**Evaluation Goal**: What are the top 5 most common business types of each local area in Vancouver?

**Introduction**

Our original plan was to build a machine learning application that would predict the type of the business based on its name. To do this, we first put all of the business names from the given CSV files into a bag of words. We treat this as a classification problem, using the business names to predict the business types. However, there are more than 60,000 businesses, and there are more than 40,000 words in our bag of words. Thus, we will need to make a matrix of size 60,000 by 40,000 to accommodate for such data. Based on such a big matrix, it takes approximately 19GB of RAM to run, which our computers cannot handle. In addition, since we are using Matlab, we cannot make good use of IBM’s Data Scientist Workbench.

As a result, we approached another goal: Of each of the local areas in Vancouver, what are the top 5 most common business types?

**Evaluation Method**

Here is a rundown of our program:

1. For each CSV file, we retrieve the local area and business type columns.
2. We clean and remove any empty rows or cells and businesses operating outside of Vancouver
3. We count the number of business types located in each local area
4. We then sort the business types based on their frequencies
5. Finally, we record these results in separate CSV files

**Findings**

We observed that the Vancouver is dominated by offices and residential business types. In the case of the west end and downtown areas, restaurants are more prominent. There are computer service businesses in downtown, which could be an indication of a growing computer software startup industry. Areas with apartments and dwelling licenses are observed to be more residential. Over the years, we noticed that in some areas, there is an increase of secondary suites – this could imply that, with regards to the escalating prices of real estate, more people are opting for less expensive housing.