# Applied Data Science Capstone

## *Introduction*

This is a project for week 4 of IBM Applied Data Science Caption.
For this particular project I am about to implement a hypothetical scenario for a concept that there may not be enough Greek Restaurants in Toronto Area. It is widely known that many Greek people moved to Canada in order to have a job. Toronto and Ottawa are the two Canadian cities with most Greeks in Canada. Apart from their history Greeks are also well known for their particular sense of having fun. One of the most important things to succeed it is the good food. So for this project I am about to locate the areas of Toronto with the least Greek restaurants and try to recommend some areas that a Greek entrepreneur might open a traditional restaurant.

## *BUSINESS PROBLEM*

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Greek Restaurant in Toronto, Canada. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : Where is the best place to open a Greek restaurant in Toronto? This will happen by examining all the territories of Toronto and locate Greek restaurants that already exist.

## *TARGET AUDIENCE*

The target audience of this project are every person who wants to open a restaurant with Greek cuisine in Toronto.

## ***Data***

The most valuable thing that we need is good and accurate data. So to be precise, for this project there are two data sources that are going to be used. The first containing information about the different areas of Toronto (i.e. postal code, city areas and broughs). This dataset is available on Wikipedia and it also used in the previous weeks. This data are going to be used to navigate in Toronto with the different area names.
Once this is possible there will be the need of some extra information about each Neighborhood. This is information is necessary to find out how many restaurants are located in each area. Foursquare is the solution to this problem as it provides information for every area only by getting the coordinates od areas name.

So to sum up, this project requires:
-List o Neighborhoods of Toronto, Canada
-Coordinates of these neighborhoods
-Venue data related to Greek restaurants. This will help us find the neighborhoods that are more suitable to open an Greek Restaurant

## ***EXTRACTING THE DATA***

- Scrapping Wikipedia webpage to retrieve Toronto neighborhoods
- Getting Latitude and Longitude data of these neighborhoods via geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

## ***Methodology***

The first thing it needs to be done in this problem is the creation of the dataset on which the whole process will be based. The first features will be scraped from Wikipedia
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
(the first table) which contains Toronto's Broughs and Neighborhoods grouped by their postal codes.

The goal of this project is to find and cluster all the Neighborhoods with Greek restaurants in Toronto. In order to succeed such a thing apart

from the Wikipedia data there is a need of all venues and their coordinates. The coordinated of each borough can be retrieved by the use of python's geocoder package. So at this point the dataset contains Toronto's Broughs, Neighborhoods and their coordinates.

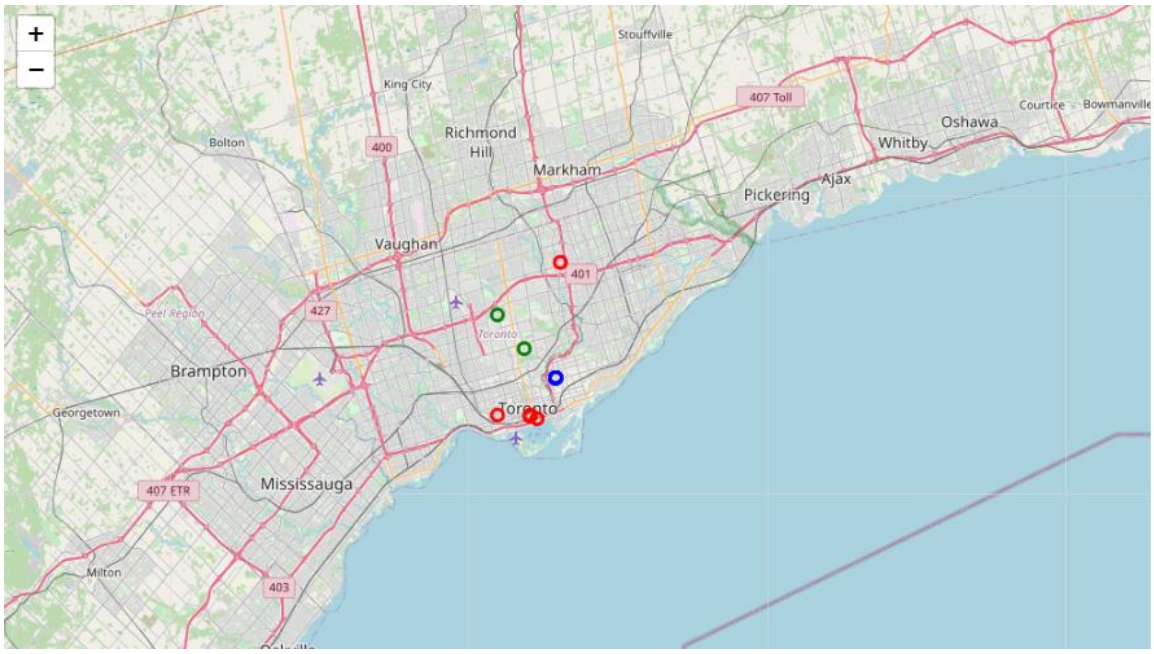
Now it is essential to locate all the venues in Toronto and filter them to find the Greek restaurants. This is possible to happen with the use of Foursquare, which provides all the necessary information for each of the available coordinates. All it demands is to send a request with the coordinates of each area set the range and it will return all the information as a json file. Foursquare is pretty easy to use as it has an API to do the hard job. The next step is to store the data in a data frame and apply one hot encoding to it in order to use it for the machine learning algorithm. Because the goal is to Cluster the Neighborhoods and not the restaurants the calculation of the mean value of each Neighborhood will help to succeed it (this is necessary because more than one restaurants might be in the same neighborhood). The next step is to get only the Greek restaurant column.

Then with a small dataset which only contains the Neighborhoods and the Greek restaurant density on each neighborhood it is time to fit the data in a ML algorithm to cluster the neighborhoods. The algorithm is K-Means with the number of k (number of clusters) equal to 3. So there will be 3 clusters of neighborhoods. So now for each neighborhood there is a cluster number that it belongs to.

Finally this cluster number is added as a separate column in the main data set with Neighborhoods, Venues and coordinates in order to visualize the clusters on a map and calculate which cluster contains the least Greek restaurants. So the cluster with least Greek restaurants will be recommended.

## ***Results***

The Results of this project can be visible by the following image.




The point of this project was to locate the areas with Greek restaurants. By this way we can locate where there are pieces of Greek culture in the city of Toronto and there it could be a good opportunity for someone o start his Greek restaurant. Even more clustering helped to recommend neighborhoods on which the competitions is the lowest possible.

Its seems that the most (7) greek restaurants are placed in cluster 1 . Up next there other biggest cluster is 0 with 5 reastaurants on Berczy Park, First Canadian Place, Underground city,Little Portugal, Trinity Toronto Dominion Centre and Design Exchange neighborhoods. Finaly the least restaurants 1 are placed in cluster 2 on Bedford Park, Lawrence Manor East and Davisville. So the most suitable place to open a greek restaurant is on Bedford Park, Lawrence Manor East and Davisville neighborhoods.

# Discussion

The following table contains information about the final merge dataframe with each row representing all the info for every Greek restaurant.

| | Neighbourhood | Greek Restaurant | Cluster | Neighbourhood Latitude | Neighbourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Berczy Park | 0.018182 | 0 | 43.644771 | -79.373306 | Alexandro's World Famous Gyros | 43.641663 | -79.375214 | Greek Restaurant |
| 1 | Fairview, Henry Farm, Oriole | 0.015152 | 0 | 43.778517 | -79.346556 | Jimmy The Greek | 43.778245 | -79.343322 | Greek Restaurant |
| 2 | First Canadian Place, Underground city | 0.010000 | 0 | 43.648429 | -79.382280 | Estiatorio Volos | 43.650329 | -79.384533 | Greek Restaurant |
| 3 | Little Portugal, Trinity | 0.023256 | 0 | 43.647927 | -79.419750 | Mamakas Taverna | 43.645908 | -79.419654 | Greek Restaurant |
| 4 | Toronto Dominion Centre, Design Exchange | 0.010000 | 0 | 43.647177 | -79.381576 | Estiatorio Volos | 43.650329 | -79.384533 | Greek Restaurant |
| 5 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Pantheon | 43.677621 | -79.351434 | Greek Restaurant |
| 6 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Mezes | 43.677962 | -79.350196 | Greek Restaurant |
| 7 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Messini Authentic Gyros | 43.677704 | -79.350480 | Greek Restaurant |
| 8 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Christina's On The Danforth | 43.678240 | -79.349185 | Greek Restaurant |
| 9 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Astoria Shish Kebob House | 43.677596 | -79.351738 | Greek Restaurant |
| 10 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Alexandros | 43.678304 | -79.349486 | Greek Restaurant |
| 11 | The Danforth West, Riverdale | 0.166667 | 1 | 43.679557 | -79.352188 | Athen's Pastries | 43.678166 | -79.348927 | Greek Restaurant |
| 12 | Bedford Park, Lawrence Manor East | 0.043478 | 2 | 43.733283 | -79.419750 | Karbouzi Greek Taverna | 43.736204 | -79.420359 | Greek Restaurant |
| 13 | Davisville | 0.027778 | 2 | 43.704324 | -79.388790 | souvlaki express | 43.707378 | -79.389848 | Greek Restaurant |

| | | Greek Restaurant |
|---|---|---|
| Cluster | Neighbourhood | |
| 0 | Berczy Park | 1 |
| | Fairview, Henry Farm, Oriole | 1 |
| | First Canadian Place, Underground city | 1 |
| | Little Portugal, Trinity | 1 |
| | Toronto Dominion Centre, Design Exchange | 1 |
| 1 | The Danforth West, Riverdale | 7 |
| 2 | Bedford Park, Lawrence Manor East | 1 |
| | Davisville | 1 |

## Conclusion

Baring in mind the above map and table the final recommendation for someone who wants to open a Greek restaurant in Toronto is cluster 2. Now as it concerns the neighborhood Bedford Park or Lawrence Manor East seems to be the best choice as their density value ('Greek Restaurant'

column) is the lowest of this cluster. **So the best choice for someone who wants to open a restaurant the best place are Bedford Park and Lawrence Manor East Neighborhoods.**