Indonesian Groceries Price Analysis During COVID-19 Pandemic

Andhika Rahadian
School of Electrical Engineering and Informatics
Bandung Insitute of Technology
Bandung, Indonesia
andhikar@students.itb.ac.id

Rahadian Yusuf
School of Electrical Engineering and Informatics
Bandung Insitute of Technology
Bandung, Indonesia
yrahadian@office.itb.ac.id

Abstract—The COVID=19 pandemic has made the several impacts on human life aspects, such as health aspects, social aspects, economic aspects, etc. The groceries price is one of the economic aspects as it is correlated to the buying power of citizens in Indonesia. However, this hypothesis should be verified by performing data analysis. In this paper, the dataset object for analysis is the daily groceries price in Indonesia and daily COVID-19 number of active cases in Indonesia. There are some methods which helps the data analysis from dataset preprocessing, using Pearson Correlation, scatter plot, polynomial regression and mean square error. The result of the data analysis with the certain method, brings the conclusion that groceries price is not really correlated with COVID-19 pandemic. Although in the certain period (June – September 2021), there is a little sign of correlation between them, but it is not significant enough to justify the hypothesis that the groceries price is correlated with COVID-19 pandemic.

Keywords—Groceries, Active Cases, COVID-19, Price, Correlation

I. INTRODUCTION

COVID-19 Pandemic gives big impact on economical aspect in the world, and Indonesia is one of the countries that faces this economic impact [1]. As we can see, there are sets of regulation in Indonesia for suppressing the transmission rate of COVID-19. One of the implementations of the regulations is decreasing of direct human interaction in some places. Places like restaurants, mall, office, schools are getting less activities and interactions by the rules and regulations of PSBB (*Pembatasan Sosial Berskala Besar*) in Indonesia [2][3].

As the nature concepts of all humans, that we are all obligatory to eat and drink as the primary needs. To get the access of resources for groceries, there needs some trading process (between countries) and because of the sets of regulation that some of country had to execute like lockdown policy, this groceries trading process would be interfered [4]. When the trading process in Indonesia is not ideal as usual, there will be a fluctuation of groceries supply and demand for Indonesia.

Besides that, in a narrower view of trading process, if someone is infected by COVID19, there should be an allocated time for self-quarantine. During this time, the interaction between groceries seller with buyer (people) is being interrupted and trading process is also interfered. This circumstances potentially degrade Indonesian people's buying power, because the demand of the groceries would be high as the supply is remain the same (or even lower).

The scope of the supply problem is about the access for the people to obtain the groceries. If the nearest market could not provide the groceries for the people, there should be another additional price for a longer distribution process that make the price of groceries itself could be higher. Refer to the statement of KPPU (*Komisi Pengawas Persaingan Usaha*), one of the main cause of groceries price increase is the long distribution process [5].

Because of these evidences, there comes a hypothesis that COVID19 active cases could make a higher price of groceries price in Indonesia for a certain time in the year 2020-2021. In this paper, we will perform simple statistical analysis from daily COVID-19 active cases in Indonesia and daily groceries price in Indonesia datasets. There will be some comparative methods that helps to analyze relationship between Indonesian groceries price during COVID-19 pandemic.

II. RELATED WORKS

There is another research about the impacts in several aspects as the COVID-19 pandemic is still happening in the year 2020-2021. Ministry of Finance in Indonesia, with the support from United Nations, build some research about Social and Economic Impacts of COVID-19 on Household in Indonesia [6]. This research has a conclusion that can be classified into several key findings, and one of them is the impact of COVID-19 on household finances has been severe. While the other key findings about the social aspects and strategic policy recommendations for Indonesia.

This research from Ministry of Finance gives an insightful result for COVID-19 impact in economical aspects in Indonesia, specifically in household finances. Besides that, this research prone to discuss the social impacts to analyze policy recommendation rather than economic impacts. In this paper, we also analyze the COVID-19 impact in economical aspects in Indonesia, but in another scope which is in specific case of groceries price trend.

III. METHODOLOGY

A. Dataset Collecting

The first step of this research is finding and collecting the relevance dataset from various references. We found the daily groceries price in Indonesia. The groceries consist of ten types of goods (and many of them has their own variants):

TABLE I. TYPES OF GOODS & EACH VARIANTS

Types of Goods	Variants
Rice	Lower Quality I&II, Medium Quality I&II, High Quality I&II
Chicken Meat	-
Beef	Level 1, Level 2
Chicken Egg	-
Red Onion	-
Garlic	-
Red Chili	Big, Curly
Chili Pepper	Green, Red
Palm Oil	Raw, Level 1, Level 2
Sugar	Premium, Regular

The groceries dataset was found from one of the web and mobile applications provided by Bank Indonesia (Central Bank of Indonesia), named "*Pusat Informasi Harga Pangan Indonesia* "[7]. Each variant has its own prices and it always updated daily. As stated before, the data consists of daily groceries price in Indonesia in general. Besides that, after we checked about the availability of the data, the application provides the dataset only from November 2020 until November 2021 (and still updating). In spite of the limitation, in this paper, we have to define time boundaries in a certain of time in November 2020 – November 2021 which still in the range of pandemic COVID-19 period in Indonesia.

Besides groceries dataset, we also search COVID-19 dataset in Indonesia to be compared each other. We found COVID-19 dataset in Indonesia from Ministry of Research & Technology of Indonesia, specific from kawalcovid19[8][9]. They provide the daily COVID-19 related data in Indonesia, e.g. active cases, total cases, death, death rate, etc. The dataset has a time boundary from the early appearing of COVID-19 in Indonesia (approx. in March 2020) and it is still updating daily (at least until November 2021, when this paper is written).

B. Dataset Cleansing

After collecting dataset from various research, we found that the data still not yet ideal to be analyzed. We perform certain techniques to do some dataset cleansing to overcome the problems of the real-world dataset.

The raw groceries data still consists of missing price value where it comes to holiday/national day and weekends. Besides that, the second problem is each variant of groceries has its own prices and it makes the data is full of information but we don't need it all. Furthermore, they provide the raw data in Microsoft Excel format (.xlsx), and contains freezing cells so it would be hard to read in Jupyter Notebook. Another notes that should be reminded that this dataset is using the Indonesian standard date (DD/MM/YYYY) and Rp is the abbreviation of Indonesian currency, which is in Rupiah.

TABLE II. RAW DATASET OF GROCERIES PRICES IN INDONESIA (CLIPPED)

No.	Komoditas(Rp)	06/11/2020		09/11/2020	
I	**Beras**	Rp	11,800	Rp	11,800
1	Beras Kualitas Bawah I (kg)	Rp	10,800	Rp	10,800
2	Beras Kualitas Bawah II (kg)	Rp	10,400	Rp	10,400
3	Beras Kualitas Medium I (kg)	Rp	11,850	Rp	11,850
4	Beras Kualitas Medium II (kg)	Rp	11,650	Rp	11,650
5	Beras Kualitas Super I (kg)	Rp	13,100	Rp	13,100
6	Beras Kualitas Super II (kg)	Rp	12,650	Rp	12,650
II	**Daging Ayam**	Rp	34,650	Rp	34,900
1	Daging Ayam Ras Segar (kg)	Rp	34,650	Rp	34,900
III	**Daging Sapi**	Rp	118,250	Rp	118,250
1	Daging Sapi Kualitas 1 (kg)	Rp	121,850	Rp	121,850
2	Daging Sapi Kualitas 2 (kg)	Rp	113,000	Rp	113,000
IV	**Telur Ayam**	Rp	25,050	Rp	25,150
1	Telur Ayam Ras Segar (kg)	Rp	25,050	Rp	25,150
V	**Bawang Merah**	Rp	36,700	Rp	37,400

When preprocessing groceries data, we perform the data structure transformation, we made another xlsx book and make a new data with the same contents but in another structure so it can be easily read by Jupyter Notebook and convert it into data frame using pandas library. Certain techniques like transposing the data value and price value are done. We also generalize the groceries price in Indonesia by determining the mean price of all types of goods groceries.

ID	Date	Rice	Chicken Meat	Beef	Chicken Eggs	Red Onion	Garlic	Red Chili	Chilli Pepper	Palm Oil	Sugar	Groceries Mean Price (Rp)
1	02/11/2020	11800	34100	117850	24850	35450	27750	43700	35550	14250	14350	35965
2	03/11/2020	11800	34200	118150	24900	35850	27900	43850	35200	14250	14400	36050
3	04/11/2020	11800	34350	118300	24900	36050	27900	43750	35300	14250	14400	36100
4	05/11/2020	11800	34500	118250	24950	36300	27950	43600	35500	14250	14400	36150
5	06/11/2020	11800	34650	118250	25050	36700	28100	43350	35700	14250	14400	36225
	07/11/2020											
	08/11/2020											
6	09/11/2020	11800	34900	118250	25150	37400	28200	42950	36400	14300	14400	36375
7	10/11/2020	11800	34900	118250	25200	37750	28250	42900	36650	14300	14400	36440
8	11/11/2020	11800	35000	118250	25300	38000	28300	42700	36650	14300	14400	36470
9	12/11/2020	11800	35100	118350	25400	38200	28400	42300	36600	14300	14400	36485
10	13/11/2020	11800	35100	118350	25550	38350	28500	42100	36350	14300	14350	36475
	14/11/2020											
	15/11/2020											

To ensure the dataset has all values everyday from November 2020 until November 2021, we insert new date row which is the subsets of holiday and weekends in Indonesia (can be seen on Table III). This insertion causing the missing value of the prices. We use the mean value of previous and next groceries price, based on the number of consecutive holidays. For example, in Table III we can see that in 07/11/2020 and 08/11/2020 has two consecutive holidays (weekends). In this case, we calculate the mean groceries price of two previous days and two next days. The mean value is used to fill the hole for groceries in holidays. We choose this method with an assumption that the groceries price within the holiday and the regular day is linear. The clipped dataset result after filling the missing value is shown by Table IV.

TABLE IV. CLEANSED DATASET OF GROCERIES PRICE IN INDONESIA
(CLIPPED)

New ID	ID	Tanggal	Harga Beras	Harga Daging Ayam	Harga Daging Sapi	Harga Telor Ayam	Harga Bawang merah	Harga Bawang Putih	Harga Cabai Merah	Harga Cabai Rawit	Harga Minyak Goreng	Harga Gula Pasir	Harga Rata Rata Sembako
1	1	02/11/2020	11800	34100	117850	24850	35450	27750	43700	35550	14250	14350	35965
2	2	03/11/2020	11800	34200	118150	24900	35850	27900	43850	35200	14250	14400	36050
3	3	04/11/2020	11800	34350	118300	24900	36050	27900	43750	35300	14250	14400	36100
4	4	05/11/2020	11800	34500	118250	24950	36300	27950	43600	35500	14250	14400	36150
5	5	06/11/2020	11800	34650	118250	25050	36700	28100	43350	35700	14250	14400	36225
6		07/11/2020	11800	34740	118250	25090	37040	28130	43200	36060	14280	14400	36299
7		08/11/2020	11800	34740	118250	25090	37040	28130	43200	36060	14280	14400	36299
8	6	09/11/2020	11800	34900	118250	25150	37400	28200	42950	36400	14300	14400	36375
9	7	10/11/2020	11800	34900	118250	25200	37750	28250	42900	36650	14300	14400	36440
10	8	11/11/2020	11800	35000	118250	25300	38000	28300	42700	36650	14300	14400	36470
11	9	12/11/2020	11800	35100	118350	25400	38200	28400	42300	36600	14300	14400	36485
12	10	13/11/2020	11800	35100	118350	25550	38350	28500	42100	36350	14300	14350	36475
13		14/11/2020	11800	35210	118340	25710	38360	28540	41880	36590	14300	14390	36512
14		15/11/2020	11800	35210	118340	25710	38360	28540	41880	36590	14300	14390	36512

For COVID19 Indonesia dataset, luckily the raw dataset provides no missing value from the time range in November 2020 until November 2021. But they provide so many information which we don't use it entirely. Transformation of dataset structure would be necessary, to separate and joint the only useful and related COVID19 dataset for our analysis (e.g. Active cases, change of active cases). But there is no need to implement fill the missing value method.

C. Visualizing & Analyzing Dataset Trends

After data cleansing process, we perform dataset visualization to see the trends of COVID-19 Active Cases and Groceries price trends in Indonesia. We use matplotlib library written in Python to perform graphs for visualizing data. The next step, we analyze the sub-period COVID-19 still exist in Indonesia (the period between November 2020 – November 2021) of both dataset trends that could be possibly most correlated each other. Price dataset analysis could be narrowed by cutting dataset and adapted with a certain period that is possibly most correlated each other.

D. Quantifying Dataset Comparison Using Pearson Correlation

To quantify the level of correlation between two datasets (groceries price and COVID19 active cases in Indonesia), we use Pearson Correlation method to determine correlation coefficient. Where correlation coefficient (symbolized by r) shows that if r close to 1 or -1, then the two dataset has a strong association/correlation. Otherwise, if r close to 0, then the two dataset has weak correlation [10].

Suppose two datasets x and y with the elements of x_i and y_i with number of elements n, the Pearson Correlation is determined by formula:

$$r = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=0}^{n}(x_i - \bar{x})^2 \sum_{i=0}^{n}(y_i - \bar{y})^2)}}$$

We use scipy.stats library written in Python language to simplify the calculation for the Pearson Correlation of two datasets.

E. Visualizating Dataset Comparison Using Scatter Plot, Polynomial Regression, and Measuring Mean Square Error

To show the deeper analysis of two dataset comparison, we use scatter plot for visualizing the values of two numerical dataset. Scatter plot consists of dots in cartesian plane that represents the value individual data point from two dataset. Using scatter plot, we can easily find the outliers of the dataset by finding the dots that came far from the dominant points in x-y plane. Scatter plot also useful for identifying patterns from two numerical datasets [11].

We also use the polynomial regression method to define the correlation model between two datasets. We use 5th order polynomial regression and the implication is that there at least six coefficients to describe the polynomial for correlation model. The mathematical notation of the correlation model would be like this:

$$y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$$

When y is groceries price, x is COVID-19 active cases in Indonesia, and a,b,c,d,e,f is the coefficients that will determined using NumPy.polyfit library in Python.

Another method that we used to quantify dataset comparison between groceries price and COVID-19 Active cases in Indonesia is by calculating the mean square error between output of polynomial regression (correlation model) and the true dataset.
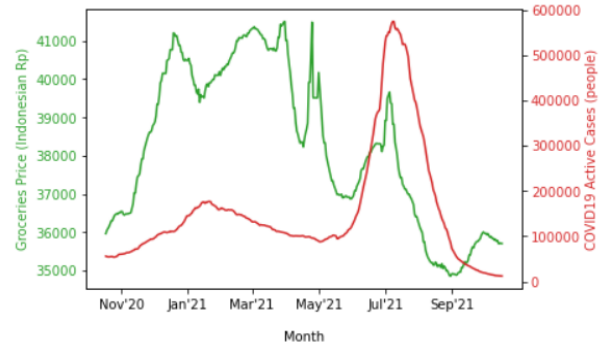
IV. RESULT & ANALYSIS

Fig. 1.   Trend comparison COVID-19 active cases with groceries price in Indonesia
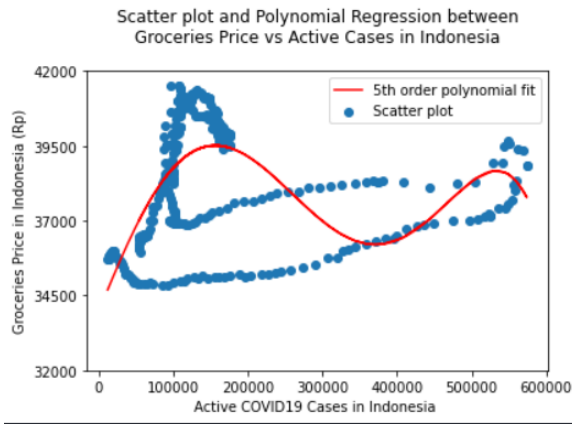
Fig. 2. Scatter plot and 5th order polynomial regression of Groceries Price vs COVID-19 Active Cases in Indonesia November 2020-November 2021

TABLE V. VALUES FOR QUANTIFYING DATASET COMPARISON & 5TH ORDER POLYNOMIAL REGRESSION COEFFICIENTS FOR NOVEMBER 2020-NOVEMEBER 2021 DATASET

Parameter	Value
Pearson Correlation	0.0645
Mean Square Error	2511495.1833
Coefficient a	-4.28125552e-24
Coefficient b	5.18898859e-18
Coefficient c	-1.73020229e-12
Coefficient d	-8.12721181e-10
Coefficient e	6.05858440e-02
Coefficient f	3.39730743e+04

From trend graph as shown by Fig.1., COVID-19 active cases dataset is not significantly correlated with groceries price, when matching the lines and curves from both graphs. To identify deeper about the correlation of both dataset, we can see the scatter plot and its 5th order polynomial regression shown in Fig.2. From scatter plot, we also don't see the patterns (ex: linearity, strong association, etc.) to show that the both dataset is correlated. When we calculate the pearson correlation of this two datasets to quantify the comparison, the result is 0.0645 (as shown on Table V) and it is close to zero which means very weak correlation between them.

The result of polynomial regression (correlation model) is shown by the red graph in Fig.2. This polynomial (with coefficient a-f) is the model that gives the prediction output. From the prediction output and true value from dataset, we get the mean square error of 2511495.1833, which is the massive value and shows that the correlation model can't represent the true value from dataset (harder to linearize the dataset).
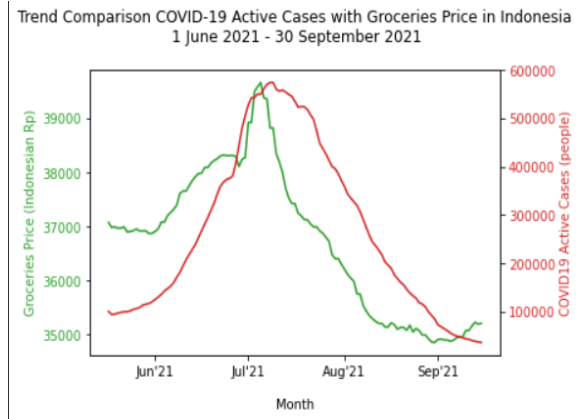

Fig. 3. Trend comparison COVID-19 active cases with groceries price in Indonesia (June 2021-September 2021)
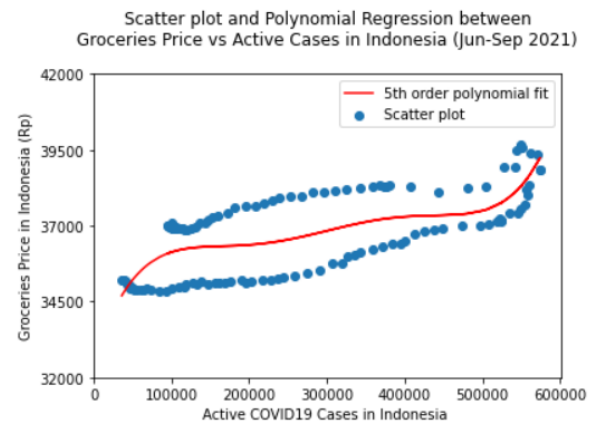

Fig. 4. Scatter plot and 5th order polynomial regression of Groceries Price vs COVID-19 Active Cases in Indonesia during June – September 2021 period

TABLE VI. VALUES FOR QUANTIFYING DATASET COMPARISON & 5TH ORDER POLYNOMIAL REGRESSION COEFFICIENTS FOR JUNE-SEPTEMBER 2021 DATASET

Parameter	Value
Pearson Correlation	0.6771
Mean Square Error	894379.9308
Coefficient a	3.15838534e-24
Coefficient b	-4.76191659e-18
Coefficient c	2.67229703e-12
Coefficient d	-6.85960596e-07
Coefficient e	8.27519627e-02
Coefficient f	3.24868454e+04

From trend graph as shown by Fig.1., we can see that in a certain period (June 2021-September 2021) both groceries price and COVID-19 active cases are the most likely to have a correlation. This is because the shape of the graph that shows linearity of both dataset (if the COVID-19 active cases also increase/decrease, then the groceries price increases/decreases).

After we perform the dataset cutting, to make an analysis in a certain period (June 2021 – September 2021), we get a new scatter plot as shown in Fig.4. This scatter plot shows the more "linear" pattern than Fig.2 (but still not significantly

linear). Furthermore, the value of Pearson Correlation of cut dataset is also higher than the value of original dataset (0.0645 to 0.6771). Besides that, the mean squared error of cut dataset is lower than the mean squared error of original dataset with the same polynomial regression order.

V. CONCLUSIONS

In conclusion, from result and analysis above have a meaning that from June 2021 to September 2021 period (which is when the policy of PPKM Darurat in Indonesia is being implemented), the level of correlation between the groceries price and active cases in Indonesia is more than the other period. However, in general (November 2020-November 2021 period) there are still not enough significant evidence for proofing that the groceries price is correlated with COVID 19 active cases in Indonesia.

REFERENCES

[1] *"Tiga Dampak Pandemi Covid-19 Bagi Perekonomian Nasional"* Accessed on Nov. 16, 2021. [Online]. Available : https://www.beritasatu.com/ekonomi/728997/tiga-dampak-pandemi-covid19-bagi-perekonomian-nasional

[2] *"PERATURAN PEMERINTAH REPUBLIK INDONESIA NOMOR 21 TAHUN 2020 - PEMBATASAN SOSIAL BERSKALA BESAR DALAM RANGKA PERCEPATAN PENANGANAN CORONA VIRUS DISEASE 2019 (COVID-19)".* Accessed on Nov. 16, 2021 [Online]. Available:https://covid19.hukumonline.com/wp-content/uploads/2020/04/peraturan_pemerintah_nomor_21_tahun_2020-2.pdf

[3] *"PSBB diperpanjang, ada dua aturan yang dilonggarkan, apa saja?"* Accessed on Nov. 16,2021 [Online]. Available : https://regional.kontan.co.id/news/psbb-diperpanjang-ada-dua-aturan-yang-dilonggarkan-apa-saja?page=all

[4] *"Dirut Bulog Akui Harga Sembako Mulai Mahal Akibat Wabah Covid-19"* Accessed on Dec. 10, 2021 [Online]. Available: https://money.kompas.com/read/2020/04/20/143037126/dirut-bulog-akui-harga-sembako-mulai-mahal-akibat-wabah-covid-19

[5] *"Ini Penyebab Naiknya Harga Bahan Pokok Menurut KPPU"* Accessed on Dec. 10, 2021 [Online]. Available: https://money.kompas.com/read/2021/04/17/040300426/ini-penyebab-naiknya-harga-bahan-pokok-menurut-kppu

[6] *"Analysis of the Social and Economic Impacts of COVID-19 on Households and Strategic Policy Recommendations for Indonesia"* Accessed on Dec. 10, 2021 [Online]. Available: https://smeru.or.id/sites/default/files/publication/covid-households_rr_en.pdf

[7] *"Data Perkembangan Harga Pangan di Indonesia"* Accessed on Nov. 2, 2021 [Online]. Available:https://hargapangan.id/tabel-harga/pasar-tradisional

[8] *"COVID-19 Indonesia Dataset"* Accessed on Nov. 2, 2021 [Online]. Available : https://sinta.ristekbrin.go.id/covid/datasets

[9] *"COVID-19 Indonesia Dataset"* Accessed on Nov. 2, 2021 [Online]. Available : https://kawalcovid19.id/

[10] *"The Correlation Coefficient"* Accessed on Dec. 16, 2021 [Online]. Available: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html

[11] *"A Complete Guide to Scatter Plots"* Accessed on Dec.16, 2021 [Online]. Available: https://chartio.com/learn/charts/what-is-a-scatter-plot/