Choosing My Favorite Place to Travel

1. Introduction
1.1 Background

- I am a Barista, so the first thing I want to figure out when I plan my travel is which place has most coffee shops. The number of the coffee shops is positively correlated with my degree of happiness. In this case, if I want to travel to Toronto, knowing which neighbourhood in Toronto has the most coffee shops matter a lot to me. So in this project, I am trying to find out how coffee shops distribute in different neighbourhoods in Toronto.
- What’ s more, since this project is for my personal travelling purpose, another important factor I need to consider is the number of hotels because the number of hotel will in a way influence my living choices and living quality.

1. Introduction
1.2 Problem

- The biggest problem for me is to find out which place has the most coffee shops and hotels. In this project, I am going to try it in Toronto and find out which neighbourhood in Toronto has the most coffee shops and hotels.

2.DATA
2.1Data needed

- Based on definition of the purpose, factors that will influence my decisions are:
- a. **number of coffee shops** in the neighborhood
- b. **number of hotels** in the neighbourhood

2.DATA
2.2Data source

- Following data sources will be needed to extract/generate the required information:
- a. the coordinate of Toronto will be obtained using **Geopy Library**
- b. number of coffee shops and location in every neighborhood will be obtained using **Foursquare API**
- c. information about Toronto Neighbourhoods will be obtained in **Wikipedia**
- d. the coordinates of neighbourhoods in Toronto will be obtain in this csv link **https://cocl.us/Geospatial_data**

2.DATA
2.3Data cleaning

- First, I downloaded or scraped data of basic information of Toronto from Wikipedia and link https://cocl.us/Geospatial_data. The data were combined into one table. However, there is a little problems with the data downloaded. For example, the basic data of neighbourhoods information in Wikipedia has missing value with "Not assigned" and duplicated data of the same postal code. In this case, I drop the missing value and combine the rows with same postal code because postal code is the unique identity of an area.
- Then, I downloaded the venues information about different neighbourhoods in Toronto using Foursquare API. And I use one-hot encoding to deal with the data obtaining from Foursquare API.

3.Methodology

- In this project, I use one-hot encoding to give values to different venue categories. The neighbourhood with highest values in "hotel" and "coffee shop" will be the neighbourhood I choose to travel to and live in.
- (*Specific data and methodology is in the notbook)

3.Methodology(partial of data)

	Neighbourhood	Coffee Shop	Café	Hotel	mean
0	Adelaide, King, Richmond	0.070000	0.05	0.03	0.050000
1	Agincourt	0.000000	0.00	0.00	0.000000
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.000000	0.00	0.00	0.000000
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0.111111	0.00	0.00	0.037037
4	Alderwood, Long Branch	0.100000	0.00	0.00	0.033333

4.Results and Discussion

- Since Woburn has the highest value of the indicators I set, it has the greatest number of hotels and coffee shops surrounding it. In this case, anybody who like coffee may choose Woburn as a travelling destination.

5.Conclusion

- Since Woburn has the highest value of the indicators I set, I suppose I should choose to live in there next time I travel to Toronto. With the highest value, Woburn would have the greatest number of coffee shops and hotels. The great number of hotels can guarantte my living quality there. And because I am a Barista, the great number of coffee shops will put me in a good mood as I can have more chances to enjoy delicious coffee.