UTM

UNIVERSITI TEKNOLOGI MALAYSIA

PROBABILITY AND STATISTICAL DATA ANALYSIS

(SEC1 2143)

PROJECT 2

(COST OF LIVING INDEX BY COUNTRY)

SECTION 8

(GROUP 4)

NAME	MATRIC ID
AMALIA SYAZANA BINTI MOHD MAZLIMIN	**A21EC0014**
FARHANAH AINA BINTI MD AYUB	**A21EC0024**
NUR DINI FATINI BINTI MOHD KAMAL	**A21EC0110**
NUR SYUHAIDA BINTI ZAWAWI	**A21EC0116**

LECTURER'S NAME : **DR. SHARIN HAZLIN BINTI HUSPI**

Table of Contents

1.0 Introduction

Most of the countries in the world went into a chaotic situation when the news of covid-19 virus are suddenly spreading all around the world. Too many changes happen during a pandemic of covid-19 and one of them is the cost of living index. Economic charts across the countries drop sharply because the countries need a large amount of funds to control this pandemic. Lockdown is also the factor for the dropped economic chart. This situation led to changes in the cost of living index in a country. Cost of living is the amount of money required to cover basic expenses such as housing, food, taxes, and healthcare in a specific location and time period. The cost of living is frequently used to compare the cost of living in one city versus another. If the expenses in that country are higher, then the cost of living is much higher compared to other countries. Our project objective is to explore the relationship between the cost of living index and the affecting factors. From the data, we expect to see that the cost of living index by country is significantly affected by its rent index, cost of living plus rent index, groceries index, restaurant index and local purchasing power index.

2.0 Dataset

Name	Variable	Data Type
Country	Name of country	Nominal
Cost of Living	Average cost of living	Ratio
Rent	Average rental rate	Ratio
Cost of Living Plus Rent	Average cost of living including rental rate	Ratio
Groceries Prices	Average price of groceries	Ratio
Restaurant Prices	Average restaurants' meal rate	Ratio
Local Purchasing Power	Relative purchasing power in buying goods and services	Ratio

Selected Variable(s)	Test	Description
Average cost of living including rental rate	Hypothesis Testing (One sample test)	**Explanation:** The variable is used to test whether the mean average cost of living including rental rate is greater than 30 at 5% of significance level **Possible Outcome:** The mean average cost of living including rental rate is greater than 30

Groceries Index, Cost of Living	Correlation analysis	**Explanation:** The variables are selected to test whether the linear relationship exists between the groceries index and the cost of living using Pearson's Product-Moment Correlation Coefficient at 95% confidence level. **Possible Outcome:** There is a strong positive linear relationship between the groceries index and the cost of living at 0.05 significance level. The higher the groceries index, the higher the cost of living.
Restaurant Prices, Groceries Prices	Regression analysis	**Explanation:** The variables are selected to test whether the restaurant meal prices depend on grocery prices, using restaurant prices as dependent variable y and groceries prices as independent variable x. **Possible Outcome:** The Restaurant Price Index depends on the Groceries Price Index. When the

		Groceries Index increases , the Restaurant Price Index Increases.
Local purchasing power index	ANOVA test	**Explanation:** A random sample is selected from the data to test the equality of local purchasing power index for all continents by analyzing the sample variances at $\alpha = 0.05$. **Possible Outcome:** The means of local purchasing power index are same for all continents at $\alpha = 0.05$.

3.0 Data Analysis

3.1 Hypothesis Testing

Based on the test, we determine whether the mean average cost of living including rental rate of 133 countries is greater than 30 per index at 5% of significance level. Since the sample size is relatively large and the variance is unknown, we assume that the sample is normally distributed. The test statistic that we applied is the Z-test.

Let μ= average cost of living including rental rate

Hypothesis statement:

H_0 : μ = 30

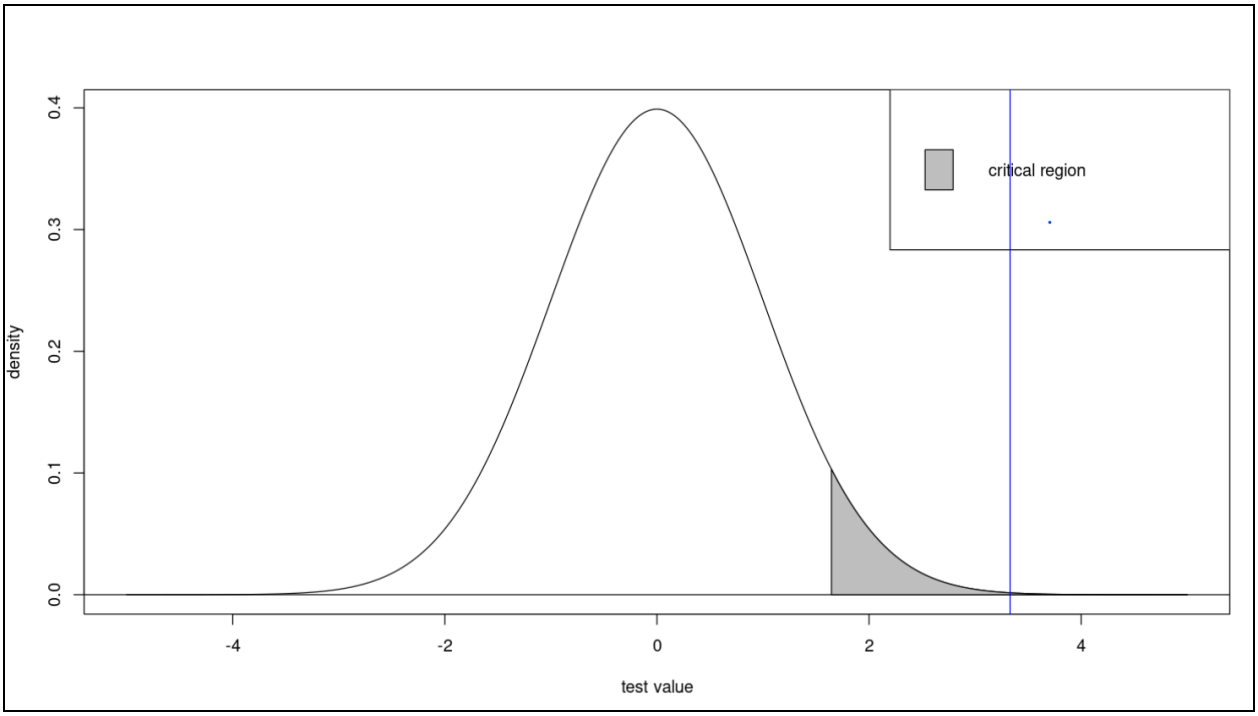
H_1 :μ > 30

Test statistic:

Significance level, α = 0.05

$Z_{0.05}$ = 1.645

alpha	0.05
mu	30
n	133
Rent	num [1:132] 87.9 70.2 74.9 55.9 58.6 ...
s	14.989051988972
xbar	34.3272727272727
z	3.32939959299635
z.alpha	1.64485362695147



Graph 1: Density of average cost of living including rental rate greater than 30 per index

Z = 3.329

According to Graph 1, the gray coloured region was labeled as the critical region, which is, $Z_{0.05}$ = 1.645 while the blue vertical line represents the test statistic value for this test. As shown in the graph, we can see that the statistical value lies within the critical region.

Since Z = 3.329 and 3.329 > 1.645, we reject the null hypothesis, H_0. There is sufficient evidence to support that mean average cost of living including rental rate is greater than 30 per index at 5% of significance level.

3.2 Correlation Test

In this correlation analysis, the variables that we used are the cost of living and the groceries index of 132 countries in the world. We will test whether there is a linear relationship between groceries index and cost of living with 0.05 significance level.

Hypothesis statement:

H_0: $\rho = 0$ (no linear correlation)

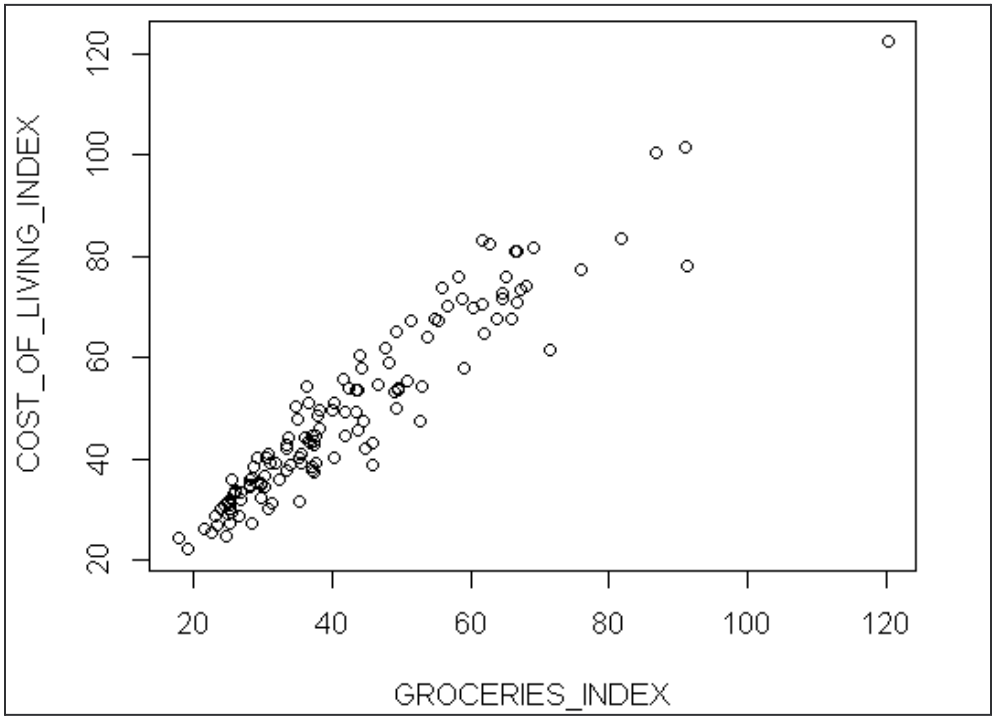
H_1: $\rho \neq 0$ (linear correlation exists)

Test statistics:

By using RStudio, we can get the result that $t = 35.219$ and $t_{0.025,130}=1.97838$. Thus, the H_0 will be rejected if $t<-1.97838$ or $t>1.97838$. Otherwise, fail to reject H_0.

Decision:

Since $t=35.219>1.97838$, we reject the null hypothesis. There is sufficient evidence to conclude that there is a linear relationship between groceries index and cost of living at 0.05 significance level.



Graph 2 : Cost of Living Index against Groceries Index

From the scatter plot graph, we can see that the points slope upward, which shows that there is a positive relationship between groceries index and cost of living, such that the higher the groceries index, the higher the cost of living.

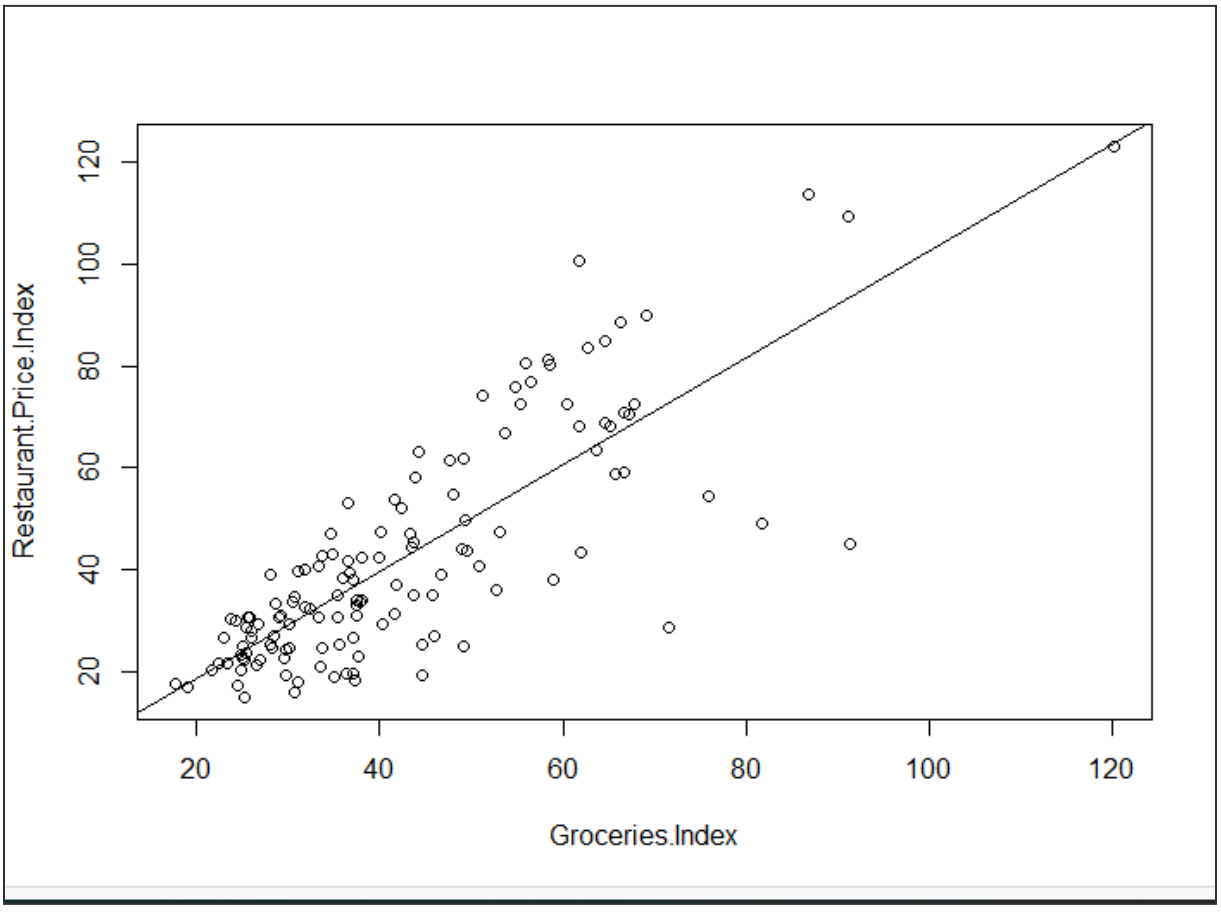
Since both variables are ratio-type data, the sample coefficient correlation is then calculated by using Pearson's technique. From the RStudio, we get $r = 0.9513872$, which indicates that there is a relatively strong positive linear relationship between groceries index and cost of living.

Through the result of the scatter plot graph and the correlation test, it shows that the groceries index is correlated with the cost of living. They have a positive linear relationship, so when the grocery index increases, the cost of living increases. It means that when a country has a higher index of groceries, the people in the country also has a higher index of cost of living, while the groceries index might be affected by the country's ability to supply products which outpaced the customers' demand during the COVID-19 pandemic and so on.

3.3 Regression Test

For this regression test, the variables we choose are the restaurant price index and the groceries price index for a random sample of 132 countries. The dependent variable(y), is the restaurant price index while the independent variable (x) is the groceries price index. This test is to determine whether the restaurant price index depends on the groceries price index.

The following is the scatter plot of the Restaurant Price Index against the Groceries Index.



Estimated Regression Equation

$$\hat{y} = -2.39712 + 1.04898x$$

Based on the plot, the line indicates that it is a positive linear relationship between the independent variable (y) and the dependent variable(x). Through analysis, we obtain the value of the intercept coefficient (b_0) is -2.39712, and the value of the estimated average Groceries Index (b_1) is 1.04898.

From the equation, we estimate that the average change in the price of the Restaurant Price Index will increase by 1.04898 for each price per index rise in the Groceries Index. From the graph, when The Groceries Index is zero, the b_0 value (-2.39712) indicates the range of the Restaurant Prices Index observed that is not explained by the Groceries Index.

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(yi - \bar{y})^2} = 0.6525$$

Based on the value calculated for the coefficient of determination R^2, which is 0.6525 represents a 65.25% increase in the average Restaurant Price Index that is some but not all of the Restaurant Price Index is explained by the Groceries Index. As a result, we can conclude that the Restaurant Price Index is moderately dependent to the Groceries Index

3.4 Anova Test

For this data analysis, ANOVA test is used to test the equality of local purchasing power index means of continents by analyzing the sample variance. The purpose of ANOVA is to test the significant differences between means. Since we assume that the populations have normal distribution and same standard deviation so the population is split into six continents and the samples are randomly selected to get an equal sample size. The significance level used to test the null hypothesis is $\alpha = 0.05$.

H_0 : $\mu1 = \mu2 = \mu3 = \mu4 = \mu5 = \mu6$

H_1 : at least one mean is different.

SAMPLE	n	$\bar{x}$	s
Asia	3	71.707	36.908
Africa	3	45.363	33.508
North America	3	58.050	35.265
South America	3	38.597	38.785
Europe	3	56.987	38.728
Oceania	3	77.713	39.265

The mean between samples, $\bar{x} = 58.070$.

The Standard deviation between samples, $S_{\bar{x}} = 14.924$

The variance between samples= 668.177

The variance within samples= 1379.080

The test statistic F = 0.485

The numerator = k-1 = 6-1 = 5

The denominator = k(n-1) = 6(3-1)= 12

F critical value = 3.11

-Since F test statistic < F critical value(0.485<3.11), we fail to reject the null hypothesis.

-There is no sufficient evidence to claim that the different continents have the same mean of local purchasing power index.

Conclusion

To conclude this report, there are various factors that affect the cost of living index for each country. Based on our analysis for the hypothesis testing (one-tailed test), the mean average cost of living including rental rate is greater than 30 per index. Besides that, we are able to obtain a positive linear relationship between the Grocery Price Index and the Cost of living by using a correlation test. Thus, when the price of groceries increases, the cost of living will also shoot up. Next, from the regression test, we discover that there is a positive linear relationship between Other than that, The Restaurant Price Index is moderately dependent on the Groceries Index. Last but not least, The ANOVA test, we claim that there is sufficient evidence to claim that different countries have the same mean of local purchasing power index.

We explore and learned how to use R programming to analyze and interpret the datasets. From the analysis and interpretation of the dataset, we are able to identify the leading factors of cost of living and how they are explained through these tests. Finally, we are able to come up with a conclusion from these analyses rather than using our own judgment to conclude. This knowledge is very useful not just for this subject but for future research, especially in the Data Science field.

With great communication and teamwork, we are able to complete the given project on time despite facing difficulties in understanding. We solved it as a team by doing discussions and making sure everyone understood their respective roles. Lastly, we would like to express our highest gratitude to our dearest lecturer, Dr. Shazlin for guiding us step by step through exercises from weekly classes. We are truly grateful that Dr. Shazlin is very patient and concerned in handling us, especially in terms of submitting the assignments and projects on time

Appendices

Raw data

Country	Cost of Living Index	Rent Index	Cost of Living Plus Rent Index	Groceries Index	Restaurant Price Index	Local Purchasing Power Index
Switzerland	122.4	50.25	87.89	120.27	123.01	119.53
Norway	101.43	36.15	70.21	91.14	109.28	88.38
Iceland	100.48	46.95	74.88	86.89	113.74	79.44
Japan	83.35	25.97	55.9	81.82	48.95	87.28
Denmark	83	31.92	58.57	61.74	100.75	100.88
Bahamas	82.51	36.36	60.43	62.65	83.66	54.18
Luxembourg	81.89	54.92	68.99	69.1	89.85	103.92
Israel	81.15	31.33	57.32	66.31	88.65	78.09
Singapore	81.1	63.27	72.58	66.75	58.99	88.96
South Korea	78.18	22.86	51.72	91.31	44.87	85.21
Hong Kong	77.22	79.57	78.35	75.94	54.36	65.32
Barbados	76.02	19.7	49.08	65.2	68.14	51.1
Ireland	75.91	43.88	60.59	58.35	81.24	80.88
France	74.14	25.39	50.82	67.9	72.54	80.36
Netherlands	73.75	35.18	55.3	55.87	80.48	90.73
Australia	73.54	34.86	55.04	67.23	70.32	107.31
New Zealand	72.53	32.09	53.19	64.69	68.8	92.66
Belgium	71.78	25.43	49.61	58.66	80.14	86.28
Seychelles	71.59	32.34	52.81	64.61	84.86	21.67
United States	71.05	40.32	56.36	66.61	70.74	109.52
Austria	70.38	26.81	49.54	61.73	68.09	82.38
Finland	70.29	26.16	49.18	56.52	76.81	99.93
Sweden	69.85	25.9	48.83	60.47	72.35	101.73
Canada	67.62	30.73	49.98	63.68	63.53	95.09
Puerto Rico	67.54	19.29	44.46	65.76	58.68	79.38
Malta	67.46	31.04	50.04	54.9	75.94	46.32
United Kingdom	67.28	29.85	49.38	51.27	74.13	91.73
Italy	67.26	21.22	45.24	55.44	72.32	65.59
Germany	65.26	27.06	46.99	49.23	61.58	102.36
Macao	64.84	43.98	54.87	62	43.37	80.9
Qatar	64.04	47.44	56.1	53.61	66.83	111.69
United Arab Emirates	61.98	41.07	51.98	47.63	61.32	91.58
Taiwan	61.37	16.42	39.87	71.51	28.49	65.67
Lebanon	60.5	24.54	43.3	43.92	58.16	44.47
Bahrain	58.94	30.95	45.55	48.11	54.55	57.88
Cyprus	57.93	20.54	40.05	44.21	63.11	57.41
Jamaica	57.82	15.25	37.46	58.96	38.12	32.37
Greece	55.67	11.68	34.62	41.63	53.74	43.68
Zimbabwe	55.3	11.02	34.12	50.8	40.79	26.03
Palestine	54.54	8.83	32.68	46.67	38.93	46.91
Ethiopia	54.39	20.9	38.37	36.3	19.37	8.34
Panama	54.16	24.76	40.09	53.03	47.27	34.23
Costa Rica	53.98	15.55	35.6	49.6	43.51	42.61
Spain	53.77	21.77	38.46	42.38	52.07	72.03

Trinidad And Tobago	53.7	17.26	36.27	49.43	49.83	45.7
Jordan	53.67	11.54	33.52	43.32	47.12	34.88
Slovenia	53.43	17.09	36.05	43.76	45.43	66.31
Mauritius	53.04	11.99	33.41	48.94	43.96	30.88
Uruguay	51.04	14.43	33.53	40.12	47.41	37.46
Estonia	50.93	15.41	33.94	36.57	52.94	71.3
Kuwait	50.37	31.21	41.21	34.68	47.08	85.59
Thailand	49.77	17.1	34.14	49.2	24.9	35.45
Croatia	49.7	13.5	32.38	39.9	42.44	50.42
Portugal	49.52	21.81	36.27	38.14	42.48	49.43
Oman	49.28	17.98	34.31	43.5	44.26	80.97
Belize	49.23	9.74	30.34	41.89	37.05	57.53
Saudi Arabia	48.34	11.39	30.67	37.89	33.51	100
Latvia	47.94	12.34	30.91	34.85	42.88	52.48
Cambodia	47.49	14.9	31.9	44.56	25.19	11.29
Fiji	47.48	18.55	33.64	52.67	35.87	33.17
Czech Republic	46.15	19.56	33.43	38.13	34.1	62.82
El Salvador	45.57	13.48	30.22	43.67	34.8	24.71
Brunei	44.71	23.08	34.36	37.11	38.06	86.95
Nicaragua	44.56	7.61	26.88	41.71	31.41	23.92
Slovakia	44.46	16.11	30.9	37.51	33.82	56.94
Lithuania	44.28	13.65	29.63	33.63	42.7	57.85
Dominican Republic	44.06	10.62	28.06	35.93	38.27	23.95
Chile	43.62	13.39	29.16	36.45	41.57	42.5
Cuba	43.46	4.84	24.99	37.4	30.95	2.18
Suriname	43.14	7.79	26.23	45.78	34.98	22.39
Namibia	43.1	16.12	30.19	36.8	39.22	55.65
South Africa	42.87	16.61	30.31	33.29	40.58	73.61
Guatemala	42.7	15.29	29.59	37.41	33	39.83
Honduras	42.17	10.6	27.07	33.33	30.55	30.26
Myanmar	42.11	23.78	33.35	44.68	19.04	10.22
Ecuador	40.98	11.96	27.1	35.46	30.66	36.08
Hungary	40.85	13.97	27.99	30.77	34.46	47.55
Brazil	40.22	10.65	26.08	29.2	30.78	32.81
Kenya	40.21	10.73	26.11	35.35	35.06	27.32
China	40.04	16.38	28.72	40.37	29.16	60.88
Poland	40.04	15.67	28.38	30.55	33.45	59.61
Russia	39.21	11.36	25.89	31.08	39.61	38.94
Botswana	39.13	11.12	25.73	31.74	39.98	58.1
Malaysia	39.12	11	25.67	37.58	22.94	64.49
Iraq	39.04	11.28	25.76	31.85	32.72	42.43
Iran	39.01	14.48	27.28	35.54	25.12	22.69
Somalia	38.68	4.75	22.45	45.88	27.04	33.24
Peru	38.65	12.78	26.28	33.8	24.69	33.95
Vietnam	38.34	13.57	26.5	37.06	19.59	28.14
Montenegro	38.23	10.38	24.91	28.63	33.16	39.8
Ghana	37.65	19.41	28.93	37.17	26.59	14.66

Philippines	37.63	9	23.94	33.46	20.84	23.48
Indonesia	37.27	10.62	24.52	37.36	18.25	25.05
Bulgaria	36.7	9.64	23.75	30.09	29.14	49.37
Albania	36.39	8.55	23.07	28.39	26.75	29.56
Bosnia And Herzegovina	35.97	6.72	21.97	28.08	25.1	41.83
Mexico	35.72	11.46	24.11	32.39	32.12	41.81
Serbia	35.72	9.08	22.98	25.46	28.63	36.72
Romania	35.31	10.05	23.22	29.03	30.5	48.86
Tanzania	35.25	14.04	25.11	29.66	22.57	17.58
Belarus	34.7	10.5	23.13	28.06	38.94	37
Turkey	34.69	6.78	21.34	29.76	24.16	40.85
Bolivia	34.56	10.93	23.26	28.09	25.11	39.34
Morocco	34.32	8.94	22.18	30.11	24.5	35.25
Moldova	33.7	8.4	21.6	25.96	27.79	27.57
Rwanda	33.35	11.88	23.08	26.83	29.17	19.32
Ukraine	33.18	10.46	22.31	26.01	26.42	31.8
Argentina	32.95	8.33	21.17	25.82	30.52	47.22
Armenia	32.84	10.91	22.35	25.61	30.43	28
Bangladesh	32.25	5.01	19.22	29.72	19.21	33.21
Zambia	31.72	7.39	20.08	26.86	22.07	54.14
Sri Lanka	31.61	7.74	20.2	35.09	18.76	24.45
North Macedonia	31.59	6.49	19.58	24.94	23.31	37.15
Paraguay	31.1	9.66	20.85	25.03	25.01	32.6
Nigeria	31	23.59	27.45	31.15	17.71	13.52
Colombia	30.66	9.58	20.58	25.05	22.44	31.12
Kazakhstan	30.64	9.78	20.66	24.31	29.87	38.36
Uganda	30.18	10.15	20.6	25.25	22.14	20.5
Algeria	30.1	5.34	18.26	30.71	15.89	28.69
Azerbaijan	29.92	9.48	20.15	23.74	30.36	30.2
Egypt	29.54	5.49	18.04	25.5	23.53	22.41
Nepal	29.05	4.77	17.44	24.91	20.36	21.9
Georgia	28.48	9.8	19.54	23.05	26.66	24.88
Kosovo (Disputed Territory)	28.47	8.47	18.91	26.53	21.04	48.9
Venezuela	27.17	6.81	17.43	28.29	24.41	2.61
Tunisia	27.04	5.51	16.74	25.25	14.95	34.45
Kyrgyzstan	26.97	9.18	18.46	23.39	21.45	24.98
Uzbekistan	26.01	7.54	17.18	21.59	20.17	24.31
Syria	25.31	5.68	15.92	22.42	21.38	11.46
India	24.58	5.68	15.54	24.55	17.17	54.3
Afghanistan	24.24	4.03	14.58	17.7	17.38	28.2
Pakistan	21.98	4.59	13.66	19.08	16.78	30.57

https://www.kaggle.com/datasets/dumbgeek/countries-dataset-2020

Process data

continent	countries	Local Purchasing Power Index
Oceania	Australia	107.31
Oceania	New Zealand	92.66
Oceania	Fiji	33.17
Asia	Russia	38.94
Asia	Malaysia	64.49
Asia	Qatar	111.69
Africa	Zambia	54.14
Africa	Ethiopia	8.34
Africa	South Africa	73.61
North America	Canada	95.09
North America	Bahamas	54.18
North America	Georgia	24.88
South America	Puerto Rico	79.38
South America	Panama	34.23
South America	Cuba	2.18
Europe	Finland	99.93
Europe	Malta	46.32
Europe	El Salvador	24.71

VIDEO LINK :**https://youtu.be/VP7Aybbt1PM**