Aruna Dorai

June 2020

Coursera Capstone Project report

Introduction

Capstone assignment - Opportunity Assessment for Capstone project using Foursquare API.
I am very excited to share that we have been hired by the Mayor of Mississauga to assist in approval for a request to set up Italian restaurant in South Mississauga neighborhood. With Covid this year it makes sense to provide people with what they need in their neighborhood. this will avoid un-necessary commute, helps isolation.

We plan to use the Foursquare Api to help the mayor with the decision. We will analyze the currently existing businesses and provide guidance with facts on the area.
This report should be reviewed in conjunction with the associated Jupyter notebook published in Github.
Publish on GitHub
The Capstone assignment on Foursquare API -part 2.ipynb file was published in the arunado/Coursera-Capstone repository.
Click #2791965 to view the commit.

Process/Hypothesis –

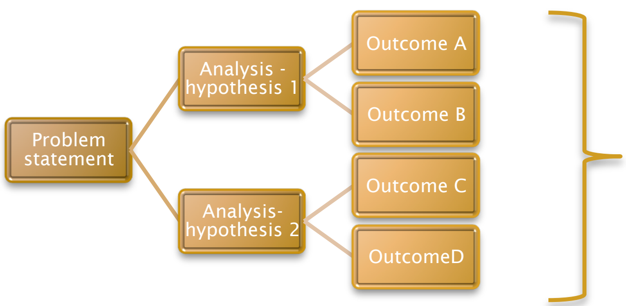
1. Do we need a restaurant in Mississauga? We will analyze the population of Mississauga, understand the ratio of people to restaurants.
Mississauga is part of Toronto, and for this exercise we will use Toronto as a representative for Mississauga.

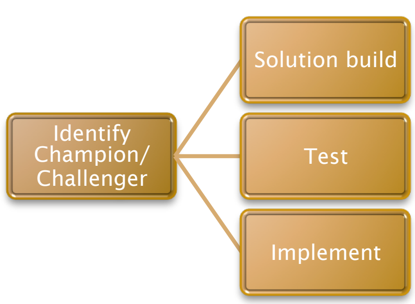
A. Analyze using Data Science if is a correlation between
the number of restaurants in Toronto today -

1. Average salary in a city and number of restaurants.
2. No of Employed people versus no of restaurants.

B. Will an Italian Restaurant be successful in this neighborhood.

The process followed -





The above-mentioned process was followed to address this problem.

Who can use this report?

This report is specifically created for the Mississauga mayors office, however can be used by zoning officers or new small businesses trying to enter the market in the south Mississauga (Port Credit) neighborhood.

This neighborhood is one of the affluent communities with beautiful custom homes and manicured gardens. The demographic is mostly double income families and has discretionary amount for going to restaurants.

Data

Various data sources were used to do this analysis.

The Coursera program helped us to be able to locate public data sources. It also enabled us to be able to extract data from HTML tables in website.

As part of the research I became educated on various sources available to data Scientists. For example Yougov - YouGov is a global provider of analysis and data generated by consumer panels in 42 markets.

However we need to exercise caution and validate that our data sources are reliable and accurate.

For the Jupyter notebook, the Coursera previous lessons, Foursquare API, Github, Snowflake and python documentation was used.

Data source world cities and no of restaurants by population The World Cities Culture Forum (WCCF), which BOP convenes, provides a way for policy makers in 38 key cities to share research and intelligence, and explore the **vital role of culture in their future prosperity.**

1. URL - http://www.worldcitiescultureforum.com/data/number-of-restaurants-per-100.000-population

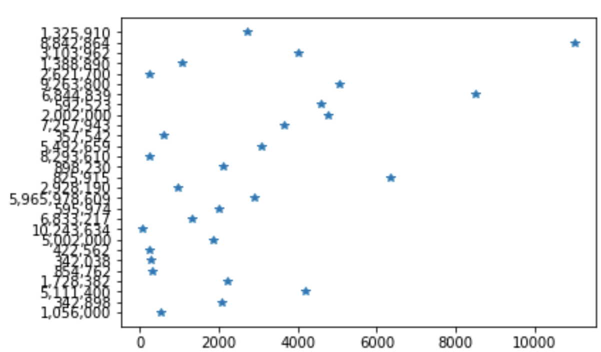
Mississauga population and employment data

2. URL – http://www7.mississauga.ca/documents/pb/opendata/2019/P_10_2019EmploymentProfileFinal.pdf
3. URL https://yougov.co.uk/topics/food/articles-reports/2019/03/12/italian-cuisine-worlds-most-popular

Methodology

1. Statistical correlation analysis was performed on 3 data elements from 3 data sets.
2. The hypothesis 1 -was is there a direct correlation between average salary in a city and the number of restaurants. – Outcome – No there is none.
3. The hypothesis 2 -was is there a direct correlation between No of persons employed in a city and the number of restaurants. – Outcome – No there is none.
4. The 3 statistics methods of Pearson, Kendall and spearman were used
5. As you can see from the scatter graph it is not correlated.

```
In [53]: plt.plot( 'Figure_x', 'Figure', data=dfX, linestyle='none', marker= '*')
         plt.show()
```

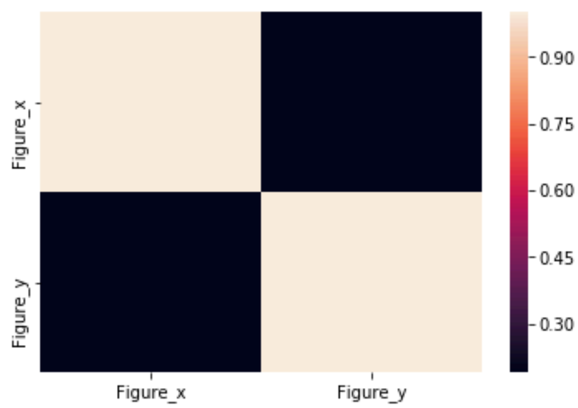


6. The Correlation analysis was not conclusive.

7. A heatgraph using seaborn was also done – showing results as below –

```
In [23]: import seaborn as sns
         corr = dfcom.corr(method="pearson")
         sns.heatmap(corr)

Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3734037c50>
```



8. The next analysis was to explore the neighborhood and explore if there was indeed an Italian restaurant in the neighborhood.
9. Italian food as we found is the most popular cuisine in the world see attached heat chart.

Italian, Chinese and Japanese cuisines are the world's most popular

% of people who have tried that cuisine in each country that say they like it

	Philippines	Singapore	Australia	Taiwan	UAE	Hong Kong	Britain	Finland	India	France	USA	Norway	Sweden	Malaysia	Vietnam	Germany	Denmark	Spain	Saudi Arabia	Italy	Indonesia	Thailand	China	Japan	AVERAGE CUISINE POPULARITY
Italian cuisine	90	89	90	81	87	81	91	87	80	93	88	83	92	74	72	89	86	94	78	99	70	69	59	85	84
Chinese cuisine	88	94	86	87	69	91	86	81	77	80	84	77	74	74	73	78	73	71	54	62	57	80	95	88	78
Japanese cuisine	90	94	74	92	58	93	57	63	57	70	74	61	66	73	81	58	61	57	43	62	72	90	54	94	71
Thai cuisine	76	91	85	76	61	84	75	77	65	70	68	74	78	91	75	68	73	56	31	52	51	98	44	48	70
French cuisine	82	79	70	71	71	73	68	74	70	96	69	74	79	58	72	62	68	70	63	60	46	62	63	68	70
Spanish cuisine	86	68	74	62	67	67	80	79	58	87	79	81	82	44	45	70	74	98	56	87	36	39	38	64	68
American cuisine	93	83	68	76	75	62	70	69	75	56	91	68	63	65	71	53	70	49	66	54	66	73	51	57	68
Mexican cuisine	85	72	77	62	71	54	76	81	70	78	86	84	77	51	46	71	78	72	64	77	38	54	42	51	67
Indian cuisine	55	77	74	60	71	48	84	71	93	71	55	75	67	70	44	62	67	52	76	57	49	27	26	63	62
Turkish cuisine	61	63	71	49	77	45	67	68	56	62	43	62	65	58	47	64	50	53	85	47	65	32	47	39	57
Korean cuisine	87	86	66	74	50	78	43	45	42	45	58	43	48	66	81	45	32	36	29	32	63	78	52	66	56
Greek cuisine	57	49	77	48	55	47	74	77	46	75	69	75	85	31	32	79	69	62	51	71	27	23	32	22	56
Vietnamese cuisine	66	67	75	63	39	77	53	55	42	74	59	56	54	45	96	57	63	38	18	36	34	67	37	44	55
Hong Kong cuisine	80	91	68	78	45	93	60	38	51	44	50	37	29	54	68	39	33	38	22	35	47	68	68	53	54
German cuisine	59	64	63	62	47	65	51	63	49	51	63	52	57	35	51	87	37	43	38	40	37	54	45	46	52
British cuisine	70	76	71	50	63	61	91	49	68	28	49	45	35	65	61	25	27	26	52	32	58	44	43	20	50
Taiwanese cuisine	68	87	59	96	42	88	37	37	42	46	48	31	25	49	62	41	28	36	24	32	40	54	64	66	50
Singaporean cuisine	72	96	75	68	53	73	55	31	62	31	34	34	24	74	64	33	29	27	26	29	55	52	54	33	49
Indonesian cuisine	59	84	71	49	47	46	54	43	48	44	39	43	44	80	39	50	44	34	45	35	97	26	31	32	49
Malaysian cuisine	65	91	76	61	52	67	58	40	56	35	34	35	30	97	44	34	32	24	37	21	66	34	41	25	48
Australian cuisine	74	74	89	49	52	58	52	47	59	35	43	41	34	54	57	36	42	27	32	35	41	48	41	28	48
Moroccan cuisine	50	49	65	44	63	38	66	49	50	81	47	48	51	45	26	42	44	53	67	50	39	25	34	16	48
Lebanese cuisine	40	50	71	35	82	31	58	42	54	67	47	48	67	43	26	35	39	43	84	45	32	22	31	13	46
Caribbean cuisine	63	50	55	46	47	39	66	57	50	54	68	54	48	31	35	50	44	54	33	49	21	29	30	22	46
Brazilian cuisine	59	45	56	43	48	41	49	47	49	58	55	48	40	26	47	49	42	56	39	66	28	29	42	33	46
Swedish cuisine	50	63	47	49	44	48	47	68	48	41	48	62	92	36	40	42	29	30	35	33	27	29	34	18	44
Argentinian cuisine	55	35	53	42	45	36	46	42	43	54	48	54	41	24	35	60	45	77	29	75	25	24	37	16	43
Danish cuisine	51	48	47	39	39	40	42	56	41	35	41	67	62	33	40	36	85	20	29	25	28	29	32	14	41
Emirati cuisine	51	43	43	48	84	28	23	22	63	30	23	24	22	51	49	27	18	27	71	24	53	29	29	14	37
Norwegian cuisine	47	48	39	41	41	42	31	52	38	40	38	81	41	31	36	33	27	27	26	27	23	30	31	17	37
Filipino cuisine	97	41	56	35	48	29	29	37	39	31	45	40	25	33	38	38	30	26	25	25	29	25	29	21	36
Saudi Arabian cuisine	47	42	41	33	67	30	18	21	60	29	24	20	22	66	28	24	22	28	89	25	64	22	31	11	36
Finnish cuisine	47	37	37	33	35	38	23	94	39	33	27	28	36	26	39	26	13	18	26	24	23	26	30	14	32
Peruvian cuisine	47	33	44	33	37	31	27	30	39	39	44	30	27	19	37	28	21	50	23	39	21	16	28	16	32
AVERAGE ATTITUDE TOWARDS CUISINES BY NATION	67	66	65	57	57	57	56	56	56	55	54	54	52	52	52	50	47	46	46	46	45	44	42	39	

10. In this survey as you can see it was found that Italian food is the most popular cuisine in the world(source -Yougov.)
11. Patriotically, the biggest fans of Italian food are Italians themselves, with 99% enjoying their national cuisine. Other big fans include Spaniards (94% of those who have tried it say they like it) and the French (92%), while the least impressed by Italian food are the Chinese (59%).

12. We will map the neighborhood of the proposed site of the new reataurant in Mississauga, for nearest Coffee shops, Restaurants and entertainment areas.
13. This will help in deciding the zoning. We will create an account in Foursquare api. We will obtain the client ID and Secret after registering with foursquare.
14. We will run a search query to see if there are any Italian restaurants in the neighborhood. This result will then be loaded on a Pandas Data frame. Using Folium Maps we will visualize this data and explore the venue. We need to ensure there are not any Italian restaurants in the neighborhood. The report with the maps will be presented to the Mayors office, so they can make a decision if another Italian restaurant is required in the area.

-

2. Using the Ratio of people to restaurants, from our Data set 1-

28	Sydney	400.4
29	Taipei	307.6
30	Tokyo	1,099.5
31	Toronto	272.5>

3. Toronto’s average has 272 per 100,000 people.

4. Mississauga has a population of 775,000. So to keep up the Toronto average we need – 272 * 775000 = 2108 restaurants as least.

5. As you can see in chart below “Food services in Mississauga is 1747.

NUMBER OF BUSINESSES BY BUSINESS CLASSIFICATION FOR TOP 10 SECTORS	
2,992	RETAIL TRADE
2,416	MANUFACTURING
2,364	OTHER SERVICES
2,175	WHOLESALE TRADE
1,975	PROFESSIONAL + TECHNICAL SERVICES
1,747	ACCOMMODATION + FOOD SERVICES
1,708	HEALTH CARE + SOCIAL ASSISTANCE
1,058	TRANSPORTATION + WAREHOUSING
859	FINANCE + INSURANCE
784	EDUCATIONAL SERVICES

6. Where as Mississauga fell short by ~361 restaurants. **So yes we can increase the number of restaurants in Mississauga.**
7. Now to examine the restaurants currently in place in the proposed site.
8. Using the Foursquare api we mapped the businesses in this neighborhood.

9. As you can see this map has various categories and businesses(Blue dot) near the proposed restaurant location (Red Dot)

Shawanaga Trail
Mississauga Road
Fleet Street
North Sheridan Way
South Sheridan Way
Mesa Crescent
Vanier Drive
Caldwell Avenue
Woodeden Drive
Indian Road
Canyon Street
Chippewa Trail
Stonehouse Crescent
Queen Elizabeth Way
Saginaw Crescent
Glenhill Crescent
Streambank Drive
Tecumseh Park Drive
Oak Lane
Tenoga Drive
Obeck Crescent
Creditdale Road
Narva Road
Birchview Drive
Algonquin Drive
Springhill Drive
Aldo Drive
Lorne Park
Lornewood Creek
Queen Street West
Lakeshore Road West
Richard's Memorial Park
QEW
Wembury Road
Winslow Road
Crescent Road
Lorne Park Road
Queen Victoria Avenue
Whittier Crescent
Truscott Drive
Lorne Park Secondary School
Lorne Park Estates
Camelford Road
Randor Drive
Merrow Road
Garden Road
Whiteoaks Park
Birchwood Drive
Birchwood Creek
Glen Road
GO Transit - Oakville Subdivision
Bramsey Drive
Jalna Avenue
Spring Road
Greenoaks Drive
Echo Drive
Parkland Avenue
Porcupine Avenue
Jack Darling Memorial Park
Agnew Road
Mazo Crescent
Clarkson Road North
Kos Boulevard
Andros Boulevard
Bexhill Road
Glen Leven

```
In [49]: df.groupby('categories').count()
```

Out[49]:

categories	name	address	cc
Arts & Crafts Store	1	0	1
Assisted Living	2	1	2
Automotive Shop	2	2	2
BBQ Joint	1	0	1
Bakery	1	1	1
Bank	1	0	1
Bistro	1	1	1
Boat or Ferry	1	0	1
Breakfast Spot	1	0	1
Building	2	0	2
Bus Line	1	0	1
Chinese Restaurant	2	2	2
Church	2	0	2
Coffee Shop	1	0	1
College Classroom	1	1	1
College Football Field	1	1	1
Convenience Store	1	1	1
Cosmetics Shop	1	1	1
Design Studio	1	0	1

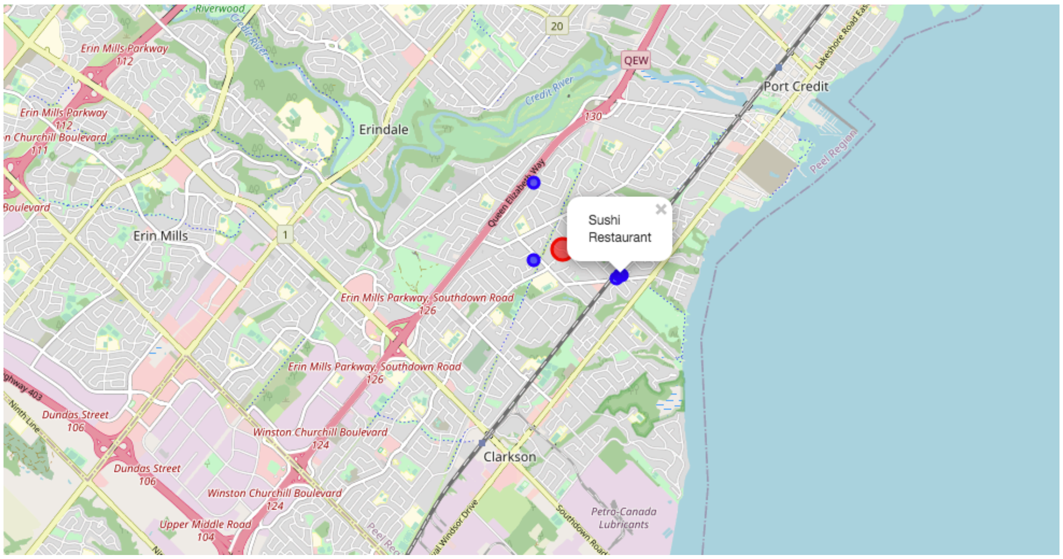
10. The data set was further filtered to only restaurants and we got a list as follows –

```
In [54]: df.head()
```

Out[54]:

	name	address	c
categories			
Chinese Restaurant	Timjas Thai Chinese	17-1107 Lorne Park Rd	C
Sushi Restaurant	Ichiban Sushi House	NaN	C
Chinese Restaurant	James Wok	1107 Lorne Park Road	C
Xinjiang Restaurant	Ta'am	1107 Lorne Park Rd	C
Sushi Restaurant	Orange Fish Sushi House	NaN	C

There is definely room to add an italian restaurant.



Results

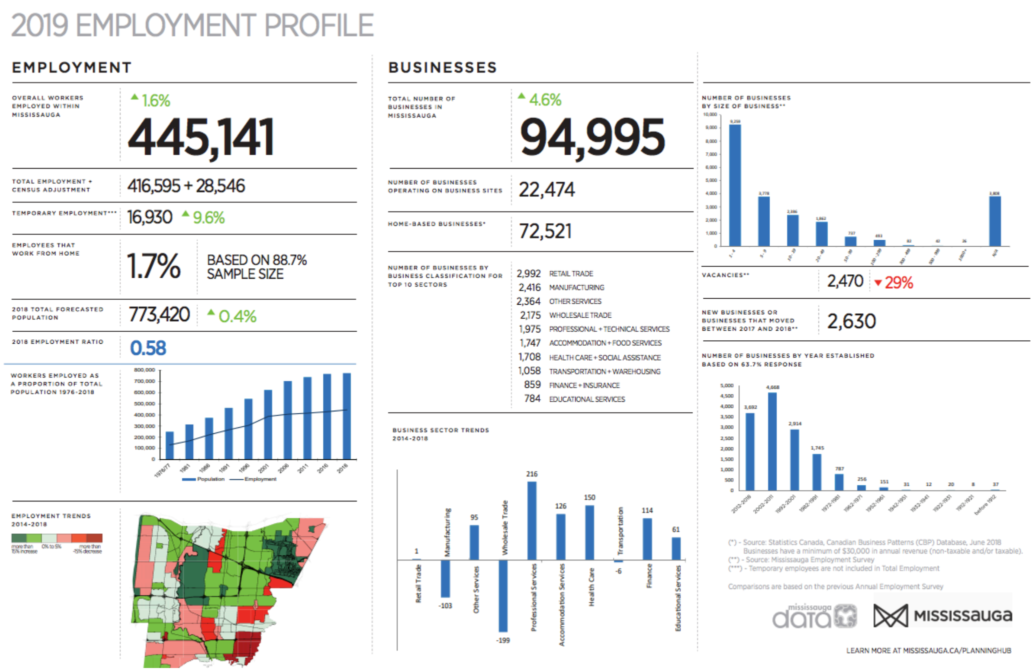
As seen in this report above the Mayor is advised to **approve** the restaurant and proposed **Italian** would be perfect for the Mississauga neighborhood.

Proposed location – ***1300 Queen Victoria Avenue, Mississauga, L5H3H2***

Applied data science was very useful to help arrive at this conclusion.

Appendix –

1. Missisauga Data



3. Data Set 1

Out[1]:

	City	Figure	Per capita	Date	Source	Notes
0	Amsterdam	53.4	NaN	2013	Trade Association for Horeca and Catering	NaN
1	Austin	207.2	NaN	2018	Restaurant Guide - Austin Chronicle	NaN
2	Bogotá	418.6	NaN	2017	Cámara de Comercio de Bogotá	NaN
3	Brussels	360.4	NaN	2013	FOD Economie ADSEI	NaN
4	Buenos Aires	221.3	NaN	2016	FEHGRA	NaN

Read a CSV file that shows no of restaurants per 100,000 people Convert it to a data frame locate Toronto data.

4. Data set 2

```
In [6]: body = client_590928f7f599463397dc68aaee279193.get_object(Bucket='arunasfirstproject-donotdelete-pr-gw4xfiuqoj149x',Ke
y='Average_income_per_capita_per_year_(ppp)_5112018.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df_data_4 = pd.read_csv(body)
df_data_4.head()
```

Out[6]:

	City	Figure	Per capita	Date	Source	Notes
0	Amsterdam	$19,271	NaN	2012	Statistics Netherlands/TNO	NaN
1	Netherlands	$17,492	NaN	2009	Statistics Netherlands /RIO	NaN
2	Austin	$39,103	NaN	2016	Census Reporter	ACS 2016 1-year
3	United States	$28,155	NaN	2013	US Census Bureau	NaN
4	Bogotá	$9,004	NaN	2014	Observatorio de Desarrollo Económico de Bogotá...	Base de Datos Dinámica Económica - Calidad de ...

5. Data Set 3

```
In [29]: body = client_590928f7f599463397dc68aaee279193.get_object(Bucket='arunasfirstproject-donotdelete-pr-gw4xfiuqoj149x',Ke
y='Working_age_population_181111.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df_data_5 = pd.read_csv(body)
df_data_5.head()
```

Out[29]:

	City	Figure	Per capita	Date	Source	Notes
0	Amsterdam	1,056,000	NaN	2013	Statistics Netherlands/TNO	NaN
1	Netherlands	7,392,000	NaN	2011	Statistics Netherlands	NaN
2	Austin	342,898	NaN	2016	Census Reporter	NaN
3	United States	182,402,740	NaN	2016	US Census Bureau	NaN
4	Bogotá	5,111,400	NaN	2018	Departamento Nacional de Estadística	NaN