Applied Data Science Capstone Project

- *Coursera*

Finding the perfect neighborhood for your family trip to Canada

Overview

Proposal:

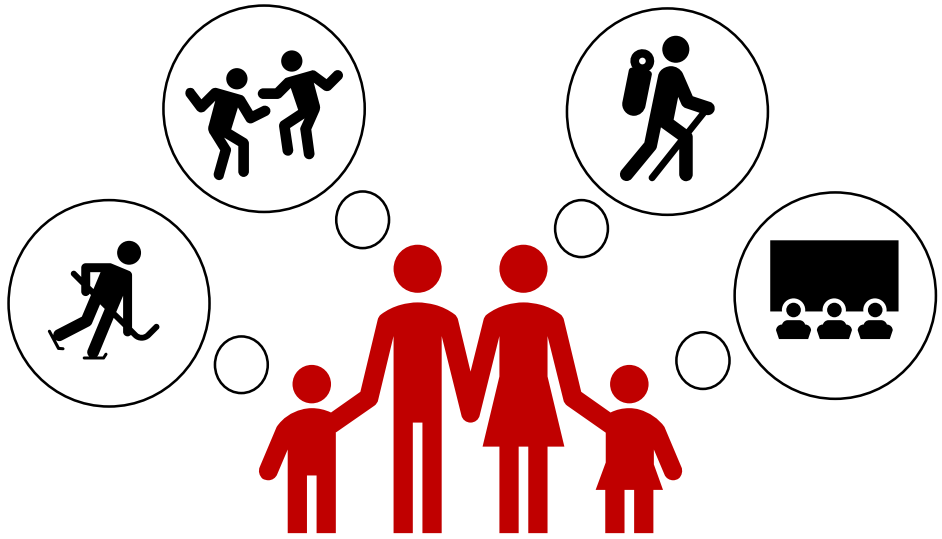
- Let‘s go to Canada for a few days to relax!

Problem:

- Where exactly should we go so everyone is happy?

Solution:

- Location ranking!



Steps

- Retrieving data on canadian neighborhoods
- Cleaning and preparing the data
- Visualizing on a Map
- Utilizing Foursquare venues
- Ranking the neighborhoods
- Finding the neighborhood that is just right

Data collection

- Wikipedia – scraping postal codes and neighborhood names[1)]
- Geonames – latitude and longitude by postal code
- Foursquare – venues and their location

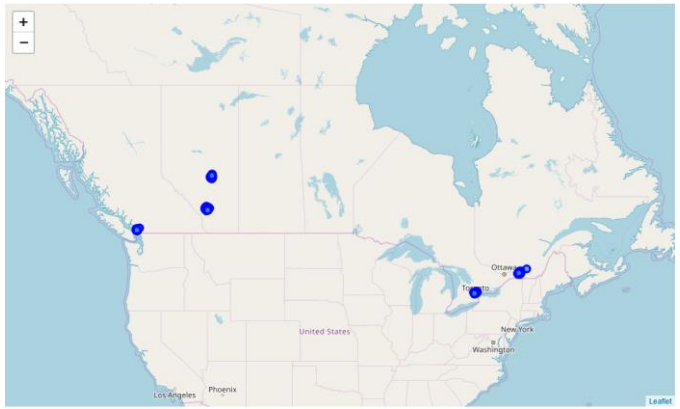
```
# Toronto postal codes and neighborhoods
source = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M').text
soup = BeautifulSoup(source, 'html5lib')

columns = ['PostalCode', 'Borough', 'Neighborhood']
toronto_data = pd.DataFrame(columns=columns)

for table_cell in soup.find_all('tr'):

    y = table_cell.text.replace('\n','').split(',')

    try:
        if len(y[1]) == 3:
            toronto_data.loc[soup.find_all('tr').index(table_cell)] = y[1:4]

        else:
            pass

    except:
        pass
```

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights
4	M7A	Queen's Park	Queen's Park

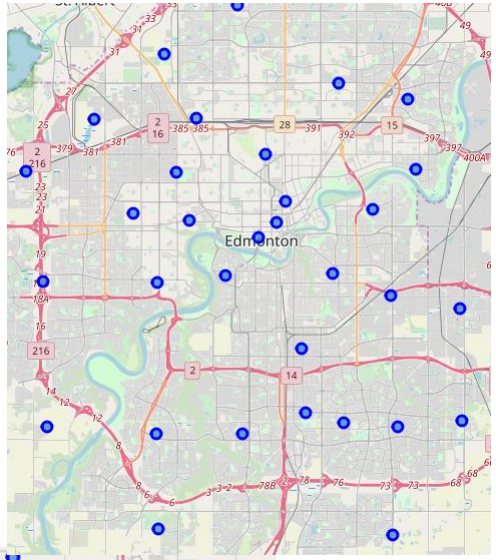
	PostalCode	Borough	Neighborhood	City	latitude	longitude
0	M3A	North York	Parkwoods	Toronto	43.7545	-79.3300
1	M4A	North York	Victoria Village	Toronto	43.7276	-79.3148
2	M5A	Downtown Toronto	Harbourfront	Toronto	43.6555	-79.3626
3	M6A	North York	Lawrence Heights	Toronto	43.7223	-79.4504
4	M7A	Queen's Park	Queen's Park	Toronto	43.6641	-79.3889

1) Focusing on Toronto, Montreal, Vancouver, Edmonton and Calgary

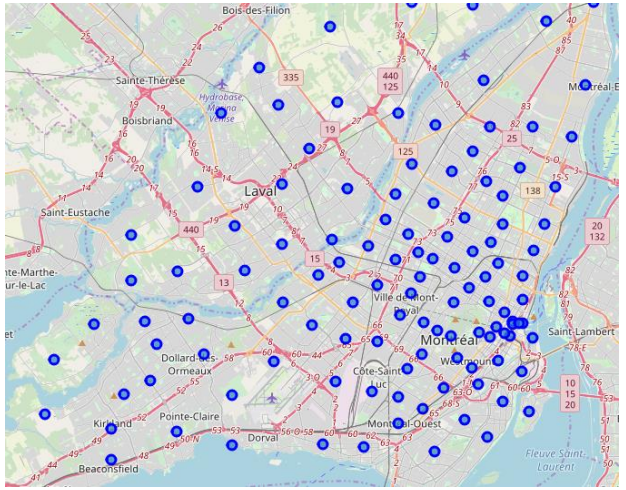
Neighborhoods on a map



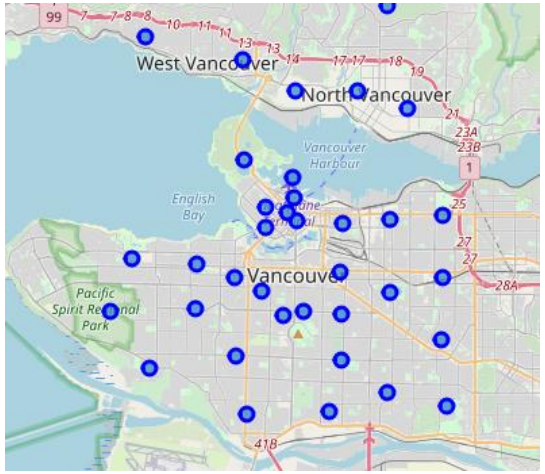
Canada



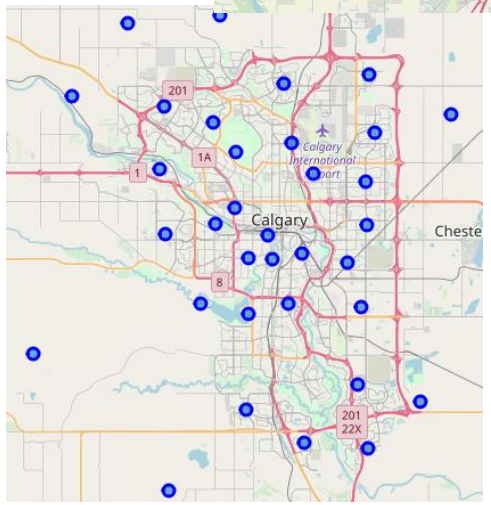
Edmonton



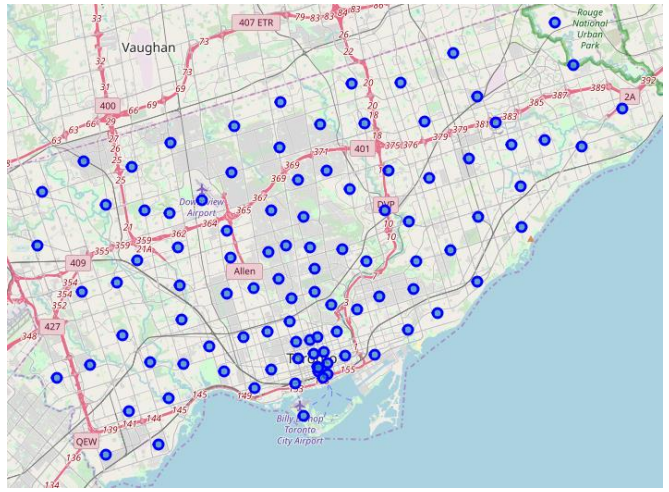
Montreal



Vancouver



Calgary



Toronto

Foursquare venues

- Top 100 venues within a 750m radius
- One-hot-encoding to find venue categories by neighborhood

```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    radius,
    LIMIT)
```

Venue	Venue Latitude	Venue Longitude	Venue Category
Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
Brookbanks Park	43.751976	-79.332140	Park
Tim Hortons	43.760668	-79.326368	Café
A&W Canada	43.760643	-79.326865	Fast Food Restaurant
Pizza Pizza	43.760231	-79.325666	Pizza Place

FOURSQUARE

Ranking the neighborhoods

	Priority-Parents	Mom	Dad	Priority-Kids	Son	Daughter
0	40	Yoga Studio	Golf Course	50	Hockey Arena	Recreation Center
1	30	Café	Bagel Shop	40	Nightclub	Accessories Store
2	20	Museum	Karaoke Bar	30	Bar	Movie Theater
3	10	Jewelry Store	Wine Bar	20	Comedy Club	Shopping Mall
4	5	Shoe Store	Steakhouse	10	Pizza Place	Gym

- 'Stereotypical' preferences
- Score neighborhoods[2)]
- Find highest score

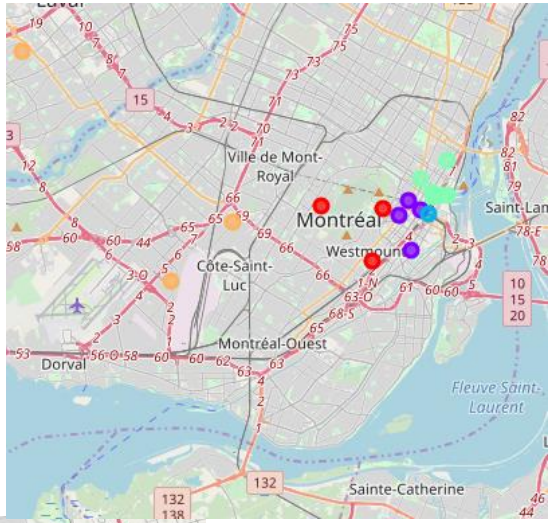
Neighbourhood	Mom pts	Dad pts	Son pts	Daughter pts	Family pts	Minimum indiv pts
North Downtown	50.0	5.0	130.0	60.0	245.0	5.0
Downtown East	90.0	5.0	60.0	60.0	215.0	5.0
SW Downtown	70.0	30.0	50.0	40.0	190.0	30.0
Queen's Park	100.0	0.0	80.0	10.0	190.0	0.0

2) Only includes neighborhoods with a venue of category hotel / hostel / inn. Occurances of a venue category within a neighborhood are scored only once per family member.

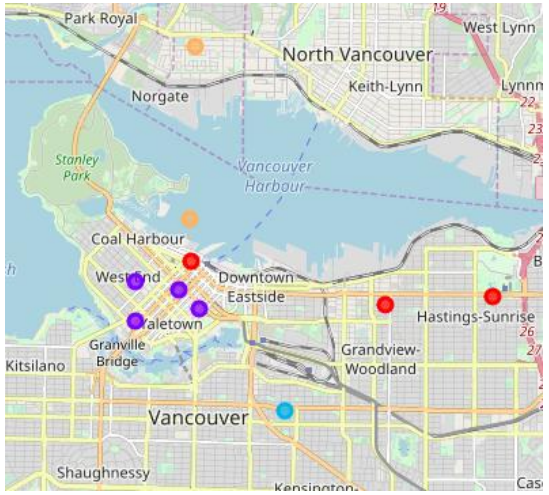
Cluster the neighborhoods

- K-Clusters[3]: 5
  - Highest score
  - Above average score
  - Average score
  - Below average score
  - Lowest score

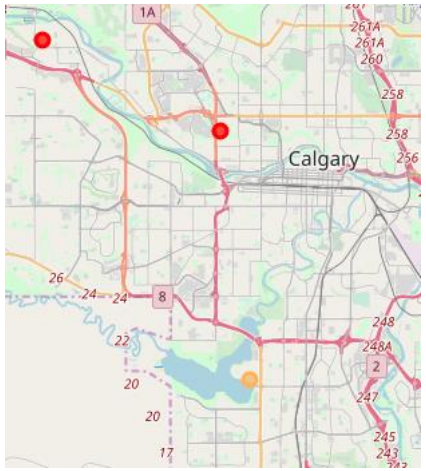
Edmonton



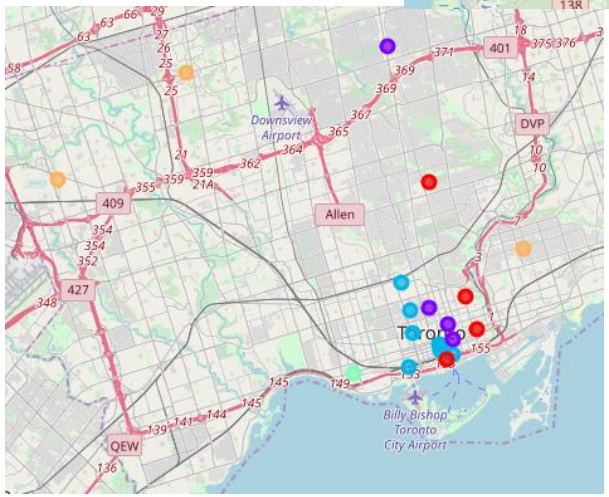
Montreal



Vancouver



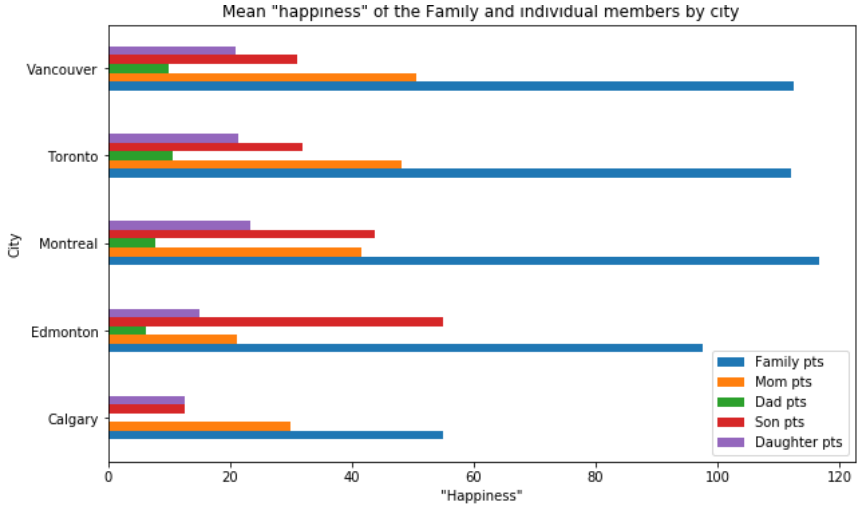
Calgary

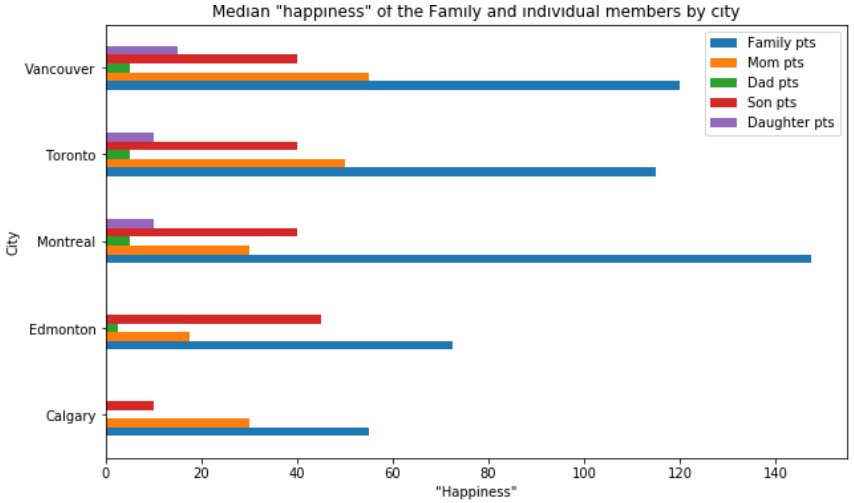


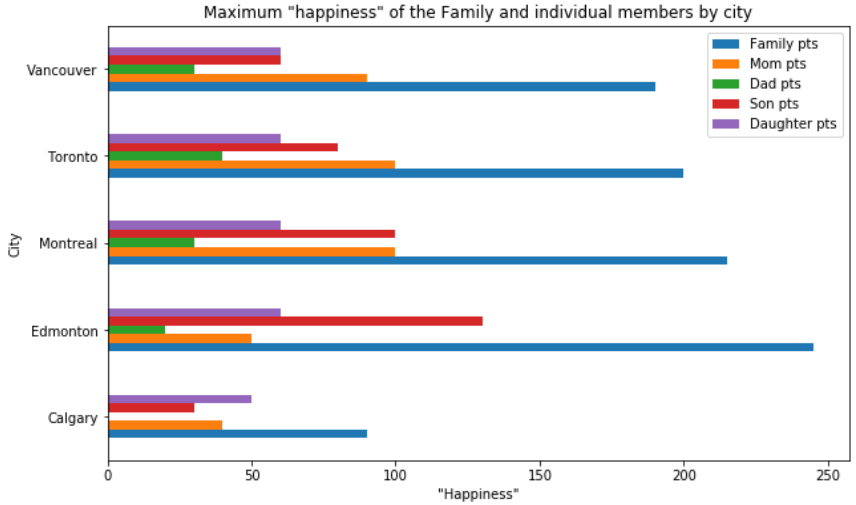
Toronto

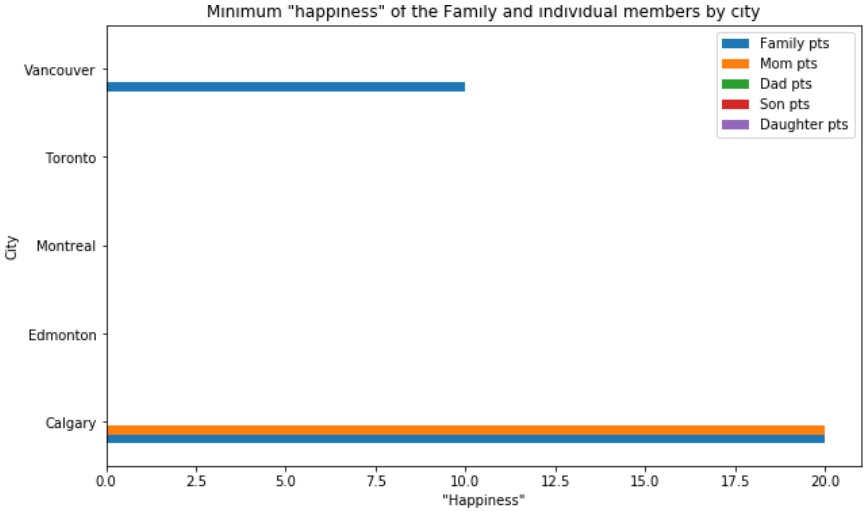
3) Clustering based on individual family members‘ as well as combined scores.

Happiness visualization by city









Conclusion

The optimal neighborhood to find a place to stay, for the hypothetical family, is North Downtown, in the city Edmonton.
For a maximized minimum of preferences met, the family should find a place to stay in SW Downtown, in the city Vancouver.
With regards to the optimal city, picking a random neighborhood, the highest family median 'happiness', defined as preferences met, is obtained in Montreal.