Measuring the Economics of a Pandemic:
How People Mobility depict Economics? An Evidence of People's Mobility Data towards Economic Activities.

Rendra Achyunda Anugrah Putra
BPS – Statistics Indonesia
rendra.ap@bps.go.id

Silvia Arini
BPS – Statistics Indonesia
arini.silvia@bps.go.id

Abstract

COVID-19 pandemic has dramatically changed the world economy in just a few weeks, including economic activities, such as consuming and working from home, patterns of demand, organization of production, global value chains, employment dynamics, and income distribution as financial conditions and stability. Social restriction to reduce the virus spread not only creates an obstacle for statistics officers to obtain data to produce key leading economic indicators but also changes social behaviour in economic activity, which needs to be measure. At the same time, reliable and punctual data are crucial when the economy faces turbulence caused by extraordinary situations. Therefore, the national statistical offices need to modernize the data life cycle of compilation, production, and dissemination to make it robust to shocks, including pandemic.

This study tries to explore the usefulness of a different proxy for measuring economic activity through the changes of people's activities using people mobility data, Google Mobility and Apple Mobility data. Cluster analysis is used to classify regions based on their source of economic growth. Since Indonesia imposes social restriction, some industries remain operated and people's mobility will have a different impact on every industry, such as mining, transportation, and trade. This study establishes models to predict regional GDP based on mobility and considering the industry contribution to regional GDP. Furthermore, the estimated regional GDP will be used to estimate the national GDP.

This study finds a consistently strong positive correlation between regional change on average people's mobility and the economic growth among 34 provinces in Indonesia. The Indonesia national GDP growth in 2020 quarter 1 is 2.64%, while the actual GDP growth in the same period is 2.97%. Additionally, the estimated Indonesia GDP growth in 2020 quarter 2 is -5.17%, while the real growth GDP in 2020 quarter 2 is -5.36%.

Keywords: People mobility, Google mobility, Apple mobility, economic growth

I. Introduction

The WHO has been declared COVID-19 as a pandemic due to the alarming levels of spread and severity on March 11, 2020. Before COVID-19, the human race has already witnessed several pandemics such as H1N1 (2009), polio (2014), Ebola (2014 and 2019), Zika (2016). According to the report of the World Health Organization (WHO as of August 15, 2020), the number of COVID-19 confirmed cases is 21,026,758 people and killed more than 755,786 people in more than 213 countries around the world and two international conveyances (WHO, 2020). Indeed, the COVID-19 pandemic has greatly changed the world economy in just a few weeks. This extreme change will likely see long-lasting effects on the functioning of the economy, including such things as consuming and working from home, patterns of demand, organization of production, global value chains, employment dynamics, income distribution as well as financial conditions and stability (IMF, 2020).

In crisis times, there is, even more, a need for high-quality data to be available on a timely basis (CCSA, 2020). Reliable and punctual data are crucial in times when the economy faces turbulence caused by extraordinary situations. The disruption in data collection affects or will impact worldwide access to key economic indicators, including GDP growth. In Indonesia, COVID 19 caused some survey activities could not be conducted in some regions due to social activities restrictions, especially in most large cities with a high number of COVID cases. Indeed, this social limitation affects field surveys that are regularly conducted by official statistics. For instance, some stores are closed, which leads to missing prices when constructing the consumer price index. This matter causes some commodities cannot be recorded and left them as missing values, which lead to bias.

In the pandemic era, GDP statistics become more critical to evaluate economic activity, and around the world pays more attention to these statistics. GDP statistics give the signal to economists that a recession may be looming or give confidence to businesses that consumers are freely spending (Moyer & Dunn, 2020). However, even before the pandemic, GDP often publishes with a delay. The pandemic not only creates an obstacle for statistics officers to obtain data to produce key leading economic indicators but also changes social behavior in economic activity, which needs to be measure, such as working from home.

The national statistical offices have to respond to new data needs at the time of a pandemic. Some research already tried to identify proxy variables to measure economic activity, which robust to shocks. Therefore, the demand of providing new methodology or new potential data sources to deal with the pandemic situation is highly needed. Therefore, statistical agencies must find the best practices in statistical business continuity. In other words, national and international statistical organizations need to modernize the data life cycle of compilation, production, and dissemination, which robust to shock, including pandemic.

Social restrictions or mobility restrictions policy affect economic activity. This policy affects differently towards a country's economic activities as well as declining GDP since it depends on the severity and duration of the policy (Canaks et al., 2020). Indonesia imposes a social restriction to reduce the spread of the virus. This regulation is different from total lockdown, which obliges all industries to close except the essential one, such as health and groceries. Under social restrictions, the industries can still run their activity with new procedures, although some industries which is highly related to mobility will suffer most significant.

This study aims to explore the usefulness of a different proxy for measuring economic activity through the changes of people's movements that can be observed by using people's mobility data. This study also examines that people's mobility changes give a useful proxy for measuring GDP growth by considering the industry contribution to GDP since each industry responds differently to mobility change. This study provides useful insight into some possible opportunities for innovation regarding new potential data sources using big data to measure economic activities during the pandemic.

II. Literature Review

A. Mobility and Economics Activities

There has been a spate of empirical evidence that derives econometric relationships between people's mobility and various indicators of economics. People's mobility becomes a potential data source for measuring economic activities. Big data on traffic lows can reveal the status of regional economic development, which is derived from the information of human mobility (B. Li et al., 2020). Also, billions of geo-positioning data produced by mobile phones are used to measure employment trends in industrial parks and consumer activity in commercial areas (Dong et al., 2017).

Human mobility has a significant association with social interaction and economic development. In addition, the diversity of human movement is an important aspect to determine the socio-economic status of a region in which both diversity and socio-economic has a two-directed relationship. A High-developed region provides a wide range of mobility, and the higher mobility individual diversification drives to higher economic opportunity, which leads to higher socio-economic development (Pappalardo et al., 2016).

Indeed, social restrictions and people's mobility restriction policy will affect economic activity. This policy will affect differently towards a country's economic activities as well as declining GDP since it depends on the severity and duration of the lockdown. The strictest lockdown measures have been imposed to lead the countries to suffer the most significant decline in economic activity (Canaks et al., 2020).

B. Nowcasting Economic Activity using Big Data

Providing a real-time assessment of economic activity is beneficial for stakeholders and policy-maker. Since the official data published with a delay, nowcasting success to be combined with large data set to overcome timeliness issues and has been applied to some sectors, such as sales, employment, and GDP (Guidotti et al., 2016). Many types of big data have already been exploited in many scientific fields, including economic. Furthermore, it is useful for macroeconomic nowcasting (Buono et al., 2017). The model using the combination of big data such as credit card transaction and Google data and official statistics data, such as labor force, wage, and the price, is expected to increase the accuracy and early read (Moyer & Dunn, 2020).

Mobile data contains beneficial information that provides information about human behavior (Buono et al., 2017). It can be used to measure some economic indicators. Compared to traditional survey-method, mobile data has some advantages since it encompasses real-time, low-cost, and high coverage data. Additionally, it provides a more significant benefit in increasing accuracy and completeness of measurement when it is combined with official statistics data (Dong et al., 2017).

Nowcasting economic activity has been proposed by some researchers by using high-frequency basis data (Ferrara & Simoni, 2019). The high-frequency data is expected to reduce the uncertainty for nowcasting economic activity on the effect of COVID-19. Google mobility index and Apple mobility index are examples of high-frequency data that depict people's mobility. The Google Mobility Index has become a new good data source to predict industrial production, which use to assess the GDP growth.

The quality of national data, which uses as a basis to measure the GDP, differs among countries in the world. In some regions, especially developing countries with low-quality national accounts data, the GDP itself is often poorly measured and lack accuracy. In addition, some states even do not have a national account. Therefore, the night light, one of the satellite data, can be used to improve or even replace the measure of economic growth in low quality national account data. Otherwise, it gives little value in improving the measurement (Henderson et al., 2012).

Estimation of economic activity using the night light can also overcome the problem associated with data collected through field survey. The night light dataset is easily available and can be frequently retrieved on a daily basis. Therefore, this dataset can be used to estimate GDP, which is usually obtained through an economic census conducted every 5-10 years (Ghosh et al., 2010). Moreover, the night light satellite becomes a potential tool for monitoring spatial and temporal social economics parameters. The study in some countries shows that the per capital lighting correlated with per-capita GDP (S. Li et al., 2017). Furthermore, the estimated GDP using night light has a higher value than the official GDP because it covers the informal economy activity (Ghosh et al., 2010).

The usage of night light data is also useful for analyzing sub- and supranational regional levels within countries (Henderson et al., 2012), although the relationship is relatively unstable. The regional instability relationship between night light and GDP across regions occurs because of various reasons. Thus, adding the control variables, which account for the regional characteristics, can stabilize the relationship between night light and GDP (Bickenbach et al., 2016).

III. Data and Methodology

A. Data

This study combines both traditional and non-traditional data sources. These data are used to measure economic activities during the pandemic, which cannot be measured by official statistics due to limitations, such as a social restriction or low statistical capacity. In addition, the new data sources are more reliable to any disruptions, including social activities restriction, since society will still produce information through big data, which represents their economic activities. The data involved in this study encompasses:

1) **Non-traditional Indicator (Big Data)**

- **Apple Mobility Trend Report**

  The Apple mobility trend report is published daily and reflect requests for directions in Apple Maps. This data shows a relative volume of directions requests per country/region, sub-region, or city compared to a baseline volume on January 13, 2020, consistent with normal, seasonal usage of Apple Maps (Apple, 2020). This study uses the Apple mobility report to provide data from January 13, 2020, to June 30, 2020.

- **Google Community Mobility Report**

The Google Community Mobility Reports show movement trends by region, across different categories of places. The Google reports chart movement trends over time by geography, across different types of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential (Google, 2020). These mobility reports are created with aggregated, anonymized data sets from users who have turned on the Google location history.

The datasets show how visits and length of stay at different places change compared to a baseline. The changes are calculated by using the same kind of aggregated and anonymized data used to show popular times for places in Google Maps. According to Google documentation, the baseline is the median value, for the corresponding day of the week, during the 5 weeks from January 3 to February 6, 2020, which is during the normal condition in almost all countries except in Hubei, China. This study uses a daily Google mobility report from February 15, 2020, to June 30, 2020.

- **Night-time light (NTL)**

The NTL data use Visible Infrared Imaging Radiometer Suite (VIIRS) Night-time Lights data. This satellite layer presents a night-time view of the earth that provides a daily informational view of our planet through satellite imagery. The U.S. Department of Commerce's NOAA National Geophysical Data Centre (NGDC) takes data from the Suomi National Polar Partnership satellite flown by NASA and NOAA satellite and performs extensive processing and includes an algorithm to produce a night-time light emission database.

We use NTL data in order to estimate countries' GDP, which has poor quality data or low statistical capacity. Therefore, the night-time light data becomes a valuable proxy for measuring economic activity. However, the NTL data also has some limitations since it cannot measure at the micro-level, especially people activities that potentially generate outputs. NTL will also change on the minimum level for a short-term period. NTL is more suitable for measuring economic changes for a more extended term period.

2) **Official Statistics Data**

- **Gross Domestic Product (GDP) and Regional Gross Domestic Product (Regional GDP)**

Official GDP and GDP at the regional level at normal conditions (pre-COVID) are used as a real initial value. Later, its value is used to forecast economic activity. In this study, we use quarterly GDP and Regional GDP data based on the value-added approach as well as industry contribution to regional GDP for each province in Indonesia.

B. Methodology

As explained before, people's mobility is closely related to economic activities. The Apple Mobility Trend Report and Google Community Mobility Report are transaction data which are used as a proxy to measure economic activities. People's mobility data can also provide timely information regarding human activities in which the information provided will reflect in nearly real-time activities. On the other hand, the NTL data can help to depict the socio-economic status of a region. This data will also be beneficial to estimate economic indicators when there is no other reliable information available, especially in countries with low statistical capacity.

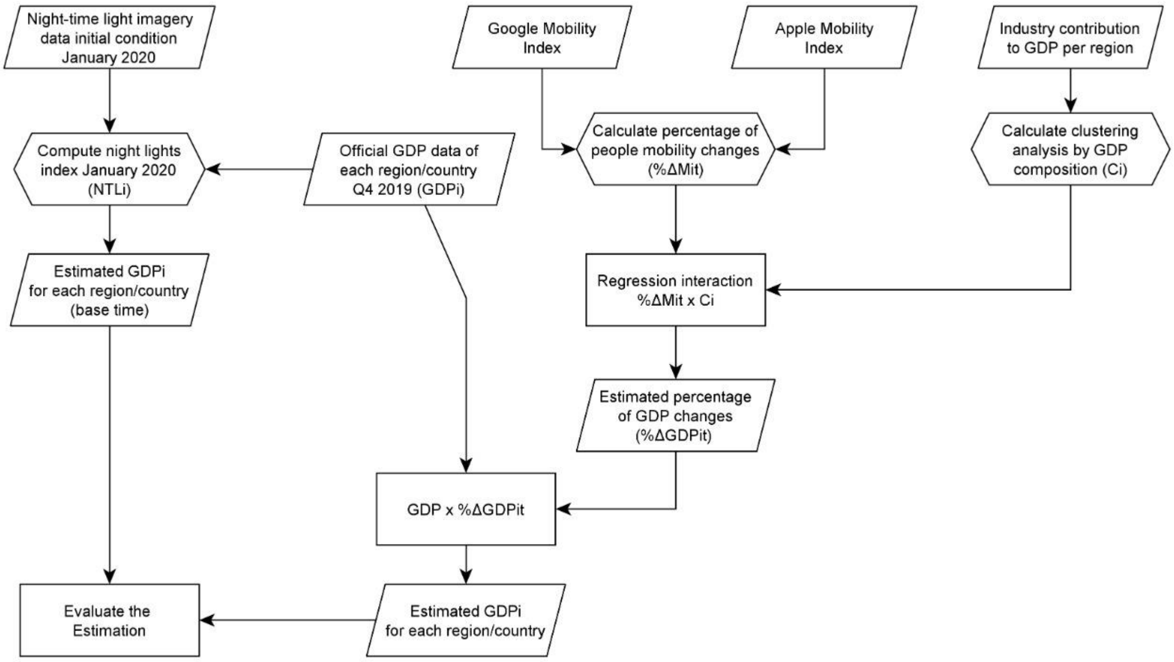


Figure 1. Flow diagram showing the step of mobility report in estimating economic growth

Initially, this study established the cluster, which represents the industry contribution per region. The cluster is involved in predicting the GDP as the interaction term with mobility. This study analyzes at the provincial level for Indonesia cases (34 provinces) using Regional GDP data in 2019 as the baseline to measure the economic activities. The regression is conducted to establish the best model to estimate the regional GDP in 2020 Q1 and Q2 using the mobility change during Q1 (January 2020 – March 2020) and Q2 (April 2020 – June 2020) at the regional level. Then, this study estimate GDP for each region based on the year on year growth. This result is used to estimate the GDP for the country level.

1) Google Mobility

The Google community mobility data already define the daily trend from several place categories, such as grocery & pharmacy, parks, transit stations, retail & recreation, residential, and workplace.

Table 1. Place Categories of Google Mobility

No	Place categories	Description
1	Grocery & Pharmacy	Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies
2	Parks	Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens
3	Transit stations	Mobility trends for places like public transport hubs, such as subway, bus, and train stations.
4	Retail & recreation	Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theatres
5	Workplaces	Mobility trends for places of work.
6	Residential	Mobility trends for places of residence.

Source: Google LLC, "Google COVID-19 Community Mobility Reports".

2) Clustering using K-means methods

This study implements K-means clustering to identify a group of provinces based on industries' contribution to regional GDP. Febri et al. (2017) state that the K-Means algorithm is a clustering algorithm into some groups based on the similarity of data features provided by determining the central value (Centroid) for each cluster or group.

This algorithm aims to divide data into K clusters so that the within-cluster sum of squares is minimized (Hartigan & Wang, 1978).

Muhammad and Adam (2017) also explain that the K-Means algorithm classifies data to the nearest neighbor use the Cosine Similarity Method to calculate the distance of the training data from the classified data. First, the algorithm determines the cluster from the data. Then, each datum that has similarities based on the distance between datum will be assigned to the cluster. After that, the new mean will be calculated for each cluster. This process will be repeated until this is no change to the cluster.

3) Regression Model

This study compared the regression model to predict the GDP using mobility. The comparison aims to evaluate the best model to predict the GDP and find the best model with better accuracy.

a. Model 1

$$Growth \mid x = \beta_0 + \beta_1\ Mobility\ change + \varepsilon$$

Model 1 only involves mobility change to predict GPD Growth. This model is established under the assumption that mobility change is related to GDP Growth.

b. Model 2

Model 2 involves mobility change and industry contribution to predict regional GDP. The industry contribution to GDP is involved with the assumption that different industry contributions within the regional GDP will respond differently to regional GDP growth caused by mobility changes. In this study, the industry contribution is indirectly involved in the model through clustering.

Model 2 aims to explain how the mobility change can be a good predictor for regional GDP growth and identify the effect of clustering in representing the industry contribution to GDP in the relationship between mobility and growth regional GDP.

<u>Full Model</u>

$$\begin{aligned} Growth \mid x = {} & \beta_0 + \beta_1\ Mobility\_change + \beta_2\ Cluster_1 + \beta_3\ Cluster_2 \\ & + \beta_4\ Cluster_3 + \beta_5\ Cluster_4 + \beta_6\ Cluster_1 * Mobility\_change \\ & + \beta_7\ Cluster_2 * Mobility\_change + \beta_8\ Cluster_3 \\ & * Mobility\_change + \beta_9\ Cluster_4 * Mobility\_change + \varepsilon \end{aligned}$$

<u>Reduced Model</u>

Cluster 1 → $Growth \mid x = (\beta_0 + \beta_2) + (\beta_1 + \beta_6)Mobility\_change + \varepsilon$

Cluster 2 → $Growth \mid x = (\beta_0 + \beta_3) + (\beta_1 + \beta_7)Mobility\_change + \varepsilon$

Cluster 3 → $Growth \mid x = (\beta_0 + \beta_4) + (\beta_1 + \beta_8)Mobility\_change + \varepsilon$

Cluster 4 → $Growth \mid x = (\beta_0 + \beta_5) + (\beta_1 + \beta_9) Mobility\_change + \varepsilon$
Cluster 5 → $Growth \mid x = (\beta_0) + (\beta_1) Mobility\_change + \varepsilon$

4) Estimated GDP

This study calculates the percentage of GDP change by calculating the change of mobility for all place categories. The percentage GDP change formula is as follow:

$$\Delta\%GDP_{it} = GDP_{i(t-1)} * Wi * \sum_{i=1}^{n} Mobility_{it}/n$$

$GDP_{i(t-1)}$	: Gross Domestic Product region-i on previous period (t-1)
Mobility	: Weighted people mobility changes during a specific time (quarterly)

The following equation calculates the estimated GDP at the regional level:

$$Est\ Regional\ GDP_{Qi,Yt} = \left(100 + \widehat{Growth}_{Qi,Yt}\right) * Regional\ GDP_{Qi,Yt-1}/100$$

The estimated regional GDP quarter-i at year (t) is calculated using the growth in quarter-i at year-t and actual regional GDP quarter-i at year (t-1). The comparison using the same quarter for the previous year to overcome the seasonality of the regional GDP each quarter. Estimated GDP and GDP Growth at the country level are calculated using the following formula.

$$Estimated\ GDP = \sum_{j=1}^{34} Est\ Regional\ GDP_j$$

$$Est\ Growth\ GDP_{Qi,Yt} = (Est\ GDP_{Qi,Yt} - Est\ GDP_{Qi,Yt-1})/Est\ GDP_{Qi,Yt-1} * 100$$

IV. Data Analysis

A. Mobility and Level of Economic Change

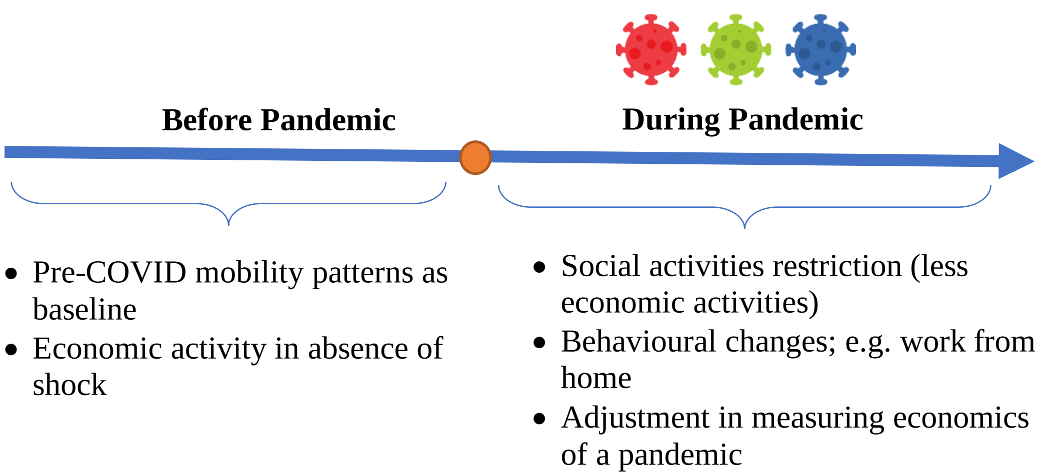
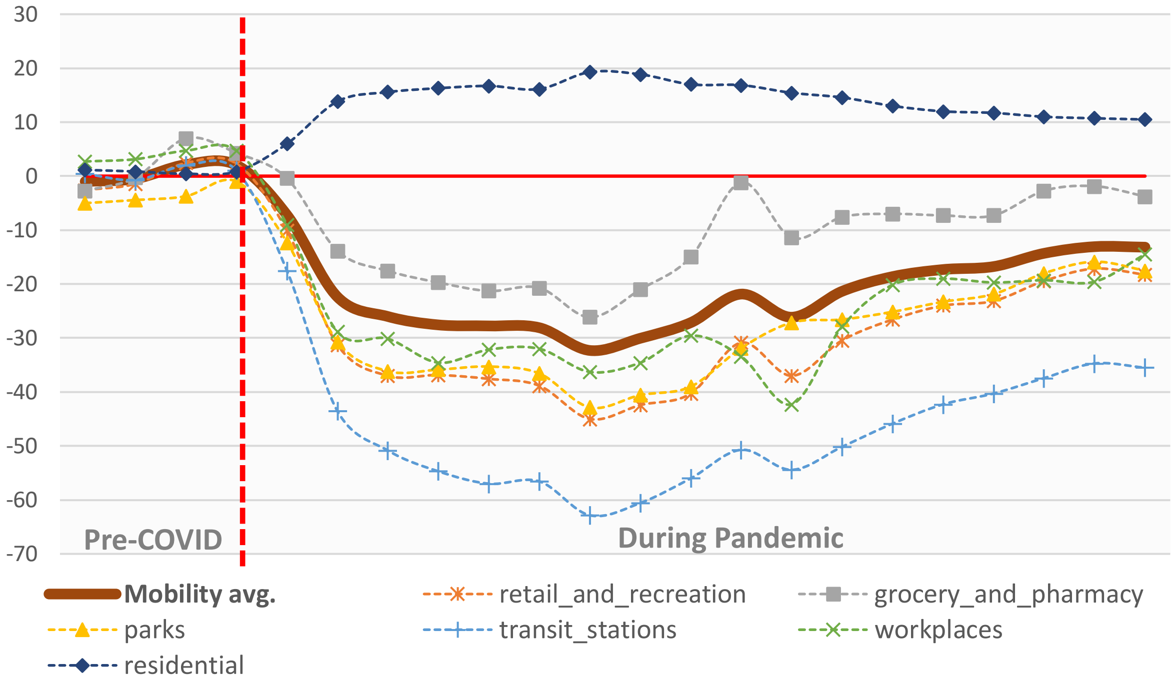


Figure 2. Mobility Change Before and During Pandemic



Mobility represents the people's activity on a daily basis. Figure 2 shows the mobility change before pandemic (Pre-COVID) and during the pandemic. Before the pandemic, society's economy level is in the normal condition shown by the mobility near the baseline. The mobility pattern before the pandemic is used as a baseline of the mobility change. In this situation, everyone can do their activity without any restrictions. When the COVID hit, government impose social activities restrictions. In this situation, people are more difficult to do their activities, which leads to a decrease in economic activity. Mobility in retail and recreation, grocery and pharmacy, parks, transit stations, and workplace decreases significantly during the pandemic. Otherwise, the activity in place of residential (home) is increasing. Additionally, some economic activities change the behavior to adapt to the pandemic, such as work from home.

Figure 3. Heat map of people mobility among 34 provinces for Retail & Recreation

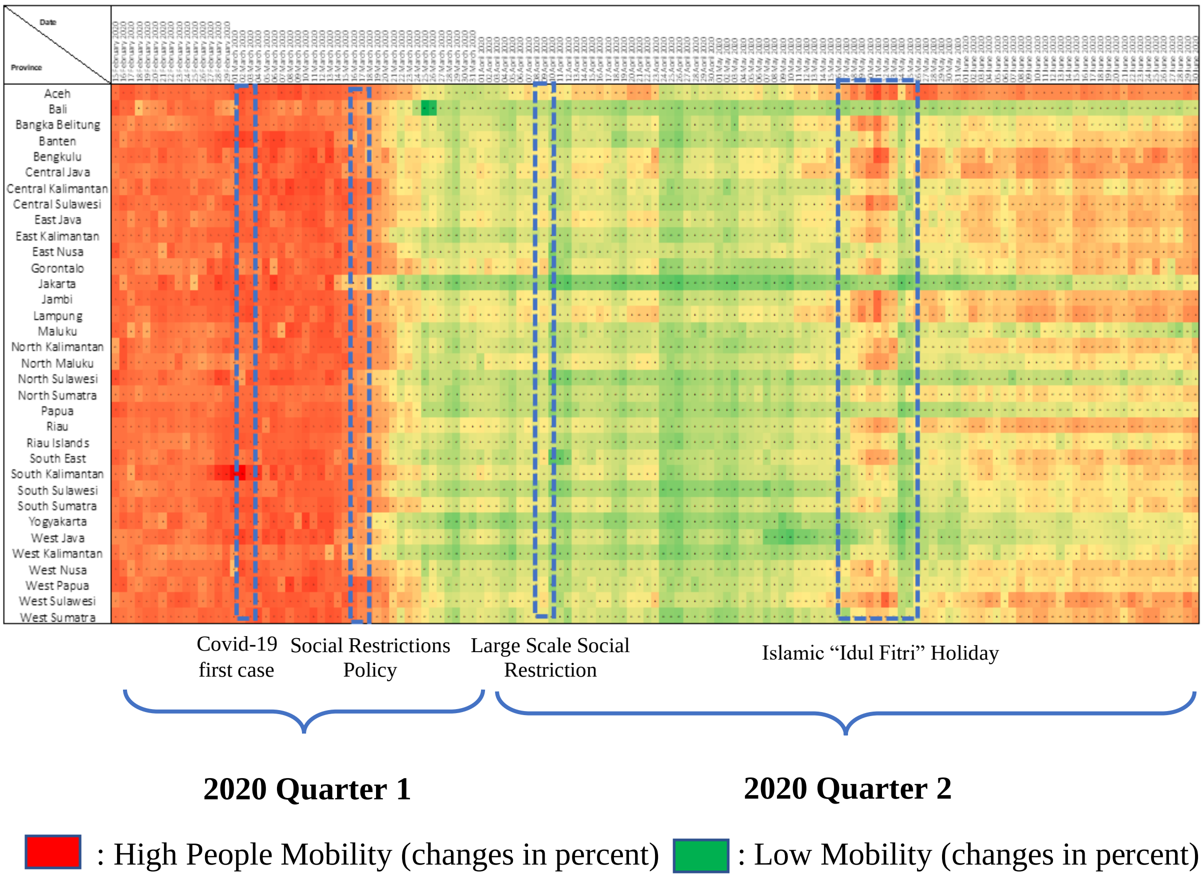


Figure 3 shows an example of people's mobility in Retail & Recreation. The level of mobility in 34 provinces in Indonesia drops significantly after the social restriction policy. However, during the public holiday (Idul Fitri) people's mobility is slightly increasing before it drops again in the next following days.

B. Cluster Analysis

Indonesia consists of 34 provinces with various contributions of industries to its GDP. Indeed, each province's different characteristics regarding the industry contribution to regional GDP will respond differently to its economic growth. Industries that closely relate to mobility, such as transportation and trade, will suffer if the government implement social restriction (lower mobility). Hence, cluster analysis is conducted to eliminate bias caused by the variety of industry contributions to regional GDP share among regions. The cluster is established based on the industry contribution to regional GDP to create a homogeneous group. This method will group regions that have similar industry contribution's composition to regional GDP into the same group.

The cluster analysis uses K-means, which find each province's similarity in terms of industry contribution to regional GDP. After considering some parameters, 5 clusters are established in which each group has a different number of provinces, in detail shown by the table in figure 4.

Figure 4. Cluster analysis of regions based on industry contribution to regional GDP

Cluster	Number of provinces
I	12
II	3
III	5
IV	2
V	12

C. Regression Model

1) Mobility to predict the Regional GDP Growth without considering industry contribution to Regional GDP

Figure 5. Regression model result without differing industry contribution to Regional GDP

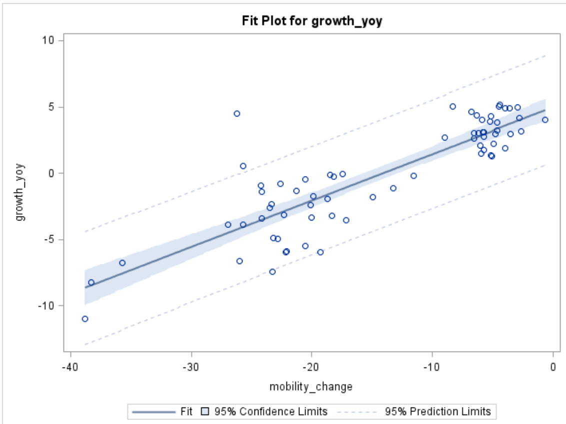
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	792.774564	792.774564	191.97	<.0001
Error	66	272.562186	4.129730		
Corrected Total	67	1065.336750			

R-Square	Coeff Var	Root MSE	growth_yoy Mean
0.744154	-3108.837	2.032174	-0.065368

Source	DF	Type I SS	Mean Square	F Value	Pr > F
mobility_change	1	792.7745636	792.7745636	191.97	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
mobility_change	1	792.7745636	792.7745636	191.97	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	4.971420989	0.43918681	11.32	<.0001	4.094556236	5.848285742
mobility_change	0.351484335	0.02536833	13.86	<.0001	0.300834821	0.402133849



Pearson Correlation Coefficients, N = 68 Prob > |r| under H0: Rho=0		
	growth_yoy	est_growth_yoy
growth_yoy growth_yoy	1.00000	0.86264 <.0001
est_growth_yoy	0.86264 <.0001	1.00000

The regression model explains the relationship between Mobility and Regional GDP Growth without considering industry contributions to Regional GDP. The results show that the mobility significantly associated with the Growth of Regional GDP (p-value < 0.05) with R-square 0.7442, which means the model explains 74,42% of its variability. Based on the model, the estimated growth values are compared with the actual Regional GDP. The result shows that the correlation between estimated and actual growth is 0.86. Considering the R-square and the correlation, this study concludes that the model is good for prediction, but it still needs improvement. Therefore, the industry contribution is included in the model to enhance the model.

2) Mobility to predict regional GDP by including the industry compositions.

Figure 6. Regression model result with clustering

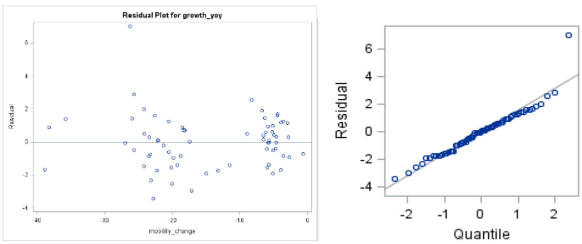
Dependent Variable: growth_yoy growth_yoy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	897.258613	99.695401	34.40	<.0001
Error	58	168.078136	2.897899		
Corrected Total	67	1065.336750			

R-Square	Coeff Var	Root MSE	growth_yoy Mean
0.842230	-2604.226	1.702322	-0.065368

Source	DF	Type I SS	Mean Square	F Value	Pr > F
mobility_change	1	792.7745636	792.7745636	273.57	<.0001
cluster_sector	4	65.2575914	16.3143978	5.63	0.0007
mobility_*cluster_se	4	39.2264584	9.8066146	3.38	0.0148

Parameter	Estimate		Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	4.653314153	B	0.63554235	7.32	<.0001	3.381137910	5.925490397
mobility_change	0.271992663	B	0.03897653	6.98	<.0001	0.193972658	0.350012668
cluster_sector 1	0.339476801	B	0.88880159	0.38	0.7039	-1.439652878	2.118606481
cluster_sector 2	1.063900229	B	1.57429883	0.68	0.5019	-2.087401266	4.215201724
cluster_sector 3	1.065658425	B	1.22535400	0.87	0.3881	-1.387154110	3.518470961
cluster_sector 4	-0.693055099	B	1.78559417	-0.39	0.6993	-4.267310169	2.881199970
cluster_sector 5	0.000000000	B	.	.	.	.	.
mobility_*cluster_se 1	0.054519174	B	0.05576468	0.98	0.3323	-0.057105965	0.166144313
mobility_*cluster_se 2	0.115040416	B	0.06487092	1.77	0.0814	-0.014812848	0.244893679
mobility_*cluster_se 3	0.259399178	B	0.07434276	3.49	0.0009	0.110585968	0.408212388
mobility_*cluster_se 4	0.158074885	B	0.12376913	1.28	0.2066	-0.089675948	0.405825717
mobility_*cluster_se 5	0.000000000	B	.	.	.	.	.



Pearson Correlation Coefficients, N = 68
Prob > |r| under H0: Rho=0

	growth_yoy	est_growth_yoy
growth_yoy growth_yoy	1.00000	0.91773 <.0001
est_growth_yoy	0.91773 <.0001	1.00000

The results show the model is fit, and the model can explain nearly 84,22% of the total variability of the response. In addition, the residuals are satisfied, which met the homogeneity assumption and normal distribution.

The results show that mobility changes are significantly associated with Regional GDP growth. Additionally, the industry cluster representing the industry contribution to GDP also affects the relationship between mobility change and regional GDP growth. It differs from the magnitude of the association between mobility and GDP Growth. To compare the result, we correlate the estimated regional GDP growth with the actual regional GDP growth, and the correlation between estimated growth and actual growth is 0.92.

By comparing model 1 and model 2, this study concludes that considering the industry contribution to regional GDP can improve the regional GDP variability. The model without involving the industry contribution to regional GDP explain 74,42%, and it increases to 84,22% after adding the interaction variable. Furthermore, it improves accuracy, which shows by the increase of the correlation between estimated growth to actual growth, from 0.86 to 0.92. So, this study found that the model that involves the cluster representing the industry contribution to GDP is better to predict GDP growth than without considering those factors. The model to estimate economic growth is as follows:

- **Cluster 1** (Province assigned in cluster 1 follow the equation):

$$\widehat{Growth} = 4.99 + 0.32\, Mobility\_change + \varepsilon$$

- **Cluster 2** (Province assigned in cluster 2 follow the equation):

$$\widehat{Growth} = 5.71 + 0.39\, Mobility\_change + \varepsilon$$

- **Cluster 3** (Province assigned in cluster 3 follow the equation):

$$\widehat{Growth} = 5.72 + 0.53\, Mobility\_change + \varepsilon$$

- **Cluster 4** (Province assigned in cluster 4 follow the equation):

$$\widehat{Growth} = 3.96 + 0.43\, Mobility\_change + \varepsilon$$

- **Cluster 5** (Province assigned in cluster 5 follow the equation):

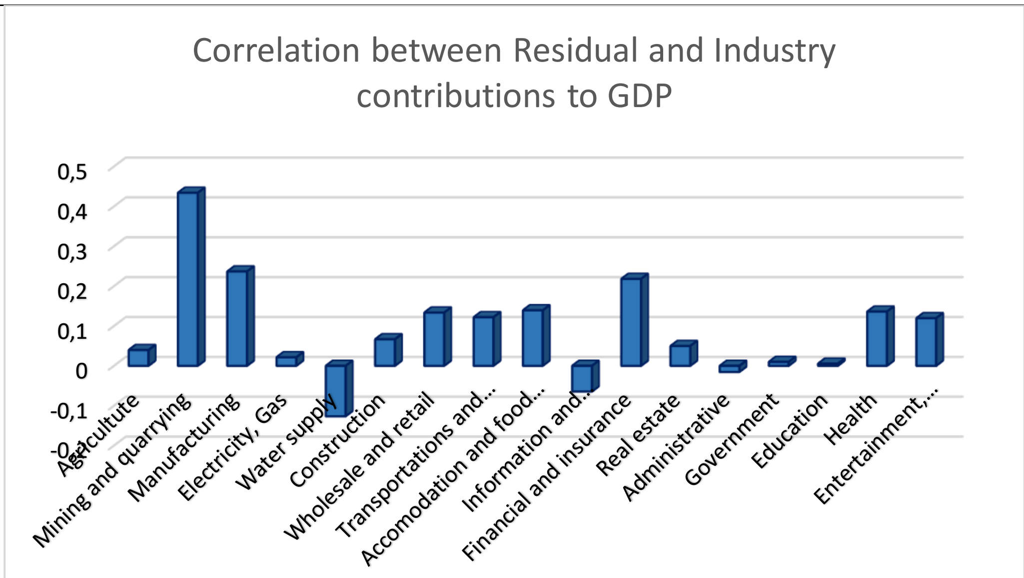
$$\widehat{Growth} = 4.65 + 0.27\, Mobility\_change + \varepsilon$$

For each province assigned in Cluster 1, each one percent increase of average mobility will increase 0.32 percent of growth, and vice versa, the decreasing of one percent of average mobility will drop 0.32 percent of growth. For each province assigned in Cluster 2, each one percent increasing of average mobility will increase 0.39 percent of growth and vice versa; the decreasing of one percent of average mobility will decrease 0.39 percent of growth. For each province assigned in Cluster 3, each one percent increase of average mobility will increase by 0.53 percent of growth, and vice versa, the decreasing of one percent of average mobility will decrease by 0.53 percent of growth. For each province assigned in Cluster 4, each one percent increase of average mobility will increase 0.43 percent of growth, and vice versa. The decreasing of one percent of average mobility will decrease 0.43 percent of growth. For each province assigned in Cluster 5, each one percent increase of average mobility will increase by 0.27 percent of growth, and vice versa, the decreasing of one percent of average mobility will decrease by 0.27 percent.

D. Residual Evaluation

The error of the model causes the gap/difference between the actual growth with the estimated growth, called residual. It might randomly occur or for some reason. Therefore, we try to identify the gap between actual growth with estimated growth.

Figure 7. Evaluation of residual model and industry contribution



Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-0.09767	0.17611	-0.55	0.5811
sector_B	sector_B	1	0.40862	0.10438	3.91	0.0002

Pearson Correlation Coefficients, N = 68 Prob > |r| under H0: Rho=0		
	residual_sector	sector_B
residual_sector residual_sector	1.00000	0.43409 0.0002
sector_B sector_B	0.43409 0.0002	1.00000

The correlation between residual data and industry contribution to GDP shows that Mining and Quarrying have the highest value compared to other industries. In addition, this industry is the only industry that has a significant correlation with residual data. In other words, the contribution from Mining and Quarrying to GDP Growth is associated with the error from the model.

In some regions, such as East Kalimantan, the mining and quarrying industry has a high impact and contribution to their economic growth. Still, this industry is less affected by people's mobility change. The mining and quarrying are mostly operated by big business and not directly affected by social mobility. Since Indonesia does not impose lockdown regulation to reduce the spread of COVID-19, the Mining and Quarrying activities are still operated during the pandemic.

E. Estimated Regional GDP, Estimated GDP, and Estimated GDP Growth

Figure 8. Evaluation of estimated regional GDP and official regional GDP Growth

Pearson Correlation Coefficients, N = 68
Prob > |r| under H0: Rho=0

	GDRP	EST_GDRP
GDRP GDRP	1.00000	0.99992 <.0001
EST_GDRP EST_GDRP	0.99992 <.0001	1.00000

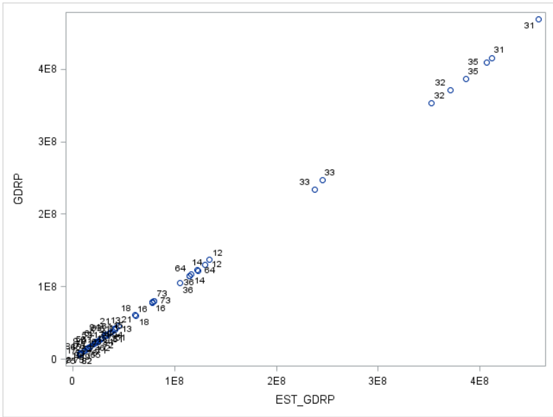


Figure 8 shows the correlation between the estimated regional GDP to actual regional GDP. The correlation between both values is high.

Figure 9. Mobility and Weighted GDP Growth

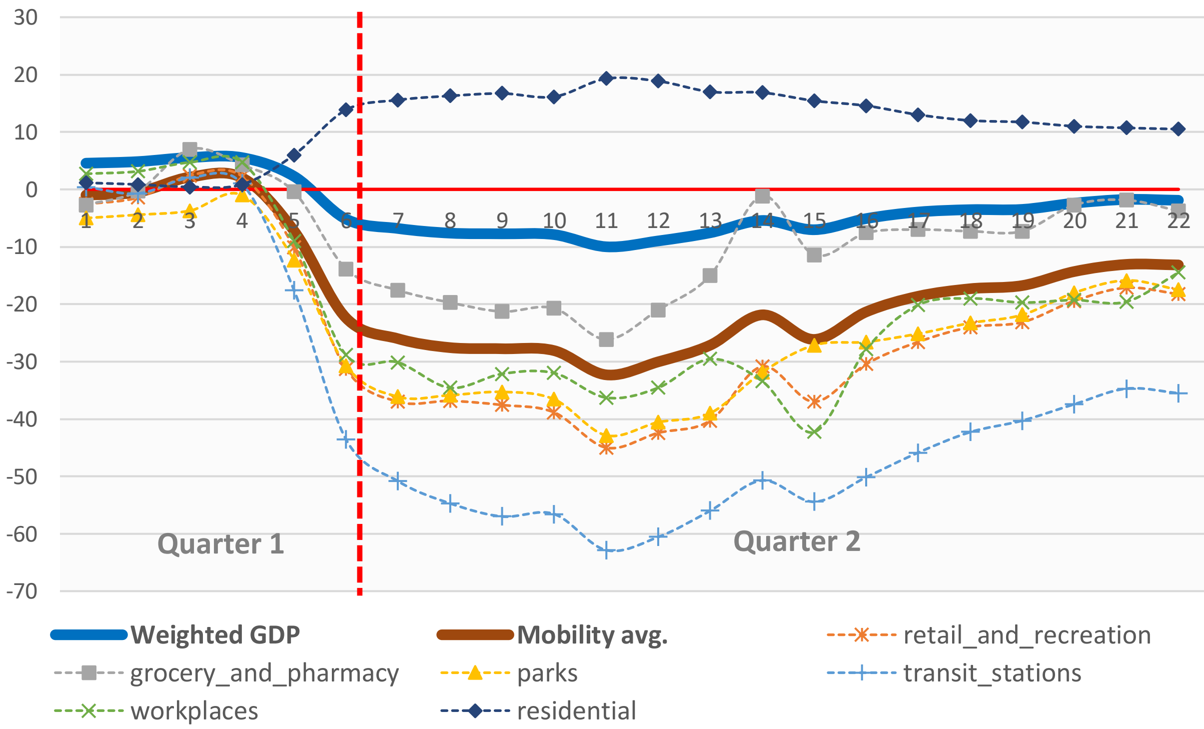


Figure 9 shows weekly mobility change and GDP weighted average growth using the model. The relationship between mobility and economic output can be seen in that figure. The figure shows that the average mobility change has a similar pattern with estimated GDP change. In quarter 1, the mobility tends to be stable, but it decreases in the last two weeks near the end of the Q1 period due to the implementation of social restriction policy. During the second semester, the level of mobility change is negative, and it is also followed by negative estimated GDP growth.

Finally, based on our estimation, the Indonesian national GDP growth in 2020 quarter 1 is 2.64, while the actual GDP growth in the same period is 2.97. In addition, the estimated Indonesia GDP growth in 2020 quarter 2 is -5.17, while the actual growth GDP in 2020 quarter 1 is -5.36.

Table 1. Comparison of Actual and Estimated GDP Growth in Indonesia

Time	Official GDP Growth	Estimated GDP Growth
2020 Quarter 1	2.97	2.64
2020 Quarter 2	-5.36	-5.17

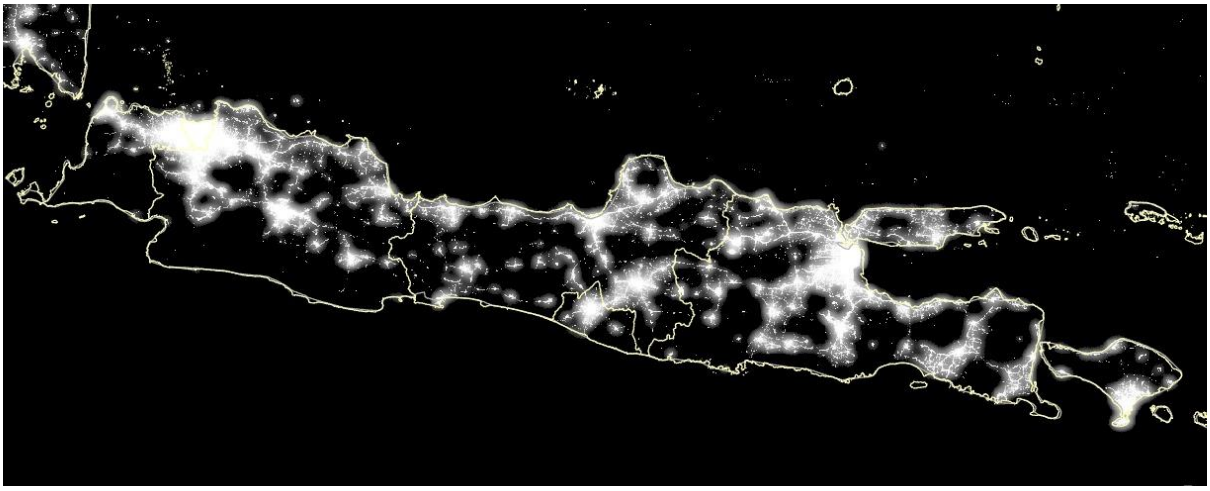
F. Evaluation of Nightime Light Data

We use NTL data as a proxy to estimate a countries' GDP, which has poor quality data or low statistical capacity. Therefore, the night-time light data becomes a valuable proxy for measuring economic activity.

Figure 10. World Nightime Light during base time

Source: The Earth Observations Group (EOG) NOAA - Visible Infrared Imaging Radiometer Suite (VIIRS)

Figure 11. Java Island - Indonesia Nightime Light during base time



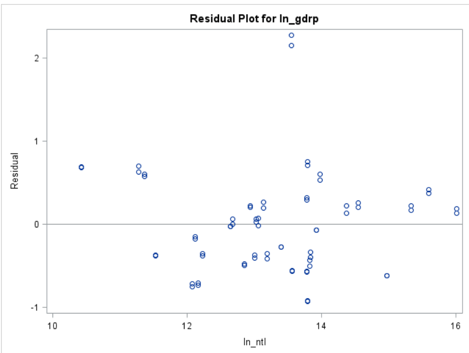
Source: The Earth Observations Group (EOG) NOAA - Visible Infrared Imaging Radiometer Suite (VIIRS)

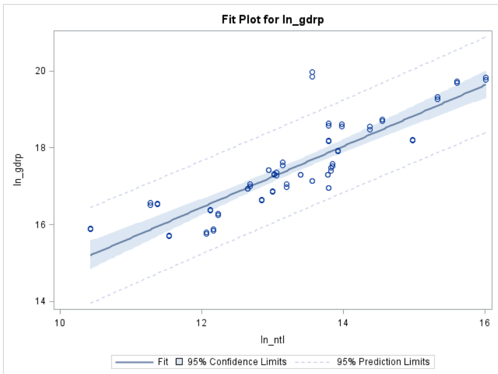
Figure 12. Regression model of NTL and Regional GDP

R-Square	Coeff Var	Root MSE	ln_gdrp Mean
0.731630	3.414919	0.596750	17.47479

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ln_ntl	1	64.07445746	64.07445746	179.93	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	6.909598172	0.79095628	8.74	<.0001	5.330403014	8.488793330
ln_ntl	0.795100545	0.05927504	13.41	<.0001	0.676754109	0.913446981



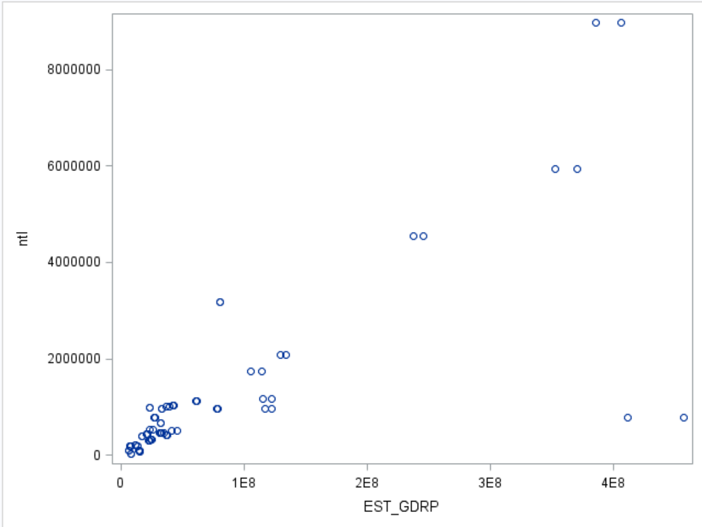


The evidence shows that NTL has associated with regional GDP and explains 73,16 of the regional GDP variation. The increase of 1 percent point of NTL is expected to lead the increment of regional GDP as 0.06 percentage. There are outliers in the data, DKI Jakarta. It needs to be observed since DKI Jakarta is the capital city of Indonesia and also the center of economic activity, while the area of DKI Jakarta is the smallest province in Indonesia. In addition, Ghosh et al. (2010) also found the outlier on the relationship of NTL in small islands and city-states.

Figure 13. Correlation between NTL and Estimated Regional GDP

Pearson Correlation Coefficients, N = 68
Prob > |r| under H0: Rho=0

	EST_GDRP	ntl
EST_GDRP EST_GDRP	1.00000	0.76474 <.0001
ntl ntl	0.76474 <.0001	1.00000



Some research has been examined the usage of NTL to predict the GDP for countries with low statistical data quality. Henderson et al. (2012) explain that NTL represents proxy GDP in low-income countries with low-quality national account data. Bickenbach et al. (2016) state that the control variable, which accounts for the regional characteristics, is needed to stabilize the relationship between night light and GDP. Figure 13 shows that the NTL and estimated GDP using mobility have a strong enough correlation. The comparison supports the result of the usage of mobility to predict the GDP.

V. Result and Conclusion

Social restrictions policy has change people's mobility behavior in Indonesia. The policy is aimed to reduce the virus's spread, which is begun on March 2020, two weeks before the end of Quarter 1. Since that time, people’s activity in the place of residence has increased while decreasing in other places, such as workplaces, retail, parks, and recreation.

Indonesia has 34 provinces with various industry contributions to regional GDP. Industries that closely relate to mobility, such as transportation and trade, will suffer if the government implement social restriction (lower mobility). However, some industries are not too affected by mobility, for example, Mining and Quarrying. Those industries are still operated during the pandemic, but with lower capacity.

Cluster analysis assigned each province to a group based on the similarity on the industry contribution to regional GDP to eliminate bias due to mobility change effects on the industry. This study found that involving the industry contribution through cluster increases the explanatory power of mobility for regional GDP. It also provides a better fit model to estimate the regional GDP, which increases the accuracy of estimated national GDP. In addition, this study found that the industry contribution to regional GDP provides different effects on the relationship between mobility change and regional GDP growth.

This study finds a consistently strong positive correlation between regional change on the average mobility report and the percentage changes of regional GDP in 34 provinces in Indonesia. The Indonesia national GDP growth in 2020 quarter 1 is 2.64%, while the actual GDP growth in the same period is 2.97%. Additionally, the estimated Indonesia GDP growth in 2020 quarter 2 is -5.17%, while the real growth GDP in 2020 quarter 1 is -5.36%.

According to the official release of GDP in 2020 Quarter 1, the Indonesian economy based on the amount of Gross Domestic Product (GDP) based on constant prices in the first quarter of 2020 reached Rp 2,755.71 trillion, whereas our finding is Rp 2,739.65 trillion. In Quarter 2, the Indonesian economy based on the amount of Gross Domestic Product (GDP) based on constant prices is Rp 2,599.15 trillion, whereas our finding is approximately Rp 2,601.65 trillion. The potential difference is due to the timeliness of measurement during the end of Q1. The COVID-19 effect on people's mobility was begun in the last two weeks before the end of Q1. Hence, the official GDP release tends to be overestimated than actual economic conditions since it probably cannot measure the changes in the last time just before GDP Q1 2020 was released. On the other hand, in quarter 2, people's mobility experiences a slight increase during the Islamic Idul Fitri holiday, which leads to our finding that our estimation is higher than the real value.

The night-time light data in Indonesia has good associated with regional GDP. However, there is an outlier in DKI Jakarta as the smallest area in Indonesia, the center of economic activities. Therefore, the NTL data is useful for predicting the regional GDP, especially for regions with a low quality of National account data. This data can adjust the baseline GDP to make sure comparability aspects among countries. In addition, the estimated GDP using mobility has strong correlation with NTL. It supports the usage of mobility to predict the GDP for regions with a low quality National account data.

VI. Future Works

Mobility potentially has a different effect on GDP growth among countries, such as underdeveloped, developing, and developed countries. The incidence of lockdown and social restriction measures will vary by the country that will inevitably affect people's mobility. The effect of mobility declining on economic output potentially different among countries since it depends on their economic structure and the industry contribution to GDP. For instance, countries that rely on labor-intensive industries requiring high people mobility may have a greater impact on social restriction policy.

In addition, the country's formal and informal workers' composition will also determine the degree of the effect of mobility. On the one hand, formal sectors can still produce output with lesser mobility, such as working from home. On the other hand, some jobs may require movement and hard to be done remotely.

References

Apple. (2020). *Apple Mobility Trend Report*.

Bickenbach, F., Bode, E., Nunnenkamp, P., & Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, *152*(2), 425–447. https://doi.org/10.1007/s10290-016-0246-0

Buono, D., Luigi Mazzi, G., Kapetanios, G., & Marcellino, M. (2017). Big data types for macroeconomic nowcasting. *Eurona*, *1*(July), 93–145. https://ec.europa.eu/eurostat/cros/system/files/euronaissue1-2017-art4.pdf

Canaks, C., Garcia-Arenas, J., Jimeno, E., Camarasa, P. (2020). The COVID-19 dilemma : mobility and economy. *International Economy, June*, 16–17.

CCSA. (2020). How COVID-19 is changing the world : a statistical perspective. *Committee for the Coordination of Statistical Activities*, 1--90.

Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C., & Wu, H. (2017). Measuring economic activity in China with mobile big data. *EPJ Data Science*, *6*(1). https://doi.org/10.1140/epjds/s13688-017-0125-5

Febri, B. A., Teguh E. P. & Adhistya. (2017). Comparison of the Decision Tree Method and K-Means Clustering in Overcoming the Cold-Start Problem of New Users. *Department of Electrical Engineering and Information Technology, FT UGM, Yogyakarta*.

Ferrara, L., & Simoni, A. (2019). When Are Google Data Useful to Nowcast Gdp? An Approach Via Pre-Selection and Shrinkage. *SSRN Electronic Journal*, *April*. https://doi.org/10.2139/ssrn.3370917

Ghosh, T., Powell, R., Elvidge, C., Baugh, K., Sutton, P., Anderson, S. (2010). Shedding Light on the Global Distribution of Economic Activity. *The Open Geography Journal*, *3*(1), 148–1661. https://doi.org/10.2174/1874923201003010147

Google. (2020). *Google Community Mobility Reports*.

Guidotti, R., Coscia, M., Pedreschi, D., & Pennacchioli, D. (2016). Going beyond GDP to nowcast Well-Being using retail market data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9564*, 29–42. https://doi.org/10.1007/978-3-319-28361-6_3

Hartigan J & Wang M. (1978) A K-means clustering algorithm. Appl Stat. 1978;28:100-108.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

IMF. (2020). *8th IMF Statistical Forum: Measuring the Economics of a Pandemic*.

Jiang, M., Kim, E., & Woo, Y. (2020). The relationship between economic growth and air pollution—a regional comparison between China and South Korea. *International Journal of Environmental Research and Public Health*, *17*(8), 11–14. https://doi.org/10.3390/ijerph17082761

Kang, H., Suh, H., & Yu, J. (2019). Does Air Pollution Affect Consumption Behavior? Evidence from Korean Retail Sales*. *Asian Economic Journal*, *33*(3), 235–251.

https://doi.org/10.1111/asej.12185

Li, B., Gao, S., Liang, Y., Kang, Y., Prestby, T., Gao, Y., & Xiao, R. (2020). Estimation of Regional Economic Development Indicator from Transportation Network Analytics. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-59505-2

Li, S., Zhang, T., Yang, Z., Li, X., & Xu, H. (2017). Night time light satellite data for evaluating the socioeconomics in Central Asia. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2W7), 1237–1243. https://doi.org/10.5194/isprs-archives-XLII-2-W7-1237-2017

Moyer, B., & Dunn, A. (2020). Measuring the Gross Domestic Product (GDP): The Ultimate Data Science Project. *Harvard Data Science Review*, *2*(1), 1–9. https://doi.org/10.1162/99608f92.414caadb

Muhammad, R., & Adam, M. B. (2017) Implementation of the Closest KNeighbor Algorithm in Classifying Twitter Followers using Bahasa Indonesia. *Journal of Information Systems.* Vol. 13, no. 1, pp. 31-37.

Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., & Giannotti, F. (2016). An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, *2*(1–2), 75–92. https://doi.org/10.1007/s41060-016-0013-2

Sampi Bravo, J. R. E., & Jooste, C. (2020). *Nowcasting Economic Activity in Times of COVID-19 : An Approximation from the Google Community Mobility Report* (Vol. 9247, Issue May).

WHO. (2020). *WHO Coronavirus Disease (COVID-19) Dashboard*.