



Mama Mia Pizza

BATTLE OF THE AMSTERDAM NEIGHBORHOODS
COURSERA CAPSTONE PROJECT FOR THE IBM DATA
SCIENCE CERTIFICATION

WILJO MEIJNHOUT

JANUARY 2020



Audience

- ▶ In Rome, Italy a famous franchise company called 'Mama Mia Pizza' want to spread their wings across the European capital cities, starting with Amsterdam in the Netherlands



Business Problem

- ▶ *"Start a Mama Mia Pizza Restaurant in a neighborhood in Amsterdam, crowded of Airbnb hosts but where the restaurant market is less saturated compared to other neighborhoods"*

Data Gathering



Airbnb host data via
InsideAirbnb.com/amsterdam



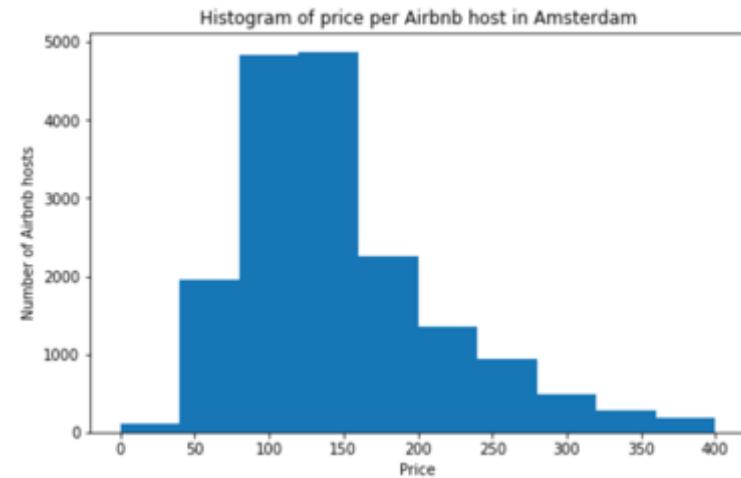
Geographical data from the
Amsterdam neighborhoods
via maps.amsterdam.nl



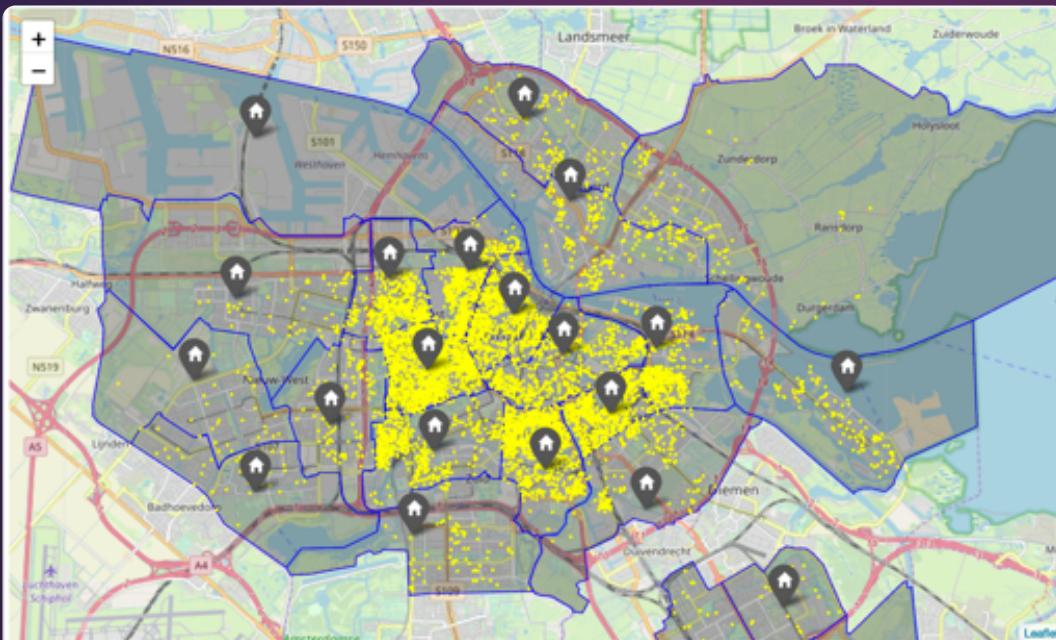
Nearby venue data from
FourSquare

Data Cleaning Airbnb

- ▶ Decrease Airbnb dataset from ~20.000 to ~7000 records
- ▶ Focus on price distribution for further cleaning
- ▶ Create new dataframe with only the Airbnb hosts between 100 – 150 Euro/night



Airbnb hosts
plotted on the
Amsterdam
neighborhoods



FourSquare Data Cleaning

Only focus on
restaurants in
Amsterdam

Create an array of all restaurants

```
In [28]: array = []
search = 'Restaurant'
for i in amsterdam_venues.venue_category :
    if search in i:
        array.append(i)

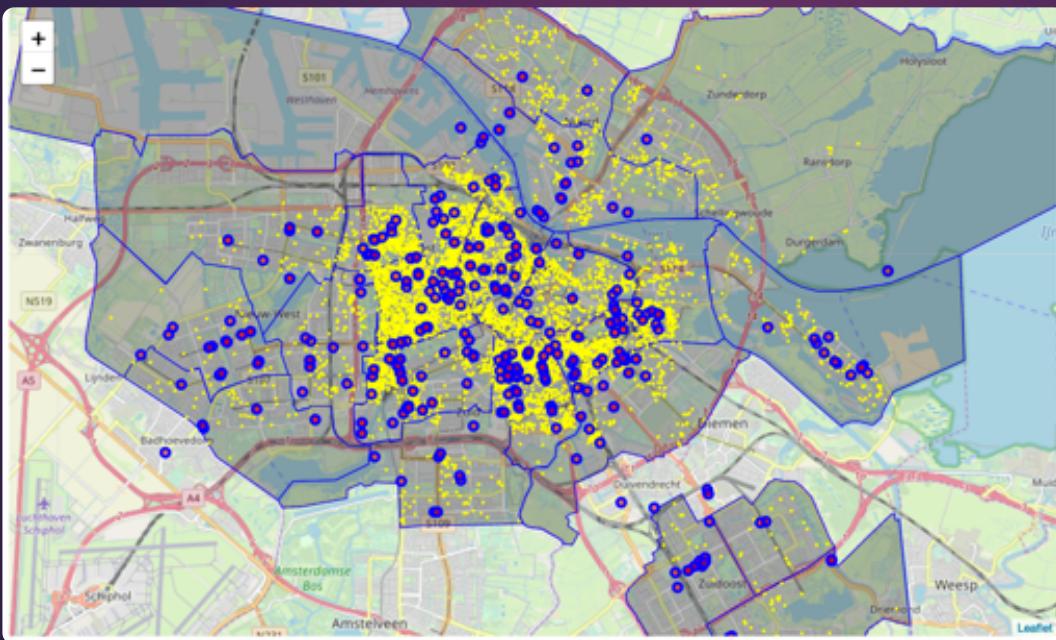
# and add Pizza place because this is also a restaurant
array.append('Pizza Place')
```

Create a new dataset of all restaurants in Amsterdam

```
In [29]: amsterdam_restaurants = amsterdam_venues.loc[amsterdam_venues['venue_category'].isin(array)]
amsterdam_restaurants.head()
```

	neighborhood	neighborhood_latitude	neighborhood_longitude	venue	venue_latitude	venue_longitude	venue_category
4	Westpoort	52.411465	4.807319	KFC	52.427470	4.820170	Fast Food Restaurant
8	Westpoort	52.411465	4.807319	McDonald's	52.427396	4.820760	Fast Food Restaurant
19	Bijlmer-Oost	52.319564	4.976832	De Smetkroes	52.322755	4.974752	South American Restaurant
22	Bijlmer-Oost	52.319564	4.976832	Pasta di Mamma	52.314779	4.955087	Italian Restaurant
24	Bijlmer-Oost	52.319564	4.976832	Margherita Tutta La Vita!	52.329054	4.955773	Pizza Place

Restaurants
plotted on
the map



Clustering restaurants

Cluster0 fits the best

```
In [33]: amsterdam_grouped_clustering = amsterdam_restaurant_grouped.drop('neighborhood', 1)

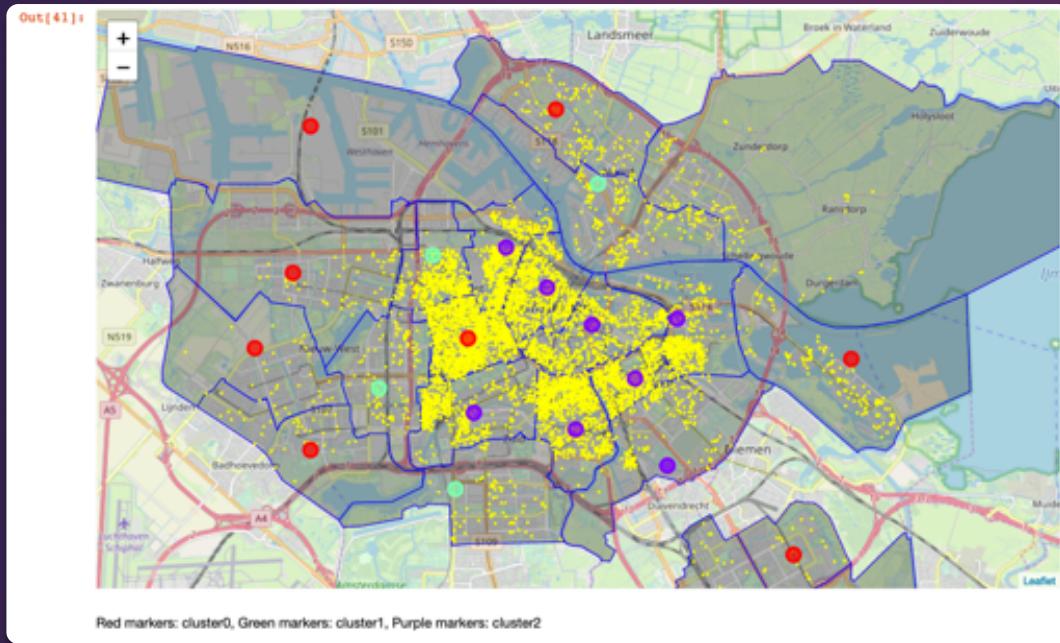
for n_cluster in range(2, 10):
    kmeans = KMeans(n_clusters=n_cluster).fit(amsterdam_grouped_clustering)
    label = kmeans.labels_
    sil_coeff = silhouette_score(amsterdam_grouped_clustering, label, metric='euclidean')
    print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))

For n_clusters=2, The Silhouette Coefficient is 0.22844391429425503
For n_clusters=3, The Silhouette Coefficient is 0.25396818256927134
For n_clusters=4, The Silhouette Coefficient is 0.24056222210026956
For n_clusters=5, The Silhouette Coefficient is 0.21517349726221882
For n_clusters=6, The Silhouette Coefficient is 0.21888426417376594
For n_clusters=7, The Silhouette Coefficient is 0.1956549358008158
For n_clusters=8, The Silhouette Coefficient is 0.20301791030757166
For n_clusters=9, The Silhouette Coefficient is 0.19429071778259674
```

```
In [35]: amsterdam_results = pd.DataFrame(kmeans.cluster_centers_)
amsterdam_results.columns = amsterdam_grouped_clustering.columns
amsterdam_results.index = ['cluster0','cluster1','cluster2']
amsterdam_results['Total Sum'] = amsterdam_results.sum(axis = 1)
amsterdam_results
```

	Satay Restaurant	Scandinavian Restaurant	Seafood Restaurant	South American Restaurant	Southern / Soul Food Restaurant	Spanish Restaurant	Sushi Restaurant	Tapas Restaurant	Thai Restaurant	Turkish Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Total Sum
0000e+01	0.20	0.200	0.200	-6.938894e-18	1.000000e-01	0.300	2.776558e-17	0.200	2.000	0.100	-6.938894e-18	13.70	
8894e-18	0.00	0.875	0.125	-6.938894e-18	-6.938894e-18	0.125	5.000000e-01	1.125	0.375	1.125	1.250000e-01	25.75	
1000e+00	0.25	0.750	0.250	2.500000e-01	0.000000e+00	0.000	2.500000e-01	1.250	1.500	0.500	0.000000e+00	25.75	

Plot
cluster on
the map



Focus on Cluster0 only

```
In [44]: # drop all neighborhoods from cluster0 and cluster2
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Bos en Lommer'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Oud-Noord'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Slotervaart'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Oud-Zuid'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Westerpark'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Centrum-West'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Centrum-Oost'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Watergraafsmeer'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Buitenveldert - Zuidas'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Zuid'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'Oud-Oost'].index, inplace=True)
airbnb_cluster0.drop(airbnb_cluster0.loc[airbnb_cluster0['neighbourhood'] == 'De Pijp - Rivierenbuurt'].index, inplace=True)
# And display it's shape
airbnb_cluster0.shape
```

```
Out[44]: (2245, 16)
```

```
In [47]: airbnb_grouped = airbnb_cluster0.groupby('neighbourhood').count().reset_index()
```

```
In [49]: # And display the airbnb data in relation to cluster0, sorted with the highest value on top
airbnb_grouped.sort_values(by=['id'], ascending=False)
```

```
Out[49]:
```

neighbourhood	id	name	host_id	host_name	neighbourhood_group	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
3 De Baanjes - Oud-West	1351	1350	1351	1344		0	1351	1351	1351	1351	1351
9 Oostelijk Havengebied - Indische Buurt	369	369	369	362		0	369	369	369	369	369
6 IJburg - Zeeburgerland	123	123	123	123		0	123	123	123	123	123
8 Noord-West	122	122	122	121		0	122	122	122	122	122
7 Noord-Oost	88	88	88	88		0	88	88	88	88	88
5 Geuzenveld - Slotervaart	59	59	59	59		0	59	59	59	59	59
2 De Aker - Nieuw Sloten	40	40	40	40		0	40	40	40	40	40
10 Osdorp	29	29	29	29		0	29	29	29	29	29
4 Gaasperdam - Driemond	28	28	28	27		0	28	28	28	28	28
1 Bijlmer-Oost	21	21	21	21		0	21	21	21	21	21
0 Bijlmer-Centrum	15	15	15	15		0	15	15	15	15	15

Conclusion

- ▶ The Amsterdam neighborhood 'De Baarsjes – Oud – West' fits the best to start a Mama Mia Pizza Restaurant
 - ▶ 'De Baarsjes – Oud – West' is in cluster0, therefor the less occupied restaurant market
 - ▶ 'De Baarsjes – Oud – West' has 1351 active Airbnb hosts

