The Battle of the Neighborhoods

1. Introduction
    a. With the increasing cost of housing in New York City, some workers have sought new cities in which to continue or start their career. A city such as Toronto will be offered as a comparison, due to its dense population and standing as a sought-after location for companies. New York City and the city of Toronto share a diversity of inhabitants, which is reflected in the makeup of their neighborhoods.
2. Problem Description
    a. As cities, such as Toronto and businesses located in cities, such as Toronto seek to lure potential inhabitants and workers, they must examine what their city offers.
    b. Comparisons of population and employment offerings are not enough as people will want to see information on venue offerings, such as restaurants and bars.
    c. Neighborhoods in both cities will be clustered, segmented and compared.
    d. We must take into account the different offerings of each city and each neighborhood to ensure there can be both an incentive and seamless transition for potential job seekers.
    e. Once complete, this information can be of use to companies located in Toronto to draw talent from across the globe.
3. Data
    a. The data set for all necessary information regarding the city of New York is https://geo.nyu.edu/catalog/nyu_2451_34572
    b. The data set for postal codes in the city of Toronto is available at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
4. Data acquisition
    a. The data for New York city will be pulled from the website mentioned in section three, this contains all neighborhood and location information.
    b. The Toronto data will be scraped from the link in section three to obtain neighborhood information
    c. Finally, Four-Square will be used to source venues within each neighborhood, using latitude, longitude, radius (1,500 meters), limit (100), section (drinks and coffee) to explore the geo-location, name and category of each.
    d. The parameters will be as follows:
        i. VENUE_PRIME = [‘bar’, ‘lounge’, ‘restaurant’]
        ii. VENUE_SECONDARY = [‘coffee’, ‘tea’, ‘donut’]
        iii. MAX_VENUES = 20
        iv. MAX_WALK = 1
            1. To keep the walk between a mile and one half a mile
        v. PRIORITY_ORDER
            1. We will use a combination of Rating, Count and Likes
    e. Once compiled and cleaned the data will be plotted and listed to show the comparison between both cities and the neighborhoods within.