

# Key Factors Influencing Cost of Living Across the World

Nidhi Joshi

Data Science, School of Science and Technology,  
City, University of London,  
London, United Kingdom  
e-mail: [nidhi.joshi@city.ac.uk](mailto:nidhi.joshi@city.ac.uk)

**Abstract**— In recent years, there have been a lot of trend-shifting events across the world. These events have led to a lot of disparity in the earning as well as spending pattern across continents. The nations realized the importance of being self-reliable in terms of production as the supply chains were heavily disrupted resulting in price spikes in almost every sector. This paper analyses the impact of increased cost of living across countries and their responses towards the same in terms of expenditure. It assesses various factors encompassing a basic quality of life and draws conclusions on which factors have the maximum impact on the index. Also, how the impact and response changes with countries.

**Keywords**— Cost of Living, Purchasing Power, KMeans Clustering

## I. INTRODUCTION

“Unprecedented price surges mean that for many people across the world, the food that they could afford yesterday is no longer attainable today,” said Achim Steiner, UNDP Administrator[1] In a world already reeling from natural calamities and issues like global warming, the COVID-19 pandemic worsened the situation. With disruptions in food and energy supply-chain, everyone across the globe felt the tremors in form of inflated prices and much higher cost of living than before. Furthermore, the wars between Ukraine-Russia and Israel-Gaza have worsened an already challenging situation “pushing millions more people into poverty”[1] This research provides with a peek into the effects of and on Cost of Living with respect to various related factors like Rentals, Food-related indices, Purchasing Power etc. It would help readers understand the quality of life based on various indices across countries. The dataset for the analysis were gathered from Numbeo[2] who crowd-source their data globally. The multiple index factors involved in the research are measured by correlation to find out dependencies on each other and certain external components like average salaries in that region. Research particularly tries to focus on analyzing Local Purchasing Power in the wake of higher Cost of Living and whether the higher cost of living affects the Purchasing Power in any way. A Cost of Living Index considers all basic needs of living like housing, food, transportation, education, healthcare etc. and hence an all-round metric reflective of life quality in that country.

## II. ANALYTICAL QUESTIONS

The targeted focus areas of this research are measuring the economic state of countries on the basis of direct or indirect impact of Cost of Living indices or the lack of it. In pursuit of analysing quality of life in countries across the world, this research correlates how The Local Purchasing Power Index varies with Cost of Living Index. The individuals referring to this paper would be assisted with answering the following questions:

- A. *What factors play a significant role in shaping the cost of living across diverse countries, and how does this variation manifest on a global scale?*
- B. *How does the Local Purchasing Power Index align with the overall cost of living and Restaurant Prices or Groceries Index, shedding light on the quality of life in various cities?*
- C. *In countries with a high cost of living, are there any instances where residents possess a greater local purchasing power?*

In the end, this paper will include clustering the results based on Cost of Living Index and Local Purchasing Power Index through machine learning techniques like KMeans Clustering method. This model aims to quantify vectors and group together the available values into a given number of clusters based on proximity with their respective mean values. This gives insight into how the behavior of Local Purchasing Power changes with respect to Cost of Living based on the respective cluster groups.

## III. DATA (MATERIALS)

This research is based on analyzing four different datasets from Numbeo[3][4] and Kaggle[5] respectively. The two separate datasets from Numbeo contained quantitative index values related to overall quality of life in multiple countries which were then combined to make sure maximum available data is utilized for the analysis. The dataset from Kaggle further added on to the merged data, the features like individual item or service prices and average monthly salary details. The index details for 549 cities across multiple

country locations were merged with the individual price details available for 4956 cities in different countries.

Since, all the indices were important in our analysis, the extra cities in the second dataset (with individual pricing) were filtered out and only that data were merged and kept which mandatorily contained with the index records. In the end, the final dataset consisted of 488 records with 63 features, 61 of whom were quantitative and 2 qualitative ('city' and 'country'). This combined dataset was apt for a comprehensive analysis around all aspects surrounding Cost of Living. It is worth noting that all the indices are relative to that of New York City (NYC) with NYC having 100% baseline value [6]

Although, there were certain limitations with the obtained datasets. The data were in a constant crowd-sourcing phase and hence, incomplete at the time of analysis. So, it cannot be assumed that the comparisons apply to all the existing cities in the world or even all existing countries.

#### IV. ANALYSIS

##### A. Data preparation

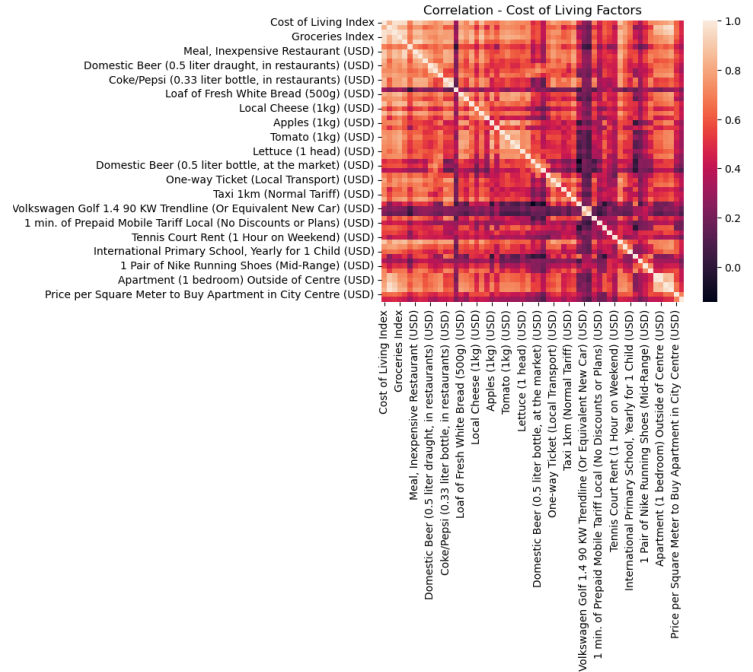
Since the aim of this research paper is to correlate various factors surrounding the quality-of-life indices, the basic index values for all the available cities (with their respective countries) were required as a basic rule. Hence, when merging with the Kaggle data, only the records which were available in the Numbeo table were merged and rest of the extra records were discarded. Missing values were checked for only the index related columns. In the final merged dataset, since the price columns were in numerical form (X1, X2, X3...), hence, they were mapped with their proper names using data from Kaggle again to display relevant names for our analysis later. The column indicating the data quality and rank of cities were dropped because the data quality was same for all the rows and ranks were not required for our further analysis. Some important indices and their calculations are explained as below:

TABLE I. INDEX DEFINITIONS

Index Name	Definition[6]
Cost of Living Index	Prices of consumer goods and services excluding housing rent relative to NYC
Rent Index	Rentals in a city compared to NYC
Cost of Living Plus Rent Index	Prices for rent, consumer goods and services prices with respect to NYC
Restaurant Prices Index	Compares meal/drink prices while eating out, with that of NYC restaurants and bars
Groceries Index	Grocery prices with relevant weights for each item compared to those of NYC
Local Purchasing Power Index	Purchasing power on an average salary relative to NYC

Furthermore, the Local Purchasing Power Index is calculated as:

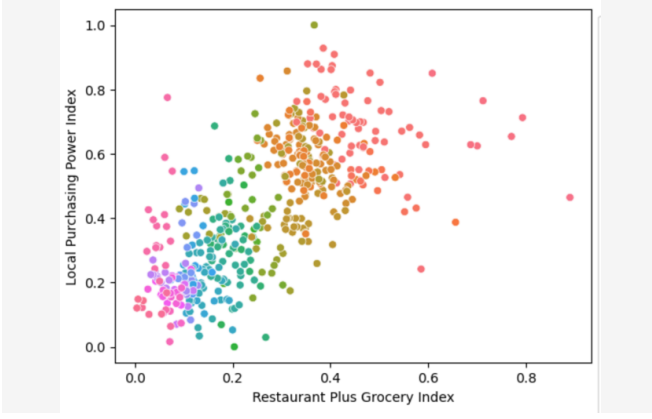
(Average Disposable Salary / Sum (Price in the city \* (cost of living factor + rent factor)))[7]



##### B. Data derivation

Some of the quantitative values were on different scales as compared to the index values and hence had to be scaled down to a standard level to make justified comparisons. The method used to scale the values was MinMax Scalar function which brought all the numerical column values within [0,1] interval.

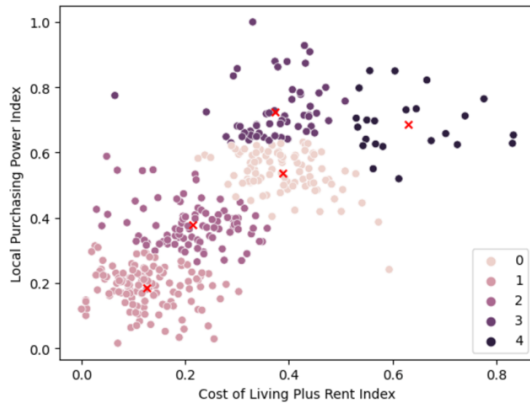
When looking at the pair grid plots, it was determined that Restaurant Prices Index was linearly related to Groceries Index. Since both were an indicator towards food related Index, a composite variable was calculated for the same with higher weightage to Restaurant Prices Index than Groceries Index. This derived variable 'Restaurant Plus Grocery Index' was then used to compare with Local Purchasing Power Index to analyze the correlation between the two. When they were both plotted against each other with countries as hue, different clusters were formed with some countries with high Restaurant Plus Grocery Index as well as high Local Purchasing Power moving in a somewhat linear fashion but with many scattered datapoints as well. The comparison was such that countries like Nepal, India, Nigeria, Libya, Egypt, Pakistan et al were low on Restaurant Plus Grocery Index but also low on Local Purchasing Power while countries like Bermuda, Switzerland, Bahamas, United States, Iceland etc. indicated high values for both the indices.



### C. Construction of models

It was observed that Cost of Living Plus Rent Index was a better measure of an index as compared to only Cost of Living which excluded rent in its calculation. So, both these features were selected into the model dataset which was split into 80-20% for training and testing sets.

The mentioned variables in the training set were passed to the model to fit and form 5 clusters using KMeans clustering method which is used for unsupervised learning. Here, the clusters were to determine the changing relationship of the two indices with each cluster and to determine the centroid points for each cluster to assess them relative to each other. After training the model, the clusters were plotted on a scatterplot to see how the model fared.



### D. Validation of results

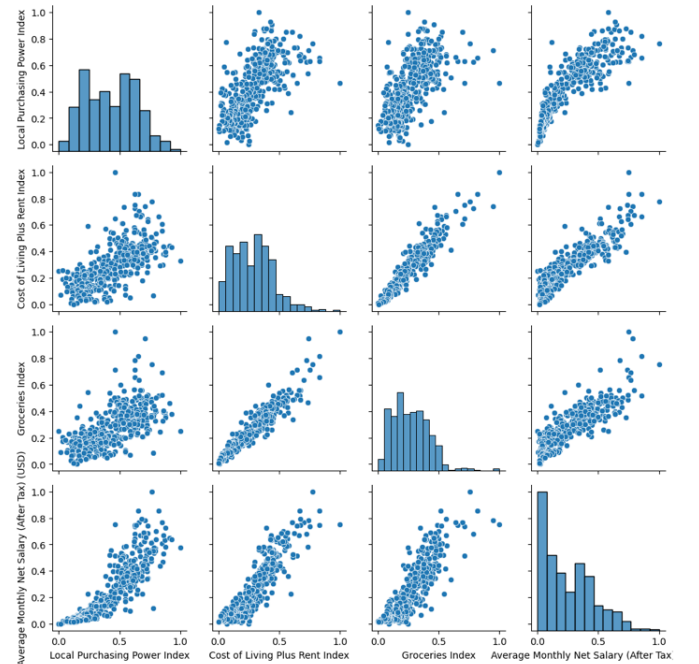
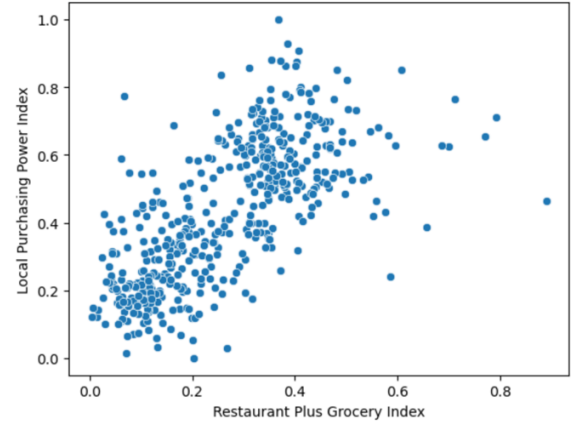
The trained KMeans cluster model is further validated by predicting on the test set. The scatterplots are plotted to examine the clusters formed and centroids are plotted on top of the scatterplot. Silhouette Score is calculated for both the trained model and the prediction. Silhouette score is a “measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from  $-1$  to  $+1$ , where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.”[8]

## V. FINDINGS, REFLECTION AND FUTURE WORK

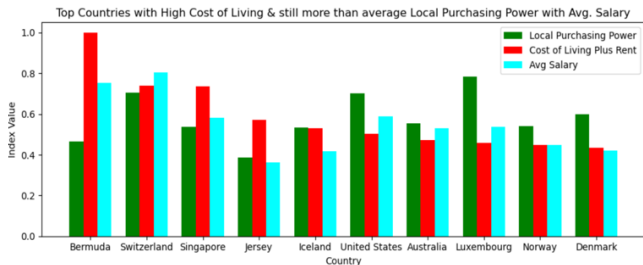
While correlating the relevant indices on pair plots, following observations were made - Cost of Living Plus Rent Index was highly correlated to Groceries Index and Restaurant Price Index.

It was also concluded through another pair plot that when Groceries become expensive, they have a direct and strong impact on the overall Cost of Living expenses. Furthermore, it was observed that the Average monthly salary was the only one common factor influencing almost every other index.

While trying to establish a credible relationship between Local Purchasing Power and Restaurant Plus Grocery Index (the composite index which was calculated earlier), it could be inferred that even though there was a vague linear direction but the datapoints were very scattered and hence, it can be said that both the indices were not really dependent and that even if Restaurant Plus Grocery become expensive, the Local Purchasing Power can be almost independent of the impact.



To determine if there were evidence of high Local Purchasing Power despite a high Cost of Living in any of the countries, mean of both the indices were taken as the threshold beyond which all countries were filtered and merged. The countries which were both with High Purchasing Power as well as Cost of Living were further filtered and sorted as per descending Cost of Living. Top 10 records were finally plotted with both the indices along with Average Monthly salary for a better comparison.



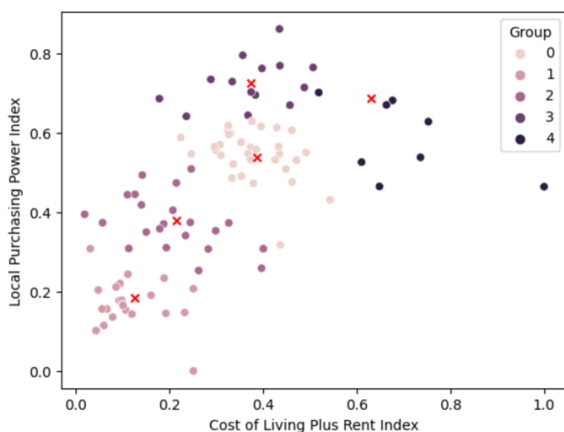
We can observe that in case of countries like Luxembourg, United States, Denmark, Norway, and Australia respectively, the Local Purchasing power exceeds the Cost of Living Plus Rent. This may be attributed to higher Average Salaries in most of those countries.

However, it is interesting to note that despite much higher salaries, people in Switzerland exhibit lesser Local Purchasing Power in the wake of high Cost of Living Plus Rent.

Furthermore, through the Clustering method, the obtained scores were as follows:

TABLE II. KMEANS CLUSTER SCORES

Dataset	Silhouette Score
Training	0.41085
Test	0.39819



Since, both the scores are above 0.25, it can be concluded that even if not distinguishably far, the clusters are still decently distant from each other to identify similar behavior.

This research can be used to understand different socio-economic factors among different countries. While some countries are not affected by any difference in living expenses and continue to spend, some countries are extra sensitive towards any fluctuations towards the same. It was interesting to see certain countries like Switzerland aiding their economy better by spending in a balanced capacity. This research can be a deciding factor for economies to take corrective measures towards a better economic status.

Future work can be based on a larger and more comprehensive dataset for maximum cities and, cost of living and purchasing power trends can be analyzed through years to understand different regions and mindsets better.

## REFERENCES

- [1] K. Whiting, "The cost-of-living crisis is having a global impact - here's what countries are doing to help," *World Economic Forum*, 2022. <https://www.weforum.org/agenda/2022/09/cost-of-living-crisis-global-impact/>
- [2] "About This Website - Numbeo.com," *www.numbeo.com*. <https://www.numbeo.com/common/about.jsp>
- [3] NUMBEO, "Cost of Living Index 2019 Mid-Year," *Numbeo.com*, 2019. <https://www.numbeo.com/cost-of-living/rankings.jsp>
- [4] Numbeo, "Current Cost of Living Index by City," *Numbeo.com*, 2019. [https://www.numbeo.com/cost-of-living/rankings\\_current.jsp](https://www.numbeo.com/cost-of-living/rankings_current.jsp)
- [5] "Global Cost of Living," *www.kaggle.com*. <https://www.kaggle.com/datasets/mviera101/global-cost-of-living/data> (accessed Dec. 24, 2023).
- [6] Numbeo, "About Cost of Living Indices At This Website," *Numbeo.com*, 2009. [https://www.numbeo.com/cost-of-living/cpi\\_explained.jsp](https://www.numbeo.com/cost-of-living/cpi_explained.jsp)
- [7] "Methodology and motivation for Numbeo.com," *www.numbeo.com*. [https://www.numbeo.com/common/motivation\\_and\\_methodology.jsp](https://www.numbeo.com/common/motivation_and_methodology.jsp)
- [8] Wikipedia Contributors, "Silhouette (clustering)," *Wikipedia*, Aug. 09, 2019. [https://en.wikipedia.org/wiki/Silhouette\\_\(clustering\)](https://en.wikipedia.org/wiki/Silhouette_(clustering))

TABLE III. WORD COUNT OF SECTIONS

Section Name	Word Count
Abstract	121
Introduction	241
Analytical Questions	230
Data (Materials)	244
Analysis	717
Findings, Reflections and Further work	472
Total	2070