**Final Project of Applied Data Science Capstone**

**- Find a new living area in Toronto**

1. Introduction

It is quite common to move to a new place for various purposes, such as get a new job, go to college, or even for retirement. However, it is difficult to foresee the living situation in the new places, especially for a totally strange place that you have never been to or even in another country. Although there is a lot of information on the internet that you can access to evaluate the city or county that you are moving to, you cannot collect all the details in an easy way since the information is mostly fragmented.

Essentially, the quality of life (QoL) can be determined by many factors. For example, how far the supermarket is? Is there a hospital near me? Will it be loud in the night due to the prosperous bar business? Do I have the access to any green area like parks or river for jogging? All these things can play a role in influencing the QoL when people live in a place. Nevertheless, the apartment or the house itself is also important, but that is out of the scope of this project.

Therefore, a quick and thorough way to gather all the necessary information and then present it to the people with the need before they move would be a great help in this context. The aim of this project is to preliminarily present a simple system to help the users choose their future residential area based on their subjective requirement. Based on their requirement, the system will go through the internet to retrieve all the locational information and categorize them into clusters so to recommend some suitable areas to the users.

To make the project closer to the realistic scenario, we assume that we are a real estate agency and providing the service to help our clients find their ideal residence in a brand-new area. Here is requirement from our client, David, who just got a new job in Toronto and will move there within 2 months. Therefore, he contracted us with this task to find an ideal place and an apartment for him.

David is a software engineer. He works mostly from home and only needs to go to the office for meetings sometimes. Therefore, the distance between his home and office is not that important to him. However, the quality of life he will have in the new neighborhood.

David doesn't like to cook because he hates the cooking smell inside the house, so the most important thing for him is to have access to any restaurants right down the street within 200 meters. In addition, he likes drinking coffee and working out in a gym. These two are equally important. However, David is allergic to many things so it would be better to avoid living in an area close to parks, flower stores, pet stores, or something similar. One more thing,

2. Data for this project

To help David find the ideal place to live, we will need the facilities data from Toronto, which we will get via Foursquare. After retrieving the data, we will perform a clustering process to filter the top 3 areas for David to choose and that will be the end results of this project.

3. Methods

In this sectoin, we will implement the following steps:

Retrieve necessary location data from Foursquare

Clean up data

Perform analaysis, clustering, and visulization

Find the top 3 areas that match the customer's needs