Lima, Perú: What places are more exposed to crime?

David Rodrigo Sánchez Navarro

May 2020

I. INTRODUCTION

Lima is the capital of Perú. This city has a population of around 10 million inhabitants, which represents a third of the population of the entire country. Lima is divided into 2 main regions: Región Lima and Región Callao. In these last two there are 50 main districts that represent about 90% of the population in the capital and they're considered as just one region due to their geographical proximity. When referring to the capital of Lima, we are mainly addressing to those 50 districts which together are known as Lima Metropolitana. As it has a third of the population in the country, it's the city that has more active business. These are of all kinds and are distributed across all districts.

However, Lima has a problem: crime. It is a city with lots of inequality and poverty, there are districts with income levels much higher than others and these brings social problems; one of these being crime. As we do know which districts are more dangerous, through the Foursquare location data we can recognize which venues are more exposed to these crime rates. Results may be useful for people looking forward to open a business. Furthermore, a georeferencing-based analysis becomes a useful tool for policy makers.

II. DATA

We will be using data from the 50 districts of Lima Metropolitana. These data include their geographical coordinates as well as crime rates by district and population. We will be using the Foursquare API as well in order to get information of frequent venues by district. Since we have these two sources, we can match information of Lima's districts with information coming from Foursquare easily. It is important to highlight that all sources of data in this report comes from official sites from Perú. All of them are public and available and because of this, not much data cleaning process was required in order to have a dataset ready to be used.

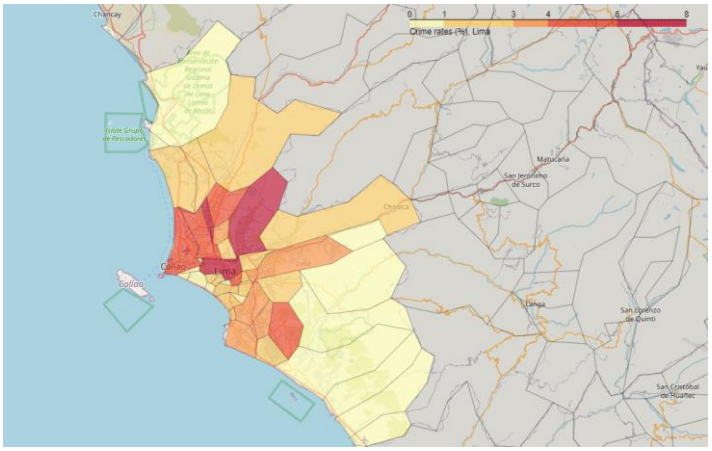
III. METHODOLOGY

To carry out this work, we take into account geo-referenced information from the districts in Metropolitan Lima. We use a source of public information that the national statistical institute of Perú (Instituto Nacional de Estadística, INEI) publishes as part of its common activities. Likewise, this same institution is in charge of carrying out official calculations of crime rates by department, province and district in the interior of the country. This being so, our information sources do not require any database cleaning process, since the data comes from an official and complete source.

As our main objective is to determine which venues or places are the most susceptible to experience a crime, we first need to find a way in which we are able to distinguish between the most frequent establishments by district. In order to accomplish this, we address this issue with the K-means clustering strategy; the main reason being that it will let us understand similarities between districts and help us identify is there are groups or clusters that may have similar crime rates, and similar venues.

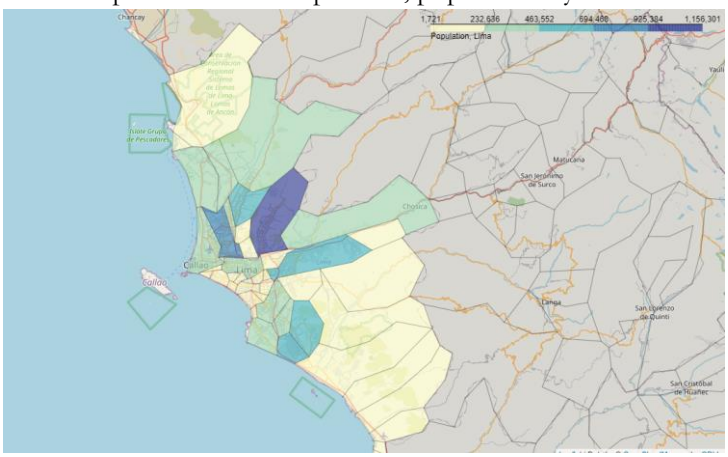
The first graph shows crime rates by districts in Lima Metropolitana. It is important to remark that higher rates are related to districts with less household income and high density in population. Official data cannot be found regarding levels of income by district as INEI databases cannot make inference of variables at that level through its national survey of conditions in households (Encuesta Nacional de Hogares, ENAHO).

Graph 1. Lima Metropolitana, crime rate by district[1]



Source: Own elaboration based on INEI's information on crime rates.

Graph 2. Lima Metropolitana, population by district



Source: Own elaboration based on INEI's information on crime rates.

As we mention previously, intuitively we would argue that high crime rates are related to high density in population. Graph 2 shows this relation, it is easy to recognize that districts with more population are that ones that have more crimes reported. As this situation emerges, it is important to notice that these places also have great number of business that could suffer from delinquency. In the next section we apply our databases from crime rates and information from the Foursquare's API to a K-means model to determine which locations or venues may be more endangered by crime. The idea behind this is that we could group districts taking into account their venues and with the information from crime rates realize which venues are more likely to suffer from crime.

[1] The crime rate by district is defined as the number of crimes reported in a district over the total number of crimes reported in Lima Metropolitana.

IV. RESULTS

Table 1. Venues most exposed to crime by district

District	Crime rate (%)	Population	1st Venue	2nd Venue	3rd Venue	4th Venue	5th Venue	6th Venue
Cluster 1								
Puente Piedra	2.400	378910	Construction & Landscaping	Yoga Studio	Dog Run	Food	Fish & Chips Shop	Fast Food Restaurant
Cluster 2								
Lima	7.802	269858	Restaurant	Art Gallery	Peruvian Restaurant	Sandwich Place	Museum	Seafood Restaurant
Los Olivos	7.322	390742	Park	Soccer Field	Restaurant	Seafood Restaurant	Burger Joint	Market
San Juan de Lurigancho	6.802	1156300	Restaurant	Dessert Shop	BBQ Joint	Farmers Market	Seafood Restaurant	Scenic Lookout
San Martin de Porres	5.087	741417	Dog Run	Shopping Mall	Peruvian Restaurant	Market	Food	Yoga Studio
Comas	5.063	545685	BBQ Joint	Bar	Clothing Store	Pet Store	Fish & Chips Shop	Fast Food Restaurant
Cluster 3								
Carabayllo	2.574	322936	Supermarket	Athletics & Sports	Yoga Studio	Diner	Fish & Chips Shop	Fast Food Restaurant
Villa El Salvador	2.318	489583	Peruvian Restaurant	Fast Food Restaurant	Bar	Yoga Studio	Dog Run	Fish & Chips Shop
Rímac	2.255	165451	Shopping Mall	Peruvian Restaurant	Athletics & Sports	Park	Dog Run	Yoga Studio
Mi Peru	0.276	63542	Shopping Mall	Peruvian Restaurant	Athletics & Sports	Yoga Studio	Dog Run	Fish & Chips Shop
Cluster 4								
Lurigancho	1.306	232902	Playground	Yoga Studio	Diner	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
Cluster 5								
Pucusana	0.109	18284	Beach	Park	Harbor / Marina	Italian Restaurant	Cajun / Creole Restaurant	Dog Run
Santa Maria del Mar	0.001	1721	Harbor / Marina	Beach	Seafood Restaurant	Yoga Studio	Diner	Fast Food Restaurant

V. DISCUSSION

Table 1 shows our results. We form 5 clusters of districts considering similarities with respect to the places around them. We find those venues which are more frequent by each district in each cluster. Cluster 2 becomes the one in which we need to focus as this is the one that concentrates most of the districts, higher crime rates, and most of the population as well. We'll notice that places for districts in cluster number 2 are the ones with a higher chance of being nearby a crime. Precisely, restaurants and markets in these districts are the ones that suffer the most from crime nearby; this is, these businesses get more exposed to crime and could mean possible damages to people running them or their clients. These results may be summarized in the following statement: "If you go to a restaurant in San Juan de Lurigancho, there's a good chance that you get robbed". Additionally, our findings are in accordance with the reality reported by the media and official sources regarding crimes that occur in the capital, which gives us a good signal.

It is important to recognize that these results may be very general. We might be able to make more refinements in the collection of data and get better results. For example, our approach was meant for a district level, but if we use databases from the national police station census of Perú and get even more specific data about location of crimes and collection of venues around those places. By doing this, we can get better results and provide better conclusions.

VI. CONCLUSIONS

In this report we use information from districts in Lima Metropolitana regarding crime rates, population, and we use the Foursquare API to get information from nearby places in each district. With these data, we cluster districts considering similarities in those venues and then compare which places are more exposed to crime by district. The results show that in Lima Metropolitana, in districts with the highest levels of population and crime rates, small business such as restaurants or markets are the ones that become more affected by crime, which impacts negatively in people run those business, people who work in there and clients. We discuss how we could improve our results and get more precise with our conclusions. Finally, results coming from a even more rigorous study, could help policy makers and people who seek to open a business and help them make better decisions.