Moving Abroad - Choosing a new city to live

IBM Data Science Professional Certificate - Applied Data Science Capstone

Rafael Henrique da Silva Santos

rafacto@gmail.com

contents

- Introduction;
- Data;
- Methodology;
- Results;
- Discussion;
- Conclusion.

introduction

- Moving abroad and choosing a new city to live is not easy;
- Sister Cities: are cities that establish a bond of cooperation on many factors;
- Different factors like cultural life, attractions, language, climate, jog market, and life cost can impact the life experience in the new city;
- Goal: this study aims to be a tool for the process of finding a new place (city and neighborhood) to live, taking into consideration all those factors.

data

- Sister Cities: Wikipedia city page;
- Living parameters: Numbeo;
- Top venues: Foursquare API;
- City's neighborhoods, their coordinates, boundaries, and safety index: municipal webpage.

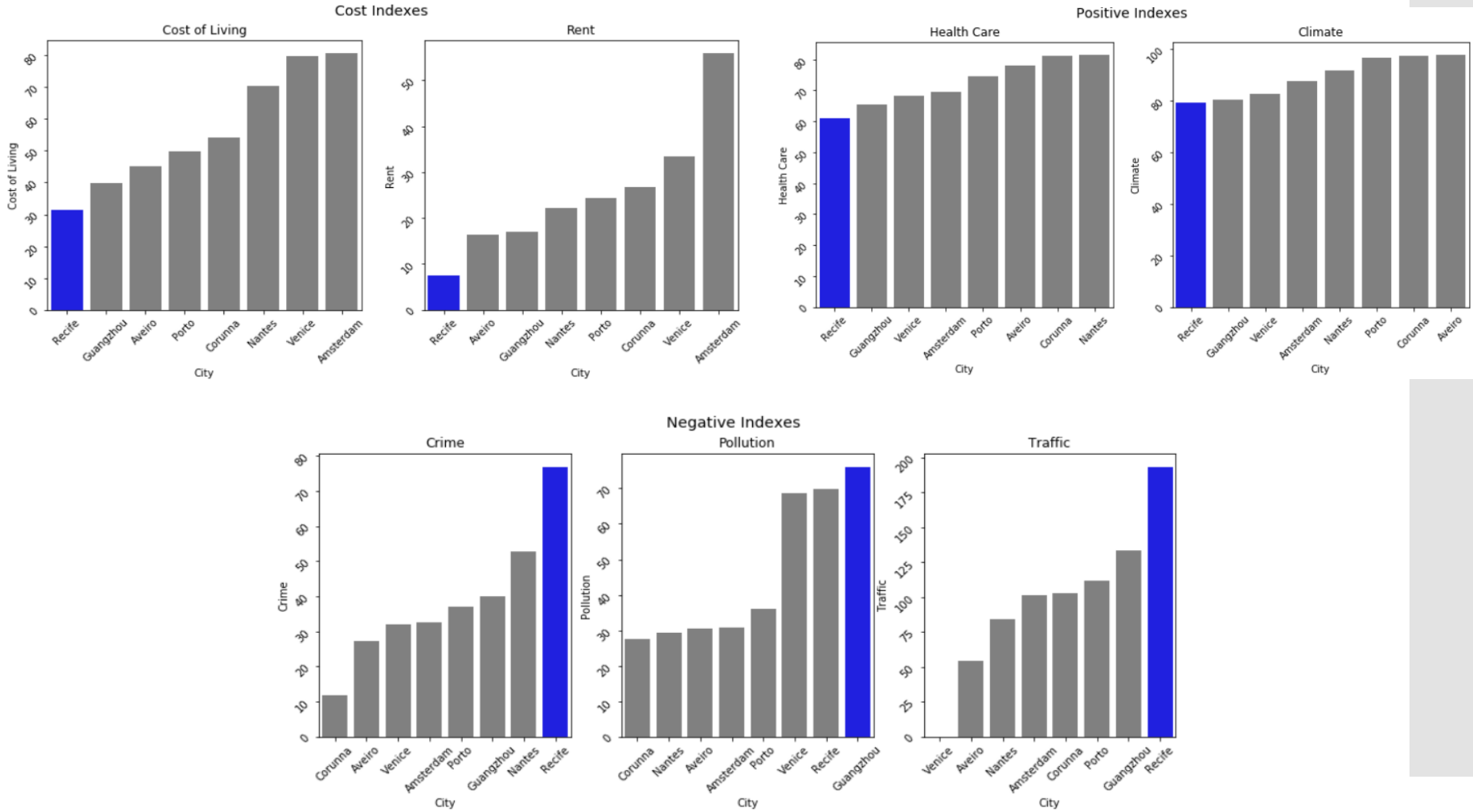
methodology

- 1) Analyze, compare, and segment, through exploratory analysis and clustering, the hometown's sister cities according to living parameters;
- 2) Segment the neighborhoods of the chosen city according to their top venues, aiming the grouping and further characterization of them;
- 3) Safety information of each neighborhood is joined to help in the final decision of each neighborhood to move in.

	City	Cost of Living	Rent	Property Price	Crime	Health Care	Pollution	Traffic	Climate
0	Recife	31.31	7.51	20.56	76.81	60.95	69.85	193.11	79.28
1	Amsterdam	80.74	55.94	10.70	32.44	69.45	30.90	100.88	87.45
2	Aveiro	45.19	16.40	8.27	27.16	78.12	30.42	54.22	97.64
3	Guangzhou	39.78	17.11	34.11	40.08	65.29	76.05	133.40	80.30
4	Corunna	54.33	26.77	11.83	11.73	81.27	27.66	102.95	97.21
5	Nantes	70.40	22.09	8.95	52.75	81.52	29.26	84.27	91.59
6	Porto	49.82	24.35	12.62	36.87	74.62	36.02	111.47	96.61
7	Venice	79.90	33.49	13.50	31.99	68.35	68.78	0.00	82.39

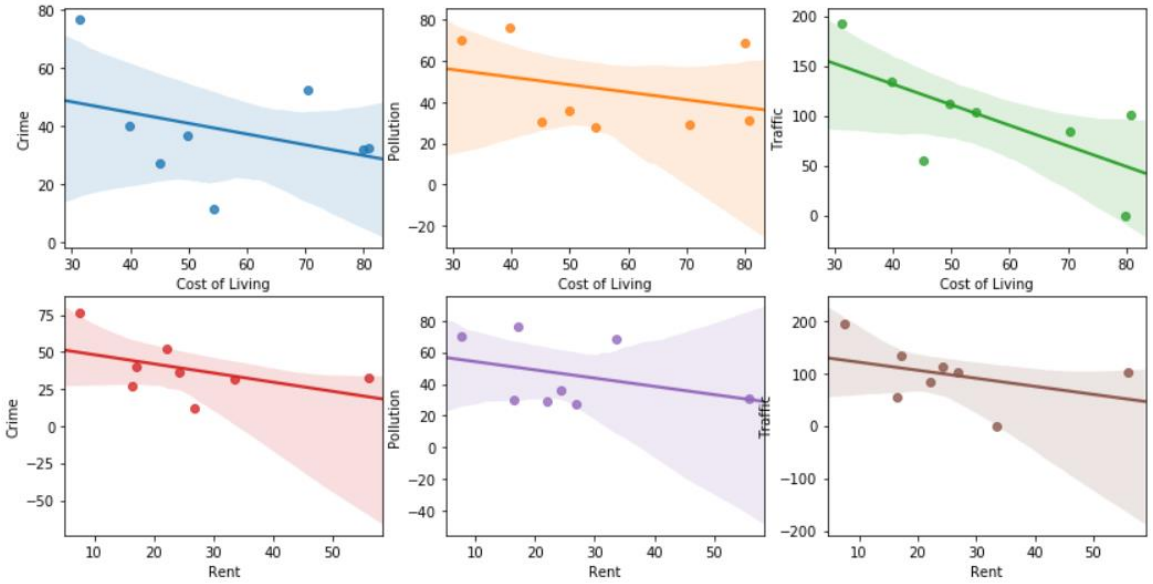
methodology

- When cheap becomes expensive: Recife is the cheapest city to live in. However, this comes with a price.



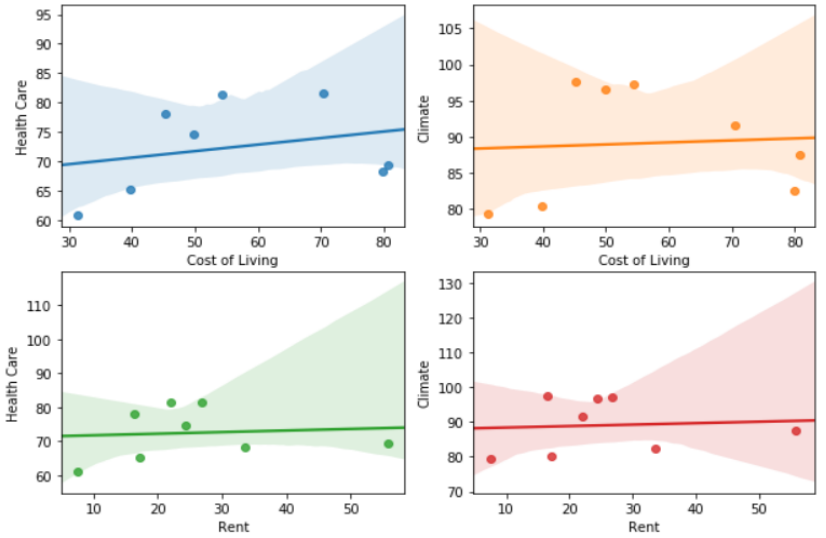
methodology

- ~~the more expensive the city the better the quality of life indicators~~



	Cost of Living	Rent	Health Care	Climate
Cost of Living	1.000000	0.833731	0.271370	0.066953
Rent	0.833731	1.000000	0.089006	0.078445
Health Care	0.271370	0.089006	1.000000	0.885980
Climate	0.066953	0.078445	0.885980	1.000000

	Cost of Living	Rent	Crime	Pollution	Traffic
Cost of Living	1.000000	0.833731	-0.356941	-0.319578	-0.682947
Rent	0.833731	1.000000	-0.466883	-0.355645	-0.395162
Crime	-0.356941	-0.466883	1.000000	0.469532	0.587757
Pollution	-0.319578	-0.355645	0.469532	1.000000	0.207473
Traffic	-0.682947	-0.395162	0.587757	0.207473	1.000000



methodology

- Since the relationships among the variables are not trivial, to help in the decision of choosing a new city to move in, a clustering algorithm is applied to segment the cities according to the similarity of those indexes:
  - Normalization
  - K = 3
- Once the city is chosen, its neighborhoods are clustered taking into consideration their venues: top 100venues within a radius of 1km;
- Finally, a choropleth map of Amsterdam neighborhoods is generated taking into consideration the safety index.

results

- Cities segmentation:

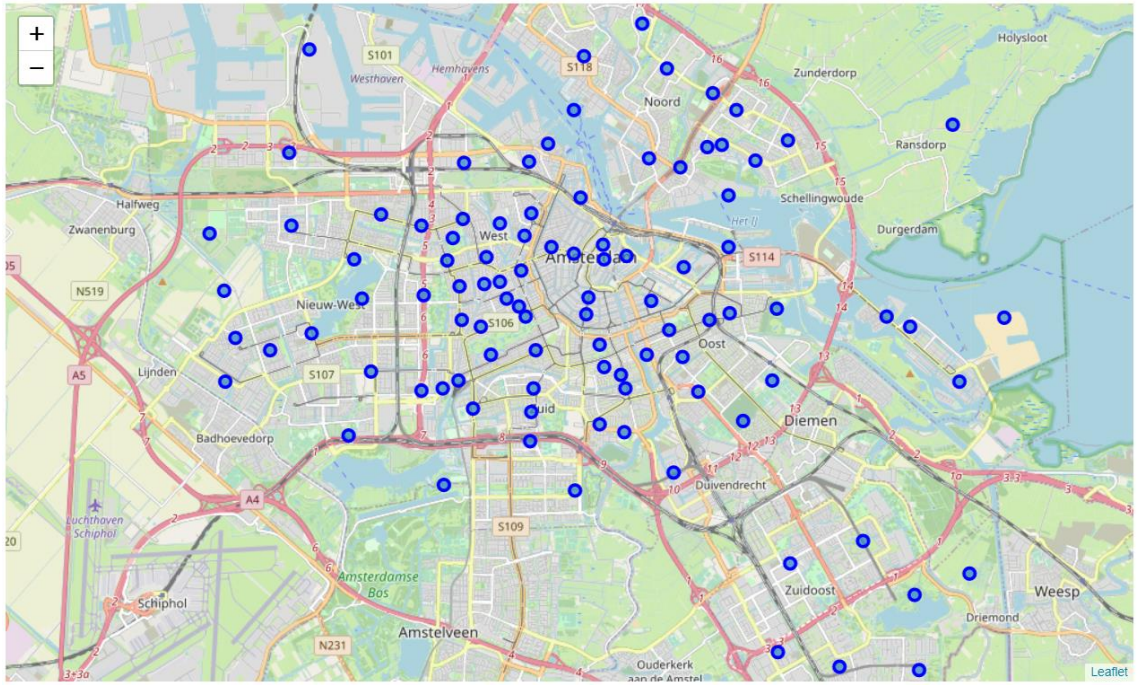
	City	Cost of Living	Rent	Property Price	Crime	Health Care	Pollution	Traffic	Climate	group
0	Recife	31.31	7.51	20.56	76.81	60.95	69.85	193.11	79.28	1
1	Amsterdam	80.74	55.94	10.70	32.44	69.45	30.90	100.88	87.45	2
2	Aveiro	45.19	16.40	8.27	27.16	78.12	30.42	54.22	97.64	2
3	Guangzhou	39.78	17.11	34.11	40.08	65.29	76.05	133.40	80.30	1
4	Corunna	54.33	26.77	11.83	11.73	81.27	27.66	102.95	97.21	2
5	Nantes	70.40	22.09	8.95	52.75	81.52	29.26	84.27	91.59	2
6	Porto	49.82	24.35	12.62	36.87	74.62	36.02	111.47	96.61	2
7	Venice	79.90	33.49	13.50	31.99	68.35	68.78	0.00	82.39	0

group	Cost of Living	Rent	Property Price	Crime	Health Care	Pollution	Traffic	Climate
0	79.900	33.49	13.500	31.990	68.350	68.780	0.000	82.39
1	35.545	12.31	27.335	58.445	63.120	72.950	163.255	79.79
2	60.096	29.11	10.474	32.190	76.996	30.852	90.758	94.10

- Group 0 (Venice): peculiar city with particular features;
- Group 1 (Recife and Guangzhou): cheap but mediocre quality of life;
- **Group 2 (Amsterdam, Aveiro, Corunna, Nates, and Porto): moderate to high life cost but good life quality**.

results

- Aveiro x **Amsterdam**
- Which neighborhood to live in?
- Amsterdam 99 neighborhoods:



results

- Total number of the obtained venues per neighborhood:

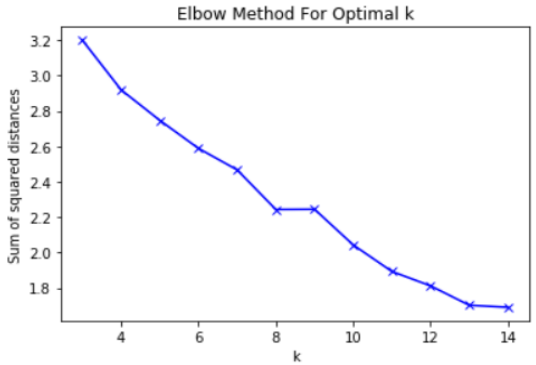
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Amstel III/Bullewijk	46	46	46	46	46	46
Apollobuurt	95	95	95	95	95	95
Banne Buiksloot	19	19	19	19	19	19
Bedrijventerrein Sloterdijk	12	12	12	12	12	12
Betondorp	34	34	34	34	34	34
...	...	...	...	...	...	...
Westlandgracht	78	78	78	78	78	78
Willemspark	100	100	100	100	100	100
Zeeburgereiland/Nieuwe Diep	29	29	29	29	29	29
Zuid Pijp	100	100	100	100	100	100
Zuidas	96	96	96	96	96	96

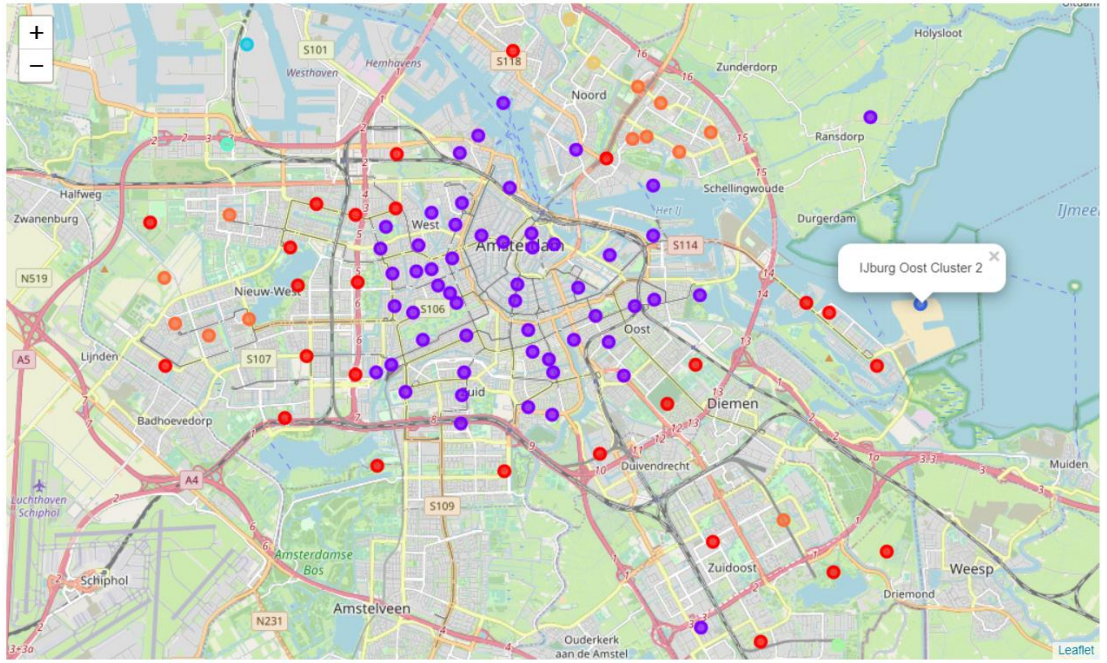
- 319 venue categories;
- 10 most common venue categories:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amstel III/Bullewijk	Hotel	Furniture / Home Store	Restaurant	Coffee Shop	Metro Station	Café	Fast Food Restaurant	Scandinavian Restaurant	Sandwich Place	Science Museum
1	Apollobuurt	Hotel	Restaurant	Bakery	Italian Restaurant	Art Gallery	Plaza	Coffee Shop	Bistro	Bar	Steakhouse
2	Banne Buiksloot	Bus Stop	Soccer Field	Hockey Field	Gym / Fitness Center	Sports Club	Supermarket	Bakery	Farm	Restaurant	Grocery Store
3	Bedrijventerrein Sloterdijk	Motorcycle Shop	Furniture / Home Store	Hardware Store	Gas Station	Auto Workshop	Restaurant	Racetrack	Rental Car Location	Breakfast Spot	Sporting Goods Shop
4	Betondorp	Soccer Field	Playground	Stadium	Café	Tennis Court	Tram Station	Nightclub	Bus Stop	Scandinavian Restaurant	Pizza Place

results

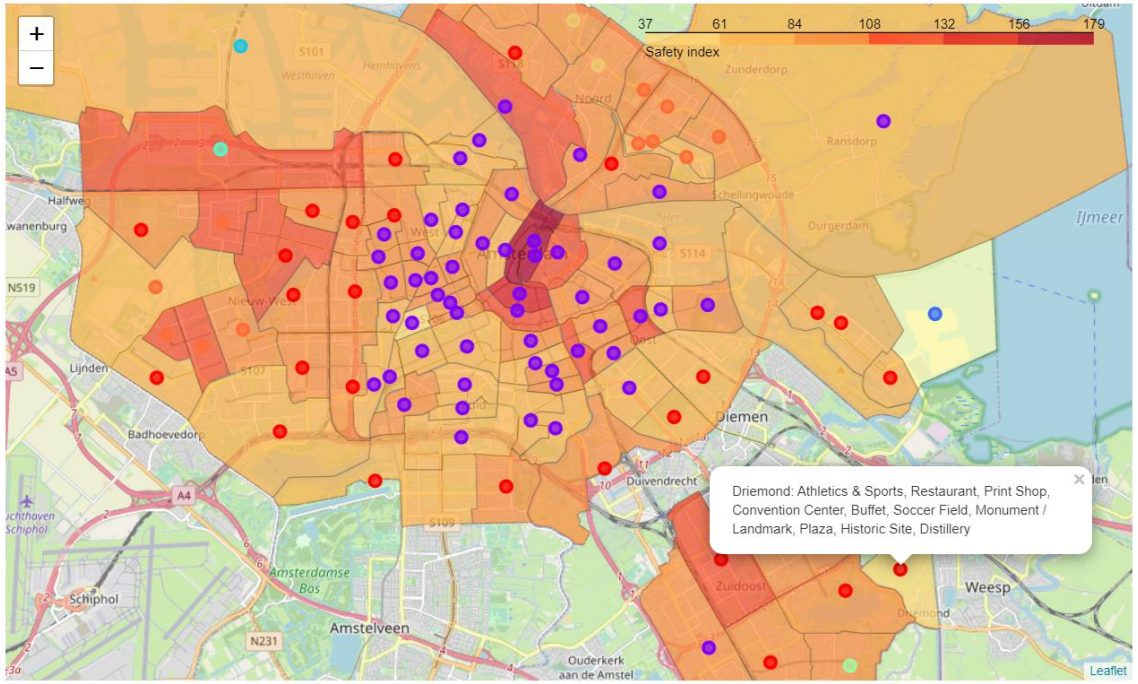
- Neighborhoods segmentation according to the similarity of venue categories





results

- Information about the safety of each neighborhood was obtained to create a choropleth map
- The choropleth map was then merged with the map of the clustered Amsterdam neighborhoods. The top 10 venues were added to the label of each point to concise all the needed information in one map, facilitating the decision about which neighborhood to live.



discussion

- Characterization of the neighborhoods clusters:
  - **Cluster 1 (26) - Area with many restaurants, cafes, and supermarkets. Some hotels but not many, indicating relative proximity of touristic areas. Some bars and clubs, indicating certain nightlife**;
  - Cluster 2 (54) - Many restaurants, cafes, and supermarkets. Many hotels and museums, indicating an area close to touristic attractions. Many bars and clubs suggesting a vivid nightlife. It seems to be the cultural center of the city;
  - Cluster 3 (1) - Beach area;
  - Cluster 4 (1) - Isolated neighborhood probably rich/expensive since it presents a heliport and a harbor;
  - Cluster 5 (1) - Isolated regular area with venues specialized in vehicles;
  - Cluster 6 (1) - Isolated area with nature-related attraction;
  - Cluster 7 (2) - Area with venues for the practicing of different sports;
  - Cluster 8 (13) - A regular residential area far from the center (the main spots are bus stops and tram stations), but with all basic stores (supermarket, bakery, restaurants, etc.). It also presents options for recreation and sports.

conclusion

- The present study applied data analysis and clustering to help in the decision processes of finding a new city and neighborhood to move in;
- First, the target city was chosen among the sister cities of the current city (Recife, Brazil): comparison and segmentation given living parameters;
- Second, neighborhood selection: neighborhoods were compared and segmented according to the most popular venues;
- Third, a choropleth map of Amsterdam neighborhoods was generated taking into consideration the safety index ;
- Last, a final map that shows the neighborhoods segmented by venues similarity, their top ten venue categories, and their safety index is generated.