Determining the best location for a new restaurant

Introduction

- Charlie, a data scientist, is looking to invest money into a new business. His business partner and long-time friend is a chef and had decided to open a new Chinese restaurant in Toronto. Charlie is has decided to leverage his data-science skills to identify the perfect location (postal code) for the new restaurant.
- Charlie must consider the market conditions including the current competition and population of the location to make the best choice when choosing the restaurant location.

Determining the best location for a new restaurant is valuable for business success

- A restaurant is a highly competitive business
- Location matter – getting people into the restaurant is difficult
- Finding the right location can make-or-break the business

Strategy

- The venues in a postal code will be used to rank them. Each venue will be given a point value.
- Attractors such as schools will be weighted more heavily than detractors such as competition
- Population will not play a significant role in determining the ranking of a postal code. Data will be normalized to level the playing field.

Strategy

Assumptions

- A larger population will attract more venues – skewing data

Data Acquisition

- Toronto postal codes is retrieved from Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Population data for postal codes in Canada: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A
- Venue data is retrieved with Foursquare API

Methodology

Set-up

- Postal codes are gathered and compiled such that there is only one entry for each postal code
- Location information (Latitude and Longitude) is added to each postal code
- Population data is retrieved and added to each postal code. Sice the restaurant relies on the local population, only postal codes with greater than 25000 population are considered
- Near-by venue data is retrieved from foursquare for each postal code

Methodology

Ranking Venues

- Venues are ranked according to perceived benefits of that venue to the restaurant.
- Points for each venue can be between -10 and +5
- Example:
  - A Bar creates foot traffic and therefore is valued at two points
  - A school creates even more foot traffic and is valued at five points
  - Other restaurants are competitors are values at minus two points
  - Since it is a family establishment, an adult store is valued at minus ten points

Methodology

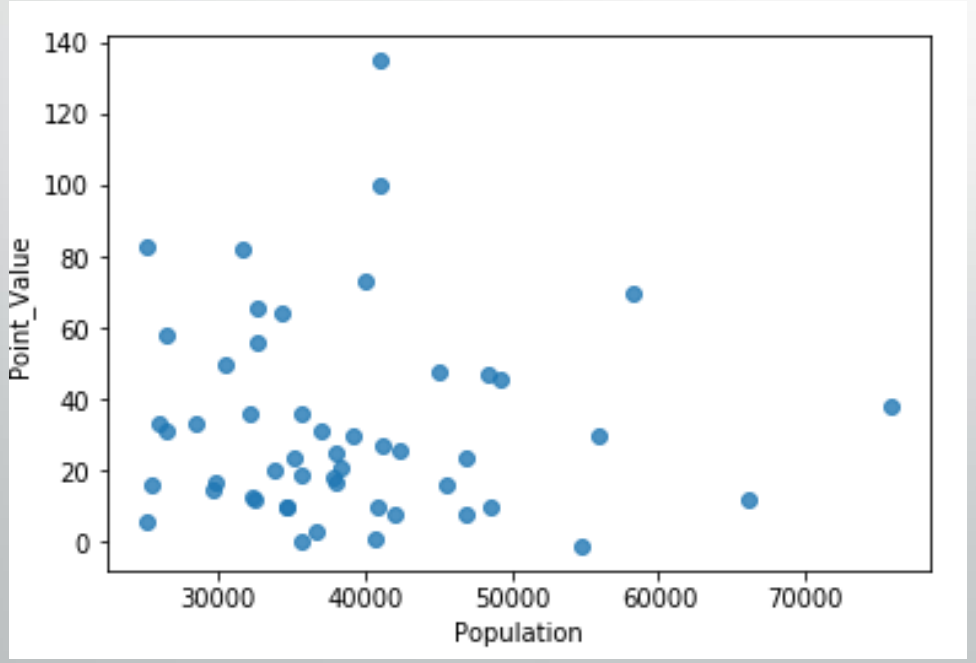
- Each venue is ranked based on the venue category as retrieved from foursquare
- Venue points are wrapped-up to the postal code and each postal code is assigned the sum of the venue points
- Since the population can vary significantly, the population data is normalized to between zero and one. Hear by referred to np.
- The point value is multiplied by (1 – np) to re-weight the points and level the playing field.
- The new point value is recorded for each postal code

Results

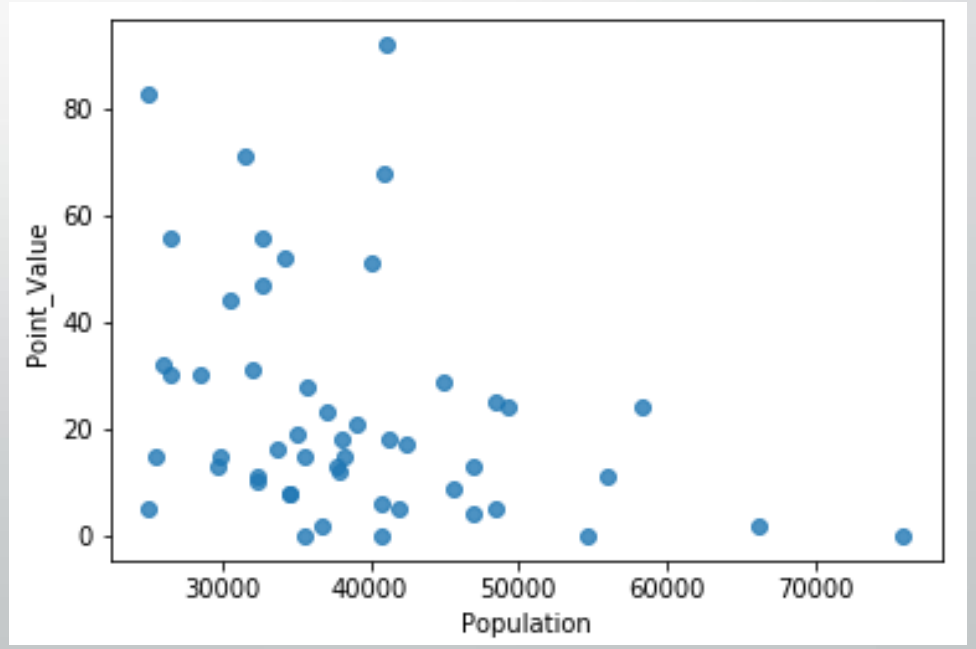
- The postal codes are ranked by the point value
- The top three are identified as:
  - 1. M5A – Population 41,078
  - 2. M4E – Population 25,044
  - 3. M4K – Population 31, 583

Results

Raw Points for each Postal Code



Normalized point values



Conclusion

Population did not have a significant impact on the points for each venue. From some analysis on the data, it seems a higher population actually brought more competition and detractors.

The best location had a mid-range population whereas the second best choice was on the low-end of the population range.

The best place for a new restaurant in the city of Toronto is in the M5A area code.