Unboxed Store Auntie M, Karlsruhe, Germany

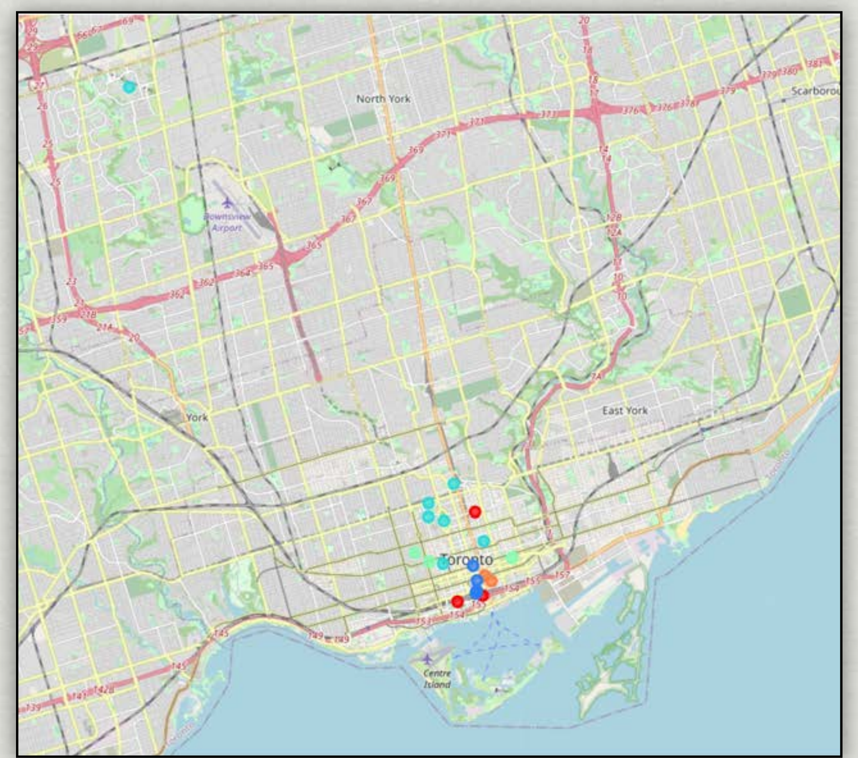
Source: *www.oekolandbau.de / Copyright BLE / Nina Weiler*

SETTING UP A CHAIN OF UNBOXED FOOD STORES IN TORONTO, ON

AN EARLY-STAGE EXPLORATION OF FEASIBILITY BASED ON THE ANALYSIS OF AVAILABLE DATA

Contents

- Introduction
- Data used
- Methodology
- Results
- Discussion
- Conclusion and Recommendations



Introduction

Project assigned to me as a data scientist by German chain of unboxed food stores (customer):

* Unboxed food stores are part of a zero-waste culture, no packaging and buy only what you need
* The idea is that the chain wants to set up 3 – 5 stores in Toronto, ON
* The task is to perform an early-stage analysis on available data, by characterizing suitable neighborhoods through their

    1. Pedestrian traffic,
    2. Number of friendly venues nearby – positive ecosystem for the stores
    3. Number of competitors nearby

    (Ranked by importance of parameter to customer)

Data Used

Data needed for

- List of Toronto Neighborhoods – scraped Wikipedia table with Toronto postal codes
- Neighborhood geographical coordinates – obtained via Bing Maps API
- Number of friendly venues and their coordinates – acquired via the Foursquare API
- Number of competitors – also via Foursquare API
- Pedestrian traffic – Excel file with pedestrian and vehicle traffic counts at signaled crossings during 8 peak hours, data from 2011 – 2019, provided as download by the Toronto city council

Methodology

1. Extracted neighborhoods form the column of the same name in the Wikipedia Toronto postal codes table
2. Created a table in Python (pandas data frame) with 210 neighborhoods
3. Added neighborhood geographical coordinates via Bing Maps API calls, removed duplicates, i.e. excluded a neighborhood if it had identical coordinates with another higher up in the table
4. Added Number of friendly venues and competitors in a radius of 500 m around the neighborhood through Foursquare API calls
5. Extracted geographical coordinates and pedestrian traffic counts from the Excel table into a new data frame

Methodology (Continued)

6. Converted the traffic counts from absolute numbers into 5 categories (0 – 4) from very little to very high pedestrian traffic
7. Searched data frame for all crossings in a radius of 500 m around a neighborhood and summed their category values
8. Added sum to the original data frame
9. Performed data analysis via machine learning, k-means clustering

Methodology (Continued)

	Neighborhood	Latitude	Longitude	Friends	Foes	Pedestrians
0	Parkwoods	43.755997	-79.329544	1	0	0.0
1	Victoria Village	43.728336	-79.314789	1	1	0.0
2	Regent Park	43.660652	-79.360489	5	7	0.0
3	Harbourfront	43.639191	-79.384201	1	6	1.0
4	Lawrence Manor	43.725235	-79.439537	0	0	0.0
5	Lawrence Heights	43.721920	-79.450676	1	0	0.0
6	Queen's Park	43.663425	-79.391914	47	11	4.0
7	Ontario Provincial Government	43.651894	-79.381714	26	21	35.0
8	Islington Avenue	43.639015	-79.547791	0	0	0.0
9	Humber Valley Village	43.670109	-79.521339	0	0	0.0

First ten rows of the data frame used for machine learning,
k-means clustering was performed over the variables in the columns Friends, Foes and Pedestrians

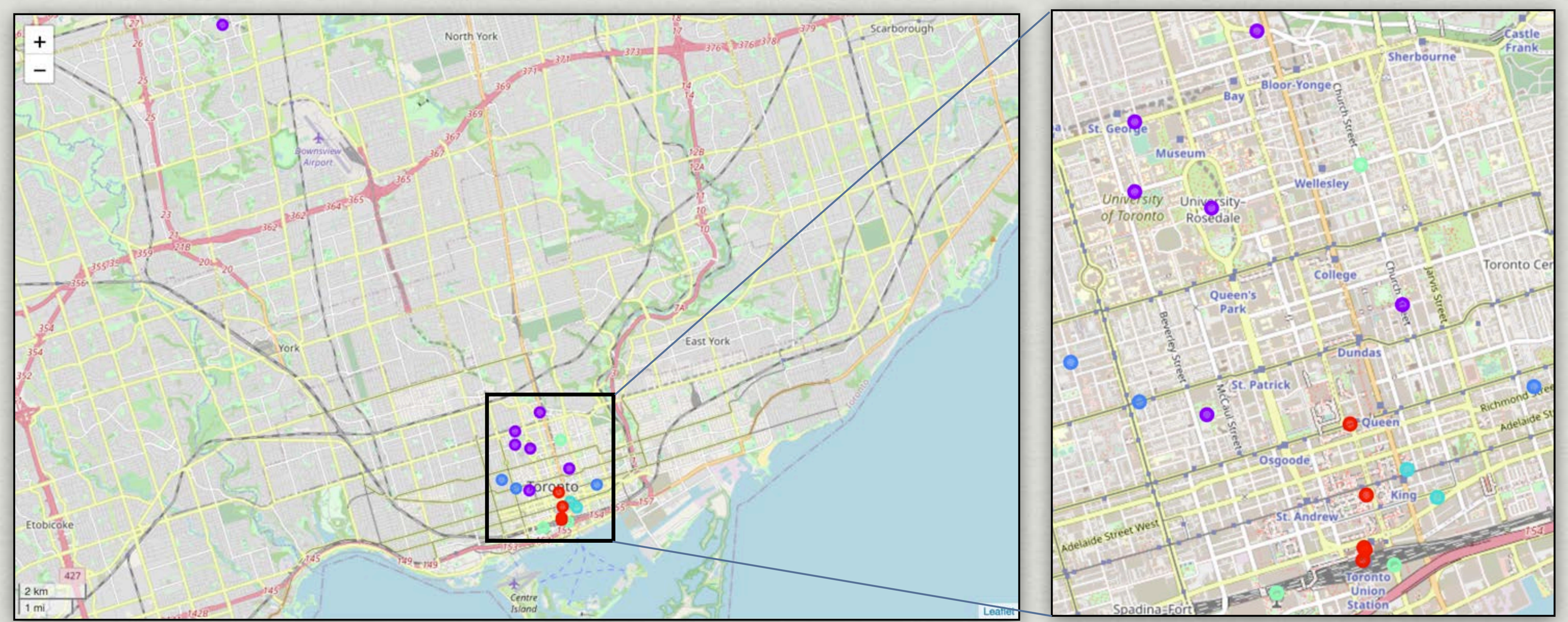
Results – k-Means Clustering

- Elbow method showed most sensible number of clusters was 7
- Excluded clusters 0 and 6 from the further evaluation because their pedestrian traffic (most important variable) is 0, friends and foes also low, if not 0 as well
- Left 20 from 210 neighborhoods in the remaining clusters 1 - 5

	Pedestrians	**Friends**	**Foes**
Cluster			
0	0.0	3.5	6.0
1	29.0	12.0	20.0
2	8.0	47.0	11.0
3	4.0	37.0	38.0
4	31.0	16.5	42.5
5	11.0	6.0	16.0
6	0.0	0.0	0.0

Median values of the variables in the 7 clusters

Results – Clusters on Toronto Map



Discussion

Cluster 4 – most attractive

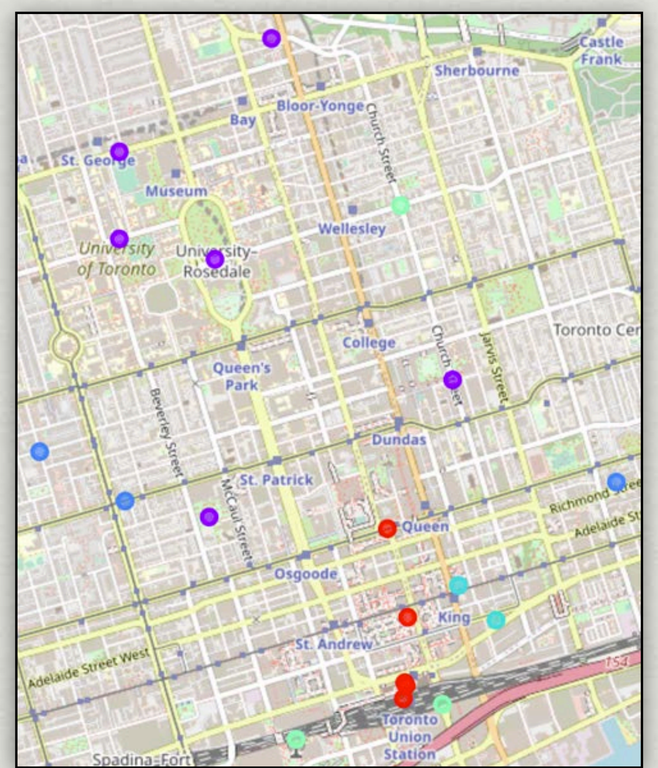
- 2 members, light blue markers in map
- Highest in pedestrian traffic, unfortunately also in number of competitors

Cluster 1 – attractive

- 5 members, three so close that they actually have to be counted as one, red markers
- Almost as high in in pedestrian traffic as cluster 1, also high number of foes

Cluster 5 - somewhat attractive

- 3 members, light-green markers
- Lower in pedestrian traffic, more foes than friends



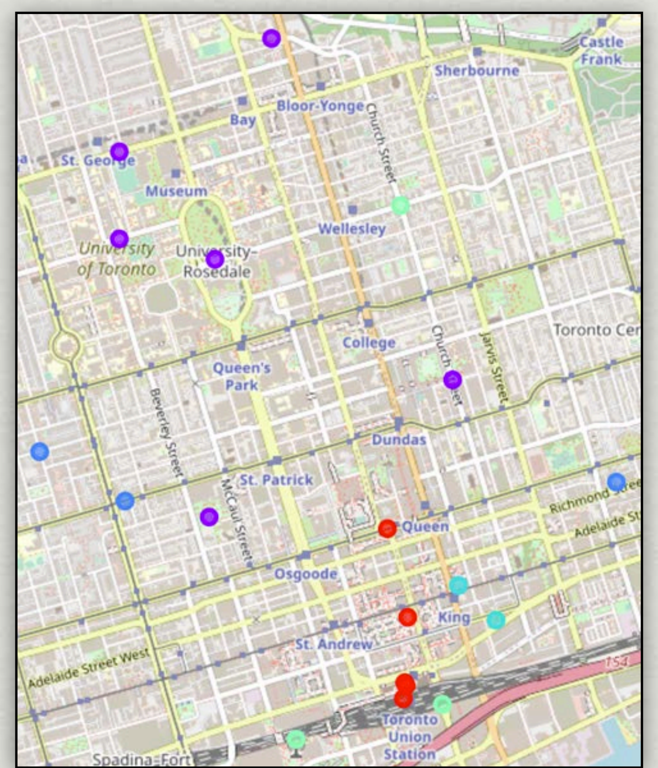
Discussion

Cluster 2 – fourth place

- 7 members, purple markers markers (one, York University, not on map)
- Limited pedestrian traffic, very high in friends, at least double than foes

Cluster 3 – fifth place

- 3 members, blue markers
- Even more limited pedestrian traffic, very high and almost equal number of friends and foes



Conclusion and Recommendations

Conclusion

- We were able to narrow down on 20 most attractive Toronto neighborhoods for setting up unboxed food stores from a total of 210
- These 20 neighborhoods are clustered and ranked by degrees of attractiveness as prescribed by the customer

Recommendations

- Perform further analysis on the determined neighborhoods using other parameters, e.g. property prices and availability of store space
- Finally, a physical visit to the city, if the venture is still attractive

Thank you!

– For more explanations and detailed information please refer to the Jupyter Notebook.