What factors influences the doubling rates of COVID-19 cases in Metro Manila?

Michelle Pernia

5.12.20

1. INTRODUCTION

1.1. Background

The National Capital Region, also known as Metro Manila, is the hotspot of COVID-19 cases in the Philippines. On the 16th of March 2020, the whole Island of Luzon, has been placed under Enhanced Quarantine Measure (EQM) in attempt to control the spread of the virus. EQM is the equivalent of a lock down which is characterized by suspended public transportation, closure of schools, limitation of establishments to essential services like health facilities, and other measures ensuring physical distancing. Local authorities were given the function to implement the EQM. Metro Manila is composed of 16 cities and 1 municipality, and these local government units (LGUs) have varying means and interpretation of the operationalization of the EQM. Like in many parts of the world, there were many roadblocks in the smooth implementation of the EQM and most of these are related to the economic impacts of quarantine on people.

When the EQM was declared, the total number of cases in Metro Manila was only 142. It has now grown up to 10,343 as of May 7 the last day of database used in the analysis here.. The LGUs followed slightly different trajectories in the growth of confirmed cases which maybe a result of the nuances in approaches and the physical and socioeconomic characteristics of the local units. It very important to know the possible factors that may affect this growth to be able to formulate effective measures for flattening or managing the curve in the future.

1.2. Framework

In this project, I want to analyze doubling rate of COVID-19 cases in the different local government units in Metro Manila. The five main aspects potentially affecting the growth indicator I want to investigate are:

1. Population density. The population density of a city is a good indicator of the challenges of doing physical distancing. A densely populated community, for instance, may mean difficulty in home quarantining or more possibility for a close contact with a case.

2. Measures of poverty. Lack of means to shop for essential needs is one of the reasons why people break the quarantine protocols. I am interested to know if scale of poverty in a locality is a factor in the spread of the virus.
3. Financial capacity of the local government. Different local authorities have different financial capacity in managing the spread in the community, and providing aids or subsidies to their residents. I am interested to know if financial capacity has any effect in controlling growth the growth.
4. Number of asymptomatic cases . The number of confirmed asymptomatic cases are widely acknowledged as possible indication why the virus is spreading undetected.
5. Essential shopping venues . Shopping areas are possible transmission points as people still converge in volume these areas during the ECQ.

2. DATA SOURCES AND PROCESSING

2.1. Data sources

The LGU level data that will be used in the analysis in this project come from multiple sources. The main source of data is from the Department of Health (DOH) of the Philippines that can be found here. Other data are scraped from Wikipedia page of Metro Manila here and government official data provided by the Philippine Statistical Authority (PSA) 2018 Poverty Statistics here, and Commission on Audit (COA) 2018 Annual Financial Report of the local government here. The details of the data are listed in the following table:

Table 1 Data description and sources

Source of the data	Form of data	Data extracted
DOH data drop 2020	google drive csv download	COVID19 cases Percent of asymptomatic patients per city
Wikipedia page	html table	2015 Population, area, and population density for latest census year
Philippine Statistical Authority - 2018 Poverty Statistics	Excel table	Poverty incidence 2018 Number of poor families 2018
Comission on Audit -2018 Annual Financial Report for Local Government	pdf report	Net income of local government, 2018
Foursquare API calls	data frame convertible json	Main shopping venues

2.2. Data processing of data from DOH

The data from DOH are structured by cases. Each row contains a confirmed COVID -19 patient, his/her date of test confirmation, including the possible date of recovery or death. It also contains the place of residence, up to three levels of aggregation, the smallest unit of which is the LGU level which will be used in this report. It also has information on the health status of the patients. The minor cleaning that were performed in the data are wrongly encoded place of residences and formatting of the name of the LGUs. A lot of the cases are also undergoing validation, so their places of residence were not yet indicated in the database.

Transformation of the data were mainly done using the date on the horizontal axis, and cases classified according to types or level of government units on the vertical axis. Using the DOH data, the dependent variable doubling time since the first confirmed case can be computed.

2.3. Other LGU level data

Other city-level data were scraped from Wikipedia html site and downloaded from official government sites in the form of excel files and pdf. The main issue in putting the data together is the difference in the format or encoding of the name of the city. For example, some datasets will encode 'Taguig City', some will call it the 'City of Taguig'. The data that were scraped from various sources include latitude and longitude, population, area, population density, poverty incidence, and number of poor families.

2.4. Foursquare API data

Foursquare API data were used to determine the number of essentials facility per city. Although the data downloaded seems to not reflect the actual number of essential venues within the city, it can provide some indication of the number of the real number. Only essential facilities, namely, supermarkets, market, farmers market, and grocery store will be considered as essential facilities in this project as these are the ones that can generate significant foot traffic and interactions. Other venue categories such as pharmacy or convenience store will not be considered in the data as they only provide some measures and may cause unnecessary weight as they proliferate in some cities.

3. DATA EXPLORATION AND ANALYSIS

This section will discuss the initial steps I undertook to understand the data, and prepare a database for analysis, and the results of the regression model performed on the data set.

3.1. Visualizing COVID-19 cases trends in the Philippines

To understand the characteristics of the data from the DOH data. I generated some graph showing the trends in COVID-19 cases in the Philippines classified according to regions and cities. I will present some of these charts in the proceeding discussions. Figure 1 below shows the daily trend of COVID-19 confirmed infections, death and recoveries in the Philippines. It can be seen here that from last week of March, the period also when testing capacity of the Philippines significantly increased, the trend of cases has been seemingly dancing with no clear trend.

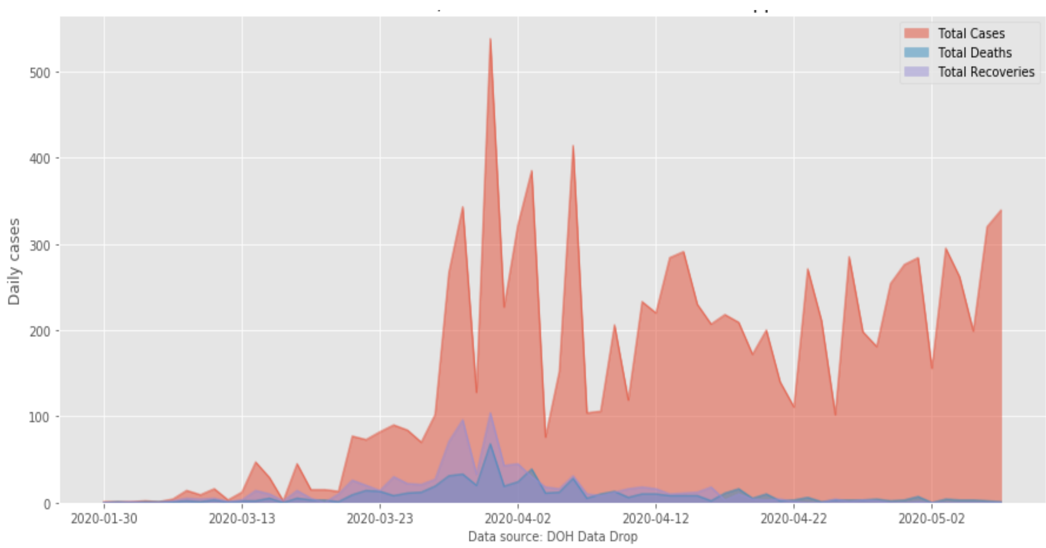

Figure 1. Daily confirmed COVID-19 cases, recoveries and death

Figure 2 shows the accumulative count of COVID-19 confirmed cases, deaths, and recoveries in the Philippines from as early as January 30 to May 7. Here, you would see that the growth in the number of cumulative cases has been continuing without a sign of slowing down. The number of recoveries has surpassed the number of deaths since last week of March.

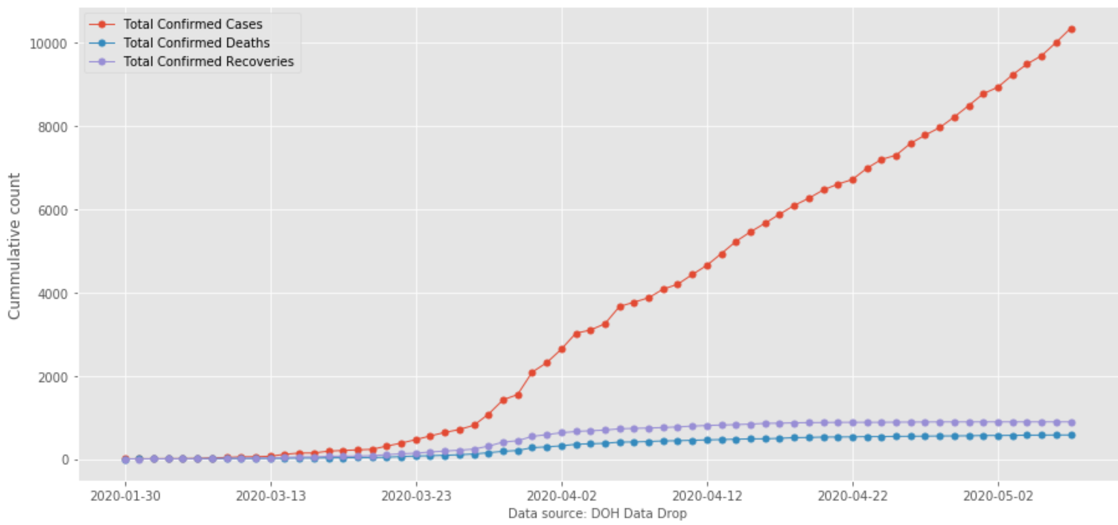

Figure 2. Cumulative COVID-19 confirmed cases, recoveries and death

Classifying the cumulative confirmed cases by region, it can be seen that the concentration of cases can be found in Metro Manila followed by the nearby Region IVA in the south and Region III in north (Figure 3 refers). Cases are also increasing in the Region VII where the second biggest urban center, Cebu, is located.

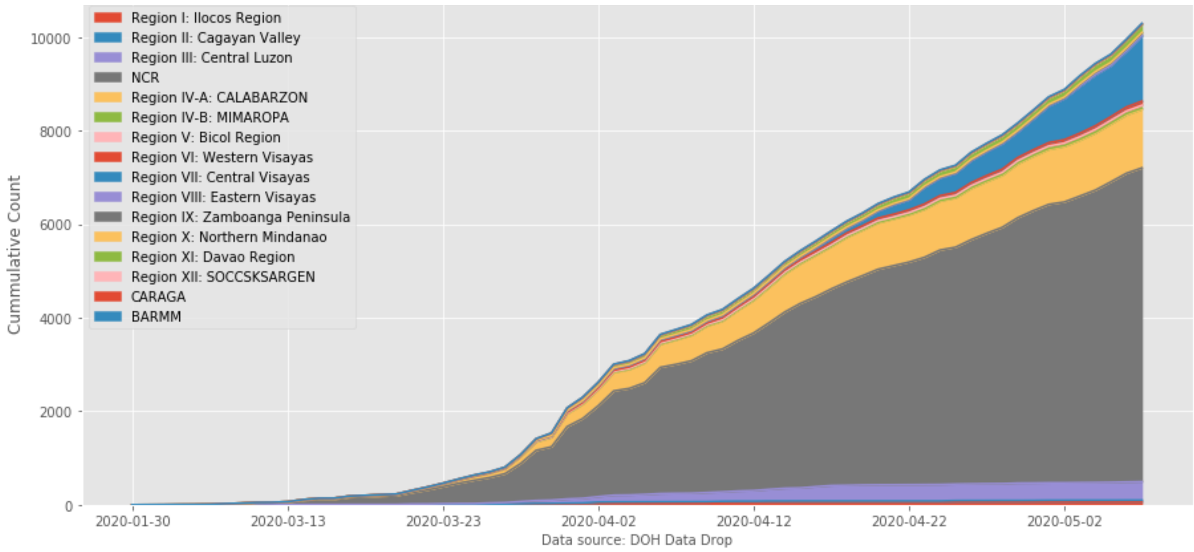

Figure 3. Cumulative COVID-19 confirmed cases by region

Figure 4 shows the cumulative number of confirmed cases per LGU within the Metro Manila. The cases of infection seem to be proportionally increasing according to LGU, but no clear pattern can be identified. The trend of the cases in Metro Manila seems to be flattening on the top.

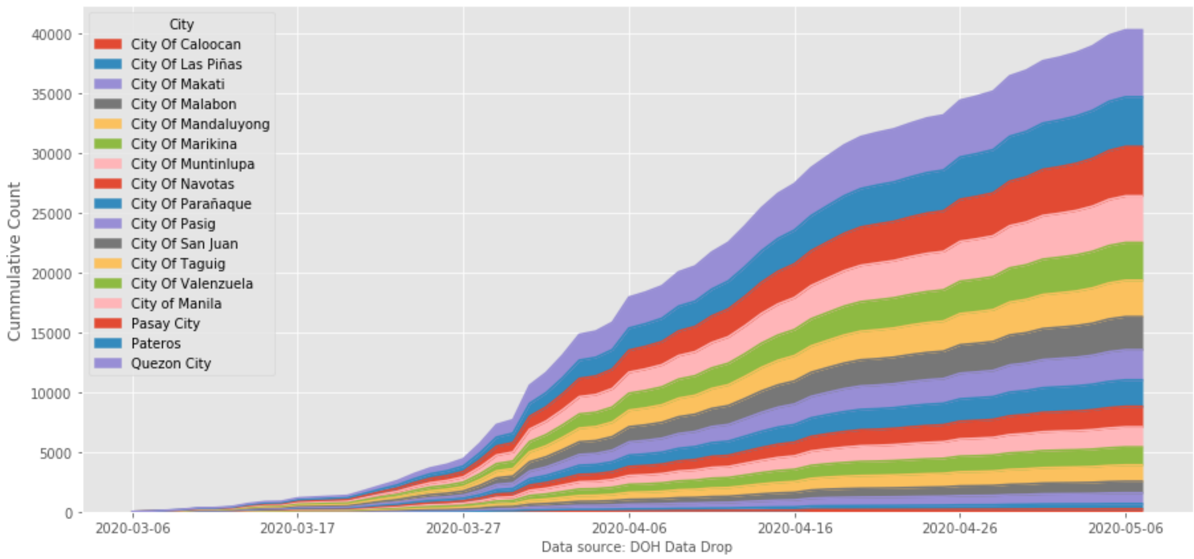

Figure 4. Cumulative COVID-19 confirmed cases by LGU

To visualize the rate of how cases in each LGUs are increasing, the cumulative number of cases is plotted on the logarithmic vertical axis against the number of days after the first infection on the horizontal axis. The code to generate this chart is guided by this [link](). Logarithmic scale means that a unit increase on the vertical axis represent a ten-fold increase. Additionally, the doubling rate at 1 day, 3 days, 5 days and 10 days were plotted on the same graph. Exponential growth is experienced at doubling rate of about 1 to 4 days. The following figure shows that the different LGUs seems to be following the same path on just different scale of cases volume. Most of the cities have now surpassed 5 days doubling time.

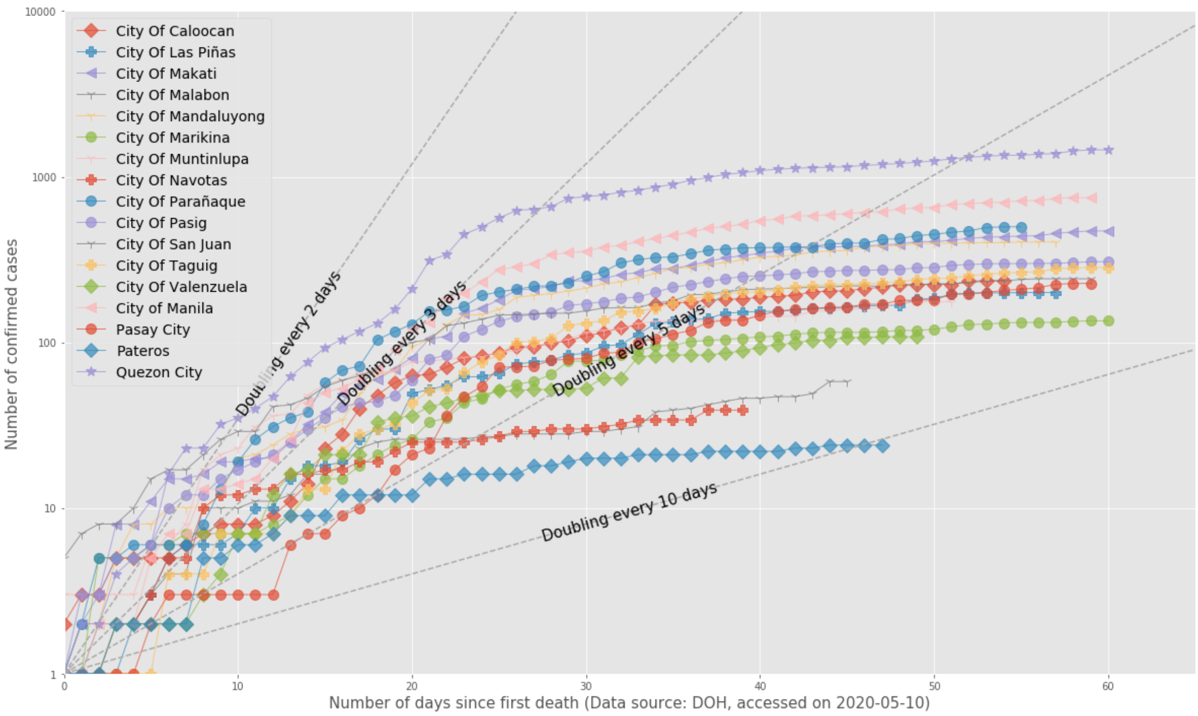

Figure 5. COVID-19 case trajectories by city

3.2. Extracting essential facilities from Foursquare API.

Using Foursquare API, venues inside Metro Manila were downloaded. The resulting number of venues were about 1,647 with a total of _______. The figure below shows the venues extracted from foursquare. Although It seems that the data is limited, the variable is not readily dismissed as it might still offer meaning to the analysis.

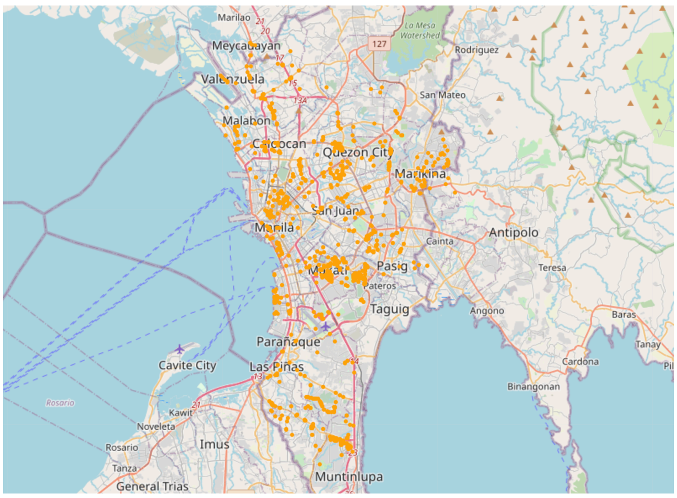

Figure 6. Venues extracted from Foursquare API

The categories that were selected from the Foursquare extract are Farmers Market, Supermarket, Grocery store and Market. These venues were reckoned to be venues that can still attract a large number of people converging even on the lockdown. The table below shows the essential venues extracted per LGU.

Table 1. Number of essential venues extracted from Foursquare API

City	Farmers Market	Supermarket	Grocery Store	Market	Total
City Of Caloocan	0	3	1	0	4
City Of Las Piñas	0	1	0	1	2
City Of Makati	0	2	1	0	3
City Of Malabon	0	2	0	0	2
City Of Mandaluyong	0	3	0	0	3
City Of Marikina	0	0	0	0	0
City Of Muntinlupa	0	2	0	0	2
City Of Navotas	0	0	0	0	0
City Of Parañaque	0	0	0	1	1
City Of Pasig	0	3	1	1	5
City Of San Juan	0	1	0	0	1
City Of Taguig	0	2	1	0	3
City Of Valenzuela	0	0	2	0	2
City of Manila	0	1	1	0	2
Pasay City	0	1	0	0	1
Pateros	0	3	1	0	4
Quezon City	1	2	1	0	4

After integrating the data extracted from different sources, the following Table 2 shows the final database used for the analysis.

Table 2. Dataset for analysis

LGU	DT	popden	pov_incidence	poor_fam	net_income_pop	per_asym	ess_venues
City Of Caloocan	6.977355	28.29	3.2	12	0.16795	0.12288	4
City Of Las Piñas	7.587793	18.4	1	2	1.01244	0.07	2
City Of Makati	6.869692	21.58	0.2	0	8.13122	0.13588	3
City Of Malabon	7.852535	22.85	1.3	1	0.47685	0.06897	2
City Of Mandaluyong	6.698843	35.12	0.9	1	2.31666	0.13614	3
City Of Marikina	8.619696	19.6	1.5	2	0.57384	0.11111	0
City Of Muntinlupa	6.971041	12.01	0.8	1	1.80041	0.06818	2
City Of Navotas	7.568014	20.79	2.4	2	1.00799	0.17949	1
City Of Parañaque	6.247981	14.15	0.6	1	1.16074	0.16834	1
City Of Pasig	7.383115	24.36	1.8	3	4.41604	0.11075	5
City Of San Juan	7.576845	20.36	0.4	0	2.58418	0.08678	1
City Of Taguig	7.612351	17.89	0.5	1	1.80432	0.0636	3
City Of Valenzuela	7.402048	13.49	0.5	1	0.1651	0.25926	3
City of Manila	6.283498	41.4	1.9	9	2.24964	0.14686	2
Pasay City	7.666215	23.14	1.5	2	1.04075	0.08811	1
Pateros	10.469006	31.92	1.8	0	0.42284	0.08333	4
Quezon City	5.806859	17.69	1.5	10	2.22494	0.10324	4

Table 3 shows the variables and their respective descriptions.

Table 3. Description of the variables

Variable	Description
DT	The number of days it takes to double the number of cases
popden	Population density, population (in thousands) per square kilometer
pov_ incidence:	The percentage of families that have income lower than the poverty threshold
poor_fam	The number of poor households in the cities (in thousand)
net_income_pop	The local government net income divided by the population
per_asym	Percent of the cases that are asymptotic
ess_venues	Number essential venues divided by the total area

As the number of LGUs is limited to 17, before a linear regression is performed, it may be necessary to reduce the number of independent variables to avoid overfitting the data. To do this, a factor analysis was done in the six explanatory variables. The variable poverty incidence (pov_incidence) and percent of confirmed asymptotic (Per_asym) has the biggest loadings in separate factors and were used for the reduced model. Table 4 is the results of the regression analysis performed on doubling time and the assumed explanatory variables using the statsmodels.api. Because the full model shows evidence of overfitting due to low significance of variables but high R-squared value, the reduce model is endorsed in this project.

Table 4. Results of Linear Regression

Variables	FULL MODEL		REDUCED MODEL	
	coef	t-stat	coef	t-stat
per_asym	15.484	1.813	34.4363	2.552
pov_incidence	1.6261	1.542	1.9549	3.839
popden	0.1382	1.941		
poor_families	-0.4538	-2.302		
net_income_pop	0.0361	0.111		
ess_venues	0.5007	1.156		
R-squared (uncentered):	0.946		0.867	
Adj. R-squared (uncentered):	0.917		0.849	

Below are the regression plots on two of the variables identified affecting doubling time. The regression plots can be more tighter if more samples are added to the analysis.

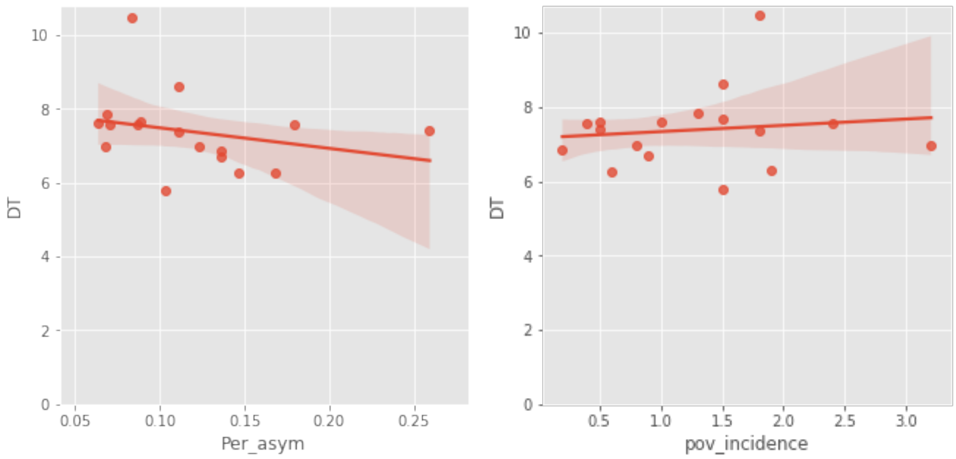

Figure 6. Regression plots on per_asym and pov_incidence vis-à-vis doubling time

4. CONCLUSION AND FINDINGS

This project presented a data integration illustration for the analysis of COVID 19 doubling rate using data extracts from different sources. While the actual data used in the analysis might be limited, the method might be useful for other places who has more extensive data on places.

Based on the analysis in this project, the two factors must be watched out by policy makers in trying to flatten the curve in Metro Manila – (1) the percent of asymptotic cases and (2) poverty incidence. The results reflect what issues are putting in the front presently. The potential key to flattening the curve are mass testing and policies to support the economically impacted households.

Aside from increasing the sample size, the project presented here have many rooms of improvement. The data can be a bit more granulated to achieve a better understanding of the doubling rate. Using this the number of essential venues from API extracts can be more meaningful, if smaller units like barangay level data would be used.