CAPSTONE PROJECT

FINDING THE BEST LOCATION IN TORONTO TO OPEN A CHINESE FOOD MARKET

QIAN LI

JULY 29, 2020

AGENDA

- Part 1: Intro/Business problem
- Part 2: Data and Methodology
- Part 3: Finding areas most populated by Chinese immigrants
- Part 4: Decide on area that is least competitive
- Part 5: Decide on area with lightest leasing cost
- Part 6: Modelling and Testing
- Part 7: Result
- Part 8: Discussion
- Part 9: Conclusion

PART 1: INTRO/BUSINESS PROBLEM

- Mr. Chen would like to open **a supermarket targeting Chinese immigrants.**
- His requirements for the ideal place are:
- ***1. large Chinese population;***
- ***2. few competitors;***
- ***3. reasonable rent.***

PART 2: DATA AND METHODOLOGY

- Data Sources: **Toronto public data portal,** **Foursquare** and **CBRE**.
- The whole research will be carried out in 4 main steps:
- 1. Use Toronto data to target the main candidate regions
  (Pandas and Matplotlib)
- 2. Use data from Foursquare and decide on the number of competitors and population coverage per competitor
  (Folium)

PART 2: DATA AND METHODOLOGY

- Use CBRE to see the leasing costs
- Crunch relevant data into one data frame, and import reference data, which is used to do supervised machine learning in deciding which region has the best chance for success.

PART 3: FINDING AREAS MOST POPULATED BY CHINESE IMMIGRANTS

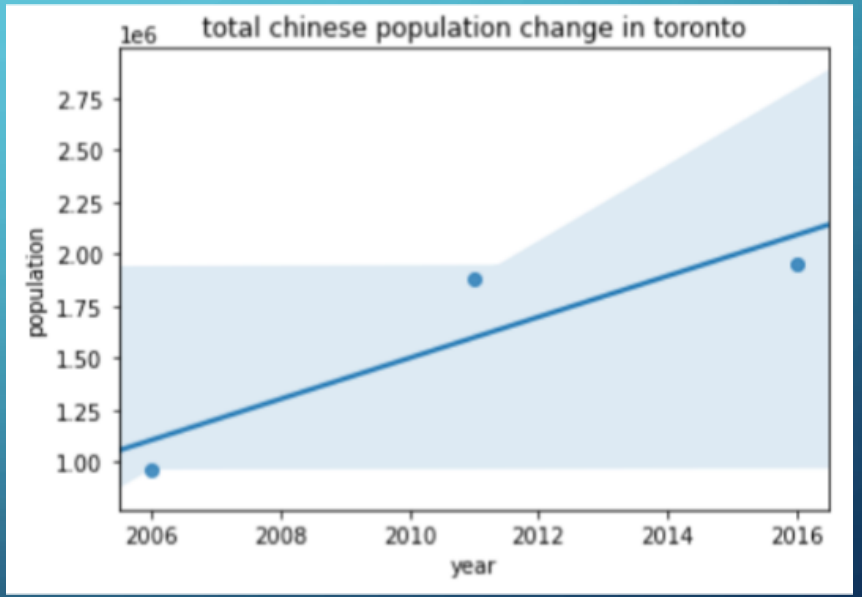
- Resource: public data portal for Toronto(https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/)
- There are 48 wards in Toronto. Data for Chinese population in each ward of 3 years are found: 2006, 2011, and 2016.

year	Ward 1	Ward 2	Ward 3	Ward 4	Ward 5	Ward 6	Ward 7	Ward 8	Ward 9	Ward 10	...	Ward 39	Ward 40	Ward 41	Ward 42	Ward 43	Ward 44	Ward 45	Ward 46	Ward 47	Ward 48
2006	1125.0	665.0	1530.0	900.0	1250.0	1070.0	1145.0	3805.0	1645.0	1480.0	...	26315.0	16000.0	33980.0	6935.0	1105.0	2085.0	0.0	0.0	0.0	0.0
2011	665.0	610.0	1410.0	885.0	1360.0	1230.0	1110.0	2545.0	1470.0	1170.0	...	27305.0	15390.0	35685.0	6180.0	1260.0	1950.0	11900.0	68780.0	50450.0	103840.0
2016	600.0	565.0	1365.0	935.0	1745.0	1715.0	1345.0	2100.0	1360.0	1620.0	...	26995.0	15185.0	35505.0	5955.0	1395.0	2000.0	12780.0	74625.0	102130.0	54425.0

PART 3: FINDING AREAS MOST POPULATED BY CHINESE IMMIGRANTS

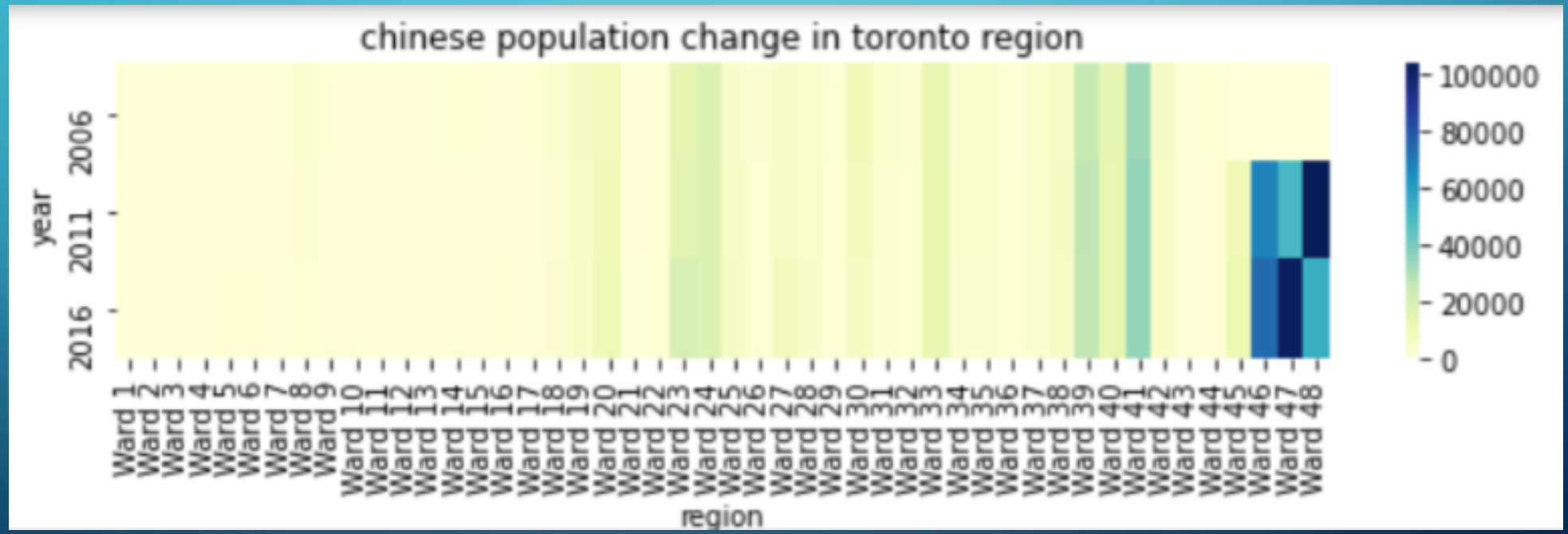
- A snapshot of total Chinese population change in whole Toronto:

year	total
2006	962140
2011	1879760
2016	1951620



PART 3: FINDING AREAS MOST POPULATED BY CHINESE IMMIGRANTS

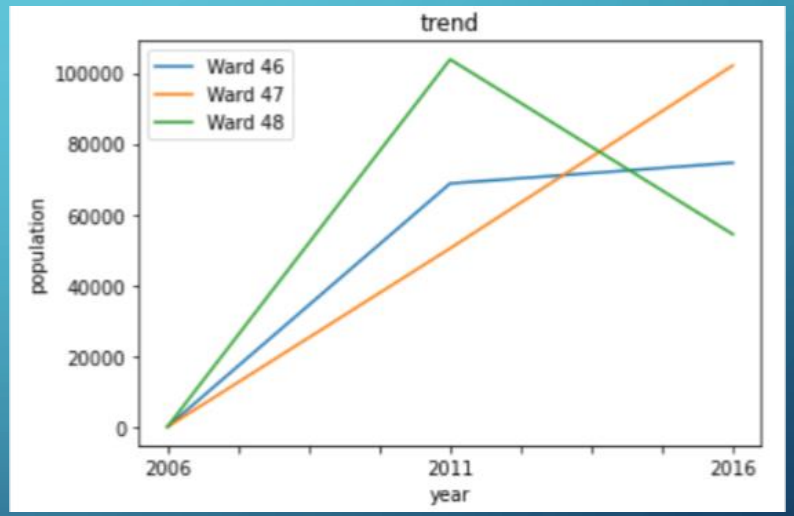
- Which region has the largest Chinese population? Wards 46,47,48.



PART 3: FINDING AREAS MOST POPULATED BY CHINESE IMMIGRANTS

- Chinese population trend in these 3 wards:

year	Ward 46	Ward 47	Ward 48
2006	0.0	0.0	0.0
2011	68780.0	50450.0	103840.0
2016	74625.0	102130.0	54425.0



- Data for 2006 is missing but it doesn't matter a lot for our analysis.
- When referring back to the original data from Toronto portal, we know that **Ward 46 is North York, Ward 47 is East York and Ward 48 is Scarborough.**

PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

- Using the geographical info of those wards, we put it into Foursquare and explore those areas.
- We use geolocator to decide the ward centers' latitude and longitude.
- North York geo location:
- 43.7543263 -79.44911696639593
- East York geo location:
- 43.6685545 -79.3072382
- Scarborough geo location:
- 43.773077 -79.257774

PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

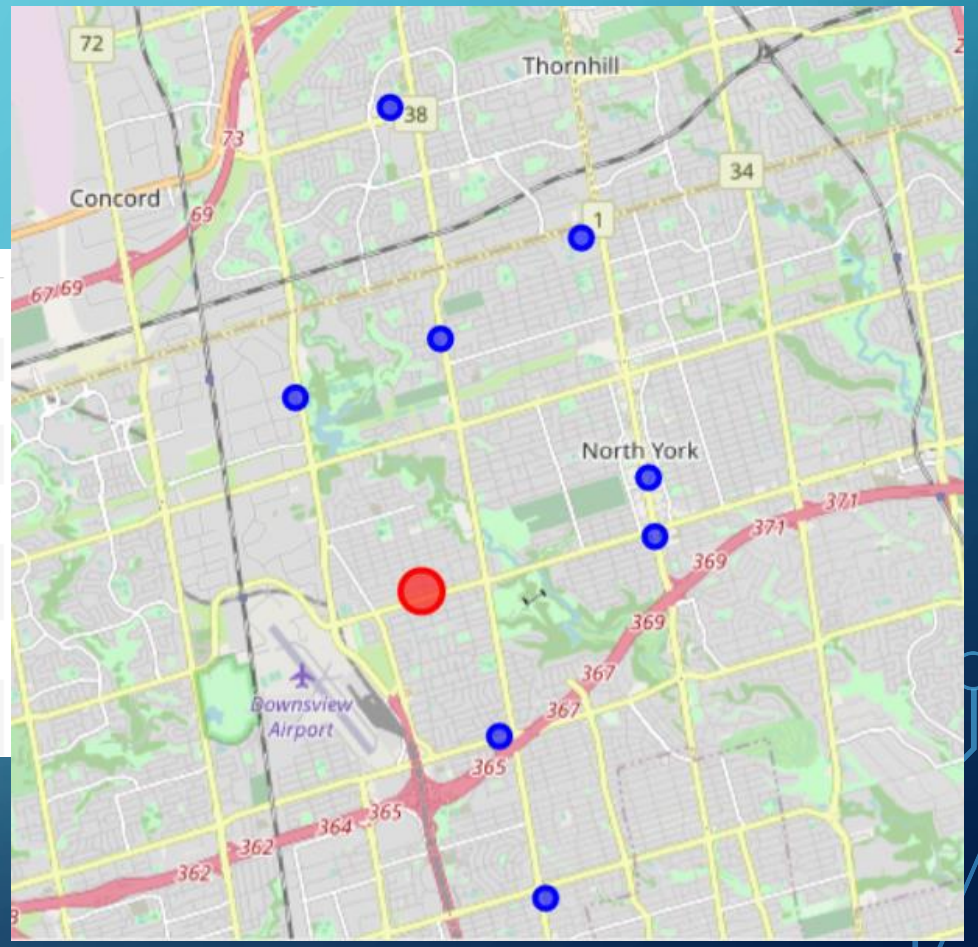
- Explore competitor markets around the areas: When doing research on Foursquare, the searching keyword is set to be “Asian Market” because all Asian markets are considered as competitors.

- We got lists of competitors and their locations on map as follow:

PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

- North York

	name	categories	lat	lng	labeledLatLngs	distance	cc	country	formattedAddress	address	postalCode	city	state	crossStreet	neighborhood	id
1	Whole Foods Market	Grocery Store	43.760662	-79.411044	[{'label': 'display', 'lat': 43.76066201105053...	3141	CA	Canada	[4771 Yonge Street, North York ON M2N 5M5, Can...	4771 Yonge Street	M2N 5M5	North York	ON	NaN	NaN	57ffd7b2d67c8ac5e8b34dd9
2	Asian Gourmet	Asian Restaurant	43.795940	-79.422880	[{'label': 'display', 'lat': 43.79594, 'lng': ...	5089	CA	Canada	[Centrepoint, Toronto ON, Canada]	Centrepoint	NaN	Toronto	ON	NaN	NaN	4c4743d576d72d7fad9b3c4d
7	Asian Course	Asian Restaurant	43.811346	-79.454070	[{'label': 'display', 'lat': 43.81134645979908...	6359	CA	Canada	[Vaughan ON, Canada]	NaN	NaN	Vaughan	ON	NaN	NaN	4c8a5f181797236a0afc6088
11	Coppa's Fresh Market	Supermarket	43.776956	-79.469547	[{'label': 'display', 'lat': 43.77695629281025...	3007	CA	Canada	[4750 Dufferin St (at Martin Ross Ave), North ...	4750 Dufferin St	M3H 5S7	North York	ON	at Martin Ross Ave	NaN	4b4a5a41f964a520ed8426e3
18	Savours Fresh Market	Market	43.718137	-79.428909	[{'label': 'display', 'lat': 43.718137, 'lng':...	4344	CA	Canada	[Canada]	NaN	NaN	NaN	NaN	NaN	NaN	5766c802498e925815e0724f
19	DPY Mama's Food Market	Grocery Store	43.737232	-79.436357	[{'label': 'display', 'lat': 43.73723200550416...	2161	CA	Canada	[322 Wilson Ave (Collinson Blvd), North York O...	322 Wilson Ave	M3H 1S8	North York	ON	Collinson Blvd	NaN	5d50879d32ac8d0008be69d5
20	Bathurst Village Market	Supermarket	43.784063	-79.445984	[{'label': 'display', 'lat': 43.784063, 'lng':...	3319	CA	Canada	[Toronto ON M2R 1Z1, Canada]	NaN	M2R 1Z1	Toronto	ON	NaN	NaN	5c2ff5b4ab42d9002cc04c65
23	Head Research (market research)	None	43.767594	-79.412041	[{'label': 'display', 'lat': 43.76759418909818...	3326	CA	Canada	[5075 Yonge St, Toronto ON, Canada]	5075 Yonge St	NaN	Toronto	ON	NaN	NaN	4d000b1121ea6ea88cce3f9f



PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

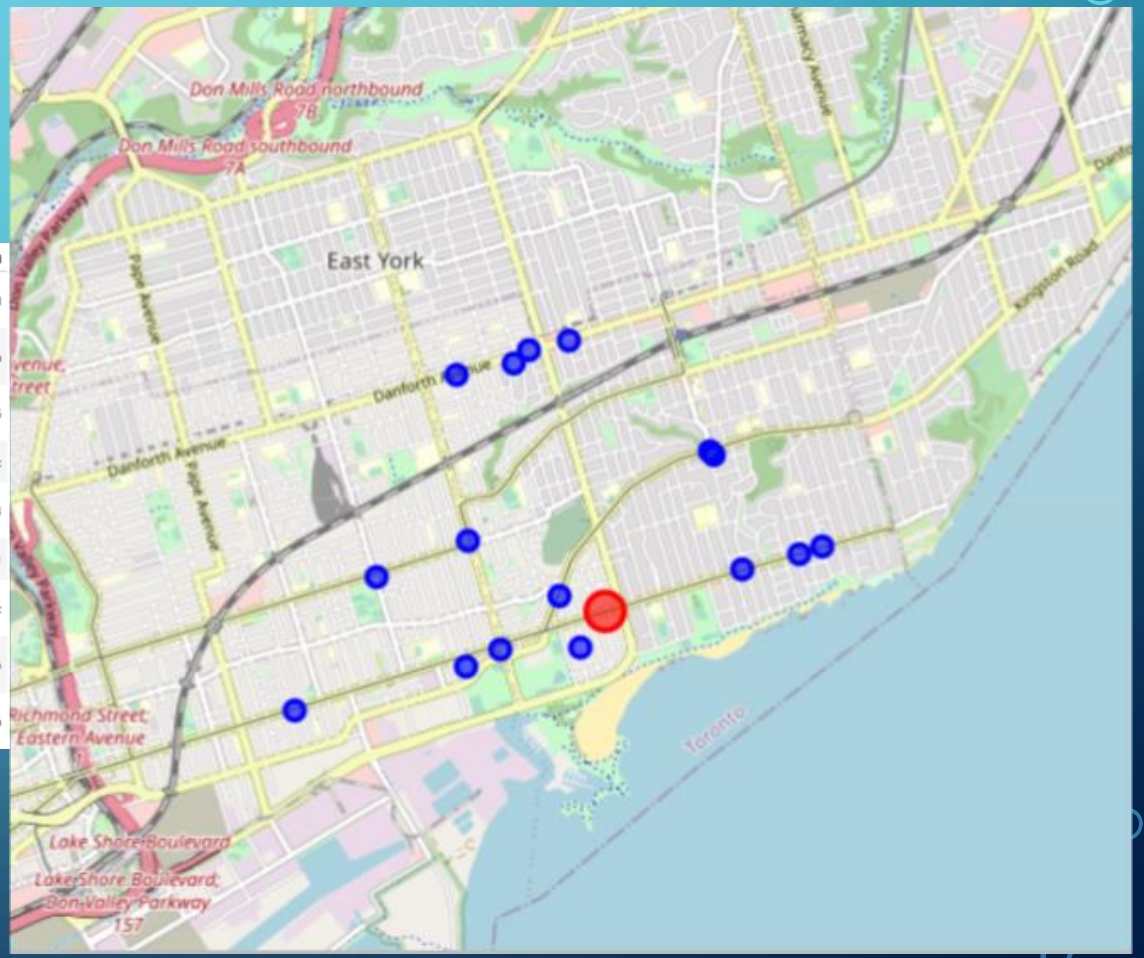
- Scarborough

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	cc	city	state	country	formattedAddress	postalCode	neighborhood	id
4	Asian Specialty	Vietnamese Restaurant	2301 Brimley Rd.	at Huntingwood Dr.	43.797819	-79.270737	[{'label': 'display', 'lat': 43.797819, 'lng':...	2944	CA	Toronto	ON	Canada	[2301 Brimley Rd. (at Huntingwood Dr.), Toront...	M1S 5B8	NaN	4f79f9c8e4b0f60afaff9f5d
5	Asian Gourmet	Chinese Restaurant	Lawrence	Markham	43.758263	-79.227862	[{'label': 'display', 'lat': 43.75826275954416...	2915	CA	Toronto	ON	Canada	[Lawrence (Markham), Toronto ON, Canada]	NaN	NaN	4c14fee382a3c9b63009fef8
7	Jocelyn's Asian Cuisine	None	NaN	NaN	43.785086	-79.260463	[{'label': 'display', 'lat': 43.785086, 'lng':...	1354	CA	NaN	NaN	Canada	[Canada]	NaN	NaN	4f021289a69d45461fab5ea7
12	Edgecombes Marketing And Promotions (warehouse)	Office	2250 Midland Ave, Toronto, ON M1P 3E6	NaN	43.775696	-79.274705	[{'label': 'display', 'lat': 43.77569613234338...	1391	CA	Toronto	ON	Canada	[2250 Midland Ave, Toronto, ON M1P 3E6, Toront...	M1P 3E6	NaN	4f819361e4b02b9569ed08a7
20	Kostas Meat Market	Greek Restaurant	259 Ellesmere Rd	Warden Ave	43.760605	-79.301830	[{'label': 'display', 'lat': 43.7606051148729,...	3804	CA	Toronto	ON	Canada	[259 Ellesmere Rd (Warden Ave), Toronto ON M1R...	M1R 4E4	NaN	4cd56bc194848cfa2b1ae6b1
21	Adlytic Marketing	Advertising Agency	2190 Ellesmere Road	NaN	43.777470	-79.230201	[{'label': 'display', 'lat': 43.7774695819856,...	2269	CA	Toronto	ON	Canada	[2190 Ellesmere Road, Toronto ON M1G 3M5, Canada]	M1G 3M5	Scarborough	5e4ccbb7e1617b00085ca334
23	Marketplace Ministries	Church	1776 Midland Ave	Midland And Ellesmere	43.762053	-79.270081	[{'label': 'display', 'lat': 43.76205334249128...	1576	CA	Scarborough	ON	Canada	[1776 Midland Ave (Midland And Ellesmere), Sca...	M1P 3P2	NaN	4f3fc3b2e4b0ae06556b2e0f

PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

- East York:

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	id
1	raise the root organic market	Fruit & Vegetable Store	1164 Queen Street east	Jones Avenue	43.666180	-79.309384	[{'label': 'display', 'lat': 43.66618035086522...	315	NaN	CA	Toronto	ON	Canada	[1164 Queen Street east (Jones Avenue), Toront...	54b1a7ba498e8d9a417cc4bd
2	Beaches Wholesome Market	Health Food Store	2234 Queen St E	Beech Ave	43.672661	-79.287567	[{'label': 'display', 'lat': 43.67266063441251...	1648	M4E 1G2	CA	Toronto	ON	Canada	[2234 Queen St E (Beech Ave), Toronto ON M4E 1...	4c547b89f5f3d13a5a2c3afa
3	Toby's Food Market	Convenience Store	86 Kingston Rd	NaN	43.669536	-79.311346	[{'label': 'display', 'lat': 43.669536, 'lng':...	348	M4L 1S6	CA	Toronto	ON	Canada	[86 Kingston Rd, Toronto ON M4L 1S6, Canada]	4da1dec12939b1f7e2fefd56
4	Leslieville Farmers Market	Farmers Market	NaN	NaN	43.664901	-79.319784	[{'label': 'display', 'lat': 43.66490139800862...	1088	NaN	CA	Toronto	ON	Canada	[Toronto ON, Canada]	4deb8ba688774880e3387c0c
5	The Big Carrot Natural Food Market	Health Food Store	125 Southwood Dr	NaN	43.678879	-79.297734	[{'label': 'display', 'lat': 43.678879, 'lng':...	1380	M4E 0B8	CA	Toronto	ON	Canada	[125 Southwood Dr, Toronto ON M4E 0B8, Canada]	4ad4c062f964a52011f820e3
7	Farmer's Fruit Market	Farmers Market	NaN	NaN	43.673050	-79.319514	[{'label': 'display', 'lat': 43.67304987074939...	1107	NaN	CA	NaN	NaN	Canada	[Canada]	4f6f9f06e4b08d8da7d60d11
8	The Dog Market	Pet Store	2116 Queen St E	Hammersmith Ave	43.671163	-79.294773	[{'label': 'display', 'lat': 43.67116322361954...	1044	M4E 1E2	CA	Toronto	ON	Canada	[2116 Queen St E (Hammersmith Ave), Toronto ON...	4ed4f69f93ad512d8df0511c
11	The Big Carrot Natural Food Market	Grocery Store	125 Southwood Dr	NaN	43.678730	-79.297478	[{'label': 'display', 'lat': 43.67873, 'lng': ...	1378	M4E 0B8	CA	Toronto	ON	Canada	[125 Southwood Dr, Toronto ON M4E 0B8, Canada]	5b82eb3501bc5a002ce55ebb
12	Rowan Homespun Market	Market	2196 Queen St E	Balsam Ave	43.672218	-79.289780	[{'label': 'display', 'lat': 43.67221832275390...	1463	M4E 1E6	CA	Toronto	ON	Canada	[2196 Queen St E (Balsam Ave), Toronto ON M4E ...	55ac0cb1498e89610f1d76aa



PART 4: DECIDE ON AREA THAT IS LEAST COMPETITIVE

- North York has 8 competitors, Scarborough has 8, and East York has 16.
- The competitiveness will need to factor in the population in those areas.
- divide the population by competitor number to see how large is the population each competitor cover, which we will name it as population coverage in the following part.

	ward	population	competitors number	population coverage per competitor
0	North York	74625	8	9328.125
1	Scarborough	54425	8	6803.125
2	East York	102130	16	6383.125

- North York having the biggest population coverage.

PART 5: DECIDE ON AREA WITH LIGHTEST LEASING COST

- Apart from the population coverage, Mr. Chen also wants to know the leasing cost, and uses both factors to determine the location choice. Here it is:

Submarket	Inventory (SF)	Vacancy Rate (%)	Sublet Space (% of Vacant Space)	Net Absorption (SF)	YTD Net Absorption (SF)	New Supply (SF)	Under Construction (SF)	Asking Net Rent ($/SF)
CENTRAL	87,701,047	2.0	19.7	139,957	139,957	0	8,807,072	33.85
Downtown	73,323,320	2.0	21.4	19,065	19,065	0	8,807,072	35.61
Financial Core	26,652,388	2.6	15.0	55,392	55,392	0	820,000	40.88
Greater Core	19,424,114	1.9	32.3	-91,993	-91,993	0	2,091,149	30.26
Downtown South	6,611,748	0.9	35.5	52,178	52,178	0	3,415,000	32.18
Downtown North	7,592,893	0.5	25.9	31,401	31,401	0	120,000	27.90
Downtown East	3,137,032	1.6	18.3	19,315	19,315	0	460,000	29.47
Downtown West	7,663,755	3.4	19.1	-54,822	-54,822	0	1,467,923	27.81
Liberty Village	2,241,390	0.3	27.3	7,594	7,594	0	433,000	27.14
Midtown	14,377,727	2.0	11.3	120,892	120,892	0	0	22.54
Bloor / Yonge	7,382,226	1.1	7.2	79,075	79,075	0	0	30.01
St. Clair / Yonge	2,194,054	1.5	0.0	-5,183	-5,183	0	0	24.20
Eglinton / Yonge	4,801,447	3.8	15.1	47,000	47,000	0	0	18.85
SUBURBAN	77,139,356	11.0	11.8	981,517	981,517	350,880	514,526	16.07
East	26,899,966	10.2	15.9	268,757	268,757	0	0	14.56
Scarborough	3,704,513	16.3	31.8	2,988	2,988	0	0	12.20
Markham N. / R. Hill	7,975,084	5.8	13.6	210,612	210,612	0	0	16.65
Markham South	4,017,526	11.5	3.9	-52,649	-52,649	0	0	15.05
E. York / D. Mills S.	2,663,480	10.9	14.8	-487	-487	0	0	12.52
Don Mills North	2,998,544	13.9	21.0	-26,422	-26,422	0	0	16.10
Consumers Road	3,876,535	10.0	5.2	137,644	137,644	0	0	14.97
G. Baker / Vic. Park	1,664,284	6.7	10.4	-2,929	-2,929	0	0	16.06
North	12,786,821	7.1	4.0	292,937	292,937	220,000	166,500	19.69
North Yonge	7,641,580	7.4	4.9	49,049	49,049	0	70,000	21.56
North York West	2,191,279	9.3	0.0	7,837	7,837	0	0	13.72
Vaughan	2,953,962	4.7	6.1	236,051	236,051	220,000	96,500	20.15

- (https://www.cbre.ca/en/research-and-reports/Toronto-Office-MarketView-Q1-2020)

PART 5: DECIDE ON AREA WITH LIGHTEST LEASING COST

- The corresponding leasing price for the 3 wards are included in the following chart with the population coverage information:

	ward	population coverage per competitor	leasing cost
0	North York	9328.125	21.56
1	Scarborough	6803.125	12.20
2	East York	6383.125	12.52

PART 6: MODELLING AND TESTING

- The final answer we want to get is whether the business will be successful or not, by analyzing the information we got.
- This is very similar to a prediction model in machine learning. Hence classification will be the method we use. We will try to use data features to predict the business in each ward will be successful (labelled as Y) or not (labelled as N).
- We found some historical data for supermarkets running in Toronto areas, which can be used to train the machine learning model. We use train-split-test to get the best model and then apply the model to real data for prediction.

PART 6: MODELLING AND TESTING

- Here is the data for modelling. For the convenience of modelling, we change the Y/N values into 1/0 as follow:

	Name	population coverage per competitor	leasing cost	Success or not
0	Market 1	10500	27	N
1	Market 2	7500	20	N
2	Market 3	6800	10	Y
3	Market 4	6600	9	Y
4	Market 5	7700	15	Y
5	Market 6	8000	13	Y
6	Market 7	9900	28	N
7	Market 8	11000	24	Y
8	Market 9	9800	19	Y
9	Market 10	8800	23	N

	Name	population coverage per competitor	leasing cost	Success or not
0	Market 1	10500	27	0
1	Market 2	7500	20	0
2	Market 3	6800	10	1
3	Market 4	6600	9	1
4	Market 5	7700	15	1
5	Market 6	8000	13	1
6	Market 7	9900	28	0
7	Market 8	11000	24	1
8	Market 9	9800	19	1
9	Market 10	8800	23	0

PART 6: MODELLING AND TESTING

- We use 80% of the data to train the model and 20% to test it. We process the "competitor" and "leasing cost" into numpy array as X, "success or not" as Y. KNN will be the method for use. We then use accuracy metrics to test how accurate it is. Setting K=2, we got the following result:

```
[51]: from sklearn.neighbors import KNeighborsClassifier
      #Start testing by using k as 2.
      k = 2
      #Train Model and Predict
      neigh = KNeighborsClassifier(n_neighbors = k).fit(X_train,Y_train)
      neigh

      /home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_laun
      the shape of y to (n_samples, ), for example using ravel().
        """

[51]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                           metric_params=None, n_jobs=None, n_neighbors=2, p=2,
                           weights='uniform')

[52]: #We use the model to predict the test result.
      Yhat = neigh.predict(X_test)
      Yhat

[52]: array([1, 0])
```

- Yhat is 1 and 0, by referring back to original data, the prediction is correct. Accuracy for train set is 0.875 and test set 0.5.

PART 6: MODELLING AND TESTING

- The result looks nice, but we want to find the best model, so we try to test every K value possible, from 1 to 9. The calculation tells us that the best K is 1, when train set accuracy is 1 and test set 0.5. Then we redo the model with k=1 and apply the model to actual data of the 3 shortlisted wards.

```
[57]: #so the best k is 1. Redo the model.
      k = 1
      neigh = KNeighborsClassifier(n_neighbors = k).fit(X_train,Y_train)
      neigh

      /home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:3: Da
      the shape of y to (n_samples, ), for example using ravel().
        This is separate from the ipykernel package so we can avoid doing imports until
[57]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                           metric_params=None, n_jobs=None, n_neighbors=1, p=2,
                           weights='uniform')

[58]: #imput the data that we need to predict for the 3 wards.
      X1=df[['population coverage per competitor','leasing cost']].values
      X1=preprocessing.StandardScaler().fit(X1).transform(X1.astype(float))

[59]: #now let's see using the KNN model we have how's the result for each ward.
      Y1hat = neigh.predict(X1)
      Y1hat

[59]: array([0, 1, 1])
```

- It gives a final result as an array [0,1,1] for the 3 wards.

PART 7: RESULT

- We transform that Boolean values back into Y/N values and integrate them into the data frame:

	ward	population coverage per competitor	leasing cost	Success or not
0	North York	9328.125	21.56	N
1	Scarborough	6803.125	12.20	Y
2	East York	6383.125	12.52	Y

- The analysis for the 3 wards has a preliminary conclusion:
- To open a Chinese market in the 3 wards, North York will be not successful, Scarborough and East York will be successful.

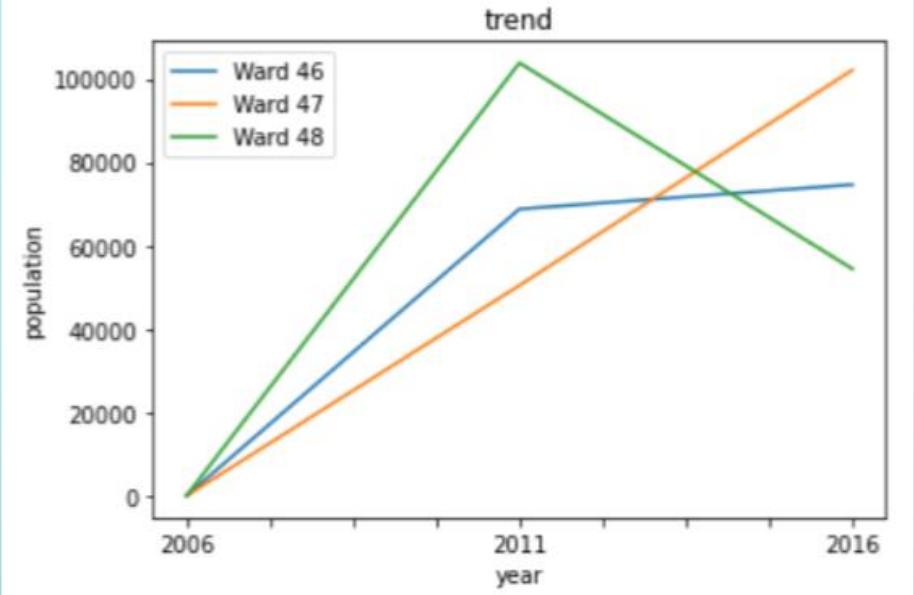
PART 8: DISCUSSION

- Unfortunately, North York fails the test.
- For Scarborough and East York:

1. Scarborough slightly wins over East York in population coverage (6803 over 6383, the difference is about 5%)
2. Scarborough wins by 0.32 CAD per square foot (12.20 compared with 12.52 in East York, the difference is about 2%).
3. However in population trend, Scarborough is much less favorable to East York.

PART 8: DISCUSSION

- Now recall the following plot:



- Just to remind that Ward 47 is East York and Ward 48 is Scarborough. We can see that the Chinese population in Scarborough is decreasing sharply by 47.5%! While that in East York is growing by 102%! We can conclude that Chinese immigrants are leaving Scarborough and more like to settle down in East York.

PART 8: DISCUSSION

- Business needs to look in to future, population trend means future, which is very important for business.
- Population is the most important feature we need to consider for potential market. Besides, when comparing with the population trend change, the difference of leasing costs and population coverage look small. Therefore, even though East York is slightly less advantageous to Scarborough in leasing cost and population coverage, the former's population trend benefit is much better than the latter. **East York is a better choice**.

PART 9: CONCLUSION

- Weighing in the population trend, population coverage and leasing cost, the final winner is:

East York!!!