Find the optimal location to start a restaurant in Toronto

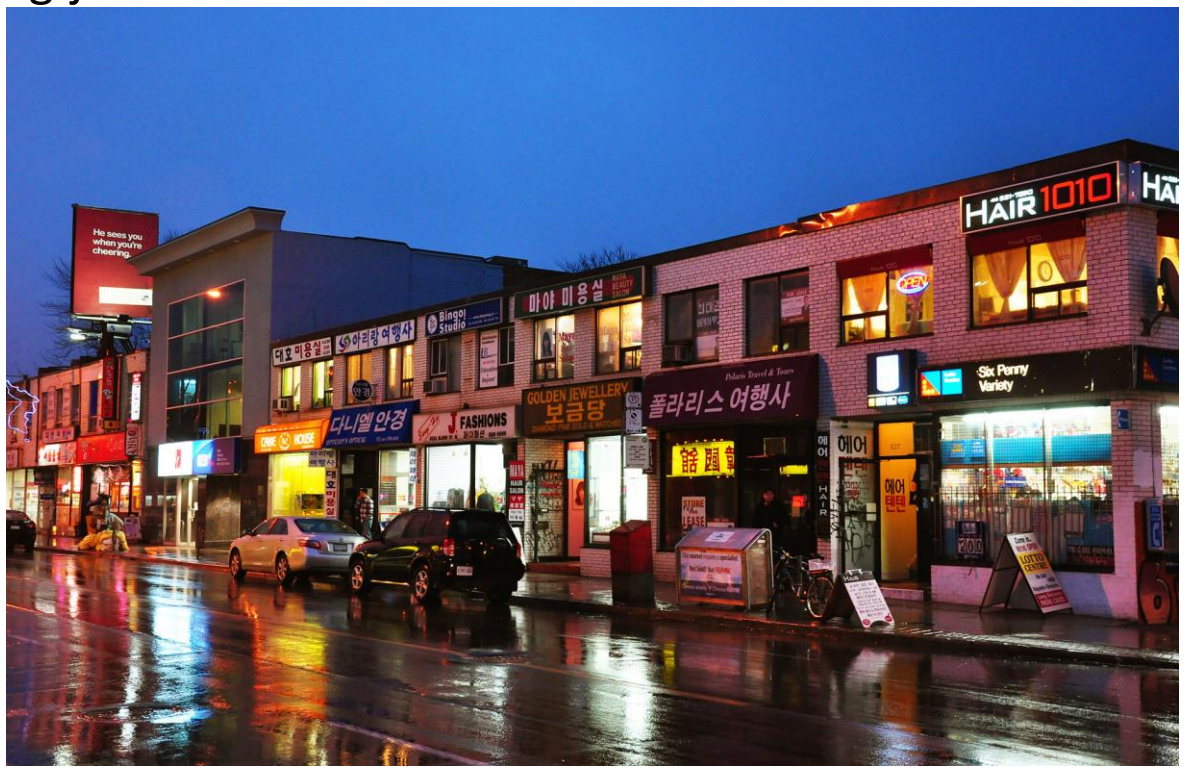
Coursera Capstone Project

By Ye Tian

08/30/209

Introduction

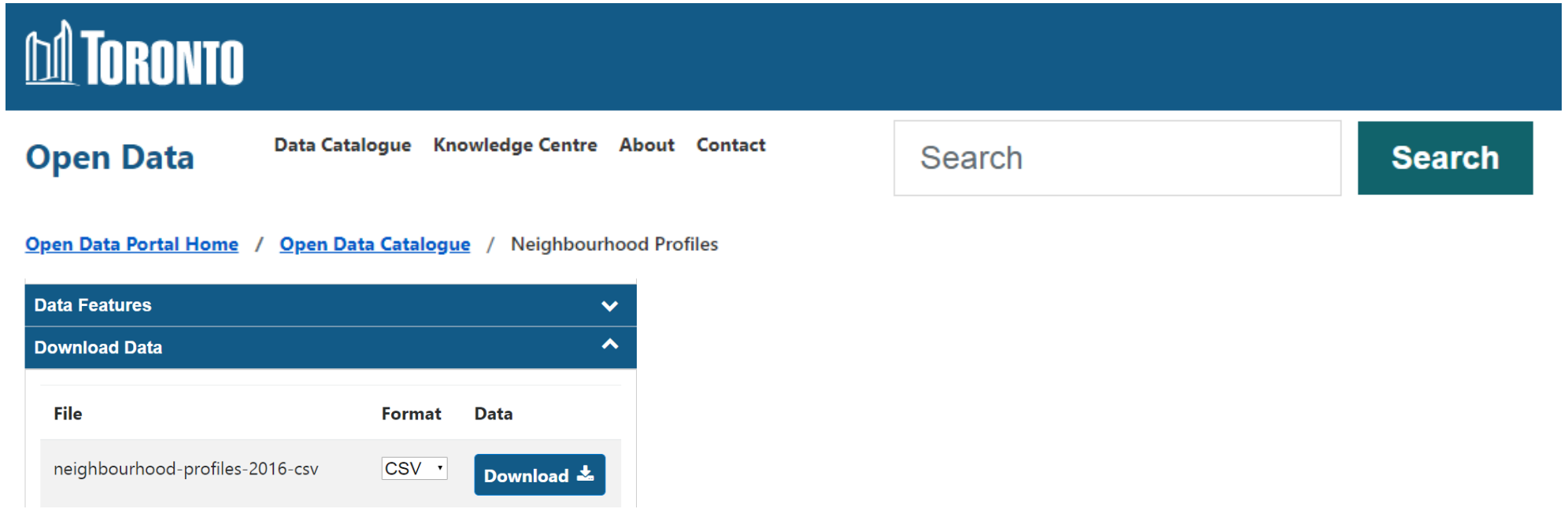
- Dong-jun is a new immigrant from Korea to the Toronto city, and he would like to start a Korean restaurant.
- According to South Korea's Ministry of Foreign Affairs and Trade, there were 240,942 ethnic Koreans or people of Korean descent in Canada as of 2017, around 60,000 of them live in Toronto.
- We would like to find the optimal location for Dong-jun



Wikipedia: Korean businesses and restaurants in Toronto

Data collection

- First, we need to obtain the neighborhood and its geospatial data of Toronto from a publicly available database e.g., https://open.toronto.ca/dataset/neighbourhood-profiles/. We download the CSV file and uploaded to IBM cloud.



Data processing

- We would like to know our potential customers for Dongjun's business. Hence, we would focus on residents with Korean origin (who can speak Korean, or who characterized themselves of Korean ancestry) in a neighborhood.
- Average individual income is also important.

	Neighbourhood	Total Population	Mother tongue Korean	Speaking Korean at home	Knowledge of Korean	Visible minority population	Ethnic origin as Korean	Using Korean at work	Average individual income
0	Agincourt North	29113	50	35	60	70	100	0	30414
1	Agincourt South-Malvern West	23757	80	70	105	140	145	0	31825
2	Alderwood	12054	30	25	45	65	75	0	47709
3	Annex	30526	420	235	520	605	635	10	112766
4	Banbury-Don Mills	27695	420	280	465	560	575	0	67757
5	Bathurst Manor	15873	270	200	285	305	305	20	45936
6	Bay Street Corridor	25797	1120	720	1260	1340	1370	60	56526
7	Bayview Village	21396	1395	1090	1625	1655	1685	135	52035
8	Bayview Woods-Steeles	13154	620	470	665	710	730	95	47246
9	Bedford Park-Nortown	23236	145	80	150	195	210	0	123077

Data processing

- We the normalize the index with total population and sort it by descending order
- We found several neighbourhoods with very high index, say 'Newtonbrook East', korean is the 3rd most speaking language, after English and Chinese in Newtonbrook East.

	Neighbourhood	Total Population	Mother tongue Korean	Speaking Korean at home	Knowledge of Korean	Visible minority population	Ethnic origin as Korean	Using Korean at work	Average individual income	Korean index	Normalized index
85	Newtonbrook East	16097	1630	1290	1845	1920	1960	225	45212	49860	3.097472
130	Willowdale West	16936	1745	1295	1810	1930	1955	125	44576	49100	2.899150
129	Willowdale East	50434	4630	3440	4985	5255	5395	470	45326	134620	2.669231
86	Newtonbrook West	23831	1925	1600	2150	2210	2275	250	34904	58740	2.464857
7	Bayview Village	21396	1395	1090	1625	1655	1685	135	52035	41900	1.958310
8	Bayview Woods-Steeles	13154	620	470	665	710	730	95	47246	18660	1.418580
67	Lansing-Westgate	16164	775	570	810	875	910	90	72371	22580	1.396931
6	Bay Street Corridor	25797	1120	720	1260	1340	1370	60	56526	31560	1.223398
111	St.Andrew-Windfields	17812	665	475	760	820	830	45	100516	19480	1.093645
58	Islington-City Centre West	43965	1480	1145	1535	1620	1625	65	52787	41160	0.936199

Nominatim API

- Like other geocoder, Nominatim API can return the coordinate with input name, say Agincourt North, Its return latitude and longitude are 43.808038, -79.266439.

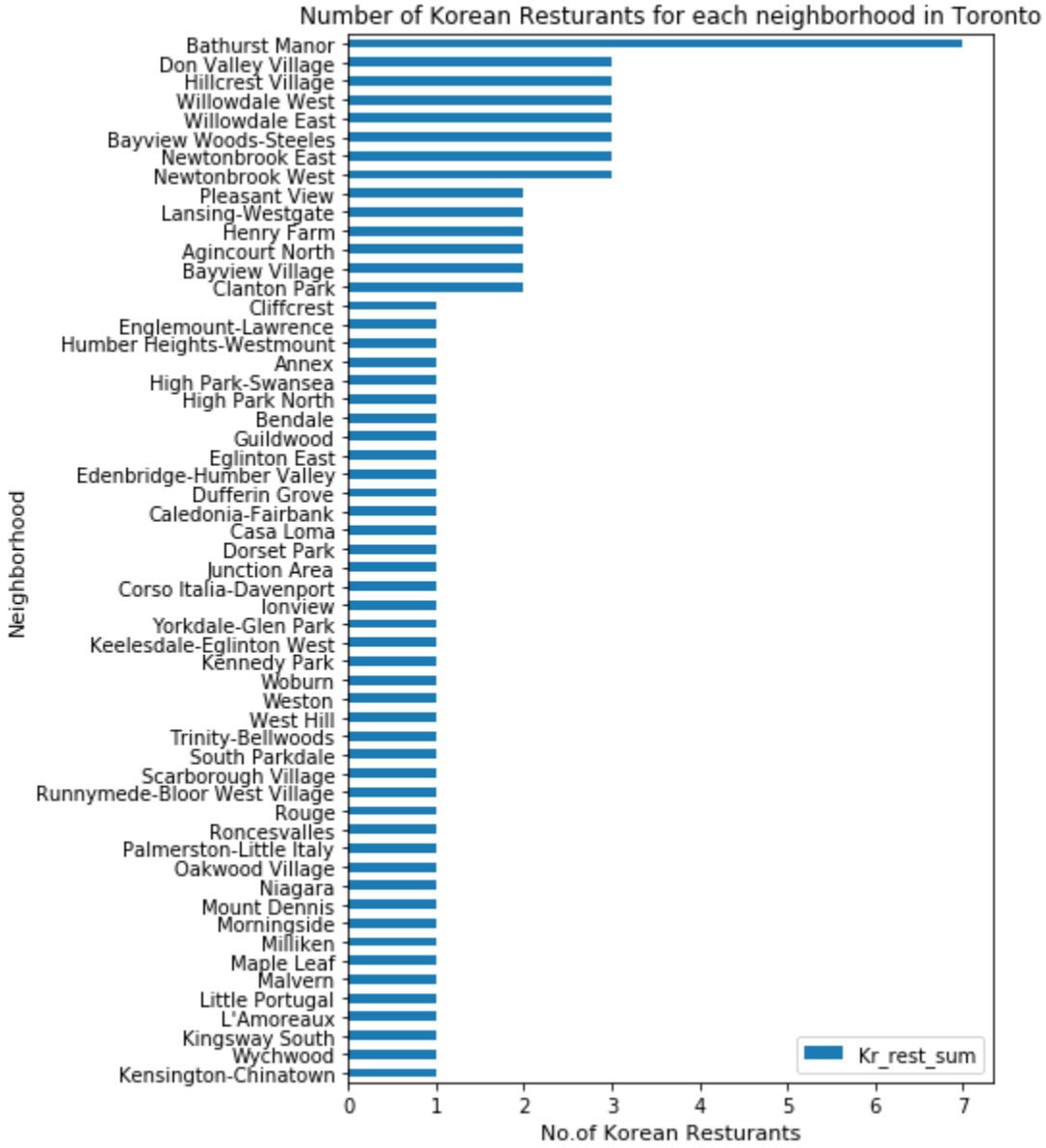
- But as a free API, we only obtain coordinate for 105 out of 140 neighbors in Toronto. (Google keeps returning None in my case.)

	Neighbourhood	Total Population	Mother tongue Korean	Speaking Korean at home	Knowledge of Korean	Visible minority population	Ethnic origin as Korean	Using Korean at work	Average individual income	Korean index	Normalized index	latitude	longitude
0	Agincourt North	29113	50	35	60	70	100	0	30414	1600	0.054958	43.808038	-79.266439
1	Alderwood	12054	30	25	45	65	75	0	47709	1180	0.097893	43.601717	-79.545232
2	Annex	30526	420	235	520	605	635	10	112766	12440	0.407521	43.670338	-79.407117
3	Banbury-Don Mills	27695	420	280	465	560	575	0	67757	12000	0.433291	43.734804	-79.357243
4	Bathurst Manor	15873	270	200	285	305	305	20	45936	7660	0.482580	43.763893	-79.456367

Foursquare API

- We then use the Foursquare API to obtain detailed information on existing Korean restaurants in each neighborhood with known coordinates and find its competitors in 105 neighborhoods.

```
( 1 / 105 ) Korean Restaurant in Agincourt North, :2
( 2 / 105 ) Korean Restaurant in Alderwood, :0
( 3 / 105 ) Korean Restaurant in Annex, :1
( 4 / 105 ) Korean Restaurant in Banbury-Don Mills, :0
( 5 / 105 ) Korean Restaurant in Bathurst Manor, :7
( 6 / 105 ) Korean Restaurant in Bay Street Corridor, :0
( 7 / 105 ) Korean Restaurant in Bayview Village, :2
( 8 / 105 ) Korean Restaurant in Bayview Woods-Steeles, :3
( 9 / 105 ) Korean Restaurant in Bendale, :1
( 10 / 105 ) Korean Restaurant in Black Creek, :0
( 11 / 105 ) Korean Restaurant in Broadview North, :0
( 12 / 105 ) Korean Restaurant in Cabbagetown-South St. James
( 13 / 105 ) Korean Restaurant in Caledonia-Fairbank, :1
( 14 / 105 ) Korean Restaurant in Casa Loma, :1
( 15 / 105 ) Korean Restaurant in Centennial Scarborough, :0
( 16 / 105 ) Korean Restaurant in Church-Yonge Corridor, :0
( 17 / 105 ) Korean Restaurant in Clanton Park, :2
( 18 / 105 ) Korean Restaurant in Cliffcrest, :1
( 19 / 105 ) Korean Restaurant in Corso Italia-Davenport, :1
( 20 / 105 ) Korean Restaurant in Danforth, :0
( 21 / 105 ) Korean Restaurant in Danforth East York, :0
( 22 / 105 ) Korean Restaurant in Don Valley Village, :3
( 23 / 105 ) Korean Restaurant in Dorset Park, :1
( 24 / 105 ) Korean Restaurant in Dufferin Grove, :1
( 25 / 105 ) Korean Restaurant in East End-Danforth, :0
( 26 / 105 ) Korean Restaurant in Edenbridge-Humber Valley, :1
( 27 / 105 ) Korean Restaurant in Eglinton East, :1
( 28 / 105 ) Korean Restaurant in Elms-Old Rexdale, :0
( 29 / 105 ) Korean Restaurant in Englemount-Lawrence, :1
( 30 / 105 ) Korean Restaurant in Etobicoke West Mall, :0
( 31 / 105 ) Korean Restaurant in Flemingdon Park, :0
( 32 / 105 ) Korean Restaurant in Forest Hill North, :0
( 33 / 105 ) Korean Restaurant in Forest Hill South, :0
( 34 / 105 ) Korean Restaurant in Glenfield-Jane Heights, :0
( 35 / 105 ) Korean Restaurant in Guildwood, :1
( 36 / 105 ) Korean Restaurant in Henry Farm, :2
( 37 / 105 ) Korean Restaurant in High Park North, :1
( 38 / 105 ) Korean Restaurant in High Park-Swansea, :1
( 39 / 105 ) Korean Restaurant in Highland Creek, :0
( 40 / 105 ) Korean Restaurant in Hillcrest Village, :3
```



Combine the data

- Lastly, we will combine data from foursquare with demographic data to find a solution to Dongjun’s question.
- We join it with previous demographic dataframe 'toronto_kr’ obtained from City of Toronto by column “Neighbourhood".
- We can then visualize the data.

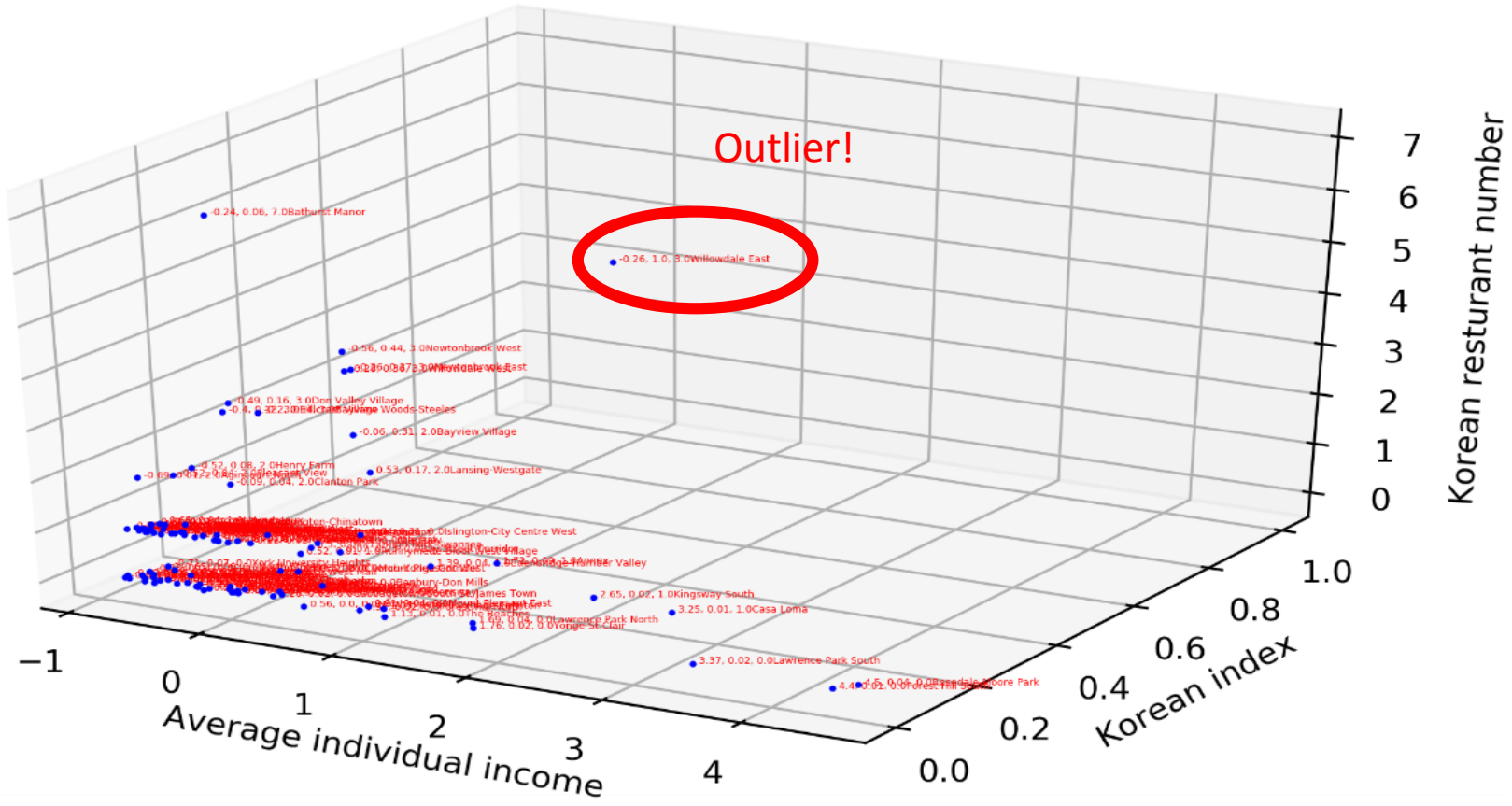
	Neighbourhood	Kr_rest_sum	Korean index	Average individual income
0	Agincourt North	2.0	1600	30414
1	Alderwood	0.0	1180	47709
2	Annex	1.0	12440	112766
3	Banbury-Don Mills	0.0	12000	67757
4	Bathurst Manor	7.0	7660	45936

Data visualization with scatter plot

- Ideally, we want to find a neighborhood with high Korean_index, high income, but low Kr_rest_sum. But also, can accept a neighborhood with high Korean_index, high income, also high Kr_rest_sum.

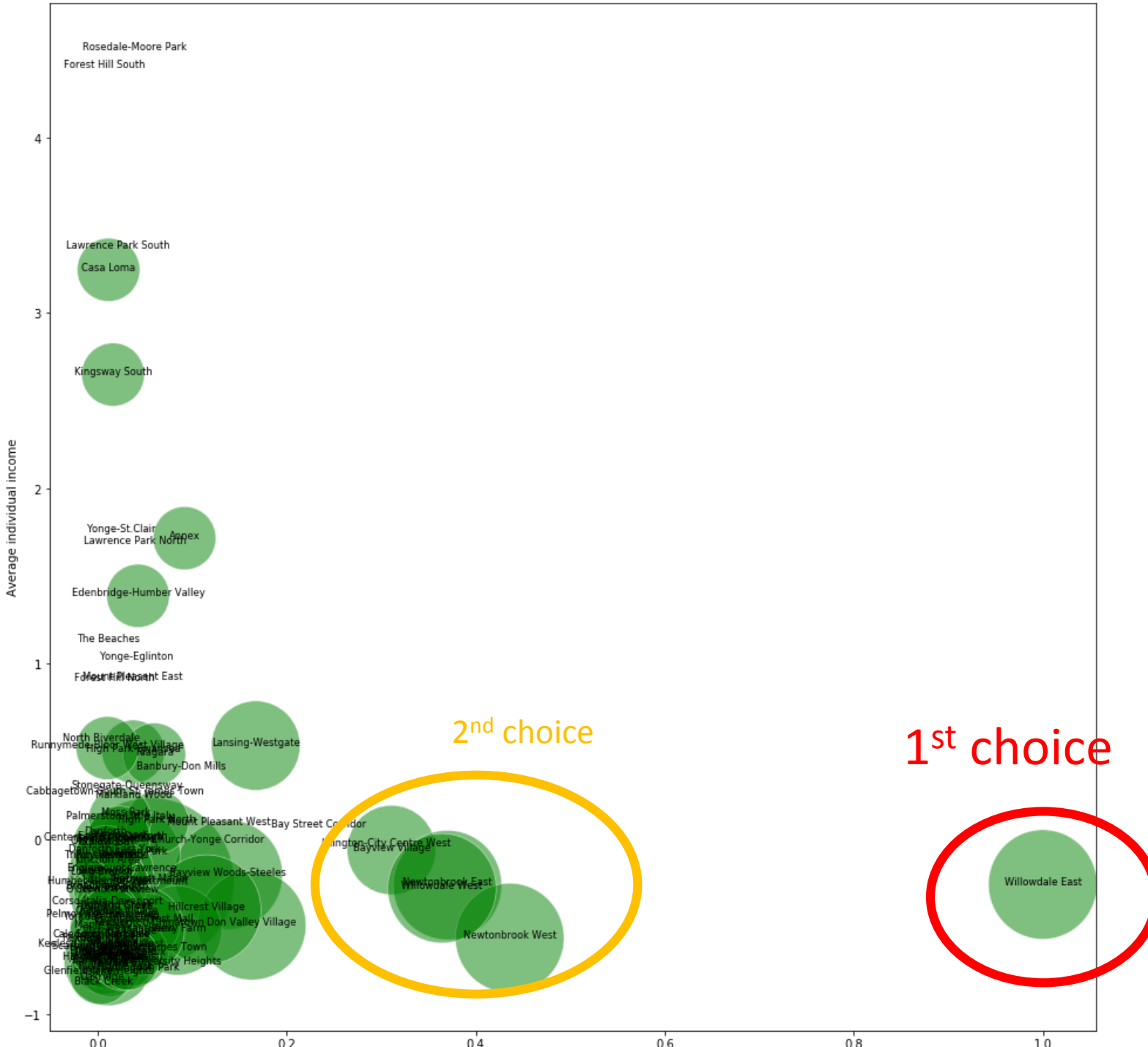
- We would like to avoid a neighbourhood with low Korean_index, low income, also low Kr_rest_sum. Hence, we can plot all the neighborhood with these three values as coordinates in the 3D space.

  Here we found an outlier, Willowdale East, which has very high Korean index, relatively good average income



Data visualization

- Again, Willowdale East will still stand out, Hence, we would suggest Dongjun to start his restaurant there.
- Newtonbrook West, Newtonbrook East, and Willowdale West could be his second group choice, who also has a slight less Korean index, but similar level of average income.



Conclusions

1. We first used publicly available data, and processed the data using Pandas. We then create a so called “Korean Index” to quantify the level of Korean-related of resident in each neighborhood. Besides, we also obtain the average individual income of each neighborhood.
2. We then combine Nominatim and Foursquare API to summarize the number of Korean restaurants in each neighborhood of Toronto.
3. Finally, we obtain the data we need and visualize them in both bubble and scatter plot.
4. The first choice is Willowdale East, which has a few high Korean index and relatively good average individual income, the number of other Korean restaurants is 3 which is smaller than 7 in Bathurst Manor. And second-best group of location are Newtonbrook West, Newtonbrook East, and Willowdale West. who also have a slightly less Korean index, but similar level of average income and few competitors as 3.