Global Cost of Living Exploration

Studies on the global cost of living are important to determine countries that are being pushed into poverty and need help from other countries to continue their everyday life. On the other hand, the policies implemented in countries that thrive due to their low living expenses can also be examined by governments that wish to ease the impact of high living expenses or inflation in their country. However, implementing good policies is only part of it since a country's natural resources and income from import export activities play a big role in determining their wealth, which then influences their cost of living. By comparing the cost of living across various types of expenses, we can gain insights on the gap of expenses' costs being faced by people from different parts of the world.

Thus, this report aims to carry out a simple analysis on the global cost of living dataset obtained from Kaggle. The main objectives of this report are:

1. To compute and provide a brief overview on the descriptive statistics of the global cost of living across salary and selected expenses.
2. To determine whether mean or median is a better measure to make comparisons for the cost of living based on the presence of outliers.
3. To determine countries with the highest cost of living based on certain expenses.

The significant findings for objective one and objective three will be highlighted to help readers understand the cost of living situation in different countries. Readers can also go through the Jupyter Notebook titled "task3.py" to view the codings utilised and brief explanations provided. Also note that references are provided for readers who want to get more information on the particular topics.

Data

The data utilised for the analysis carried out in this report is the "Global Cost of Living" dataset that was obtained from Kaggle. This dataset contains information on the cost of living in 4956 cities across countries all over the world. The data was gathered by Miguel Piedade through the scraping of Numbeo's website, the world's largest cost of living database. The original dataset has

about 4956 observations across cities in many different countries worldwide. There are 58 variables provided, where 3 of them are categorical variables (city, country and data quality), whereas another 55 are numerical variables involving cost of living expenses and salary. The data_quality variable is a binary categorical variable with value 0 if Numbeo considers that more contributors are needed to increase data quality, and value 1 otherwise. Therefore, the dataset is filtered to include only observations with good data quality for the analysis to be carried out. Besides that, the dataset is filtered to include only 11 variables so that more focus can be put on certain living expenses. Next, missing values are dropped from the data set before renaming the variables to enhance clarity for readers. The final dataset after cleaning contains 836 observations with 12 variables, which are explained in the table below. All the observations of the numerical variables are recorded in USD (except mortgage_IR in percentage) to enable comparison across countries with different currencies.

Original Variable Name	Renamed Variable	Description
x50	aprtmnt_city_centre	Cost for an apartment with 3 bedrooms in the city centre
x51	aprtmnt_out_centre	Cost for an apartment with 3 bedrooms outside of the city centre
x41	cinema	Cost for an international movie release ticket
country	country	Name of the country
x39	fitnessclub	Monthly fee for fitness club membership per adult
x33	gasoline	Gasoline cost per litre
x38	internet	Cost for Internet (60 Mbps or More, Unlimited Data, Cable/ADSL)
x3	mcd_meal	Cost for a McMeal or equivalent combo meal at McDonalds

x55	mortgage_ir	Yearly mortgage interest rate for 20 years fixed rate (%)
x54	salary	Average monthly net salary after taxes
x43	school	Yearly fees for international primary school per child
x36	utilities	Cost for basic utilities including electricity, heating, cooling, water and garbage for an $85m^2$ sized apartment

Note that country and salary are not cost of living expenses. Country is the only categorical variable in this dataset, the other variables are all numerical.

Results and Discussion

Objective 1

The table below shows the descriptive statictics including maximum, minimum, mean, median, standard deviation, skewness and kurtosis for all the countries combined. This is more of a global view rather than by country. The "country" column is excluded from this table as it is a categorical variable and descriptive statistics cannot be computed. The table below is obtained straight from the output of Jupyter Notebook.

Descriptive Statistics	aprtmnt_city_centre	aprtmnt_out_centre	cinema	fitnessclub	gasoline	internet	mcd_meal	mortgage_IR	salary	school	utilities
maximum	7146.840000	4325.000000	21.370000	168.570000	2.770000	715.630000	17.630000	50.000000	7935.380000	59421.680000	454.060000
minimum	103.930000	72.640000	1.120000	7.800000	0.030000	3.330000	2.290000	0.800000	35.750000	0.000000	12.370000
mean	1476.554677	1087.895395	8.994641	39.539151	1.481292	39.119474	7.282117	6.386806	1984.429438	9695.825467	144.796806
median	1239.200000	912.720000	8.890000	36.620000	1.375000	34.580000	7.380000	5.140000	1558.620000	7587.805000	140.275000
std	1061.786680	787.223607	4.139808	17.792918	0.499958	35.217683	2.589363	5.125079	1611.681415	7682.529410	77.811360
skewness	1.432257	1.194505	0.229901	1.821299	-0.014883	9.916923	0.711770	2.506130	0.854264	1.738058	0.387472
kurtosis	2.625398	1.317619	-0.696087	6.328059	-0.729904	173.073629	1.445890	11.384205	-0.010202	4.754038	-0.466740

Firstly, we observe how the cost for an apartment with 3 bedrooms in the city centre and outside the city centre have a huge gap. This is to be expected because living expenses in the city are sure to be higher as more jobs with higher pays are usually located in the city centre. Even the kurtosis value for apartments in the city centre are higher, indicating higher tendencies for outliers to occur, giving a wider range of prices as is also portrayed by its higher standard deviation value. International movie prices per ticket have nearly identical mean and medians of $8.99 and $8.89, meaning that for all countries combined, one ticket could cost on average $8.99. Interestingly, the cost of gasoline per litre is slightly skewed to the left at a value of -0.015. A negative skew would mean that the left tail of the gasoline price distribution is slightly longer relative to the right tail, indicating more countries with gasoline prices below the mean of $1.48. It is some good news indeed. On the other hand, we have cities with the price of a McMeal in McDonalds (well-known

fast food chain) being 238.89% above the median price. Seems like not everyone in the world can easily enjoy fast food anytime they want to.

Next, observe how certain variables such as "fitnessclub", "internet", "mortgage_IR", "salary", "school" and "utilities" have a huge range value (maximum value – minimum value). Their range values are $160.77, $712.30, 49.2%, $7899.63, $59421.68 and $441.69 respectively. It is highly alarming that the cost of expenses range so widely across countries. With such high prices to join a fitness club, even people who want to stay healthy would need to think twice! Additionally, the high disparity of mortgage interest rates (range of 49.2%) indicate that it is much more expensive to take loans to buy properties in certain countries. In that case, we observe from the dataset that Argentina is the country with the mortgage interest rate of 50%, causing their citizens to have to pay so much more for a property than its face value. The interest rates are hiked up to combat inflation, but this of course crushes the dreams of potential homebuyers (Gillespie 2018). Furthermore, view the salary discrepancy among countries which is absolutely saddening. There are cities with a monthly salary of only $35.75, which is only 2.29% of the global median salary. It is a known fact that salaries/income are usually skewed to the right where majority of people earn a small salary, but a minority of wealthy individuals earn big salaries. This can be described by the 80/20 rule coined by Vilfredo Pareto.

Another finding from this table to emphasise is the fact that there is a certain city with free education for international primary school level. By going through the dataset, it is observed that a city in Finland called Jyväskylä, does not charge anything ($0.00) for its international primary school fees. Apparently, the education services of the City of Jyväskylä are responsible for basic education, which is free of charge for comprehensive school (first to ninth year). Comprehensive school is intended for children and teenagers aged 7 to 16, even including immigrants who permanently reside in Finland. More on the education in Jyväskylä can be obtained through this article on City of Jyväskylä's website (Jyväskylä t.th.).

Objective 2

For objective 2, boxplots for each numerical variable are plotted to observe if the values contain outliers or not. This is so that the proper central tendency, either mean or median, can be chosen for analysis needed in objective 3. The boxplots can be seen in the diagram below.
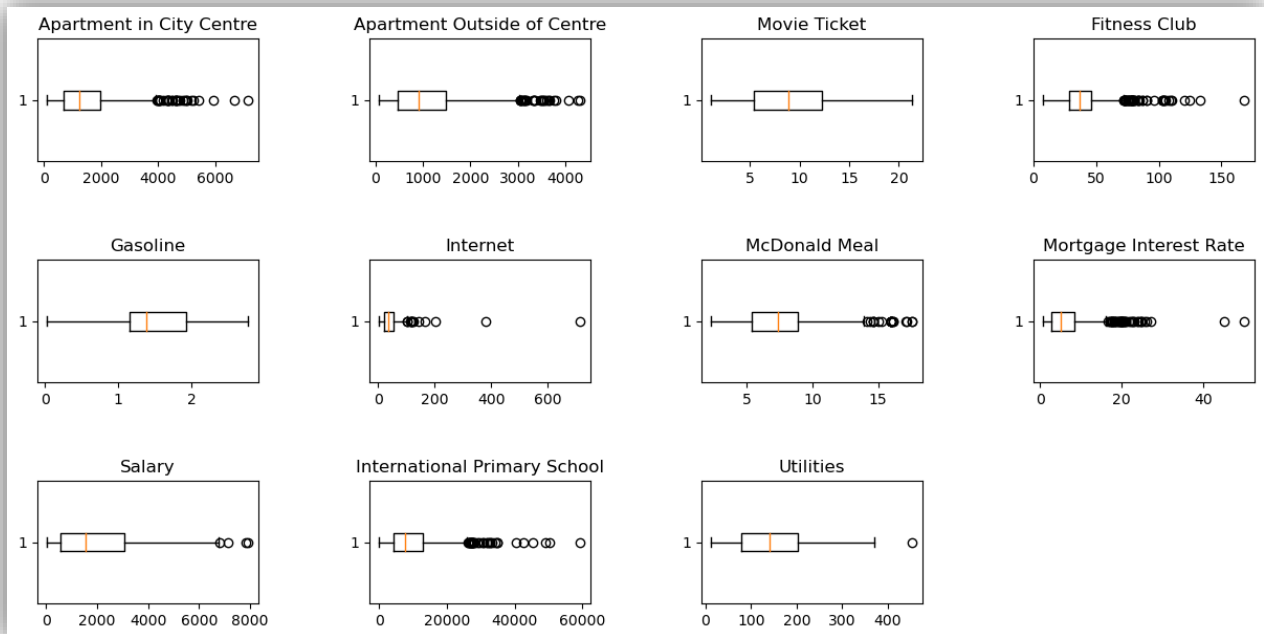

Outliers are represented in boxplots as circles that lie outside of the smallest and largest values within the inner fences. From the diagram above, it is observed that variables such as "aprmnt_city_centre", "aprmnt_out_centre", "fitnessclub", "internet", "mcd_meal", "mortgage_IR", "salary", "school" and "utilities" all have outliers that lie outside the largest value within the inner fence. It can be said that there are quite many outliers in these variables. Anyway, even by looking at the table in objective 1, it is noticeable that most variables have positive kurtosis. The distributions exhibit a leptokurtic behaviour, deviating from the normality assumption and thus leading to the tendency of having outliers. Therefore, in objective 3 which involves further analysis of the cost of living, comparisons are made based on the median cost instead of mean cost. This is because mean is particularly susceptible to the influence of outliers that could drag the mean up (skewed to the right), causing the value to be a lot higher and not

really representative of the true average. The median is a more robust and accurate measure when data is skewed.

Objective 3

In this final objective, barplots are used to represent the top 10 countries with the highest living cost expenses (across several variables) and salary. This way, we can identify countries that have a high cost of living in general or are maybe going through bad inflation. Certain interesting findings are highlighted in this subsection.
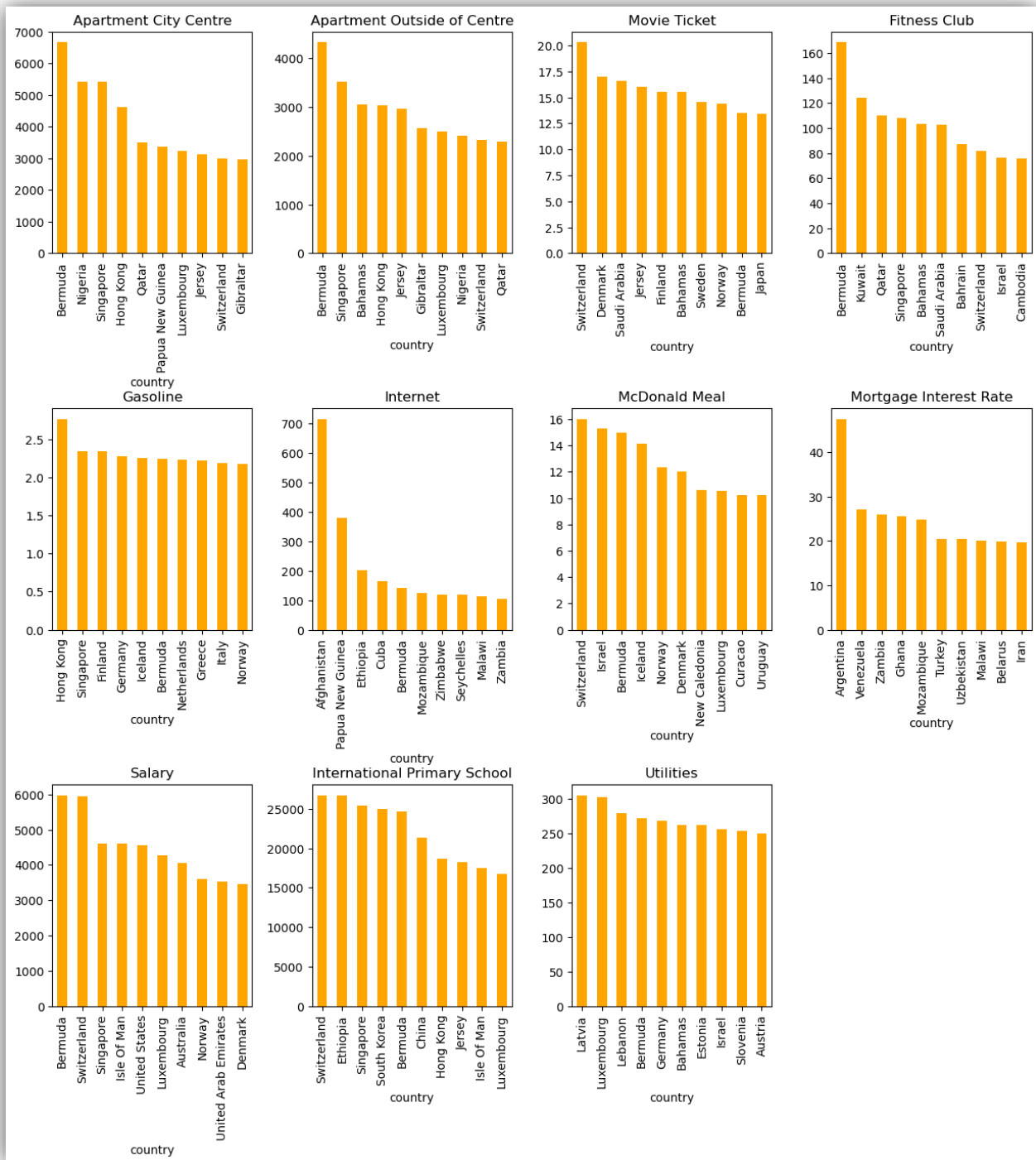
The most obvious finding here is that Bermuda occurs in the top 10 expenses for 9 out of 10 types of expenses available in the barplots. It even occurs as the country with the most expensive cost for apartment in the city centre, outside of the city centre and fitness club membership at $6666.67, $4325.00 and $168.57 respectively. This could be due to them having the highest median net salary

of $5973.81. Bermuda is said to have one of the highest per capita incomes in the world, mainly due to their offshores financial services. According to this dataset, this British island territory in the North Atlantic Ocean has the highest salaries but also a very high standard or cost of living. By looking at the barplots closely, we can also see that Switzerland has the second highest median salary in the world ($5955.26), but their cost of living is also quite high as they appear in the top 10 for 6 types of expenses. On the other hand, it is also worth noting that the United States appears in fifth place for countries with the highest salary, but they do not appear in any of the top 10 highest expenses. Considering the variables involved, this could indicate that Americans might have more discretionary income and may be able to have more savings. Thus, the Bermuda and United States differ even though both countries have high salaries. One country has high salary with high cost of living, while the other has high salary with what can be considered as a more comfortable cost of living.

Surprisingly, one of the least developed countries in the world, namely Afghanistan, has the highest median cost for internet at $715.63, which is so much higher than Papua New Guinea at $379.33. An article by Hakimi (2021), also states about Afghans complaining about their super high priced but low quality Internet. Even the country at the tenth spot, Zambia, has median cost for internet of $103.74, three times the global median cost ($34.58). Ironically, 8 out of 10 countries (besides Bermuda and Seychelles) in the top 10 for Internet costs, are poor, underdeveloped countries with somewhat small economies. With such high Internet prices, the citizens which are mostly under the poverty line may not be able to gain easy access to the Internet. Internet subscription might be a luxury for them on which they cannot simply spend money.

Moving on, Bermuda, Nigeria, Singapore, Hong Kong, Qatar, Papua New Guinea, Luxembourg, Jersey, Switzerland and Gibraltar appear as the top 10 for median cost in apartments in the city centre. The same countries except for Papua New Guinea, which is replaced by the Bahamas, also have the highest median cost for apartments outside of the city centre. The results portray that these nine countries have expensive properties, and it might not be that easy for every citizen to buy homes. Contrary to the city of Jyväskylä in Finland having free international primary school, we see the top 10 countries listed in the barplot having median school costs of $16759.70 to $26711.26. Education can certainly be very expensive. Lastly, looking at the utilities barplot, it is seen that the top 10 values range from a median of $249.88 to $304.70. The citizens of these

countries have to pay so much just for their basic home needs, whereas the country of Venezuela, has to only pay $12.37 for their basic utilities. The citizens might really need to save electricity and water in their everyday lives to save some money.

Conclusion

The findings of this report are truly eye-opening in that there are many people out there living a much more comfortable life than us, but there are also many others suffering due to high inflations and extreme cost of living. We should be more aware of the disparities in cost of living among countries to be more informed and insightful people. We cannot simply assume that the cost of living expenses follow a normal distribution since there tends to be many outliers in the data, which is then better measured using median values. Certain countries do not even have high median salaries, but they appear as countries with the highest cost of living for certain aspects. The large discrepancies in cost of living worldwide is something that governments should work together in order to bridge the gap.

Much deeper analysis can be carried out to gain further insights into the global cost of living. Many interesting conclusions can be made from many different aspects using this dataset if one intends to.

References

Hakimi, B.A. 2021. Afghans Complain of High Priced, Low Quality Internet. *TOLOnews*. https://tolonews.com/business-175098 [16 January 2023].

City of Jyväskylä. t.th. The education services of the City of Jyväskylä. https://www.jyvaskyla.fi/en/daycare-and-education/basic-education/education-services-city-jyvaskyla [16 January 2023].

Gillespie, P. 2018. The world's worst currency is shattering homebuyers' dreams. *BNN Bloomberg*. https://www.bloomberg.com/news/articles/2018-10-25/the-world-s-worst-currency-is-shattering-home-buyers-dreams?leadSource=uverify%20wall [16 January 2023].