

How Child-Friendly is Toronto?

Capstone Project Report

Introduction / Business Problem

An interesting question for city management and families living in a city alike is about the quality of life for our youngest. Access to adequate care and education services, to leisure activities as well as proper and affordable housing are important aspects of a child-friendly city.

However, resources for addressing potential shortcomings in these areas are always limited in the budget of a city, and demands might differ in different parts of a city. It is hence of crucial importance to find out more about the living situation of children and their parents in the city in order to better inform decision making regarding, for example, new schools and child services, new playgrounds, subsidized housing, and more.

The core business question at hand can hence be summarized as “What is the living situation of children in different parts of the City of Toronto like? Are there areas with similar characteristics and challenges?” Potential stakeholders interested in the answer are the city administration/management as well as the children & parents living in Toronto, current ones as well as future ones!

We will look at this question considering several categories of geographical and demographic data.

Data

We are going to use two data sources for this task:

1. Data about businesses and other forms of venues provided by the Foursquare API
2. Demographical data provided by the city of Toronto via their Open Data portal at <https://open.toronto.ca/>

The Foursquare API is going to be used to retrieve information about two important aspects regarding the life of children, i.e. access to education & care and leisure activities. The search functionality of the API enables us to search for different types of venues by category which allows us to easily search for daycare services, preschools, playgrounds, candy stores, etc. in the vicinity of a given coordinate. In connection with the demographical data, this enables us to provide detailed neighbourhood profiles describing how close the access to education and how much fun it is to live in a neighbourhood as a child!

The Open Data Portal of Toronto provides a huge variety of datasets that can be freely accessed and downloaded. We are going to leverage two of them:

1. Neighbourhoods is a dataset providing geographical information about the 140 neighbourhoods (as defined in the social planning department of the city) such as their boundaries and their (centre) locations.

2. Neighbourhood profiles is a dataset largely aggregating 2016 census data containing basic information about age structure, income and workforce structures, unemployment and benefits, ethnic structure, and many more at neighbourhood granularity.

Please note that the data regarding locations does not correspond exactly with what you might have seen in the Capstone Course so far (merged neighbourhoods by postal codes). Unfortunately, the data retrieved from the City of Toronto is incomplete such that neighbourhood locations have to be added manually – a completed dataset (with locations) complemented by geographic information from Bing will be available from the Github repo of this project.

Of course a full discussion of the business question considering all aspects of reality is out of the scope of this project. Please be aware of several simplifications:

- There are more aspects regarding quality of life for children and their parents than just the access to schools and the like and options for leisure time. Safety, traffic, and environmental issues play a big role, too, but are neglected here for the sake of brevity.
- The interpretation of the data would be best done by experts knowing Toronto a bit better than an online student from some remote part of the world! Local knowledge would be useful of course to make more reliable statements about reasons for certain values.

Methodology

After cleaning the data and deriving some features that were not explicitly contained in the data, we end up with the following features per neighbourhood

- Neighbourhood name, neighbourhood code, and geographical coordinates
- Population and land area
- Relative number of children in the overall population
- Average income
- Relative number of not-suitable housing, defined as housing with less than one room per person
- Relative number of affordable housing, defined as housing for which less than 30% of the income is used for
- Relative number of inadequate housing, defined as housing required major repairs.

First, we explored the data via boxplots to see if any attributes showed only very little variance which would then make them less suited for clustering neighbourhoods into similar areas. Outliers were kept in the data set due to either very few instances or because there was no reason to doubt their plausibility. A correlation analysis was performed to identify attributes that could have a confounding effect. We noticed a quite strong correlation between average income and non-suitability of housing and decided to drop the latter before clustering.

We applied K-Means clustering to cluster the neighbourhoods and applied the elbow method to decide a good number of clusters. It turned out that $k=4$ would be a good value, although the “elbow” is quite smooth and not very pronounced. After that, clusters were compared visually via maps and statistically via the mean values in the clusters

Results

The four resulting clusters differ quite a lot in their characteristics. Two clusters, that could be summarized as the poorer outer suburbs, not only have a lower average income but also limited close access to education and leisure venues (Fig 1 and 2). Another cluster, being located west of the city centre, around the city centre and spreading across the northern shore, has a high average income and also scores well in “fun and education” (Fig 3). The city centre, essentially the fourth cluster, scores even better but a low number of children living there (Fig 4).

A more detailed description can be found in the notebook of the project. Below you find figures of the locations of the clusters.

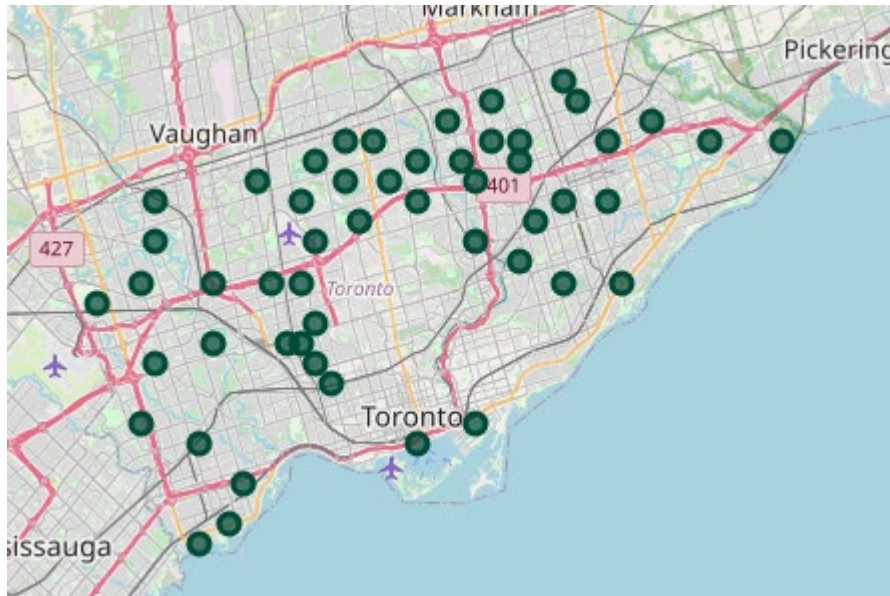


Fig 1: Lower income and limited access to education and leisure



Fig 2: Lower income and limited access to education and leisure, part 2



Fig 3: Highest income and better access



Fig 4: Where fun and education reside: the city centre

Discussion

The first main finding is related to the first two clusters. They include larger areas of the outer surroundings of Toronto with low average income. In these areas, access to education institutions and leisure venues seems to be less close than in the other parts of the city. The fact that less access to these two things and low income match in these two clusters should be investigated further as poverty and limited access to education might cause more problems. If the city had to prioritize investments to education and activities for children in general, they should maybe focus on these neighbourhoods.

The second finding is about the structure in down-town Toronto with very high scores regarding education and fun places for children on the one hand, the lowest number of children in the overall population on the other hand. It would be interesting to learn more about the flows of children to schools between neighbourhoods to better understand whether or not children have long ways to school (possibly located down-town) - this might have effects on health of children as well as traffic and ecological aspects. The city might also want to find out if it might pay off to make the city centre more attractive to young families.

Conclusion

The initial question of “How Child-Friendly is Toronto?” is of course a bit tabloid-ish and cannot be answered by this data analysis alone. It is however interesting to see that income and (direct) access to are somewhat connected and visible in the clusters identified. It can furthermore only reflect some aspects of what is important for a happy life as a child in the city as discussed earlier.

Further analysis would be exciting to better understand how these aspects are related to the geographic areas of Toronto – and other cities!