**Introduction/Business Problem**

My friend Mark wants to open a new coffee shop in Toronto city, base on his business experience, people often would like to buy a cup of coffee before he/she go into stores, like book store, baby store, or even super mall. So Mark would like to find some good places for his new coffee shop. As still in very first stage, and Toronto is a big city with lots of neighborhoods, Mark hope to firstly select some neighborhoods which have better business potential for the new coffee shops. So the objective of this project is to help Mark to find out which neighborhoods have best business potential for opening a new coffee shops base on geographic data and his business experience.

**Data** **Description.**

To help Mark find the best neighborhoods for opening a new coffee shop, we need data of all neighborhoods of Toronto, and latitude/longitude data for each neighborhood as well. And we also need to get venues data from Foursqure.com which contains popular nearby venues for each neighborhood, then we will select all stores information and existing coffee shops as well to analyze competition.

**Methodology.**

To 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.

Firstly, we will use data from wiki website to get all neighborhoods of Toronto, and then parse the data into dataframe, and add latitude/longtitude data into neighborhoods datafrma. To complete the first step, we will use folium map to visualize all neighborhoods data within the map.

Secondly, use Foursquare to retrieve top popular venues for each neighborhood in nearby, to to select all stores and coffee shops data for further analysis.

Then, by combining neighborhoods data with venues data, we get complete stores/coffee shops competition data into one dataframe. We will calculate two indicators: Rate is the number of stores per existing coffee shop, and Diff is the difference of numbers of these two features. The higher the Rate and Diff means there are more stores and less existing coffee shops, also means there more business potential for open a new coffee shop.

Finally, we choose to visualize the neighborhoods with folium map, and to highlight the neighborhoods with Rate > 1 and Diff > 3 to be the best choice for open a new coffee shops as recommendation to Mark.

**Results.**

By adopting above methodology and with data analysis. We have achieved the result of choosing about six neighborhoods for Mark to open a new coffee shop. We visualized these neighborhoods with others in the map by marking with different size and color, the red marked neighborhoods in the map are within top list of our recommendation.

**Discussion.**

The project is just a initial analysis for location recommendation, we only select stores and existing coffee shop as features to analysis the completion. Basically base on the analysis from this project, we can do further analysis more data to get more accurate information for the analysis, such as popularity of each stores and coffee shops, or number of office buildings and population of each neighborhood.

**Conclusion.**

Summarily, this project can give a initial analysis and advice for Mark to open his new coffee shop. We successfully narrow down the candidate neighborhoods from more than 60 to less than 10.