



JOURNEY PLANNER CAPSTONE PROJECT

25 March 2019
Erika Agostinelli

AGENDA

1. Problem Statement
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

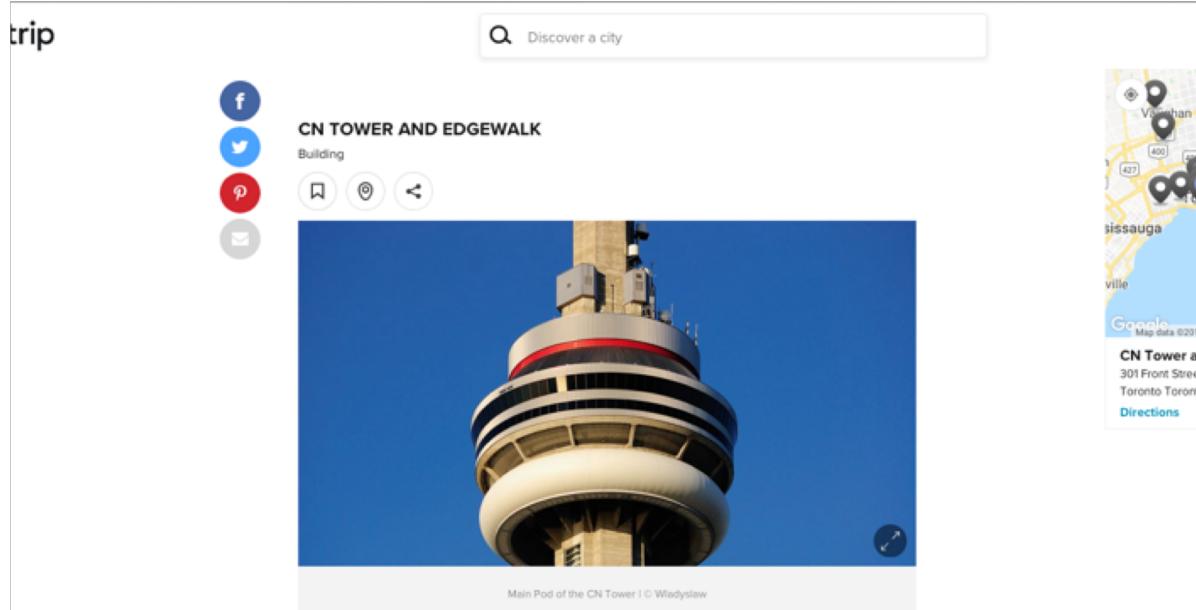
PROBLEM STATEMENT

This analysis has the aim to facilitate the organization and scheduling of a hypothetical 10-days-trip to Toronto.

- ✓ Find a method to cluster my Point Of Interest (POI), so that I can visit each day, places that are located in the same area, minimising the usage of the car and travel time.
- ✓ 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



CN Tower and Edgewalk

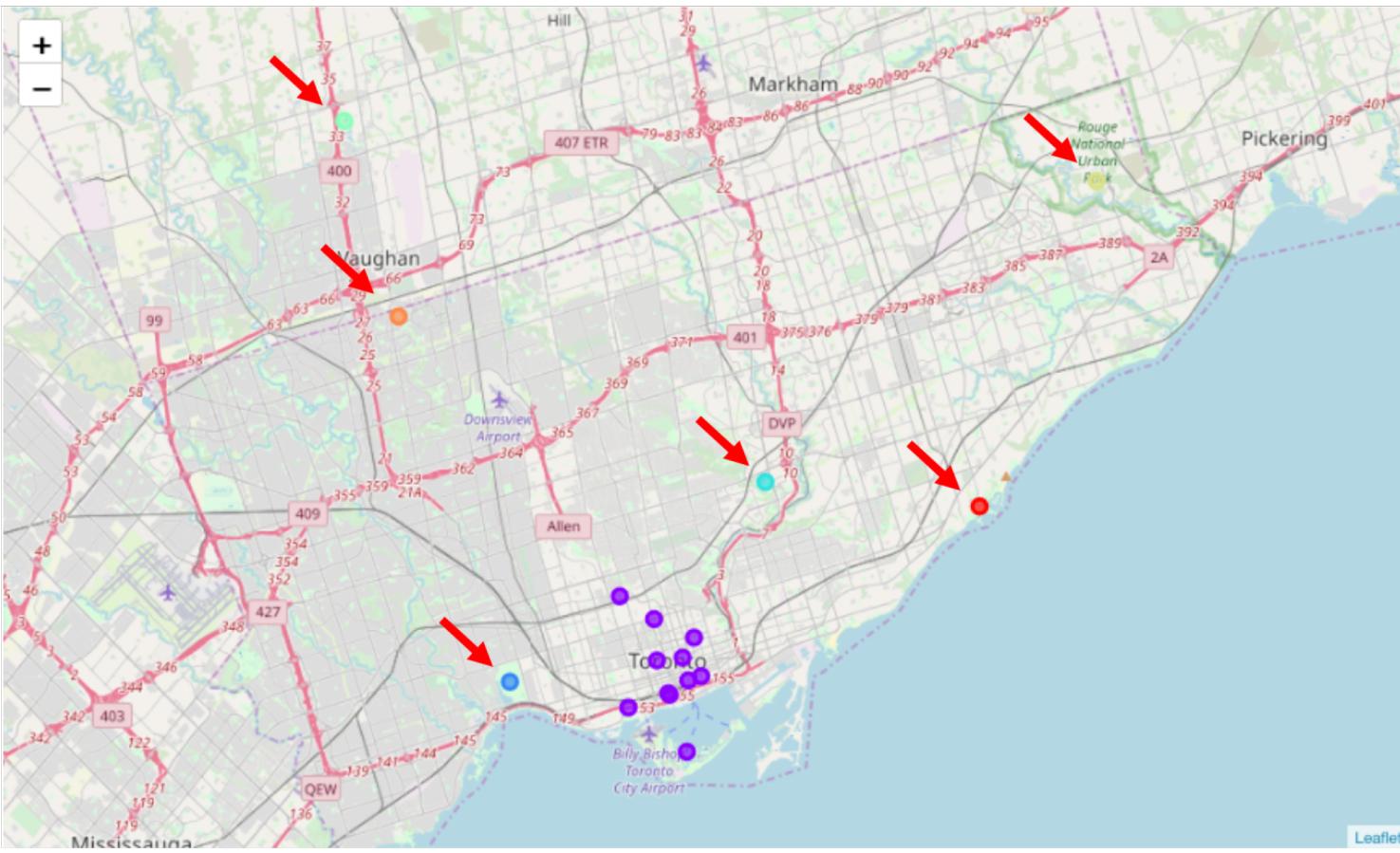
Experience Toronto from dizzying heights! Listed as one of the Seven Wonders of the Modern World, the [CN Tower](#) stands at 553.2 meters (1,815 feet) tall and is the city's most prominent attraction and remains an enduring symbol of the city. You can enjoy the cityscape and the view of Lake Ontario and beyond from the LookOut

METHODOLOGY – CHOOSE POI

Let's use BeautifulSoup for scraping the website ("must visit attractions in toronto") and retrieve the titles of each article to create a collection (array) of places to visit.

Website scraped:

<https://theculturetrip.com/north-america/canada/articles/20-must-visit-attractions-in-toronto/>



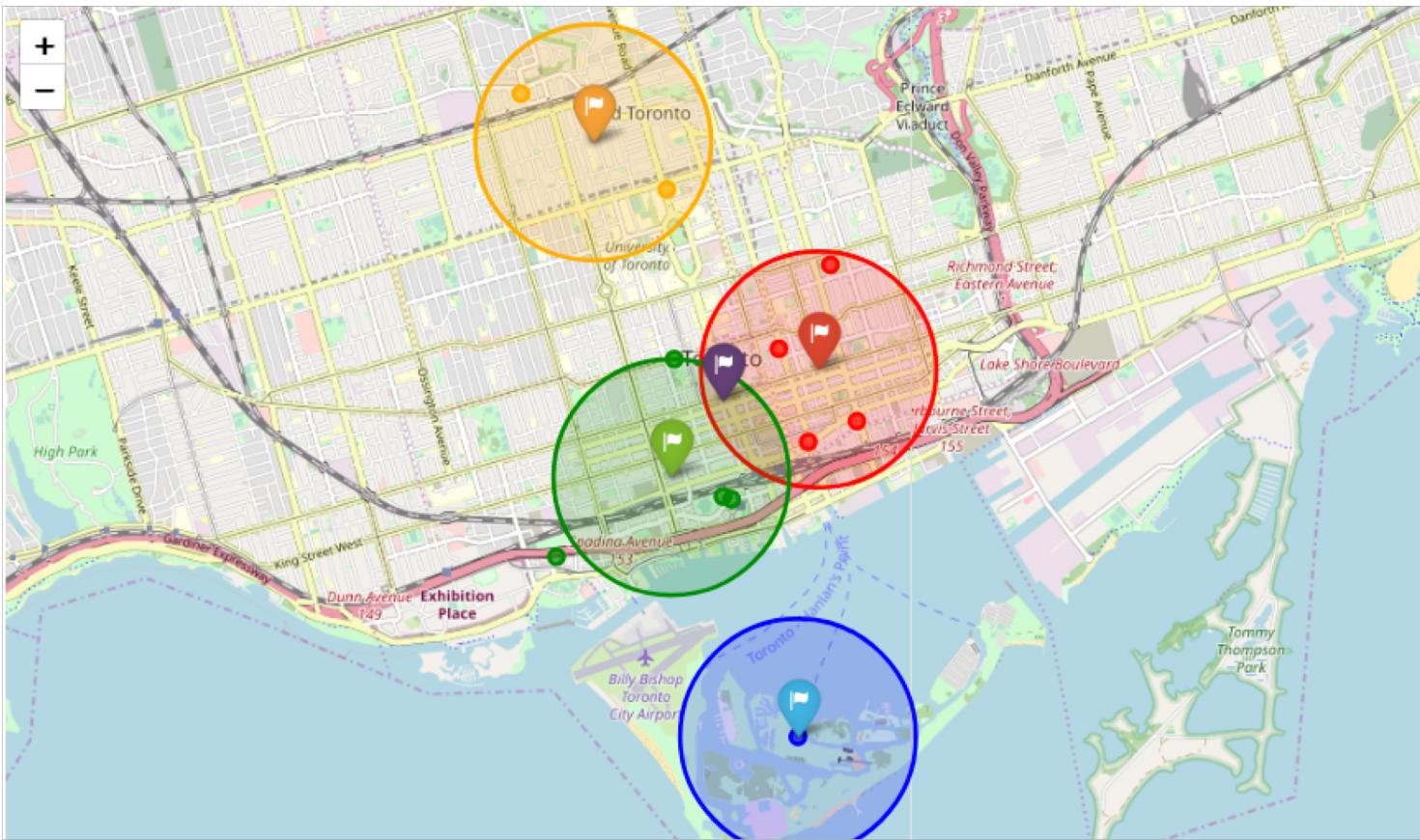
METHODOLOGY – LOCATION & CLUSTERING

Using Foursquare APIs we can retrieve the coordinate of the locations extracted in the previous point.

Then I will apply k-means for clustering.
Initial clustering groups k=7.

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

The city centre is the biggest group that needs to be divided further more.

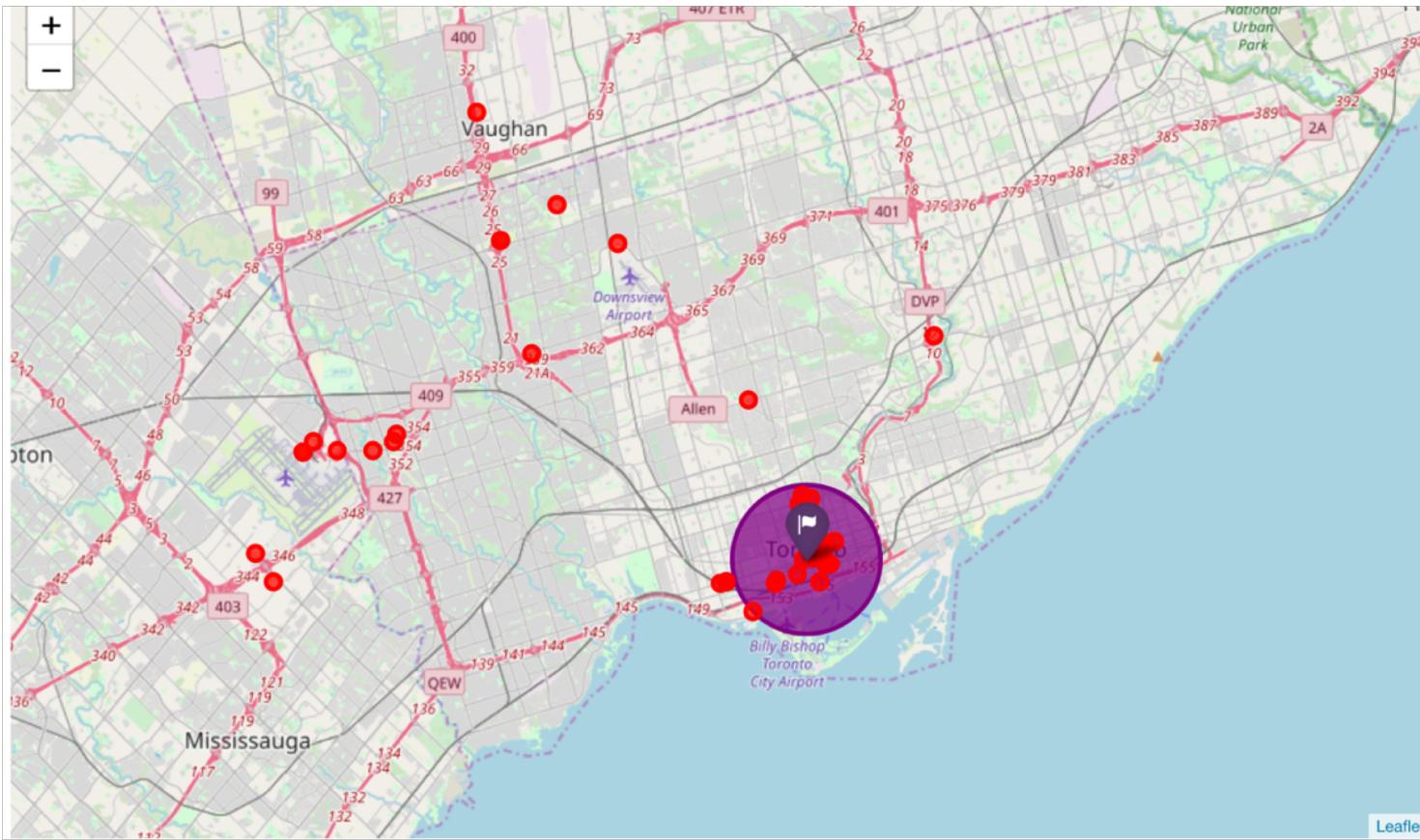


METHODOLOGY – CLUSTERING THE CITY CENTRE

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

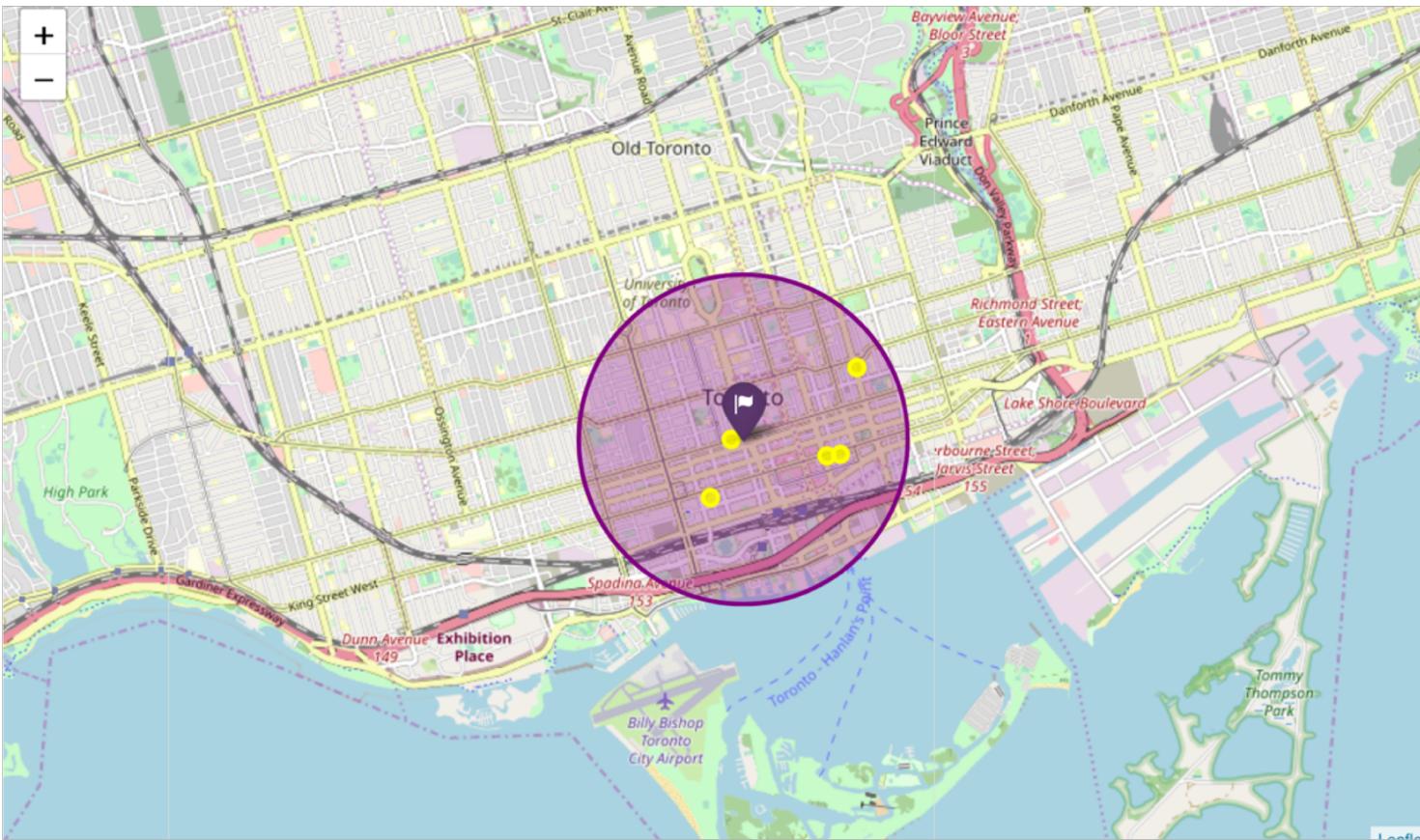


METHODOLOGY – FIND HOTELS

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

And Collect the following information on the hotels: name, address, description, ratings, photos prefix, suffix, width and height.

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



METHODOLOGY – TOP 5 HOTELS

The strategy is 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.

I am using the **Google Maps - Distance Matrix API** to calculate the travel time.

PICK THE RIGHT HOTEL

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, availability, checking more reviews from the customers etc.).

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



DISCUSSION

1. 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 not an easy task.
2. 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 different real scenario, this might not be the case.
3. I could have used and experimented with different clustering techniques. For this 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, 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.