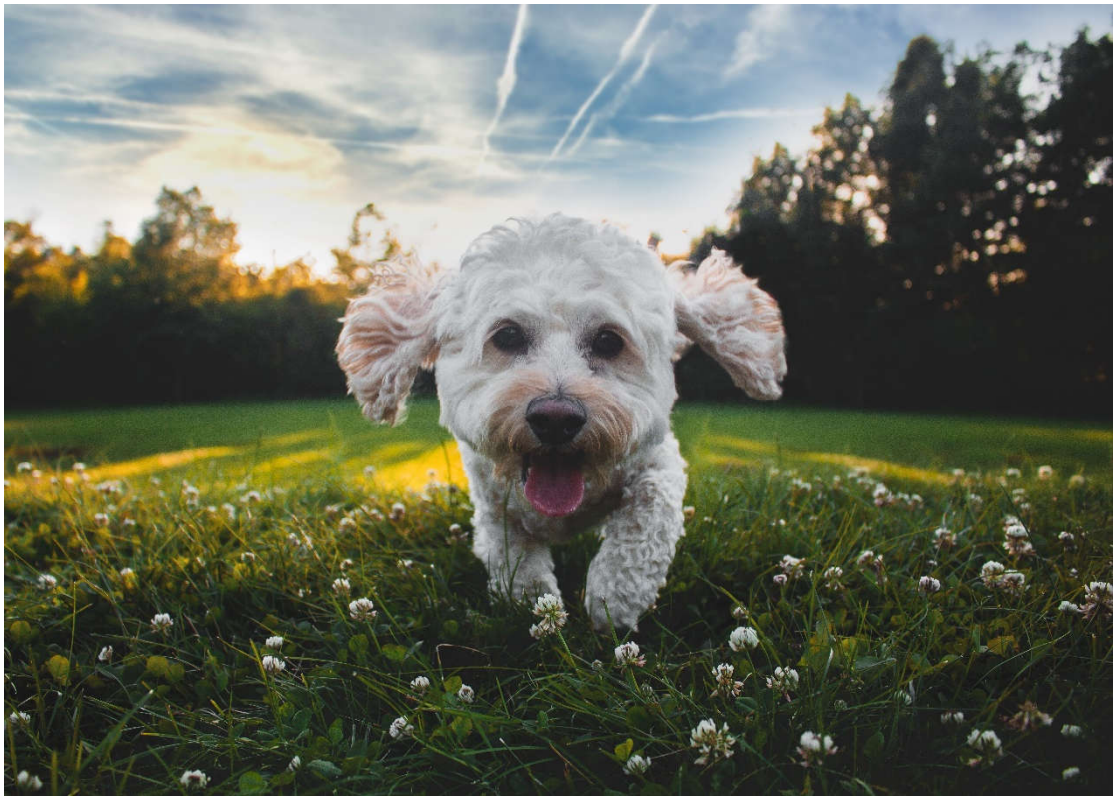


Photo by Joe Caione on Unsplash

Exploring Toronto neighbourhoods – where to open the next Doghotel

CAPSTONE PROJECT

Karin Leisink | IBM Coursera Data Science Introduction | September 2019

0 Preface

This report describes the research I performed for the Datascience Capstone project to complete the Data Science introductory course offered by IBM and Coursera. A problem regarding neighbourhoods is addressed. Solving this problem requires location data. In this first draft of the report I will outline the problem under investigaton and mention the stakeholders and the people who may possibly take an interest in the outcome of the research or in the way the resarch is performed. The datasets used during the analysis phase will be stated and briefly explained. At a later stage I will produce a second and final draft of this report in which I will describe the methodology and present the results. My observations and recommendations will also be discussed. A final conclusion will complete the report.

1 Introduction

1.1 Business Problem and Background Discussion

My friend Chelsea, who lives in Toronto, found out that I enrolled in a Data Science course and asked me to help her realise a childhood dream of hers, namely establish a Doghotel that offers boarding and sitting services to dogs. It means dogowners can bring their dog for a day or a couple of days on occasions where they are not able to take care of the dog themselves. It enables them to have the hospital treatment, do the home renovation or go on holiday, knowing that their dog is in goods hands.

Now, Chelsea is wondering whether there is a neighbourhood in Toronto where she could open the very first dog hotel. And not only that, she hopes that the surrounding neighbourhoods are also deprived of any doghotels. Chelsea has given me two extra requirements:

- there should be 'enough' potential clients for her in that neighbourhood, that means dogs who are chipped (registered) and vaccinated.
- the owners of the clients ought to be able to afford the service. The cost of living in Toronto is very high [1]. It is estimated you need to earn at least $50,000 a year to meet your expenses. Keeping a dog is not cheap ($2,000 a year provided you have a healthy dog) and the high quality dog hotel service which Chelsea is going to offer, will cost around $40 for a stay of one day and $60 for an overnight stay. Chelsea expects that dog owners, who receive a yearly income of $70,000 or more, are willing to spend that amount of money for their dog's stay in the hotel.

The following stakeholders and interested parties can be identified:

1. My friend Chelsea who wants to establish a dog hotel

2. People who would like to establish a dog hotel (preferably not in Toronto just yet, let Chelsea set up a business first)
3. Future Data Scientist for whom this research might be an interesting case study.
4. My colleagues who are software engineers and take an interest in the domain of Data Science and in the Data Science Introductory course that I am taking. This research gives them an idea of what the course is about.

1.2 Datasets and usage of data

Nr	Dataset	Source
1	List of Toronto postal codes with corresponding Neighbourhood and Borough names	Scraped from a wikipage[2]
2	List of Geospatial Codes of Toronto neighbourhoods.	Supplied by Coursera[3]
3	List of registered cats and dogs in Toronto (2019)	Toronto Open Data[4]
4	List of neighbourhood profiles in Toronto from the 2016 census	Toronto Open Data[5]
5	List consisting of postal codes and Neighbourhood names to link the neighbourhoods mentioned in the neighbourhood profiles list to the scraped Toronto postal codes	Self-constructed with the help of wordpostalcode.com [6]
6	List of dog boarding and dog sitting services in Toronto.	Self-constructed with the help of the google search engine[7].

Some remarks have to be made about the acquisition of the datasets and the processing of the data:

- The postal codes were scraped with the help of the BeautifulSoup package. Postal codes without an assigned borough and neighbourhood were discarded. Postal codes without an assigned neighbourhood received the borough name as neighbourhood name.
- The cats were omitted from Toronto open data dataset with the registered cats and dogs.
- The list with the 'Neighbourhood data from Toronto open data' contains lots of data about the inhabitants of the various neighbourhoods. I am only interested in the annual income data of households (income data of individuals were also given, but dogs are kept in a household and households can have double incomes, hence the household income data were used). In the dataset the income data is divided into income categories like the following:

  Under $5,000

  $5,000 to $9,999

  $10,000 to $14,999

  $15,000 to $19,999

  ……

$60,000 to $69,999

$70,000 to $79,999

.....

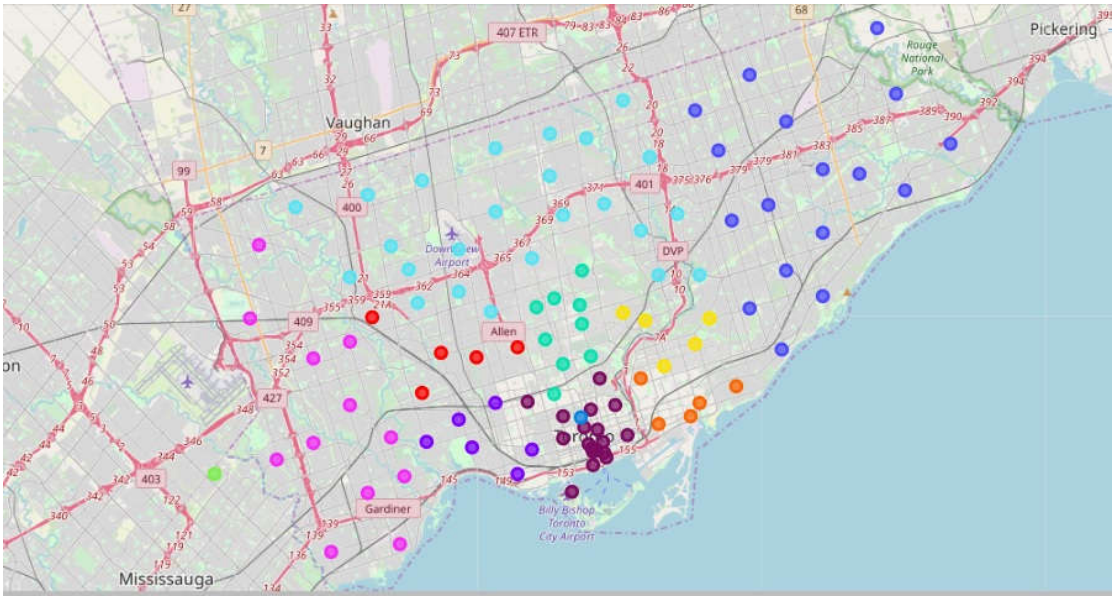
$200,000 and over

For each category and each neighbourhood the number of households are given. At first I kept all the household income columns. But in order to calculate the percentage of households with an income of $70,000 or more, I first calculated the total number of households per neighbourhood. Then I added up all the households in the categories of $70,000 to $79,999 and the categories above. With these two totals the percentage of households with an annual income of at least $70,000 could easily be determined:

*(1 / total amount of households) * total amount of households earning at least $70,000*

- The neighbourhoods in the dataset with the Toronto neighbourhood data differ from the 'scraped' neighbourhoods and sadly, no postal code was given. In order to join the two datasets I produced a postal code with every neighbourhood name in the dataset with the Toronto neighbourhood data. The neighbourhood names with the looked up postal codes were stored in a dataset.
- Initially I wanted to use Foursquare to retrieve the dog sitting and dog boarding venues [8]. However, dog sitting and dog boarding are not a specific venue category with Foursquare. The venue category 'pet service' comes closest. I retrieved the pet service venues for every Toronto neighbourhood, but the venue name only could point out whether it was a dog sitting or dog boarding facility. With names like 'My Pet Food 'N More','Paws & Affection','Velvet Paws', it is just not clear what their business is. If I want confirmation I should look up these venues on the internet. After filtering the dog carers I had 9 venues . In a city of 2,5 million inhabitants and over 50,000 registered dogs, there have to be more venues. So I decided to look up all the dog sitting and dog boarding facilities in Toronto with the help of a search engine [7]. I decided to ignore the venues that I retrieved with Foursquare and use the results of my thorough search on the internet to construct my own dataset with dog boarding and dog sitting venues.

With the scraped neighbourhood data, the geospatial codes and the Folium library, the area of activity can be shown in a map with the neighbourhoods depicted as circles. Neighbourhoods belonging to the same borough have identical colours.



Now that all the datasets are collected, the data can be analysed.

References

[1] Cost of living in Toronto

[2] List of postal codes of Canada:_M

[3] Toronto Geospatial codes

[4] List of registered dogs and cats published on Toronto open data (updated 23 July 2019)

[5] List of Toronto neighbourhood profiles from the 2016 census published on Toronto open data

[6] Postal Codes Search by Address, Country, City

[7] Google search engine

[8] Foursquare