JULY 2020

ITALIAN TAKE AWAY IN THE HEART OF AMSTERDAM

CAPSTONE PROJECT

1. Introduction

1.1 Background

Italian food is famous worldwide and many people travelling around the world often look for Italian restaurants, especially if travelling in Europe.

This is true also for the locals, who increasingly often want to try something different from their daily diet.

However, as I've been told by the Italian owner of a “Pizzeria” in a small village in the Netherlands, if you really want to make the Italian food appreciated by locals, you need to adapt it a little to the locals' taste.

I do not totally agree with this sentence, as I am a strong supporter of each country’s tradition and for example, I would prefer a traditional dish instead of a traditional but revisited one. But I understand that the globalization got us used to have all at hand and the possibility of trying new experiences often prevails on the real sense of the experience itself.

Aside from these philosophical thoughts, which are not properly linked to data science and to this project but might help in getting my point of view on the problem, the project topic is introduced in the next paragraph.

1.2 The problem

Italians are used to have a structured lunch with more than one course, while most people especially in northern Europe usually have a rapid lunch with a sandwich or a slice of pizza. In general, Italian restaurants outside Italy provide for the customers to have a sit and take some time to eat and have a conviviality moment. But this sometimes does not fit with the real needs. Thus, instead of trying to adapt Italian dishes’ taste to that of the locals, one could think of trying to serve the Italian food in a more “local style”.

The aim of the project is to predict the best location for an Italian take away in the heart of Amsterdam. The shop will be a place where you can buy sandwiches or lunch boxes all made with 100% imported Italian raw materials.

Data that might contribute to determining the best position for the shop will include the proximity to offices in the city centre, but also the distance from universities or parks where people could decide to consume the food.

The most promising neighbourhoods will be found based on these criteria and using the data science instruments. Advantages of each area will be finally expressed so that the best possible final location can be chosen by stakeholders.

1.3 Interest

This study is targeted to stakeholders who want to open an Italian restaurant in Amsterdam trying to conciliate the Italian food authenticity with the local culture in an innovative way. Lots of tourists and local workers could be interested in having the possibility of eating some fresh and genuine Italian food during their lunch without having to give up to an open-air break.

2. Data acquisition and usage

2.1 Data acquisition

Based on the problem definition, factors that will influence the decisions are:

- number of existing restaurants in the neighbourhood (fast-food)
- number of and distance to Italian restaurants in the neighbourhood (if any)
- distance of neighbourhood from city centre
- distance of neighbourhood from universities and offices
- distance of neighbourhood from main parks and green areas

A regularly spaced grid of locations, centred around the city centre, will be used to define the neighbourhoods.

To extract and generate the required information, the following data sources will be used:

- centres of candidate areas will be generated using algorithms and approximate addresses of these centres will be obtained using geopy, a Python client for several popular geocoding web services
- the number of restaurants, their type and location in each neighbourhood will be obtained using Foursquare API
- the number of offices, parks and universities as well as their location in each neighbourhood will be also taken from Foursquare API
- the coordinates of Amsterdam centre will be obtained using geopy and the address for the geocoding will be set as that of the well-known Dam Square.

Latitude and longitude coordinates for centroids of the candidate neighbourhoods will be obtained by creating a grid of cells covering the area of interest which is approximately 3x3 kilometres centred around Dam Square.
All data extracted from Foursquare will be cleaned and processed with some basic explanatory data analysis in order to find the best neighbourhood in which stakeholders can invest in an Italian take away, given all the conditions specified.

2.2 Data cleaning

Prior to downloading all the required data from Foursquare, areas of interest have been identified.

Latitude and longitude of Dam Square have been retrieved using geopy geocoder and a grid of possible neighbourhoods has been created by dividing a 3x3km area around Dam Square in squared areas with side equals to 300m (Figure 1a). Centres of these zones have been approximately calculated using mathematical formulas and the corresponding addresses were obtained using geolocator reverse geocoding. Addresses of area centres, their latitude and longitude as well as their distance from Dam Square have then been placed into a Data Frame.

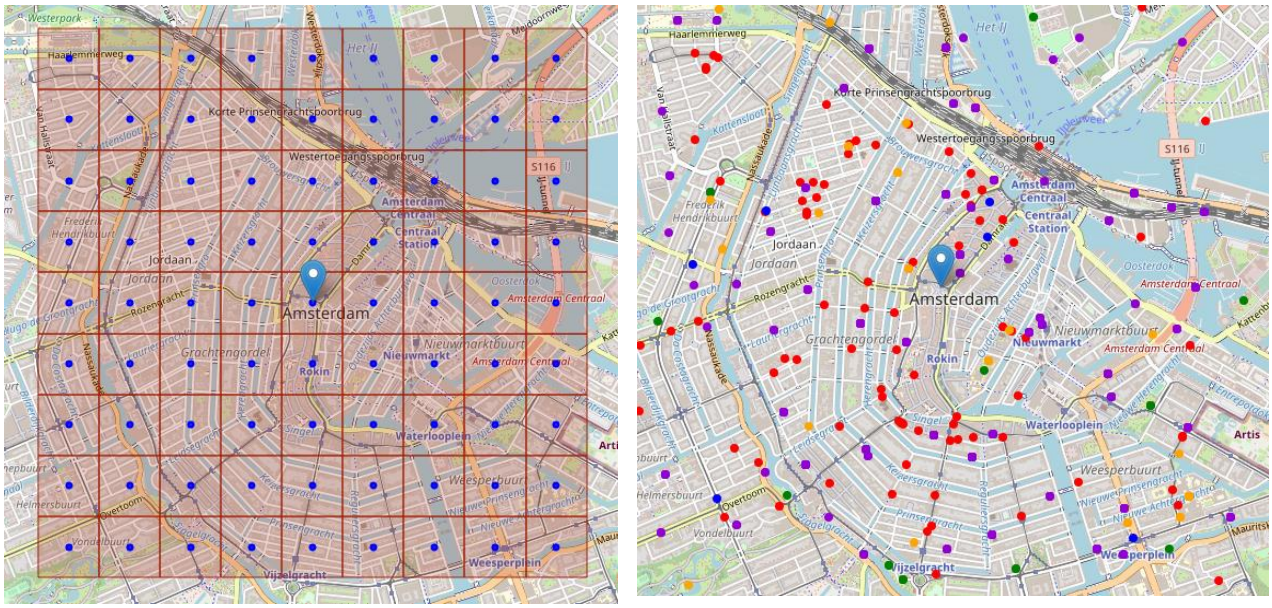
Next, relevant data for the analysis have been downloaded from Foursquare API. The interest was in venues in 'food' category, but only those that are not real restaurants, as the new location is a take away. However, as the take away would serve Italian food, Italian restaurants have been included in the search. In addition, the search has comprised parks, universities and offices within the defined neighbourhoods.

All venues retrieved using Foursquare are plotted in Figure 1b.

One can distinguish each venue category using the following colour legend:

- Red dots: Italian restaurants
- Blue dots: fast food
- Orange dots: universities
- Green dots: parks

- Violet dots: offices



(a) *(b)*

Figure 1 Folium plots of designated areas of interest in the heart of Amsterdam. Grid of square shaped neighbourhoods around Dam Square (a) and Foursquare venues within the defined areas (b)

Venues have been searched within a radius of 500m around each area centre, to ensure not to miss data. Any duplicates have then been removed.

3. Methodology

Basic instruments given by exploratory data analysis have been used to inspect the collected data, with the objective of detecting areas around Amsterdam centre that have a low Italian restaurants density and are close to green areas, universities and offices at the same time.

Considering the areas defined in the previous paragraph, number of Italian restaurants, fast food, universities, offices and parks has been calculated for each candidate neighbourhood and results have been collected and added to the previously defined Data Frame.

The same procedure has been repeated to get the distance to the closest venue in each category, regardless of how distant it is.

Heatmaps have then been used to identify a few promising areas close to Dam Square with low number of Italian restaurants and high density of university, parks and offices. These neighbourhoods have been further analysed.

Comparing the two pockets obtained by the first part of the analysis, both meeting the requirements for the candidate locations, attention has been focused on the one located south-east from Dam Square. This small area, which is part of the Centruum Oost neighbourhood, is roughly centred around Nieuwmarkt Square.

The new area has been divided in small neighborhoods and clustered using k-means algorithm, to find the best locations for the Italian take away. The addresses obtained at the end of the process constitute the list of promising location to be used as a starting point for a final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis and Results

First, the number of Italian restaurants, offices, parks and universities has been calculated for each candidate neighbourhood and added to the Data Frame (DF). The first 10 records in the DF are shown in Figure 2.

	Address	Latitude	Longitude	Distance from centre (m)	Restaurants in area	Parks in area	Universities in area	Offices in area
0	124, Vondelstraat, Vondelbuurt, Amsterdam	52.361261	4.872937	1697.0569	0	0	1	5
1	Helmersbuurt, Amsterdam	52.364256	4.872937	1592.6799	1	0	0	4
2	53-3, Kinkerstraat, Oud-West, Amsterdam	52.367252	4.872937	1513.7249	2	0	0	1
3	11-1, Allard Piersonstraat, Da Costabuurt, Ams...	52.370248	4.872937	1464.3098	3	1	0	3
4	De Poort, 18, Hugo de Grootkade, Frederik Hend...	52.373243	4.872937	1447.4631	1	2	1	3
5	33-H, Eerste Hugo de Grootstraat, Frederik Hen...	52.376239	4.872937	1464.3094	2	3	2	5
6	228, Van der Palmkade, Staatsliedenbuurt, Amst...	52.379234	4.872937	1513.7242	2	2	1	3
7	79-H, Groen van Prinstererstraat, Staatslieden...	52.382230	4.872937	1592.6790	6	0	0	2
8	141-H, Haarlemmerweg, Amsterdam	52.385226	4.872937	1697.0557	5	1	1	9
9	17, Roemer Visscherstraat, Vondelbuurt, Amsterdam	52.361261	4.877832	1401.2045	0	1	0	10

Figure 2 Data Frame with all candidate locations address, coordinates, distance from Dam Square and n# of venues within the area

Next, the distance from each area centre to the closest venue in each category has been evaluated and saved to the DF (Figure 3).

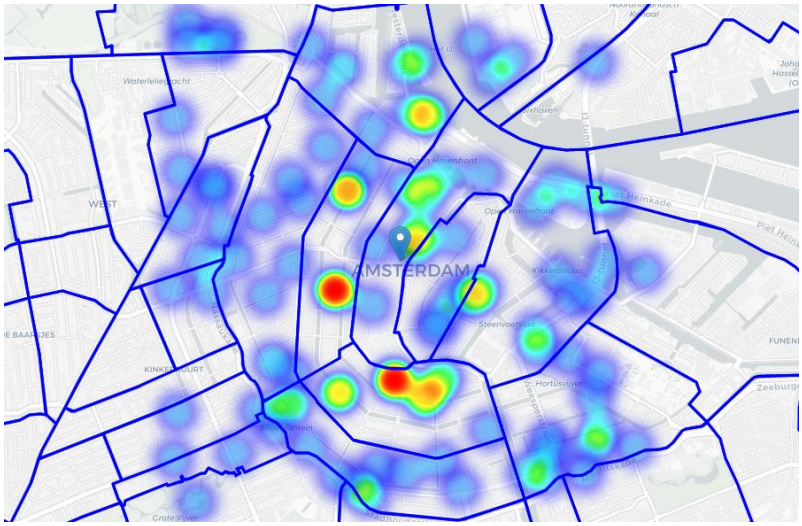
	Address	Latitude	Longitude	Distance from centre (m)	Restaurants in area	Parks in area	Universities in area	Offices in area	Distance to Italian restaurant	Distance to universities	Distance to parks	Distance to offices
0	124, Vondelstraat, Vondelbuurt, Amsterdam	52.361261	4.872937	1697.0569	0	0	1	5	211.9465	165.0499	583.2698	168.9166
1	Helmersbuurt, Amsterdam	52.364256	4.872937	1592.6799	1	0	0	4	245.7676	386.3762	566.0717	145.9762
2	53-3, Kinkerstraat, Oud-West, Amsterdam	52.367252	4.872937	1513.7249	2	0	0	1	280.9397	607.8141	375.7659	259.5786
3	11-1, Allard Piersonstraat, Da Costabuurt, Ams...	52.370248	4.872937	1464.3098	3	1	0	3	114.0842	406.3952	213.9512	129.1034
4	De Poort, 18, Hugo de Grootkade, Frederik Hend...	52.373243	4.872937	1447.4631	1	2	1	3	125.2630	231.7740	220.1286	23.2327
5	33-H, Eerste Hugo de Grootstraat, Frederik Hen...	52.376239	4.872937	1464.3094	2	3	2	5	232.9967	148.9331	167.1853	128.0786
6	228, Van der Palmkade, Staatsliedenbuurt, Amst...	52.379234	4.872937	1513.7242	2	2	1	3	86.0326	174.7848	154.7782	126.6486
7	79-H, Groen van Prinstererstraat, Staatslieden...	52.382230	4.872937	1592.6790	6	0	0	2	141.4246	363.0891	343.6102	167.9922
8	141-H, Haarlemmerweg, Amsterdam	52.385226	4.872937	1697.0557	5	1	1	9	58.7600	196.3155	174.1612	98.7240
9	17, Roemer Visscherstraat, Vondelbuurt, Amsterdam	52.361261	4.877832	1401.2045	0	1	0	10	180.6070	406.1813	253.5164	107.3537

Figure 3 First 10 rows of the DF with the new calculated distances values

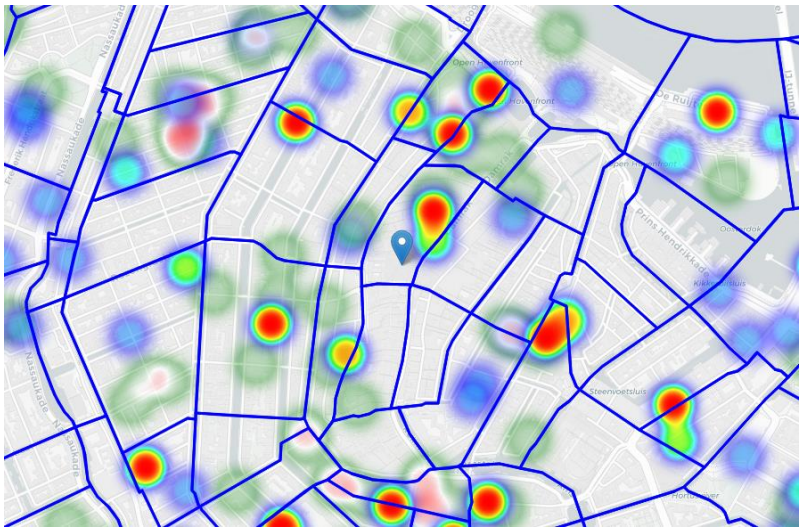
The obtained results have been used to generate *heatmaps* showing the density distribution of each venue category within the investigated area (at geographical points).

Comparing the plots in Figure 4(a-b-c), two pockets representing good locations for the Italian take away have been identified south-east and south-west from Dam Square, corresponding to the neighbourhoods **Centrum West** and **Centrum Oost** (Figure 5).

(a) Italian restaurants (green to red gradient) and fast food density



(b) Universities, parks and offices density



(c) Superposition of (a) and (b)

Figure 4 Heatmaps of Amsterdam city centre

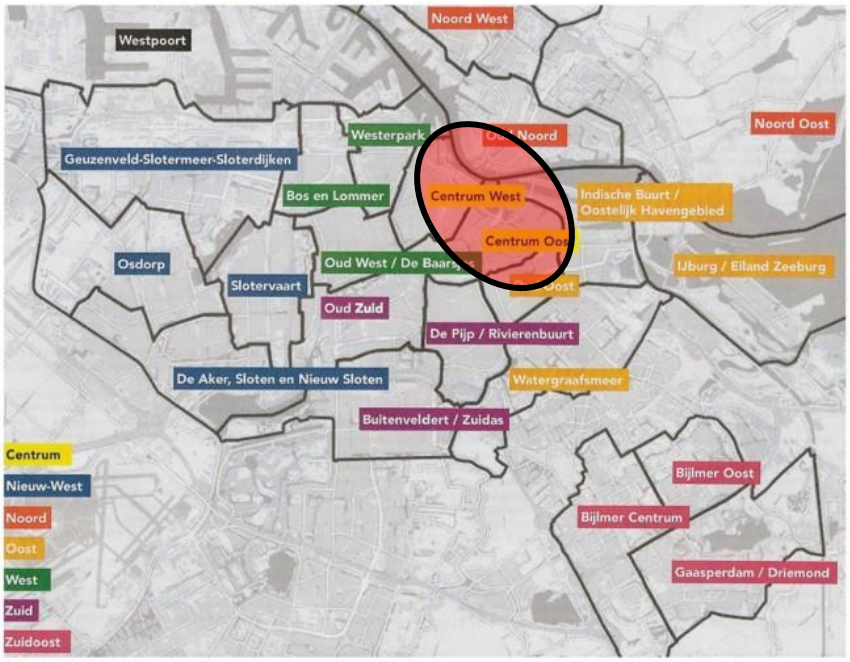


Figure 5 Map showing Amsterdam neighbourhoods

Carefully looking at both areas, **Centrum Oost** seemed to have a lower Italian restaurants density, especially south-west from Prins-Hendrikkade. A narrower region of interest has then been defined, delimited by Prins-Hendrikkade, Dam Square and the Amstel river. Almost in the middle of this area there's another quite famous square of Amsterdam, called **Nieuwmarkt**, whose coordinates have been used as the centre for this new area (Figure 6).

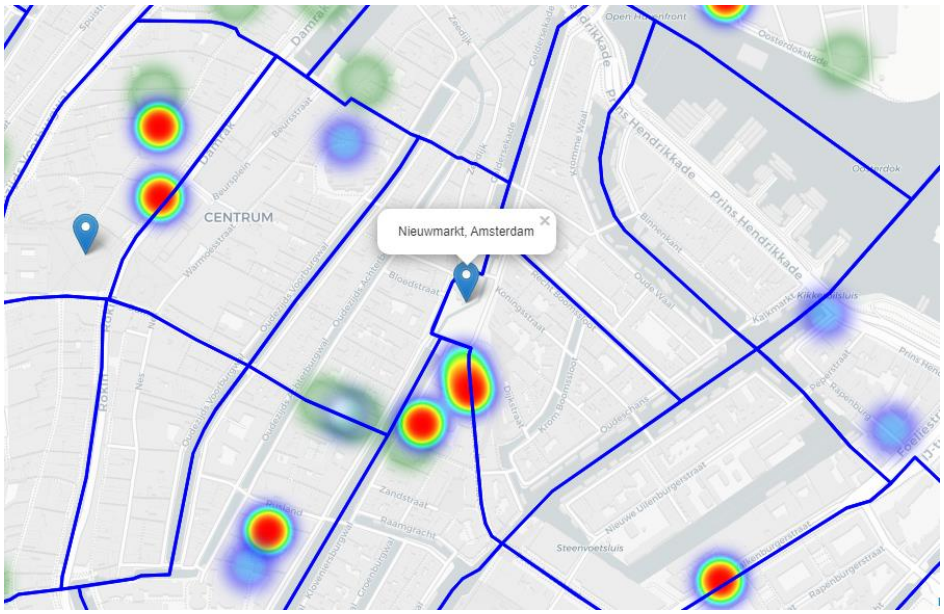


Figure 6 Nieuwmarkt position with respect to Dam Square

The new area, centred in Nieuwmarkt and expanding till Dam Square, has been divided in a new group of squared areas with side equals to 200m. 16 candidate neighbourhoods, that can be seen in Figure 7, have been generated with the same procedure used in the first step.

Figure 7 New candidate areas around Nieuwmarkt Square

For each of the new candidate locations, distances to closest Italian restaurant and number of food venues, offices, parks and universities within a radius of 250 meters have been computed and collected into a new DF (Figure 8).

	Latitude	Longitude	Distance to Italian restaurant	Italian Restaurants nearby	Universities nearby	Parks nearby	Offices nearby
0	52.369262	4.894863	236.9795	1	4	1	0
1	52.371509	4.894863	254.9652	0	4	1	0
2	52.373756	4.894863	175.8758	1	0	0	17
3	52.376002	4.894863	89.4272	6	0	0	27
4	52.369262	4.898534	112.4569	4	5	0	10
5	52.371509	4.898534	50.6084	3	1	0	19
6	52.373756	4.898534	166.3877	1	0	0	2
7	52.376002	4.898534	62.8808	2	0	0	2
8	52.369262	4.902205	335.4052	0	0	0	0
9	52.371509	4.902205	301.2514	0	0	0	9
10	52.373756	4.902205	408.7989	0	0	0	0
11	52.376002	4.902205	392.2411	0	0	0	0
12	52.369262	4.905876	580.4004	0	0	0	14
13	52.371509	4.905876	471.7030	0	0	0	10
14	52.373756	4.905876	297.7162	0	0	0	2
15	52.376002	4.905876	229.9443	1	0	0	5

Figure 8 New Data Frame comprising the coordinates of the new areas' centres and all new calculated data

At this point, locations have been filtered to consider only the ones with no more than two restaurants in radius of 200m and at least two venues among parks, universities and offices within the same radius.

The output of this step is reported below, and the three resulting locations are summarized in Figure 9.

```
Locations with no more than two italian restaurants nearby: 14
Locations with more than one park nearby: 2
Locations with more than one university nearby: 4
Locations with more than one office nearby: 11
Locations with both conditions met: 3
```

	Latitude	Longitude	Distance to Italian restaurant	Italian Restaurants nearby	Universities nearby	Parks nearby	Offices nearby
0	52.369262	4.894863	236.9795	1	4	1	0
1	52.371509	4.894863	254.9652	0	4	1	0
5	52.371509	4.898534	50.6084	3	1	0	19

Figure 9 Final optimal locations for the Italian take away

The three zones have been clustered using k-means algorithm, to create centres of zones containing good locations for the Italian take away. The obtained clusters nicely group the candidate locations and their centres are well placed in the middle of the zones 'rich' with location candidates.

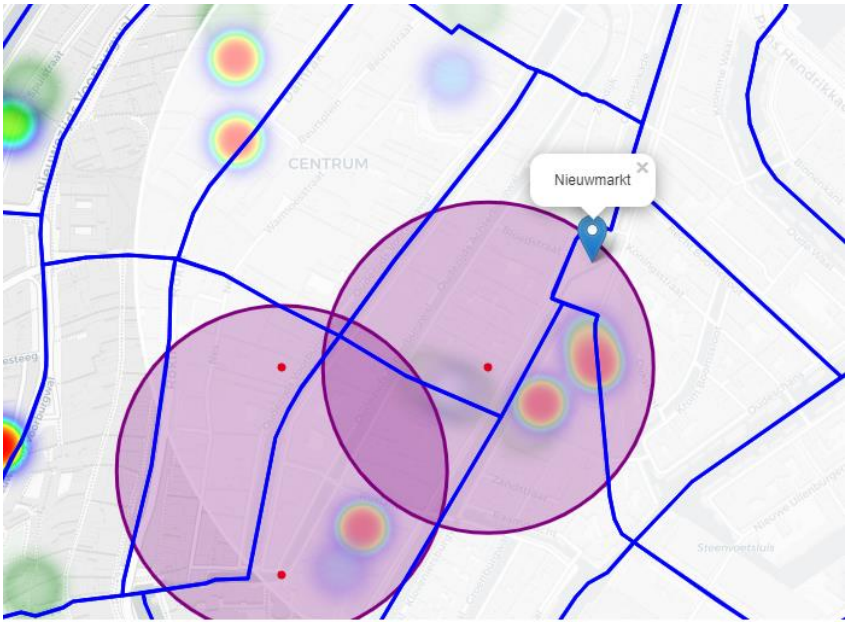


Figure 10 Cluster generated by k-means algorithm

Addresses of those cluster centres are a good starting point for exploring the neighbourhoods and to find the best possible location based on neighbourhood specifics.

Reverse geocoding has been used to obtain candidate area centres addresses to be presented to stakeholders.

```
============================================================
Centres of area candidates recommended for further analysis
============================================================

217H, Oudezijds Voorburgwal, Burgwallen-Oude Zijde, Amsterdam  => 0.3km from Dam Square
7, Boerensteeg, Burgwallen-Oude Zijde, Amsterdam               => 0.5km from Dam Square
```

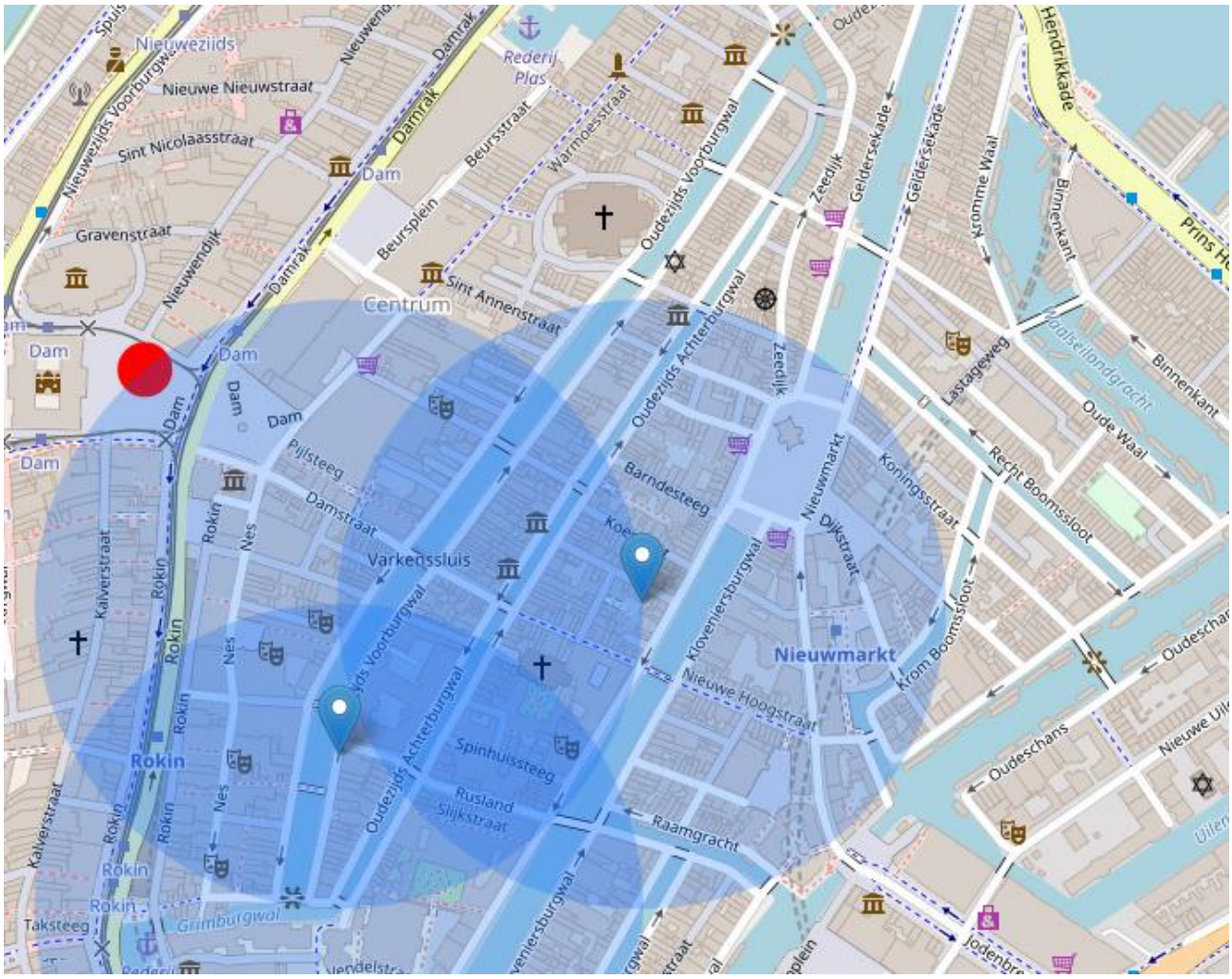


Figure 11 Position of the two clusters centres with respect to Dam Square (red dot)

5. Discussion of results

The analysis conducted to the identification of two addresses in the heart of Amsterdam, representing the centres of zones containing few Italian restaurants (less than 3) and at least two venues among parks, universities and offices in radius of 200 metres.

The good locations found are shown on map in Figure 10 with a radius of ~200 meters (violet circles) and are in the Oude Centrum neighbourhood, south and south-east from Dam Square. Their centres/addresses can be considered as a good starting point for exploring area neighbourhoods in search for potential locations for the Italian take away.

Thus, the global result consists of two small areas located 0.3 and 0.5 km south-east from Dam Square. These areas contain three location where an Italian take away could open, based on number of Italian restaurants in the neighbourhood and number of parks, universities and offices nearby. This, of course, does not imply that those zones are optimal locations for a take away. The purpose of this analysis was to provide information on areas in the heart of Amsterdam not already crowded with Italian restaurants but with offices, universities and parks in the vicinity, to allow people to take their food and have their lunch outside. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has no nearby competition.

6. Conclusions

The purpose of this project was to identify an area in the heart of Amsterdam with few Italian restaurants in the vicinity but close to offices, universities and parks, in order to help stakeholders in narrowing down the search for optimal locations for an Italian take away. The idea of the project was indeed to provide a new way to taste Italian food, trying to mix the real Italian recipes with a more European way of having lunch.

By calculating Italian restaurants density distribution from Foursquare data, neighbourhoods that justify further analysis were first identified. The choice of the neighbourhoods was also driven by the number of universities, parks and offices in the areas. Then, venues in these areas that satisfy the aforementioned prerequisites were collected and clustered to create major zones of interest. Finally, addresses of these areas' centres were obtained for future usage as a starting point for further exploration by stakeholders.

Final decision on optimal location will be made by stakeholders, based on specific characteristics of neighbourhoods and other locations in every recommended zone, considering also additional factors such as zones attractiveness.