Name: Le Van Cong

IRN: 2132300013

MIS 311

Individual Assignment

Assignment#1: Personal Portfolio Site

**1. Data Overview**

The "Cost of Living" dataset contains 200 entries across 5 columns (Country, Year, Average Monthly Income, Cost of Living, and Region), providing a global overview of income and living expenses from 2000 to 2023 across six major regions, including Asia, Europe, Africa, North America, South America, and Oceania. These data might come from government statistical agencies. The context of this dataset seems to focus on economic indicators related to income and living expenses, which could be useful for comparative analyses of affordability, economic trends, or regional disparities.

**2. Data Cleaning**

**Missing Value**

In this file, I found 2 missing values ​​in the average monthly income column and 2 missing values ​​in the region column. I chose the technique of deleting these missing values ​​for the following reasons:

Deleting missing values ​​is reasonable because the dataset has 200 rows but only 4 missing values ​​(2%), which is insignificant to affect the analysis.

Avoid bias because there is no need to guess replacement values, especially important for categorical columns like Region.

The data after deletion is still representative because the missing values ​​are randomly distributed, not concentrated in any specific group.

The method is simple and effective when the missing rate is low, helping to maintain the integrity of the original data.

**Duplicate Value**

I found 2 duplicate values ​​and I used the removal technique for the following reasons:

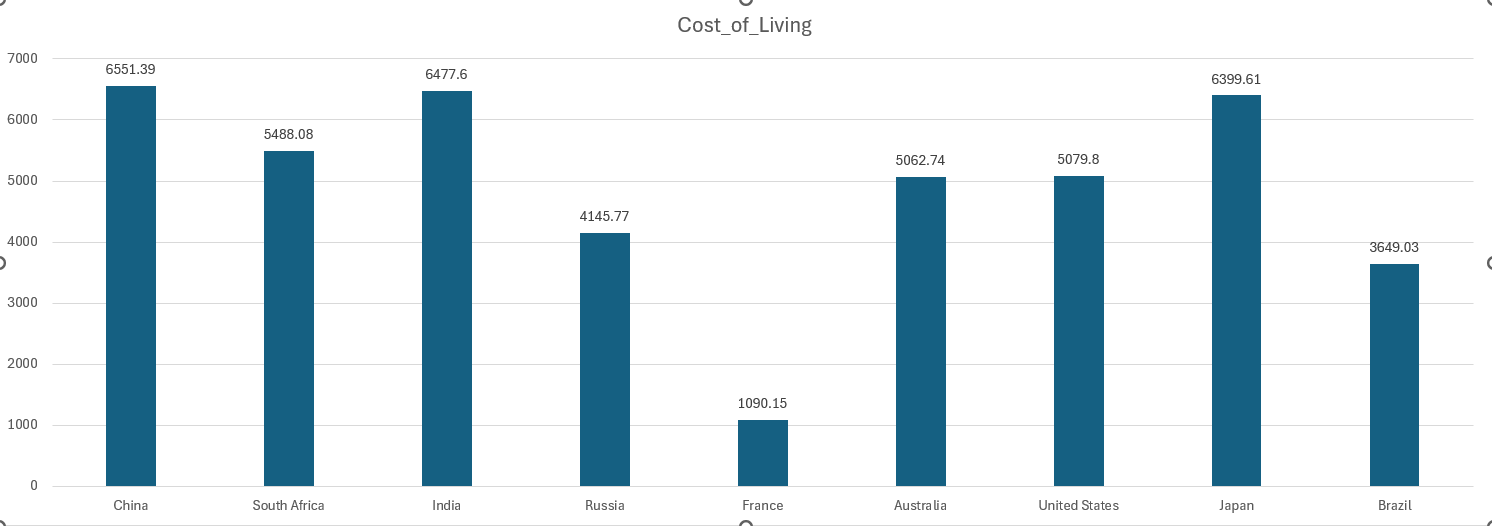
Avoid skewing the analysis results: Duplicate values ​​can cause statistics to be counted multiple times, leading to incorrect conclusions.

Ensure data consistency: Each record should represent a unique observation, duplicates break this principle.

Save processing resources: Removing duplicate data helps reduce the size of the dataset without losing important information.

Suitable for most analysis problems: Except for special cases where duplicates need to be kept, removal is the optimal way to handle identical records.

**3. Descriptive Statistics**



**Key Insight 1: Cost of Living Disparities Across Regions**

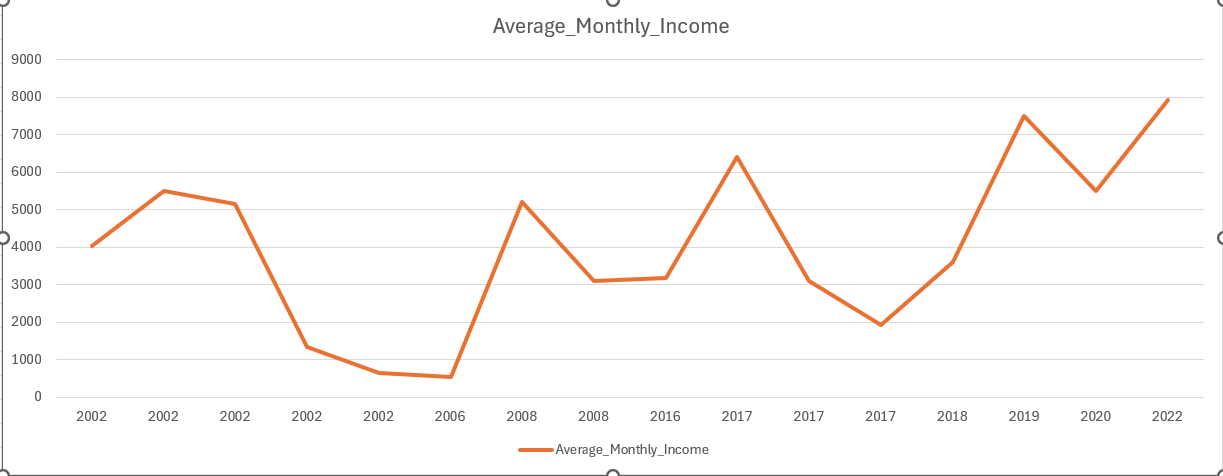
Finding: The data reveals significant variations in the Cost\_of\_Living across different regions in 2020, with Asian countries like China, India, and Japan exhibiting the highest costs, while European nations like France show the lowest.

**Implication:**

The high cost of living in Asian countries like China, India, and Japan (from 6399.61 to 6551.39) suggests that residents may face greater financial burdens for basic necessities, which could impact their overall quality of life and savings potential.

Conversely, the lower cost of living in France (1090.15) indicates a more affordable lifestyle, potentially attracting expatriates or businesses looking to reduce operational expenses.

Policymakers and businesses should consider these disparities when making decisions about wages, pricing, or market entry strategies to ensure competitiveness and sustainability in high-cost regions.



**Key Insight 2: Income Volatility in China Over Time**

Finding: China's Average Monthly Income demonstrates dramatic fluctuations between 2002 and 2022, with values ranging from as low as 635.83 to as high as 7,924.70, reflecting a 12-fold increase over two decades.

**Implication:**

The rapid rise in income highlights China's remarkable economic growth, which has likely lifted millions out of poverty and expanded the middle class, contributing to global economic shifts.

However, the extreme volatility, especially in early years (e.g., 2002 data with multiple inconsistent entries), raises concerns about data reliability, such as potential measurement errors or inconsistent reporting methods.

Analysts should verify the data sources and methodologies to ensure accuracy, as flawed data could lead to misleading conclusions about economic trends or policy effectiveness in China.