Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

1. In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea.
2. In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

**Review criteria**

This capstone project will be graded by your peers. This capstone project is worth **70%** of your total grade. The project will be completed over the course of **2 weeks**. Week 1 submissions will be worth **30%** whereas week 2 submissions will be worth **40% of your total grade**.

For this week, you will required to submit the following:

1. A description of the problem and a discussion of the background. (**15 marks**)
2. A description of the data and how it will be used to solve the problem. (**15 marks)**

For the second week, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. (**15 marks**)
2. A full report consisting of all of the following components (**15 marks**):

* Introduction where you discuss the business problem and who would be interested in this project.
* Data where you describe the data that will be used to solve the problem and the source of the data.
* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.
* Results section where you discuss the results.
* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
* Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (**10 marks**)

**Description of the problem**

* Can you combine it with something?
  + Look at this paper where they link it to ambulance calls and show that there is a trend between daytime check-ins, ambulance calls, and the types of calls they see based on the things in that area - <https://arxiv.org/pdf/1801.09524.pdf>
* An idea might be to measure a cities walkability by measuring the total number of services provided and their density, such as by seeing how many services are available in a city based on its density, and then see if cities that have a higher density of services have less chronic health issues
  + The problem with this is that if the calls on foursquare would be exceeded by trying to call every type of service in a city, it might be too much for the system to handle
    - Another idea might be to utilize the most frequent venues tool to show what’s best for different people, make an algorithm that clusters into different groups and then look to see what’s in each cluster to help recommend to people where to live
    - Can also use this to compare cities, and is very similar to algorithm already have and done in the lab, can just get new neighbourhood data from Wikipedia using scraping
* Can this all be connected to a health outcome? Maybe can connect types of venues in different regions of the city to different health outcomes by neighbourhood – foursquare api tutorial videos in corusera module can help with this