Capstone Project: Hospitals in Lima districts

Name: Alexis Martínez

1. Introduction:

The coronavirus COVID-19 pandemic is the defining global health crisis of our time and one of the greatest challenges we have ever faced. Since its beginnings in Asia, it has spread to most countries in the world, making the presidents order a long quarantine to keep it at bay. Unfortunately, the economy of every nation has plummeted while the positive cases and deaths keep increasing. There doesn't seem to be a cure in the near future, leaving us with the only option of coexist with the virus.
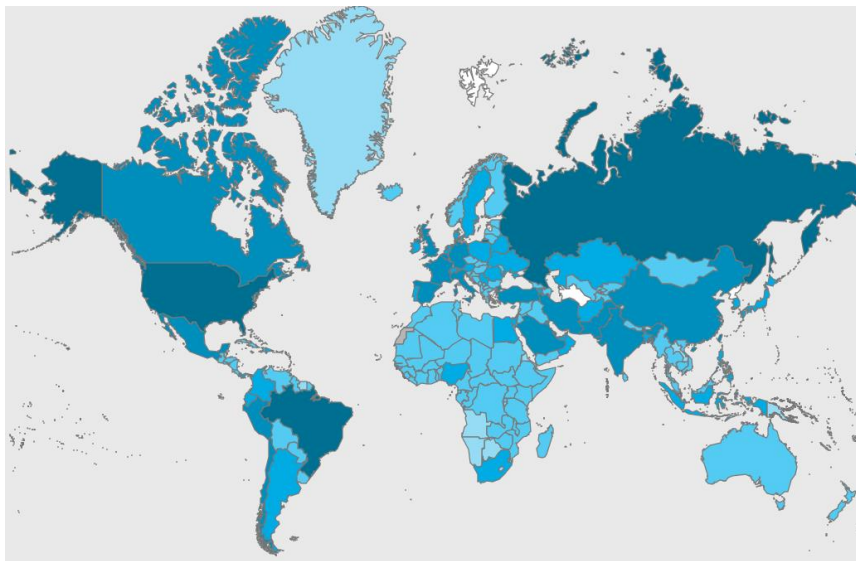

Figure 1: Choropleth map of the world according to Covid-19 cases

Peru is one of the most affected nations by COVID-19, being in the top 10 countries with most cases as of June 2, 2020. The Peruvian president Martín Vizcarra declared a national state of emergency on March 15, when the country had just 71 confirmed cases of COVID-19. The order closed Peru's borders and banned Peruvians from leaving the house except to access essential goods or perform essential work. But public health experts say living and working conditions in the country of 33 million, where a fifth of people live on only around $100 a month, has made it near impossible for many Peruvians to comply with quarantine measures.

1.1. Business Problem:

The government of Peru has organized a group of data scientists to analyze the COVID-19 positive cases in the city of Lima and select a district to open a new hospital. Lima is the capital city of Peru, and it holds 60% of the total positive cases in the country. Analyzing the data, which districts would the team recommend opening a new hospital?

2. Data

To solve this problem, the team has set up to find the following data:

- List of districts in Lima, Peru. This encompass the whole project.
- Latitude and Longitude of the districts in Lima, Peru. We need this information to plot the map and get the hospitals nearby.
- COVID-19 positive cases in Lima per district.
- Hospitals in Lima per district.

2.1. Source of Data

To gather the data needed, we are going to use the following sources:

List of Districts in Lima:

The Wikipedia Page https://en.wikipedia.org/wiki/List_of_districts_of_Lima has the data we need in a table format. The following table is a dictionary of each column:

Column Name	Definition
Districts	Name of the District
UBIGEO	Unique code only used in Peru
Area	Area in Km of the district
Population	Population of the district (2005)
Population Density	Population density of the district (2005)
Created	Date of Creation
Postal Code	Postal Code of the District
Location	Map of Lima showing the District's location

To get the data into Python, we'll use web scraping techniques. The columns we'll need are Districts, UBIGEO and Postal Code.

Latitude and Longitude of the districts in Lima:

This information can be retrieved with Python package Geocoder. This package can be very unreliable, and Google Maps API isn't really an option since 2018. Because of that, we have a CSV file with the correct latitude and longitude of each district, in case the Python package doesn't work.

COVID-19 positive cases in Lima:

The government of Peru has the information open to everyone in their webpage https://www.datosabiertos.gob.pe/dataset/casos-positivos-por-covid-19-ministerio-de-salud-minsa. The dataset they provide is updated every day, and has the positive cases per department, province and district of Peru. The following table is a dictionary of each column in the dataset.

Column Name	Definition
UUID	Unique Identifier
DEPARTAMENTO	Department in Peru
PROVINCIA	Province in Peru
DISTRITO	District in Peru
METODODX	Type of test applied
EDAD	Age
SEXO	Gender
FECHA_RESULTADO	Date of the Positive Case (DD/MM/YYYY)

In this case, we are only going to use the columns 'DISTRITO' and 'METODODX'.

Hospitals in Lima per district

To get this data, we are going to use Foursquare API. Using this API, we can make RESTful API calls to retrieve data about hospitals in different districts of Lima. To use it, we need a developer account active. Take into account that the account has a limited amount of normal and premium calls per day. In the following link, we can find the venue categories they have and the ID to call them https://developer.foursquare.com/docs/build-with-foursquare/categories.