

Finding Locations to Open a Bar in Amsterdam

IBM DATA SCIENCES CAPSTONE PROJECT

Qingting.Song April 2019

Introduction and Business Problem

Amsterdam is one of the most popular European cities for travellers. According to the prediction of Netherlands Board of Tourism and Conventions, 24 millions of people has visited Amsterdam in 2017, and this number will reach 31 millions in 2030. The popularity of this city, has made it an ideal place for services oriented businesses such as hotel, restaurants and bars.

A wealthy retired banker is looking for a good location in Amsterdam to open up a bar, and hope that the revenue generated by this business can support his retirement life in a Greek island. Since living cost in southern European is lower than western European, he doesn't need to open a big bar. As long as the cash flow is steady, he would be happy.

In order to find the best location of a small bar, he has asked data science team to help him out.



Data Collection

- We need to collect names of all the neighbourhoods of Amsterdam. This can be achieved by scraping the Wiki page:
https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam

```
url = requests.get('https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam').text
soup = BeautifulSoup(url,'html.parser')
```

We use BeautifulSoup package to conduct our Web Scraping and store data as Dataframe.

- Get Geo data via Geocoder, <https://geocoder.readthedocs.io/index.html>. This will request and pull all the latitude and longitude data of Amsterdam Neighbourhoods which will be used to create map of Amsterdam. As usual, all data that we collected will be converted to pandas Dataframe first.

```
def get_latlng(neighborhood):
    # initialize your variable to None
    lat_lng_coords = None
    # loop until you get the coordinates
    while(lat_lng_coords is None):
        g = geocoder.arcgis('{}, Amsterdam, Netherlands'.format(neighborhood))
        lat_lng_coords = g.latlng
    return lat_lng_coords

# call the function to get the coordinates, store in a new list using list comprehension
coords = [get_latlng(neighborhood) for neighborhood in AMS_df['Neighborhood'].tolist()]
# create temporary dataframe to populate the coordinates into Latitude and Longitude
df_coords = pd.DataFrame(coords, columns=['Latitude', 'Longitude'])
df_coords.to_csv("df_coords.csv", index=False)

# merge the coordinates into the AMS_df dataframe
AMS_df['Latitude'] = df_coords['Latitude']
AMS_df['Longitude'] = df_coords['Longitude']
AMS_df.head(15)
```

	Neighborhood	Latitude	Longitude
0	Admiralenbuurt	52.372728	4.856362
1	Amsteldorp	52.360420	4.905250
2	Amsterdam Oud-West	52.365390	4.870220

- Request Foursquare data of all the venues in Amsterdam(via Foursquare developer API). We will merge this dataset with geographic data of Amsterdam and use this combined dataset as input of clustering analysis, and find out the most suitable location for opening a small bar.

Methodology

In this project, we will use simple K-means clustering to conduct the unsupervised machine-learning task. K-mean identifies k number of centroids and allocate nearby data points according to certain similarities. We have imported machine-learning library: sklearn to conduct K-mean clustering.

We will achieve our goal of finding a suitable location by following those steps

1. Import python libraries that are needed for performing analysis
2. Fetch and Consolidate data from different data sources
3. Explore data
4. Conduct K-mean clustering
5. Inspect the results and draw conclusion

Results and Discussion

After we consolidated three data sets: Amsterdam Neighbourhood name, geographic data and Foursquare venues data. We did some simple analysis. Hotel is the most popular hospitality business, followed by Coffee Shop, restaurant, café and Bar. Opening a bar seems to be a reasonable idea, it has less overhead costs than Hotel and restaurants and this business is less competitive. In one night, people only have dinner once and sleep in one hotel, but they do tend to go to couple of different bars.

	Count
Hotel	667
Coffee Shop	454
Restaurant	449
Café	375
Bar	373
French Restaurant	292
Park	223
Bakery	203
Breakfast Spot	169
Bookstore	153

NOTE: Coffee Shop in Amsterdam doesn't sell coffee, they actually sell cannabis for personal consumption

Our client doesn't want to open a bar in the most competitive neighbourhood. However, he also intends to avoid the area where nobody wants to open a bar, since people love to hop from one bar to another during their night out. Majority of fun-seeking people tend to have drinks in different bars in one night. So, his business needs certain amount of neighbours.

Neighbourhoods such as :Houthaven and Watergraafsmeer have highest number of bars. Neighbourhoods such as : Bos en Lommer and Spaarndammerbuurt have second highest number of bar. Those are the neighbourhood which we don't recommend.

The following table shows Neighbourhood with the highest amount of Bars

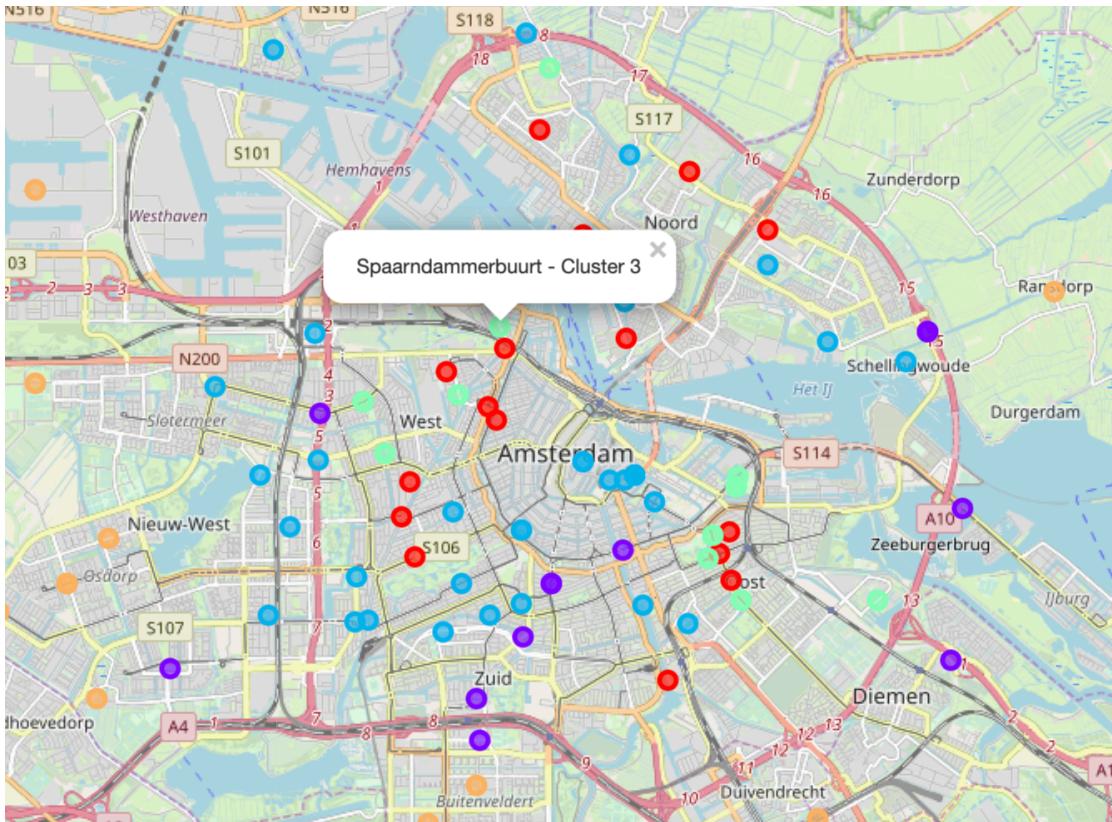
	Count
Houthaven	8
Amsterdam Science Park	8
Watergraafsmeer	8
Admiralenbuurt	8
Westelijke Eilanden (Amsterdam)	7
Bos en Lommer	7
Staatsliedenbuurt (Amsterdam)	7
Spaarndammerbuurt	7
Museumkwartier (Amsterdam)	7
Oosterparkbuurt (Amsterdam)	7

If we take a look of K-mean cluster analysis, those most popular neighbourhoods belong to Cluster 3. The bright green dots on Amsterdam map are in cluster 3. We recommend our client to avoid those areas.

```
ams_merged.loc[ams_merged['Cluster Labels'] == 3]
```

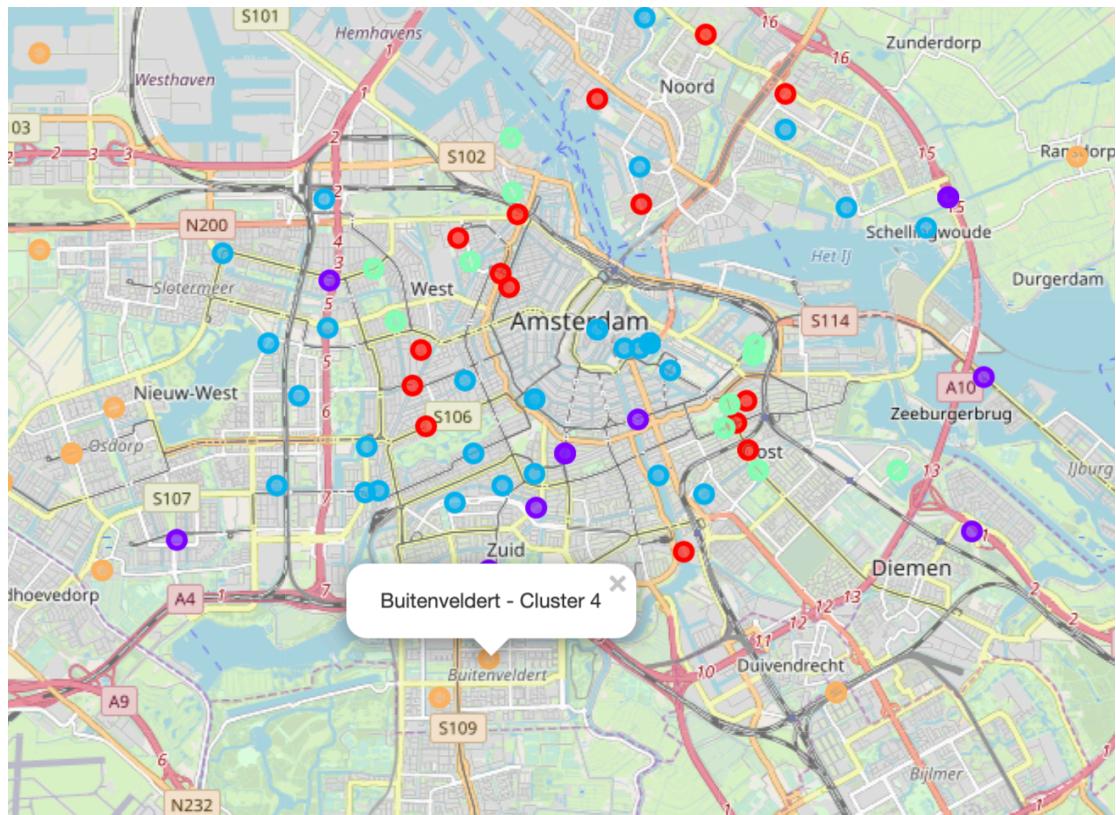
	Neighborhood	Bar	Cluster Labels	Latitude	Longitude
4	Amsterdam Science Park	0.080000	3	52.354320	4.958030
83	Spaarndammerbuurt	0.070000	3	52.388683	4.879929
84	Staatsliedenbuurt (Amsterdam)	0.070000	3	52.380050	4.871490
0	Admiralenbuurt	0.080000	3	52.372728	4.856362
9	Bos en Lommer	0.070000	3	52.379190	4.851740
62	Oosterparkbuurt (Amsterdam)	0.070000	3	52.359470	4.922980
60	Oostelijke Eilanden	0.070000	3	52.368570	4.928600
98	Westelijke Eilanden (Amsterdam)	0.070000	3	52.368570	4.928600
51	Museumkwartier (Amsterdam)	0.070000	3	52.362290	4.923910
50	Molenwijk (Amsterdam)	0.069767	3	52.421005	4.890154
18	Czaar Peterbuurt	0.070000	3	52.369716	4.928944
35	Houthaven	0.080000	3	52.395367	4.879504
78	Ruigoord	0.071429	3	52.409994	4.749248
94	Watergraafsmeer	0.080000	3	52.354280	4.929670

The bright green dots on Amsterdam map are in cluster 3



We want to avoid Neighbourhoods with the lowest number of Bars. Those orange dots on the map are cluster 4. They are far away from Amsterdam centre and close to highway. Cluster 1 is also bad location for open a bar.

	Count
Stadionbuurt	2
Bijlmermeer	1
Nieuw Sloten	1
Zuidas	1
Tuindorp Nieuwendam	1
Ruigoord	1
Nieuwendam	1
Burgwallen Nieuwe Zijde	1
Gaasperdam	1
Bullewijk	1



Conclusion

Recommended Neighborhoods for opening a small bar:

- Cluster 1
- Cluster 2

The best location for opening a small bar is in cluster 1 and 2, they are the purple dots and blue dots on map respectively. As we can see from the map, they are located close to Amsterdam city centre and close to tourist attractions. Neighbourhood such as Uilenburg and Plantage are nice areas that are in the center of city, but are not too competitive.

