

APPLIED DATA SCIENCE CAPSTONE FROM IBM BY COURSERA

CAPSTONE PROJECT

Author: Carlos A. Evangelista Busso

1. Business Problem:

The present project is born in the middle of a problem caused by the virus called: Coronavirus or Covid-19, which shakes the entire world in 2020, in most countries there is quarantine given by governments that forced people to be at home to take care of their health. Most of the day or in some cases like mine, which I spent almost the entire period of quarantine inside my home for the type of the company where I work, mining company, all activities were suspended.

The days I stayed at home I had the opportunity to observe with my family the activity carried out by the business located in front of my home, a pharmacy, where merchandise arrived day after day and this was something that we did not notice despite its years of existence. At first sight it is a very profitable business, it was at that moment where together with my family we asked ourselves the following question: What place or areas are profitable to start a pharmacy business?

The city of Metropolitan Lima is the capital of Peru and has around 9.5 million people [1] and in this there are approximately 198 first-level health centers [2]. It is in this city where this project is focused and an analysis will be carried out using geolocation to determine the location where a pharmacy business could be started and justify the reasons for it to be profitable.

2. Data Description:

To solve the problem, an analysis was made of the areas around the main hospitals in Lima. Logically, it is through these places that people who are treated and subsequently prescribed to acquire medicines by doctors pass through.

- In order to obtain the coordinate information of these places, first we must have a list of the names of the hospitals with which we will work. This data collection was performed through a web search [3] and subsequent use of the method: Web Scraping.
- Once the list of all hospital names has been obtained, information is extracted from the coordinates of each one, this is done using the Google Maps API: Places API [4].
- With the coordinates, the Foursquare API was used to obtain the most common places around each hospital within a radius of 500 meters [5].