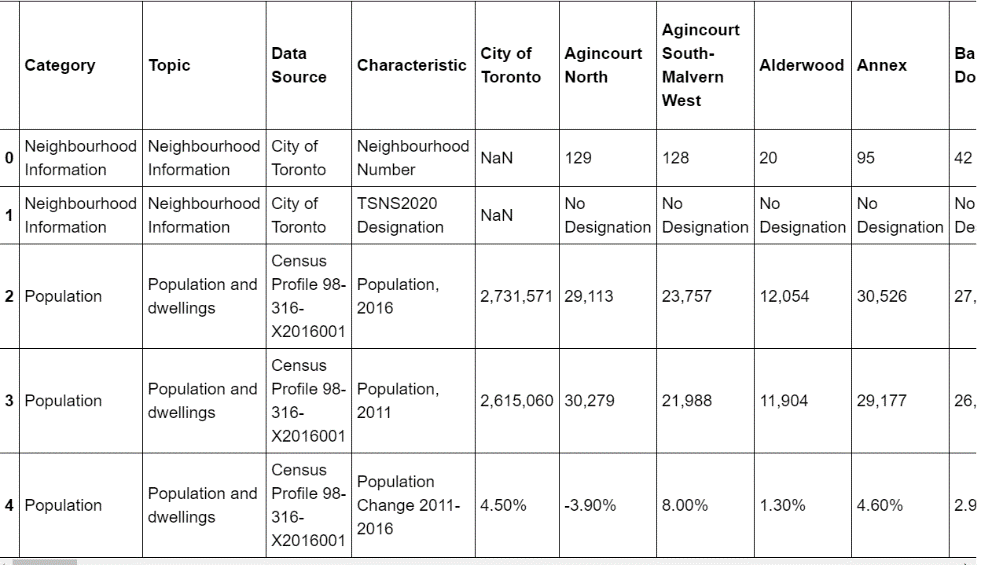
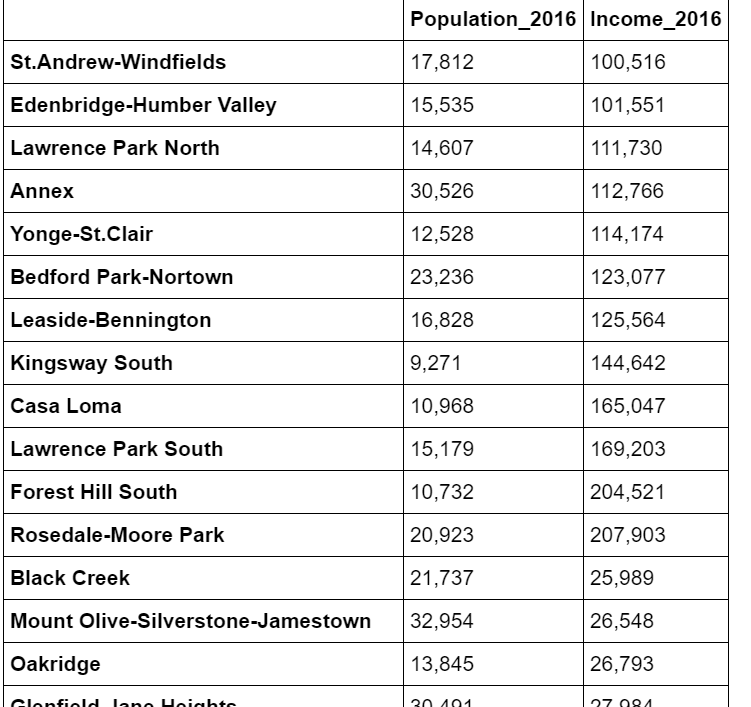
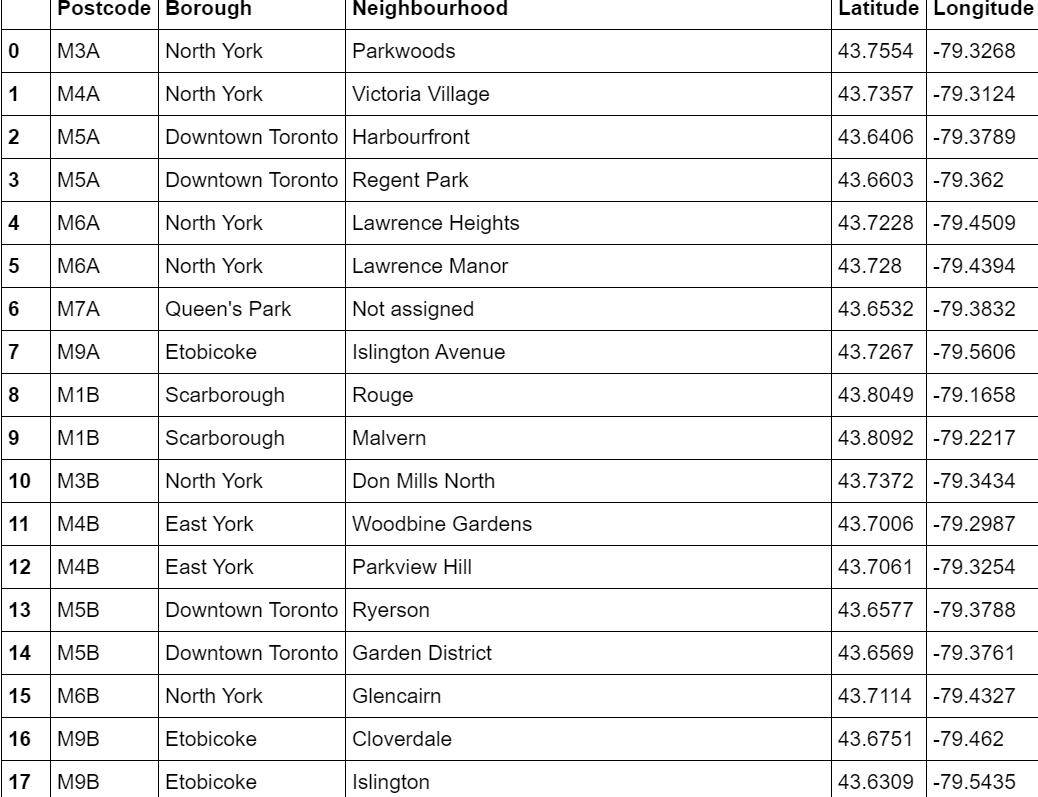
**Report-Toronto Neighbourhoods**

* **Introduction-Business Problem**
* The target audience is someone who already has a restaurant in a different area and would like to move to Toronto. They need of course to maintain themselves so they would be interested in opening a new restaurant in Toronto. Hence, they would benefit from this case study as it will be provide a clustering of different areas based on similarities so that the moving person can identify a similar area to their hometown so they can settle in comfortably. Moreover, the case study will provide information on the incomes and populations of each neighbourhood so that they can identify their client base but also what kind of money they might earn, and where customers might give the largest tips.
* **Data**
* In order to do this analysis reliably, the data used will be taken from the Toronto 2016 Census containing population, average income per neighbourhood etc. and this will be combined with the Foursquare API which is to collect the information on existing restaurants in these neighbourhoods which might be competitors. For example, if there is already a successful restaurant in a particular area it may not be the best idea to settle there. From the dataframes extracted, graphs will be produced to make the data more visual.
* **Methodology**
* First, the census data from 2016 was used to collect the income per neighbourhood from Toronto
* ****
* Then a list of names was collected
* ****
* Next, a new dataset is created with the incomes and populations displayed for each neighbourhood, with the table sorted in order of income for easier use.
* 
* Alongside this, K-means clustering was used on the Foursquare API data for Toronto to group the neighbourhoods, the latitude and longitudes were added to each section in order to be able to plot a map, to which the restaurants in competition could be added.
* The clustering method is a simple form of machine learning use, while the production of tables comes under the category of exploratory data analysis as I didn’t know what results might be produced before.
* 
* The reader now has a lot of data displayed which it can use to draw its own conclusions, depending on what kind of neighbourhood they are looking for and the kind if salary they want.
* **Results and discussion**
* The interpretation of the results depends of course on the user, but we can assume that they would want to move to an area with maximal chance of a high salary with a large threshold population and minimal competition. From the sorted table, we can see that the most recommendable areas in terms of purely salary are Rosedale-moore Park, Forest Hill South and Lawrence Park South. The latter is the one with the highest population which may mean more clients so this may be the best area, however the 3 areas all have a similar population. It is in the area of midtown Toronto, which we can then use our map to explore.
* **Conclusion**
* These results should be taken with a pinch of salt because a high average salary in the area may imply a high cost of living or high rent for the restaurant. The user may want to have the restaurant in a rich area while living further away in a more affordable area. This will require more data as it will be necessary to investigate transport to ensure good transport to work is possible. Moving area is of course a decision that requires many more factors analysed.