

Capstone project – Battle of the neighborhoods

JAI Borst, 19 nov 2019

1. Introduction

Curaçao is a relatively small (444km²) island located in the southern Caribbean and although a sovereign state, part of the Kingdom of the Netherlands. In 2011, Curaçao had a population of about 150.000 people and steadily growing over recent years. In 2019 Curacao reported to have a population of over 164.000 people¹. The population growth is mainly induced by immigration. A large portion has Dutch roots but also Columbia and Haiti have a fair share in the population growth.

About 90 percent of Curaçao's population had the Dutch nationality in 2011. Varying shares of foreign-born persons have obtained the Dutch nationality, ranging from almost 20 percent of Haitian-born persons to about 76 percent of Surinamese-born persons. The share of the population that has never lived abroad decreased from almost 70 percent in 2001 to 56 percent in 2011.²

Curacao mainly has a strong tourism economy, roughly 20% of it's GDP and is heavily depended on it's foreign travels. On average Curacao has over 400.000 stayover arrivals and well over 3,5 million visitor night per year. The top markets are The Netherlands, United States, Venezuela, Germany, Canada, Columbia and Brazil.³

This unique combination of multicultural tourism and population on a sunny, tropical island, could make for a splendid location of an Italian restaurant. Question is though; Where on the island would be a good place to open up such an Italian restaurant? For both the Tourism Board Curacao as well as any entrepreneur looking for a great opportunity this is very relevant information.

In order to answer this question, one should also take into account that the average household income on Curacao fluctuates heavily per neighborhood. People below a certain averga income are less likely to go to an Italian restuarant. And since the infrastructure does not accomodate large volumes of traffic, location is most likely a key to the possible performance of the restaurant.

Also the neighborhoods with the most foreign-born population might be of severe importance as the Curacao-born population tends to eat more local Creol food instead of forein food such as Italian.

1 <https://www.macrotrends.net/countries/CUW/curacao/population>

2 http://www.caribbeanelections.com/eDocs/statistics/cw_stats/cw_population_housing_census_2011.pdf

3 <https://issuu.com/curacaotouristboard/docs/ctb-annual-2017-digital>

So in order to find the best possible spot for an Italian restaurant we'll have to take the location into account with respect to the average income and clusters of foreign-born population. Lastly, we'll have to look at competition. Whether or not an Italian restaurant is already located there.

2. Data

In order to find the best location for an Italian restaurant on the island of Curacao, we'll need a couple of resources. Firstly we'll use data of the Curacao Census 2011⁴, which is an extensive research at the population indexes, such as household population, fertility rates, poverty rates, marital status, migrant population and much more made by the Central Bureau of Statistics Curacao. This large dataset containing 291 rows and 34 columns will be loaded as a CSV file and referred to as 'Census'.

Specifically the data on average household income (column 'meanHHinc') per neighborhood is used in this report as well as the percentage of foreign-born population (column 'forborn') in a certain area versus the Curacao-born (column 'curborn') population. To find the optimal location for the Italian restaurant we have to maximize both the average household income levels as well as the percentage foreign-born population in each neighborhood.

For displaying that information on a map, we need the appropriate neighborhoods in a Json file. As this was not readily available, this report makes use of a parsed set of Polygons depicting the neighborhood sizes, derived from OpenStreetMap, and loaded as a GeoJson file⁵. However OpenStreetMap only recognizes 65 different neighborhoods, whereas the Census files knows 291 different subneighborhoods.

In order to minimize dataloss and missing values, part of the original Census file was adjusted, i.e. the neighborhood "Brievengat Zuid" (Brievengat South) was grouped together with "Brievengat". Also different methods of writing the same neighborhoods were adjusted, for example "Aciento" was spelled as "Asiento". Lastly, the data in Census was formatted, in this case capitalized, to make a match with the GeoJson file.

Finally, the Foursquare API⁶ is used to pull location information of the chosen neighborhood, in terms of other available restaurants to find out whether there is another competitor in the form of an Italian restaurant. Also the density of nearby restaurants is a factor to take into account.

4 <https://cbcuracao.maps.arcgis.com/apps/webappviewer/index.html?id=6ed83880519f40ddb5d98d8b73671bcd>

5 https://github.com/Jaikej/Data-Science-Assignments/blob/master/Curacao_AL8.GeoJson

6 <https://foursquare.com/>

3. Methodology

This report makes use of the Curacao Census 2011 dataset in order to find the best location for an Italian restaurant. Firstly we have converted the dataset to be used in conjunction with the GeoJson file by adjusting the neighborhood names to capitalized strings. Below a quick overview of the dataset.

FID	NAME	HH	pop	hhsz	sexratio	p0_14	p15_64	p65u	curborn
1	Paradijs bij hel	23	59	2.6	0	0.000	0.000	0.000	0.000
2	Abrahamsz	220	573	2.5	85	0.162	0.651	0.187	0.713
3	Parasasa	78	171	2.2	82	0.164	0.655	0.181	0.520
4	Maria maai	245	676	2.5	78	0.195	0.612	0.192	0.768
5	Parera	22	56	2.5	0	0.000	0.000	0.000	0.000
...
61	Zwaan	23	34	1.5	0	0.000	0.000	0.000	0.000
62	Chango bij groot kwartier	45	109	2.4	95	0.165	0.697	0.138	0.661
63	Schelpwijk	382	966	2.5	78	0.151	0.678	0.171	0.740
64	Chango bij korporaal	27	55	2.0	0	0.000	0.000	0.000	0.000
65	Scherpenheuvel	176	488	2.7	86	0.193	0.684	0.123	0.762

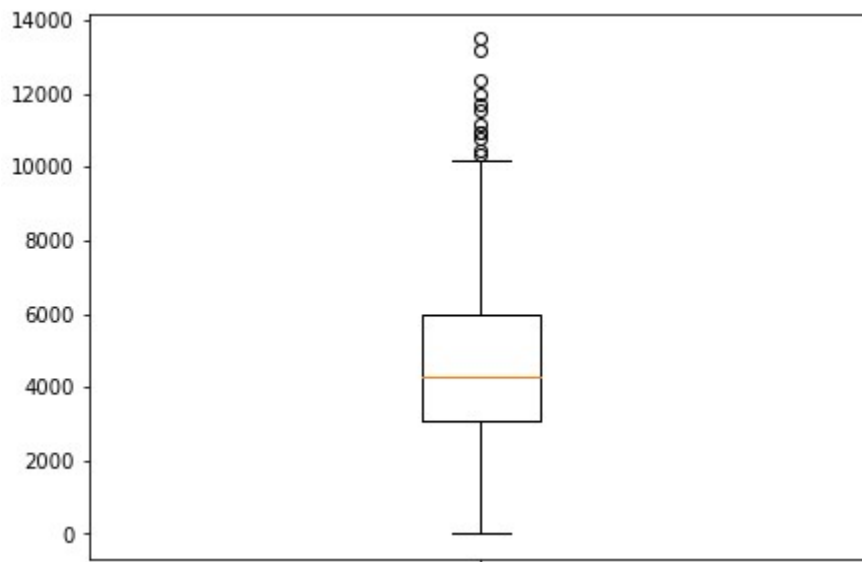
In an exploratory analysis of both the Census dataset as well as the neighborhood map in the GeoJson file is found that not all neighborhoods are in both lists. Therefore we use only the comparable part of the dataset.

To see whether is an appropriate sample, a describing overview is generated with Pandas Dataframes:

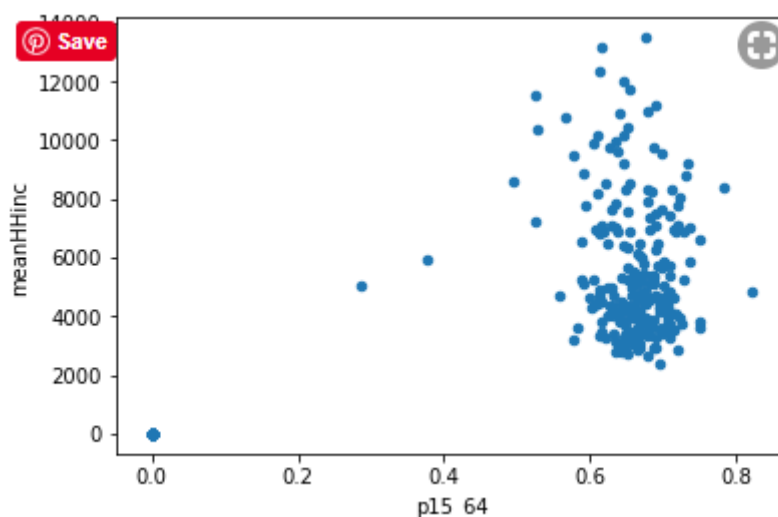
	FID	HH	pop	hhsz	sexratio	p0_14	p15_64	p65u	curborn	forborn	...	airpolut
count	291.000000	291.000000	291.000000	291.000000	291.000000	291.000000	291.000000	291.000000	291.000000	291.000000	...	291.000000
mean	146.000000	188.783505	516.731959	2.582818	67.130584	0.153546	0.527997	0.119137	0.590570	0.210117	...	16.787973
std	84.148678	241.905377	681.345786	0.475228	34.668630	0.084837	0.268170	0.087312	0.310049	0.141562	...	19.718940
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	73.500000	44.000000	119.000000	2.500000	68.500000	0.136500	0.592500	0.060500	0.558000	0.121500	...	5.500000
50%	146.000000	98.000000	270.000000	2.600000	81.000000	0.180000	0.652000	0.126000	0.714000	0.215000	...	9.700000
75%	218.500000	233.000000	602.000000	2.800000	88.000000	0.207500	0.681500	0.171000	0.805000	0.313000	...	18.250000
max	291.000000	1455.000000	3892.000000	5.000000	113.000000	0.329000	0.822000	0.651000	0.953000	0.550000	...	100.000000

From this overview it has become clear that the average household size is around 2,5 persons, of which 53% are between the age of 15 and 64, which is relevant for our question. On average around 15% is younger than 15 years of age, and 12% older than 64 years of age. Of this population 21% on average is born in another country than Curacao. So we are looking for neighborhoods where foreign born is maximized as well as household income.

In order to look at the average household income, we can generate a boxplot from column 'meanHHinc'. The average household income is rounded to 4408 ANG, the local currency, for approximately 2,5 persons in the household.



However for the purposes of this report, the persons between the age of 15 and 64 are more interesting as they are more likely to spend money on our restaurant than for example persons under 15 years of age. To see how this is divided within average household income, the scatterplot below is generated.



When selecting the neighborhoods with the 5 highest average household income in descending order, we'll get this overview stating 'Mahaai' on top with an average income of 12.358 ANG, followed by 'St. Willibrordus', 'Groot Piscadera', 'Kwarchi' and 'Mon Repos'.

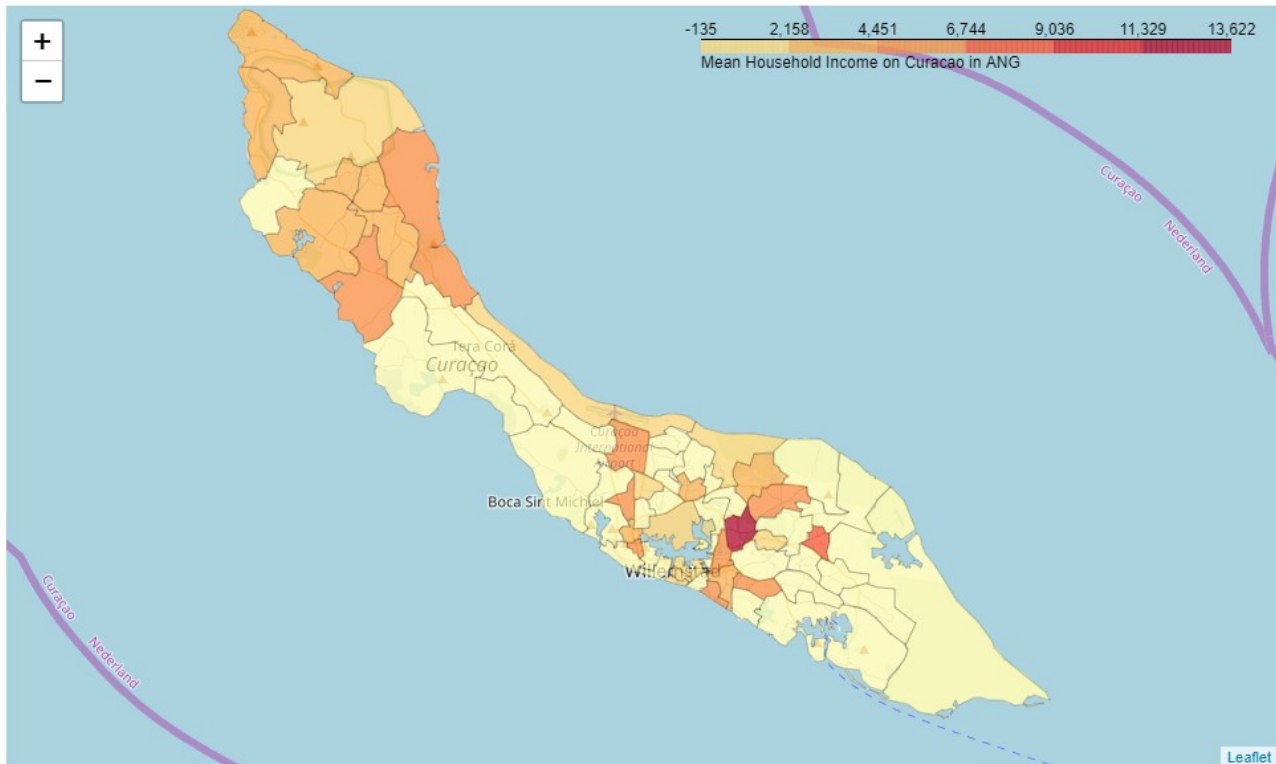
	NAME	HH	pop	hhsiz	sexratio	p0_14	p15_64	p65u	curborn	forborn	...
237	Mahaai	152	389	2.6	80	0.211	0.612	0.177	0.513	0.487	...
109	St. willibrordus	240	588	2.5	85	0.192	0.646	0.162	0.630	0.370	...
134	Groot piscadera	240	749	2.7	94	0.124	0.784	0.092	0.721	0.279	...
216	Kwarchi	234	670	2.9	95	0.197	0.681	0.122	0.740	0.260	...
258	Mon repos	93	270	2.6	80	0.144	0.693	0.163	0.770	0.230	...

The same goes for the percentage foreign-born population per neighborhood. To dermine the top largest percentage, we have selected column ‘forborn’ with descending order. The neighborhoods ‘Mahaai’, ‘Salina’, ‘St Willibrordus’, ‘Zeelandia’ and ‘Wishi’ have the largest number of foreign born population.

	NAME	HH	pop	hhsiz	sexratio	p0_14	p15_64	p65u	curborn	forborn
237	Mahaai	152	389	2.6	80	0.211	0.612	0.177	0.513	0.487
251	Saliña	151	361	2.4	88	0.186	0.632	0.183	0.627	0.373
109	St. willibrordus	240	588	2.5	85	0.192	0.646	0.162	0.630	0.370
174	Zeelandia	88	135	1.5	39	0.067	0.378	0.556	0.630	0.370
171	Wishi	328	841	2.6	75	0.205	0.669	0.126	0.695	0.305

4. Results

In order to pick the best location for an Italian restaurant in Curacao, it is important to factor two variables into account, namely average household income and foreign born population in the different neighborhoods. When looking at the top 5 results in terms of average household income (column 'meanHHinc'), the neighborhood 'Mahaai' scored top result with an average household income of 12.358 ANG. On the map 'Mahaai' is visible in dark red color:



Evidently when looking at the neighborhood with the highest percentage foreign born population also the neighborhood 'Mahaai' scored top results with almost 49% being foreign born versus 51% being native Curacao born population.

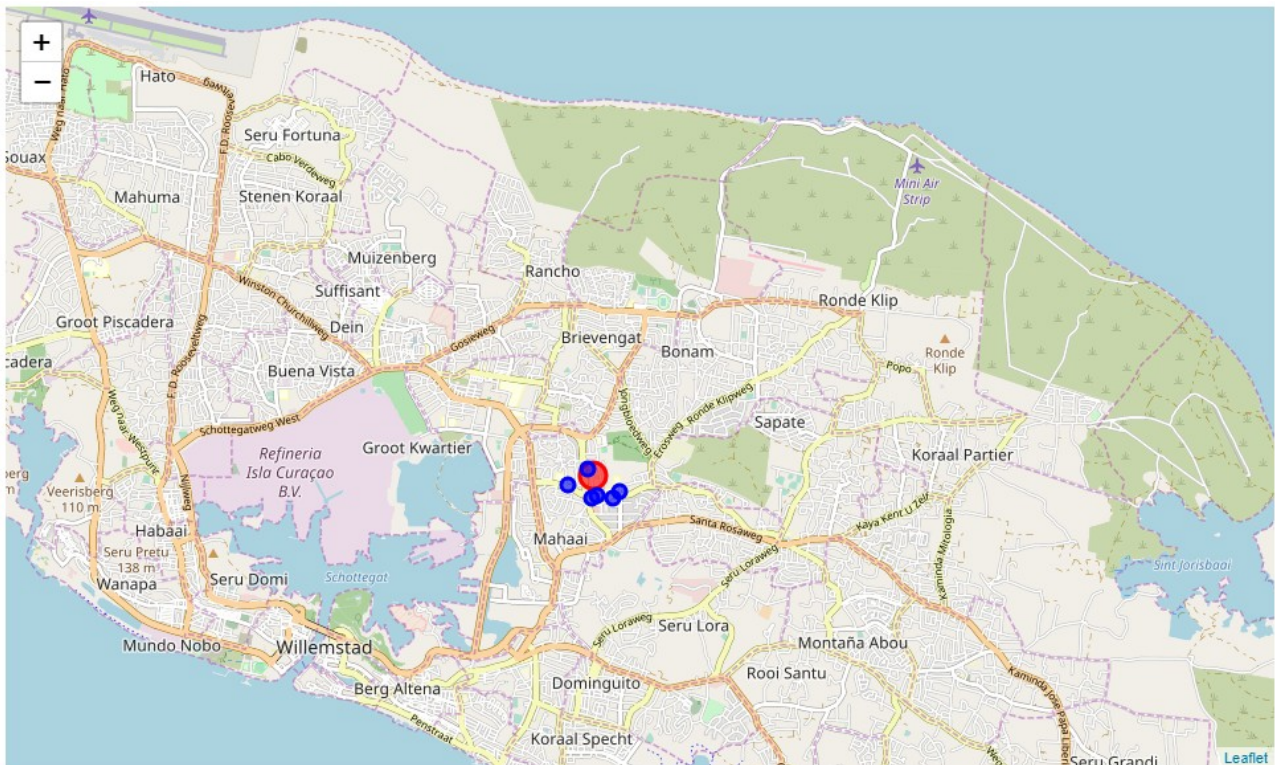
When connected to the Foursquare API to check whether there are other conflicting venues nearby, the result showed only 6 venues in total nearby of which 5 restaurants.

Restaurant Brakkeput Mei Mei
Hai Ba Restaurant
Pincho Garden Restaurant And Steak Boutique
Supreme Restaurant Equipment & supplies
Al-Diwan Restaurant
Ling Hong

Given the fact that none of these restaurants are serving Italian food, this would be an ideal location.

	name	categories	lat	lng	labeledLatLngs	distance	cc	city
0	Restaurant Brakkeput Mei Mei	Brazilian Restaurant	12.129700	-68.900010	[{"label": "display", "lat": 12.1297, "lng": -68.900010}]	360	CW	Jan Sofat / Spaanse water, Willemstad
1	Hai Ba Restaurant	Chinese Restaurant	12.131669	-68.897461	[{"label": "display", "lat": 12.131669, "lng": -68.897461}]	115	CW	NaN
2	Pincho Garden Restaurant And Steak Boutique	Portuguese Restaurant	12.127910	-68.896980	[{"label": "display", "lat": 12.12791, "lng": -68.896980}]	320	CW	Willemstad
3	Supreme Restaurant Equipment & supplies	Video Store	12.128302	-68.896328	[{"label": "display", "lat": 12.12830190538713, "lng": -68.896328}]	283	CW	Willemstad
4	Al-Diwan Restaurant	Middle Eastern Restaurant	12.128024	-68.894249	[{"label": "display", "lat": 12.12802410125732, "lng": -68.894249}]	421	CW	NaN
5	Ling Hong	Chinese Restaurant	12.128720	-68.893423	[{"label": "display", "lat": 12.12871950042185, "lng": -68.893423}]	442	CW	NaN

These venues are closely clustered together on the map, which might even indicate a shopping mall or at least cluster of venues, making the start of an Italian restaurant very viable.



4. Discussion

Although the dataset has some missing values, an estimate can be given of an ideal location to start a new Italian restaurant. However this might be suboptimal when the information of the complete dataset would be available. This would be achievable when all the neighborhoods will have comparable naming. In order to do so, extensive local knowledge is needed to put certain areas in the correct borough.

Nevertheless, the outcome of an ideal startup location in the vicinity of the 'Mahaai' region is evident based on the top results in both the average household income, thus the ability to spend a certain amount of money on dinner in a restaurant, and the top result in most percentage of the population being foreign born, and therefore most likely to be interested in a less local cuisine.

5. Conclusion

Curacao is a relatively small island, but knows nonetheless quite some differences within its neighborhoods in terms of best places to start a new Italian restaurant. When looking at the different variables depending on finding a good location, we have incorporated information from the Census dataset, which is concerned with information about demographic variables on the population of Curacao.

With this dataset is calculated which neighborhood had the highest average household income to be most likely to spend it in a restaurant, being the area of 'Mahaai'. Also is checked which neighborhood had the highest percentage foreign born population, which also resulted in the area of 'Mahaai'.

When looking at the information pulled from Foursquare, can be concluded that no serious competitor is in the direct vicinity of 'Mahaai' and thus a great place to start this new venue.