Alyssa Mendenhall

3/13/20

Yoga in Toronto

IBM Capstone

**Table of Contents**

[1. Introduction 2](#_Toc34990784)

[2. Business Problem 2](#_Toc34990785)

[3. Data Collection 2](#_Toc34990786)

[4. Methodology of Analysis 3](#_Toc34990787)

[5. Results 4](#_Toc34990788)

[6. Discussion 6](#_Toc34990789)

[7. Conclusion and Future Directions 6](#_Toc34990790)

[8. References 7](#_Toc34990791)

# Introduction

As more individuals learn about the mental and physical benefits of mindfulness and yoga, studios to practice these activities are becoming more common. Although almost everyone can benefit from yoga, many do not have the resources to partake in formal yoga classes. Thus, in order to open a successful new yoga studio in Toronto, and analysis must be undertaken to understand what neighborhoods a venue such as this would succeed in. This can be a difficult task. In rural areas, individuals may have to travel longer distances to participate in a yoga class. These areas would be sparse in existing yoga studios due to the hassle of travel to them. However, the opposite could be true of more urban areas. With the activity becoming increasingly popular, it could be difficult to enter the market and succeed without becoming crowded out by other studios. Thus, an ideal location for a yoga studio would likely be a place where yoga has already made an impact (is somewhat popular as a venue) but is not the most or second most popular activity in the area. This project is most relevant to those who are passionate about yoga but practical in understanding that in order to succeed, they must choose a location suitable to their business.

# Business Problem

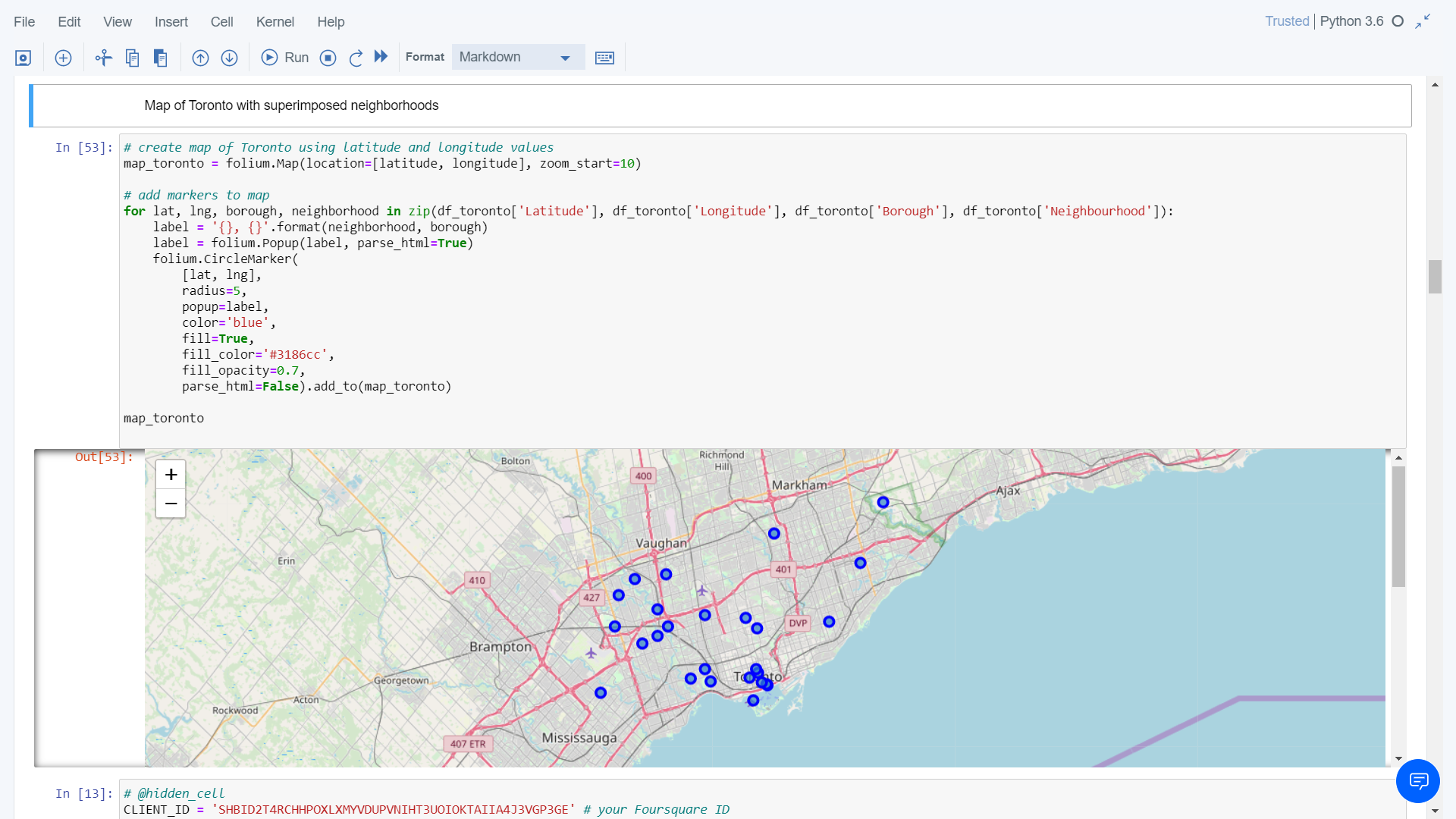
As a potential small business owner, the goal of the creation of a yoga studio is, in addition to providing a service, creating a source of revenue for the owner. In Canada, it is estimated that the creation of a studio, including employee salaries for the first 3 months, is approximately $1 million CAD[1]. Thus, such an investment of capital would need to be made in an area that is conducive to its success. Toronto is a sizable city in Canada that contains a diverse population of individuals. Its neighborhoods vary greatly in their venue composition. Exploring neighborhoods in this city as possible locations for a studio is the first step to determining if a location in Toronto matches the ideal conditions for a yoga studio. An oversaturated neighborhood, such as one that has yoga studios as its first and second most common venues, would be unlikely to be able to support another studio. However, an area with no yoga studios would be more difficult to attract customers to. Creating a yoga studio in an area where a yoga culture has been established, but not oversaturated, would result in the highest net profit and therefore a more successful business.

# Data Collection

The data used for this project will be collected from Foursquare, a user-sourced dataset that is commonly used for venue data[2]. To determine what neighborhoods will be used for this problem, a table of neighborhoods and their respective postal codes will be scraped from the Wikipedia page[3]. Then, this data will be merged with the location data of these neighborhoods using Google Maps Geocoding API. Using Foursquare combined with the neighborhood locations, a list of nearby venues will be obtained. From a one-hot matrix of these values, the most common venues can be procured.

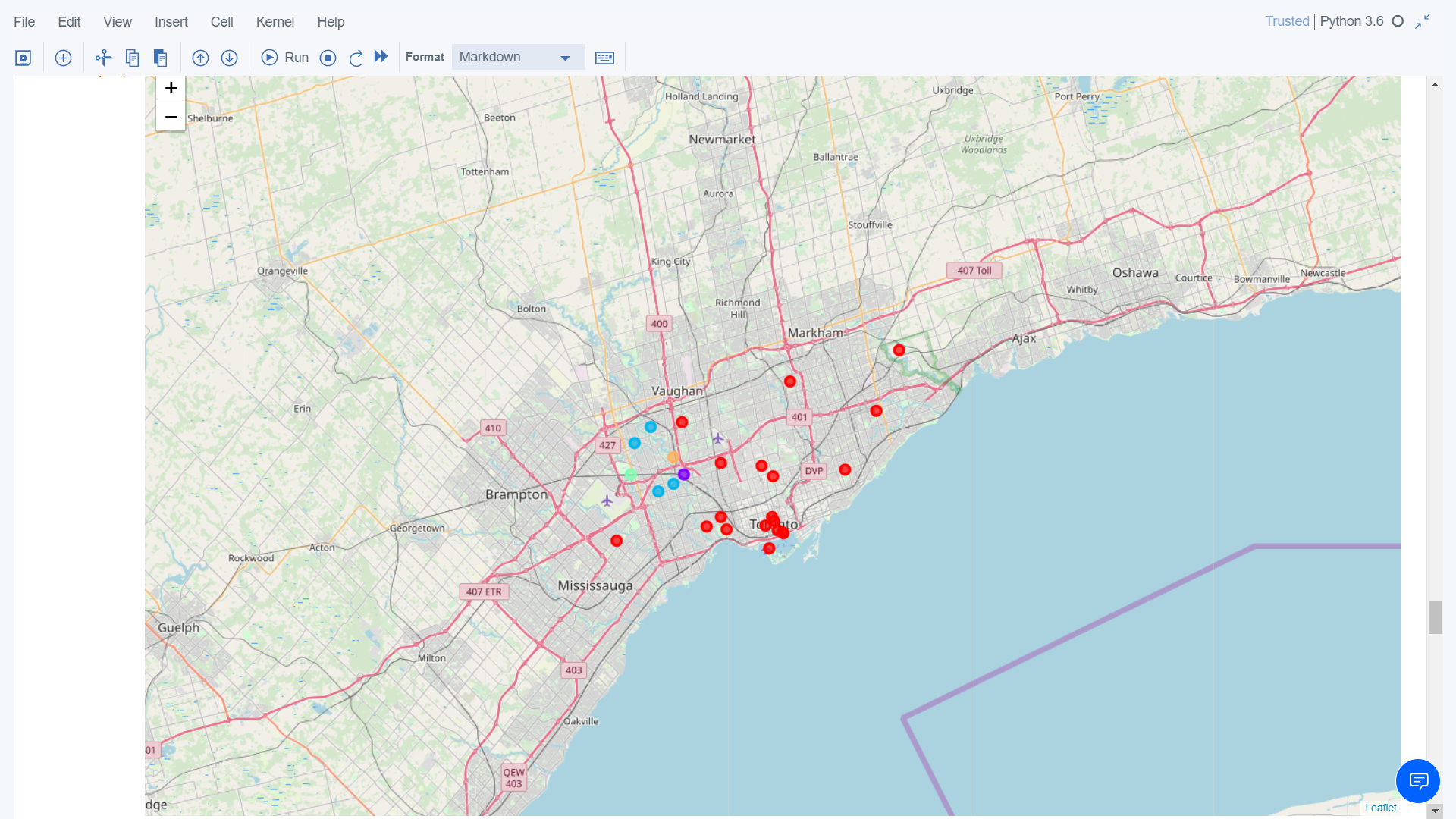
# Methodology of Analysis

To ensure data, methods, and principles accuracy, a thorough analysis of the methods and postulates made have been used. Since the Geocoding API is prone to erroneous readings, a second csv file of curated data was cross referenced for the longitude and latitude coordinates. Initial data visualization was performed using folium to determine neighborhood locations as well as spot outliers in data collection that might have led to erroneous results as shown in Figure 1. Folium combines two powerful ecosystems, Python and Leaflet, in order to create accurate visualization maps that are simple to create and update. This visualization package was also used to visualize the results of the cluster analysis to give the user a sense of how the venue similarity of each neighborhood correlates with its geographic location. Please refer to the included Jupyter Notebook for the documented breakdown of the use of these tools.



**Figure 1:** Map of the Toronto area neighborhoods used in the analysis.

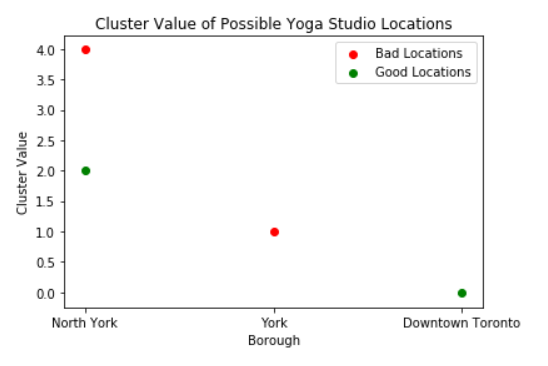
A k-Means analysis was performed in order to group neighborhoods according to their similarity in venue composition. This analysis was performed due a variety of advantages such as the data set size, which is moderate, is ideal for the kMeans cluster analysis. Additionally, the algorithm is very efficient, allowing for optimization of the value of k with relatively little time and computational power. Upon performing this analysis for the range of number of clusters tested, it was found that 5 clusters produced the groups with the most intra-cluster similarity and inter-cluster differences. Thus, the analysis was performed using 5 as the number of clusters. This analysis is documented in the associated Jupyter Notebook, which should be referenced for more detail. These clusters were then visualized using Folium, and the top 5 most common venues was assessed for each neighborhood in the cluster (Figure 2). Optimal clusters were those that contained yoga studios as the 3rd-4th most common venue. Clusters with yoga studios as the most and second most common venues were removed from consideration.



**Figure 2:** Clustered Map of the Toronto Area

# Results

By examining data values after the processing described above, I found 2 locations that would be optimal for a yoga studio, as shown in Figure 3.



**Figure 3:** Good locations for a yoga studio (those with a moderate amount of yoga venues already in existence) are located in North York and Downtown Toronto.

The North York and Downtown Toronto neighborhoods would be ideal locations for a budding entrepreneur to create a yoga studio. There is some interest in yoga in these locations; however, the market is not oversaturated. In these locations, the most popular or second most popular venue in the area is not a yoga studio, unlike in York. Some areas of North York may be oversaturated, as shown by the red point in Figure 3. Thus, Downtown Toronto would be the best neighborhood for an individual to create a yoga studio.

If other factors were present that did not allow for Downtown Toronto to be the location of interest, this analysis also aids potential studio makers in selecting an optimal secondary neighborhood for their studio. By also performing a cluster analysis of the neighborhoods of Toronto using the kMeans clustering algorithm, possible secondary locations that would be optimal for yoga studios can be generated. This allows for generalization of the data even if it is incomplete. Assuming that communities that support yoga studios also support a similar spread of venues, if Downtown Toronto is a less than ideal location for reasons other than venue popularity, choosing a location within a similar cluster, cluster 0, would be the next best solution. A breakdown of cluster 0 is presented in Table 1, showing yoga studios as being somewhat popular amongst a select group of other neighborhoods.

**Table 1**: Breakdown of the first cluster (Cluster 0) by frequency of yoga studios in the neighborhood.



# Discussion

In an ideal situation, a yoga enthusiast would be able to create a studio in Downtown Toronto with a reasonable amount of starting funds. However, in the real world, other factors are at play. This analysis aims to give a limited perspective on the popularity of yoga studios as a venue in Toronto. Before action would be taken, it is recommended that yoga practitioners conduct their own analysis on other impactful factors in a successful business.

If an individual were unable to create a yoga studio in Downtown Toronto, this analysis would also be beneficial in determining an appropriate second choice of location. As shown in relation to Table 1, consideration of other venues within the same cluster as determined by kMeans cluster analysis gives a fair amount of other locations, labeled in green, where yoga studios rank within the top 10 venues, without being one of the first two venues. This allows for generalization of the data even if it is incomplete. Assuming that communities that support yoga studios also support a similar spread of venues, if Downtown Toronto is a less than ideal location for reasons other than venue popularity, choosing a location within a similar cluster, cluster 0, would be the next best solution. Again, this analysis is limited in that it only considers frequency of venues within a neighborhood, and not additional factors such as cost or access to public transportation

In order to produce a conclusion, this analysis makes a few assumptions. First, this model assumes that a neighborhood with yoga studios as its first or second most common venue is oversaturated. This doesn’t take into account growing communities that could have a real use for additional yoga studios or size of the studio itself. Size of the yoga studio could be important because many smaller yoga studios may hold the same patronage as one large studio, thus skewing the data towards assuming an area is oversaturated or undersaturated.

Additionally, this model assumes that areas that do not have yoga studios within the top 10 most common venues don’t have an interest in yoga. It could be the case in such an area that individuals are not familiar with yoga but would practice if they were exposed to it. Additionally, these individuals may already practice yoga at home, but are unable to find a studio that fits into their lifestyle. In these instances the model would break down, so additional considerations should be taken into account before any decision is made regarding yoga studio location.

# Conclusion and Future Directions

Although it is a simple model, this analysis demonstrates that Downtown Toronto, or other locations within the first analyzed cluster, would be ideal locations for an individual looking to start their own yoga studio. These locations would likely meet the need of a neighborhood where yoga is popular, but not an oversaturated market.

However, this model, like any, has its assumptions and limitations. This model assumes that a moderate frequency of yoga studios is optimal, with the high extreme displaying an oversaturated marked and the low extreme showing lack of interest. Additionally, this model is limited to frequency analysis of venues within the Toronto area. This area could be expanded, or other factors such as area costs and access could be analyzed to overcome these limitations.

Before making a final decision regarding yoga studio location, a broad spectrum analysis of many factors should be performed. In addition to the presented analysis, it is recommended at minimum that entrepreneurs explore variables such as local public transportation access, especially in such an urban area as Toronto. An analysis of costs within the area would likely be of use as well. These additional analyses would give the individual a larger picture of the many complex factors at play in the creation of such a high value business as a yoga studio.

# References

[1] Martins, A. T. (2020). Cost to Start a Yoga Studio in Canada. (accessed 11 March 2020).

[2] Foursquare (2015b) Foursquare for Developers – Venues Service. Available at: https://developer.foursquare.com/overview/venues.html (accessed 11 March 2020).

[3] List of postal codes of Canada: M. (2020, February 27). (accessed 11 March 2020).