**Battle of the Cities & Neighborhoods - NY & Toronto**

**Introduction**

IT companies world-wide continue to face the challenge of shortages in skilled and productive Data Scientists to be deployed for AI, Machine Learning, & Deep Learning related projects that can truly deliver technology driven business value to enterprises and institutions globally.

To address this challenge some large IT companies have established Data Science and Cognitive Computing Centers in major cities in India and continuing to hire hundreds of new graduates to be on-boarded in each of these overseas centers.

One of the core requirements of the induction and Base level training program for the newly hired Data Scientists is to be able to bring them to the Large Strategic Training Centers in North America for a period of 6 months to enable the smooth delivery of domain and technology

training. Management has come to recognize that most of the

Data Scientists from India are vegans and require suitable and quality

living environment while they are in North America for their 6 months training. It’s important that their specific requirements with regards to food and other aspects of Vegans lifestyles are respected and enabled to ensure optimum learning and skills acquisition by the new Data Scientists.

IT companies face the "Business Problem" of helping the new

Data Scientists in a logical manner with their choice of either to live in NY or Toronto which are the two major training centers in North America. It is envisaged that relevant recommendations capturing the core features (i.e. Vegan restaurants, & Community Centers) of the 2 cities with respect to Vegans Lifestyle would be very helpful for the new-hires.

Alternatively, the problem statement can be captured by the Question below...

" How can IT companies help their overseas Vegan staff to choose between which Neighborhoods in NY / Toronto to live during their 6-months

training? "

The core objective of this Data Science project is to conduct a

neighborhood comparative analysis for NY and Toronto with specific emphasis on the 2 lifestyle features of "Eating Places/restaurants" and "Community/Social Centers" and enable relevant & valuable recommendations about the City.

### ****Trusted Data Sources****

The following trusted data sources will be used to conduct Data exploration and other forms of data analysis as part of the methodology for the project.

1) NY Neighborhood Data - https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json & https://geo.nyu.edu/catalog/nyu\_2451\_34572

2) Toronto Neighborhood Data - 'https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M'

3) Foursquare Location Data (Venues [Restaurants, Community Centers], TOP tips, Favorites, User Experience, etc) will be used to cluster, segment, target, and position to craft recommendations for the vegans end-user community.

Using the data available in the above 3 trusted sources, I will be conducting clustering and neighborhood based analysis leveraging primarily Foursquare APIs and tools such as KNN and relevant unsupervised learning methods to deliver recommendations options to the target user community primarily comprising of the newly hired vegans.

Pls see below some relevant examples of tailored data that will benefit the vegans and enable them to make meaningful decisions whether to live in NY or Toronto and in which TOP 5 Neighborhoods within each of the cities.

1) There are many vegan eating places in NY with the top rating being 9.3 for abcV Restaurant.

2) Similarly there are a large number of vegan eating places in Toronto with 'Fresh on Bar" scoring 9.0.

3) There are also tens of Neighborhoods in both cities that truly deliver pristine eating and relaxing environment for vegans.

4) Leveraging the recommender systems capabilities enabled by Foursquare to conduct compare / contrast the data points to derive the recommendations.

Let me also provide below the specific steps that will be deployed to work with the data derived from Foursqaure and the approach will cover both Toronto and NY and also the category of "Vegan Community Centers"

A) Search for a specific venue = "Vegan Restaurant"

B) Perform the GET Request and examine the results for Vegan eating places

C) Get relevant part of JSON and transform it into a pandas dataframe & review the data for patterns and obvious insights

D) Explore Venues (= Vegan related) with Rating of 8.5 (out of 10) or more

E) Get the number of tips for each Venue and focus on what could be a source of valuable information for the Data Scientists (new-hires)

F) Gather frequent user feedback and comments

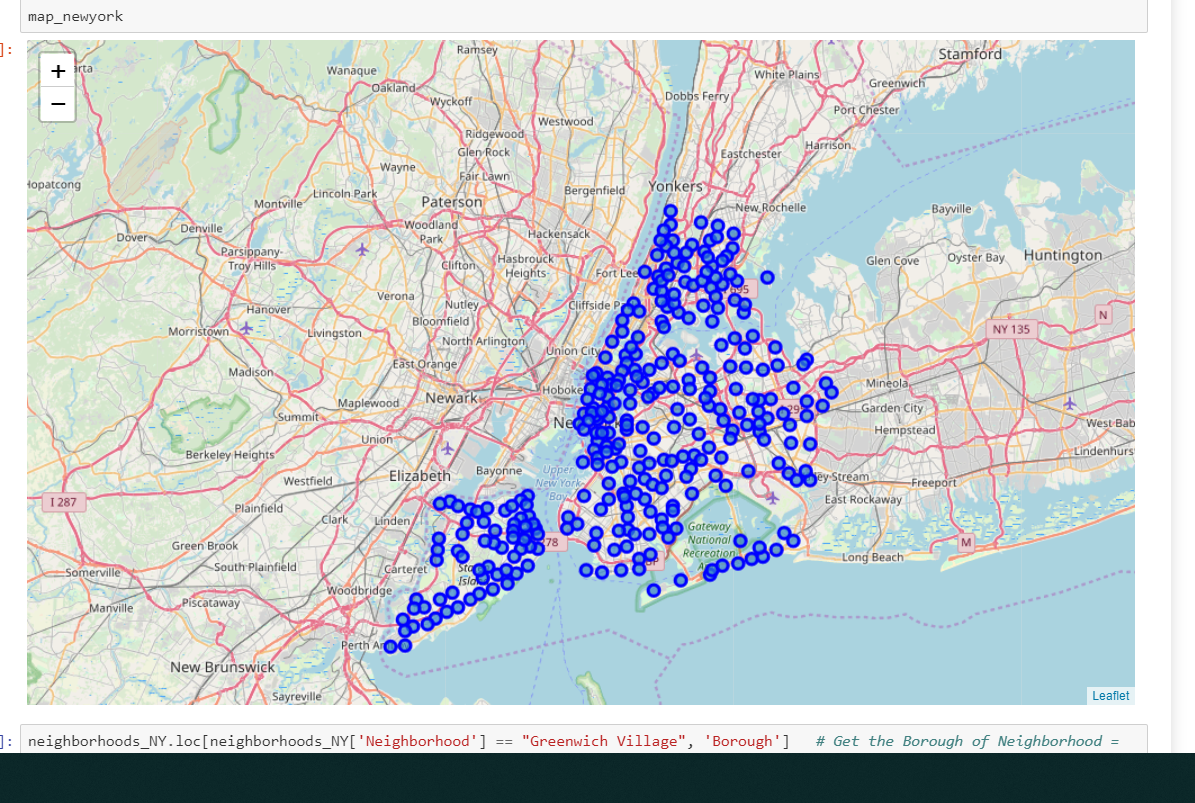
G) Narrow down the venues and the associated neighborhoods for each city that can form the set of recommendations

### ****STUDY****

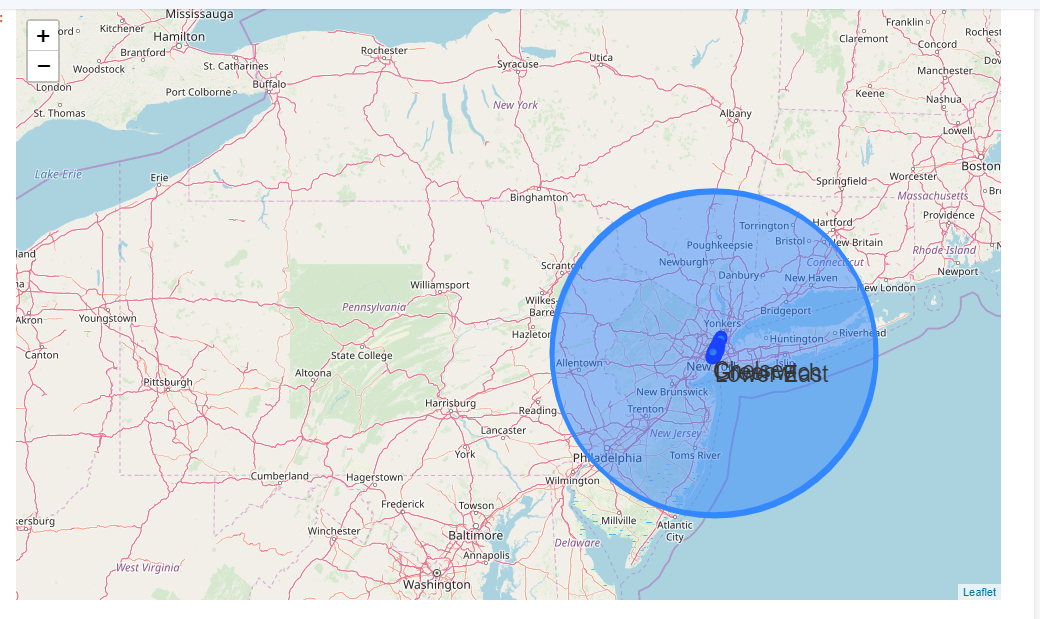
The methodology employed is aligned with the Data Science 10 steps program that was discussed during an earlier module of thecourse. Let me outline below the core areas of focus and tasks performed as part of the methodology to address the problem of "helping the Vegans" for make the right choice regarding their training centers in NY and Toronto.

Obtained relevant vegan places data from Foursquare and clean it for data understanding and grouping etc. Explored the data for clusters & patterns of Neighborhoods in Boroughs. Grouped the places of high interest into relevant neighborhood and Borough pairings. Generated the markers on both NY & Toronto maps to highlight the neighborhoods that are in focus. Used Foursquare venue, Categories etc to enable the profiling the primary & focus neighborhoods. Generateg the short-listed neighborhoods for both NY & Toronto

## Map of Vegan eateries in New York



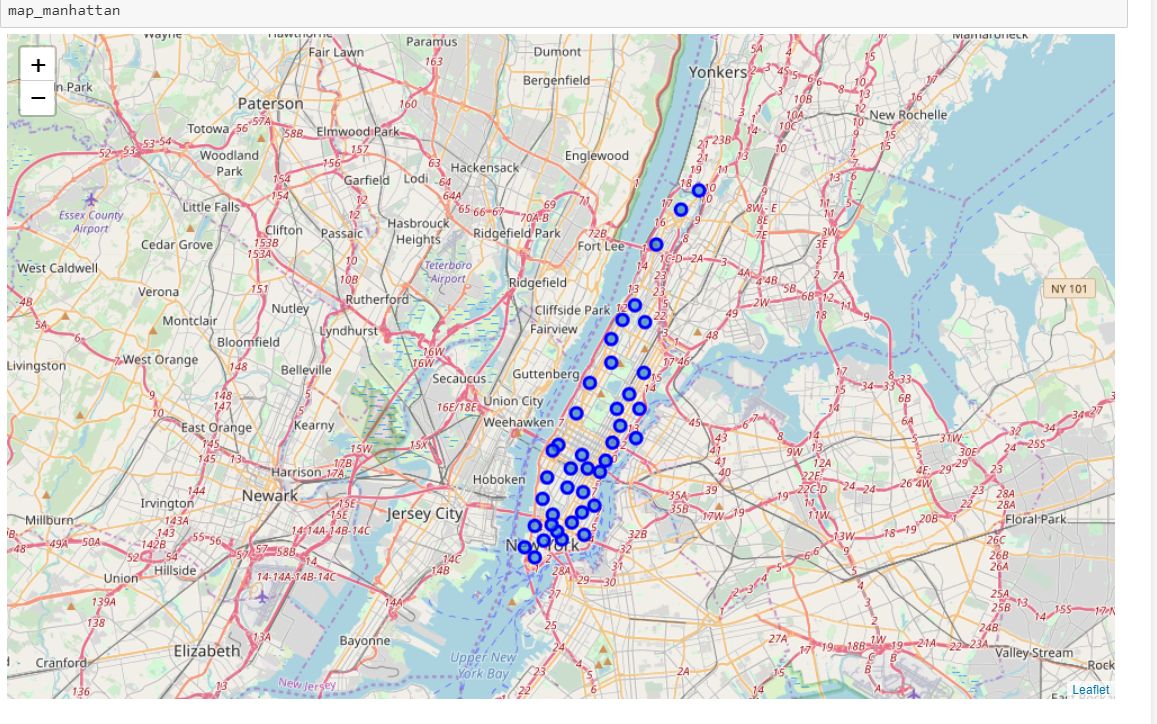
## Map of Vegan eateries in Toronto



### ****Results****

It was pleasing to see that there were specific neighborhoods that truly catered for Vegans and have gained high levels of reputation for Vegan friendly services and customer care.

# Vegan Restaurants in Manhattan



## Here is the short-list of neighborhoods for Vegans in NY

Chelsea and Clinton 6, Gramercy Park and Murray Hill 1, Greenpoint 2, Greenwich Village and Soho 5, Lower East Side 2, Northwest Brooklyn 3, Upper East Side 1

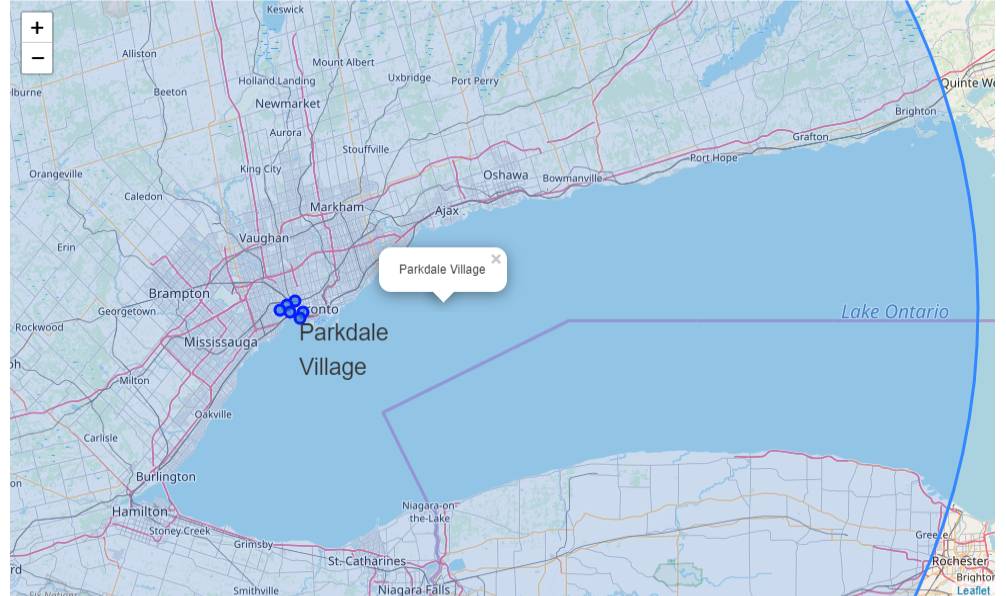
### It was clear that Greenwich, Chelsea, and Northwest were highly favored for Vegans in NY

## Let me now summarize the findings for Toronto

### Get the TOP 3 Vegan Places

Bloorcourt, East Toronto, Kensington Market, Parkdale Village

# Map of Top 3 Vegan Restaurants in Toronto



### It was clear that Parkdale Village is highly favored for Vegans in Toronto

While this project was able to derive specific recommendations in terms of places to live in NY & Toronto for Vegans, it has limitations given that the assumptions for the model are simplistic in nature. The main variable of focus was only Food & Leisure where as Vegans may be interested in a number of other amenities and interests that will make it a multi-variable focused data science analysis. From my perspective the project has afforded me a tremendous opportunity to explore and apply most of the knowledge and insights gained during the previous 8 modules of this course.

### ****Conclusion****

It is very evident that a Data Scientist will need to establish clarity and crisply identify the "Business Problem" or “Idea of Research" that he/she is undertaking and be able to articulate clearly the value being delivered to the beneficiaries / stakeholders / end-users etc.

Once the Business Problem / Idea has been confirmed, its also equally important to explore / research for the trusted data sources that can be leveraged to undertake the Data Science framework-based problem solving.

Next comes the critical step of choosing the methods (Unsupervised or Supervised and within each of them the specific methods for machine learning / model depevelopment and training etc ) for data analysis/mining/patterns recognition etc...

It is also vital that relevant presentation and visualization techniques are used to share and present the findings to the project sponsor / stakeholders / end-users etc. It may be also appropriate to comment any future work that may be valuable to augment the solution and enrich it.