Coursera Capstone Final

Business Problem

To divide cities and towns of Canada into groups according to infrastructural facilities such as medical, educational, transport, business, etc.

Stakeholders

This project may be interesting to investors or companies. That they can estimate benefits of different cities/towns in Canada and choose one where they can open new offices/branches.

Data

List of data sources used in this project:

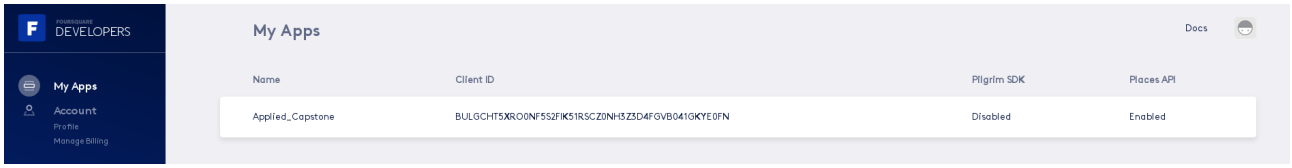
1 List of cities/towns from:
https://simplemaps.com/static/data/world-cities/basic/simplemaps_worldcities_basicv1.5.zip,
where data looks like:

	city	city_ascii	lat	lng	country	iso2	iso3	admin_name	capital	population	id
0	Selkirk	Selkirk	50.1500	-96.8833	Canada	CA	CAN	Manitoba	NaN	9986.0	1124499880
1	Trepassey	Trepassey	46.7370	-53.3633	Canada	CA	CAN	Newfoundland and Labrador	NaN	398.0	1124344083
2	Schefferville	Schefferville	54.8000	-66.8167	Canada	CA	CAN	Québec	NaN	471.0	1124403382
3	Whitehorse	Whitehorse	60.7167	-135.0500	Canada	CA	CAN	Yukon	admin	23276.0	1124348186
4	Trout River	Trout River	49.4837	-58.1166	Canada	CA	CAN	Newfoundland and Labrador	NaN	452.0	1124733447
5	Yorkton	Yorkton	51.2171	-102.4665	Canada	CA	CAN	Saskatchewan	NaN	15172.0	1124108820
6	Antigonish	Antigonish	45.6269	-61.9982	Canada	CA	CAN	Nova Scotia	NaN	6739.0	1124839247
7	Gander	Gander	48.9500	-54.5500	Canada	CA	CAN	Newfoundland and Labrador	NaN	3345.0	1124310517
8	Berens River	Berens River	52.3666	-97.0333	Canada	CA	CAN	Manitoba	NaN	892.0	1124806860
9	Port-Menier	Port-Menier	49.8226	-64.3480	Canada	CA	CAN	Québec	NaN	263.0	1124109240

Picture 1 — Dataset with cities of Canada

The dataset includes 250 cities of Canada with next fields:
- city name;
- latitude;
- longitude;
- state;
- population.

2 Foursquare API for extracting facilities information within 5 km radius by cities.



Picture 2 — Created Application in Foursquare API

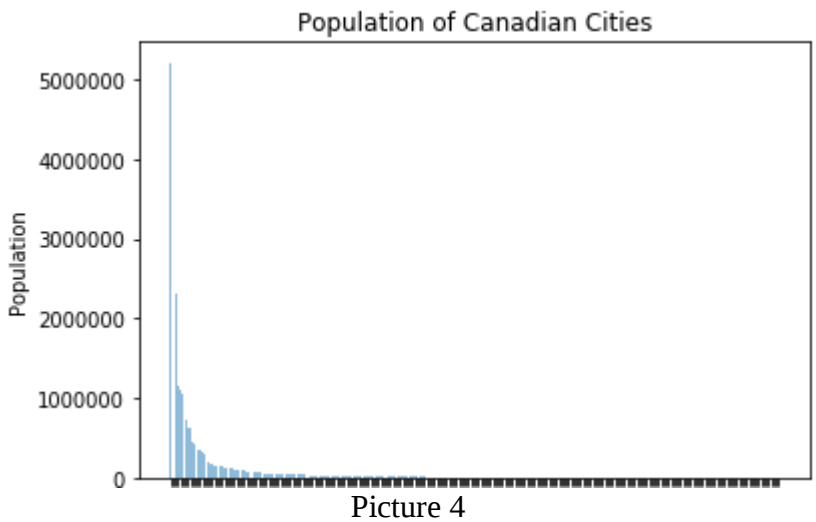
	city	medical	education	food	shops	transport
0	Selkirk	11	0	4	6	1
1	Trepassey	1	0	0	0	0
2	Schefferville	0	0	0	0	0
3	Whitehorse	8	3	5	9	5
4	Trout River	0	0	0	0	0
5	Yorkton	5	1	4	5	1
6	Antigonish	2	1	0	7	0
7	Gander	2	0	0	3	0
8	Berens River	0	0	0	0	0
9	Port-Menier	0	0	0	0	0
10	Wetaskiwin	2	0	1	5	1

Picture 3 — Facitily Information gathered using Foursquare API

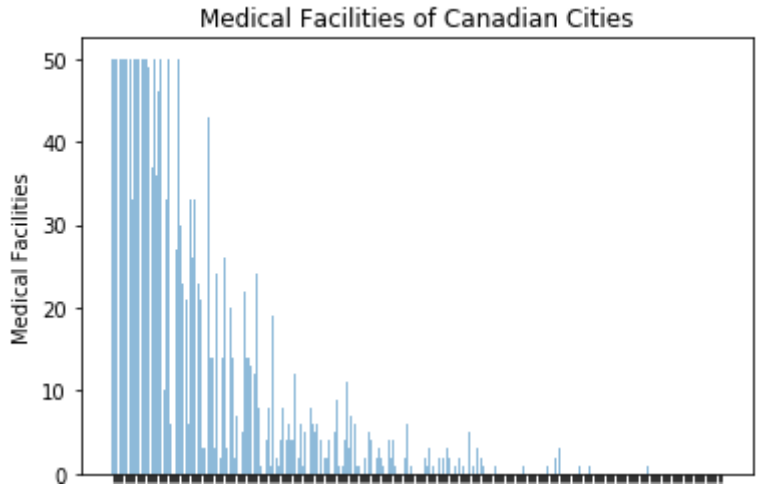
Methodology

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

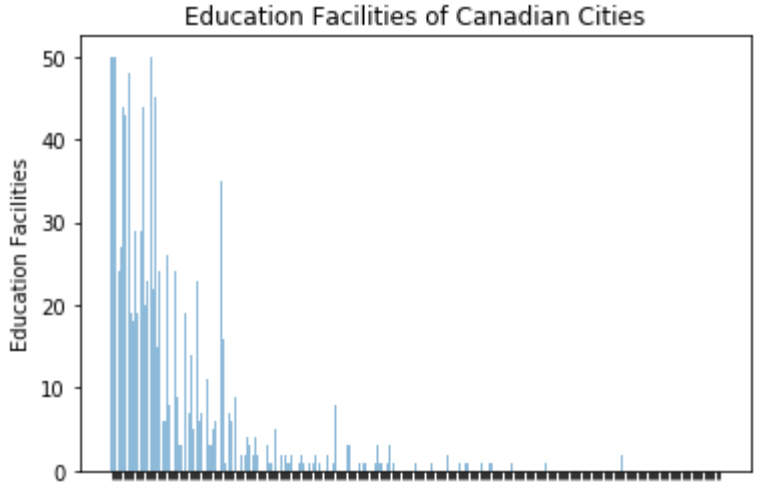
Data Visualization using Bar Graphs:



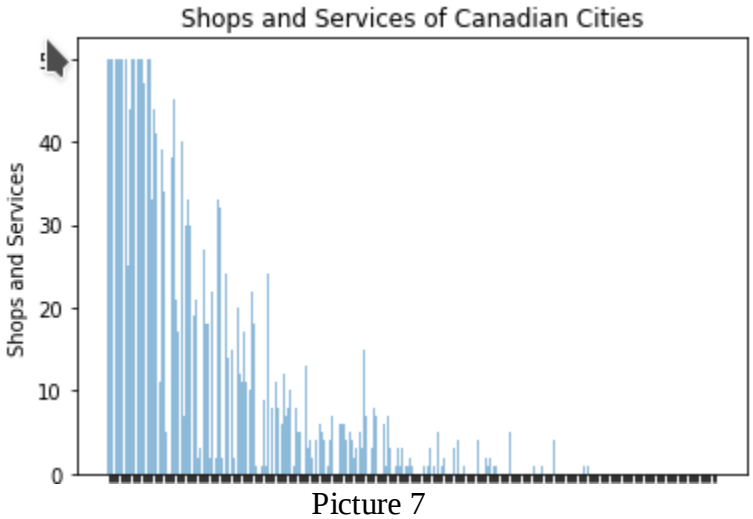
Picture 4



Picture 5



Picture 6



Picture 7

Based on these graphs number of groups to divide cities is chosen to be 4 in K-Means Clusterization.

Picture 8 — Folium Map with Canadian Cities

All gathered data will be used in analyzing cities using Kmeans algorithm and divided by 4 Groups.

Results

```
k_means = KMeans(init="k-means++", n_clusters=4, n_init=12)
k_means.fit(X)
k_means_labels = k_means.labels_
k_means_labels
k_means_cluster_centers = k_means.cluster_centers_
k_means_cluster_centers
```

```
array([[ 1.18324607,  0.41884817,  0.73298429,  1.46073298,  0.30366492],
       [36.11111111, 18.55555556, 15.94444444, 36.        ,  5.77777778],
       [49.13333333, 36.6       , 48.33333333, 50.        , 25.66666667],
       [14.57692308,  4.15384615,  7.46153846, 15.80769231,  2.88461538]])
```

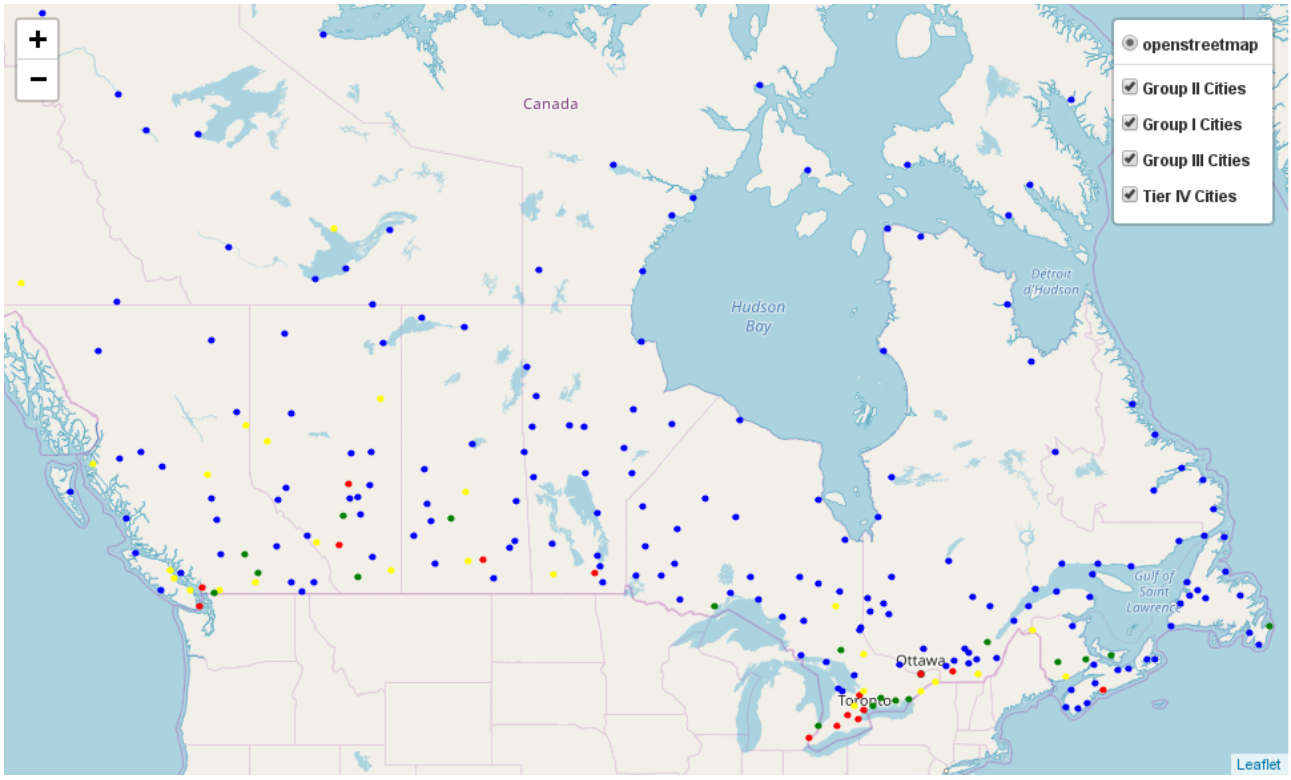
```
k_means_labels
```

```
array([2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 1, 1, 1, 3,
       1, 1, 0, 1, 1, 1, 3, 3, 1, 0, 1, 1, 1, 3, 3, 0, 0, 1, 3, 3, 0, 3,
       0, 1, 1, 0, 3, 3, 0, 3, 0, 3, 3, 3, 3, 3, 3, 3, 3, 0, 0, 0, 3, 0,
       3, 0, 0, 0, 0, 0, 3, 0, 0, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

Picture 9 — K-Means centres and centres labels for the 4 clusters

	city	lat	lng	admin_name	population	medical	education	food	shops	transport	tier
0	Toronto	43.7000	-79.4200	Ontario	5213000.0	50.0	50.0	50.0	50.0	50.0	2
1	Montréal	45.5000	-73.5833	Québec	3678000.0	50.0	50.0	50.0	50.0	25.0	2
2	Vancouver	49.2734	-123.1216	British Columbia	2313328.0	50.0	50.0	50.0	50.0	50.0	2
3	Ottawa	45.4167	-75.7000	Ontario	1145000.0	50.0	24.0	50.0	50.0	31.0	2
4	Calgary	51.0830	-114.0800	Alberta	1110000.0	50.0	27.0	50.0	50.0	37.0	2
5	Edmonton	53.5500	-113.5000	Alberta	1058000.0	50.0	44.0	50.0	50.0	37.0	2
6	Hamilton	43.2500	-79.8300	Ontario	721053.0	50.0	43.0	50.0	50.0	12.0	2
7	Winnipeg	49.8830	-97.1660	Manitoba	632063.0	50.0	48.0	50.0	50.0	30.0	2
8	Québec	46.8400	-71.2456	Québec	624177.0	33.0	19.0	4.0	25.0	0.0	1
9	Oshawa	43.8800	-78.8500	Ontario	450963.0	50.0	18.0	28.0	44.0	10.0	1
10	Kitchener	43.4500	-80.5000	Ontario	417001.0	50.0	29.0	50.0	50.0	17.0	2
11	Halifax	44.6500	-63.6000	Nova Scotia	359111.0	50.0	19.0	50.0	50.0	20.0	2
12	London	42.9700	-81.2500	Ontario	346765.0	50.0	29.0	50.0	50.0	14.0	2
13	Windsor	42.3333	-83.0333	Ontario	319246.0	50.0	44.0	50.0	50.0	18.0	2
14	Victoria	48.4333	-123.3500	British Columbia	289625.0	50.0	20.0	50.0	50.0	18.0	2
15	Saskatoon	52.1700	-106.6700	Saskatchewan	198958.0	49.0	23.0	25.0	47.0	13.0	1
16	Barrie	44.3838	-79.7000	Ontario	182041.0	37.0	50.0	25.0	50.0	11.0	2
17	Regina	50.4500	-104.6170	Saskatchewan	176183.0	50.0	22.0	50.0	50.0	15.0	2
18	Sudbury	46.5000	-80.9666	Ontario	157857.0	36.0	45.0	14.0	33.0	3.0	1
19	Abbotsford	49.0504	-122.3000	British Columbia	151683.0	46.0	15.0	21.0	44.0	7.0	1
20	Sarnia	42.9666	-82.4000	Ontario	144172.0	50.0	24.0	11.0	41.0	1.0	1
21	Sherbrooke	45.4000	-71.9000	Québec	139652.0	10.0	6.0	2.0	11.0	1.0	3
22	St. John's	47.5850	-52.6810	Newfoundland and Labrador	131469.0	33.0	6.0	8.0	39.0	7.0	1
23	Kelowna	49.9000	-119.4833	British Columbia	125109.0	50.0	26.0	22.0	34.0	7.0	1

Picture 10 — Canadian Cities and predicted labels by K-means



Picture 11 — Labeled by Groups map of Canadian Cities

Discussion

Here is the average vlues for each group:

	Tier	Avg_Population	Avg_Medical	Avg_Education	Avg_Food	Avg_Shops	Avg_Transport
0	Tier 1	1197361	49	36	48	50	25
1	Tier 2	151394	36	18	15	36	5
2	Tier 3	41518	14	4	7	15	2
3	Tier 4	7204	1	0	0	1	0

Inferences:
1. Group I and II cities have high population due to higher facilities
2 Group III and IV cites have lower number of them.
3 For interested ones cities of Group III and IV are looking good due to their lower rate of development and well opportunity to invest there.
4 But the transport situation must be considered in Group IV cities.

Conclusion

Here you can see the implementation of K-Means Clusters algorithm on Canadian Cities on facility availability using Foursquare API.