

Haifa's Business Exploration

By: Roee Groiser

Introduction

We want to test the feasibility of starting a big business of some type in Haifa. Haifa is heterogenous city, it has different neighborhoods differ by culture, economic level, the nature of the neighborhood and the residents. Starting a business may be a good choice in some neighborhood while it might not be the same in other place. We want to build a model that takes all neighborhoods, divides them to some clusters of similar neighborhoods and by this division to see which neighborhoods are fit to an entrepreneur's goals. The assumption of my research is the fact that the answer whatever to start a new business or not will be the same for other similar neighborhood. It is important to note that the answer for the asked question is binary – yes or no, and there is no measure of the probability that the business will succeed or not.

Data

We have 2 main components of data:

The neighborhoods: we should collect Haifa's neighborhoods by mining them from a html code of Wikipedia page containing the list of the neighborhoods. For each neighborhood name, we should get the geolocation (pair of latitude and longitude) of the neighborhood using Geopy python package.

The venues: after we collected foreach neighborhood its name and geolocation, we can use FourSquare and OpenStreetMap APIs in order to collect location data of venues located in Haifa, and bind them to the closest neighborhood. The relevant data we use from those APIs are the geolocation of the venues and their categories e.g. café, hotel, restaurant etc.

Binding between each venue to its neighborhood is done by using the geolocations of them and binding to the closest neighborhood.

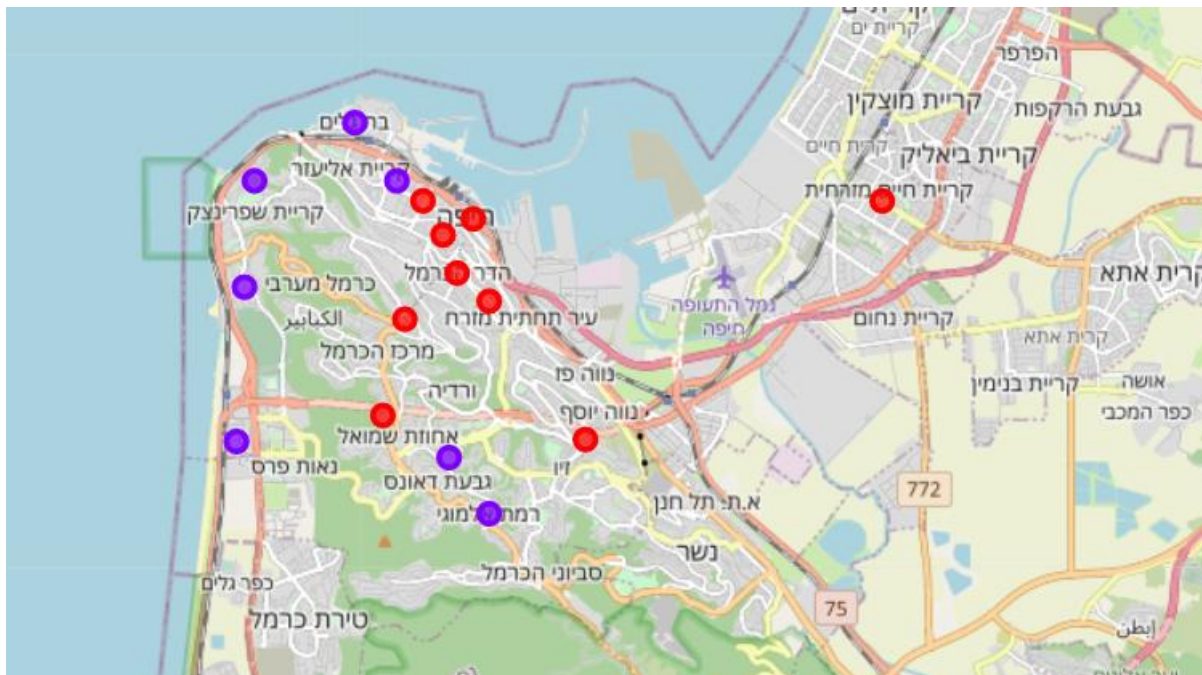
Methodology

After we collect all the data above, we join between the data types i.e. making binding between each venue to its closest neighborhood as explained above. After binding data, we should start the data preparation stage – we should explore our data and explore only neighborhoods with enough data in order to get reliable results. We decided to take only the upper median of the neighborhoods with most venues. Then, we encode the categories of the venues as 1-hot vector encoding which helps us to build a profile of each neighborhood. Next, we group our data by

neighborhoods, normalize it with calculating the mean of each category occurrence for each category and neighborhood and apply the data with KNN clustering algorithm when $k=2$, to areas where it's recommended to start a new business and other areas.

Results

The best way to emphasize the results is by viewing this map



The red marks symbolize the areas in Haifa where it's recommended to start new business since they are well-known areas in Haifa for culture, leisure and food (See observations section). The purple marks symbolize resident neighborhoods with less attractions.

I compare the results to my own ground truth – my knowledge about Haifa (as someone who grew up in this city). The confusion matrix is as described below:

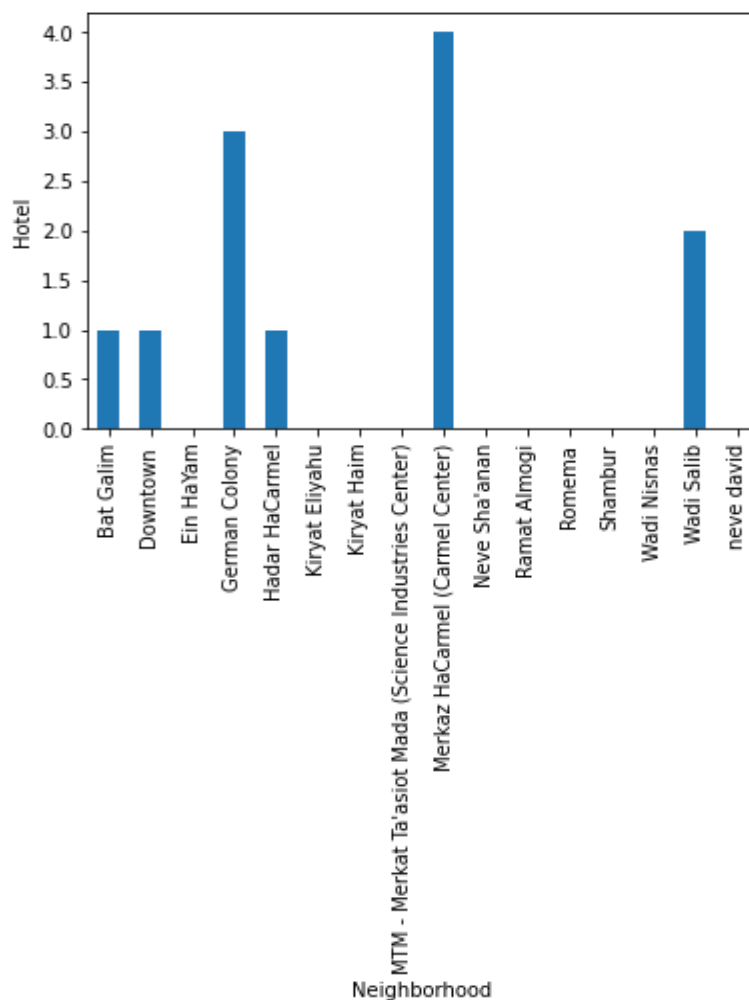
	Positive	Negative
True	7/16	6/16
False	2/16	1/16

To sum up this section, the built model was accurate on ~81% (TP+TN) of the examined neighborhoods.

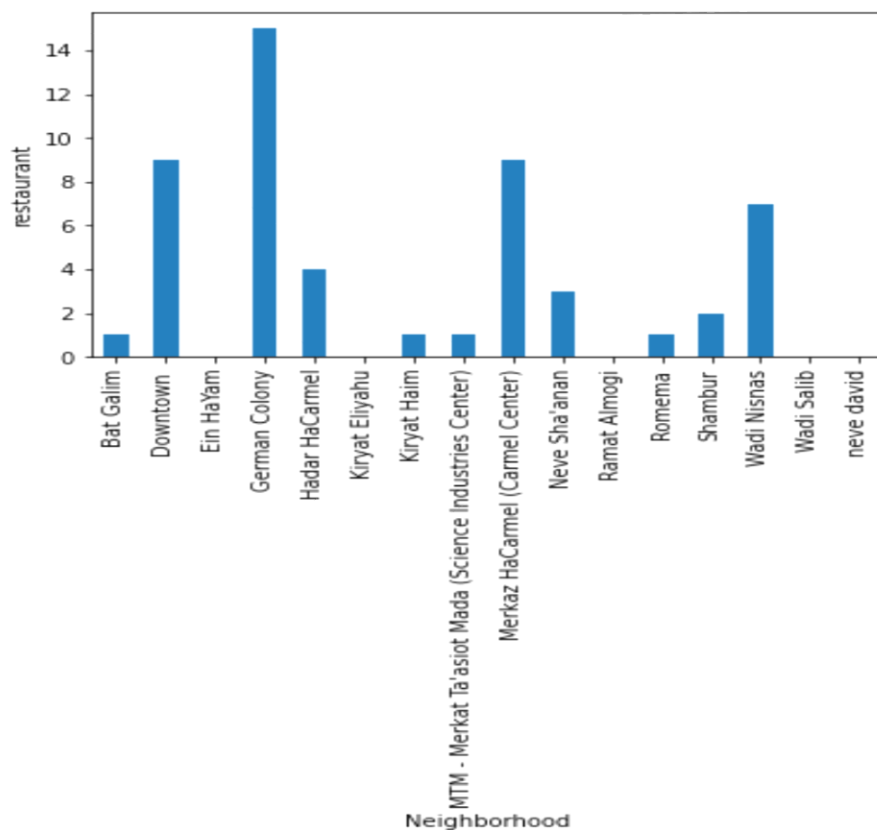
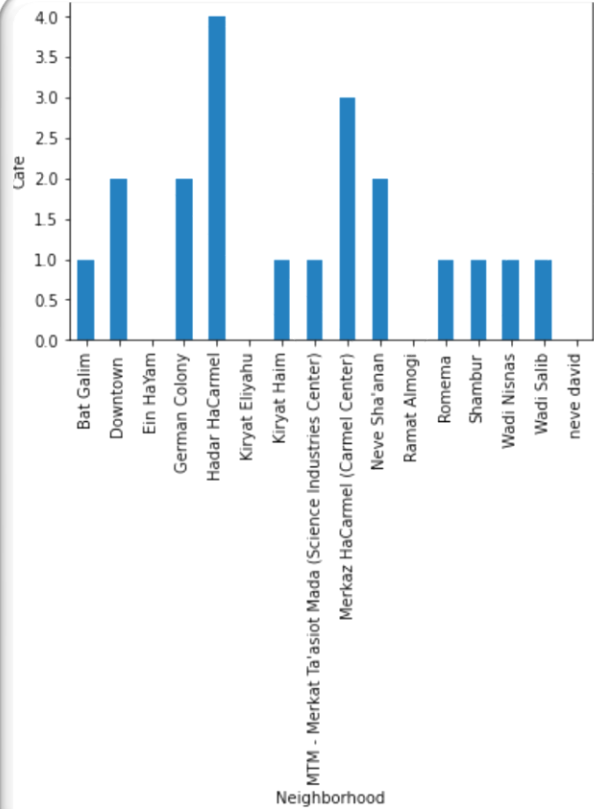
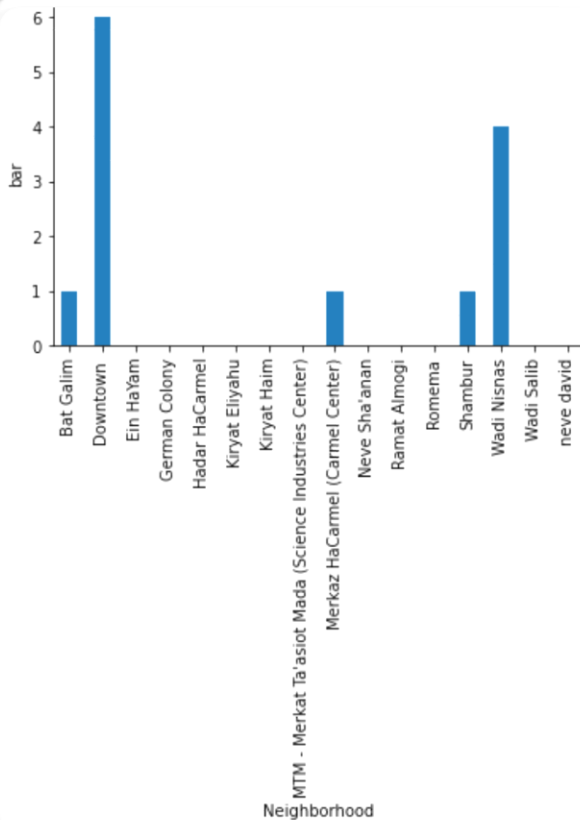
Observations

During my exploration I observed in some interesting facts that are reflected with my knowledge:

A. The specific location of the hotels – Haifa do not have a lot of hotels, and those who are in the city are located in some central areas, or in locations with a wide view to the Galilee and Haifa's bay. This fact was reflected when I search for the number of the hotels per neighborhood.



B. The relation between the number of food businesses, hotels and cafes to the number of POIs located in those areas – I observed that areas with cafes, restaurants etc. considered to be more popular – based on my knowledge and the data I collected and showed in the following figures:



The figures above show the number of café, restaurant and bar places per neighborhood. We can see that areas like Wadi Nisnas, Downtown, German Colony and Merkaz HaCarmel. has much more businesses then other neighborhoods, and they are considered popular based on the following table shows how many businesses and POIs located in each area:

Neighborhood	count
Merkaz HaCarmel (Carmel Center)	70
Downtown	62
German Colony	47
Wadi Nisnas	43
Hadar HaCarmel	32
Bat Galim	19
Shambur	17
Neve Sha'anan	16
Kiryat Haim	10
Wadi Salib	10
Ramat Almogi	9
Romema	9
neve david	8
MTM - Merkat Ta'asiot Mada (Science Industries...	8
Kiryat Eliyahu	7
Ein HaYam	7

Conclusions:

This was an amazing and interesting research. I discovered that I don't know Haifa as much as I thought I know since I saw that Haifa has many areas, squares, and neighborhoods. I even discovered that Haifa is more various than I knew. On the other hand, I think that the results I got in my classifier are fit to the reality and what I know about Haifa and its well-known areas for leisure, vacation and tourism areas. That's my chance to say 'thank you' to Coursera and IBM for a great, interesting and instructive course. It opened my eyes and my mind for an interesting and newly topic in science and I will take another courses of this topic, striving to get a job related to Data Science. Thanks for reading.