1. 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. Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.*

Prospective new store owners need to take into account many different factors when deciding where to open their business, even moreso for a business such as a restaurant. Knowing the location and neighborhood density of different genres of food would help owners know better neighborhoods to target, since opening the 2nd Italian restaurant in a neighborhood likely has better staying potential and earning potential than the 8th Italian restaurant in a separate part of Toronto.

There are additional factors a prospective restaurant owner might utilize, such as data within their specific cuisine about the targeted audience and price range for restaurants in the area. Additionally the quality of the existing restaurants in the area may provide information about potential in opening another restaurant of the same cuisine in a given neighborhood. All of this data can be scraped or extrapolated from Foursquare and similar databases.

For this project we’ll only be focusing on coffee shops in and around Toronto, the density of those shops around the city, and the best place to open a new coffee shop in Toronto.

* *Data where you describe the data that will be used to solve the problem and the source of the data. Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.*

The data needed would be the separate neighborhoods in Toronto, data of a few different types of restaurants (ie. Italian, Chinese, German) and which neighborhoods they are in, and from that we would need to find the relative density of those types of restaurants in each neighborhood compared to the neighborhood's population. The location data for restaurants around Toronto can be used to identify where there is a relative lack in cuisine in certain neighborhoods, and from there extrapolate possible "untapped markets" for a new restaurant opening.

Further, we can gather a Foursquare dataset that includes the average review rating of each of these restaurants to also highlight where there might be, for example, "no good Chinese place" in a certain area. There would be potential to capitalize on a local population that could be more receptive to another Chinese restaurant moving into the area even when there are a decent number of similar but unpopular restaurants of the same cuisine already established.

There may even be conclusions we can draw from Foursquare data extracted about the average price range a customer may expect at the given restaurants in the area. If there are only upscale Italian restaurants in a particular neighborhood, it's reasonable to assume there's potential in a "Mom and Pop" type family-friendly Italian restaurant in the neighborhood.

* *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, if any, and what machine learnings were used and why.*

After gathering Toronto neighborhoods, boroughs, and zip codes from Wikipedia, some data cleaning was done to remove unnecessary columns and blank rows. Afterwards we connected those boroughs to latitude and longitude coordinates in order to place and cluster them on a Folium map.

Next, I used FourSquare to extract all the coffeeshops near Toronto into a dataframe for further analysis. Then k-means clustering was used to cluster the boroughs into 5 different groups, in order to get an idea of neighborhood density in Toronto. This was followed by getting a tally of the number of coffeeshops within each cluster, and then a graph was made to show the number of coffeeshops within each cluster of boroughs.

* *Results section where you discuss the results.*

The data shows that cluster 4, the orange cluster on our Folium map located north of downtown Toronto, has the lowest density of coffeeshops. This indicates that it would face the least competition from other similar coffeeshops, and thus would be most likely to be successful.

* *Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.*

This analysis is not perfect, as it is based upon zip codes and borough density – this is probably correlated but not directly linked with population itself, as some zip codes may have a much lower population density. This would skew our results, as a more populous zip code would be able to support a higher number of coffeeshops. However, the number of coffeeshops within cluster 4 is such a stark difference from most of the other clusters, that I am relatively confident that opening a coffeeshop within that area would still be able to succeed based on the lack of competition there.

* *Conclusion section where you conclude the report.*

In conclusion, the FourSquare API was found to be very useful in addressing and analyzing our business problem. Machine learning was used to cluster the boroughs in an efficient and fair way which allowed us to quickly investigate where there was a relative lack of competition for our business. The pandas library provided a quick and fluid way to clean and organize our data. The matplotlib and folium libraries allowed us to visualize the data that we pulled and more easily draw conclusions.