# Introduction

## Background

This report is part of the IBM Data Science Professional Certificate program submitted for the final Capstone project. The programme is a 10-course series created by IBM and provided online on Coursera. The topic of this report is open for the learner. He or she are to select a problem and try to extract insights, and possibly offer recommendations, by leveraging location data, interactive maps, and the Foursquare API – all while adhering to the IBM data science lifecycle and best practices. Additional tools such as web scraping and the *k-means* machine learning algorithm can be used if required.

## Problem Description

Exactly a year ago from the time of writing of this report, a massive blast ripped through the port of Beirut, the capital city of Lebanon, on August 4, 2020. The explosion was apocalyptic in its magnitude, affecting 50 percent of the city. [1] The shock-wave essentially left much of Beirut’s city center in ruins.

These are not good times in Lebanon. In the latest blow, inflation is running at more than 200% for food. [2] These twin tragedies pile yet another layer of hardship on people already reeling from an array of crises. As such, the international community and numerous local non-governmental organizations (NGOs) joined hands to support those hardest hit by the disasters.

The result was a large scale and global response. Organizations providing assistance to residents, foremost food distribution, operated within humanitarian response zones designated by the United Nations. However, due to the accumulation of crises and the scale of the explosion, coordinating such an extensive humanitarian response grant the risk of unexpectedly marginalizing or inadequately supporting particular zones in dire need of greater help.

Therefore, this report maps the socioeconomic status (poverty level) of people residing in these *operational zones* with the corresponding number of organizations operating in the same zone to distribute food parcels. The goal is to explore if any zone or group of people would have required more assistance. Also, the report further investigates two specific neighborhoods in Beirut which are identified throughout our analysis to be of different socioeconomic levels and are receiving uneven assistance level.

## Target Audience

This project intends to serve three groups of audience:

1. **Researchers:** Provide them with findings that bridge the socioeconomic status of zones with the delivered assistance level – an observation not commonly prevalent in literature and publications on the topic.
2. **International and Local Humanitarian Organizations:** Help organizations better identify zones to pinpoint areas of action, and promote data science technologies as a tool to streamline future response and recovery efforts.
3. **Supporters:** Supply unfunded and interested supporters of the Beirut blast recovery efforts useful information, maps and insights which can help them both raise their awareness on the situation and plan their intervention accordingly.

# Methodology

## Data Collection

Three datasets of various formats were utilized. First, the platform Humanitarian Data Exchange (HDX)[[1]](#footnote-1) hosted by the Office for the Coordination of Humanitarian Affairs (OCHA) provided us the option to download two datasets into our local machine:

1. The UN-HABITAT dataset (.CSV format) classifying the operational zones as per the socioeconomic status of the respective citizens.
2. The geospatial data of the UN operational zones around the port of Beirut in the form of a zipped file. Out of the various data formats provided within the folder, we utilized the shape file (.SHP).

Second, the interactive map also provided by OCHA showing the number of organizations operating within particular zones was used as an online web tool. For the scope of this project, the *Food Security* parameter only was considered.

Finally, the Foursquare API provided two JSON scripts, and eventually transcribed into a pandas data frame, of the venues 500m away from two designated neighborhoods in Beirut: Ashrafieh at zone number 62 and Bourj Hammoud at zone number 89.

## Data Description

The main CSV dataset from UN-HABITAT pinpoints the socioeconomic status (poverty levels) for the 188 operational zones according to six aggregated factors: Not residential, none poor, minority poor, half poor/half not poor, majority poor, and all poor. The original dataset before data pre-processing (wrangling) can be shown in figure 1:

Figure : The original CSV dataset imported into a pandas data frame.

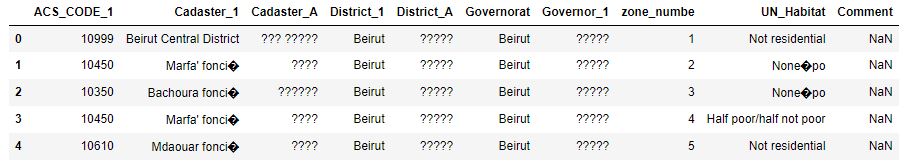
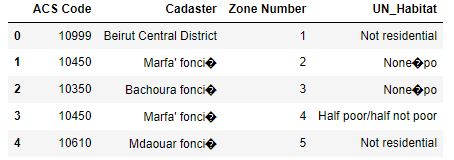


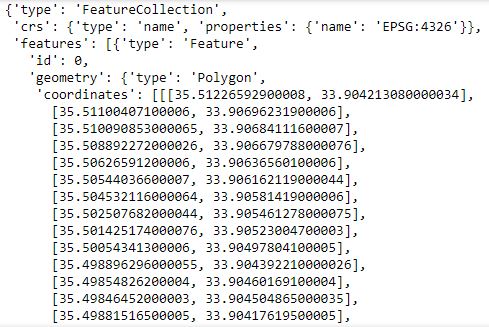
Figure 2 shows the same data frame further processed to drop unneeded columns and errors in data entries. The display of several characters in the “UN\_Habitat” column was addressed to drop miscellaneous icons at a later stage through the analysis after viewing the GeoJSON data.

Figure : The data frame after data pre-processing.



The SHP shape file provided us with the GeoJSON script necessary to generate a choropleth map via the *Folium* library. The fields of interest in our project are those providing information on the zone number, object id, and the corresponding coordinates of each polygon constituting the area of each zone. Figure 3 provides a snapshot of those fields:

Figure : Snapshot of the relevant part of the GeoJSON script.



This script was extracted by an ArcGIS built-in module throughout early analysis and later at an advanced stage of analysis a python-based approach via *Geo Pandas* yielded the same results.

Second, the online web tool is an interactive choropleth map designed by the vendor via Power BI. It displays the number of organizations operating in the region as well as for their names upon the selection of a particular zone.

Finally, the data frames generated by the Foursquare API are discussed in the *Analysis* section of this report.

## Work Process

The work process of this report localizes the IBM data science lifecycle to meet the project scope. Hereafter, the methodology elements within the generic IBM lifecycle will be referred to as *stages* and the specific activities of this projects as *steps*. The steps showcased in figure 4 start after the “Business Understanding” stage:

Figure : The 9-steps work process utilized through the project development.

Probing deeper into these 9 steps, the first three are inline with the *Data requirements* up until *Data preparation* stages. Steps four to seven fall under the *modeling* stage with several reiterations with the previous stages. The remaining steps fall under the *evaluation* stage. However, as we are not preparing a commercial solution, both final stages of *deployment* and *feedback* are not adopted.

## Analysis Tools

Table : Tools utilized through the project development.

|  |  |  |
| --- | --- | --- |
|  | Tool | Description |
| Software | ArcGIS | Extract GeoJSON script from a shape file (.SHP) |
| Jupytr Notebook | Run the python script |
| Anaconda Desktop | Initiaite an instance of local and online Jupytr Notebook. |
| Libraries | Folium | Used to generate map visualization (Choropleth)via geospatial data to illustrate the socioeconomic status of every zone. |
| Pandas | Store and manipulate data. |
| Geo-Pandas | Store and manipulate geospatial data, and generate a GeoJSON script from a SHP file. |
| Matplot | Used to generate plot visualizations. |
| Data Sources | Foursquare API | The *explore* endpoint provided venues around designated latitude and longitude locations. |
| HDX | Provided CSV data on the socioeconomic status of operational zones. |
| Online Power BI | Provided data on the number of organizations operating in specific zones in the format of an interactive choropleth map. |
| Packages | GDAL | Allow the *!pip* command to import the Geo-pandas library within the project’s Jupytr Notebook. |
| Pyproj |
| Fiona |
| Shapely |
| Geopandas |

1. HDX platform and the corresponding data available at: <https://data.humdata.org/dataset/beirut-port-explosion-operational-zones> [↑](#footnote-ref-1)