16 August 2020

Avoid Risk Neighbourhoods: General Insecurity & COVID-19

Introduction/Business Problem

Every year is unique and particular. But, 2020 brought the world the special planetary pandemic challenge of COVID-19. It spread and penetrated rapidly into different parts of the globe. And, the autonomous city of Buenos Aires (CABA: Ciudad Autonoma de Buenos Aires) is not an exception.

In this particular setting, I contemplated for the capstone project a hypothetical corporate client from abroad, an international trading firm (The Client), that is planning to relocate their representative family to the city of Buenos Aires, once the pandemic-related restrictions are lifted. Very concerned with the two risks—the general security risk (crime) and the pandemic risk (COVID-19)— the Client wants to exclude high risk neighbourhoods from the selection of the location of the relocation. At the same time, the Client wants to understand the characteristics of neighbourhoods by popular venue categories such as restaurants, shops, and sports facilities. In this context, the Client hired me as an independent consultant to conduct a preliminary research for the future plan. The Client sets the following three objectives.

1. Identify outlier high risk neighbours (the Outlier Neighbourhood/Cluster) in terms of these two risks—the general security risk (crime) and the pandemic risk (COVID-19).
2. Segment non-outlier neighbours into several clusters (the Non-Outlier Neighbourhoods/Clusters) based on a quantitative risk metrics.
3. Characterize the Non-Outlier Clusters/Neighbourhoods regarding popular venues.

For the first two objective, I would run 'clustering' methods, unsupervised machine learning techniques, since there is no label in the data. In other words, the clusters are not empirically known: for the first objective, we have no empirical idea about the outlier neighbourhoods; for the second, about the cluster structure among the neighbourhoods.

For the third objective, I would exploit Foursquare API to explore popular venue categories for Non-Outlier Neighbourhoods.

Michio Suginoo

16 August 2020

Buenos Aires is a densely populated city: the total population of approximately 3 million in the area of 203 km^2. And each neighbourhood has its own distinct size of area and population. The city is divided into 48 administrative division, aka 'barrios', to which I will refer as 'neighbourhood' in this report.

These two given risks of our interest are likely affected by the population density profiles. Especially, the fact that as 'social distancing' is a key to the prevention of COVID-19 suggests that population density is a significant attribute for the pandemic risk. In other words, the higher the population density, the higher the infection rate. The similar would be the case for the general insecurity. Obviously, this preconception needs to be assessed based on the actual data.

So, in the conduct of this research, I would likely need to incorporate into the analysis the differences in the population density among neighbourhoods.

In addition, since the Client wants to analyse the neighbourhoods from the perspective of two risks together, this poses another challenge. I would like to explore the construction of a compound risk metric of these two risks to meet the demand.

This is a quick introduction of the project.

Michio Suginoo