

|  |
| --- |
| Economic Performance Index  2023 |
|  |
| May 8  Lee Xuan Ong  Student ID: D00253282 |

# Title Heading

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |

Contents

[Title Heading 2](#_Toc197635482)

[Subtitle Text Here 2](#_Toc197635483)

Executive Summary

Why Economic performance measures matter ?

As the world evolves and the economic integrates with the countries deeper, it is important that the countries need to compared with each other and knowing what was the factor that affect their economic performance. While comparison cannot be only looking at GDP we need to consider more factors that was affecting the countries like trade, education, infrastructure, environment and good governance. These factors is interrelated and affecting each other in a country. This project measures aim to measure with these factors together by creating a Composite Economic Performance Index.

Why a composite Index?

Economic performance of countries is not depending on a single factor of GDP and so composite index is needed for visualization to have a better understanding of how it affect a countries. The indexes that are used included macroeconomic, human capital, infrastructure and technology,

How was the study conducted ?

In this study I collected the data by using different datasets that is published by world Bank, DataBank, ASEANstats and Transparency International. All the data will be merged together by using the country following the year 2023 columns. With Appropriate and logical statistic techniques helps in imputing the missing value. Merged data will be normalized and standardized while preparing for the multivariate analysis. The method used for multivariate analysis is Principal Component Analysis(PCA) and cluster analysis. Adjusting the weight of the data will occurred based of the multivariate analysis. Hence, Composite Index PC1 and Composite Index Weight PC1 is made. The difference between these two index values are one is made entirely using a fair weight given to each of the sub index values, another is weighted with differently on each sub index value. From there on, each country will have their own composite index values and will be able to rank them based on it.

Key findings

By using the composite index that I found after going through the process of weighting and aggregation. Surprisingly I found that the country that ranked Bhutan has the highest composite index. Once I dig deeper into it I realized that the sub index of Environmental and sustainability has the highest value as and for the other sub index is medium while comparing to other countries.

The lowest value composite index that I found out is Afghanistan. It have the lowest value as low as 0.10. The reason of having this low is because that it has low sub index in trade and investment, institutions and political and macroeconomic. These are some factors that is affecting this country getting a very low composite index value.

Other than that, countries that have a balanced scores across all the sub index or indicators tended to perform well overall. So even if they didn’t have any top scores in particular categories, it will tend to perform well.

In conclusion, these result showed that how useful it is for using a composite index to measure and identifying the factor and the performance of the country.

Results

|  |  |  |
| --- | --- | --- |
| Rank | Country Name | Composite Index Value |
| 1 | Bhutan | 0.87 |
| 2 | Fiji | 0.87 |
| 3 | Samoa | 0.86 |
| 4 | Macau | 0.84 |
| 5 | Papua New Guinea | 0.83 |
| 6 | Solomon Islands | 0.83 |
| 7 | Nepal | 0.81 |
| 8 | Marshall Islands | 0.80 |
| 9 | Kiribati | 0.79 |
| 10 | Cambodia | 0.79 |
| 11 | Brunei | 0.78 |
| 12 | Laos | 0.77 |
| 13 | East Timor | 0.77 |
| 14 | Nauru | 0.77 |
| 15 | Singapore | 0.76 |
| 16 | Vanuatu | 0.75 |
| 17 | Mongolia | 0.74 |
| 18 | India | 0.73 |
| 19 | Tonga | 0.73 |
| 20 | Hong Kong | 0.72 |
| 21 | Vietnam | 0.72 |
| 22 | Sri Lanka | 0.72 |
| 23 | Indonesia | 0.69 |
| 24 | Malaysia | 0.69 |
| 25 | Maldives | 0.68 |
| 26 | New Zealand | 0.68 |
| 27 | Philippines | 0.67 |
| 28 | Japan | 0.65 |
| 29 | Pakistan | 0.64 |
| 30 | Thailand | 0.64 |
| 31 | South Korea | 0.63 |
| 32 | China | 0.62 |
| 33 | Bangladesh | 0.62 |
| 34 | Australia | 0.62 |
| 35 | Palau | 0.61 |
| 36 | Myanmar | 0.57 |
| 37 | Afghanistan | 0.10 |

A map of asia with different colors of continents

AI-generated content may be incorrect.

This is my Geographical Composite index that was shown on the asia map and pacific map. The colour with more towards yellow it having more higher composite index. The colour toward more darker blue is having lower. As you can see south east asia area is more likely to be in the area of medium level composit index around 0.6+-. While looking at Afghanistan nearing middle east, we noticed that it started to have a low composite index. The economic performance in the Asia is performing well and moderate.

# Theoretical Framework

Introduction

The purpose of this theoretical framework is to create a composite index which measures the Economic Performance . This is a multi-dimensional phenomenon where it cannot simple analyzed nor measured with a single variable.

This index is intended to provide a measurement tool for comparing Asia and Pacific countries based on all the indicators or variables. It enabled us to identify the strengths and weaknesses of each country and also found out that what factor is affecting the country for their performance.

Structure of Sub Groups

To comprehensively capture the complexity of Economic Performance, the composite index is broken down into multiple group representing the sub index from different specific dimensions within the data sets:

* Macroeconomy
  + This measures the country of GDP per capital, growth rate, final consumption expenditure, government debt and inflation. It explain the economic outputs of each country.
* Trade and Investment
  + This measures the country of import and export of goods and services and the current account balance that was public. This will help in exploring more external drives for GDP.
* Human Capital
  + This measures the labor force, birth rate expectancy, school enrollment and unemployment. This help in explaining the quality of life of the people of the country.
* Infrastructure and Technology
  + This measures the access to electricity and internet, people that have subscription to the mobile cellular and the logistic score. This explain the infrastructure and technology of the country have and how good it is.
* Institution and Political
  + This measures the political stability, regulatory quality and the corruption score. This explain the safety and corruption of the countries.
* Environmental and sustainability
  + This explain the carbon dioxide that emits and how much they have utilizing the renewable energy.

These sub-groups will be normalized, weighted, and aggregated into a single composite score for each country.

Identifying the selection criteria

Input phase

In the input phase it include:

1. Data selection – finding the data to represent my indicators
2. Imputing Missing value- filling up or dropping the missing value for the row of my data set

This is where it will be getting the initiated or raw data.

Processing Phase

To prepare the data for multivariate analysis, the datasets is required to be standardized with using z-score normalisation for Principal components analysis and min/max normalisation for clustering analysis. Once these clustering is done, now the columns of data will be decided is gonna be weighted less or dropped.

Output Phase

After Weighting and aggregation occurred, this is where the composite index is output. Now based on the composite index, the countries can be ranked and now available to know which perform the best and which perform the worst in this index.

Data Selection

In the process of selecting data selection, I gone through some website that published the data free online like World Bank, ASEANStats or Transparencry International. This is some of the website I gone through and trying to find the dataset that I want for my subgroup. Below table is the data that I found and will be the indicators:

|  |  |  |
| --- | --- | --- |
| Category | Indicator | Source |
| Macroeconomic | GDP Growth Rate | [Link](https://databank.worldbank.org/asean-gdp-gr-annual-%28%29/id/46303a24) |
| GDP Per Capita | [Link](https://data.worldbank.org/indicator/NY.GDP.PCAP.CD) |
| Final Comsumption Expenditure | [Link](https://data.worldbank.org/indicator/NE.CON.TOTL.ZS) |
| Government Debt | [Link](https://www.imf.org/external/datamapper/CG_DEBT_GDP@GDD/SWE) |
| Inflation Rate | [Link](https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=Z4) |
| Trade and Investment | Current account balance | [Link](https://data.worldbank.org/indicator/BN.CAB.XOKA.GD.ZS?locations=Z4) |
| Imports goods and services | [Link](https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS) |
| Export goods and services | [Link](https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS) |
| Human Capital | Labour force | [Link](https://kidb.adb.org/explore?filter%5Bindicator_id%5D=1100028&filter%5Beconomy_code%5D=AFG%2CARM%2CAUS%2CAZE%2CBAN%2CBHU%2CBRU%2CCAM%2CCOO%2CFIJ%2CFSM%2CGEO%2CHKG%2CIND%2CINO%2CJPN%2CKAZ%2CKGZ%2CKIR%2CKOR%2CLAO%2CMAL%2CMLD%2CMON%2CMYA%2CNAU%2CNEP%2CNIU%2CNZL%2CPAK%2CPHI%2CPLW%2CPNG%2CPRC%2CRMI%2CSAM%2CSIN%2CSOL%2CSRI%2CTAJ%2CTAP%2CTHA%2CTIM%2CTKM%2CTON%2CTUV%2CUZB%2CVAN%2CVIE&filter%5Byear%5D=2000%2C2001%2C2002%2C2003%2C2004%2C2005%2C2006%2C2007%2C2008%2C2009%2C2010%2C2011%2C2012%2C2013%2C2014%2C2015%2C2016%2C2017%2C2018%2C2019%2C2020%2C2021%2C2022%2C2023%2C2024%2C2025&grouping=indicators&showRegions=1) |
| Life Expenctancy at birth | [Link](https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=Z4) |
| School encrollment | [Link](https://data.worldbank.org/indicator/SE.SEC.ENRR) |
| Unemployment rate | [Link](https://www.imf.org/external/datamapper/LUR@WEO/VNM/THA/SGP/PHL/MYS/IDN) |
| Infrastructure and Technology | Access to Electricity | [Link](https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS) |
| Access to Internet | [Link](https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS) |
| Mobile Subscription | [Link](https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS) |
| Logistics performance index | [Link](https://lpi.worldbank.org/) |
| Institution and Political | Corruption Perception Index | [Link](https://www.transparency.org/en/cpi/2023) |
| Political stability | [Link](https://data.worldbank.org/indicator/PV.PER.RNK) |
| Regulatory Quality | [Link](https://www.worldbank.org/en/publication/worldwide-governance-indicators) |
| Environmental and Sustainability | CO2 emssion per capita | [Link](https://data.worldbank.org/indicator/EN.GHG.CO2.PC.CE.AR5) |
| CO2 emssion total | [Link](https://data.worldbank.org/indicator/EN.GHG.CO2.RT.GDP.PP.KD) |
| Renewable energy consumption | [Link](https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS) |

I have a total of 6 category for my sub group or subindices most of it have a minimum of at least 3 indicators. Now all the indicators will merged together into one data set by using country name and getting the 2023 data inside. For those data that have no data for 2023 then I will try getting the nearest value like 2022.

In this process there several things to consider too, for example missing data does not represent an empty cell in a csv it can be a “..”, “…” and “no data”. These is all the representations of no data in the data sets. So what I did is by tracing these value and changing the data to null value to represent a standard of missing value.

Other than that, country name is not consistent too. Hence, I standardised the country name so that it can be joined together in the dataframe and will not have duplication.

There is a scenario that the data sets having duplication of data in the same year. The reason behind this data is because it repeated for 4 times a year where it represent the data quarterly so what I did is getting the mean of these data and combine it into 1 row.

Imputing of missing data

Even though during the process of data selection, getting the nearest year value is done. There are still a lot of cases where the there are no data for the entire row. Hence what we can do is using mean and median. To determine mean and median for filling up the missing values. Indicators will be checked with using skew to see if have outliers. If there are outliers, then median will be used to fill the missing data else using mean / average value for it.

However, before we start filling up the data we will be checking the data if there is any row that is have more than 5 columns of data in the same row is missing. If there is then we delete specific row. This is because I want to maintain the data set quality and reliability. If there are too many data is missing in the same row and impute it, it could bias the data set and weakening the accuracy of composite index

Multivariate Analysis

Before starting into the PCA analysis and Clustering analysis, It required my data to be standardized (z score normalized) for PCA analysis and min max normalization for my clustering analysis. There are a total of 4 steps of doing the multivariate analysis:

1. Explore the correlation of each indicators within the category
2. Explore Scatter matrix to find the relationship between the indicators
3. Doing principal component analysis
4. Clustering analysis

Sub Index 1: Macroeconomic

A screenshot of a graph

AI-generated content may be incorrect.

Insight

This is the Macroeconomic category indicators correlation matrix. The highest correlation is 0.43 which is GDP growth rate indicator and GDP per capita indicator. This proved that countries higher-income tend to experience higher economic growth. The next highly correlated on Governement debt with GDP per capita (0.25) and inflation rate (0.23). This indicates that if government tend to have higher debt then the prices rise and income rise too.

Sub Index 2: Human Capital

A screenshot of a graph

AI-generated content may be incorrect.

In this Human Captial group the overall is mostly medium correlation. While i realised that the school\_enrollment\_rate and unemployment\_rate are having the lowest correlation of 0.07. School enrollemnt rate and life expectancy at birth have a high correlation that represents if there are more life being birth then there will be more students in the school. Another highly correlated on labour force participation with school enrolment (0.36) and life expectancy at birth( 0.32) this indicates that if there are more labour force, more students and life birth will increase.

Sub Index 3: Infrastructure and Technology

A screenshot of a graph

AI-generated content may be incorrect.

In the heat map majority of the indicator is positively correlated this indicated that if one of the indicator value increase, it will affect the other indicators too. The highest correlated indicators are Logistics performance index(LPI) with access to internet, this suggested that if the value increase on number of people access the internet the better the LPI score it will be. Other than that, it will be number of people access to internet and electricity is the next highly corelated 0.55, this proven that if the number of people have access to electricity increase it will increase the number of people have access to internet too.

Sub Index 4: Trade and Investment

A screenshot of a graph

AI-generated content may be incorrect.

In this heatmap correaltion matrix, we can see the exports and imports goods and services have high corelation in it. Hence when there are more exports, imports will increase too. This proven that if a country have high exports in goods and services, it will have high imports too.

Sub Index 5: Institution and Political

A screenshot of a graph

AI-generated content may be incorrect.

In this correlation matrix heatmap we can see that the regulatory quality have a high correlation with the Corruption perception index, which indicates having a high regulatory quality tends to have high corruption in the government based on the data.

Sub Index 6: Environmental and Sustainability

A graph of a graph with numbers and a number of red squares

AI-generated content may be incorrect.

In this correlation matrix most of the indicators are weakly correlated. There is a moderate negative correlated indicators which are co2 emission per capita with renewable energy consumption. This indicates that if the more the co2 emits in a capital the less the renewable energy is used.

Principal Component Analysis (PCA)

PCA is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data preprocessing. Thie data is linearly transformed onto a new coordinate system such that the directions capturing the largest variation in the data can be easily identified.

PCA was applied to standardized indicators grouped under six sub-indices. The contribution (loading) of each indicator to PC1 and PC2 was used to assess its importance in explaining variance. Indicators with low contribution to PC1 and strongly contribution in PC2 will be removed or weighted lower in the composite index. Here is category that I have use their standardized values to work on PCA:

Sub Group 1: Macroeconomy

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| gdp\_per\_capita\_2023 | 0.6586 | 0.1487 |
| gdp\_growth\_rate\_2023 | 0.5963 | 0.1569 |
| inflation\_rate\_2023 | 0.4521 | 0.5066 |
| government\_debt\_2023 | 0.0754 | 0.6668 |
| final\_consumption\_expenditure\_2023 | 0.0252 | 0.502 |

Interpretation

Government debt and final consumption expenditure is having a very low contribution to PC1 and higher in PC2. Hence, it indicates that they do not drive the primary driver here. These indicators will be dropped.

Sub Group 2: Human Capital

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| life\_expectancy\_at\_birth\_2023 | 0.5788 | 0.1292 |
| labour\_force\_participation\_rate\_2023 | 0.5455 | 0.2089 |
| school\_enrollment\_rate\_2023 | 0.5221 | 0.5472 |
| unemployment\_rate\_2023 | 0.3081 | 0.8001 |

Unemployment rate has a low contribution in PC1 and is strong in PC2, displayed that it does not contribute to the primary driver in human capital. It will be dropped.

Sub Group 3: Infrastructure and Technology

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| access\_to\_internet\_2023 | 0.6228 | 0.2905 |
| lpi\_score\_2023 | 0.5506 | 0.1424 |
| access\_to\_electricity\_2023 | 0.5202 | 0.1541 |
| mobile\_subscriptions\_2023 | 0.1957 | 0.9336 |

Mobile subscription contribute strongly in PC2 but low in the primary driver (PC1). Hence, it is less aligned with the core dimension of infrastructure and technology performance. It will retained but weighted lower.

Sub Group 4: Trade and Investment

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| exports\_2023 | 0.6831 | 0.0603 |
| imports\_2023 | 0.634 | 0.4457 |
| current\_account\_balance\_2023 | 0.3625 | 0.8932 |

Current account balance will be removed as it has strong contribution only on the PC2 and not on the primary trade and investment performance dimension. Hence, it will be removed.

Sub Group 5: Institution and Political

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| cpi\_score\_2023 | 0.6204 | 0.1466 |
| regulatory\_quality\_2023 | 0.5785 | 0.5778 |
| political\_stability\_2023 | 0.5296 | 0.8029 |

All these indicators are meaningful contributions to PC1 equally, especially CPI score have a slightly higher contribution.

Sub Group 6: Environmental and Sustainability

|  |  |  |
| --- | --- | --- |
| Indicator | PC1 Contribution | PC2 Contribution |
| co2\_emissions\_per\_capita\_2023 | 0.6899 | 0.2596 |
| renewable\_energy\_consumption\_2023 | 0.6829 | 0.0766 |
| co2\_emissions\_2023 | 0.2404 | 0.9627 |

Total CO2 emissions have low contribution in PC 1 but dominate in PC2. Since it does not strongly represent the main sustainability performance dimension, it was removed from the final composite.

Clustering Analysis

Sub Group 1: Macroeconomic

A graph with a line

AI-generated content may be incorrect.A graph of a number of numbers

AI-generated content may be incorrect.

In the elbow method the SSE stops decreasing significantly on the “elbow point” when the k which is the cluster increased. The elbow appear when cluster increase to 3. As you can see there are a total of 4 cluster in the PCA clustering. Cluster 2 is tightly packed points near the origins possibly the average or central group. Cluster 0 ,1 and 3 are spread out, indicating more unique or could be outlier groups. PCA helps reduce dimensionality for visualization, but the clustering is done in the full feature space.

Sub Group 2: Human Capital

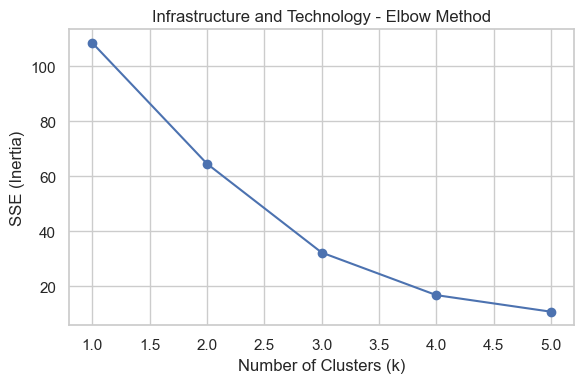
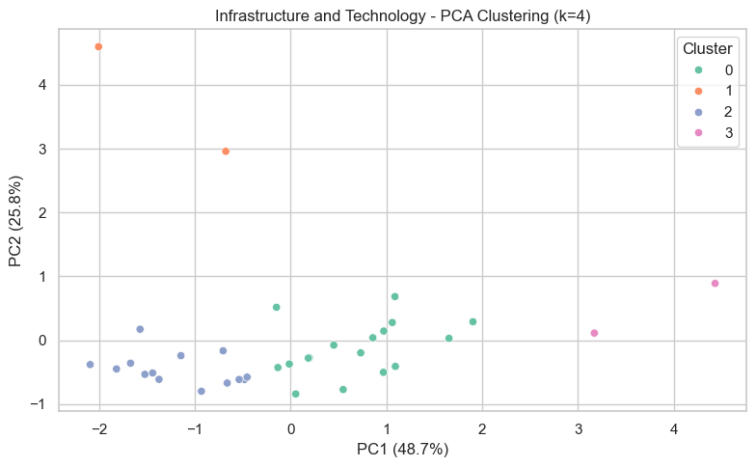
A graph with a line

AI-generated content may be incorrect. A graph with colored dots

AI-generated content may be incorrect.

In the elbow method is shows the same again where k = 3 the elbow point appeared and the SSE stop decreased significantly starting from there. For PCA clustering we can see that there are a total of 3 groups of cluster. Cluster 0 and cluster 1 are well spread along PC1 and PC2 at the left and right. This present us that the distinc groups with different human capital profiles. Where cluster 2 has only a few point on the top of the plot and quite far from each other, I would say this is an outliers. This is a balanced distribution Cluster pot and meaningful.

Sub Group 3: Infrastructure and Technology

Is the same with previous elbow method that the SSE stop decreasing after the cluster reached 3 clusters. There are still some drops when cluster increased up to 4 clusters. So choosing k = 4 is justifiable. For the PCA clustering cluster 0 is located at the middle and likely to be the average group. Cluster 2 is well populated but located near the left. Cluster 1 and 3 only have a few points and is far from the average group. Hence this could highly be a outliers. In conclusion this PCA cluster reveals a core group of countries which are cluster 0 and 2, and a few outliers which are clusters 1 and 3. The outliers could be a very advanced economies or very underdeveloped with no technology developed.

Sub Group 4: Trade and Investment

A graph with a line

AI-generated content may be incorrect.A graph with green and blue dots

AI-generated content may be incorrect.

The elbow point is at k = 3 where it stop decreasing significantly. Hence suggested cluster is 3 clusters for the PCA cluster plot. In the PCA Cluster plot, cluster 0 is packed at the central which is most likely the average group. While cluster 2 is above cluster 0 and near the center. However cluster 1 is far from the majority clusters and is most likely to be outliers.

Sub Group 5: Institution and Political

A graph with a line

AI-generated content may be incorrect. A graph with colored dots

AI-generated content may be incorrect.

In this elbow method, the elbow point is k = 4 where is suggested to have 4 clusters for the PCA clustering plot. In this plot we can see that cluster 0 is packed in the middle which can be the mid-range group, while cluster 1 is packed at the left and represent weaker on it. Cluster 2 is packed at the right-shifted indicates strong institutions or highly developed political environments, possibly advanced democracies or well-regulated economies. Lastly cluster 3 is packed at the top center and represent moderate institutional quality with relatively higher stability.

Sub Group 6 : Environment and Sustainability

A graph with a line

AI-generated content may be incorrect.A graph with red dots

AI-generated content may be incorrect.

In this last clustering analysis, the elbow point is k = 3 where the suggested cluster is 3. In the PCA clustering plot, astonishing cluster 1 dominates the entire data forming a tight horizontal linear near PCA2 = 0. This is the main group that the trending of environmental and sustainability. Cluster 0 and 2 is most likely to be a outlier where one of the indicators is extremely good or bad.

Multivariate Analysis Conclusion