1. **Introduction**
   1. **Background**

Suppose a client Ava got two job offers from two different companies. Company A’s office is located in Toronto and is approximately 20 min drive away from Ava’s apartment. Ava lives in the central bay street neighborhood in Toronto. Company B’s office is in Vancouver, which means Ava has to move to Vancouver in order to work in Company B.

However, Company B is offering a higher salary to Ava compared to Company A. One thing that makes Ava hesitate to move to Vancouver is the convenience of living in her current neighborhood, which has various types of venues around and makes her life much comfortable and enjoyable.

* 1. **Problem**

If Ava were to move over to Vancouver, she would like to find a neighborhood around Company B (maybe 20 min drive away) that’s similar to her current one, so as to achieve a good work-life balance as she does now. Thus, this project aims to look for a neighborhood around Company B in Vancouver that’s similar to the central bay street neighborhood in Toronto in terms of the venue categories.

1. **Data**
   1. **Data Sources**

Dataset of information about neighborhoods in Vancouver such as their names and geocodes came from the dataset called “Local area boundary” on the website:

<https://opendata.vancouver.ca/explore/dataset/local-area-boundary/table/?location=12,49.2474,-123.12402>.

Geocode of the central bay street neighborhood could be found in this file <https://cocl.us/Geospatial_data>.

Data of information about all neighborhoods in Toronto are scraped from the Wikipedia page: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Information about venue categories of neighborhoods can be obtained through API calls made to Foursquara.com.

* 1. **Data Description**

Data scraped from the Wikipedia page contains missing values, for example, some postal codes were not assigned with any borough’s or neighborhood’s name. Those rows with data under ‘Borough’ not assigned were removed. In order to get the geocodes for the neighborhoods, the csv file <https://cocl.us/Geospatial_data> was imported and got combined with our first data frame.

Dataset of neighborhoods in Vancouver was downloaded directly from opendata.vancouver.ca. as a csv file. Only columns of the neighborhoods’ names and geocodes (latitude and longitude) are retained. There were 22 neighborhoods of Vancouver in the dataset. Their geocodes would be put into API calls to the Foursquare website to get nearby venues.