**Introduction/Business Problem**

As a growing tech company in Vancouver, you are tasked with opening a new office on the other side of Canada. Moving across the country can be difficult for your employees, so you want to make the transition as painless as possible. Because your employees are generally young and active, you will need to find the best possible neighborhood for your new headquarters. The Chinatown area where you are located is young, dynamic, and fun, with a great food scene. The challenge will be to recommend a new location to your boss that best matches the current location where the employees are happy.

**Data**

The major data source for this project will be foursquare. By looking at the venues we can create a profile. First, we will clean up the venue data a bit. For example, “burrito place” and “Mexican restaurant” are categorically similar enough to be grouped under a single label. Next, we will create a detailed profile of our current neighborhood. To do this properly we will need to also perform some qualitative analysis. It’s not enough to just have a Japanese restaurant in both places, if the one in Vancouver is exceptional and the one in Toronto isn’t. Once we have our detailed profile, we will then map that profile onto the different Toronto neighborhoods to come up with the best possible fit.

**Initial process**

The first step was to categorize Chinatown. This was done by using Foursquare. We pulled their venues and built a profile of our neighborhood:

Text

Description automatically generated

We then did a similar process for all the neighborhoods in the Toronto area and plotted them on the map:

Map

Description automatically generated

We then added our neighborhood to this data set to see where it would cluster. The initial results were a little less than encouraging. The first clusters were very apparent, i.e. venues that included Airport related things, or stadium-related venues were easily clustered, but it became hard to continue to distinguish or eliminate areas.

Diagram

Description automatically generated

We took the above 9 neighborhoods and tried to better digest the data to create a better dataset for re-clustering. After re-clustering we finally narrowed it down to two:

Map

Description automatically generated

**Conclusion**

The final assessment was that either the Garden District or Hardbord would be good places to look. While this was able to very quickly identify these two places, there are some drawbacks to taking this approach at face-value. The conclusion was solely based on data provided from Foursquare. The entire neighborhood of Chinatown had just 22 venues, which isn’t really enough to form a full picture of the area. Additionally, the categorization of the data misses a bit of data underneath, such as “Asian Restaurant” vs “Thai Restaurant” or “Chinese Restaurant.” For these reasons, it was worthwhile to confirm that the data was properly matching Chinatown to these two neighborhoods. Sure enough, both areas are young, dynamic, and have bustling food scenes and night life. In this case, k-nearest neighbor did indeed identify the nearest neighborhoods from across a continent.