quality of Life Index is strongly linked to Purchasing Power Index and negatively related to Pollution Index and Traffic Commute Time Index.

Purchasing Power Index is positively associated with both Quality of Life Index and Cost of Living Index.

Safety Index and Health Care Index have a moderate positive relationship with Quality of Life Index.

Cost of Living Index is positively correlated with Quality of Life Index but negatively with Property Price to Income Ratio.

Property Price to Income Ratio inversely relates to Quality of Life Index and Cost of Living Index.

Traffic Commute Time Index and Pollution Index both negatively impact Quality of Life Index.

Climate Index shows minimal correlation with the other indices.

9. Parallel Coordinate Plot:

In this parallel coordinate plot of 6 indicators for the top 10 countries, Luxembourg starts with a high Quality of Life Index but shows a significant decline in the Cost of Living Index. The Netherlands and Denmark have relatively stable transitions across most indicators. Oman stands out with a sharp drop from a moderately high start in some indices to a lower value in the Traffic Commute Time Index. Switzerland has a complex pattern, with a notable peak in the Cost of Living Index. Overall, the plot reveals diverse relationships and trends among different countries across these various quality - of - life - related indicators.

10. Stacked Bar Chart:

Each bar is segmented into four colors, each representing one of the indices. Switzerland shows the highest overall quality of life, while Qatar has a strong purchasing power but lower scores in safety and health care. Luxembourg and the Netherlands have high purchasing power scores.

In summary, this visualization effectively highlights the quality of life strengths and weaknesses in each country, providing a clear overview of where improvements may be needed.