Capstone project – week 2

[Document subtitle]

Contents

[FROM PROBLEM TO APPROACH 2](#_Toc533844651)

[What is the Problem we are trying to solve? 2](#_Toc533844652)

[Problem 2](#_Toc533844653)

[Background 2](#_Toc533844654)

[Stakeholders/Audience 2](#_Toc533844655)

[High level solution approach 2](#_Toc533844656)

[Using Data to Answer the Question 2](#_Toc533844657)

[How & Why? 2](#_Toc533844658)

[WORKING WITH DATA 3](#_Toc533844659)

[What data we use in resolving the puzzle? 3](#_Toc533844660)

[DERIVING THE ANSWER 3](#_Toc533844661)

[Analytic Approach 3](#_Toc533844662)

[Step – 0 3](#_Toc533844663)

[Step – 1 3](#_Toc533844664)

[Step – 2 4](#_Toc533844665)

[Step – 3 4](#_Toc533844666)

[Step – 4 4](#_Toc533844667)

[Step – 5 5](#_Toc533844668)

[Step – 6 6](#_Toc533844669)

[Step – 7 6](#_Toc533844670)

[Result 7](#_Toc533844671)

[Conclusion 7](#_Toc533844672)

# FROM PROBLEM TO APPROACH

## What is the Problem we are trying to solve?

#### Problem

What would be the ideal Business (i.e. Coffee Shop, Bakery, Gym etc.) to open in a given neighborhood considering the least satisfied needs in the area. If you find such business, then where would be the best location in the region to set up the operations for competitive advantage.

#### Background

In commercial neighborhoods you find businesses spreading across many different categories. Each category satisfies a unique need and start of a new business in any category will always change the prevailing business landscape and its dynamics in the region. The model defines in this exercise would help building an insight on above business problem by evaluating the current business landscape real time. It provides tools to figure out what business to start in the given region and where to setup its operations.

#### Stakeholders/Audience

Whoever who wants to start a new business in a given neighborhood can utilize the results to scan the business landscape and decide where to place (geographically) the business based on the customer ratings and the number of similar businesses existing.

#### High level solution approach

“Fulfil the needs scarcest and use other’s lower customer rating to exploit your business opportunity”.

## Using Data to Answer the Question

#### How & Why?

Building awareness on existing business types and venues is a primary need in setting up a new business within any region. An insight over the number of businesses in the same category (i.e. how many coffee shops, book shops etc.) would be crucial. Depending on that count one can decide whether that unique need is already fulfilled or not (how scarcest). Therefore, data on the given region to evaluate current business landscape is a mandatory need in this study.

Knowing customers’ view over existing businesses and venues add value to the study as it’s a deciding factor in the end. So, data on customer’s view (in a form of customer rating) on each business venue would be a major necessity (other’s lower customer rating to exploit your business opportunity) to execute our strategy.

In summary we need data on

* Existing business category and venue data in the region of interest
* Customer ratings data on each business venue
* Supporting data, to collect additional details on geographic locations and neighborhoods.

# WORKING WITH DATA

## What data we use in resolving the puzzle?

Business related data in the neighborhood we collected through Foursquare API. It provides a comprehensive set of details on business venues and the customer ratings. While Foursquare provides business details the Wikipedia provided some location specific data (such as post codes and neighborhoods) together with geospatial data in Cognitive class. Further details you can find in upcoming sections in the report. In summary the data is sourced from

* Foursquare API – Through API calls
* Wikipedia – Through web scraping API

The business data collected through Foursquare API provides rich set of real time information adding a great value. So, we can assure the data we use in the study represents the problem we try to solve. However, fundamentally it depends on crowdsourcing so, whatever shortcomings come along with crowdsourcing will influence the credibility of the result.

Few data manipulations were done when pre-processing Wikipedia data on postal codes, especially when you have missing data (in a form of ‘Not assigned’). But certainly, it won’t create any significant impact at the end.

# DERIVING THE ANSWER

## Analytic Approach

In this study we use visualization analytic approach to derive the answer. How the data can be visualized in getting the answer step by step is described below. (The technical implementation of each step corresponds the sections in Jupyter notebook)

### Step – 0

We have datasets fetched from ‘https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M’ and ‘http://cocl.us/Geospatial\_data’. After pre-processing we have a combined dataset includes boroughs, neighborhoods and geospatial statistics (latitudes, longitudes).



### Step – 1

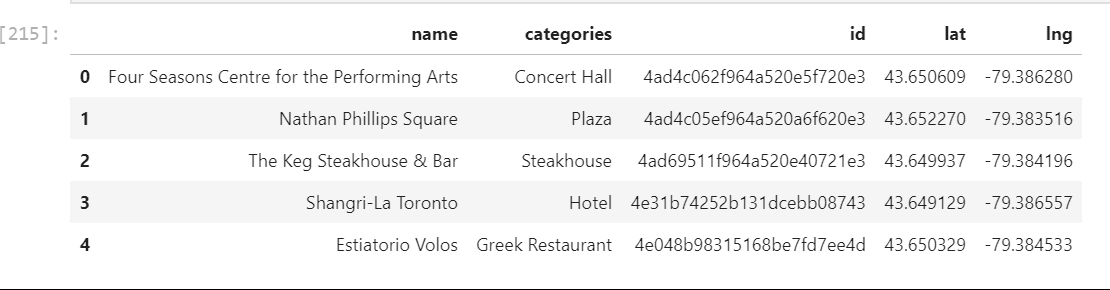
Select a neighborhood for the study from a given borough/post code. Within 500m perimeter from that given neighborhood search for all the business types using Foursquare API. For simplicity, limit the venues fetched maximum to 100.

In the study we selected the neighborhood ‘Adelaide,King,Richmond’. From the business perspective it means this is the area we analyze in order to setup our business. To evaluate the 500m perimeter to retrieve the business details we used Foursquare API. (‘ <https://api.foursquare.com/v2/venues/explore?&client_id=NZG3JJSPA2TKFZTUGIITA3ZWMBHL45KKUSXBVUAJHNKONKYR&client_secret=GZ33INJ512B130DIBZJ424041FTAXGMM3SIPZHRBCQWYTNTE&v=20180605&ll=43.65057120000001,-79.3845675&radius=500&limit=100>’).

### Step – 2

Filter the required details such as venue name, venue category, venue’s latitude and longitude returned from Foursquare API.

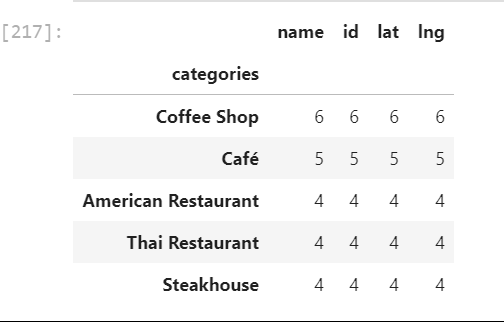
The resulting output after preprocessing the .json file is as below. As you can see you there are business venues in different categories with respective geospatial data.



### Step – 3

Consolidate and summarize the venue categories returned from the API. This will give you the number of businesses in each category in the given neighborhood. (ex. 6 Coffee Shops, 5 Cafes etc.)

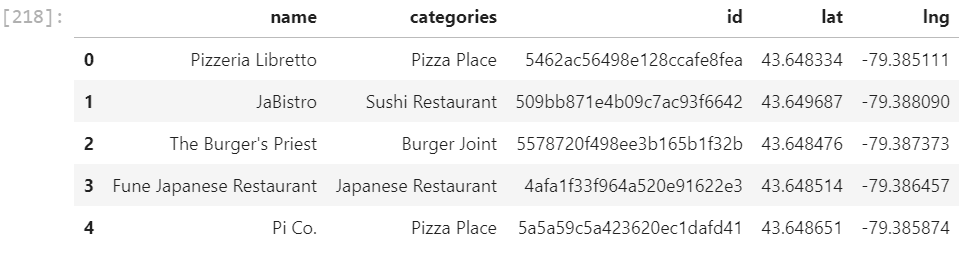
The returned data is as follows. This way you can group the venue categories in the region



### Step – 4

Define a threshold level. If the number of businesses in the given category exceeds the threshold level in the neighborhood we will not consider setting up a similar business. (i.e. In simple terms, if we have more than 2 (threshold) Coffee shops in the neighborhood we will not consider in setting up a new coffee shop. Because the assumption is there are enough coffee shops in the region if it exceeds threshold) – Remember the solution approach? - Fulfil the needs scarcest

When considering the above condition, we achieve following results. For the simplicity we limit our choice to first 5 categories fetched.



As an example, you can see we can start Pizza place, Sushi restaurant etc. as they fall into our threshold condition. In summary such categories are as follows.

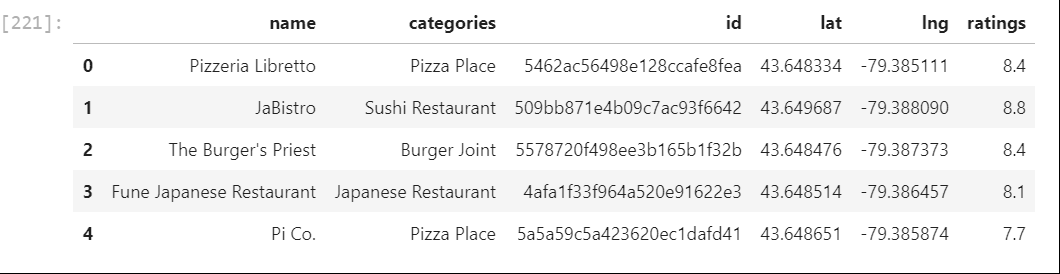


Please note that we take only the first 5 records for the simplicity. So now you know what businesses you can start in the region since there are not ‘much places’ in the vicinity.

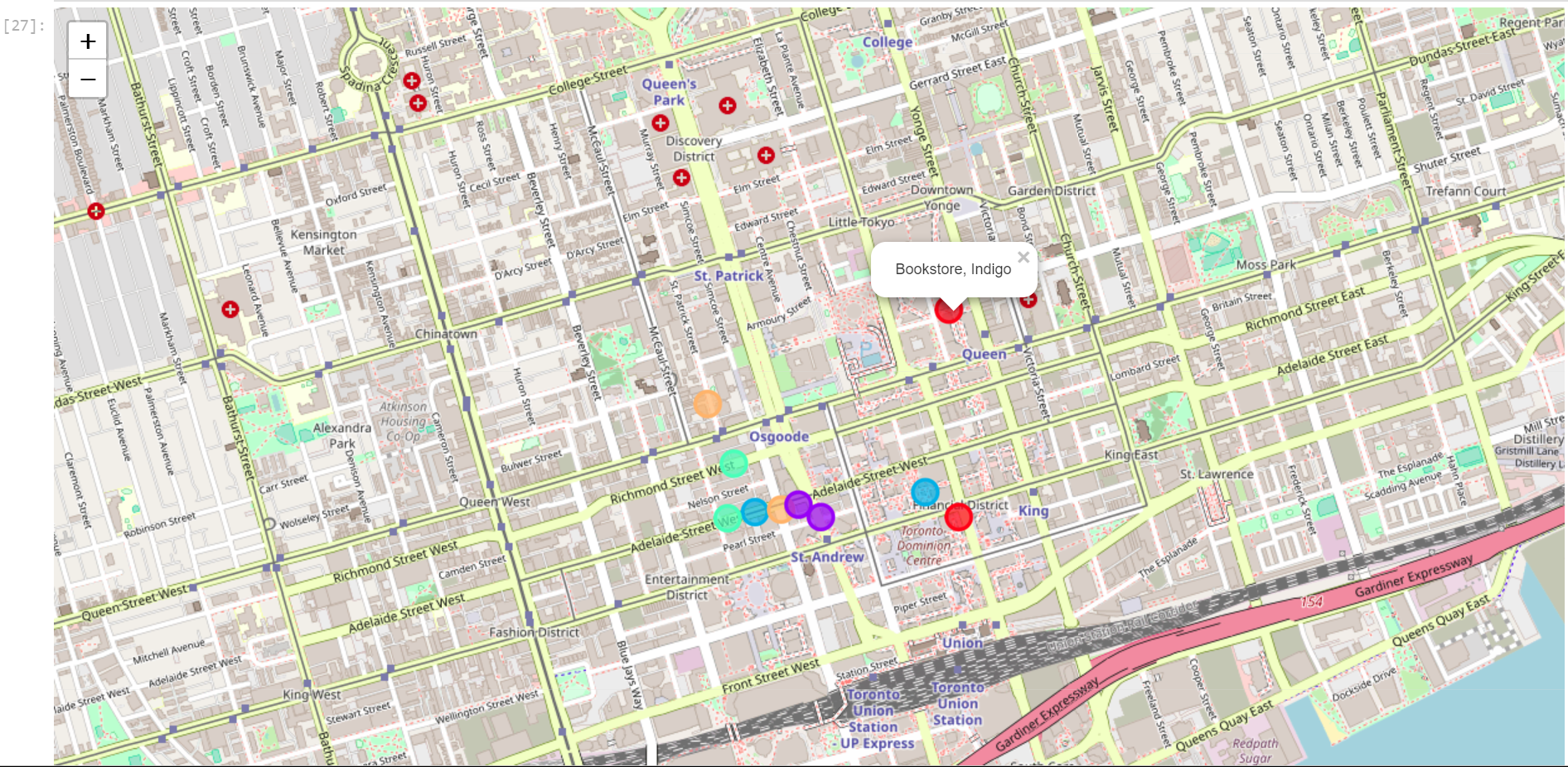
### Step – 5

Visualize the venues on the neighborhood map under different categories so you see the spread of businesses on the landscape.

Map the businesses with above categories on the location map. See the data below.



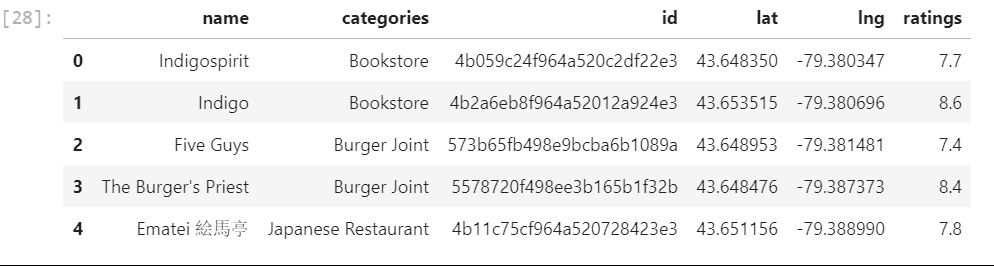
Visualize it on the map. You can see different colors assigned are for different categories. There are 10 business venues from our categories of interest.



### Step – 6

Get the customer rating for each business venue. Evaluate which business venue has the lower customer rating under each business type.

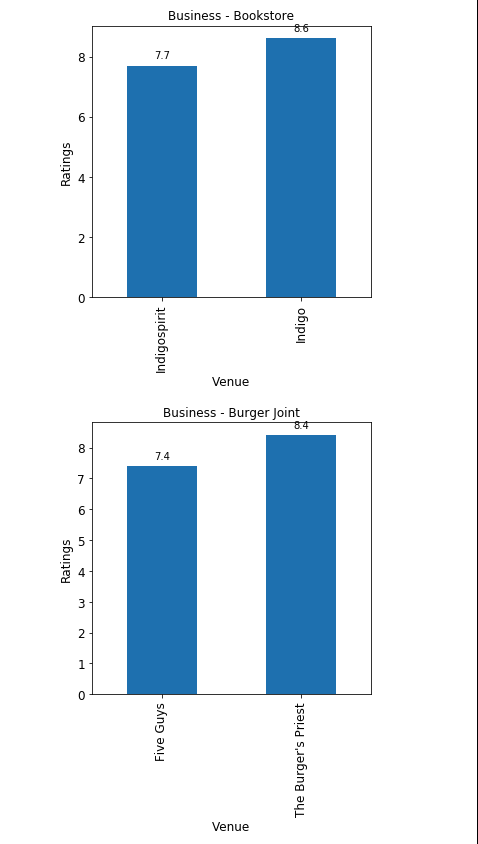
List the customer ratings for all businesses under the radar.



### Step – 7

Visualize lower customer rating under each category. Set up the operations (locate your office/shop) close to the venue where you have lower customer rating depending on your selected business type. Remember the solution approach? - use other’s lower customer rating to exploit your business opportunity.

For quick identification of the business venues with lower customer ratings you can use the following bar graphs as there is a graph for each business type. Only 2 graphs were selected into this report.

If you want to start a book store you need to position it close to the location where the ‘Indigosprint’ located. The ‘Indigo’ bookshop already has a good reputation from customer rating perspective. Positioning close to ‘Indigosprint’ will help you to exploit its lower customer rating for your competitive advantage.

## Result

The answer we seek to our problem will be guided through above steps. It simply helps to figure out what scarcity you would possibly fulfill in a form of a new business and where to setup its operation to achieve a competitive advantage. An example scenario you can find in Step – 8 over how to interpret the results. Of course, you could refine the solution further to generate additional insights by drilling down the contents of individual customer reviews (sentimental analysis). But this study did not reach that level.

## Conclusion

The model can be directly put into practice by the anyone in our target audience. However, in real practice adding local business context awareness is useful on top of the result, so you adjust the output sensibly. This model implementation can be easily extended and adjusted to any geographical location where you have sufficient volume of data (from Foursquare API) for a credible result. However, building a feedback loop to evaluate the model efficiency over the time would be essential and critical for further development and sustainability of the model.

.