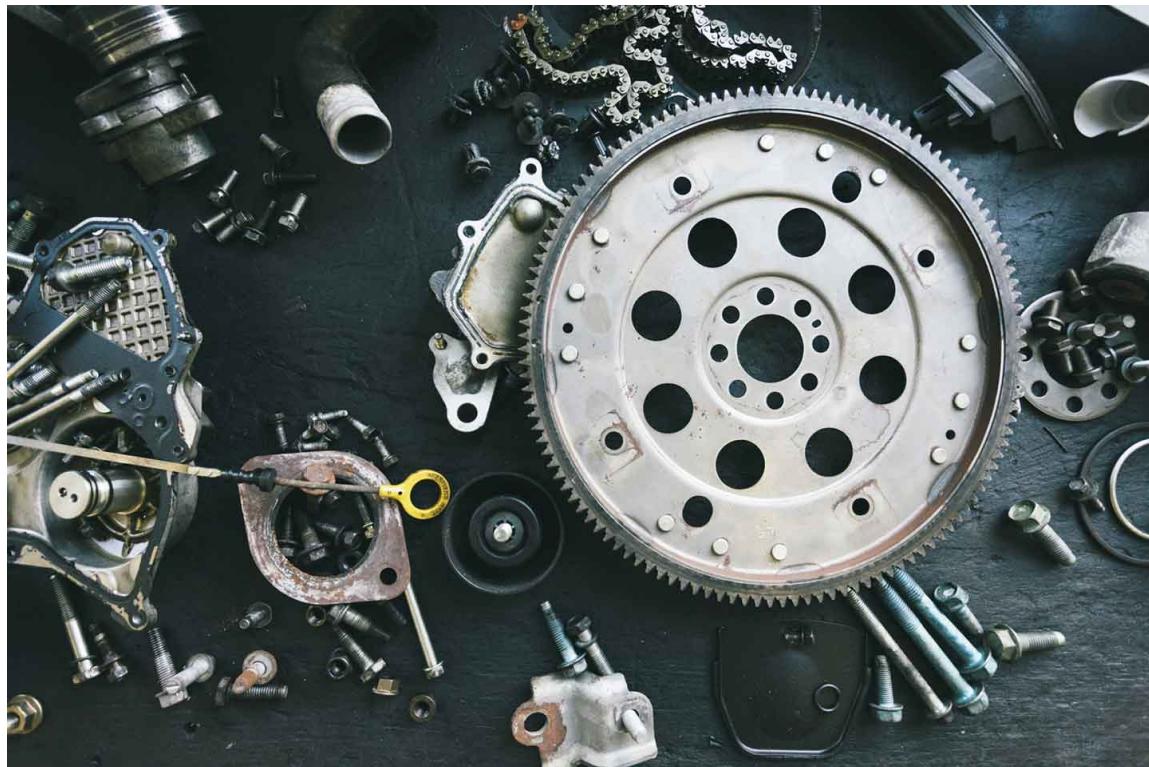


JOURNEY PLANNER

Capstone Project

25 March 2019

Helping travellers to organize their trip... faster!



Erika Agostinelli

TABLE OF CONTENTS

INTRODUCTION.....	3
DATA.....	3
METHODOLOGY	4
1. COLLECTION OF POI.....	4
2. FINDING LATITUDE AND LONGITUDE OF THE POI.....	4
3. CLUSTERING THE POI.....	5
<i>Clustering Results</i>	5
4. CLUSTERING THE CITY CENTRE POI.....	6
<i>Clustering Results</i>	6
5. FIND THE HOTELS AROUND CITY CENTRE.....	7
6. FIND DETAILS OF THE HOTELS: RATINGS, WEBSITE ETC.....	7
<i>Filtering Hotels</i>	8
7. FIND THE BEST LOCATED HOTEL.....	8
<i>Small Correction:</i>	9
RESULTS	11
DISCUSSION	12
CONCLUSION	12

INTRODUCTION

Every time I want to visit a new city, I want to make sure that I can visit the most representative places (POI - Point of Interests). Moreover, I want to find a strategic place for my hotel (starting point) and grouping the visits per each day in the most efficient manner. In my hypothetical trip to Toronto, I have 10 days of travel to schedule, to make the most out of it. Therefore, with this project I want to:

- Find a method to cluster my POI, so that I can visit each day, places that are in the same area.
- Find a method to search for the best located hotel

DATA

1. Collection of POI from the following website: <https://theculturetrip.com> - **Web scraping using BeautifulSoup**
2. Finding latitude and longitude of the POI found above - **Foursquare API call (Regular calls)**
3. Clustering the POI - **K-means Clustering**
4. Clustering the city centre POI - **K-means Clustering**
5. Find the hotels around city centre - **Foursquare API call (Regular calls)**
6. Find Details on the hotels - **Foursquare API call (Premium calls)**
7. Find the best located hotel - **Google Maps, Directions API**
8. Show photos of the top 5 hotels - using data collected in point 6

METHODOLOGY

1. Collection of POI

What are the best places to visit in Toronto?

Since I am not familiar with this city, the first step is to start a research on Internet. Usually after reading 4 or 5 websites, you start to have a good idea of what are the "must-see" attractions that also fits your interest. For this project, I am going to pick a website that collects the most iconic places in Toronto which they recommend to visit. Let's use BeautifulSoup for scraping the website and retrieve the titles of each article to create a collection (array) of places to visit. I will analyse the `html` of the website using the "inspect element" tools and localise the titles of the places. (fig.1) Website scraped: <https://theculturetrip.com/north-america/canada/articles/20-must-visit-attractions-in-toronto/>

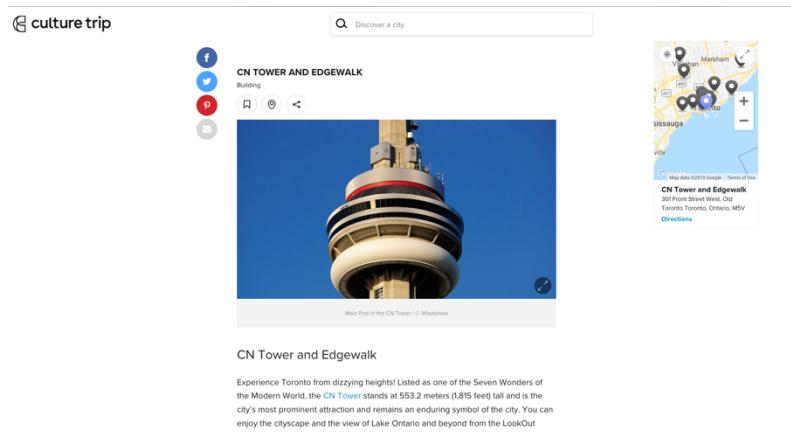


Figure 1 - Page that was scraped for this project. The title of each article represents the name of the places that we want to visit.

2. Finding latitude and longitude of the POI

Once we have created the array of the locations on the previous point, we need to find the latitude and longitude of each places. These coordinates will help us cluster the locations into groups in the following step (#3). Let's use the **Foursquare API (Regular calls)**¹ in order to get this information.

¹ <https://foursquare.com/developers>

3. Clustering the POI

We have 17 locations that we want to visit.

How do we cluster them? I decided to use their coordinates so that closer locations can be visited in the same day (cluster) - Technique used: Partitional Clustering - **K-means**

Clustering.²

The initial number for clustering is $k = 7$ because I am expecting that most of the attractions are going to be in one big group (in the city centre) and consequently, I will need to divide them furthermore, afterwards.

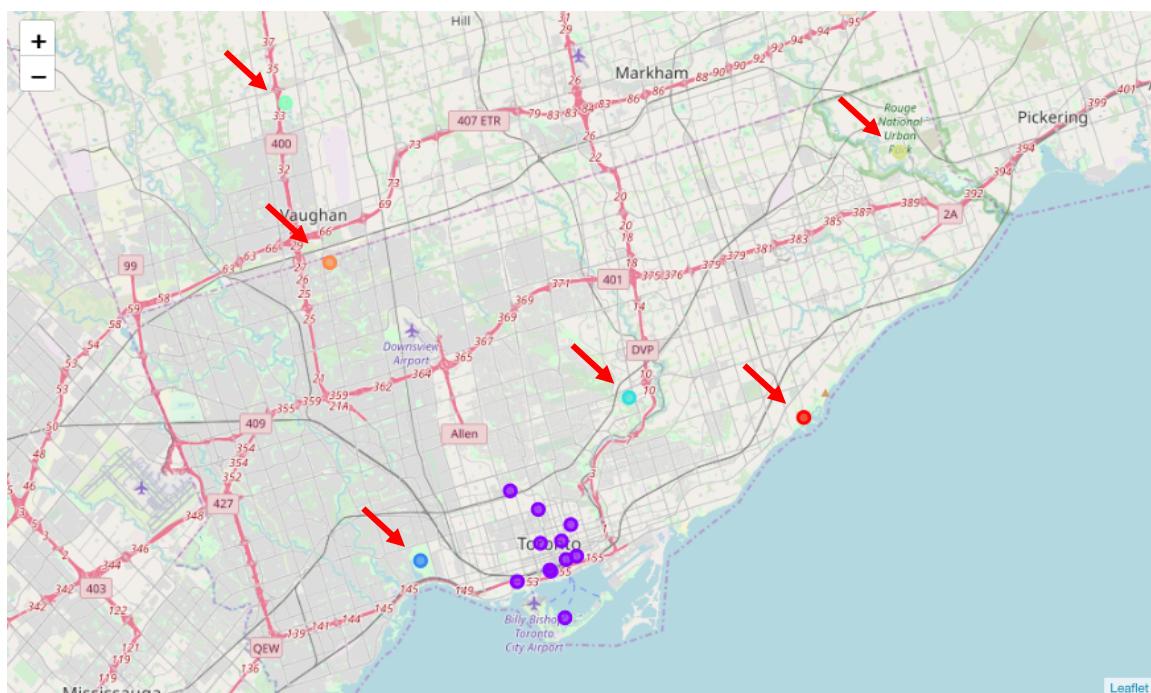


Figure 2 - Clustering Toronto's POI

Clustering Results

As anticipated, most of the clusters are positioned far away from the city centre.

- Day 1: Black Creek Pioneer Village (Orange dot)
- Day 2: Canada's Wonderland (Green Dot)
- Day 3: High Park (Blue dot)
- Day 4: Scarborough Bluffs (Red Dot)
- Day 5: Toronto Zoo (yellow dot)

² <https://en.wikipedia.org/wiki/k-means-clustering>

- Day 6: Ontario Science Centre (light blue)
- the rest: City Centre (purple dots)

From day-1 to day-6 all POI are fortunately destinations where spending the whole day is considered appropriate. (e.g. Canada's wonderland, Ontario Science Centre etc). Moreover, I could rent a car for these days and the remaining 4 days in the city centre I can use the public transportation, therefore avoiding to rent a car for the whole stay.

"Day 7" can be further down divided into more clusters since a lot of destinations were clustered together (step #4).

4. Clustering the city centre POI

The city centre locations (purple points in Fig. 2) are numerous and difficult to visit all in one day. Let's cluster them in 4 remaining clusters (remaining 4 days of travel).

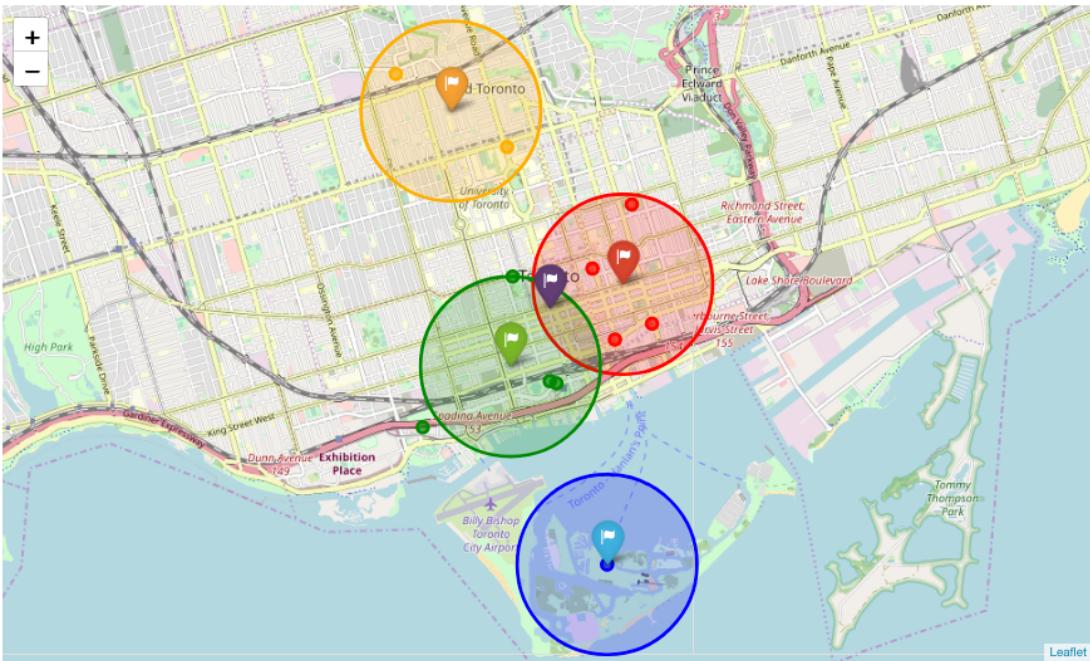


Figure 3 - Clustering the city centre in 4 clusters

Clustering Results

The purple flag is the centroid of the City centre group discovered in the previous step.

The city centres clusters found were the following:

- Day 7: Orange Group : Royal Ontario Museum , Casa Loma
- Day 8: Red Group : Allen Gardens, Toronto Eaton Centre, St. Lawrence Market, Hockey Hall of fame

- Day 9: Green Group: Art Gallery of Ontario, Ripley's Aquarium, CN Tower and Edge Walk
- Day 10: Blue Group: Toronto Islands & Centreville

In the centre of each group, I have highlighted the centroids with a flag icon. These centroids will be important for step #7, when we will try to calculate the distance between the potential hotel and the location of interests.

It seems that we were able to divide all Toronto's POI in a realistic manner. Now it's time to find a hotel.

5. Find the hotels around city centre

Search for the 40 hotels around the city centre centroid (purple flag).

This task was done by using one regular call using Foursquare API and the aim was to collect all the hotels near Toronto, thus creating a starting point for our search.

6. Find Details of the hotels: ratings, website etc...

In order to choose my hotel, I want to know the ratings, description, the website, and one image for each hotel of the area (prefix, suffix, width and height information). This is done by foursquare API (premium calls).

The next step will be to filter these hotels based on my necessity.

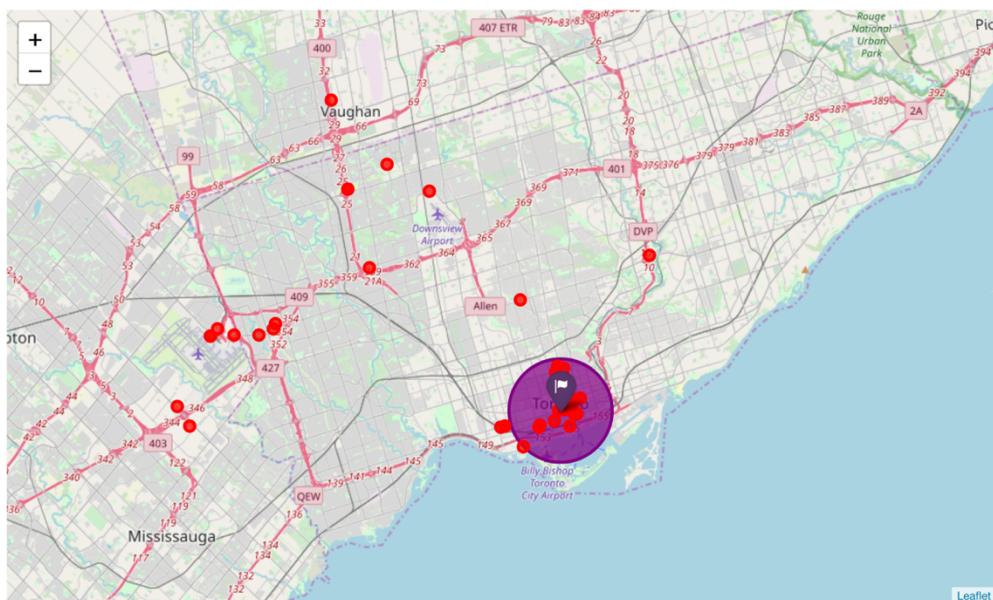


Figure 4 - Hotels in Toronto and the city centre (in purple is the centroid of the city centre cluster) - one of the requirements is that the hotel needs to be less than 18km from the city centre centroid (purple area in the map)

Filtering Hotels

The requirements for filtering the hotels are the following:

1. all the hotels with **no distance or rating** information will not be used.
2. The hotel needs to be **close to the city centre**: 18 kilometres max from the centroid of city centre (purple flag)
3. The minimum rating needs to be **at least 8**

7. Find the best located hotel

I reduced the number of the possible hotels, based on my necessity.

Now, I want to find the best located hotel based on my 4 clusters that I need to visit.

The strategy is that I am going to **calculate the travel time (in minutes) from the hotel to the centroids of each group**, and the best performing hotels will be selected: select a hotel that can **minimise the travel time** during my holiday.

Since it's not recommended to use the car in the city centre but better to use the public transportation³, I am going to calculate the travel time using the "**transit**" mode. In order to calculate the travel time, I am using the **Google Maps - Distance Matrix API**. ⁴

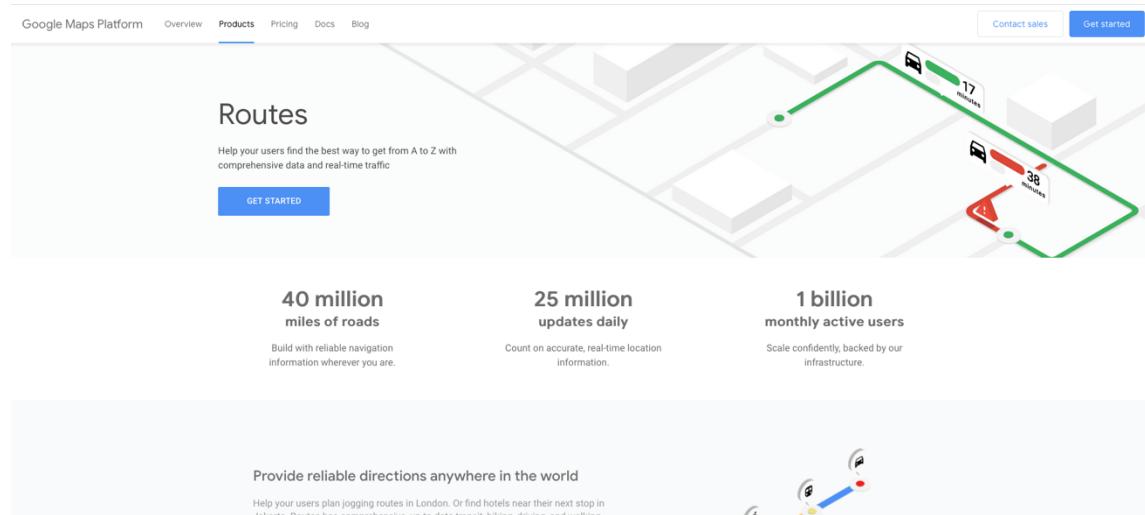


Figure 5 - Google Routes includes also Distance Matrix API functionality

³ https://www.tripadvisor.co.uk/ShowTopic-g155019-i55-k6150900-Car_rental_or_public_transport-Toronto_Ontario.html

⁴ <https://developers.google.com/maps/documentation/distance-matrix/intro>

Small Correction:

The centroid of Toronto Islands & Centreville is positioned on the islands themselves. (see Blue cluster in Fig. 3) – However, Google API struggles to calculate the distance from this centroid probably because it does not have the information on the ferry services that connects the city to the islands. Therefore, I have manually changed the position of the centroid to Jack Layton Ferry Terminal, 9 Queens Quay W, Toronto, ON M5J 2H3, Canada where the Ferry to the island will depart. The coordinates are taken from Wikipedia⁵.

Based on the filtering done in step # 6 and the calculation of the distance between the hotels and the centroids of our four clusters, I have picked the top 5, best located hotels in Toronto.

hotel	id	address	latitude	longitude	distance	ratings	description	to_cluster_1	to_cluster_2	to_cluster_3	to_cluster_4	total_duration	
Cosmopolitan Toronto Centre Hotel & Spa	4ad4c05cf964a520bbf520e3	8 Colborne St (at Yonge St) M5E 1E1	43.649064	-79.377598	17814	8.1	NA	...	5.6	10.1	11.8833	19.8	47.3833
The Rex Hotel Jazz & Blues Bar	4b68aed1f964a520de862be3	194 Queen St W (Queen & St. Patrick) M5V 1Z1	43.650505	-79.388577	17143	8.1	"More Great Jazz than anywhere else, all the t..."	6.85	12.8167	11.1833	17.5667	48.4167	
The Omni King Edward Hotel	4adf7d0bf964a520127b21e3	37 King Street East (b/wtn Victoria St. and Tor...)	43.649191	-79.376006	17884	8.2	Located in the heart of Downtown Toronto, this...	7.63333	10.3167	13.5	22.1167	53.5667	
Le Germain Hotel Toronto Mercer	4ad4c05cf964a520baf520e3	30 Mercer St M5V 1H3	43.645669	-79.391044	17460	8.3	Choose Le Germain Hotel Toronto for elegance, ...	13.25	20.3167	4.18333	27.75	65.5	
The Grand Hotel & Suites Toronto	4b7d9098f964a52014c72fe3	225 Jarvis St. (at Dundas St. E.) M5B 2C1	43.656449	-79.374110	17367	9.0	NA	...	7.11667	19.35	22.7333	30.1333	79.3333

Figure 6 - Top 5 - Best located hotels in Toronto (Data frame) – the last 5 columns contain the travel time (in minutes) from the hotel to each cluster.

Example: **Cosmopolitan Toronto Centre Hotel & Spa** has an overall rating of 8.1 and the distance from the cluster one is 5.6 minutes, from cluster two around 10 min, from cluster three around 12 min and from cluster four around 20 min, using public transportation.

⁵<https://tools.wmflabs.org/geohack/geohack.php?pagename=Jack Layton Ferry Terminal¶ms=43 38 25 N 79 22 31 W region:CA-ON type:landmark>

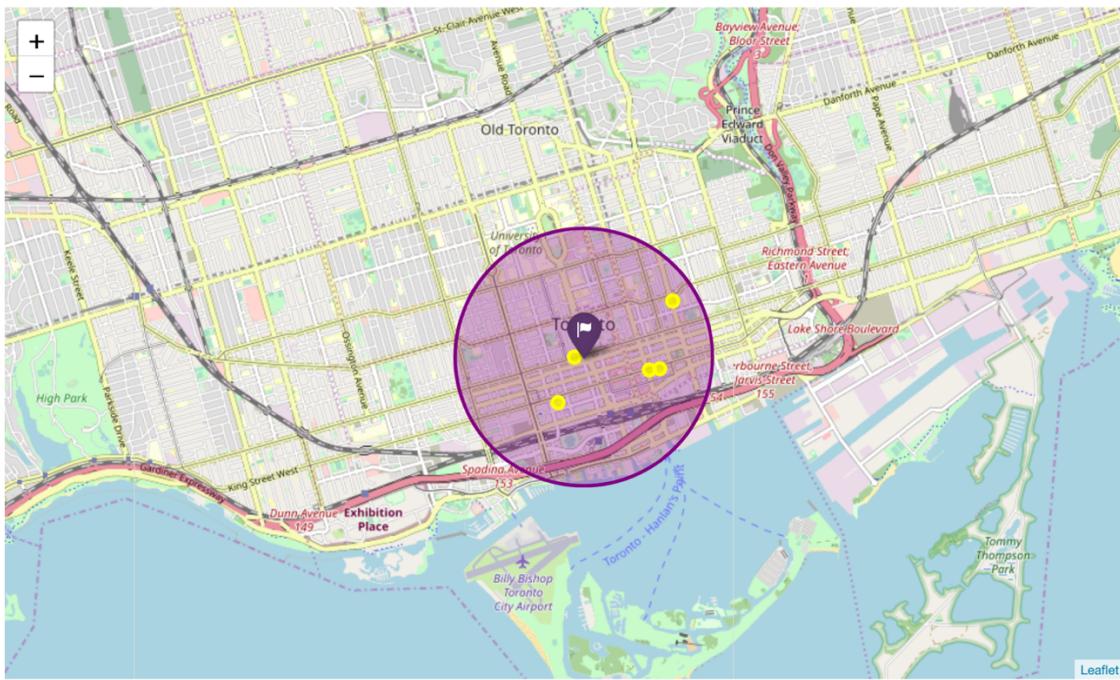


Figure 7 - How the top 5 hotels are located compared to the city centre centroid (Purple flag)

Cosmopolitan Toronto Centre Hotel & Spa
website: NA
rating: 8.1



The Rex Hotel Jazz & Blues Bar
website: <http://therex.ca>
rating: 8.1



The Omni King Edward Hotel
website: <https://www.omnihotels.com/hotels/toronto-king-edward>
rating: 8.2



Figure 8 - Using the data collected in step # 6, we can show one image per each hotel in order to facilitate our choice (Prefix, suffix, width and height of the photo)

RESULTS

This analysis had the aim to facilitate the organization and scheduling of a hypothetical trip to Toronto. We trusted a website that recommended to visit 17 places in Toronto <https://theculturetrip.com>, and we have mapped those destination and clustered them based on their coordinates (Foursquare API used to collect the coordinates).

The initial clustering was done by only using 7 clusters using the k-means method, because the destinations were already far apart from each other and we were already planning to spend the whole day on each of the POI far from the city centre. Moreover, we could group the days where we needed to rent a car in order to reach the distant destinations and to save money. (Fig. 2)

- **Day 1:** Black Creek Pioneer Village (Orange dot)
- **Day 2:** Canada's Wonderland (Green Dot)
- **Day 3:** High Park (Blue dot)
- **Day 4:** Scarborough Bluffs (Red Dot)
- **Day 5:** Toronto Zoo (yellow dot)
- **Day 6 :** Ontario Science Centre (light blue)

The remaining days are dedicated to explore the city centre. Since we are going to use public transportation, we can save money and reduce the rental of the car to 6 days instead of 10 days. (Fig. 3)

- **Day 7 - Orange Group :** Royal Ontario Museum , Casa Loma
- **Day 8 - Red Group :** Allen Gardens, Toronto Eaton Centre, St. Lawrence Market, Hockey Hall of fame
- **Day 9 - Green Group:** Art Gallery of Ontario, Ripley's Aquarium, CN Tower and Edge Walk
- **Day 10 - Blue Group:** Toronto Islands & Centreville

The next step is to find the best hotel. For me, it is very important that the rating of the chosen hotel is reasonable high (≥ 8) and located very close to my POI of the city centre, since I want to minimise the travel time. First thing, I chose the hotels located in the premises of the city centre (18km max) and in order to calculate the travel time I used Google Map's Distance Matrix API to check how long was going to take from each hotel to reach each of the centroids of the city centre clusters. The top 5 best performing hotels are collected in the data frame at point 8 (see Fig. 6).

In conclusion, if I were to choose a hotel from the top 5, I would be inclined to choose **The Grand Hotel & Suites Toronto**. The travel time from each clusters are:

- to cluster 1: 7min
- to cluster 2: 19min

- to cluster 3: 22min
- to cluster 4: 30min

which are reasonable distances/durations and the rating being the highest amongst the others (9.1). However, realistically more research is necessary in order to reach a final conclusion (more photos, checking the websites, checking more reviews from the users etc.).

DISCUSSION

I assume I am not the only person that dreamed of an automatic "Journey Planner". There are some interesting services online like <https://www.inspirock.com/>, <https://www.routeperfect.com/trip-planner>. However some "manual" tweaking is always necessary.

Organising a trip can be very stressful especially because you want to have the best experience (and see everything!). Moreover, it is a very subjective topic since we all have different taste and interests when it comes down to choose places to visit.

In this particular analysis, the strong assumption was that the website <https://theculturetrip.com> could offer a good choice of destinations for any user, but the reality is that most of the time, choosing the POI is already not an easy task (especially if you travel with group of people). In a real use case scenario, creating the "array of the destination" might be more difficult than shown in point 1.

A second strong assumption is that, distant destinations were worth visiting for the whole day - In this particular case, it seems that indeed the further destinations were worthwhile spending an entire day (Canada's Wonderland etc.) but in some other real scenario, this might not be the case. i.e. you could spend half a day in one distant location, and drive to a further different part of the city and spend the remaining day.

Another point worth mentioning is that I could have played more with different clustering techniques. For the analysis I used the method that was introduced during this course.

CONCLUSION

This analysis can be seen as an attempt to create an automatic journey planner that could help in the initial phase of the organization of a trip. However, as discussed in the previous step, it is difficult to satisfy all the necessities and taste of all users. Realistically, this analysis can be used as a template to cluster and organise the POI based on the distance and help you scraping information around hotels that you might be interested in. (& it was a good occasion to play with various APIs :))