

1. Data Overview

The “Cost of Living” dataset offers detailed insights into average monthly income and living expenses across countries and years. By including both earnings and the cost required to sustain a basic standard of living, it provides a more comprehensive view of financial conditions across regions and time periods.

Size: 202 rows × 5 columns.

Columns:

- **Country:** Name of each country (e.g., Australia, India, Russia, South Africa, Brazil).
- **Year:** The year of observation, ranging from 2000 to 2023.
- **Average_Monthly_Income:** The average monthly income per person, measured in USD.
- **Cost_of_Living:** The estimated average monthly living expenses per person, also in USD.
- **Region:** The geographical region to which each country belongs (e.g., Asia, Africa, Europe, Oceania, South America, North America).

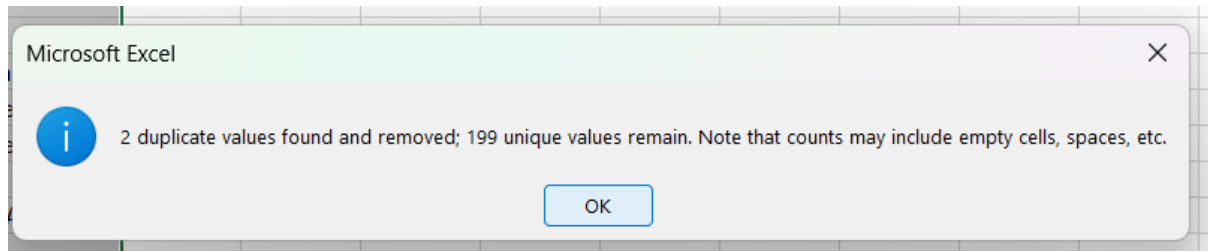
1	Country ▾	Year ▾	Average_Monthly_Income ▾	Cost_of_Living ▾	Region ▾
2	Australia	2013	3483.92	1106.07	Oceania
3	India	2019	7771.03	5422.78	Asia
4	Russia	2004	6991.3	3972.36	Europe
5	South Africa	2011	6628.04	6755.75	Africa
6	Brazil	2015	2434.27	2656.36	South America
7	Australia	2015	1781.66	4575.3	Oceania
8	China	2020	5514.82	6551.39	Asia
9	Japan	2006	7470.32	1076.56	Asia
10	Australia	2003	4675.72	6585.71	Oceania
11	Russia	2000	4787.1	4940.05	Europe
12	Russia	2004	2599.84	847.72	Europe
13	South Africa	2022	6271.2	2386.36	Africa
14	Brazil	2009	1902.83	5073.94	South America
15	India	2021	2927.59	844.51	Asia
16	South Africa	2004	3690.77	4242.33	Africa
17	South Africa	2003	4307.08	2682.83	Africa
18	Japan	2001	2318.07	4498.04	Asia
19	Canada	2019	1361.28	701.9	North America
20	Brazil	2009	5079.65	6152.14	South America

Although the dataset does not specify its source, it is likely compiled from reputable global databases. Organizations such as the World Bank, IMF, and ILO regularly publish

standardized data on income and living costs. Additionally, platforms like Numbeo aggregate cost-of-living information from both official and crowd-sourced inputs. Given the dataset's structure — covering average income, living expenses, and regional breakdowns over time — it appears to be derived from a combination of credible international and governmental sources. These institutions are known for providing reliable and up-to-date economic data.

2. Data Cleaning

Duplicate Records



The dataset did contain 2 duplicates: Go to Data and Remove duplicate. I found 2 duplicates in this data, I decided to remove them because they can affect the result of analysis causing inaccurate outputs. Based on the notification, two duplicate records were identified and removed, resulting in a final dataset of 200 distinct observations. We removed duplicates at the initial stage to minimize potential impact on the results when handling missing values. Retaining these duplicates could have distorted the summary statistics, visual representations, and overall analysis. By ensuring that each country-year combination appears only once, the dataset maintains greater accuracy, consistency, and interpretability in the findings.

Missing Values

1	Country	Year	Average_Monthly_Income	Cost_of_Living	Region
160	Australia	2012		4692.15	Oceania
176	Mexico	2006		4219.26	North America

Rows 160 and 176 contain missing values in the Average_Monthly_Income column.

1	Country	Year	Average_Monthly_Income	Cost_of_Living	Region
160	Australia	2012	4266.46	4692.15	Oceania
173	Canada	2006	4266.46	6918.49	North America
176	Mexico	2006	4266.46	4219.26	North America

I selected the median as the preferred measure of central tendency because it more accurately reflects the typical income level across countries in the dataset. Unlike the mean, which can be heavily influenced by a small number of high-income countries, the median is resistant to such outliers. As the midpoint of the income distribution, the median offers a clearer picture

of the general economic condition, especially in datasets with significant income disparities. This makes it a more robust and reliable indicator for global analysis.

1	Country	Year	Average_Monthly_Income	Cost_of_Living	Region
26	Mexico	2003	888.68	562.27	
38	Mexico	2000	3588.81	5645.74	

Rows 26 and 38 have missing values in the Region column, with no region names provided for these countries.

1	Country	Year	Average_Monthly_Income	Cost_of_Living	Region
26	Mexico	2003	888.68	562.27	North America
38	Mexico	2000	3588.81	5645.74	North America

In the dataset, the Region column is missing for certain rows. To address this, I reviewed other entries and identified that Mexico belongs to North America. By assigning "North America" to the missing values, I ensured consistency and accuracy within the dataset. This approach helps prevent data fragmentation and supports a clean, standardized classification system for regional analysis.

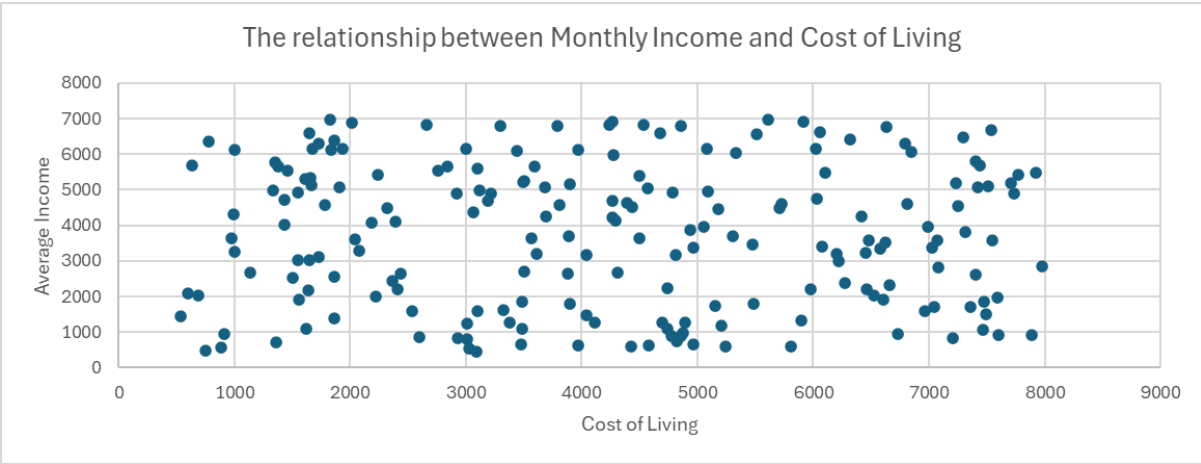
3. Descriptive Statistics

Average_Monthly_Income		Cost_of_Living	
Mean	4244.193216	Mean	3705.127286
Standard Error	150.0444918	Standard Error	140.515613
Median	4266.46	Median	3688.09
Mode	4266.46	Mode	#N/A
Standard Deviation	2116.638031	Standard Deviation	1982.216654
Sample Variance	4480156.553	Sample Variance	3929182.864
Kurtosis	-1.166687671	Kurtosis	-1.272006829
Skewness	0.055307916	Skewness	-0.04103302
Range	7441.82	Range	6516.53
Minimum	534.74	Minimum	464.49
Maximum	7976.56	Maximum	6981.02
Sum	844594.45	Sum	737320.33
Count	199	Count	199
Confidence Level(95.0%)	295.8903602	Confidence Level(95.0%)	277.0992448

The average monthly income is \$4,244, while the average cost of living is \$3,705. Both variables show high standard deviations (Income: \$2,116.64; Cost: \$1,982.22), indicating substantial variability across countries. This suggests significant economic disparities, where some countries enjoy a healthy income-cost gap, while others may struggle with low income or high living expenses.

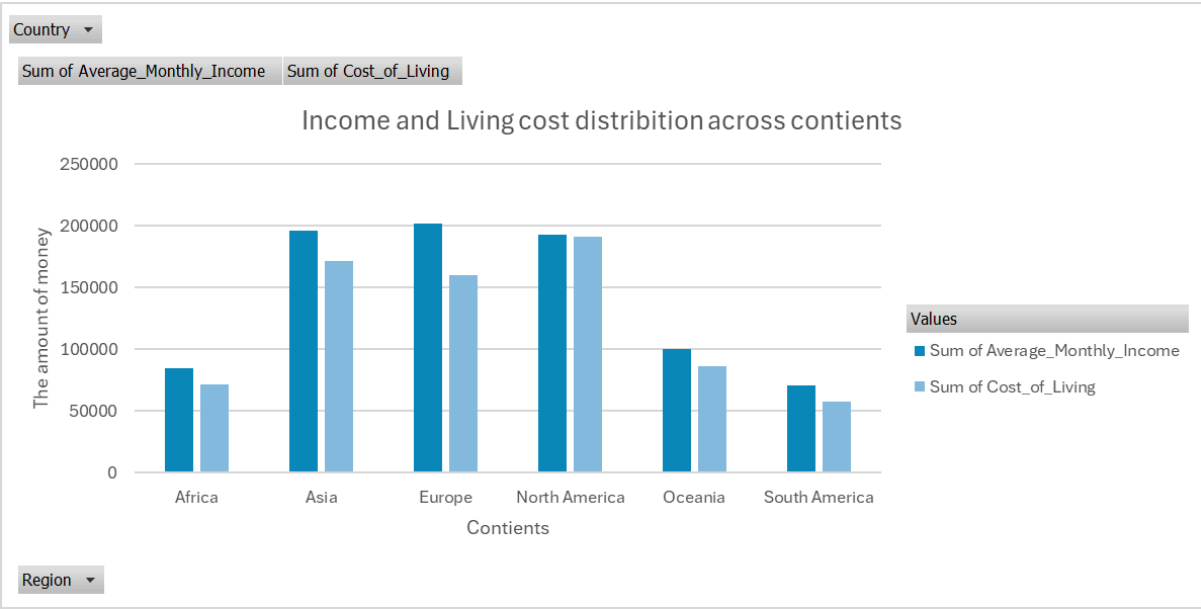
4. Insights

Insight 1: The Relationship Between Income and Cost of Living



The scatter plot illustrates the relationship between average monthly income (y-axis) and cost of living (x-axis). The data points are widely scattered without a clear linear pattern, indicating a weak or no strong correlation between the two variables. Notably, some countries with high living costs do not correspondingly have high incomes, and vice versa. This suggests that income and cost of living tend to vary independently across different countries.

Insight 2: Income and Living Cost Distribution Across Regions



The bar chart compares total income and total cost of living across regions. Europe and Asia stand out with the highest overall income and expenses, reflecting strong economic activity. In contrast, North America shows a smaller gap between income and living costs, suggesting tighter budgets. These differences highlight varying levels of financial comfort and savings potential among regions.