From Coffeeshop to Restaurant: Maximizing One's Amsterdam Experience

Yinlu Deng

Jan 13, 2020

1. Introduction

1.1. Background

Amsterdam, capital of the Netherlands, is Ill-known for its vibrant, unique and exquisite and food venues. Almost Iekly a new place to eat opens in this city of amazing history, architecture, art, canals, red lights, and, Ill, coffeeshop* culture. The latest is, by all means, one of the greatest tourist attractions from all over the world.

**Hereinafter by coffeeshops, as the Dutch do, I mean places where one can smoke Ied or consume it in the form of pastry sIets, etc. (although quite often there's also real coffee to drink with your marijuana cookie).*

It is a scientifically and empirically proven fact that Ied consumption, no matter in which form, increases one's appetite – in other words, you get very strong food cravings – "the munchies" – from it. There are indeed few places in Amsterdam where you can smoke green and eat food all in one place, but quite often these offer mostly snacks and even rarely – medium to poor cuisine. Thus it is strongly advised by the locals and experienced tourists to plan your coffeeshop experience beforehand – first, consume the marijuana product of choice safely and responsibly at the coffeeshop and then head directly to the food venue nearby. But how does one avoid performing a double search for such a winning combination? Our research is to fill this niche.

1.2. Business Problem and Interest

Different groups of people would benefit from our project's results, namely:

1. Tourists planning their first experience as described above.
2. Accomplished Amsterdam visitors or even locals looking for new ideas for their following from coffeeshop-to-restaurant tour.
3. Tourists or locals that, on the contrary, want to avoid dining at a place close to a coffeeshop.
4. Potential restaurant owners that consider location proximity of their food venue to a coffeeshop as one of the important success factors.

To achieve this, I will create a short and simple guide on where to smoke and eat in Amsterdam based on Foursquare likes, restaurant category and geographical location data for restaurants and coffeeshops. I will also cluster all the restaurants of Amsterdam by their proximity to coffeeshops so that our user could easily determine what is the best duo of places of their interest.

2. Data Acquisition and Cleaning

2.1. Data Sources

For this assignment, I will be utilizing the Foursquare API to pull the following location data on restaurants and coffeeshops in Amsterdam:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Count of Likes

Another file I used was .csv file with geographical and statistical data about all the neighborhoods of Amsterdam which was downloaded from an official government website and used as a DataFrame.

2.2. Data Cleaning

Data I acquired was clean enough to work with as it came from an official source. I had to drop only one row and ignore some columns (mostly all the information about the inhabitants of the neighborhoods), but decided not to drop it in order to be potentially used during further work on this project.

	subject	region_name	regio_type	region_code	ninhabitants	nmen	nwomen	nage_0_to_15	nage_15_to_25	n
1	Burgwallen-Oude Zijde	Amsterdam	Wijk	WK036300	4280	2340	1935	255	675	2
2	Kop Zeedijk	Amsterdam	Buurt	BU03630000	1020	570	445	50	140	5
3	Oude Kerk e.o.	Amsterdam	Buurt	BU03630001	670	365	300	30	130	3
4	Burgwallen Oost	Amsterdam	Buurt	BU03630002	1610	880	730	120	250	7
5	Nes e.o.	Amsterdam	Buurt	BU03630003	370	185	180	25	70	1

An example of a non-relevant column I decided to keep: age groups of the inhabitants.

```
Data columns (total 37 columns):
subject                       578 non-null object
region_name                   578 non-null object
regio_type                    578 non-null object
region_code                   578 non-null object
ninhabitants                  578 non-null int64
nmen                          578 non-null int64
nwomen                        578 non-null int64
nage_0_to_15                  578 non-null int64
nage_15_to_25                 578 non-null int64
nage_25_to_45                 578 non-null int64
nage_45_to_65                 578 non-null int64
nage_65_older                 578 non-null int64
nunmarried                    578 non-null int64
nmarried                      578 non-null int64
ndivorced                     578 non-null int64
nwidowed                      578 non-null int64
nimmigrant_western            578 non-null int64
nimmigrant_nonwestern         578 non-null int64
nimmigrant_marokko            578 non-null int64
nimmigrant_antiles_aruba      578 non-null int64
nimmigrant_surinam            578 non-null int64
nimmigrant_turkey             578 non-null int64
nimmigrant_other_non_western  578 non-null int64
nhouseholds                   578 non-null int64
nhh_single_person             578 non-null int64
nhh_no_children               578 non-null int64
nhh_with_children             578 non-null int64
ave_househ_size               561 non-null float64
populatio_density             543 non-null float64
area_total                    578 non-null int64
area_land                     578 non-null int64
area_water                    578 non-null int64
urbanisation_grade            573 non-null float64
address_density               573 non-null float64
geojson                       578 non-null object
lon                           578 non-null float64
lat                           578 non-null float64
```

All the available data in our data set.

3. Methodology

I started our exploratory analysis examining our dataset of 578 rows. The most relevant columns appeared to be latitude and longitude of the neighborhoods.

```
The geograpical coordinate of Amsterdam, Nederland 52.3745403, 4.89797550561798.
```

Using these parameters, I grouped our neighborhoods into 10 clusters by their latitudes and longitudes.

```
# group neighbourhoods by coordinates
n_clusters = 10
neigbourhoods_grouped = KMeans(n_clusters=n_clusters, random_state=0).fit(df_data_1[['lat', 'lon']])
neigbourhoods_grouped.cluster_centers_
```

```
array([[ 52.35418413,   4.8981259 ],
       [ 52.3656447 ,   4.99573439],
       [ 52.35477417,   4.81402566],
       [ 52.35873673,   4.93480968],
       [ 52.35146221,   4.86206048],
       [ 52.3084988 ,   4.97114465],
       [ 52.39572502,   4.93289053],
       [ 52.37687063,   4.84651167],
       [ 52.37696239,   4.78780509],
       [ 52.38700674,   4.88262484]])
```

Using our Forsquare credentials, I obtained by API all the necessary information about coffeeshops of Amsterdam, which is a huge city with hundreds of buurts and wijks (neighborhoods in Holland), so I have to limit the scope of our search and focus on top-30 coffeeshops.

Then I described the top-30 venues list to see whether there's much variance in the values.

```
count    30.000000
mean      8.476667
std       0.522384
min       7.200000
25%       8.100000
50%       8.500000
75%       8.875000
max       9.400000
Name: score, dtype: float64
```

The content itself appeared to be as expected:

	id	score	category	name	address	postalcode	city	href	latitude	longitude	n_cluster
0	4a2705a4f964a52052881fe3	8.8	Аптека марихуаны	Grey Area Coffeeshop	Oude Leliestraat 2	1015 AW	Amsterdam	/v/grey-area-coffeeshop/4a2705a4f964a52052881fe3	52.374641	4.888839	10
1	4bff1a50480eef3b4f6e9e7f	8.5	Аптека марихуаны	Coffeeshop IBIZA Amsterdam	Hemonystraat 16	1074 BP	Amsterdam	/v/coffeeshop-ibiza-amsterdam/4bff1a50480eef3b...	52.357405	4.902060	1
2	4a270064f964a52051831fe3	8.9	Аптека марихуаны	De Dampkring	Handboogstraat 29	1012 XM	Amsterdam	/v/de-dampkring/4a270064f964a52051831fe3	52.367759	4.890478	1
3	4b78952df964a5205ed82ee3	8.4	Аптека марихуаны	Amnesia	Herengracht 133	1015 BG	Amsterdam	/v/amnesia/4b78952df964a5205ed82ee3	52.375631	4.888934	10
4	4a270344f964a520f3841fe3	8.3	Аптека марихуаны	Coffeeshop Easy Times	Prinsengracht 476	1017 KG	Amsterdam	/v/coffeeshop-easy-times/4a270344f964a520f3841fe3	52.364452	4.885096	1

Having backed up everything properly, I moved to the restaurants surrounding our coffeeshops.

Their interesting featured Ire the following:

```
# The column names for the restaurants dataframe
restaurants_columns = ['id',
                       'score',
                       'category',
                       'categoryID',
                       'name',
                       'address',
                       'postalcode',
                       'city',
                       'latitude',
                       'longitude',
                       'venue_name',
                       'venue_latitude',
                       'venue_longitude',
                       'n_cluster']
```

with radius = 500 and limit = 10.
This way I got 188 restaurants, looking like:

	id	score	category	categoryID	name	address	postalcode	city	latitude	longitude	venue_name	venue_latitude	venue_longitude	n_cluster
0	5b918a0460d11b002c3228e1	8.1	Italian Restaurants	4bf58dd8d48988d110941735	Cecconi's	210 Spuistraat	1012 VT	Amsterdam	52.372017	4.888873	Grey Area Coffeeshop	52.374641	4.888839	10
1	4a27db82f964a52035941fe3	5.9	Fast Food Restaurants	4bf58dd8d48988d16e941735	McDonald's	Nieuwendijk 212	1012 MX	Amsterdam	52.373864	4.892855	Grey Area Coffeeshop	52.374641	4.888839	10
2	4a27db82f964a52036941fe3	5.6	Fast Food Restaurants	4bf58dd8d48988d16e941735	McDonald's	Damrak 92	1012 LP	Amsterdam	52.373805	4.893692	Grey Area Coffeeshop	52.374641	4.888839	10
3	4a26ffccf964a52012811fe3	8.2	Creperies	52e81612bcbc57f1066b79f2	The Pancake Bakery	Prinsengracht 191	1015 DS	Amsterdam	52.377594	4.886235	Grey Area Coffeeshop	52.374641	4.888839	10
4	4a27db7ef964a5201e941fe3	6.5	Fried Chicken Joints	4d4ae6fc7a7b7dea34424761	KFC	Damrak 87-88	1012 LP	Amsterdam	52.373967	4.894076	Grey Area Coffeeshop	52.374641	4.888839	10

Having this data, I observed that 74 of 188 restaurants Ire unique and 20 of the top 30 coffeeshops/venues had > 5 restaurants nearby.

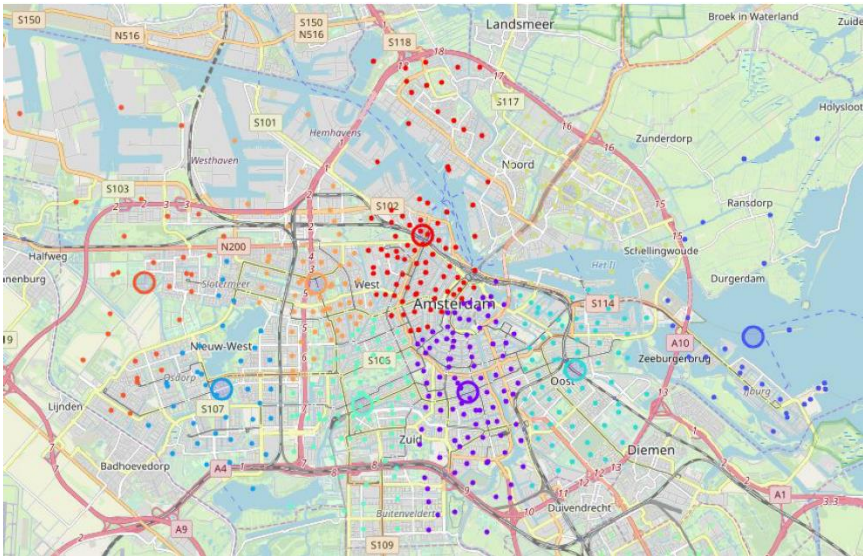
Out of 42 unique restaurant categories, top-10 unique were:

```
category
Coffee Shops              35
Fast Food Restaurants     20
Cafés                     19
Bars                      15
Burger Joints             13
Creperies                  7
Bakeries                   6
Restaurants                6
Breakfast Spots            5
Fried Chicken Joints       5
```

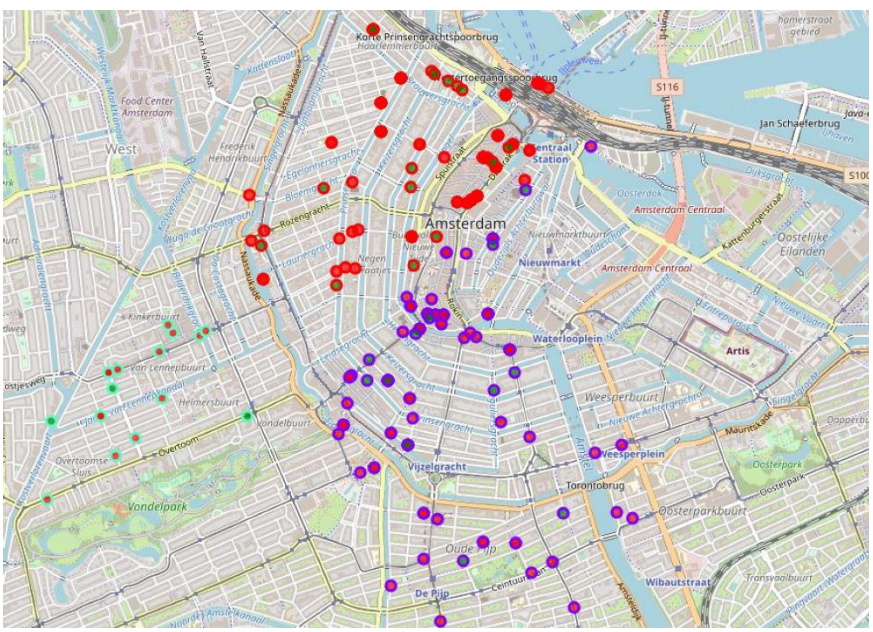
The following appeared to be places with the highest average score:

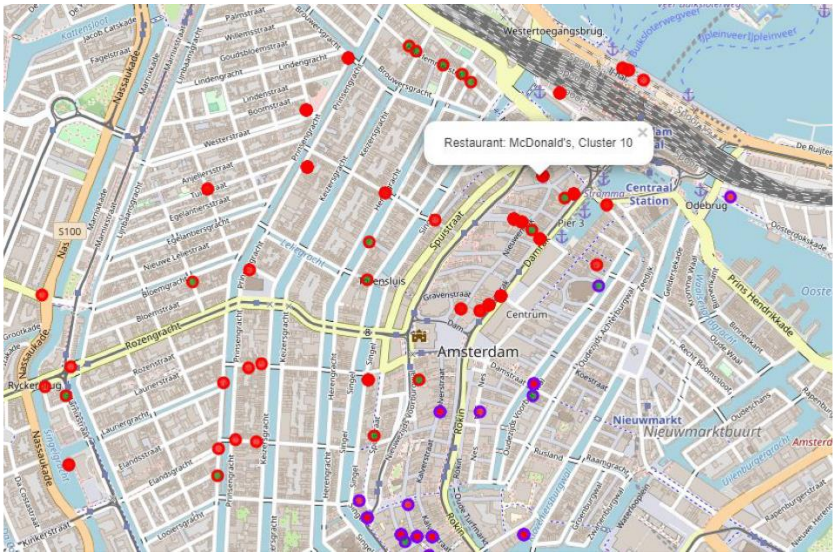
```
category
Dessert Shops             9.4
Pizza Places              9.2
Cocktail Bars             9.2
Diners                    9.1
Food Courts               9.1
French Restaurants        9.1
Caribbean Restaurants     9.1
Seafood Restaurants       9.1
Moroccan Restaurants      9.0
Steakhouses               9.0
Name: score, dtype: float64
```

The next logical step was to see all the ready for an analysis data on a map. Using Folium, I set up a map of clustered Amsterdam neighborhoods, which looked like this:

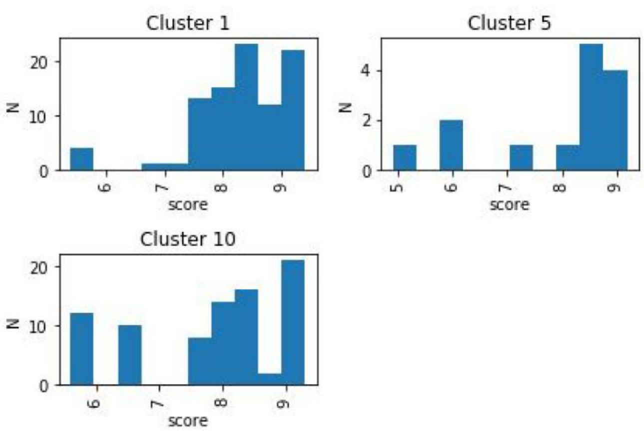


The following step was to add coffeeshops and restaurants to the same map, which I did, using red marker for restaurants and green for coffeeshops respectively. Below are our resulting maps in different close-up levels:

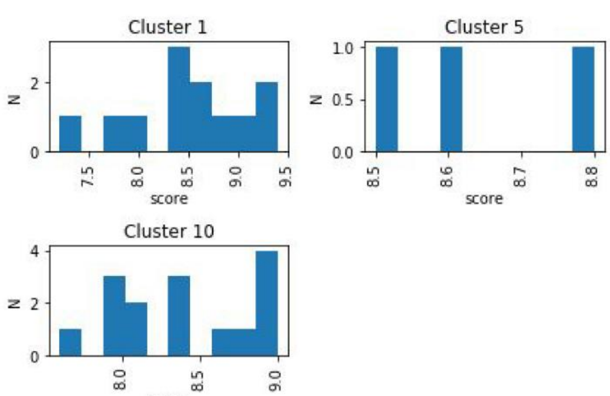




As a final touch, I visualized our clusters distribution on histograms, with these graphs representing restaurants by their number and score in relevant clusters:



and the following shoving top-venues/coffeeshops by their number and score:



4. Results

There's a variety of top-level restaurants in the centrum of Amsterdam which offer different types or food to one's taste. Great number of them is located nearby (within 500m) top-rated from coffeeshops, which proves our main hypothesis stated in Introduction. Neighbourhood clusters which have the biggest number of said top venues are 1, 5 and 10, which are, as expected, located in the centrum of the city, close to Amsterdam Central railway station, popular tourist attractions and so on. On the contrary, the further from the center, the less coffeeshops with enough good restaurants around them I observe. So if a visitor of the city seeks this specific kind of experience with maximal proximity and minimal commute, it is strongly advised for him or her to choose one of the neighborhoods from the clusters 1, 5 or 10.

5. Discussion

As mentioned in the Results section, all the top-30 coffeeshops of Amsterdam belong to three clusters and that brings us a new theory to discuss whether or not it might be economically beneficial to widen the geography of said venues by creating new or moving old ones to different neighborhoods. Such an action would make it possible to attract more tourists to other areas located out of historical centrum of the city, as it is currently an outspoken priority of the local government due to enormous overcrowding.

6. Conclusions

As our research shows, Amsterdam is indeed a capital of top coffeeshops and excellent food venues. Observing them clasterized made it easier to draw conclusions, such as extreme centralization of tourist attractions, which typically causes huge problems with housing, congestion, unemployment, air pollution, social problems and energy tension.

While our research is focused on Food Venues only, other possible categories can also be used for the same implementation (e.g. proximity to coffeeshops) such as Nightlife, Hotels etc. I have chosen to limit the scope of our research due to Foursquare API daily limit of free user queries. There are also other limitations such as the fact that the accuracy of data I used purely depends on the data provided by FourSquare.