Venue composition of top-rated bars in Copenhagen, Denmark.

Manuel Ernesto Arreola Sevilla

05 of June 2020

Introduction/ Business Problem

- If you have been working in the service industry, you have noticed that the success of a venue isn't entirely on what you can offer but in great part, the location. The venues around and how easy is to access would be a key factor to consider.
- The aim of this project is to define the venue structure around the best rated beer bars in Copenhagen, Denmark. A cluster analysis of the venue composition of all the available bars would be done to help us understand if it’s correlates with their rating. The insight might help to better decide in the future, a suitable location to open a bar in Copenhagen.

Data

- Acquired exclusively using the Foursquare API, location data and ratings.
- A search query with different key words was done but ‘bar’ was the one that brought more venue results, and so it was decided to be use.
- A radius of 10,000 meters around the main coordinate of Copenhagen, Denmark was traced.
- The *venue composition* of the obtained bars was created delimitating a radius of 200 meters around each location.

Methodology

- **K-means clustering with Machine Learning**

- “K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes”[1].
  - **Elbow method**
- “The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters”[3].

Results

- From the 28 venues in Copenhagen rated and categorized in the Foursquare API as “bar”, exploring a radius of 100 meters around, it was obtained a total of 253 venues including the main “bar”. In total, 87 unique venue categories. After the exploratory query of the venue composition, 1 of the main venues were removed since the data from the venues around was not enough to perform a cluster analysis.

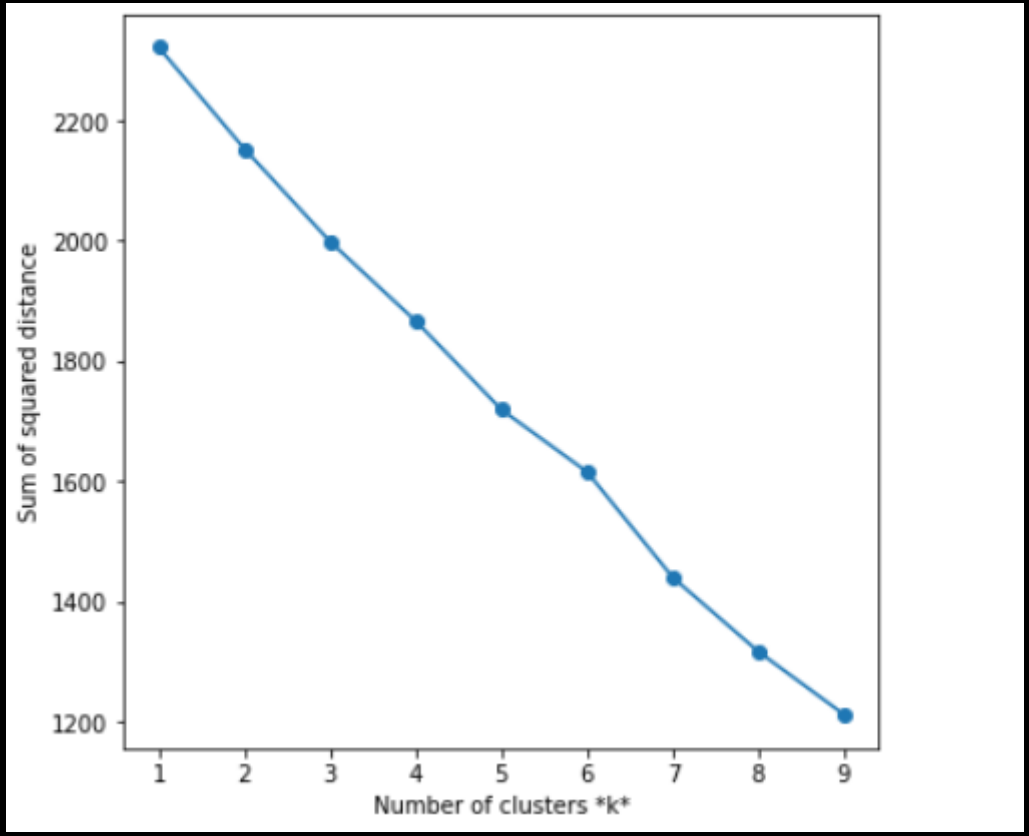
Venue composition of top-rated bars

name	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
Mikkeller Bar Viktoriagade	Scandinavian Restaurant	Hostel	Convenience Store	Szechuan Restaurant	Thai Restaurant	Middle Eastern Restaurant	Burger Joint	Beer Bar	Grocery Store	Dumpling Restaurant
Ved Stranden 10 - Vinhandel & Bar	Scandinavian Restaurant	Wine Bar	Lounge	Coffee Shop	Steakhouse	Furniture / Home Store	Dive Bar	Dumpling Restaurant	Farmers Market	Fish Market
Garden Restaurant & Bar	Tapas Restaurant	Art Gallery	Indie Movie Theater	Bistro	Breakfast Spot	Music Store	Gluten-free Restaurant	Farmers Market	Fish Market	Food & Drink Shop
Democratic Coffee Bar	Café	Library	Coffee Shop	Women's Store	Farmers Market	Fish Market	Food & Drink Shop	French Restaurant	Fried Chicken Joint	Furniture / Home Store
Lord Nelson Bar	Women's Store	Scandinavian Restaurant	Art Gallery	Australian Restaurant	Coffee Shop	Dessert Shop	Italian Restaurant	Jazz Club	Pedestrian Plaza	Yoga Studio
Bo-Bi Bar	Dive Bar	Shoe Store	Men's Store	Clothing Store	Candy Store	Tea Room	Furniture / Home Store	Dumpling Restaurant	Farmers Market	Fish Market
K Bar	Scandinavian Restaurant	Wine Bar	Plaza	Lounge	Coffee Shop	Restaurant	Fried Chicken Joint	Dive Bar	Dumpling Restaurant	Farmers Market
Berlin Bar	Department Store	Indian Restaurant	Record Shop	Men's Store	Beer Garden	Fish Market	Women's Store	Gay Bar	Farmers Market	Food & Drink Shop
Charlies Bar	Clothing Store	Juice Bar	Men's Store	Scandinavian Restaurant	Shoe Store	Coffee Shop	Candy Store	Tea Room	Miscellaneous Shop	Vegetarian / Vegan Restaurant

Table 1. 10 most common venues at the top-rated bars in Copenhagen.

Elbow method

Fig 1. The elbow method was used with k values of 1 to 9.



Clustering analysis of venue composition corresponding to ratings

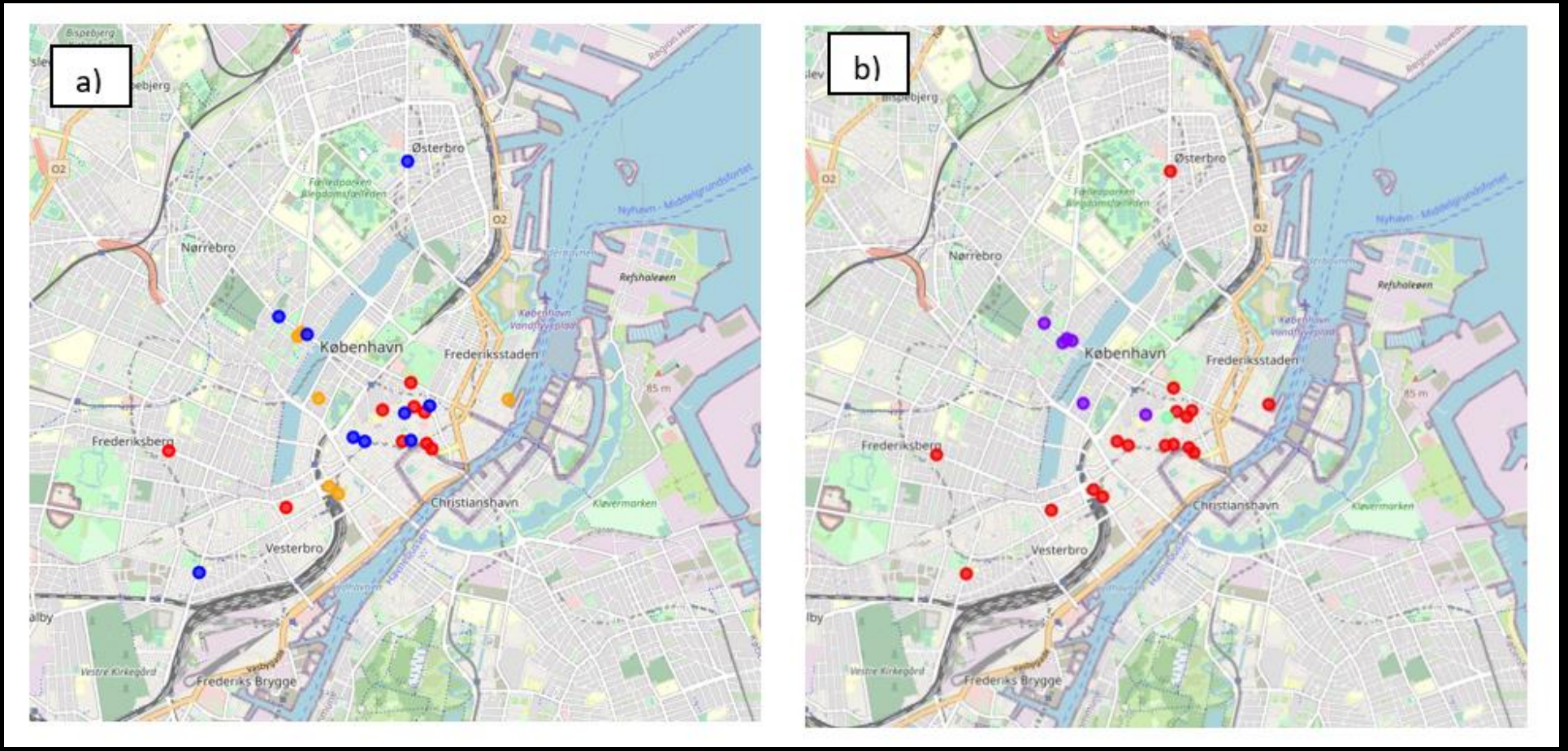
name	Level	Cluster Labels
Mikkeller Bar Viktoriagade	top	0
Ved Stranden 10 - Vinhandel & Bar	top	0
Garden Restaurant & Bar	top	0
Democratic Coffee Bar	top	1
Lord Nelson Bar	top	0
Bo-Bi Bar	top	0
K Bar	top	0
Berlin Bar	top	0
Charlies Bar	top	0
Harbo Bar	middle	1
Library Bar	middle	0
Nimb Bar 'n' Grill	middle	0

name	Level	Cluster Labels
Grillen Burger Bar Nørrebro	middle	1
Gorki Cafe Bar	middle	1
SALT Bar & Restaurant	middle	0
Cosy Bar	low	0
Ravnsborg kitchen and bar	low	1
Temple Bar	low	1
Benni's Bar	low	0
Hornsleth Bar	low	2
Hoppes Cafe & Bar	low	0
The Australian Bar	low	0
The South African Bar	low	0
LA Tequila Bar	low	0

Table 2. Bars rated. The header Level reflects the rating obtained: top = higher than 8; middle = higher than 7 but lower or equal to 8; low = 7 or less. The cluster labels resulted from the k-means clustering with machine learning, using k = 3.

Map visualization by Rating vs Clustering Labels

Fig 2. a) A map of Copenhagen showing the rated bars. Red circles correspond to high-rated bars (>8); orange for medium-rated bars (>7); blue for low-rated bars (=<7); b) A map of Copenhagen of the venues analyzed with a color code for the Cluster Labels obtained from the machine learning algorithm. Red = 0, Purple = 1, Light green = 2.



Discussion

- As can be seen in **Table 2**, the resulted cluster labels did not matched the expected rating labels, corresponding to high rating (>8.0), medium (<8.0 and >7.0), and low (<7.0), which might be due the big amount of variables or that k-means is not the adequate machine learning algorithm. Further work could be done on analyzing the composition of the venues using fewer specific categories. For example, instead of having Women's, Men's, Shoe's Stores might be more helpful to keep them as a single category: Clothing Store.
- **Figure 2** shows that none of the k values is good enough for the analysis of the data. If narrowing the categories does not bring the expected results, might be helpful to try a different algorithm for clustering them.

Conclusion

- To better understand how venues get a high rating, might be useful to complement with more diverse location data, as location of public transportations, or complement with what is available as premium user in the Foursquare API: menu, tips and comments from users.

References

- https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1. [21/06/20]
- https://en.wikipedia.org/wiki/K-means_clustering#Algorithms. [21/06/20]
- https://www.scikit-yb.org/en/latest/api/cluster/elbow.html#:~:text=K%2Dmeans%20is%20a%20simple,number%20(k)%20of%20clusters.&text=The%20elbow%20method%20runs%20k,average%20score%20for%20all%20clusters. [21/06/20]