|  |
| --- |
| IBM Data Science Professional – Capstone Project |
| Food Distribution in the Zones Affected by the Beirut Port Blast |

|  |
| --- |
| Haytham Tibni  2021 |

Table of Contents

[1 Introduction 2](#_Toc80103974)

[1.1 Background 2](#_Toc80103975)

[1.2 Problem Description 2](#_Toc80103976)

[1.3 Target Audience 2](#_Toc80103977)

[2 Methodology 3](#_Toc80103978)

[2.1 Data Collection 3](#_Toc80103979)

[2.2 Data Description 3](#_Toc80103980)

[2.3 Work Process 5](#_Toc80103981)

[2.4 Analysis Tools 5](#_Toc80103982)

[3 Analysis and Results 6](#_Toc80103983)

[3.1 Exploratory Data Analysis 6](#_Toc80103984)

[3.2 Choropleth Map 7](#_Toc80103985)

[3.3 Comparison 8](#_Toc80103986)

[3.4 Foursquare API 8](#_Toc80103987)

[4 Discussion 10](#_Toc80103988)

[5 Conclusion 10](#_Toc80103989)

[References 10](#_Toc80103990)

[Table of Figures 11](#_Toc80103991)

Food Distribution in the Zones Affected by the Beirut Port Blast

# 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.

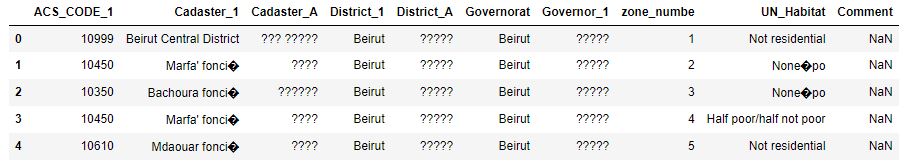
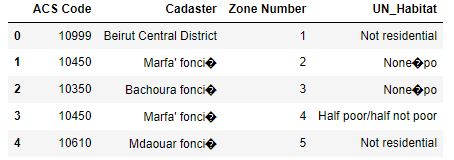


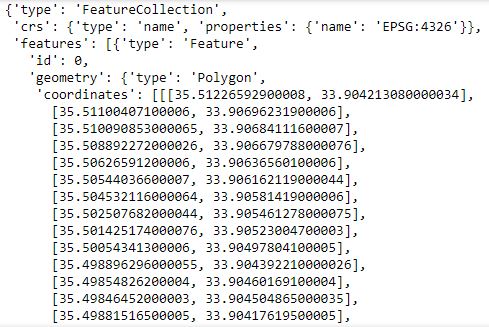
Figure 2 shows the same data frame further processed to drop unneeded columns and errors in data entries. The display of several characters in the “UN\_Habitat” column was addressed to drop miscellaneous icons at a later stage through the analysis after viewing the GeoJSON data.

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 1 below enumerates the tools utilized throughout the abovementioned work process:

Table 1: Tools utilized through the project development.

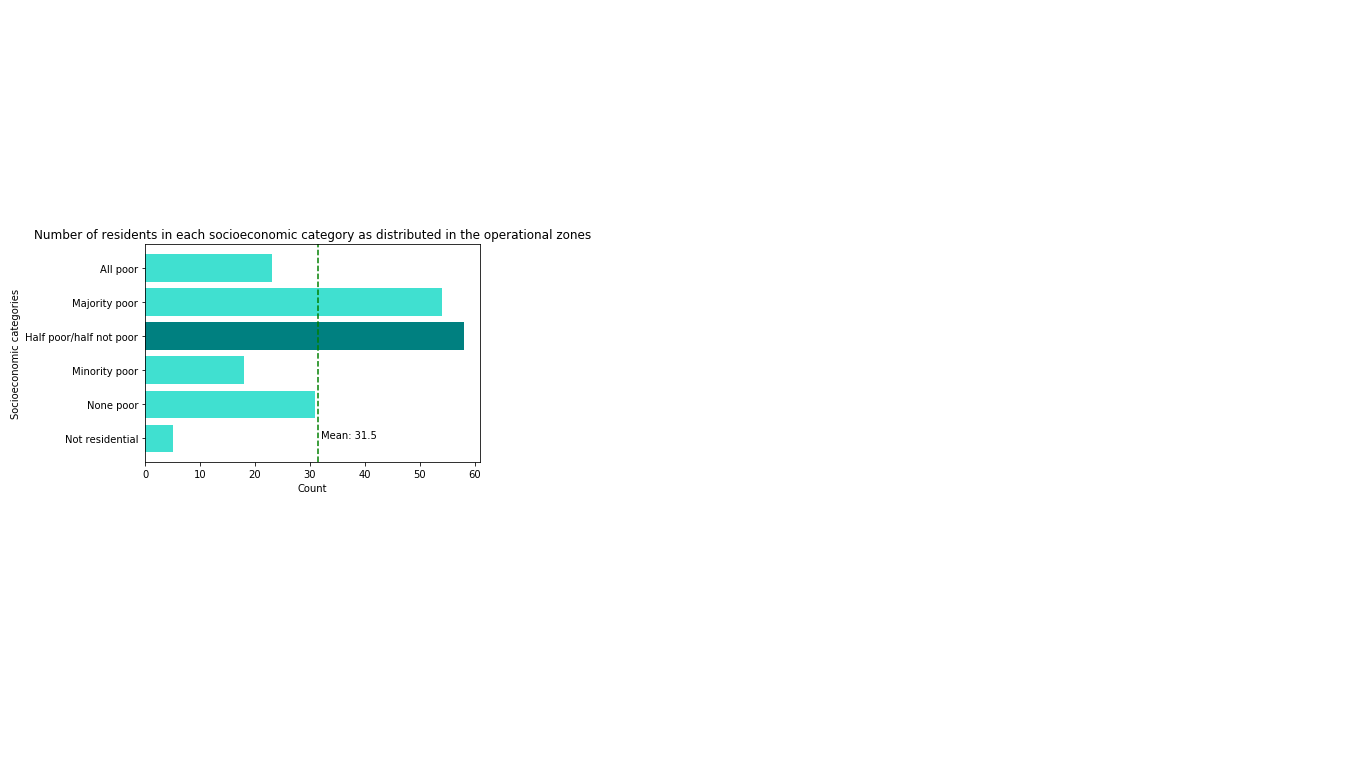
|  |  |  |
| --- | --- | --- |
|  | Tool | Description |
| Software | ArcGIS | Extract GeoJSON script from a shape file (.SHP) |
| Jupytr Notebook | Run the python script |
| Anaconda Desktop | Initiaite an instance of local and online Jupytr Notebook. |
| Libraries | Folium | Used to generate map visualization (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 |

# Analysis and Results

## Exploratory Data Analysis

The CSV file was manipulated as a data frame by the *pandas* library to keep only the required columns (ACS Code, Cadaster, Zone Number, and UN\_Habitat). The count of the six socioeconomic categories of the “UN\_Habitat” column were then visualized as a bar graph:

Figure : Count of the 6 socioeconomic categories.



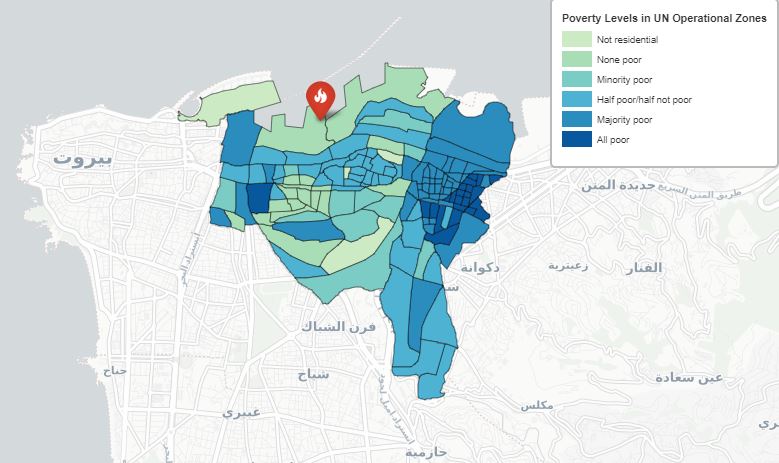
As can be seen, zones having either the majority or half of the residents to be poor are above the mean value. Also, a substantive proportion of the zones are either not poor or have a minor population to be poor, with several zones not residential at all. This suggests that the operational zones are spread between two extremes and that humanitarian agencies should plan their activity accordingly to ensure even and fair distribution of food parcels.

The Shapefile provided the coordinates (latitude and longitude) of each zone forming a polygon. These coordinates will create the boundaries of the zones enabling the display of data on a *folium-leaflet* map. Since it is a GeoJSON script, we converted it into a data frame to view the attributes. This step was necessary to ensure the numbering of the zones in the GeoJSON script are exact to those of the main data frame. Otherwise, we will not be able to project the socioeconomic categories of the zones on the generated zones map. In our work, the needed JSON property (FID) was mutual to the data frame column (Zone Number) and no further data wrangling was required.

## Choropleth Map

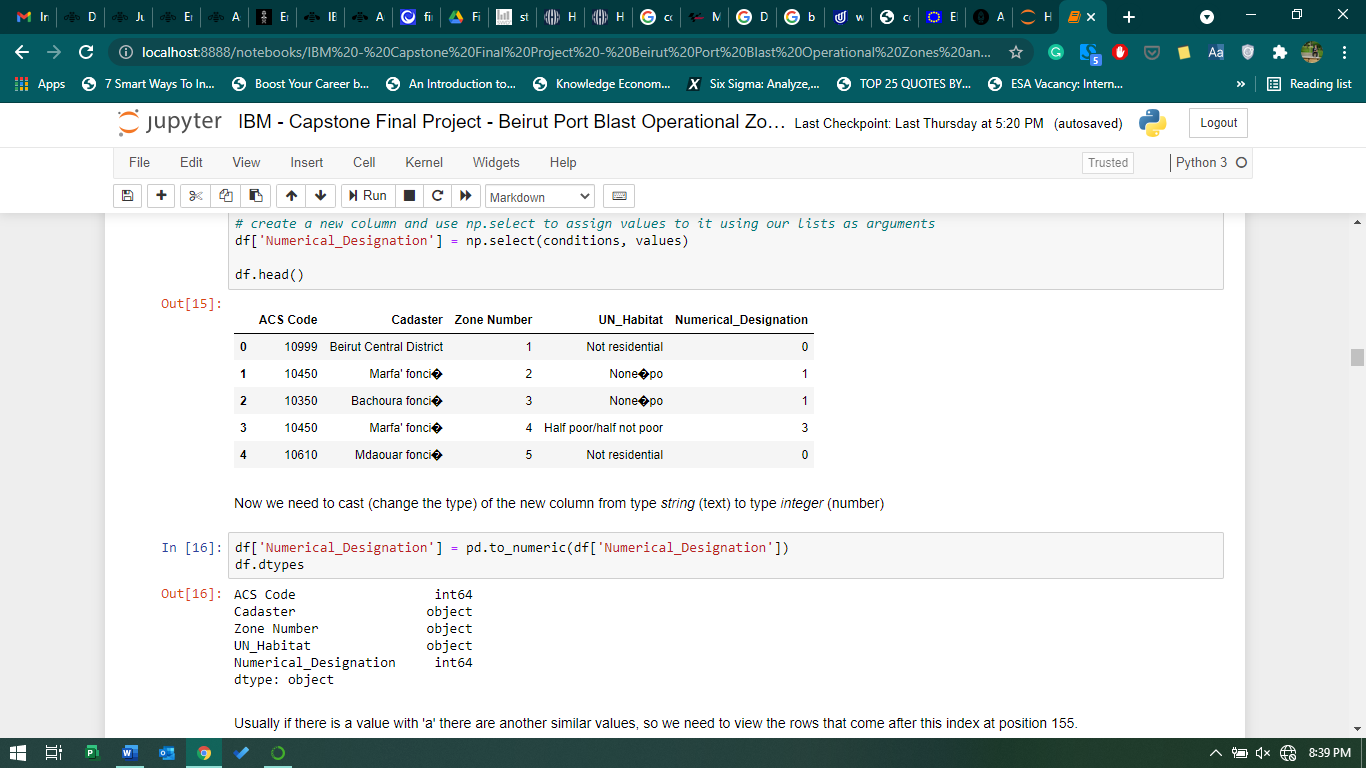
The processed data frame of the seriocomic categories is added as a layer on top of the generated zones map. As shown in the below figure, we were able to pinpoint the distribution of the six socioeconomic categories across the zones where the red label corresponds to the port explosion area:

Figure : Choropleth map pinpointing the six socioeconomic categories across the operational zones.



One main challenge was displaying categorical variables on a map. The *folium* library requires numerical data as a parameter to generate a choropleth map. Our approach was designating each category a numerical value where the basic category *Not Residential* was given the value 0 and the maximum category *All Poor* was given the value 6. Similarly, every categorical value within this range was designated to the corresponding numerical level. These designations were generated via an automated process which populated a new column “Numerical\_Designation” with the new numerical values by mapping the socioeconomic category in the *UN\_Habitat* column.

Figure : The generated Numerical Designation column as mapped with the UN\_Habitat column.



Another simple challenge was displaying the categorical legend with the correct color intensity. An existing function that incorporates Cascading Style Sheets (CSS) elements was utilized having the generated *folium* map object as the main argument.

## Comparison

The generated choropleth map was compared with the interactive map provided by OCHA[[2]](#footnote-2) – showing the number of organizations operating within particular zones – to identify if poor areas received higher attention from humanitarian organizations in comparison with affluent areas. To facilitate comparison, the color scheme of our map was selected to alter between green (minimum) and blue (maximum) to resemble that of the interactive map.

The comparison demonstrated that more than approximately threequarters of the zones received greater assistance manifested by an increased number of organizations as poverty levels also increased.

However, we selected two zones from the remaining zones for further analysis as they did not reflect the same trend. The first was zone number 62 located in “Ashrafieh” which includes few poor residents and received a relatively high number of assistance. On the other hand, the second zone was zone number 89 located in “Bourj Hammoud” which all of its residents are poor and received minimal number of assistance.

## Foursquare API

These two locations were examined in detail by passing their latitude and longitude as an API request with the “explore” endpoint to receive a list of surrounding venues in a pre-defined radius. [3] Identifying the existing venues will provide us with a more detailed look into these zones to determine if the zone categorized as poor and received minimal assistance (zone #89) is actually loosely populated with minimal business activity thus can a few number of humanitarian organizations is sufficient. The opposite is true with zone #62.

First, the GeoJSON coordinates constituting the polygon of the two zones were obtained as a Python variable of type *list* after importing the *Geo-pandas* library (We had to install the five packages listed in the Analysis Toolssection on Anaconda desktop prompt).[[3]](#footnote-3) Second, we calculated the centroid from the values in the *list*. These centroids are now the latitude and longitude of the two zones. Finally, we passed these centroids into Google maps to visually validate the results. The results were accurate and we can move forward with our analysis.

The search radius parameter of the API was set to 500m away from the centroids. The received response in the format of a JSON script was converted into a data frame via the *json* library and only columns of interest were kept as shown in the below figure:

Figure : Top 10 trending venues within 500m radius of the two zones.

|  |  |
| --- | --- |
| **Zone Number 62 in Ashrafieh** | **Zone Number 89 in Bourj Hammoud** |
|  |  |

These venues were further analyzed grouping the top categories in both zones and visualized the result as per the below two bar graphs:

Figure : Top 5 venues' categories for both zones.

|  |  |
| --- | --- |
|  |  |

As figure 9 demonstrates, the affluent zone 62 in Ashrafieh has business activities and services that are labor-intensive and require larger support such as assisted living and a building accommodating several people whereas the poor zone 89 in Bourj Hammoud has smaller businesses such as a café and a bakery.

# Discussion

As the previous analysis demonstrated, the trend observed was higher number of organizations operating in a particular zone as the number of poor residents also increased. However, visually presenting the poverty level of the zones and comparing it with a choropleth map of the similar color scheme pointed out several zones that did not accommodate this trend.

Hence, two zones representing the outliers from the trend were further processed by the Foursquare API. The intent was to explore the categories of venues within a 500m radius proximity from the centroid of these two zones.

The grouping of venues by categories and viewing their presence in the two zones (count) indicated that the more affluent zone has bigger and more labor-intensive business activities, and hosts larger residence areas. This indicates that this zone needs greater number of people assisting – in our project greater number of people accounts to greater number of organizations – to support recovery efforts. The opposite is true for the less affluent zone.

# Conclusion

The report answers two questions. The answer to the first and main question posed from the beginning of the analysis: Food distribution assistance was adequate to support the residents in a particular zone. The number of operating organizations reflected the socioeconomic status (poverty level) of the residents in every zone.

The second question was posed as we moved forward in our analysis. Some zones seemed to be marginalized while others seemed to receive undeserved assistance. Probing deeper, our answer turned out to be that the marginalized zones are actually less populous and with humble business activity, hence fewer organizations were able to provide support sufficiently. And the zones seemingly to be receiving undeserved assistance required greater number of organizations as the recovery needs were greater.

This small percentage of outliers even showed that assistance in the form of food distribution considered factors beyond poverty level such as the population, layout, and human activity in the zone during the recovery process.

# References

|  |  |
| --- | --- |
| [1] | M. Yahya, "Can Lebanon's Fires Be Put Out?," Carnegie Middle East Center, Beirut, August 6, 2020. |
| [2] | The Economist, "Beirut after the blast: the crunch of glass, acrid smoke and stairs slick with blood," The Economist, August 5, 2020. |
| [3] | Foursquare, "Foursquare Developers: Explore Our Endpoints," [Online]. Available: https://developer.foursquare.com/docs/places-api/endpoints/. [Accessed 19 July 2021]. |

# Table of Figures

[Figure 1: The original CSV dataset imported into a pandas data frame. 4](#_Toc80103900)

[Figure 2: The data frame after data pre-processing. 4](#_Toc80103901)

[Figure 3: Snapshot of the relevant part of the GeoJSON script. 4](#_Toc80103902)

[Figure 4: The 9-steps work process utilized through the project development. 5](#_Toc80103903)

[Figure 5: Count of the 6 socioeconomic categories. 6](#_Toc80103904)

[Figure 6: Choropleth map pinpointing the six socioeconomic categories across the operational zones 7](#_Toc80103905)

[Figure 7: The generated Numerical Designation column as mapped with the UN\_Habitat column. 8](#_Toc80103906)

[Figure 8: Top 10 trending venues within 500m radius of the two zones. 9](#_Toc80103907)

[Figure 9: Top 5 venues' categories for both zones. 9](#_Toc80103908)

1. HDX platform and the corresponding data available at: <https://data.humdata.org/dataset/beirut-port-explosion-operational-zones> [↑](#footnote-ref-1)
2. The OCHA Partners Presence Interactive Dashboard map available at <https://bit.ly/3k3NAGf> [↑](#footnote-ref-2)
3. Python extension packages available at: <https://www.lfd.uci.edu/~gohlke/pythonlibs/> [↑](#footnote-ref-3)