CapStone Data Science Project - Travel Explorer

1. Introduction & Business Problem

Travel to various locations can be for pleasure, learning or both. Often, identifying the destination to visit with points-of-interest of one's choice can be a challenge & time-consuming. Especially so, if locations are not popular, well-known or visited often. These locations could be local, with less facilities/conveniences, nevertheless most suitable for the person looking for locations for a specific activity or a purpose.

It would be helpful to narrow down areas, places, points-of-interest to customer specifications, given minimal information such as a town, city or a state.

2. Audience for the results of the project

Businesses like travel agencies, private tours & non-business related entities such as nature study groups, photography clubs, schools or individuals looking to explore various destinations for pleasure, hobby, photography, outdoor activities, nature study, e.t.c would leverage the results of this project. The output of this project can also be customized for various other purposes to cater to customer's needs.

3. Data

The basic data needed to address the business problem contains names of places to explore, zip codes/addresses, latitude & longitude.

For this project, assumptions were made that the data will pertain to a specific city of Toronto, province of Ontario, in the country of Canada.

The basic data is available on 2 different web-sites, and will have to be combined :

PostalCodesOfCanada

```
- to retrieve names of places and its postal codes
```

Latitude/Longitude

```
- to retrieve latitude and longitude for each postal code
```

- Additional data below will be retrieved from Foursquare, a location data provider :

- Category of venue (place known for) Nature, Music, Museum, Buildings, Parks, Stadiums, e.t.c
- Optional details such as whether it is open to public, Reviews, Ratings, e.t.c

3.1 Preparing Data

The basic data is downloaded, cleaned & merged to create a data set with place, zip code, latitude & longitude.

The latitude / longitude of a location feeds into Foursquare API.

The API will output a list of venues/categories for each location, along with details.

This data is then analyzed further to determine how to solve the problem.

4. Methodology

Since clients/stakeholders will provide a rough location on the places they are looking to travel to, an assumption is made for this project, that data will pertain to city of Toronto, province of Ontario, in the country of Canada.

The methods outlined below can be repeated for any location of the customer's choice and their needs.

Note :

The latitude/longitude values for places outside of Canada, can be obtained from OpenCage Geocoder website, which takes postal codes as the input. For U.S.A., the postal codes can be extracted from the web-site. https://www.unitedstateszipcodes.org

The first step was to retrieve list of places specific to the city of Toronto, Canada, and retrieve the corresponding latitude/longitude from 2 different websites.

The 2 data sets were merged, missing data removed, producing a list of postal codes, borough, associated neighborhoods, and their corresponding latitude and longitude.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636

For this assignment, we pick up Boroughs containing the string 'Toronto'. Since we want to start with client's specific request around specific boroughs.

Depending upon our exploratory analysis, we would determine if we want to expand our search to other boroughs of Toronto. Below is a table of count of Neighborhoods for each Borough

	Borough	Neighborhood
0	Central Toronto	9
1	Downtown Toronto	19
2	East Toronto	5
3	East York	5
4	Etobicoke	12
5	Mississauga	1
6	North York	24
7	Scarborough	17
8	West Toronto	6
9	York	5

I used the **FourSquare API** location provider to explore Neighboorhoods for Boroughs containing the string 'Toronto'.

A limit in the result of **200 venues**, and a search **radius of 30 miles** of each neighborhood was used to search for venues.
My initial search is to see what category of venues will be returned.

	Venue Category	Count
0	Coffee Shop	143
1	Café	90
2	Restaurant	57
3	Italian Restaurant	41
4	Park	36
5	Hotel	32
6	Bakery	32
7	Japanese Restaurant	31
8	Pub	26
9	Bar	26

The results showed majority of the venues associated with Food, very minimal venues were meant any out-door activity.

Using FourSquare Documents on their web-site, I extracted specific categories related 'Out-doors'

I then refined my search for venues using Foursquare once more, to only include categories below, and aggregated the results

```
- Lake , Park , Outdoors & Recreation , Scenic , Beach
- Mountain , Botanical , River , Nature , Other Great Outdoors , Forest
```

	Neighborhood	Venue Category	Count
0	Berczy Park	Beach	1
1	Berczy Park	Park	1
2	Business reply mail Processing Centre, South C...	Park	1
3	Business reply mail Processing Centre, South C...	Skate Park	1
4	Central Bay Street	Park	1
5	Christie	Park	2
6	Church and Wellesley	Park	1
7	Commerce Court, Victoria Hotel	Park	1
8	Davisville	Park	1
9	Davisville North	Park	1

5. Results

Aggregating the result further, to see the breakdown of out-door categories

	Venue Category	Count
0	Park	27
1	Lake	2
2	Beach	1
3	Other Great Outdoors	1
4	Scenic Lookout	1
5	Skate Park	1

To visualize the distribution of the neighborhoods from the result set, I used Follium to map the out-door locations super-imposed on a map of the Toronto City.

6. Discussion

An analysis of category types for the neighborhood of Toronto revealed majority of **venues were related to food businesses** such as restaurants, bakery, café, e.t.c

When data was further filtered, it was observed very few neighborhoods of Toronto had venues meant for 'Outdoor' Activities.

Of these few venues, there were mostly of type 'Park', which seemed to be a playground rather than place to travel to.

If the analysis had revealed a good number of venue categories related to 'Outdoor' Activities, then a FourSquare would be used once more to drill down to retrieve addition details such as tips, reviews, photographs, e.t.c

7. Conclusion

The final analysis concludes 'Toronto' neighborhoods are preferable for a city life, businesses with **limited choice** for people that are interested in nature related **out-door** activities.

Customers with interests in photography, nature, water related sports, wildlife would prefer non-city locations.

One recommendation would be to broaden the search to other Boroughs of Toronto, or to other cities in Canada.