Identifying Sustainable Development Patterns of Nations and Cities in Three Asian Regions with Unsupervised Learning

Doyun Lee[*] · Jiyeon Park[**] · Danbi Kim[***] · Steven Jige Quan[****]

Abstract: Asia, the largest and most populous continent in the world, is in dire need of sustainable development due to its rapid urbanization; particularly, the Southeast, South, and West Asia regions demand the most attention. However, previous studies have been limited in examining sustainable development patterns in the three regions. This study identified sustainability development patterns in 23 countries and 33 major cities in three Asian regions using time series clustering, an unsupervised learning technique. Considering the three pillars of sustainability, three clusters were identified at both national and city levels based on three variables: population, GDP per capita, and total energy/electricity consumption per capita. The identified clusters suggested diverse sustainable development patterns in each of the three regions. The largest clusters at national and city levels exhibited a rapid development trend in all three aspects, indicating an urgent need to pursue sustainable development. Other clusters revealed complex trends that were closely linked to local development characteristics. Overall, the clusters at both national and city levels transcend the regions significantly, suggesting that future policymaking in the three regions should be tailored to local issues and take cues from regions with similar patterns rather than adhering to the conventional notion of Asian regions. The results provide fresh insights into sustainable development patterns in the three regions, thereby helping policymakers make better policy comparisons and more targeted policy execution decisions.
Key Words: South Asia, Southeast Asia, West Asia, Time Series Clustering, Machine Learning

*   Graduate School of Environmental Studies, Seoul National University, Korea
**   Graduate School of Environmental Studies, Seoul National University, Korea
*** Graduate School of Environmental Studies, Seoul National University, Korea
**** Corresponding author, Graduate School of Environmental Studies, Seoul National University, Korea, Associate Professor

I. Introduction

Sustainable development becomes more important in rapid global urbanization and accelerating climate change (Diesendorf, 2000). As the world's largest continent, Asia is facing more challenges. It hosts 60% of the world's current human population which is undergoing rapid urbanization. This raises big concerns on the environmental consequences of the development of Asian countries and cities. In particular, among six Asia regions, Southeast, South, and West Asia are the most fastest growing regions, and their development patterns should receive great attention (IEA, 2021). However, studies on the sustainable development of these regions are limited 〈Fig 1〉.

〈Figure 1〉 Study Areas
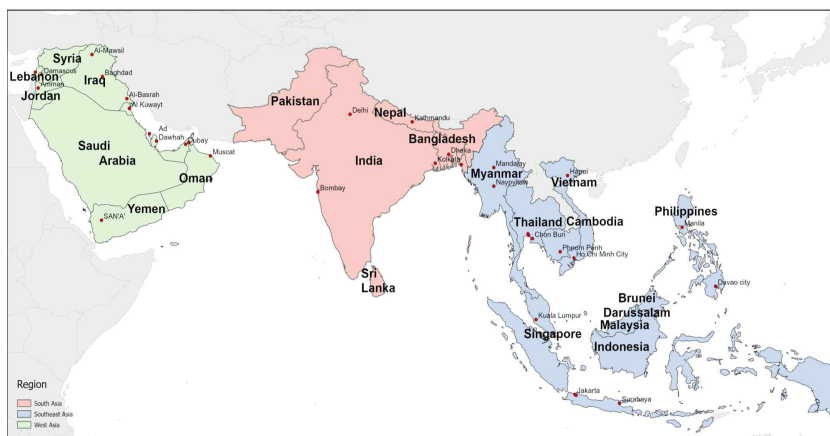
To have a concrete understanding on these three developing regions in Asia in terms of sustainability, this study investigated their development patterns with respect to the three pillars in sustainability, namely social, economic and environmental aspects

(Purvis et al., 2019). Specifically, this study discovered the development patterns in 23 countries and 33 major cities in three Asian regions with the time series clustering, an unsupervised learning technique. Three variables, population, gross domestic product (GDP) per capita, and total energy/electricity consumption per capita were used to represent three pillars of sustainability, namely, social, economic and environmental development, respectively (Chen et al., 2021). The choice of these variables were based on data availability in the three regions. Development trends measured by the three variables during 2010-2019 at national and city levels were examined using time series clustering, and three clusters were identified at both levels. The resulted clusters provided new insights into the diverse patterns of national and city-level development in the three regions.

Ⅱ. Literature Review

1. Identification of development patterns of nations and cities with clustering

The clustering method has been adopted in understanding development patterns of nations and cities. Identified clusters provide important information regarding societal variation and are a useful way to inspect intercountry relationships and further build up new theories (Gupta et al., 2002). The grouping of nations into similar clusters were often based on (Gupta et al., 2002): (a) geographic proximity (Furnham et al., 1994; Kirkcaldy and Lynn, 1994); (b) mass migration and ethnic social capital (Portes and Zhou, 1994); (c) religious and linguistic commonality (Cattell, 1950). Scholars also

applied the clustering method to understand development patterns of finer scales, e.g., the neighborhood development patterns in Chicago and LA (Delmelle, 2016). Recently there have been many studies applying clustering method to nations and cities. But since the clustering is mostly arbitrary, discussions are necessary to understand the clustering results and compare them with traditional grouping of locations often based on geographic proximity and socio-economic similarity.

2. Sustainable Development Patterns in Asian Countries and Cities

The sustainable development patterns in nations and cities in Asia, a rapidly growing area among all continents, are more of interest as they have diverse religions, cultures, large populations and energy use.

There have been clustering studies for Asian nations. Gupta et al. (2002) clustered 61 Global Leadership and Organizational Behavior Effectiveness (GLOBE) nations, including Asian counties, into 10 clusters, regarding cultural aspects. Kantar and Keskin (2013) examined electricity consumption and GDP in Asian countries using average linkage cluster analysis. Irfan and Shaw (2017) investigate clusters within South Asian countries and time period (1978-2011) on relationship between environmental pollution, energy consumption, and the level of urbanization. Kwon et al. (2018) analyze GHG emission history across 24 Asian countries coming up with six clusters, investigating the possibly reduceable carbon emission sectors. Arshad et al. (2020) examined the effects of information and communication technology, energy consumption and CO_2 emissions in the South and Southeast Asian regions using cluster analysis. Arshad et al. (2020) examined that two clusters were analyzed based

on their social development score.

However, studies on sustainable development patterns in Asia are still limited in general, especially at the city level. Even fewer studies were done for Southeast, South, and West Asia, the most rapidly growing regions.

3. Time Series Clustering for Identification of Trend Pattern

Most of previous studies on identification of clusters for nations and cities were based on static data. But development trend should be understood with time series data, which require particular clustering methods to analyze. For time series data, time series clustering is the most used approach in different fields to discover patterns. The method is powerful in that it can cluster a large amount of complex time series data with information with a straightforward visual interpretation, although challenges exist for too large and high dimensional time series data (Aghabozorgi et al., 2015).

Few studies used time series clustering method for nations and cities. Alonso et al. presents a time series clustering method for the historical data of CO_2 emissions in industrialized countries (Alonso et al., 2006). Zarikas et al. (2020) analyzed clustering of countries with COVID-19 cases. However, there are few literatures that focused on clustering developing Asian nations and cities, from a sustainable perspective.

Ⅲ. Data

1. Data Collection

The data were collected for nations and cities in Southeast, South, and West Asia. At the national level, energy consumption data were extracted from subsections of World Energy Balance provided by the International Energy Agency (IEA, 2021). GDP and population data were collected from the World Bank website (World Bank, 2022).

At the city level, population data were collected from the World Bank and the GDP and electricity consumption data were collected from the grid dataset created by Chen et al. (2021). The raw data of GDP and electricity were extracted in 1 x 1 km grid based on each city's administrative boundaries using ArcGIS (ArcGIS, 2022), and then aggregated at the city level.

2. Data Cleaning and Processing

The collected data at the nation level contained a large chunk of missing data, which is common in survey data because some nations may lack the means to collect and compile data in a consistent manner (Chi et al., 2016). This is the case particularly for the majority of country in the study area categorized as developing countries. In total, energy consumption data of six countries were missing (Bhutan, Afghanistan, Maldives, Timor-Leste, Laos and Palestine).

The city level data contained more missing data because of different levels of data coverage at the global scale in the source dataset (Chen et al., 2021). Overall, GDP and electricity consumption data were missing in 20 cities during the study period (Thimphu, Islamabad,

Karachi, Lahore, Faisalabad, Colombo, Kabul, Male, Bundar Seri Begawan, Dili, Vientiane, Singapore, Thimphu, Islamabad, Karachi, Lahore, Faisalabad, Colombo, Kabul and Male).

Energy consumption and GDP data were normalized by total population to develop the final datasets, which contained energy consumption per capita, total population, GDP per capita data for 23 nations and 33 cities ⟨Table 1⟩.

⟨Table 1⟩ Data Description

Data	Energy consumption per capita	Total population	GDP per capita
Scale	National data		
Time period	2000, 2005, 2010, 2015, 2019		
Source	IEA World Energy Balance	World Bank	World Bank
Unit	TJ / person	person	US dollar / person
Missing Data	Bhutan, Afghanistan, Maldives, Timor-Leste, Laos, Palestine		

Data	Energy consumption per capita	Total population	GDP per capita
Scale	City data		
Time period	2000 ~ 2019 (yearly)		
Source	Chen et al., 2021	World Bank	World Bank
Unit	kWh / person	person	US dollar / person
Missing Data	Thimphu, Islamabad, Karachi, Lahore, Faisalabad, Colombo, Kabul, Male, Bundar Seri Begawan, Dili, Yangon,, Vientiane, Singapore, Jiddah, Makkah, Ramallah, Masqat, Ad-Dawhah, Ar-Riyadh		

Ⅳ. Methodology

1. K-means Time Series Clustering

This study adopted the time series k-means, an unsupervised

machine learning algorithm, to cluster nations and cities in the three regions. The standard Euclidean k-means was assumed in the clustering because the time series data in this study have the same fixed intervals and lengths.

The standard Euclidean k-means problem is arguably the most widely-studied clustering problem (Dutta et al., 2017). In the general form of this problem, a set of n points in Euclidean space are given, $X = x_1, \ldots, x_n \in R^d$, and the objective is to select k center points $\mu_i \in R^d$, $i = 1$ to k, so that the sum of squared distances for each point to the nearest center is minimized (Awasthi et al., 2015). This problem can be written as:

$$\text{minimize} \ \sum_{x \in X} \min_{i \in [k]} \| x - \mu_i \|^2 \tag{1}$$

In this study, the clustering problem is more complex as it involves three variables, population, GDP per capita and total energy/ electricity consumption per capita. The three time series for the variables were concatenated to form one long time series for each sample. This resulted in a 15-dimensional data point for each sample at the nation level (five time points data of three variables) and 60-dimensional data point at the city level (20 time points with three variables). Before time series concatenation, the time series data of each variable were standardized. In this data preprocessing, the magnitude of the data was eliminated, and only the trends of sequences were considered in the clustering.

"TimeseriesKmean" module in "tslearn", a python library, was used in implementation (tslearn, 2022). The distance was assumed the

Euclidean distance in the algorithm. TimeSeriesScalerMeanVariance function in the package was used to standardize time series data for each variable separately in data preprocessing. Each clustering was repeated 50 times to deal with getting stuck in local optima due to the randomness in time series k-means.

2. Determining the optimal k

One of the most essential steps in clustering is to find the optimal number of clusters where the data points are added (Somu et al., 2021). The elbow method and silhouette score have been used to identify optimal k for k-means. In this study, the silhouette score was adopted due to its wide use and convenience in computation. The clustering was applied with k ranging from 2 to 10, and silhouette scores were calculated for those experiments based on Eq (2-3):

$$S = \frac{\displaystyle\sum_{i=1..N} s(i)}{N} \tag{2}$$

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

Where S is the silhouette score, $s(i)$ is silhouette statistic, $a(i)$ is the average distance to all the points belong to the nearest cluster other than its own cluster $b(i)$ is the average distance to points in the nearest cluster besides its own nearest defined by the cluster minimizing this average distance, N is the number of observations (Tibshirani et al., 2001).

Then the k value with the highest silhouette score was selected as the optimal k and the corresponding clustering results were chosen

for analysis.

Ⅴ. Results

1. Cluster Results of Nations

〈Table 2〉 indicates the clustering result of national population, national GDP per capita and total energy consumption per capita. Nations were clustered into three different groups 〈Table 2〉 and 〈Fig 3〉. 〈Fig 2〉 shows the correspondence between geographic regions and identified clusters.

〈Figure 2〉 Mapping of clustered nations in the three Asian regions
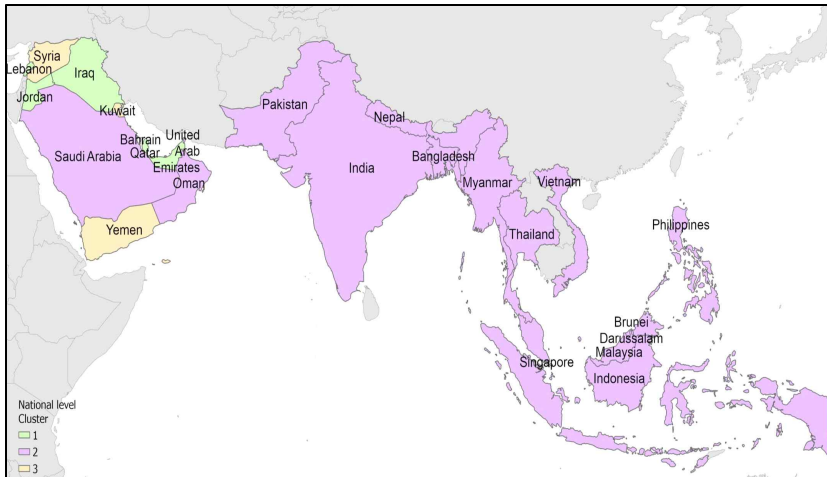
〈Table 2〉 Identified clusters of nations in three Asian regions

Cluster	Nation	Region	Cluster	Nation	Region
Cluster 1	Bahrain	West Asia	Cluster 2	Myanmar	Southeast Asia
	Iraq	West Asia		Indonesia	Southeast Asia
	Jordan	West Asia		Malaysia	Southeast Asia
	Lebanon	West Asia		Philippines	Southeast Asia
	Qatar	West Asia		Singapore	Southeast Asia
	United Arab Emirates	West Asia		Thailand	Southeast Asia
Cluster 2	Bangladesh	South Asia		Vietnam	Southeast Asia
	India	South Asia		Oman	West Asia
	Pakistan	South Asia		Saudi Arabia	West Asia
	Nepal	South Asia	Cluster 3	Kuwait	West Asia
				Syria	West Asia
	Brunei	South Asia		Yemen	West Asia

〈Figure 3〉 Flowchart of three clusters for Asian Nations

〈Figure 4〉 indicates the trends of the growth in national population, GDP per capita, and total energy consumption per capita from year 2000 to 2019 〈Fig 4〉. Table 3 shows the national level descriptive statistics of three Asian regions 〈Table 3〉.

Countries from Cluster 1 show increase in national population and GDP per capita, but decrease in total energy consumption per capita. This pattern depicts the sustainable development characteristics in parts of West Asia, which constitute Cluster 1. Such a pattern may be explained with two different mechanisms. The development of the nations might have reached to a level that does not need intensive energy consumption. An alternative explanation is that those nations are aware of energy consumption issue and reduce it during their development. Either way, Cluster 1 shows a positive outcome on sustainability

Cluster 2 is the largest cluster and shows increase in all three variables. Countries from cluster 2 contain all of the countries from South and Southeast Asia, and two countries from West Asia. This indicates that most of the Asian countries from three regions had increase in population and GDP per capita, but decrease in total energy consumption per capita, which implies that most of Asian countries need more radical policies to reach sustainable development.

Cluster 3 is characterized by a complex development pattern which has a peak in all three variables: an increase until 2010 and a decrease after 2010 in population and GDP per capita, and an increase until 2005 but a decrease abruptly after 2005 for total energy consumption per capita. Only three West Asia countries belong to Cluster 3, and the pattern suggested those nations might have undergone unusual national events around 2010 that impacted the development trend (i.e. war, economic crisis, famine, etc.).

〈Figure 4〉 Development Patterns in Population, GDP per capita and total Energy consumption per capita of Asian Nations
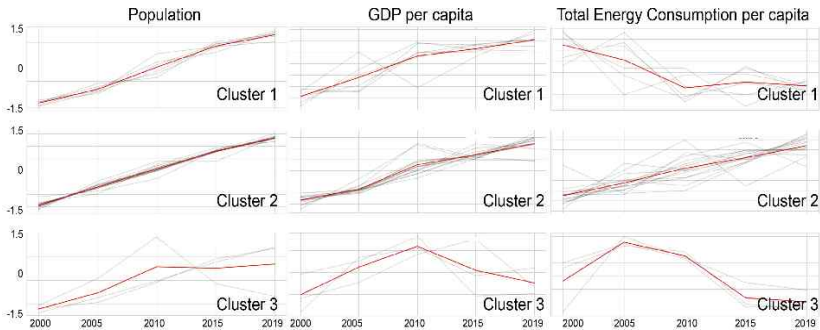

〈Table 3〉 National level descriptive statistics of three Asian regions

Variable	unit	description	mean	std	min	max
Energy consumption per capita	kWh / person	energy consumption per capita of three Asian regions	0.084	0.094	0.004	0.409
Total population	person	total population of three Asian regions	99,103,202	249,000,000	333,166	1,370,000,000
GDP per capita	US dollar / person	GDP per capita of three Asian regions	12,703.41	16,412.19	146.605	67,403.09

2. Cluster Results of Cities

〈Table 4〉 indicates the clustering result of cities considering the total population, GDP per capita and total energy consumption per capita 〈Fig 6〉. Cities were clustered to three different groups 〈Table 4〉 and 〈Fig 5〉. 〈Fig 6〉 shows the correspondence between geographic regions and identified clusters. 〈Figure 7〉 indicates the development trends of total population, GDP per capita, and total energy consumption per capita for the three clusters. Table 5 shows the national level descriptive statistics of three Asian regions 〈Table 5〉.

〈Figure 5〉 Mapping of clustered cities in the three Asian regions
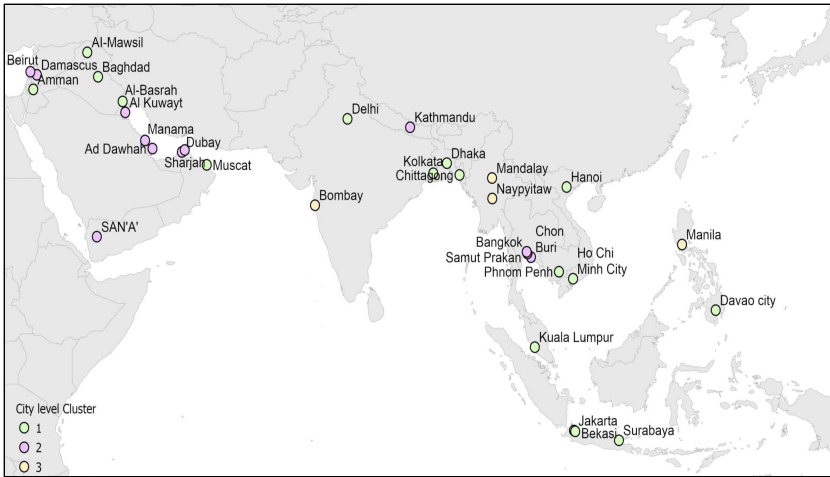
〈Table 4〉 Identified clusters in cities in the three Asian regions

Cluster	City	Region	Cluster	City	Region
Cluster 1	Hồ Chí Minh city	Southeast Asia	Cluster 2	Bangkok	Southeast Asia
	Hà Nội	Southeast Asia		Chon Buri	Southeast Asia
	Bekasi	Southeast Asia		Samut Prakan	Southeast Asia
	Surabaya	Southeast Asia		SAN'A'	West Asia
	Jakarta	Southeast Asia		Dubay	West Asia
	Kuala Lumpur	Southeast Asia		Sharjah	West Asia
	Davao city	Southeast Asia		Damascus	West Asia
	Phnom Penh	Southeast Asia		Ad Dawhah	West Asia
	Muscat	West Asia		Beirut	West Asia
	Amman	West Asia		Al Kuwayt	West Asia
	Al–Basrah	West Asia		Manama	West Asia
	Baghdad	West Asia		Kathmandu	South Asia
	Al–Mawsil	West Asia	Cluster 3	Mandalay	Southeast Asia
	Chittagong	South Asia		Naypyitaw	Southeast Asia
	Dhaka	South Asia		Manila	Southeast Asia
	Delhi	South Asia		Bombay	South Asia
	Kolkata	South Asia			

《Figure 6》Flowchart of three clusters in the three Asian regions
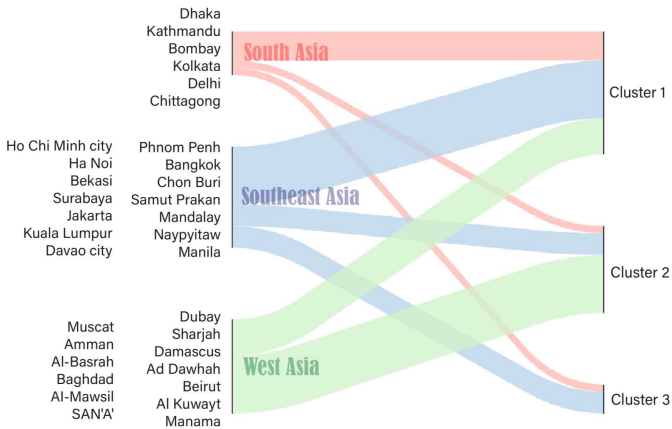

Cities from cluster 1 show increase in all three variables. As Cluster 1 is the largest cluster, containing most South and Southeast Asian cities and some West Asian cities, the pattern suggested that most cities in the three regions are developing at a radical rate, requiring stronger policies towards sustainable development.

Cluster 2 shows a pattern of an increase in population but a decrease in GDP per capita and total energy consumption per capita. This trend indicates that cities from cluster 2 has an inefficient development, which consist of a majority of cities from West Asia, a few from Southeast Asia, and one from South Asia. Specifically, cities in this cluster might have been overgrowing, with economic development left behind. The energy use per capita decreased, but it might be attributed to the rapid increase of total population and less growth of economic development and living conditions.

Cluster 3 has an opposite pattern compared to Cluster 2, showing a decrease in population but an increase in GDP per capita and energy consumption per capita. This suggests that the only a few

cities from South Asia and Southeast Asia cities in cluster 3, the smallest cluster, did not grow in size, but had an improvement in economic output and possibly living conditions. Such a pattern suggests an efficient development, however, the increase in energy consumption indicates the necessity of sustainable development policies. Further inspection is needed to understand the reason of the population decrease.

〈Figure 7〉 Development Patterns in population, GDP per capita and total Energy consumption per capita in three Asian regions
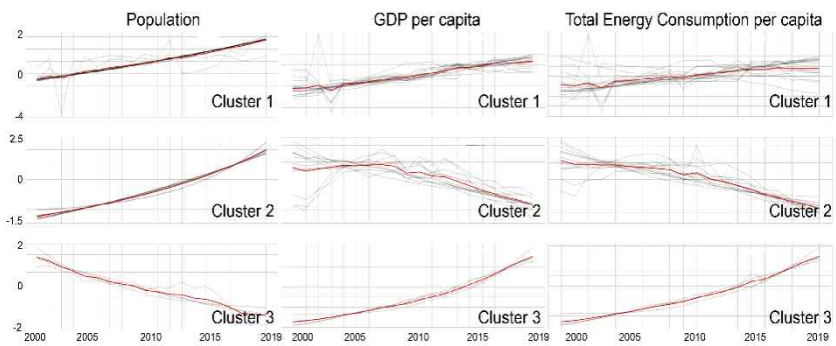
〈Table 5〉 City level descriptive statistics of three Asian regions

Variable	unit	description	mean	std	min	max
Energy consumption per capita	kWh / person	energy consumption per capita of three Asian regions	892	1121	11	6,955
Total population	person	total population of three Asian regions	2,499,539	3,288,049	60,904	18,975,420
GDP per capita	US dollar / person	GDP per capita of three Asian regions	6,860	7,589	191	42,428

VI. Conclusion

There is an urgent need for sustainable development in Asia which is expected to have great environmental challenges in rapid urbanization. In particular, the South, West, and Southeast Asia regions have the most rapid urbanization in Asia, and require most attention on their development patterns. However, previous studies on those three regions are few and the temporal dimension was little considered in identifying their development patterns.

This study identified the development patterns considering the three pillars in sustainability in 23 countries and 33 major cities in three Asian regions with time series clustering, an unsupervised learning technique. Three clusters representing three different development patterns were identified at both national and city levels based on three variables of population, GDP per capita, and total energy/electricity consumption per capita. The identified clusters suggested diverse sustainable development patterns in the three regions. Among them, the largest clusters at the nation (Cluster 2) and city level (Cluster 1) showed a rapid developing pattern. This indicates that most of the nations and cities in the three Asian regions are in urgent need of sustainable development. Other clusters revealed complex trends that were closely linked to local development characteristics. In particular, Cluster 1 at the nation level showed a growth in GDP per capita and population, but a decrease in energy consumption per capita, implying an efficient and sustainable development pattern. But further examinations are necessary to understand the local conditions and development patterns. In general, Asian cities showed more diverse developing patterns than Asian countries.

Overall, the clusters at both nation and city levels greatly transcend the regions, suggesting future policy making in the three regions should be tailed to local issues and learn from places with similar patterns rather than following the conventional notion of Asian regions. The results provide new insights into sustainable development patterns in the three regions, helping policy makers for better policy comparison and more targeted policy implementation. Moreover, there is a need to develop various assessment tools for sustainable development measurement on each cluster on national as well as city levels. Such tool development can enhance coordination between cities and nations to have more aligned efforts on sustainable development.

This study still has some limitations, such as the limited list of cities and periods examined, due to the data availability issue. Also, analysis with a mixed method can provide more comprehensive view of the development patterns at the nation and city levels for policy making. These will be examined in future research.

▌References ▌

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah, 2015, "Time-series clustering–a decade review." *Information Systems*, 53, pp.16-38.
Alonso, A. M., J. R. Berrendero, A. Hernández, and A. Justel, 2006, "Time series clustering based on forecast densities." *Computational Statistics & Data Analysis*, 51(2), pp.762-776.

Arshad, Z., M. Robaina, and A, Botelho, 2020, "The role of ICT in energy consumption and environment: an empirical investigation of Asian economies with cluster analysis." *Environmental Science and Pollution Research*, 27(26), pp.32913-32932.

Awasthi, P., M. Charikar, R. Krishnaswamy, and A. K. Sinop, 2015, "The hardness of approximation of euclidean k-means." arXiv:1502.03316

Cattell, R., 1950, "The Principal Culture Patterns Discoverable in the Syntal Dimensions of Existing Nations." *The Journal of Social Psychology*, 32(2), pp.215-253.

Chen, J., M. Gao, S. Cheng, W. Hou, M. Song, X. Liu, and Y. Liu, 2022, "Global 1 km × 1 km gridded revised real gross domestic product and electricity consumption during 1992-2019 based on calibrated nighttime light data." *Scientific Data*, 9(1), p.202.

Chi, J., E. Chi, and R. Baraniuk, 2016, "K-POD: A Method for k-Means Clustering of Missing Data." *The American Statistician*, 70(1), pp.91-99.

Delmelle, E., 2016, "Mapping the DNA of Urban Neighborhoods: Clustering Longitudinal Sequences of Neighborhood Socioeconomic Change." *Annals of the American Association of Geographers*, 106(1), pp.36-56.

Diesendorf, M., 2000, Sustainability and sustainable development. Sustainability: The corporate challenge of the 21st century, Sydney:Allen & Unwin, 2, pp. 19-37.

Dutta, A., A. Vijayaraghavan, and A. Wang, 2017, "Clustering Stable Instances of Euclidean k-means." arXiv:1712.01241

ESRI, n.d, "*ArcGIS pro*," https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm, [2022.12.30]

Furnham, A., B. Kirkcaldy, and R. Lynn, 1994, "National Attitudes to Competitiveness, Money, and Work Among Young People: First, Second, and Third World Differences." *Human Relations*, 47(1), pp.119-132.

Gupta, V., P. Hanges, and P. Dorfman, 2002, "Cultural clusters: Methodology and findings." *Journal of World Business*, 37(1), pp.11-15.

Gupta, V., G. Surie, M. Javidan, and J. Chhokar, 2002, "Southern Asia cluster: Where the old meets the new?" *Journal of World Business*, 37(1), pp.16-27.

IEA, 2021, Net Zero by 2050 – A Roadmap for the Global Energy Sector, Paris: IEA.

Irfan, M. and K. Shaw, 2017, "Modeling the effects of energy consumption and urbanization on environmental pollution in South Asian countries: A nonparametric panel approach." *Quality & Quantity*, 51(1), pp.65-78.

Kantar, E. and M. Keskin, 2013, "The relationships between electricity consumption

and GDP in Asian countries, using hierarchical structure methods." *Physica A*, 392(22), pp.5678-5684.

Kwon, Y., H. Lee, and H. Lee, 2018, "Implication of the cluster analysis using greenhouse gas emissions of Asian countries to climate change mitigation." *Mitigation and Adaptation Strategies for Global Change*, 23(8), pp.1225-1249.

Portes, A. and M. Zhou, 1994, "Should immigrants assimilate?" *The Public Interest*, 116(116), pp.18-33.

Purvis, B., Y. Mao, and D. Robinson, 2019, "Three pillars of sustainability: In search of conceptual origins." *Sustainability Science*, 14(3), pp.681-695.

Somu, N., G. M R. Raman, and K. Ramamritham, 2021, "A deep learning framework for building energy consumption forecast." *Renewable & Sustainable Energy Reviews*, 137, p.110591.

Tibshirani, R., G. Walther, and T. Hastie, 2001, "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 63(2), pp.411-423.

tslearn, n.d, *"tslearn.clustering.TimeSeriesKMeans," https://tslearn.readthedocs.io/ en/stable/gen_modules/clustering/tslearn.clustering.TimeSeriesKMeans.h tml*, [2022.12.30]

WCED, 1987, World Commission on Environment and Development: Our Common Future, Oxford: Oxford University Press.

World Bank, 2022, "World Development Indicators, Population," https://www.worldbank. org/en/about/leadership?cid=ECR_GA_worldbank_EN_EXTP_search&gclid=Cj 0KCQiAtbqdBhDvARIsAGYnXBOTLBjn63IrzkO31xUn7yDdFYUUxG6dWACLOR sI6RErOV_O-GAEH8MaAqxWEALw_wcB, [2022.12.30]

World Bank, 2022, "World Development Indicators, GDP per capita," https://www. worldbank.org/en/about/leadership?cid=ECR_GA_worldbank_EN_EXTP_s earch&gclid=Cj0KCQiAtbqdBhDvARIsAGYnXBOTLBjn63IrzkO31xUn7yDd FYUUxG6dWACLORsI6RErOV_O-GAEH8MaAqxWEALw_wcB, [2022.12.30]

Zarikas, V., S. Poulopoulos, Z. Gareiou, and E. Zervas, 2020, "Clustering analysis of countries using the COVID-19 cases dataset." *Data in Brief,* 31, p.105787.

Doyun Lee: Doyun Lee is in the master's course in urban and regional planning at Seoul National University's Graduate School of Environmental Studies. Her research interests include sustainable urban planning, smart cities, and aging cities(un2ver5e@snu.ac.kr).

Jiyeon Park: Jiyeon Park is in the master's course in urban and regional planning at Seoul National University's Graduate School of Environmental Studies (ppqq2008@snu.ac.kr).

Danbi Kim: Danbi Kim is in the master's course in urban and regional planning at Seoul National University's Graduate School of Environmental Studies (kimx5250@snu.ac.kr).

Steven Jige Quan: Steven Jige Quan is an Associate Professor and the Director of the City Energy Lab in the Graduate School of Environmental Studies at Seoul National University. His research interests include smart and sustainable urban systems, urban form and energy, AI-aided design, and sustainable zoning system(sjquan@snu.ac.kr).