Data

There are two main sources of data that is used in this study. One is the location file created using the geojson.io website and the other one is the Foursquare Places API.

In order to analyze the region, I needed a file with location points in it, in the region to be analyzed. I was not able to find such a file and decided to create by myself.

It is possible to draw markers on the desired location on a map using the geojson.io website. One thing to decide on is the distance between the markers. I decided it to be around (not less than and as close as possible to) 300 meters. In combination with this, I set the 'radius' parameter using the Foursquare Places API to be 150 meters. This way, I tried to make sure that:

- The area to be analyzed is covered as much as possible
- No venues are duplicated (a venue is counted and considered only once)

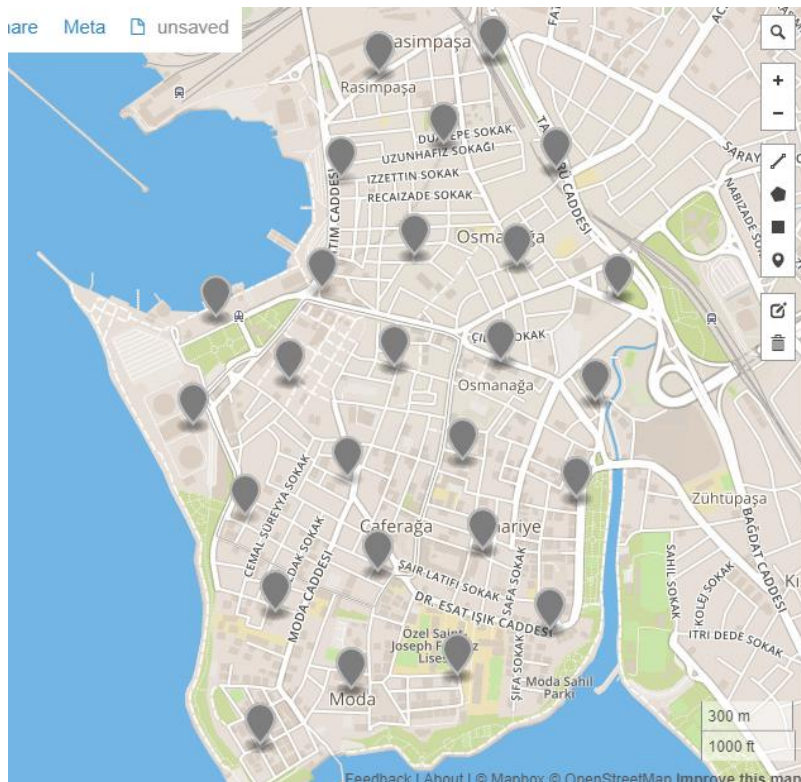The resulting map is as below:

Image 1: Map of the region to be analyzed, with markers on it

Of course it is possible to select different distances instead of 300 meters (for min distance between two points) and as 150 meters as radius for the Foursquare Places API. I decided that this is suitable for the purposes of this study.

Another thing I had to do was to name these markers, as they do not correspond to any specific neighborhoods or location units. I named them as below, starting with A1 at the bottom-left and increasing as 2,3,4 to the right and as B,C,D upwards as you can see below.
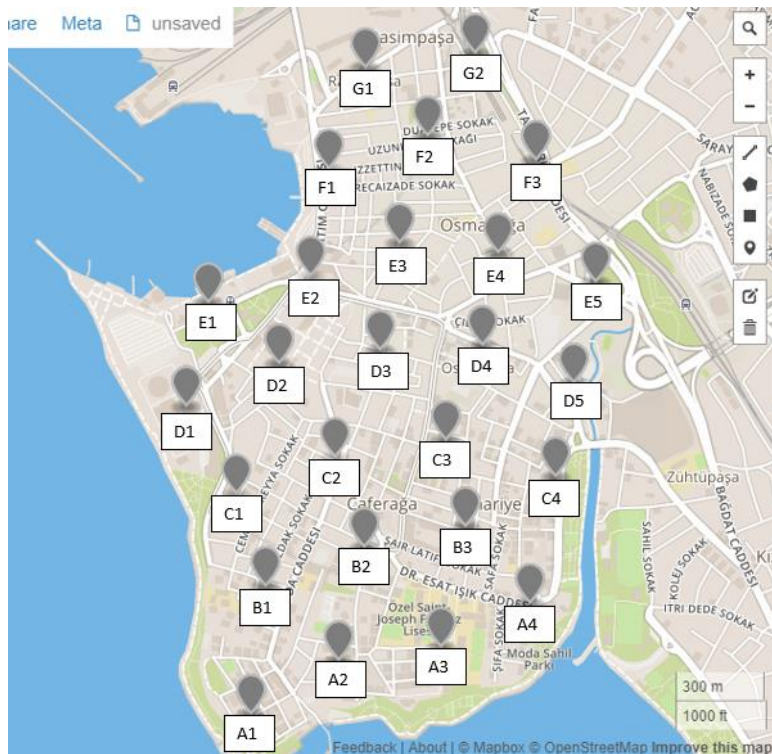
Image 2: Naming of the markers on the map

I saved this as a csv file (1) where each line in the file contains the marker name, together with the latitudes and the longitudes of the marker, which is what is needed for the analysis.

Below you can see a screenshot from this file:

	lon	lat	marker-color	marker-size	marker-symbol	name
2	29.02231693267822	40.979962866583264	#7e7e7e	medium		A1
3	29.02538537979126	40.9814045825477	#7e7e7e	medium		A2
4	29.028990268707275	40.98174475812385	#7e7e7e	medium		A3
5	29.03214454650879	40.98291106106472	#7e7e7e	medium		A4
6	29.029870033264157	40.984952041582275	#7e7e7e	medium		B3
7	29.026308059692383	40.9844013070302026	#7e7e7e	medium		B2

Image 3: Screenshot of the csv file with location data

The other data source that is used in this study is the Foursquare Places API. The code generated with the required parameters (which includes User credentials, coordinates of the point whose surrounding will be searched and the radius of the search) sends a request to this API and receives back the requested data about the venues. This data includes the name, address, category and certain other properties about the venue. Afterwards, this received data is further processed to extract meaningful information for our analysis.

References

(1) File with the location data - map - kadikoy - v5.csv
https://github.com/maydinel/Coursera_Capstone/blob/master/map%20-%20kadikoy%20-%20v5.csv