**Data Science Capstone Project**

**Final (The Battle of Neighbourhoods)**

**Introduction**

With broader application of data science arising from new capabilities and ideas led to existing businesses facing another level of threats. Start-ups are keener to leverage new technologies and innovative business models consistently looking for ways to impact customers in such a way that they influence customers on the products they are choosing and how their chosen product better impact their lives.

To remain competitive existing businesses have embraced use of location data to identify consumer patterns, tastes, and trends. They are readily available these days from several social media platforms including Google maps, Facebook check-in service, Yelp.ca platform, location review on sites like TripAdvisor and Booking, and Foursquare among other location data providers.

Working as a data scientist, one of customers have asked for consulting as to their business strategy involving data science technologies to come up with business models to keep them competitive as a start-up restaurant. Location is one most important factor when considering to opening a new restaurant and it is strongly believed that data services using data science technology will give them advantages.

**Data**

1. Population of Toronto by FSA (Forward Sortation Area)
   * 1. Source: Statistics Canada Open Portal
2. Average Income of Toronto by FSA
   * 1. Source: Statistics Canada Open Portal
3. Foursquare query looking up restaurants within each region specified by FSA
4. Geocoder library containing coordinates of each region specified by FSA.

**Methodology**

1. Clustering regions of distinct post code by the number of restaurants within specific radius will be done by using K-means clustering to group regions into “high number of restaurants with high review ratings”, “high number of restaurants with low average review ratings”, “low number of restaurants with high review ratings”, and “low number of restaurants with low review ratings”.
2. Population data for each region with distinct post code will be used to in conjunction with the number of restaurants within each region of specific radius to give priorities of each region.
3. Reviews from each sub region of the city will be evaluated from data collected from Foursquare and determine their ratings and number of potential customers estimated from the size of reviews.
4. Using all of analysis done above, customer will be advised of best type of restaurants for each region, with number of competitors around the region so ultimately give them insights on their choice of restaurants, its best location as a final answer.

**Results**

**Cluster 0**

Avg Income 125482.0

Population, 2016 31501.1

Venue Category 0.1

**Cluster 1**

Avg Income 52079.413793

Population, 2016 20333.172414

Venue Category 0.344828

**Cluster 2**

Avg Income 126949.0

Population, 2016 5277.7

Venue Category 1.5

**Cluster 3**

Avg Income 68664.500000

Population, 2016 35007.909091

Venue Category 0.045455

**Cluster 4**

Avg Income 54055.677419

Population, 2016 32403.709677

Venue Category 0.161290

**Discussion**

Cluster 0 shows that most of regions (9/10) have no restaurants within 700m with average income of more than 125k which is double the average of Toronto. Average population however in cluster 0 regions are almost half the average population of Toronto region. Therefore Cluster 0 region includes high-income neighbours, but they are not many

Cluster 1 region has most of regions not having any restaurants within 700m but region "M5H" for example has 4 restaurants when demographic shows ~20k people having high income (~180k) leading to an intuition that the region has a high demand of restaurants whether its saturated or not. Cluster 1 region overall has average income of ~52k which is less than Toronto average with only ~20k people living in. Non-luxurious type of restaurants is well suited as a start-up consideration.

Cluster region 2 is a region where people of average income of ~127k lives in but only ~5k lives in. There are on average of about 1.5 restaurants within 700m of radius which is relatively higher than cluster regions 0 or 1, which could be a sign be high demand for restaurants. Luxurious restaurants with high priced foods could potentially be the right choice for a start-up in this region.

Cluster region 3 has relatively higher population than any other clusters and they are also above Toronto average. Their average income looks to be around on par with Toronto average. Any kind of restaurants seem to work well depending on the choice of specific regions, but the demand seems to be very low thus this is not the ideal region to open a restaurant.

Cluster region 4 seem very similar to cluster 3 where average income is below Toronto average and their average population is slightly above Toronto average. Demand for foods in this region doesn’t seem to be stable and this is not the ideal location for a start-up to open a restaurant unless they strategically choose the region for a long-term success.

**Conclusion**

It is advised that depending on the strategy that the client is trying to implement the solution should be adjusted accordingly. Luxurious restaurants with high prices foods, services, and quality should consider cluster region 2 and start considering nearby venues to further narrow the choice. Restaurants with mediocre level of service and quality of foods are well suited to open in cluster region 1 as they have a group of people with average level of income and population. Client should seek further analysis of which specific areas within the region should be the optimal choice of location.