



# Correlation between Covid-19 Spreading and Vaccinations in Tangerang City

Yusuf Sudiyono<sup>1</sup>, Harjanto Prabowo<sup>2</sup>, Agung Trisetarso<sup>3</sup>, Meyliana<sup>4</sup>

<sup>1</sup>Department of Computer Science, Bina Nusantara University, Jakarta, Indonesia

<sup>2</sup>Computer Science Management Department - BINUS Graduate Program, Bina Nusantara University, Jakarta, Indonesia

<sup>3</sup>Mathematics and Statistics Department, Bina Nusantara University, Jakarta, Indonesia

<sup>4</sup>Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia

\*Correspondence to: Yusuf Sudiyono

*(Received: 16 May 2025*

*Revised: 20 June 2025*

*Accepted: 24 July 2025)*

## KEYWORDS

Clustering,  
Covid-19,  
Vaccination  
s, Tangerang  
City

## ABSTRACT:

The Covid-19 outbreak, which began in December 2019, rapidly developed into a global disaster impacting almost every region worldwide. The transmission caused by Covid-19 has caused many casualties that occurred in all affected countries. To find out in depth the pattern of the spread of Covid-19, research was conducted to analyse data starting during the pandemic period to the endemic period that occurred in Tangerang City. The clustering analysis process in this study is based on the data on the spread of Covid-19 with confirmed cases in 13 districts combined with data on the age category of Covid-19 patients. From the clustering results, it produces a pattern of the spread of Covid-19 that occurs in each of the closest districts with the largest number of population affected by Covid-19. In addition, clustering based on age categories results in people of productive age with the age categories of early adulthood and late adulthood entering into the age category that contributes the highest number of confirmed Covid-19 cases. Therefore, by conducting clustering analysis, it can facilitate the vaccination of confirmed Covid-19 cases in each district with the highest cases being prioritised for vaccination. The results of the multiple correlation analysis can be concluded that there is a very strong relationship between the vaccination of a population in each region and the increase in Covid-19 confirmed cases. So that by giving vaccinations that are right on target in an area can reduce confirmed cases of Covid-19 in the Tangerang City area

## 1. Introduction

The Covid-19 outbreak began in December 2019, initially affecting several countries before spreading globally. The virus was thought to have originated from the wildlife trade in the Wuhan market, China [1]. Covid-19 spread directly or indirectly between people (through contaminated objects or surfaces), or through close contact with an infected person via nasal or oral secretions [2], [3]. These secretions include saliva, respiratory secretions, or droplets (splashes) of secretions. These secretions come out of the mouth or nose, for example when an infected person coughs, sneezes, talks, or sings. People within close proximity (1

meter) of an infected person can be exposed to Covid-19 when infectious droplets enter their mouth, nose or eyes [4].

In this study, the data on the spread of Covid-19 in confirmed cases that occurred in 13 districts of Tangerang City were analysed and classified based on cases per phase [5]. The case model per phase was then carried out a clustering analysis process in each region for confirmed Covid-19 cases based on the age category of Covid-19 patients. Thus, the pattern of the spread of Covid-19 in each district and the level of public awareness during the pandemic using the clustering model were obtained.



Furthermore, by conducting a clustering analysis process, it is obtained in each district in the form of a Covid-19 distribution cluster. Therefore, the highest Covid-19 confirmed case in each district is one of the steps in tackling the spread of Covid-19 cases by providing vaccinations. Where the spread of confirmed Covid-19 cases occurs in each region with the highest cases being a priority in vaccination.

The latest Covid-19 cases in April 2025 in several Asian regions have occurred again, such as Singapore, Bangkok, and Hong Kong. The Covid-19 cases that have occurred in Singapore have infected and recorded 14,200 cases in the week from April 27 to May 3, 2025. The Straits Times in a report, Monday, May 19, 2025, stated that the number increased from the previous week, which was recorded at 11,000 cases [6]. The Singapore Ministry of Health estimates that the spike in cases occurred due to a decrease in herd immunity due to the reduced number of vaccine recipients, which can affect immunity in a group. On May 15, 2025, 30 deaths were recorded due to Covid-19 in Hong Kong. Epidemiologists found that 83 percent of severe cases affected patients over the age of 65.

Since mid-February 2025, confirmed cases of Covid-19 globally have increased again by 11%. This spike primarily occurred in the Eastern Mediterranean, Southeast Asia, and the Western Pacific [7]. Therefore, in this study, preventive measures must be implemented, aimed at mitigating and conducting an in-depth analysis of indicators that may influence the increase in Covid-19 cases in a region. Although Covid-19 cases have entered the endemic period, vigilance and prevention efforts must remain a top priority in facing a surge in cases that will occur in the future.

## 1. Study Phases

### 2.1 Defining objectives

In this study, the author applies quantitative data types that are used as support in completing this paper. The definition and type of data used in this research object is quantitative data containing data on Covid-19 patients and the provision of vaccinations in a certain period of time, namely from January 2020 to June 2023. This type of quantitative data refers to results in the form of numbers or figures so that they can be measured mathematically [8].

This analysis was conducted comprehensively over a specific time period in each sub-district, combined with vaccination data to evaluate vaccination effectiveness and serve as a basis for learning and planning for response if similar Covid-19 cases recur in the future. The data in this study came from the Covid-19 patient monitoring and vaccination information system database managed by the Tangerang City Communication and Informatics Office [9].

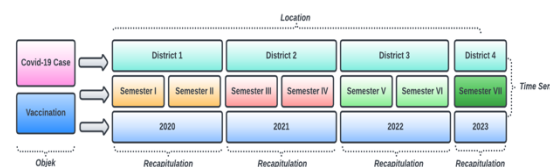


Figure 1. Types of Study Data

### 2.2 Clustering

Clustering is one of the Unsupervised Learning methods that aims to group data based on the similarity/distance between data. Clustering has a characteristic where members in one cluster have the same similarity or very close distance, while members between clusters have very different similarities or very far distance. According to the book entitled Introduction to Data Mining, clustering methods are divided into two types, namely Hierarchical Clustering and Partitional Clustering [10].

The purpose of clustering in general, both hierarchical and partitional clusterings is to create clusters that have the same characteristics within a cluster member and have different characteristics between clusters. This concept requires the cluster creation process to pay attention to the distance (dissimilarity) of all data to each cluster center. The most commonly used dissimilarity calculation methods are Euclidean distance and Manhattan distance, but it is possible to use other distance measurements, depending on the data we are analysing [11].

### 2.3 K-Means

K-means clustering is one of the unsupervised learning algorithms that belongs to the non-hierarchical cluster analysis used to group data based on variables or features. K-Means is the most popular clustering algorithm and is often used in various applications such as data analysis, pattern recognition, and data clustering. This algorithm works by dividing a dataset into k groups based on the distance between objects in the dataset.



Groups are formed for data that have similar characteristics, while data with different characteristics are grouped separately. The main goal is to minimise the variation within groups, so that the data within a group is highly similar [12]. The following steps can be followed to perform clustering using the K-Means technique [13]:

Determine the desired number of groups ( $k$ ).

Initialise the cluster centre (centroid).

The objects or data will be grouped into the closest cluster.

To calculate the distance between each data and the centroid, the Euclidean distance formula is used. By using this formula, we can find the shortest distance between each data and the centroid. The following is the first equation of the Euclidean Distance formula.

Description:

$D(i, j)$  = Distance between data at points  $i$  and  $j$

$X_{ki}$  =  $i^{\text{th}}$  data  $k^{\text{th}}$  data attribute

$X_{kj}$  = The centroid data to  $k$  point on the data attribute to  $j$ .

Formula:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

A recalculation is performed to determine the group centre by considering the current group members. The group centre is calculated by taking the average of all data/objects that belong to a particular group, using equation 2 as follows:

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk}) \quad (2)$$

Repeat step number 3 if there is a change in the cluster center.

## 2.4 Case per Phase Analysis

Case per Phase data analysis in this study is to find the cause of the high spread of Covid-19 cases that occur in each region and to determine the correlation that occurs between the factors that cause the increase in confirmed cases. Figure 4 shows the data analysis process carried

out to determine the spread of Covid-19 based on case per phase. The first phase, shows a period of 6 months, then in the second phase, it is accumulated again 6 months ahead with the previous stage, namely the first phase. The process continues until the seventh phase [5]. In each phase, the age category factor is analysed to find out how the level of public awareness is based on age category in the increase of Covid-19 cases in Tangerang City.

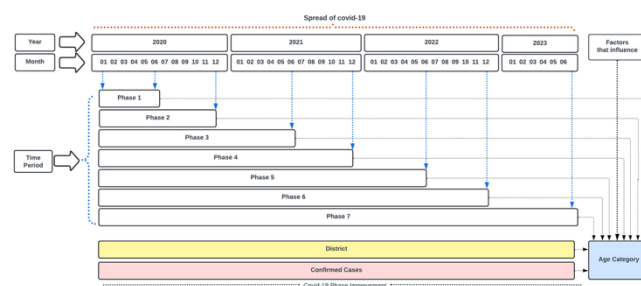


Figure 2. Classification per Phase Scheme of Covid-19 Data Analysis

The next step is to combine case data by phase using clustering analysis, which aims to determine the distribution pattern of confirmed Covid-19 cases. This enables the tracking of public awareness levels by age group in each sub-district in Tangerang City, which will be discussed in the next section on research methodology.

## 2.6 Covid-19 Multiple Correlation Analysis Method

The multiple correlation model on Covid-19 confirmed cases aims to analyse the factors of increasing Covid-19 confirmed cases in an area. In an increase in the number of Covid-19 confirmed cases, the higher the population in an area, the higher the number of confirmed cases; as Covid-19 can be transmitted through direct contact with sputum droplets from infected people (through coughing and sneezing), and if touching surfaces contaminated with the virus. People within close proximity (1 metre) of an infected person can be exposed to Covid-19 when infectious droplets enter their mouth, nose or eyes [4]. This is certainly what causes most populations to have the highest number of confirmed cases.

Thus, the distribution and administration of vaccinations need to be carried out evenly and thoroughly and re-analysed to see how the role of vaccination is in suppressing the number of confirmed Covid-19. In



addition, it is also to measure how the correlation of vaccination in an area is appropriate or not and as an evaluation material and preventive action if the Covid-19 virus pandemic period occurs again in the future.

### 2.2.1 Multiple correlation analysis

Multiple correlation analysis aims to determine the degree of correlation between the independent variable and the dependent variable. Multiple correlation is a value that gives the strength of the influence or correlation of two or more variables together with other variables [14].

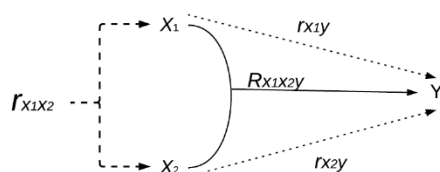


Figure 4. Correlation between independent variables (X1, X2) and a dependent variable (Y).

Description:

$r_{X1X2}$  = multiple correlation coefficient.

$X1$  = independent variables  $X1.s$

$X2$  = independent variable  $X2.$

$r_{X1Y}$  = correlation coefficient between  $X1$  and  $Y.$

$r_{X2Y}$  = correlation coefficient between  $X2$  and  $Y.$

$R_{X1X2Y}$  = correlation coefficient between  $X1, X2,$  and  $Y.$

$Y$  = Dependent variable.

### 2.2.2 Determination Analysis

Determination analysis is used to determine how much variation in the relationship between profitability, solvency, liquidity and firm value variables is able to explain variations in stock return variables. The coefficient of determination is known from the  $R^2$  value in percentage.

### 2.2.3 F test

The F test is used to test the significance of the effect of the independent variables simultaneously on the dependent variable, using a level of significance / degree of confidence of 95% or an error rate of 5% ( $\alpha = 0.05$ ). To determine the level of significance is done by

comparing sig. F with  $\alpha = 0.05$ . If the sig.  $F < 0.05$  then the independent variables simultaneously have a significant effect on the dependent variable. On the contrary, if the sig.  $F > 0.05$  level, the independent variables simultaneously do not have a significant effect on the dependent variable.

### 2.2.4 T-test

The t test is used to test the significance of the effect of the independent variable partially on the dependent variable, using a level of significance / degree of confidence of 95% or an error rate of 5% ( $\alpha = 0.05$ ). To determine the level of significance is done by comparing sig. t with  $\alpha = 0.05$ . If the sig. t level  $< 0.05$  then the independent variable simultaneously has a significant effect on the dependent variable. On the contrary, if the sig.  $t > 0.05$ , the independent variable partially does not have a significant effect on the dependent variable.

## 2.7 Age Categories

According to the Ministry of Health, a person's age is categorised into several levels, which definitely have been taken into account beforehand. The age limits of children have also been determined in the law. Therefore, the recording of children's boundaries is indeed intended so that in providing education, attention, and others, the handling will be more appropriate [15], [16].

Toddlerhood: 0-5 Years old

Childhood: 5-11 Years old

Early Adolescence: 12-16 Years old

Late Adolescence: 17-25 Years old

Early Adulthood: 26-35 Years old

Late Adulthood: 36-45 Years old

Early Elderly : Period 46-55 Years old

Late Elderly : Period : 56-65 Years old

Seniors: > 65 Years old

## 3. Methodology Research

### 3.1 Flow Diagram of the Research

The method used to achieve the research objectives is shown in figure 5 below. This study begins with a literature review of the cases to be examined by analysing the increase in Covid-19 cases. This is done to



analyse data on Covid-19 cases that have occurred during the pandemic and until the endemic. Thus the function of control, management, and evaluation in each district area in the event of similar outbreak disaster case can be carried out quickly and measurably.

The next stage is to identify the problems in this research on the object of research, namely the increase in Covid-19 confirmed cases in each district, the pattern of distribution of Covid-19 cases that occurred in each region and how the level of public awareness was during the pandemic. After the problem identification is determined, the next step is to prepare the data collection process in this research, including data on the distribution of confirmed Covid-19 cases that are integrated with the population and civil registration database and geospatial data of Tangerang city. Each case that appears is classified into several phases at the level of cases that occur within the time span per semester. Therefore, further analysis can be done later by understanding the pattern of the spread of Covid-19 cases that occur in each region.

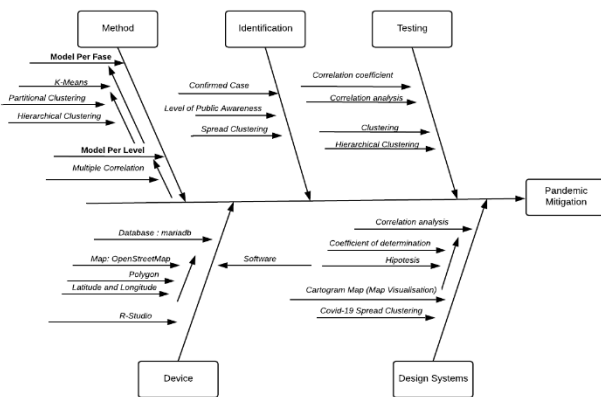


Figure 5. Fishbone Diagram Research Methodology.

### 3.2 Case per Phase K-means Clustering Analysis

The K-Means Clustering model in this study aims to determine how the severity and level of public awareness are of the spread of Covid-19 that occurs in each area of Tangerang City through the clusters formed. The cluster is done by grouping the number of confirmed cases in each district based on the severity that occurs and the level of risk of the spread of Covid-19.

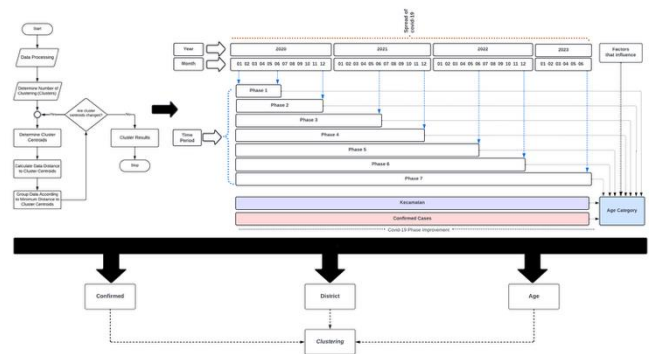


Figure 6. Clustering Image of Confirmed Cases Per Phase Based on Age.

Furthermore, in each district, researchers conducted an analysis model process using clustering based on confirmed Covid-19 cases with age category variables per each phase. Therefore, the clustering results aim to find out how the spread of Covid-19 cases is that have occurred in each region in each district. Thus, areas that experience the highest cases with the highest risk of severity based on certain age ranges can be prioritised in dealing with confirmed Covid-19 cases.

#### 3.2.1 Age Category Clustering (Phase I)

Figure 7 below is graphical information that shows the optimum number of clusters with the elbow method. In the elbow method, it can be seen that a line that experiences a break that forms an elbow at  $k = 2$ .

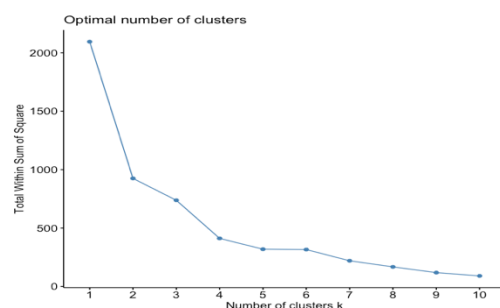


Figure 7. Elbow Method Graph (Phase I).

In this first phase is the initial period of the spread of confirmed Covid-19 cases from the results of clustering division based on age categories, namely there are 2 clustering groups in each district. In the clustering group, confirmed Covid-19 cases are divided into medium and low cases in each district.

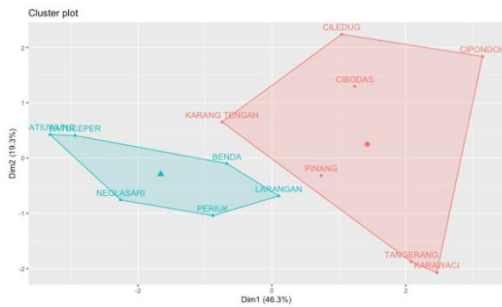


Figure 8. Cluster Results (a) k = 2 (Phase I).

In Figure 9, there is a division of clustering groups consisting of 3 groups. In cluster 3, the highest number of cases by age category occurred in the Cipondoh, Karawaci, Tangerang, Ciledug, Cibodas, Pinang and Karang Tengah districts.

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final.cl
TANGERANG	1	0	2	17	22	12	18	7	6	
PINANG	0	1	2	9	21	17	11	7	4	
KARAWACI	0	0	1	16	19	18	15	13	7	
KARANG TENGAH	1	2	0	9	15	16	9	7	0	
CIPONDH	2	3	3	12	28	17	18	16	4	
CILEDUG	3	3	2	8	20	18	12	11	2	
CIBODAS	5	1	2	11	24	13	10	11	2	
PERIUK	1	1	0	11	9	7	13	10	2	
NEELASARI	3	0	0	4	10	5	2	5	3	
LARANGAN	1	2	1	11	13	8	15	6	5	
JATTIUNG	0	2	1	4	4	6	1	3	0	
BENDA	1	1	3	11	5	10	14	4	1	
BATUKEPER	0	2	2	0	6	3	5	3	2	

Figure 9. Cluster Members (Phase I).

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final.cl
1	1.71	1.43	1.71	11.7	21.3	15.9	13.3	10.3	3.57	
2	1	1.33	1.17	6.83	7.83	5.67	8.35	5.17	2.17	

Figure 10. Average Case Variable Cluster (Phase II).

While in the average division results, it can be seen that the occurrence of Covid-19 cases in the first highest order occurred to people who have an early adult age category, which is around 26 to 35 years old. Then, in the second highest order occurred in the late adult age category aged between 36 to 45 years old. Furthermore, in the third and fourth orders, namely those who have the age category of early elderly aged between 46 to 55 years old and late adolescence aged 17 to 25 years old.

### 3.2.2 Age Category Clustering (Phase II)

In phase 2, the best cluster division is displayed on the information graph showing the optimum number of clusters with the elbow method. In the elbow method, it can be seen that a line is broken that forms an elbow at k=3.

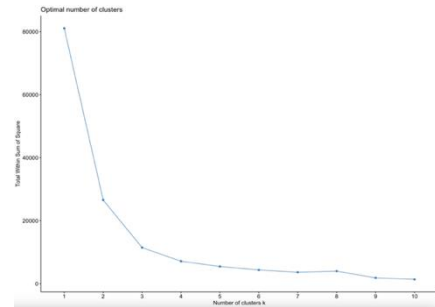


Figure 11. Elbow Method Graph (Phase II).

Entering the second phase, the spread of confirmed Covid-19 cases began to spread to several areas and resulted in the division of clustering based on age categories into 3 groups in each district. Each clustering group, divides Covid-19 confirmed cases, namely with the highest, medium and lowest cases as shown in Figure 12.

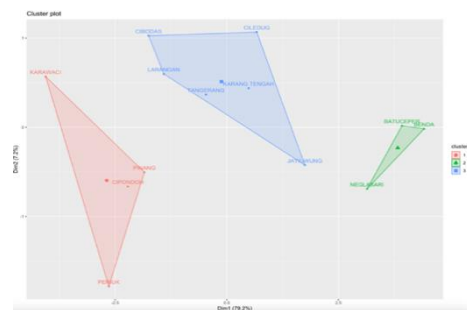


Figure 12. Elbow Method Graph (Phase II).

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final.cl
PINANG	10	11	22	85	139	121	106	72	16	1
PERIUK	6	16	24	103	132	126	151	91	17	1
KARAWACI	13	21	37	96	130	139	132	85	28	1
CIPONDH	10	21	21	100	132	107	125	65	11	1
NEELASARI	3	6	9	43	35	34	34	33	4	2
BENDA	4	6	3	12	18	23	21	7	0	2
BATUKEPER	4	4	3	11	29	29	30	21	6	2
TANGERANG	10	8	14	88	98	76	87	57	23	3
LARANGAN	10	15	25	82	120	91	87	50	23	3
KARANG TENGAH	11	7	6	56	97	76	63	59	14	3
JATTIUNG	7	10	3	50	73	73	82	19	5	3
CILEDUG	10	16	14	38	89	56	68	32	14	3
CIBODAS	8	23	19	65	112	107	84	65	32	3

Figure 13. Cluster Members (Phase II).

In the second phase clustering based on the age category, as shown in Figure 13 above, the group division of each clustering is divided into 3 clusters. The spread of Covid-19 began to spread in several areas massively. The 1<sup>st</sup> cluster group shows the highest occurrence of cases based on age categories in the Pinang, Cipondoh, Periuk and Karawaci districts.



Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors
1	9.75	17.2	26	96	134	123	128	78.2	18
2	3.67	5.33	5	22	27.3	28.7	28.3	20.3	3.3
3	9.33	13.2	13.5	61.8	98.2	79.8	78.5	47	18.5

Figure 14. Average Cases of Cluster Variables (Phase II).

After displaying the clustering members, the next is the result of dividing the average in each cluster group. Therefore, it can be seen that the first highest Covid-19 case is experienced by people based on the early adult age category, namely those aged around 26 to 35 years old, then the second is the early elderly category aged between 46 to 55 years old. Next is the late adult age category, which is between 36 and 45 years old and finally late adolescence aged 17 to 25 years old.

### 3.2.3 Age Category Clustering (Phase III)

Moreover, in the third phase, the best cluster distribution shows k=2 as shown in Figure 15. The formation of the cluster is explained based on graph information that shows the optimum number of clusters on a line that experiences a break that forms an elbow.

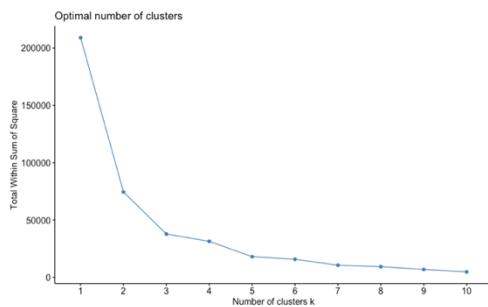


Figure 15. Elbow Method Graph (Phase III).

After the determination of the optimum cluster with the elbow method is determined in this third phase, the next step is to visually display the clustering division of Covid-19 confirmed cases based on the age category which is divided into 3 cluster groups in each district. In each clustering group, the number of confirmed cases is divided into the highest, medium and low cases as shown in Figure 16 below.

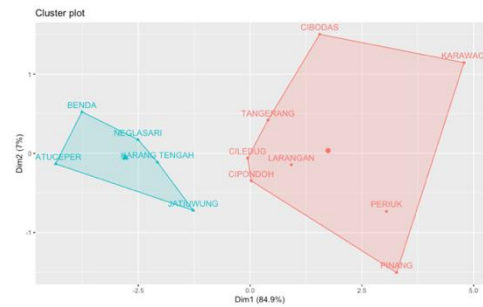


Figure 16. Cluster Results (a) k = 2 (Phase III).

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final_cluster
TANGERANG	12	23	39	147	207	198	139	100	39	1
PINANG	35	37	42	199	303	211	218	140	35	1
PERIUK	24	34	45	200	258	212	252	149	42	1
LARANGAN	17	28	35	158	258	141	158	120	42	1
KARAWACI	29	59	62	189	288	200	199	152	76	1
CIPONDOKH	13	21	27	167	226	154	150	95	33	1
CILEDUG	17	28	33	157	184	112	142	94	34	1
CIBODAS	16	35	40	158	182	188	158	122	74	1
NEGLASARI	7	15	22	117	115	79	102	58	21	2
KARANG TENGAH	12	18	14	92	173	109	92	73	25	2
JATEUNG	19	20	27	107	159	152	126	60	11	2
BENDA	4	12	25	73	75	82	56	31	9	2
BATUCEPER	9	9	7	57	84	67	58	24	8	2

Figure 17. Cluster Members (Phase III).

The formation of clustering in the third phase as shown in Figure 18 explains the division of groups from each district as many as 2 clusters. In cluster 1 is the highest occurrence of cases based on age categories that occur in the Pinang, Karawaci, Larangan, Cipondoh, Periuk, Tangerang, and Ciledug districts.

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors
1	20.4	33.1	40.4	172	238	184	177	122	46.9
2	10.2	14.8	19	89.2	121	97.8	85.2	49.2	14.8

Figure 18. Average Cases of Cluster Variables

Then in the results of the average division of the third phase, it can be seen that the occurrence of the first highest Covid-19 confirmed case occurred in people who had an early adult age category aged between 26 to 35 years old. Furthermore, the second order occurs in the late adult age category aged between 36 to 45 years old. Then, in the third and fourth orders occurred in the early elderly age category aged between 46 to 55 years old and late adolescence aged between 17 to 25 years old

### 3.2.4 Age Category Clustering (Phase VI)

Entering the fourth phase, the formation of the optimum number of clusters formed by the elbow method is described in Figure 19. The graph information shows the optimum number of clusters with the appearance of a line that experiences a break that forms an elbow at k = 2.

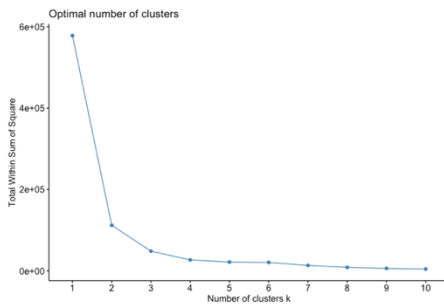


Figure 19. Elbow Method Graph (Phase IV).

After determining the optimum number of clusters in the fourth phase, the next step is to visually display each cluster group, then divide the confirmed cases with the highest and medium cases. This transformation in Covid-19 confirmed cases occurs because the spread of Covid-19 cases has begun to spread evenly in each district as shown in Figure 20.

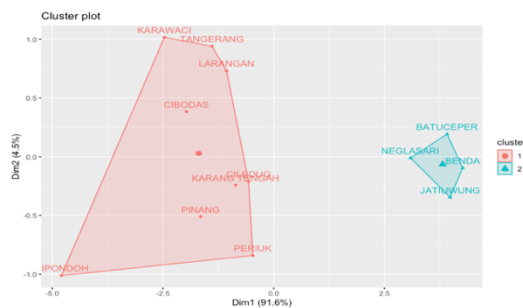


Figure 20. Cluster Results (a) k = 2 (Phase IV).

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final.clu
TANGERANG	43	64	76	221	355	340	246	167	141	
PINANG	59	75	83	234	402	315	242	190	78	
PERBUK	47	63	70	247	336	243	237	168	51	
LARANGAN	40	57	55	230	371	329	226	140	101	
KARAWACI	61	64	71	262	418	356	284	216	151	
KARANG TENGAH	54	82	80	182	307	279	204	182	86	
CIPONDOH	77	101	131	365	493	428	363	233	110	
CILEDUG	43	48	81	209	343	288	236	173	75	
CIBODAS	54	76	82	242	385	323	249	207	123	
NEGLASARI	16	26	33	123	207	135	108	66	41	
JATITUWUR	16	6	26	104	159	122	101	40	12	
BENDA	15	16	17	74	146	92	73	49	19	
BATUCEPER	15	14	18	94	143	123	93	49	36	

Figure 21. Cluster Members (Phase IV).

Furthermore, in the fourth phase, in Figure 21 above it can be seen the division of groups from each clustering as many as 2 clusters. In cluster 1, the highest number of cases by age category occurred in the Cipondoh, Karawaci, Pinang, Cibodas, Tangerang, Larangan, Ciledug, Periuk and Karang Tengah districts.

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seni
1	53.1	70	81	244	378	321	254	196	11
2	15.5	15.5	23.5	198.8	164	118	193.8	51	

Figure 22. Average Cases of Cluster Variables (Phase IV).

### 3.2.5 Age Category Clustering (Phase V)

While in the results of the average division in the fourth phase, it can be seen that the occurrence of confirmed Covid-19 cases in the first order is more experienced by people who have an early adult age category, which is around 26 to 35 years old. Then the second order occurs in the late adult age category aged between 36 to 45 years old. Furthermore, the third order occurs in the early elderly age category aged between 46 to 55 years old and finally the late adolescent age category which has an age between 17 to 25 years old.

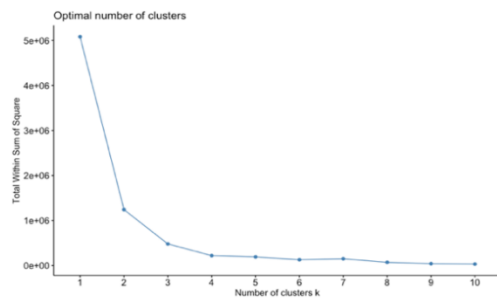


Figure 23. Elbow Method Graph (Phase V).

In the fifth phase, the distribution of confirmed Covid-19 cases from the clustering results based on the age category, namely there are 2 clustering groups in each district. In each clustering group, the confirmed cases are divided into the highest cases and medium cases. The spread of Covid-19 cases that occurred began to spread evenly in each district that has the closest regional distance as illustrated in Figure 24.

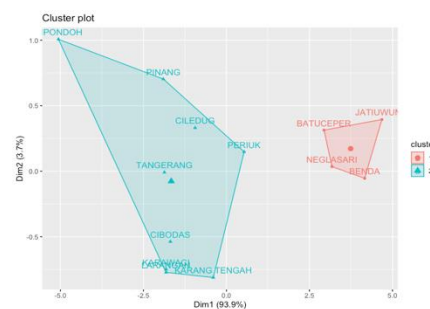


Figure 24. Cluster Results (a) k = 2 (Phase V).



	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final	cluster
TANGERANG	169	188	208	867	1197	910	526	341	164	1	
PINANG	189	205	236	781	1264	927	540	315	125	1	
PERLUK	120	97	162	697	877	542	427	269	103	1	
LARANGAN	175	169	217	648	1277	936	534	333	210	1	
KARAWACI	179	168	190	833	1282	831	464	364	203	1	
KARANG TENGAH	139	124	171	653	1050	639	421	322	178	1	
CIPONDOK	266	267	305	1140	1788	1180	824	436	170	1	
CILEDUG	150	173	189	696	1179	798	490	319	121	1	
CIBODAS	140	191	207	778	1119	821	526	308	192	1	
NEGLASARI	51	80	72	448	579	325	185	114	68	2	
JATTIUNG	35	39	37	230	321	184	134	61	16	2	
BENDIA	30	35	41	299	354	256	182	101	50	2	
BATUCEPER	71	88	100	336	499	370	242	138	55	2	

Figure 25. Cluster Members (Phase V).

In the fifth phase as shown in Figure 24 above, the group division of each clustering consists of 2 clusters. In cluster 1, the highest number of cases by age category occurred in the Cipondoh, Karawaci, Larangan, Pinang, Tangerang, Ciledug, Cibodas, Karang Tengah and Periuk districts.

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	senior
1	170.	176.	209.	788.	1225.	846.	527.	340.	163.
2	46.8	60.5	62.5	328.	438.	284.	181.	104.	47.

Figure 26. Average Cases of Cluster Variables (Phase V).

While in the results of the average division in this fifth phase, it can be seen that the occurrence of Covid-19 confirmed cases in the first highest order is more experienced by people who have an early adult age category aged around 26 to 35 years old. Furthermore, the second highest case occurred in the late adult age category aged between 36 to 45 years old, then followed by the case below occurred in the early elderly age category aged between 46 to 55 years old and finally in the late adolescent age category aged between 17 to 25 years old.

3.2.6 Age Category Clustering (Phase VI)

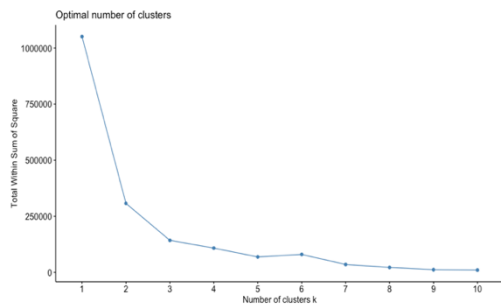


Figure 27. Elbow Method Graph (Phase VI).

Furthermore, in the sixth phase, cluster formation with the elbow method obtained graph information showing the optimal number of clusters as k=2. The formation of

the elbow is seen in a line that forms a fault forming an elbow as shown in Figure 27.

During the sixth phase, the distribution of confirmed Covid-19 cases from the results of clustering division based on age category resulted in 2 clustering groups in each district. In each clustering group, the confirmed cases are divided into the highest cases and medium cases. Changes in Covid-19 confirmed cases occur due to the spread of Covid-19 cases that are already high and begin to spread massively in each district as shown in figure 28 below.

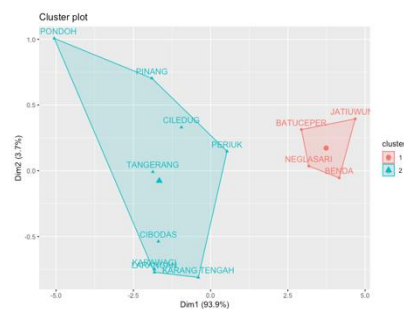


Figure 28. Cluster Results (a) k=2 (Phase VI).

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final	cluster
TANGERANG	86	117	131	322	583	442	286	170	126	1	
PINANG	90	109	128	243	512	338	214	114	61	1	
LARANGAN	76	124	103	194	491	432	251	153	148	1	
KARAWACI	87	102	99	262	473	354	217	180	133	1	
KARANG TENGAH	50	67	89	195	452	305	196	128	101	1	
CIPONDOK	112	139	187	430	742	542	357	175	92	1	
CILEDUG	43	59	65	176	423	264	189	103	55	1	
CIBODAS	50	95	121	272	372	332	238	149	108	1	
PERLUK	37	54	56	189	327	163	141	79	41	2	
NEGLASARI	32	16	24	102	197	95	60	50	46	2	
JATTIUNG	19	10	14	64	97	47	41	30	9	2	
BENDIA	25	20	23	106	180	120	76	58	31	2	
BATUCEPER	28	38	42	115	184	129	81	64	30	2	

Figure 29. Cluster Members (Phase VI).

Next, in the sixth phase in figure 29 above, it can be seen that the group division of each clustering formed consists of 2 clusters. The first cluster is the highest occurrence of cases based on age category in the Cipondoh, Tangerang, Pinang, Larangan, Karawaci, Karang Tengah, Ciledug, and Cibodas districts.

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors
1	74.2	104	115.	262.	506.	376.	241.	146.	103
2	28.2	27.6	31.8	115.	197	111.	79.8	56.2	31.4

Figure 30. Average Cases of Cluster Variables (Phase VI).



In the results of the average distribution of clustering variables in the sixth phase, it can be seen that the first highest occurrence of confirmed Covid-19 cases is experienced by people who have an early adult age category, which is around 26 to 35 years old. Then the second order occurs in the late adult age category, which is between 36 to 45 years old. Furthermore, the third order occurs in the early elderly age category between 46 to 55 years old and the last is late adolescence between 17 to 25 years old.

3.2.7 Age Category Clustering (Phase VII)

Following, the last phase in this study is the seventh phase, cluster formation with the elbow method has obtained graph information showing the optimal number of clusters as many as 2 clusters (k=2). The formation of the elbow is seen in a line that forms a fault forming an elbow as shown in Figure 31 below.

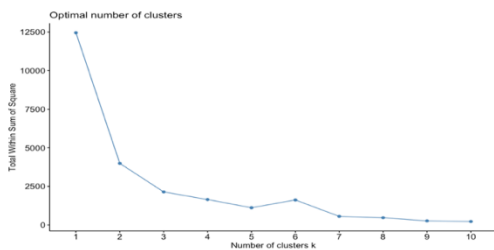


Figure 31. Elbow Method Graph (Phase VII).

During the seventh phase as shown in Figure 32 below, the formation of clustering on the distribution of confirmed Covid-19 cases based on the age category resulted in 2 cluster groups in each district. In each cluster group, the Covid-19 confirmed cases are divided into 2 clusters, namely the highest cases and medium cases. This change in Covid-19 confirmed cases occurs because the spread of Covid-19 cases has begun to spread evenly in each district.

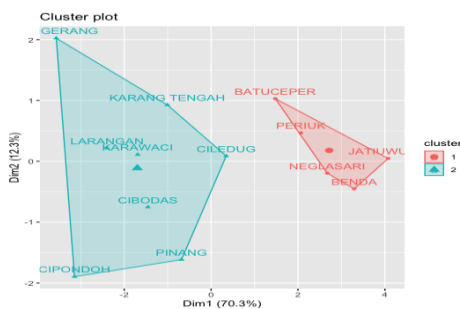


Figure 32. Cluster Results (a) k= 2 (Phase VII).

In the next clustering results in the seventh phase as shown in Figure 33 below, it can be seen that the group division of each clustering consists of 2 clusters. In cluster 1, the highest occurrence of cases based on age category occurs in the Cipondoh, Larangan, Pinang, Karawaci, Tangerang, Ciledug, Cibodas and Karang Tengah districts.

	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors	final_cluster
TANGERANG	17	5	5	15	47	48	37	28	15	1
PINANG	5	0	1	21	52	25	22	12	1	1
LARANGAN	6	5	4	24	56	30	37	18	10	1
KARAWACI	10	2	6	24	50	28	15	20	12	1
KARANG TENGAH	8	3	5	13	37	27	15	24	11	1
CIPONDOK	5	2	5	26	86	48	35	22	12	1
CILEDUG	9	2	2	16	40	26	15	11	5	1
CIBODAS	9	1	5	26	39	41	30	22	7	1
PERIUK	6	1	1	6	24	12	10	7	7	2
NEGLASARI	2	1	0	12	11	7	8	11	4	2
JATUWUL	0	0	2	2	4	2	3	2	1	2
BENKA	2	0	0	10	13	6	8	2	4	2
BATUCEPER	6	2	1	8	19	9	10	9	13	2

Figure 33. Cluster Members (Phase VII).

Cluster	toddlerhood	childhood	early_adolescence	late_adolescence	early_adulthood	late_adulthood	early_elderly	late_elderly	seniors
1	8.62	2.5	4.12	20.6	50.9	34.9	22.8	20.9	11.2
2	3.2	0.8	0.8	7.6	14.2	7.2	7.8	6.2	5.8

Figure 34. Average Cases of Cluster Variables (Phase VII).

In the seventh phase, the results of the average division of cluster variables can be seen that the occurrence of confirmed cases of Covid-19 with the first highest case is experienced by people who have an early adult age category, which is around 26 to 35 years old. Then the second order occurs in the late adult age category, which is between 36 to 45 years old. Furthermore, the third order occurs in the early elderly age category between 46 to 55 years old and the last is the late elderly aged between 56 to 65 years old as shown in Figure 34 above.

3.3 Dendrogram Diagram of Phase I to VII

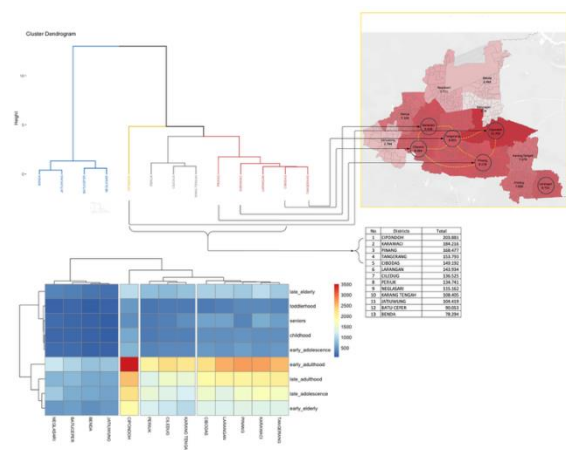


Figure 35. Dendrogram Diagram of Phases I to VII.



After the clustering analysis process is carried out, the next step is grouping with a dendrogram diagram as in Figure 35, which produces 2 cluster groupings. In cluster 1 there are 4 districts and cluster 2 consists of 9 districts. In the first cluster is the division of clusters in districts with the lowest confirmed cases of Covid-19 consisting of 4 districts, namely Batuaceper, Benda, Jatiuwung and Neglasari. Meanwhile, the cluster with the highest cases occurs in the 2<sup>nd</sup> cluster which consists of 3 sub-clusters. In the 1<sup>st</sup> sub cluster, the district with the highest case is Cipondoh district, followed by the 2<sup>nd</sup> sub cluster, which consists of 5 districts, namely Karawaci, Pinang, Tangerang, Cibodas and Larangan. Meanwhile, the 3<sup>rd</sup> sub cluster is the sub cluster with the lowest number of cases of the other 2 sub clusters, namely Karang Tengah, Ciledug and Periuk districts.

### 3.4 Correlation Structure of Confirmed Cases to Vaccination in Per-level Cases

The next stage is the correlation model of vaccination cases per level, which explains the stages of conducting in-depth correlation analysis of the variables involved in it. The variable process consists of 3 variables including 2 Independent (X) and Dependent (Y) variables.

The Independent variable consists of population size and number of vaccinations while the Dependent variable contains the number of confirmed cases. The independent variable population size (X1) contains 13 districts in the Tangerang City area. Meanwhile, the independent variable number of vaccinations (X2) has been conducted in 13 districts of Tangerang City.

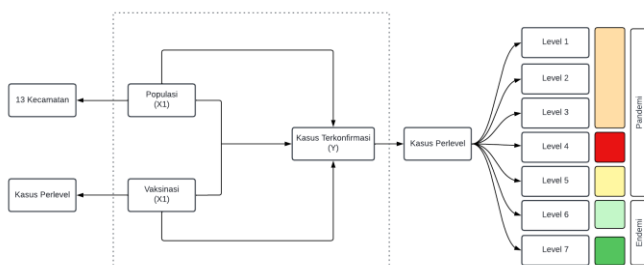


Figure 35. Correlation Structure of Confirmed Cases to Vaccination in Perlevel Cases.

Furthermore, for each variable, the calculation process is carried out between variables to determine the relationship of how much influence and correlation that occurs between variables. These variables such as the number of vaccinations given in an area with confirmed

Covid-19 cases that occur in a district based on the level of cases per level. At each level, it can be seen how high or dangerous the cases that occur to cause the highest death cases.

### 3.4.1 Covid-19 Confirmed Cases and Level 3 Vaccinations

In the table below, we can see data on the distribution of vaccinations in 13 districts with confirmed Covid-19 cases in semester 3. Vaccination was carried out in 2021 in 13 districts in Tangerang City. At this stage, the calculation process is carried out to find the correlation between each region and the Covid-19 cases that have occurred and also to determine the role of vaccination in suppressing the increase in the number of confirmed Covid-19 cases in 13 districts of Tangerang City.

In table 1, the correlation calculation for each variable is carried out on the total population (X1) and the number of vaccinations (X2) with confirmed cases (Y), namely:

Correlation of population (X1) with confirmed cases (Y) = 0.67588

Correlation of vaccination (X2) with confirmed cases (Y) = 0.77585

Correlation of population (X1) with vaccination (X2) = 0.91149

Table 1. Covid-19 Confirmed Cases And Level 3 Vaccinations

No	Districts	Population (X1)	Vaccination (X2)	Confirmed Cases (Y)
1	BATU CEPER	92.407	24.626	315
2	BENDA	83.775	18.378	367
3	CIBODAS	148.106	43.909	973
4	CILEDUG	166.308	48.351	801
5	CIPONDOH	252.136	60.116	886
6	JATUWUNG	102.297	22.010	681
7	KARANG TENGAH	118.003	38.939	606
8	KARAWACI	186.149	59.730	1.314
9	LARANGAN	166.161	51.376	957
10	NEGLASARI	117.034	30.574	536
11	PERIUK	142.522	41.525	1.216
12	PINANG	182.622	50.081	1.220
13	TANGERANG	154.394	52.500	904
Level		3		
Start Date		2021-01-01		
End Date		2021-06-30		

It can be seen from the results of correlation calculation based on the correlation index value, it is found that each variable can be seen in a positive relationship, namely between population and confirmed cases, which has a strong relationship because the spread of confirmed cases is influenced by population size. Meanwhile, vaccination with confirmed cases also has a strong relationship, this



is appropriate because vaccination is focused gradually on areas with high confirmed cases. Then, the relationship between population and the number of vaccinations has a very strong relationship because where the largest population gets a large number of vaccinations also because the spread of Covid-19 cases ranges to densely populated populations.

After determining the correlation between each variable in table 1, the next step is to calculate the multiple correlation between the variable population size (X1) and the number of vaccinations (X2) with confirmed cases (Y) as follows:

Multiple Correlation Coefficient = 0.77957

F count (Fh) = 7.74619

Ftable (ts 5%) = 4.1

From the results of the multiple correlation calculation above, it can be concluded that because Ho is rejected if the number of f count > f table, Ha (alternative hypothesis) states that there is a significant relationship between population size (X1) and number of vaccinations (X2) with confirmed cases (Y).

### 3.4.2 Covid-19 Confirmed Cases and Level 4 Vaccinations

Table 2. Covid-19 Confirmed Cases And Level 4 Vaccinations

No	Districts	Population (X1)	Vaccination (X2)	Confirmed Cases (Y)
1	BATU CEPER	92.407	90.064	585
2	BENDA	83.775	75.498	501
3	CIBODAS	148.106	154.322	1.741
4	CILEDUG	166.308	127.322	1.496
5	CIPONDOH	252.136	193.925	2.301
6	JATIUNGUNG	102.297	118.779	586
7	KARANG TENGAH	118.603	97.892	1.456
8	KARAWACI	186.149	183.477	1.875
9	LARANGAN	166.161	140.259	1.641
10	NEGLASARI	117.034	109.672	755
11	PERIUK	142.522	143.666	1.454
12	PINANG	182.622	155.795	1.678
13	TANGERANG	154.394	149.487	1.653
Level	4			
Start Date	2021-07-01			
End Date	2021-12-31			

In table 2 above, the correlation calculation for each variable is carried out on the total population (X1) and the number of vaccinations (X2) with confirmed cases (Y), namely:

Correlation of population (X1) with confirmed cases (Y) = 0.91243

Correlation of vaccination (X2) with confirmed cases (Y) = 0.87394

Correlation of population (X1) with vaccination (X2) = 0.90996

Based on the results of the correlation calculation based on the correlation index value, it is found that each variable can be seen in a positive relationship, namely between population and confirmed cases, which has a very strong relationship because the spread of confirmed cases is influenced by population size. Meanwhile, vaccination with confirmed cases also has a very strong relationship as well, this is appropriate because vaccination has been evenly distributed and focused gradually on areas with the highest confirmed cases. Then, the relationship between population and the number of vaccinations has a very strong relationship because where the largest population gets a large number of vaccinations also because the spread of Covid-19 cases ranges to densely populated populations.

After knowing the correlation between each variable in table 2, the next step is to calculate the multiple correlation between the variables of population size (X1) and number of vaccinations (X2) with confirmed cases (Y) as follows:

Multiple Correlation Coefficient = 0.91848

F count (Fh) = 26.97120

Ftable (ts 5%) = 4.1

In the results of the multiple correlation calculation above, it can be concluded that because Ho is rejected if the number of f count > f table, Ha (alternative hypothesis) states that there is a significant relationship between population size (X1) and number of vaccinations (X2) with confirmed cases (Y).

### 3.4.3 Covid-19 Confirmed Cases and Level 5 Vaccinations

Table 3. Covid-19 Confirmed Cases And Level 5 Vaccinations

No	Districts	Population (X1)	Vaccination (X2)	Confirmed Cases (Y)
1	BATU CEPER	92.868	16.863	1.899
2	BENDA	84.114	10.374	1.348
3	CIBODAS	149.093	24.718	4.369
4	CILEDUG	168.670	20.065	4.115
5	CIPONDOH	256.392	39.371	6.376
6	JATIUNGUNG	102.650	13.146	1.057
7	KARANG TENGAH	118.409	20.184	3.697
8	KARAWACI	188.124	32.656	4.514
9	LARANGAN	166.900	21.926	4.476
10	NEGLASARI	118.692	16.916	1.902
11	PERIUK	144.208	32.064	3.294
12	PINANG	185.342	31.254	4.582
13	TANGERANG	155.094	30.297	4.570
Level	5			
Start Date	2022-01-01			
End Date	2022-06-30			



In table 3 above, the correlation calculation for each variable is carried out on the total population (X1) and the number of vaccinations (X2) with confirmed cases (Y), namely:

Correlation of population (X1) with confirmed cases (Y) = 0.91707

Correlation of vaccination (X2) with confirmed cases (Y) = 0.85641

Correlation of population (X1) with vaccination (X2) = 0.86772

Established on the results of the correlation calculation based on the correlation index value, it is found that each variable can be seen in a positive relationship, namely between population and confirmed cases, which has a very strong relationship because the spread of confirmed cases is influenced by population size. Meanwhile, vaccination with confirmed cases also has a very strong relationship as well, this is appropriate because vaccination has been evenly distributed and focused gradually on areas with the highest confirmed cases. Then, the relationship between population and the number of vaccinations has a very strong relationship because where the largest population gets a large number of vaccinations also because the spread of Covid-19 cases ranges over a dense population.

After knowing the correlation between each variable in table 3, the next step is to calculate the multiple correlation between the variable population size (X1) and the number of vaccinations (X2) with confirmed cases (Y) as follows:

Multiple Correlation Coefficient = 0.92515

F count (Fh) = 29.70023

Ftable (ts 5%) = 4.1

Likewise, from the multiple correlation calculation above, it can be concluded that because Ho is rejected if the number of f count > f table, Ha (alternative hypothesis) states that there is a significant relationship between population size (X1) and number of vaccinations (X2) with confirmed cases (Y).

### 3.4.4 Covid-19 Confirmed Cases and Level 6 Vaccinations

Table 4. Covid-19 Confirmed Cases And Level 6 Vaccinations

No	Districts	Population (X1)	Vaccination (X2)	Confirmed Cases (Y)
1	BATU CEPER	92.868	2.190	711
2	BENDA	84.114	1.77	639
3	CIBODAS	149.093	3.319	1.740
4	CILEDUG	168.670	1.865	1.357
5	CIPONDOH	256.392	3.952	2.796
6	JATIUNG	102.650	2.12	331
7	KARANG TENGAH	118.409	3.133	1.583
8	KARAWACI	188.124	2.469	1.907
9	LARANGAN	166.900	1.455	1.972
10	NEGLASARI	118.692	1.737	622
11	PERIUK	144.208	3.286	1.087
12	PINANG	185.342	1.448	1.809
13	TANGERANG	155.094	2.626	2.263
Level	6			
Start Date	2022-07-01			
End Date	2022-12-31			

In table 4 above, the correlation calculation for each variable is carried out on the total population (X1) and the number of vaccinations (X2) with confirmed cases (Y), namely:

Correlation of population (X1) with confirmed cases (Y) = 0.86145

Correlation of vaccination (X2) with confirmed cases (Y) = 0.63832

Correlation of population (X1) with vaccination (X2) = 0.54839

Formulated on the results of the correlation calculation based on the correlation index value, it is found that each variable can be seen in a positive relationship, namely between population and confirmed cases, which has a very strong relationship because the spread of confirmed cases is influenced by population size. Meanwhile, vaccination with confirmed cases also has a very strong relationship as well, this is appropriate because vaccination has been evenly distributed and focused gradually on areas with the highest confirmed cases. Then, the relationship between population and the number of vaccinations has a strong relationship because where the largest population gets a large number of vaccinations also because the spread of Covid-19 cases ranges over a dense population.

After knowing the correlation between each variable in table 4, the next step is to calculate the multiple correlation between the variable population size (X1) and the number of vaccinations (X2) with confirmed cases (Y) as follows:



Multiple Correlation Coefficient = 0.88400

F count (Fh) = 17.87892

Ftable (ts 5%) = 4.1

From the results of the multiple correlation calculation above, it can be concluded that because Ho is rejected if the number of f count > f table, Ha (alternative hypothesis) states that there is a significant relationship between population size (X1) and number of vaccinations (X2) with confirmed cases (Y).

### 3.4.5 Covid-19 Confirmed Cases and Level 7 Vaccinations

Table 5. Covid-19 Confirmed Cases And Level 7 Vaccinations

No	Districts	Population (X1)	Vaccination (X2)	Confirmed Cases (Y)
1	BATU CIEPIER	90.053	441	77
2	IBENDA	78.294	23	45
3	CIBODAS	149.192	778	170
4	CILEDUG	136.525	575	126
5	CIPONDOH	203.881	1.220	241
6	JATIUNGUNG	104.419	44	16
7	KARANG TENGAH	108.405	732	143
8	KARAWACI	184.216	186	167
9	LARANGAN	143.934	73	186
10	NEGLASARI	115.162	125	56
11	PERIUK	134.741	582	74
12	PINANG	168.477	376	166
13	TANGERANG	153.793	184	207
Level	7			
Start Date	2023-01-01			
End Date	2023-06-30			

In table 5 above, the correlation calculation for each variable is carried out on the total population (X1) and the number of vaccinations (X2) with confirmed cases (Y), namely:

Correlation of population (X1) with confirmed cases (Y) = 0.82087

Correlation of vaccination (X2) with confirmed cases (Y) = 0.49453

Correlation of population (X1) with vaccination (X2) = 0.44836

According to the results of the correlation calculation based on the correlation index value, it is found that each variable can be seen in a positive relationship, namely between population and confirmed cases, which has a very strong relationship because the spread of confirmed cases is influenced by population size. Meanwhile, vaccination with confirmed cases also has a moderate relationship, this is because vaccination has been evenly distributed, focusing gradually on areas with the highest confirmed cases. Then, the relationship between population and the number of vaccinations has a

moderate relationship because where the largest population still gets a large number of vaccinations but has begun to decrease because the spread of Covid-19 cases has decreased.

After knowing the correlation between each variable in table 5, the next step is to calculate the multiple correlation between the variable population size (X1) and the number of vaccinations (X2) with confirmed cases (Y) as follows:

Multiple Correlation Coefficient = 0.83298

F count (Fh) = 11.33203

Ftable (ts 5%) = 4.1

In the results of the multiple correlation calculation above, it can be concluded that because Ho is rejected if the number of f count > f table, Ha (alternative hypothesis) states that there is a significant relationship between population size (X1) and number of vaccinations (X2) with confirmed cases (Y).

## 4. CONCLUSION

The spread of Covid-19 cases in early 2020 in Indonesia and especially in Tangerang City, has resulted in many fatalities [17]. This virus spreads very easily and quickly infects other people, thus the best way to avoid contracting the virus is by wearing a mask, washing hands with soap, avoiding crowds and maintaining distance. Public compliance in implementing health protocols plays a very important role in suppressing the spread of Covid-19 cases. Therefore, it is necessary to conduct comprehensive and in-depth research as a control and control function as preventive measures if similar cases recur.

### 4.1 Clustering Analysis Results

In this study, clustering analysis of confirmed Covid-19 cases per phase resulted in the following Covid-19 distribution pattern:

#### 4.1.1 Per district Clustering Analysis

In the results of clustering analysis on confirmed Covid-19 cases in each district with cases per phase, the spread of confirmed Covid-19 cases spread rapidly in areas that have the densest population. Furthermore, the increase in cases occurs in the areas closest to it.

#### 4.1.2 Clustering Analysis of Cases by Age Category



By conducting clustering analysis based on age categories, it can be seen that the increase in confirmed Covid-19 cases is dominated by people of productive age in the early adult age category between 26 years old to 35 years old and late adults aged 36 to 45 years old. Therefore, the need for the role of the community in maintaining health protocols during the Covid-19 pandemic needs to be improved and education to the community as preventive action steps if a disaster such as the Covid-19 outbreak recurs.

#### 4.2 Results of Vaccination Administration Analysis Per level

During the pandemic, a very worrying situation has occurred in several regions and especially in Tangerang City. By analysing per level, we can see how dangerous the cases due to Covid-19 are at each case level. So that vaccination is very important as the formation of antibodies and suppresses the increase in the number of cases of spread and mortality due to Covid-19.

In table 6 below, it results in a conclusion because  $H_0$  is rejected if the number of  $f$  count  $>$  from  $f$  table, then  $H_a$  (alternative hypothesis) states that there is a significant relationship between population size ( $X_1$ ) and number of vaccinations ( $X_2$ ) with confirmed cases ( $Y$ ) in all data in semester 3 to semester 7. Thus, by conducting multiple correlation analysis, at each level of the case results in a hypothesis at each level, namely a positive and strong relationship in the provision of vaccinations in each region, so that the provision of targeted vaccinations can reduce Covid-19 confirmed cases, especially in the Tangerang City area and switch from pandemic to endemic.

Table 1. Summary of Multiple Correlations of Covid-19 Confirmed Cases

Cases Per Level	Start Date	End Date	$r_{x_1y}$	$r_{x_2y}$	$r_{x_1x_2}$	Multiple Correlation Coefficient	F count (Fh)	Conclusion
Level 3	2021-01-01	2021-06-30	0,67588	0,77585	0,91149	0,77957	7,74619	There is a significant relationship between $X_1$ , $X_2$ and $Y$
Level 4	2021-07-01	2021-12-31	0,91243	0,87394	0,90996	0,91848	26,9712	There is a significant relationship between $X_1$ , $X_2$ and $Y$
Level 5	2022-01-01	2022-06-30	0,91707	0,85641	0,86772	0,92515	29,70023	There is a significant relationship between $X_1$ , $X_2$ and $Y$
Level 6	2022-07-01	2022-12-31	0,86145	0,63832	0,54839	0,884	17,87892	There is a significant relationship between $X_1$ , $X_2$ and $Y$
Level 7	2023-01-01	2023-06-30	0,82087	0,49453	0,44836	0,83298	11,33203	There is a significant relationship between $X_1$ , $X_2$ and $Y$
Population	$X_1$							
Vaccination	$X_2$							
Confirmed Case	$Y$							
Paabel (n 5%)	4,1							

#### References

- [1] M. Zanin and D. Papo, "Assessing functional propagation patterns in COVID-19," *Chaos Solitons Fractals*, vol. 138, Sep. 2020, doi: 10.1016/j.chaos.2020.109993.
- [2] Y. C. Wu, C. S. Chen, and Y. J. Chan, "The outbreak of COVID-19: An overview," 2020. doi: 10.1097/JCMA.0000000000000270.
- [3] Y. Sudiyono, A. Trisetyarso, H. Prabowo, and M. Meyliana, "Linear Regression Model to Predict the Spread of COVID-19 in Tangerang City," in *2021 9th International Conference on Information and Communication Technology (ICoICT)*, 2021, pp. 121–126. doi: 10.1109/ICoICT52021.2021.9527424.
- [4] World Health Organization (WHO), "QA how is COVID-19 transmitted." Accessed: Sep. 13, 2022. [Online]. Available: <https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-how-is-covid-19-transmitted>
- [5] Y. Sudiyono, A. Trisetyarso, M. Meyliana, and H. Prabowo, "HEALTH SOCIAL CREDIT SYSTEM SIMULATION OF COVID-19 PANDEMIC IN TANGERANG CITY DISTRICTS," *J Theor Appl Inf Technol*, vol. 15, no. 5, 2023, [Online]. Available: [www.jatit.org](http://www.jatit.org)
- [6] Iwan Santosa, "Kasus Covid-19 Melonjak di Singapura, Bangkok, dan Hong Kong." Accessed: Jul. 17, 2025. [Online]. Available: <https://www.kompas.id/artikel/kasus-covid-19-melonjak-di-singapura-bangkok-dan-hongkong>
- [7] World Health Organization (WHO), "COVID-19 - Global Situation." Accessed: Jul. 15, 2025. [Online]. Available: <https://www.who.int/emergencies/disease-outbreak-news/item/2025-DON572>
- [8] Will Hillier, "A Step-by-Step Guide to the Data Analysis Process." Accessed: Sep. 10, 2022. [Online]. Available: <https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step/>
- [9] Diskominfo Kota Tangerang, "Diskominfo Kota Tangerang." Accessed: Jul. 17, 2025. [Online].



Available:

<https://diskominfo.tangerangkota.go.id/>

- [10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 2006.
- [11] M. Zhang and A. Parnell, "Review of Clustering Methods for Functional Data," 2023. doi: 10.1145/3581789.
- [12] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 1, pp. 186–195, Nov. 2019, doi: 10.30812/matrik.v19i1.529.
- [13] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," 2018. [Online]. Available: <https://www.depkes.go.id/>.
- [14] Sugiyono, *Statistika Untuk Penelitian*. Bandung: CV. Alfabeta, 2013.
- [15] kemenkes.go.id, "Klasifikasi Usia Menurut Kemenkes." Accessed: Oct. 04, 2023. [Online]. Available: <https://www.kemkes.go.id>
- [16] Nul Hakim Lukman, "The Urgency of The Elderly Welfare Law Revision," *Jurnal Masalah-Masalah Sosial*, vol. 11, Jun. 2020, Accessed: Sep. 25, 2023. [Online]. Available: <http://jurnal.dpr.go.id/index.php/aspirasi/index>
- [17] S. Setiati and M. K. Azwar, "COVID-19 and Indonesia," *Acta Med Indones-Indones J Intern Med* •, vol. 52, 2020, Accessed: Feb. 01, 2024. [Online]. Available: <https://www.researchgate.net/publication/340645813>