# How can machine learning identify the potential of digital trade facilitation in bridging inequality?

Sergio Martinez Cotto
*PhD Candidate at University of St. Gallen*

**Abstract**

*This paper is the result of an exploratory application of machine learning (ML) methods for identifying the potential that digital trade facilitation can have in bridging inequality gaps between and within countries. Most well-known country classifications guiding country-level technical assistance are based on income level, socio-economic development dimensions, and geographical location of countries. While policymakers and researchers have approached these indicators as a valid criterion to characterize structural patterns that differentiate countries among themselves, more objective criteria could play a potential role in bridging effectiveness gaps in allocating technical assistance efforts among countries. ML methods, such as clustering methods, could enable adopting a more objective approach for classifying countries according to desired strategic objectives, such as leveraging digital trade facilitation for reducing between- and within-country inequality. This paper then has two objectives. First, it aims to contribute to existent country classification criteria by identifying a set of variables to cluster countries according to their levels of digital trade facilitation, inequality, and other institutional, social, and economic factors. Second, it intends to explore some possible policy implications for countries from Asia and the Pacific region. Section 1 introduces the ML analysis used and states the underlying motivation. Section 2 explains the data sources and variables used, as well as the process followed to clean and explore the data. Section 3 and 4 present two ML clustering methods applied for this paper's analysis and their corresponding results. Section 5 concludes and discusses areas for future improvements. Section 6 includes an appendix with data visualizations that resulted from the analysis presented in this paper.*

## 1. Introduction and Analytical Motivation

This report is the result of an exploratory application of clustering methods to gain insights on how best to target digital trade facilitation for bridging inequality gaps between and within countries. While doing so, it aims to raise a discussion on where countries in the Asia and Pacific region stand in their levels of digital trade facilitation and inequality because of rising trends observed in the latter over the past 20 years [UNESCAP, 2020]. According to the Organization of Economic Cooperation and Development (OECD), the Asia and Pacific region has seen a gap in the income distribution between the richest and the poorest 10% of the population that is twice as large as OECD member countries [OECD, 2022]. Studies have pointed out the concerning levels of income and wealth inequality in the Asia and Pacific region because of their prevalence despite the region's rapid growth and poverty reduction in recent decades [Zhuang, 2022].

As of now, most well-known country classifications guiding country-level technical assistance are based on income level, socio-economic development dimensions, and geographical location of countries. While these indicators have been approached by policymakers and researchers as a valid criterion to characterize structural patters that differentiate countries among themselves, more objective criteria could play a potential role in bridging effectiveness gaps in allocating technical assistance efforts among countries. Thus, clustering methods using machine learning (ML) algorithms could serve as an objective approach for classifying countries according to desired strategic objectives, such as leveraging digital trade facilitation for reducing between- and within-country inequality.

Using ML to classify countries objectively is helpful to adopt a structured approach in analysing multiple patterns that can help describe an optimal number of country groups. In particular, clustering methods that have been developed through machine learning algorithms follow a systematic sequence for generating valuable insights in characterizing country groups that could best fit a given country sample. Omran and Engelbrecht (2007) have documented applications of clustering methods for research purposes that frame the analysis of this paper to an extent [Omran, Engelbrecht and Salman, 2007].

As the linkage between digital trade policies and inequality has not been addressed sufficiently in academic and policy circles, using clustering methods can generate valuable insights on two fronts. First is to understand characteristics of country groups that can provide directions on how the linkage between digital trade policies and inequalities has taken place. Second is to provide a more objective approach for targeting digital trade facilitation efforts to bridging inequality gaps between and within countries.

As dealing with different types of data may imply particular challenges, so it is the case with country-level data. A particular challenge that often arises is the presence of missing values for some points in time because of limited or delayed data collection in some countries. This issue then makes difficult to use up-to-date data and raises the necessity to use statistical means to deal with missing values and other data inconsistencies.

## 2. Literature review

### Relationship between digital trade and inequality

Studies addressing the effects of digital trade in inequality are limited in the literature. One reason to expect this is that a universal definition of digital trade has not been recognized nor accepted [Lopez Gonzélez and Jouanjean, 2017]. As the OECD Working Party on International Trade in Goods and Trade in Service Statistics has pointed out, the complexity of what could be referred as "digital trade' lies in the little empirical and internationally comparable information that is existent to depict a full understanding of the scale and policy implications of digital trade [OECD, 2017]. However, three international organizations, namely the Organisation for Economic Co-operation and Development (OECD), the World Trade Organization (WTO) and the International Monetary Fund (IMF), have identified the degree of convergence on definitions referring to digital trade as, "all trade that is digitally ordered and/or digitally delivered [OECD, WTO and IMF, 2020; Mourougane, 2021; UNCTAD, 2022]."

A recurrent study in the literature is the one conducted by Zhu, et al (2022) that focused on the income inequality effects of digital deliverable services trade. The authors used part of the digital trade definition adopted by the OECD, WTO and IMF and found that digital deliverable services trade has a negative effect on income inequality in a panel of 100 countries with data available from 2005 to 2019 [Zhu, et al., 2022]. These results then suggested that digital trade contributes to bridge inequality gaps, although the results resulted statistically significant in high-income and middle-income countries. Other studies, however, have shown contrasting effects of dimensions of digital trade in inequality. Yin and Choi (2022) studied the income inequality effects of the interaction of digitalization with trade openness in the Group of Twenty (G20) countries over the 2002 – 2018 period. The authors found heterogeneous effects by income level. Their results showed that higher income inequality was associated to an increase of the interaction between digitalization and trade openness in high-income countries, but the opposite effects resulted in middle-income countries [Yin and Choi, 2022].

Because of contrasting results on the digital trade impacts on income inequality that have been documented in the literature, this paper aims to contribute by identifying clusters of countries with varying levels of income inequality and measures facilitating digital trade.

## 3. Data Collection and Preparation

### Description of data sources and targeted variables

The data that was used for this report consisted of the World Development Indicators Database developed by the World Bank Group [World Bank, 2023]. This database is known for providing an open access to databases that have been cleaning for analysis using machine learning and data science software, such as Python and R.

To identify the relevant variables for clustering countries according to their levels of digital trade facilitation, between- and within-country inequality, and other factors, two steps were involved. First, a group of variables from the World Development Indicators Database was selected according to their association to the following group variables:

a) between-country inequality;
b) within-country inequality;
c) social assistance;
d) economic resources generation;
e) demographic, social and economic country-level characteristics.

These group variables were defined subjectively but their consideration was provided due to their contextual relevance to understand the factors observed alongside the potential linkage of digital trade facilitation to inequality outcomes. Second, a digital trade facilitation index – an aggregate indicator of several digital trade facilitation measures adopted by countries that is reported bi-annually by the United Nations Global Survey on Digital and Sustainable Trade Facilitation – was matched to the aforementioned variables and classified in a group variable called "digital trade facilitation" [United Nations, 2021]. By concluding these two steps, a perspective on the contextual relevance for the chosen variable groups could be explain. Understanding varying levels of digital trade facilitation and between- and within-country inequality across country groups can be supported by observing how these groups differ in terms of their means for generating economic resources – such as foreign direct investment, taxes and trade), social assistance programs, and country-level characteristics in economic, social and demographic dimensions.

Panel 1 and 4 in the Appendix Section of this report include a list of all variables used for the ML task, with such list providing descriptions per variables and the variable groups where each variable was assigned.

### Data Exploration and cleaning

Once the relevant variables were identified and selected, they were merged and organized in a csv file that was later imported in a R studio file. The targeted variables for this ML task are numeric as they mostly refer to values expressed in monetary terms, percentage, or index scores. Values for these variables represent the average values that each country observation obtained during the 2010 – 2019 period. For the digital trade facilitation variable, 2019 values were used as data associated with this variable started to be collected in 2016. When describing this dataset, it was observed that all numeric variables had at least one missing value. Missing values were then dealt with imputation method which matched average values of country groups specified in terms of income level and geographical location. In other words, the average values of countries in a given region and income group were matched to countries missing any values across the variables included in the dataset. When imputation was not possible due to very low observations in country groups defined in terms of income level and geographical location, imputation was done in terms of country groups by income level. After the imputation process concluded, missing values were not longer present. The output of this process can be observed in the table included in Panel 1 in the Appendix section. This table provides summary statistics for all variables in the dataset and details the minimum, mean and maximum values of each variable, as well as the standard deviation and percentiles 25th and 75th – the latter two are equivalent to the first and third quartiles.

## 4. Data Analysis with K-means clustering

### Preparatory data transformation to train the K-means clustering model

The next step after dealing with missing values and cleaning the dataset is to transform the relevant data in a scale that can enable to train ML clustering models. As the dataset included variables with different measurement units, normalizing them in z values with mean equal to zero was chosen as the data transformation path. To do so in R, a data frame with the numeric variables only was first created, followed by a transformation of all variables in such variables in a scale of z values. The transformed data frame is then used for the k-means and hierarchical clustering methods that will be further explored in the following sections.

### Evaluating the optimal choice for clusters' size with k-means

Prior to running the k-means algorithm to obtain the results for a given number of clusters, a number of methods were considered to determine an optimal cluster' size. Three methods were considered: The Gap statistic, average silhouette and the elbow methods. The results for these methods were included under panel 2(a) in the appendix section of this report.

The Gap Statistic Method showed that 10 clusters (k = 10) was the optimal clusters' size. However, the Elbow and Average Silhouette methods obtained results that may not visually lead to conclude the optimal clusters' size is ten. Rather, the Average Silhouette Method may suggest that four clusters (k = 4) may be the optimal choice for the clusters' size as the average silhouette width is decreasing after this point according to the corresponding plot for this method. The Elbow Method may suggest a similar result as the total within-clusters sum of squares diminishes at a decreasing rate after the clusters' size is equal to four (k = 4).

As the resulting plot from applying the Gap Statistic Method visually suggests that 10 clusters is the optimal clusters' size, k = 10 was considered as the first choice. The second choice was four clusters (k = 4). These results are visually represented in Panel 2(b) in the Appendix Section. They group countries according to the cluster sizes chosen.

### Results analysis and discussion

Next, the first and second choices of cluster sizes – which resulted in k = 10 and k = 4 respectively – were added to the original dataset, so countries can now be classified by cluster in each cluster size grouping. To simplify the reporting of these results, panel 4 and panel 5(a) only document the output of classifying countries per each of the 10 clusters selected as the first choice of cluster size under the k-means method explained in the previous section. The results are described as follows.

To analyse the characteristics of each cluster in terms of the variable groups selected for this ML task, Panel 4 calculates the mean value of all variables for the ten cluster (k = 10) obtained under k-means. As explained under Section 2(a) on data sources and targeted variables, variables included in this dataset were classified under country groups to characterize each cluster in a more aggregated manner. To facilitate the interpretation of results per cluster, panel 4 applies a colour scale to differentiate among the higher and lower values per variable – the scale is applied horizontally. Red colours alike refer to the lowest values. Orange colours alike indicate lower-middle values. Yellow colours alike denote higher-middle values. And green colours point to higher values. This equivalence of variable values to scale colours applies for all variable groups except for the variables under the within-country inequality group. The latter follow a rather inverse equivalence which indicates that red-to-green colours correspond to high-to-low levels. Such inverse equivalence are more useful to explain the direction of values for within-country inequality variables. For instance, a high value for poverty rates is red because it is more concerning and indicates higher within-country inequality to a degree.

As the objective of this ML task is to gain insight on the nexus between digital trade policies and inequality gaps, the clusters represented in Panel 4 contribute to such objective by identifying how countries differ in terms of their levels of digital trade facilitation, between- and within-country inequality, economic resource generation, social assistance, and country-level characteristics in demographic, economic and social dimensions. A starting point with the interpretation of results is by looking at the levels of digital trade facilitation per cluster. Cluster 1 and 10 are the ones with countries that have adopted more facilitative measures for digital trade. At the same time, these clusters also have lower levels of within-country inequality relatively to other clusters. Looking at the results of these clusters across the other variable groups, it can be observed that such clusters have relatively higher levels of social assistance programs, electricity access and gross domestic product (GDP) both in aggregate and per capita (per individual in the country population) terms. However, these countries differ in terms of income distribution by quintile, means for generating economic resources and other country-level characteristics.

While not all country clusters with higher levels of digital trade facilitation may not be the ones necessarily having the lowest levels of inequality between and within country, clusters 1 and 10 could serve as potential country groups subject of study to identify enabling conditions that can leverage the facilitation of digital trade for bridging persistent inequality gaps. This is an insightful result obtained in accordance to the motivation that guided the present ML task of applying clustering methods.

As of now, results obtained with the k-means clustering algorithm were analysed and discussed. The next section presents an extension of this clusters' analysis by applying hierarchical clustering methods. As it could be expected, results from hierarchical clustering methods will indicate a different optimal clusters' size which can provide additional insights and complimentary follow-up actions regarding the optimal cluster size obtained with the k-means method.

## 5. Data Analysis with hierarchical clustering

### Preparatory data transformation to train the hierarchical clustering model

The data frame with all variables expressed in terms of z values with mean equal to zero was also used for the hierarchical clustering methods applied for the ML task documented in this report. Further calculations were conducted on these variables to apply two hierarchical clustering functions: the agglomerative and divisive hierarchical clustering. For agglomerative hierarchical clustering, a Euclidean method was applied to calculate a dissimilarity matrix. This then use a complete linkage and was further compared with data transformation associated to other hierarchical clustering methods. Average, single, complete and ward methods were used for assessing the hierarchical clustering method with the strongest clustering structure. After computing the coefficients for each method, we obtained the "Ward" method was the strongest with a coefficient of 0.9472443. With regard to divisive hierarchical clustering, the R function "diana" was applied and obtained a clustering structure indicated by the following coefficient: 0.8431869. As the function "diana" found a weaker clustering structure for the divisive hierarchical clustering, the "Ward" clustering structure analyzed under the agglomerative hierarchical clustering methods was chosen. Next sections will compute the performance measure and results analysis based on the "Ward" method for agglomerative hierarchical clustering.

### Evaluating the optimal choice for K clusters

Determining the optimal clusters from applying the "Ward" function of agglomerative hierarchical clustering used the similar methods that lead to determine optimal clusters for the k-means clustering method discussed in the previous section. These methods are as follows: The Gap statistic, average silhouette and the elbow methods. The results for these methods were included under panel 3(a) in the appendix section of this report.

The Gap Statistic Method showed that 3 clusters (k = 3) was the optimal clusters' size. However, the Elbow and Average Silhouette methods obtained results that may not visually lead to conclude the optimal clusters' size is three. Rather, the Average Silhouette Method may suggest that four clusters (k = 4) or even six clusters (k = 6) may be the optimal choices for the clusters' size as the average silhouette width is decreasing after these point according to the corresponding plot. The Elbow Method may suggest the optimal choice could be four clusters (k = 4) as the total within-clusters sum of squares diminishes at the same or a decreasing rate for subsequent cluster sizes.

As the resulting plot from applying the Gap Statistic Method visually suggests that 3 clusters is the optimal clusters' size, k = 3 was considered as the first choice. The second choice was four clusters (k = 4). These results are visually represented in Panel 3(b) which indicates the number of cuts – equivalent to the number of clusters – done to the cluster dendrograms generated by applying the ward function of the agglomerative hierarchical clustering that was selected as explained in the previous section.

### Results analysis and discussion

Next, the first and second choices of cluster sizes – which resulted in k = 3 and k = 4 respectively – were added to the original dataset, so countries can now be classified by the three cuts (clusters) in each cluster size grouping. To simplify the reporting of these results, panel 4 and panel 5(a) only document the output of classifying countries per each of the 3 clusters selected as the first choice of cluster size under the agglomerative hierarchical method explained in the previous section. The results are described as follows.

Unlike the optimal choice of cluster size obtained with k-means, the choice after applying the selected hierarchical clustering method resulted in a lower size. To some extent, this result can help to reduce the complexity that could have been associated in labelling each of the ten optimal clusters found with the k-means method. By taking a look into the cut tree no. 2 (cluster no. 2), the linkage between digital trade facilitation and lower inequality can be observed. Cluster no. 2 identifies countries that have higher levels of digital trade facilitation, lower levels of poverty and some other measures of within-country inequality. In addition, these countries are characterized by higher values of gross domestic product (GDP) in both aggregate and per capita terms, social assistance, and generation of economic resources through government revenues, foreign direct investment and trade. These countries also have higher levels of electricity access in both rural and urban areas, internet access, and urban population share. Cluster no. 3 could be observed as a group of countries where the linkage between digital trade facilitation and inequality is observed in medium terms. Lastly, cluster no. 1 underscore countries with the lowest levels of digital trade facilitation, inequality and other factors covered in the remaining variable groups.

Panel 5(b) includes the list of countries grouped in each of the three clusters obtained with the agglomerative hierarchical clustering method.

## 6. Discussion for further improvement of the machine learning model

To conclude, the optimal choice of clusters (cut tree = 3) found by applying the ward function of agglomerative hierarchical clustering methods seemed to have performed better in supporting the linkage between digital trade facilitation and reduced inequality that justified the ML task perform in this report. Cluster groups obtained through hierarchical clustering methods seem to be the more easily insightful for researchers and policymakers seeking to determine classification criterion to group countries according to their progress in digital trade levels and their current levels of between- and within-country inequality and other institutional, social and economic factors, such as those covered in this analysis. As a complimentary insight for decision-making in targeting digital trade facilitation efforts for bridging inequality gaps between and within countries, the optimal choice of clusters' size obtained with k-means could support to disentangle the various country-level characteristics identified in the three clusters found with hierarchical clustering.

To further improve the clustering models applied for the ML task described in this report, some calibrations could be done to the dataset used. Perhaps, more variables could be added to the variable groups on digital trade facilitation and between-country inequality. In addition, alternative measures could be specified to capture the dimension of between-country inequality. Indicators on gross domestic product (GDP) in both aggregate and per capita terms were chosen because these are the usual indicators referred in academic research describing analytical dimensions associated to between-country inequality. However, alternative measures such as the percentage of those GDP values in their corresponding maximum levels could be others worth exploring.

Lastly, sample size perhaps can be increased to see if different optimal choices for clusters' sizes can provide additional insights worth of empirical research and policy consideration. Increasing sample size for country-level observation is an effort that increases over time, so it can be worth replicating this ML task once new data and variables associated with the variable groups selected for the ML task become available in the coming years.

## 7. Conclusions and Implications for the Asia and the Pacific Region

Panels 5(a) and 5(b) in the Appendix section highlight the clusters in which countries from Asia and the Pacific region have been assigned to. For interpreting the results, looking at the panel 5(b) results could provide a more generalized diagnostic. Out of 55 countries from the region, 31 and 23 countries were assigned to clusters no. 2 and 1 respectively, with only 2 countries belonging to cluster no. 3. Panel 4 in the Appendix section provides an indication on the average levels of digital trade facilitation, income inequality and other indicators per each of the clusters in panel 5(b). According to panel 4 results, it could be interpreted that more than a half of countries in the Asia and Pacific region are in the cluster no. 2, which is the cluster with countries characterized by relatively higher levels of digital trade facilitation and lower inequality, with the latter measured by indexes of income distribution (the Gini Index) and the income share at the first income distribution quartile. Countries in cluster 2 also depict relatively lower levels of poverty, larger economic size, more social investments, and higher access to electricity and internet.

With the remaining countries from Asia and the Pacific region identified in clusters no. 1 and 3, this paper could then provide a contribution to policy efforts aiming to build technical capacity in countries that are lagging behind in digital trade facilitation and facing large income inequality gaps and socio-economic development challenges. An initial policy consideration could be setting cluster no. 2 countries as benchmark for identifying which effective digital trade facilitation measures could be replicated in countries belonging to clusters no. 1 and 3, as means for attaining desired policy objectives such as reducing income inequality. For a more disaggregated picture, panel 5(a) could also enable policymakers to look into digital trade facilitation measures observed across 10 clusters of countries in relation to levels of income inequality and other social and economic indicators.

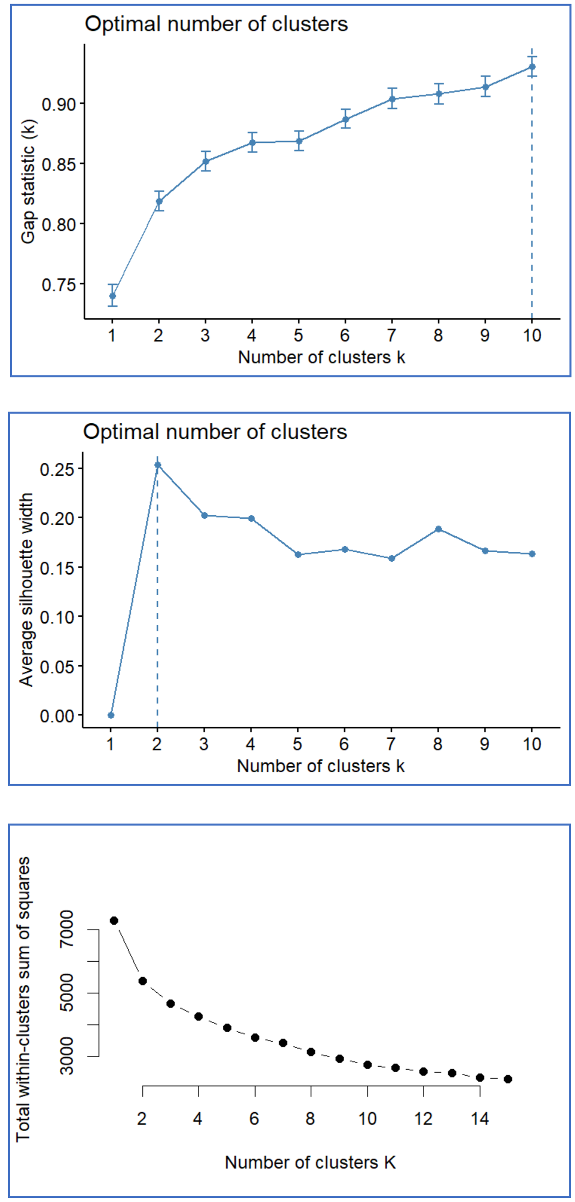



Figure 1: (a): results on optimal number of k-means clusters (gap statistic method, average silhouette method and elbow method)

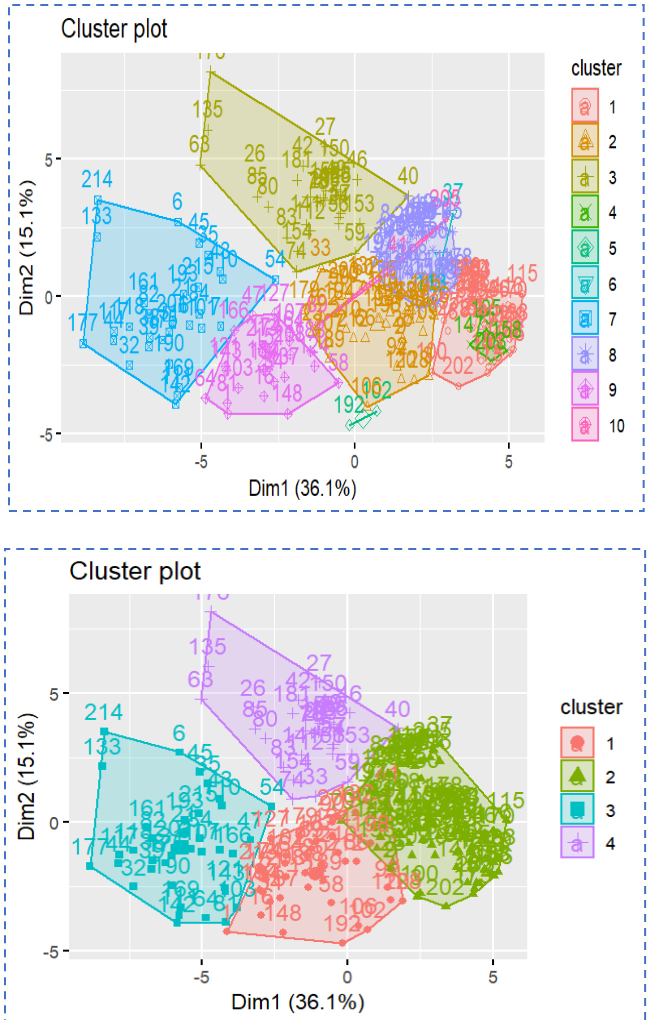
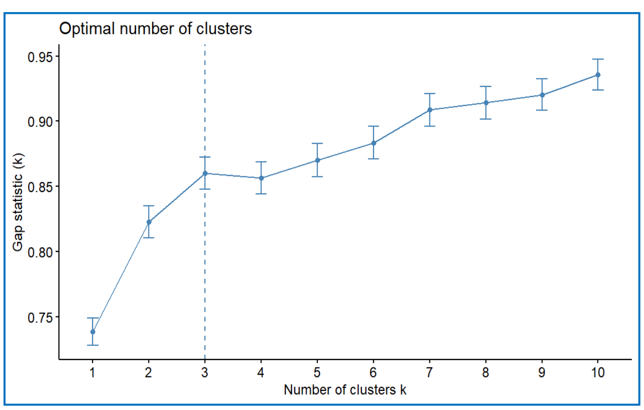
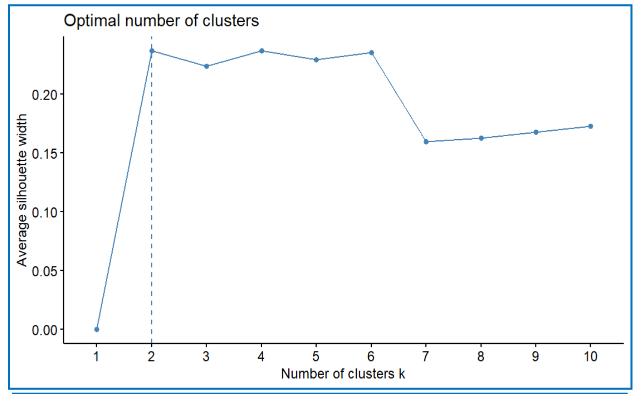
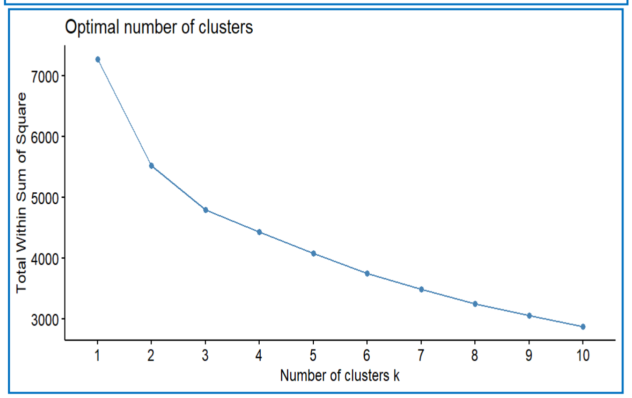
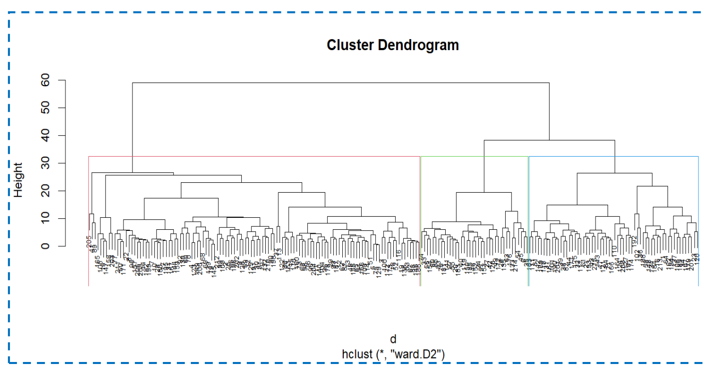



Figure 1(b): results of clustering groups with two choices of optimal k-means (fist choice k=10, second choice k-4)



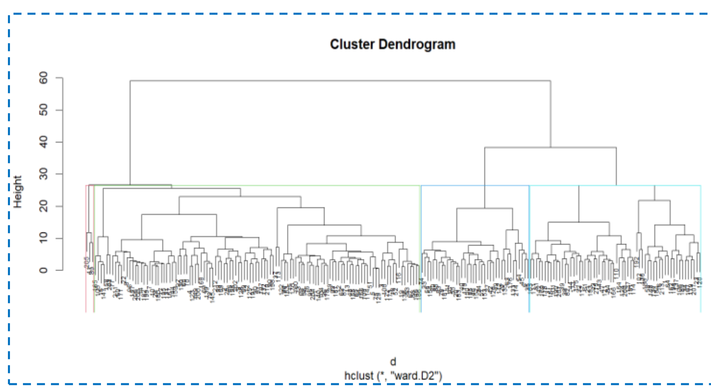




Figure 2 (a): results on optimal cluster size (gap statistic method, average silhouette method, elbow method, first choice hcl-ct-tree=3, second choice hcl-cut tree=4)

***Table 1: Summary stats of all variables after imputation was applied on missing values***

| Variable | Description | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|---|---|---|---|---|---|---|---|---|
| FDI_inflow_per_GDP | Foreign direct investment, net inflows (% of GDP) | 215 | 11.8 | 46.0 | -2.4 | 1.6 | 6.4 | 538.4 |
| GDP_capita | GDP per capita, PPP (constant 2017 international $, thousands) | 215 | 21.0 | 21.3 | 0.8 | 4.8 | 33.1 | 134.7 |
| GDP_ppp | GDP, PPP (constant 2017 international $, billions) | 215 | 547.4 | 1879.3 | 0.0 | 17.4 | 371.5 | 18500.0 |
| Capital_per_GDP | Gross fixed capital formation (% of GDP) | 215 | 23.4 | 6.7 | 8.7 | 19.9 | 26.0 | 55.4 |
| Gov_expense_per_GDP | General government final consumption expenditure (% of GDP) | 215 | 18.1 | 9.2 | 2.2 | 13.3 | 19.7 | 67.6 |
| Remittance_per_GDP | Personal remittances, received (% of GDP) | 215 | 4.4 | 5.8 | 0.0 | 0.4 | 6.4 | 34.0 |
| Trade_per_GDP | Trade (% of GDP) | 215 | 94.9 | 55.8 | 13.6 | 60.0 | 112.6 | 399.4 |
| Taxes_per_GDP | Tax revenue (% of GDP) | 215 | 16.8 | 6.4 | 0.0 | 13.1 | 18.3 | 66.8 |
| Gov_revenue_GDP | Revenue, excluding grants (% of GDP) | 215 | 26.8 | 14.9 | 0.0 | 18.4 | 32.5 | 169.8 |
| Gini_index | Gini index | 215 | 37.9 | 6.9 | 25.1 | 32.8 | 41.3 | 63.2 |
| Income_share_Q4 | Income share held by fourth 20% | 215 | 21.8 | 1.2 | 16.2 | 21.2 | 22.7 | 24.7 |
| Income_share_Q1 | Income share held by highest 20% | 215 | 45.0 | 5.7 | 34.6 | 41.1 | 47.6 | 68.6 |
| Income_share_Q5 | Income share held by lowest 20% | 215 | 6.6 | 1.6 | 2.5 | 5.3 | 7.8 | 10.1 |
| Income_share_Q2 | Income share held by second 20% | 215 | 11.1 | 1.8 | 4.8 | 10.0 | 12.3 | 14.8 |
| Income_share_Q3 | Income share held by third 20% | 215 | 15.5 | 1.6 | 8.1 | 14.7 | 16.7 | 18.7 |
| People_living_below_median_income | Proportion of people living below 50 percent of median income (%) | 215 | 13.7 | 4.8 | 3.5 | 10.6 | 18.5 | 24.8 |
| Extreme_poverty_ratio | Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population) | 215 | 10.3 | 17.7 | 0.0 | 0.5 | 12.4 | 78.5 |
| Poverty_ratio | Poverty headcount ratio at $3.20 a day (2011 PPP) (% of population) | 215 | 20.9 | 26.7 | 0.0 | 1.9 | 36.2 | 91.7 |
| Poverty_ratio_two | Poverty headcount ratio at $5.50 a day (2011 PPP) (% of population) | 215 | 35.2 | 33.4 | 0.1 | 5.7 | 64.3 | 97.9 |
| People_living_below_poverty_line | Poverty headcount ratio at national poverty lines (% of population) | 215 | 26.1 | 13.4 | 1.1 | 18.5 | 30.1 | 82.3 |
| Electricity_access | Access to electricity (% of population) | 215 | 83.2 | 26.8 | 4.2 | 73.0 | 100.0 | 100.0 |
| Electricity_access_rural | Access to electricity, rural (% of rural population) | 215 | 77.1 | 33.8 | 0.9 | 56.8 | 100.0 | 100.0 |
| Electricity_access_urban | Access to electricity, urban (% of urban population) | 215 | 91.6 | 16.7 | 10.6 | 92.6 | 100.0 | 100.0 |
| Internet_access | Individuals using the Internet (% of population) | 215 | 46.6 | 27.8 | 0.0 | 19.4 | 69.4 | 97.2 |
| Coverage_social_insurance_programs | Coverage of social insurance programs (% of population) | 215 | 25.5 | 18.0 | 0.4 | 6.3 | 44.9 | 55.7 |
| Coverage_social_protection_labor | Coverage of social protection and labor programs (% of population) | 215 | 53.7 | 23.4 | 1.9 | 41.3 | 71.0 | 96.9 |
| Coverage_social_safety_nets | Coverage of social safety net programs (% of population) | 215 | 39.3 | 18.4 | 0.6 | 27.0 | 50.6 | 93.3 |
| Population | Population, total, millions | 215 | 33.9 | 133.7 | 0.0 | 0.8 | 23.0 | 1374.6 |
| Male_population_share | Population, male (% of total population) | 215 | 49.9 | 3.2 | 45.9 | 48.7 | 50.4 | 76.1 |
| Female_population_share | Population, female (% of total population) | 215 | 50.1 | 3.2 | 23.9 | 49.6 | 51.3 | 54.1 |
| Urban_population_share | Urban population (% of total population) | 215 | 60.0 | 24.0 | 12.0 | 41.3 | 79.3 | 100.0 |
| Rural_population_share | Rural population (% of total population) | 215 | 40.0 | 24.0 | 0.0 | 20.7 | 58.7 | 88.0 |
| Net_ODA_aid_ocnstant | Net official development assistance and official aid received (constant 2020 US$, millions) | 215 | 499.9 | 835.4 | -465.9 | 28.8 | 565.8 | 5059.8 |
| Digital_trade_facilitation_index | Digital trade facilitation index (0 less; 1 more facilitation) | 215 | 0.5 | 0.2 | 0.1 | 0.4 | 0.7 | 0.9 |

*Table 2: Mean value of all variables by the first choice of cluster sizes obtained with K-means and hierarchical clustering*

| Varible Groups | Variables | Optimal cluster size with k-means (k = 10) | | | | | | | | | | Optimal cluster size with hierarchical clustering (k = 3) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 |
| Digital trade facilitation | Digital_trade_facilitation_index | 0.7 | 0.5 | 0.6 | 0.6 | 0.2 | 0.6 | 0.4 | 0.6 | 0.4 | 0.8 | 0.4 | 0.6 | 0.6 |
| Between-country inequality | GDP_capita | 47.9 | 10.7 | 12.9 | 58.9 | 2.6 | 50.7 | 2.4 | 28.6 | 4.2 | 25.2 | 3.9 | 32.7 | 11.8 |
| | GDP_ppp | 798.5 | 150.2 | 326.0 | 244.6 | 2.0 | 323.4 | 41.5 | 452.4 | 262.0 | 14183.3 | 128.0 | 844.6 | 294.6 |
| Within-country inequality | Gini_index | 30.6 | 34.9 | 48.8 | 30.9 | 28.3 | 37.8 | 42.8 | 39.0 | 34.7 | 38.9 | 37.2 | 34.7 | 48.9 |
| | Income_share_Q4 | 22.7 | 21.9 | 20.3 | 23.4 | 22.2 | 22.8 | 20.8 | 22.7 | 21.8 | 21.8 | 21.4 | 22.5 | 20.3 |
| | Income_share_Q1 | 38.9 | 42.9 | 54.1 | 38.6 | 37.9 | 44.1 | 49.3 | 45.1 | 42.8 | 45.8 | 44.8 | 42.1 | 54.1 |
| | Income_share_Q5 | 8.1 | 7.5 | 4.3 | 7.9 | 9.5 | 6.0 | 5.8 | 5.7 | 7.7 | 6.5 | 7.1 | 7.1 | 4.3 |
| | Income_share_Q2 | 13.1 | 11.8 | 8.4 | 12.7 | 13.5 | 11.1 | 9.9 | 10.8 | 11.8 | 10.8 | 11.2 | 11.9 | 8.4 |
| | Income_share_Q3 | 17.2 | 15.9 | 13.0 | 17.3 | 17.1 | 16.0 | 14.1 | 15.7 | 15.9 | 15.2 | 15.3 | 16.4 | 12.9 |
| | People_living_below_median_income | 10.1 | 10.9 | 19.9 | 11.8 | 5.6 | 16.4 | 15.0 | 17.1 | 9.8 | 12.8 | 11.4 | 12.9 | 20.0 |
| | Extreme_poverty_ratio | 0.3 | 1.5 | 6.7 | 0.0 | 11.7 | 0.9 | 46.7 | 1.2 | 16.6 | 8.9 | 27.2 | 1.1 | 12.3 |
| | Poverty_ratio | 0.6 | 10.0 | 16.2 | 0.0 | 41.1 | 2.4 | 70.6 | 3.3 | 46.6 | 24.8 | 52.1 | 4.1 | 23.5 |
| | Poverty_ratio_two | 1.6 | 33.1 | 32.9 | 0.1 | 75.2 | 7.4 | 87.4 | 10.0 | 77.5 | 40.5 | 76.5 | 12.4 | 40.2 |
| | People_living_below_poverty_line | 15.5 | 20.1 | 28.5 | 22.3 | 35.9 | 22.3 | 46.1 | 21.2 | 33.5 | 14.7 | 37.2 | 19.2 | 29.7 |
| Economic resource generation | Gov_expense_per_GDP | 18.9 | 21.2 | 16.6 | 16.9 | 66.7 | 10.9 | 14.4 | 18.1 | 15.8 | 13.8 | 17.2 | 19.0 | 16.6 |
| | Taxes_per_GDP | 19.8 | 16.4 | 16.7 | 14.4 | 44.0 | 18.0 | 12.9 | 18.6 | 13.6 | 10.2 | 14.7 | 18.0 | 16.3 |
| | Gov_revenue_GDP | 33.5 | 27.3 | 23.0 | 26.7 | 132.1 | 32.5 | 16.2 | 30.9 | 19.4 | 14.8 | 23.1 | 30.0 | 23.0 |
| | FDI_inflow_per_GDP | 15.1 | 4.6 | 4.5 | 1.9 | 1.2 | 452.6 | 5.4 | 11.5 | 3.3 | 2.0 | 3.9 | 18.0 | 5.3 |
| | Capital_per_GDP | 22.4 | 27.7 | 21.8 | 25.0 | 34.8 | 20.8 | 21.8 | 21.4 | 23.9 | 31.2 | 23.8 | 23.1 | 23.7 |
| | Trade_per_GDP | 137.4 | 92.3 | 73.8 | 122.5 | 99.4 | 133.6 | 71.2 | 104.8 | 66.7 | 39.5 | 69.1 | 111.9 | 83.2 |
| | Remittance_per_GDP | 1.3 | 9.2 | 4.9 | 0.2 | 8.3 | 1.4 | 4.4 | 3.0 | 5.0 | 1.2 | 6.8 | 3.3 | 3.9 |
| | Net_ODA_aid_ocnstant | 125.0 | 482.9 | 334.0 | 18.2 | 149.6 | 28.0 | 919.5 | 153.0 | 1535.2 | 778.2 | 1054.8 | 251.7 | 387.8 |
| Social assistance | Coverage_social_insurance_programs | 43.1 | 26.1 | 15.9 | 44.9 | 7.4 | 44.9 | 2.7 | 40.3 | 7.8 | 31.8 | 7.3 | 38.5 | 14.3 |
| | Coverage_social_protection_labor | 73.6 | 52.7 | 59.7 | 71.0 | 38.3 | 71.0 | 18.2 | 69.4 | 27.1 | 77.7 | 25.8 | 68.3 | 52.5 |
| | Coverage_social_safety_nets | 51.6 | 36.6 | 48.6 | 45.6 | 33.5 | 45.6 | 17.0 | 46.3 | 23.8 | 64.7 | 21.7 | 47.3 | 42.4 |
| Demographic, social and economic country-level characteristics | Electricity_access | 100.0 | 96.6 | 89.3 | 100.0 | 74.6 | 100.0 | 30.9 | 99.8 | 60.9 | 95.1 | 53.1 | 99.0 | 82.0 |
| | Electricity_access_rural | 100.0 | 91.5 | 79.8 | 100.0 | 66.0 | 100.0 | 13.7 | 99.7 | 49.4 | 93.2 | 42.3 | 97.4 | 69.3 |
| | Electricity_access_urban | 100.0 | 98.9 | 95.7 | 100.0 | 91.4 | 100.0 | 59.3 | 100.0 | 85.9 | 99.0 | 76.0 | 99.7 | 91.5 |
| | Internet_access | 80.0 | 40.5 | 40.5 | 82.8 | 17.0 | 82.8 | 10.0 | 64.9 | 16.5 | 47.6 | 15.8 | 65.3 | 37.6 |
| | Population | 18.0 | 15.7 | 23.3 | 4.2 | 0.6 | 0.0 | 16.6 | 16.8 | 43.6 | 998.6 | 25.5 | 41.9 | 22.2 |
| | Male_population_share | 49.1 | 50.5 | 49.6 | 66.6 | 49.9 | 47.8 | 49.7 | 48.5 | 49.9 | 50.9 | 50.0 | 49.9 | 49.8 |
| | Female_population_share | 50.9 | 49.5 | 50.4 | 33.4 | 50.1 | 52.2 | 50.3 | 51.5 | 50.1 | 49.1 | 50.0 | 50.1 | 50.2 |
| | Urban_population_share | 80.4 | 60.0 | 60.6 | 90.8 | 40.2 | 57.2 | 37.3 | 68.6 | 35.9 | 56.4 | 35.2 | 72.6 | 60.1 |
| | Rural_population_share | 19.6 | 40.0 | 39.4 | 9.2 | 59.8 | 42.8 | 62.7 | 31.4 | 64.1 | 43.6 | 64.8 | 27.4 | 39.9 |

**Note:** panel 4 applies a colour scale to differentiate among the higher and lower values per variable – the scale is applied horizontally. Red colours alike refer to the lowest values. Orange colours alike indicate lower-middle values. Yellow colours alike denote higher-middle values. And green colours point to higher values. This equivalence of variable values to scale colours applies for all variable groups except for the variables under the within-country inequality group. The latter follow a rather inverse equivalence which indicates that red-to-green colours correspond to high-to-low levels.

## Table 3: List of countries per cluster size groups

Table: 3 (a): List of countries in cluster groups obtained with optimal k-means size (k = 10)

| K = 10 | Countries per cluster group | No countries | No. AP Countries |
|---|---|---|---|
| 1 | Australia, Andorra, Austria, Belarus, Belgium, Brunei Darussalam, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Gibraltar, Hong Kong SAR, China, Hungary, Iceland, Ireland, Japan, Kazakhstan, Korea, Rep., Luxembourg, Macao SAR, China, Malta, Monaco, Netherlands, New Zealand, Norway, Poland, San Marino, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Sweden, Switzerland, Ukraine, United Kingdom. | 40 | 9 |
| 2 | Algeria, Albania, American Samoa, Armenia, Bhutan, Bosnia and Herzegovina, Cabo Verde, Equatorial Guinea, Fiji, French Polynesia, Gabon, Guam, Indonesia, Iraq, Jordan, Kyrgyz Republic, Lebanon, Maldives, Marshall Islands, Mauritius, Moldova, Mongolia, Morocco, Nauru, New Caledonia, Northern Mariana Islands, Palau, Samoa, Sri Lanka, Tajikistan, Tonga, Tunisia, Turkmenistan, Tuvalu, Uzbekistan, Vietnam, West Bank and Gaza. | 37 | 23 |
| 3 | Bolivia, Belize, Botswana, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Eswatini, Ghana, Grenada, Guatemala, Guyana, Honduras, Jamaica, Libya, Mexico, Namibia, Nicaragua, Panama, Paraguay, Peru, Philippines, South Africa, St. Lucia, St. Vincent and the Grenadines, Suriname. | 31 | 1 |
| 4 | Kuwait, Bahrain, Oman, Qatar, United Arab Emirates. | 5 | 0 |
| 5 | Timor-Leste, Kiribati. | 2 | 2 |
| 6 | Liechtenstein, Cayman Islands. | 2 | 0 |
| 7 | Benin, Angola, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Djibouti, Eritrea, Gambia, The, Guinea-Bissau, Haiti, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mozambique, Niger, Rwanda, Sierra Leone, Somalia, South Sudan, Tanzania, Togo, Uganda, Zambia, Zimbabwe. | 31 | 0 |
| 8 | Argentina, Antigua and Barbuda, Aruba, Azerbaijan, Bahamas, The, Barbados, Bermuda, British Virgin Islands, Bulgaria, Curacao, Faroe Islands, Georgia, Greece, Greenland, Iran, Islamic Rep., Isle of Man, Israel, Italy, Latvia, Lithuania, Malaysia, Montenegro, North Macedonia, Portugal, Puerto Rico, Romania, Russian Federation, Serbia, Seychelles, Sint Maarten (Dutch part), Spain, St. Kitts and Nevis, St. Martin (French part), Thailand, Trinidad and Tobago, Turkiye, Turks and Caicos Islands, Uruguay, Venezuela, RB, Virgin Islands (U.S.). | 40 | 7 |
| 9 | Bangladesh, Afghanistan, Cambodia, Cote d'Ivoire, Egypt, Arab Rep., Ethiopia, Guinea, Korea, Dem. People's Rep., Lao PDR, Mali, Mauritania, Micronesia, Fed. Sts., Myanmar, Nepal, Nigeria, Pakistan, Papua New Guinea, Sao Tome and Principe, Senegal, Solomon Islands, Sudan, Syrian Arab Republic, Vanuatu, Yemen, Rep. | 24 | 12 |
| 10 | India, China, United States. | 3 | 2 |

Table 3 (b): List of countries in cluster groups obtained with optimal hierarchical clustering cut tree (k = 3)

| **K = 3** | **Countries per cluster** | **No. Countries** | **No. AP Countries** |
|---|---|---|---|
| 1 | Afghanistan, Bangladesh, Benin, Bhutan, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Cote d'Ivoire, Egypt, Arab Rep., Eritrea, Ethiopia, Gambia, The, Guinea, Guinea-Bissau, Haiti, Kenya, Kiribati, Korea, Dem. People's Rep., Kyrgyz Republic, Lao PDR, Lesotho, Liberia, Madagascar, Malawi, Maldives, Mali, Mauritania, Micronesia, Fed. Sts., Myanmar, Nauru, Nepal, Niger, Nigeria, Pakistan, Papua New Guinea, Rwanda, Samoa, Sao Tome and Principe, Senegal, Sierra Leone, Solomon Islands, Somalia, South Sudan, Sri Lanka, Sudan, Syrian Arab Republic, Tajikistan, Tanzania, Timor-Leste, Togo, Tonga, Uganda, Uzbekistan, Vanuatu, Yemen, Rep., Zimbabwe. | 60 | 23 |
| 2 | Albania, Algeria, American Samoa, Andorra, Antigua and Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, The, Bahrain, Barbados, Belarus, Belgium, Bermuda, Bosnia and Herzegovina, British Virgin Islands, Brunei Darussalam, Bulgaria, Canada, Cayman Islands, China, Croatia, Curacao, Cyprus, Czech Republic, Denmark, El Salvador, Equatorial Guinea, Estonia, Faroe Islands, Fiji, Finland, France, French Polynesia, Gabon, Georgia, Germany, Gibraltar, Greece, Greenland, Guam, Hong Kong SAR, China, Hungary, Iceland, India, Indonesia, Iran, Islamic Rep., Iraq, Ireland, Isle of Man, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Rep., Kuwait, Latvia, Lebanon, Liechtenstein, Lithuania, Luxembourg, Macao SAR, China, Malta, Marshall Islands, Mauritius, Moldova, Monaco, Mongolia, Montenegro, Morocco, Netherlands, New Caledonia, New Zealand, North Macedonia, Northern Mariana Islands, Norway, Oman, Palau, Poland, Portugal, Puerto Rico, Qatar, Romania, Russian Federation, San Marino, Saudi Arabia, Serbia, Seychelles, Singapore, Sint Maarten (Dutch part), Slovak Republic, Slovenia, Spain, St. Kitts and Nevis, St. Martin (French part), Sweden, Switzerland, Thailand, Trinidad and Tobago, Tunisia, Turkiye, Turkmenistan, Turks and Caicos Islands, Tuvalu, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, RB, Vietnam, Virgin Islands (U.S.), West Bank and Gaza. | 117 | 31 |
| 3 | Angola, Belize, Bolivia, Botswana, Brazil, Cabo Verde, Cameroon, Chile, Colombia, Congo, Rep., Costa Rica, Cuba, Djibouti, Dominica, Dominican Republic, Ecuador, Eswatini, Ghana, Grenada, Guatemala, Guyana, Honduras, Jamaica, Libya, Malaysia, Mexico, Mozambique, Namibia, Nicaragua, Panama, Paraguay, Peru, Philippines, South Africa, St. Lucia, St. Vincent and the Grenadines, Suriname, Zambia. | 38 | 2 |

Disclaimer and acknowledgements: The opinions expressed in this paper should be attributed to the author. They are not meant to represent the positions or opinions of the United Nations Economic Commission for East Asia and the Pacific (ESCAP). This paper is a working draft version and must not be quoted nor cited. All remaining errors and omissions are the fault of the author.

## References

López González, J., & Jouanjean, M. (2017). Digital Trade: Developing a Framework for Analysis (OECD Trade Policy Papers, No. 205). OECD Publishing, Paris. https://www.oecd-ilibrary.org/trade/digital-trade_524c8c83-en

Mourougane, A. (2019). Measuring digital trade (Going Digital Toolkit Note, No. 18). Organisation for Economic Co-operation and Development (OECD). https://goingdigital.oecd.org/data/notes/No18_ToolkitNote_MeasuringDigitalTrade.pdf

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. A. (2007). An overview of clustering methods. Intelligent Data Analysis, 11(6), 583–605. https://doi.org/10.3233/ida-2007-11602

Organisation for Economic Co-operation and Development (OECD), World Trade Organization (WTO), & International Monetary Fund (IMF). (2019). Handbook on Measuring Digital Trade (Vol. 1). OECD, WTO and IMF. https://www.oecd.org/sdd/its/Handbook-on-Measuring-Digital-Trade-Version-1.pdf

Organisation for Economic Co-operation and Development (OECD). (2017). Measuring digital trade: Towards a conceptual framework (STD/CSSP/WPTGS(2017)3). OECD. https://unctad.org/system/files/non-official-document/dtl_eWeek2017c04-oecd_en.pdf

Organisation for Economic Co-operation and Development (OECD). (2022). Income inequality. In Society at a Glance: Asia/Pacific 2022. OECD Publishing Paris. https://doi.org/10.1787/7ef894e5-en.

United Nations (UN). (2021). UN Global Survey on Digital and Sustainable Trade Facilitation [Dataset]. UN. https://www.untfsurvey.org/

United Nations Conference on Trade and Development (UNCTAD). (2022). Digital trade: Opportunities and actions for developing countries (Policy Brief No. 92). UNCTAD. https://unctad.org/system/files/official-document/presspb2021d10_en.pdf

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP). (2020, February 10). The rise of the digital economy and tech-driven inequality : Can the tech revolution be an equaliser instead? unescap.org. Retrieved March 4, 2023, from https://www.unescap.org/blog/rise-digital-economy-and-tech-driven-inequality-can-tech-revolution-be-equaliser-instead#

World Bank. (2023). World Development Indicators [Dataset]. https://databank.worldbank.org/source/world-development-indicators

Yin, Z. H., & Choi, C. W. (2022). Does digitalization contribute to lesser income inequality? Evidence from G20 countries. Information Technology for Development, 29(1), 61–82. https://doi.org/10.1080/02681102.2022.2123443

Zhu, W., Li, X., Wang, H., & Sun, B. (2022). Digital service trade and income inequality: a panel data analysis for 100 countries. Applied Economics Letters, 1–5. https://doi.org/10.1080/13504851.2022.2061895

Zhuang, J. (2022). Income and Wealth Inequality in Asia and the Pacific: Trends, Causes, and Policy Remedies. Asian Economic Policy Review, 18(1), 15–41. https://doi.org/10.1111/aepr.12399