An Analysis of Craigslist Housing Listing in a Toronto Context

CSDA1050 - Advance Analytics Captstone Author: Jessee Ho

Introduction

Over the last few years, the gentrification of Toronto has occurred at an impressive rate, seeing house prices rise on average 15% - 30% year after year. The process of gentrification can bring positives such as increasing property values to the benefit of existing owners, on the other hand, it can force out local populations who can no longer afford to remain in their own neighborhoods. Real estate listings were extract from Craigslist over a 3 week period and analyzed using various geospatial representations to examine the affordability of current housing as well as locate areas of demand.

Summary Findings

Some key insights were that a majority of listings were 1 – 2 bedroom unites localized to two clusters. The first cluster located in the downtown core, mainly comprise of the Waterfront, Niagara, Trinity Bellwood, Bay Street Corridor and South Riverdale neighborhoods. The second cluster was a narrow-localized cluster in the Willowdale East neighborhood. Another interesting observation was that the price per bedroom cost dramatically decreased as the bedrooms increase, since the common area cost would be shared among the bedrooms. The most expensive houses were located in the Edenbridge-Humber Valley neighborhood, which had a dramatically different price range than surrounding neighborhoods. Despite the relatively low income of the downtown core, it scored as one of the best regions in terms of the affordability score.

Limitations

The data extraction period, while over one month, became sporadic after the first week of collection. This was due to the availability of internet while I was abroad. The data was sourced from a single website which could only pull 3000 listings per day extraction. Additionally, the turnover period for these listings was relatively long so there was a significant number of repeated data (only around 100 new listings per day). Furthermore, the data quality was not the best, with significant missing data values as there is no oversight or data control when listing a property. In terms of data content, I found that the real estate data contained both condominiums and houses which can have dramatically different price points and cause some skewing of the data. While outliers were addressed as part of the data cleaning process, I believe this had an impact on the overall aggregate results.

The strength of the analysis can be further increased by sourcing additional data that impacts a listing score such as, retail points of interest, parks and recreational area, and transit routes and stops. Additionally, a more reputable and controlled source should be used for analysis such as MLS, though I acknowledge that Craigslist has less barriers to obtaining data.

Results

Plotting all Listings by Neighbourhood

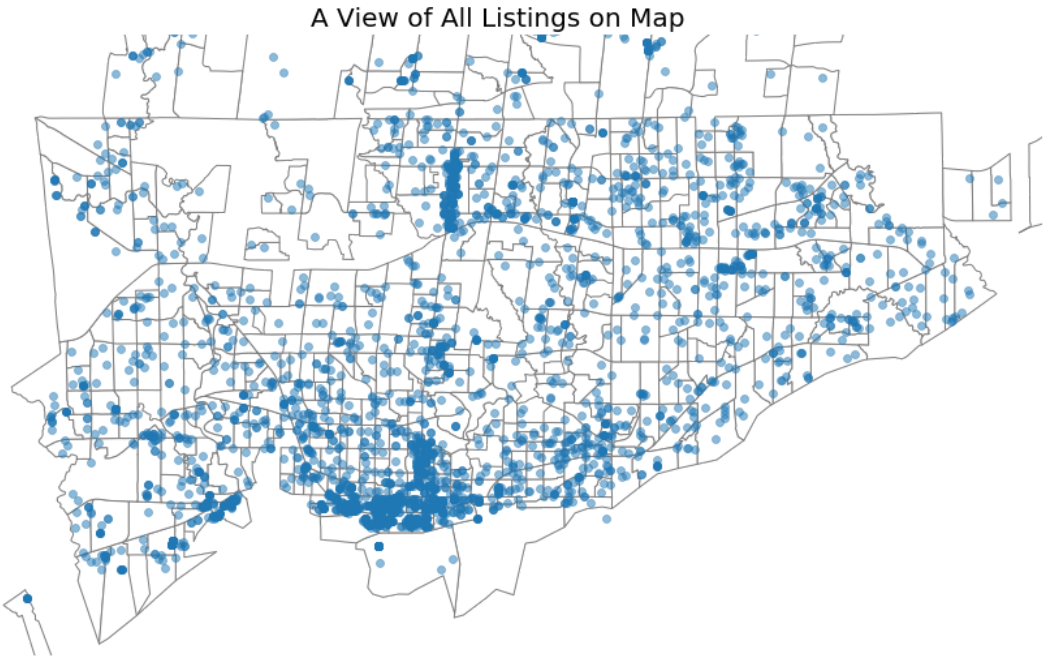
From plotting on listings on a map, two clustering trends become very apparent:

1) heavy clustering around the Downtown core, specifically neighbourhoods ("N"):

N77: Waterfront, N70: South Riversidale, N75: Church-Yonge, N76 Bay Street, N81: Trinity Bellwood, and N82: Niagara North (liberty village)

2) Surprisingly, there was a distinct narrow corridor of clustering at N51: Willowdale which was crowned in 2015 by the Toronto Star as the hottest up and coming neighbourhood in Toronto.

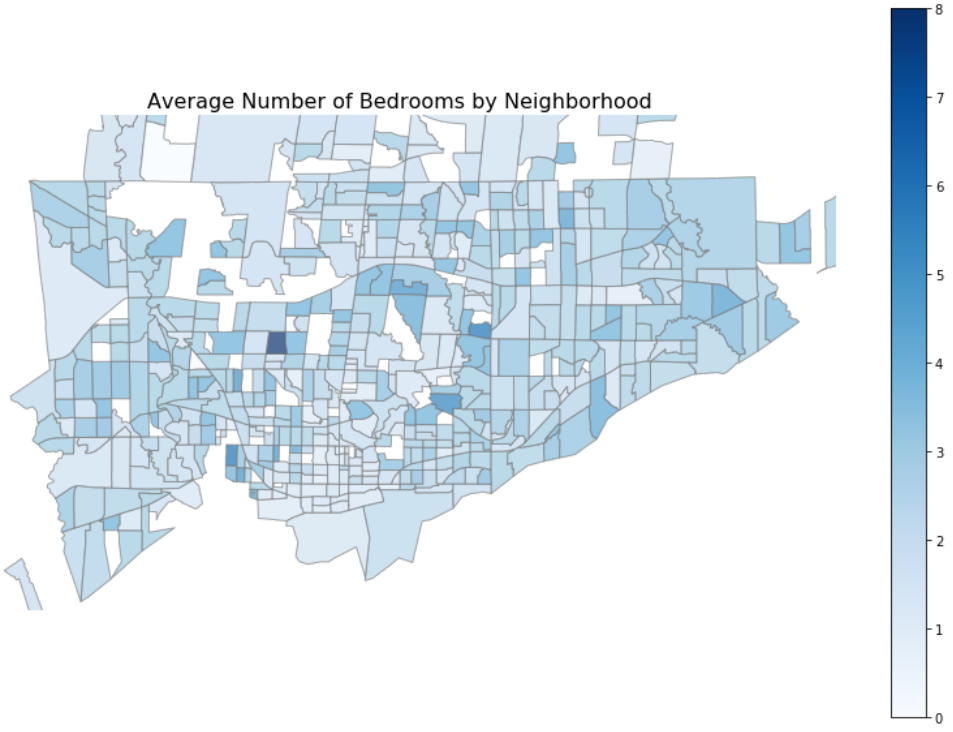
In [797]:



Average Number of Bedrooms By Neighbourhood

Confirming expectations, the heavy cluster located in the Downtown core/Waterfront regions are predominiately 1 - 2 bedroom units and most likely condos. As you move over to South Riverdale region though we can see the bedroom sizes increase, perhaps indicateing townhouse complexes. The Willowdale East cluster seems to me made up of 1 - 2 bedrooms as well. On interesting observation is South-East Yorkdale-Glen Park seems to have very large 7-8 bedrooms units. We'll see if the median household income of this neighbourhood reflects the house size in the following maps.

In [798]:

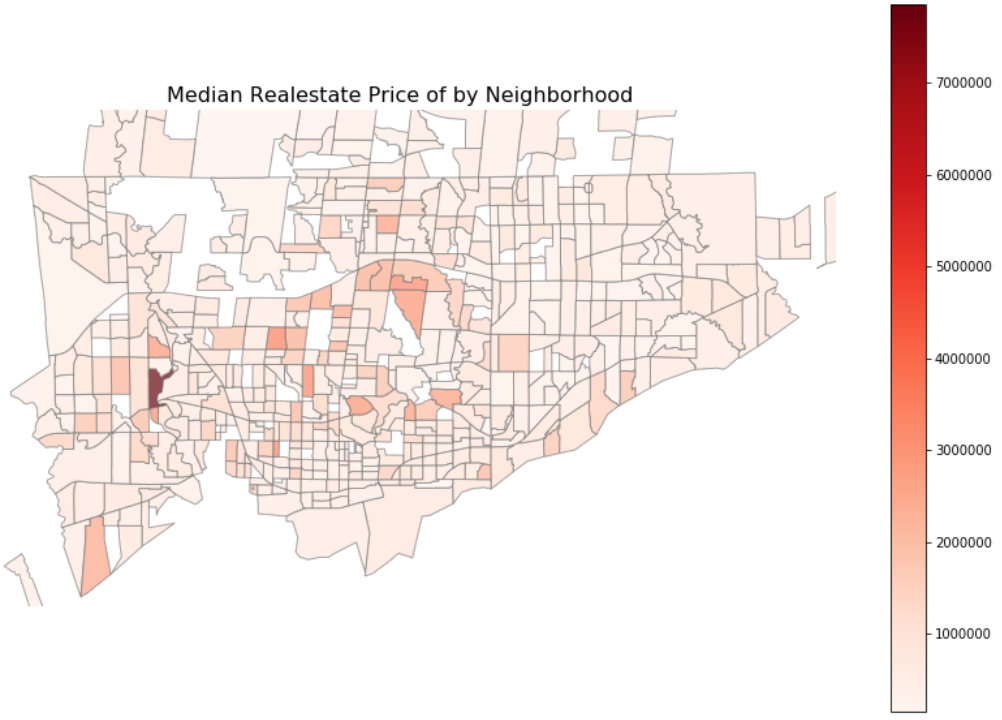


Average Realestate Prices by Neighbourhood

One obervation that immediately pops out is the extremely high house price in N9:Edenbridge-Humber Valley (only dark red region). Upon further research, Edenbridge-Humber valley is one of the wealthiest in Toronto. It is a suburban neighbourhood developed for wealthy people. In contrast to the surronding neighbourhoods like N115:Mount Dennis which was an industrial, working-class neighbourhood some distance from the centre of the city because of factories, industry and slaughter houses.

Looking at Downtown core we can see that the price is relatively lower. Lower being the $0-1.5$MM range but that makes sense due to the density of condo development.

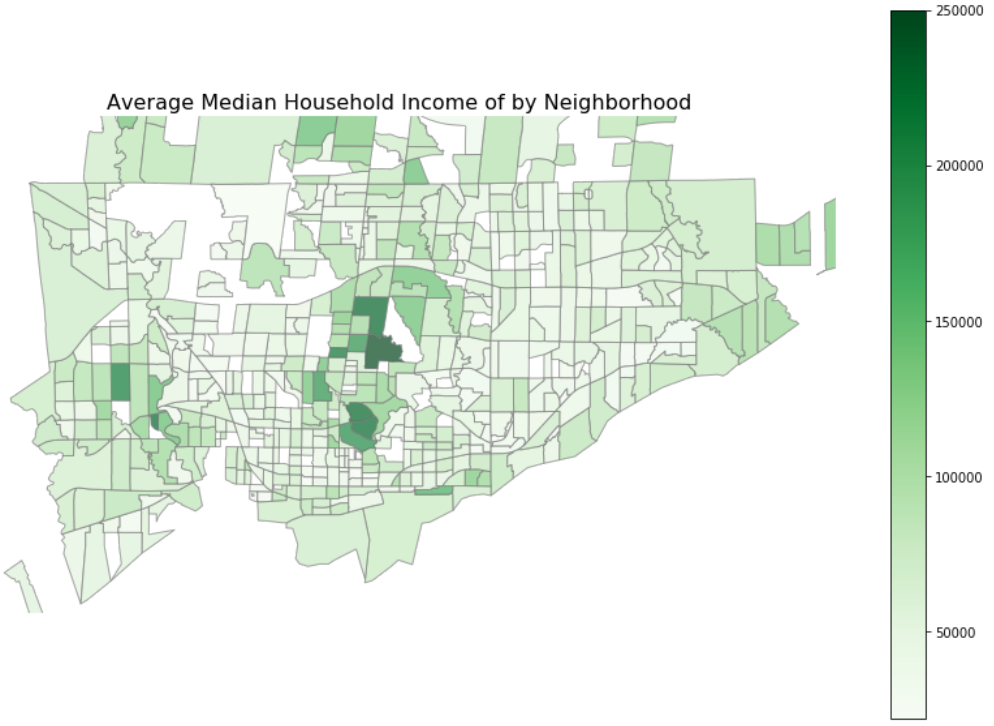
In [799]:



Average Median Household Income by Neighbourhood

Looking at average median household income, it is no surprise that the N41: Bridle Path-Sunnybrook-York Mills neighbourhood is the highest. While it's a little odd that the house prices in this neighbourhood were relatively low, when compared to income; however, this data came from craiglist, which may not be the go to source of realestate listing for these areas. Another observation is that the observed income of residents in the downtown core is relatively low, somewhere between ($) 50,000 to 100,000.

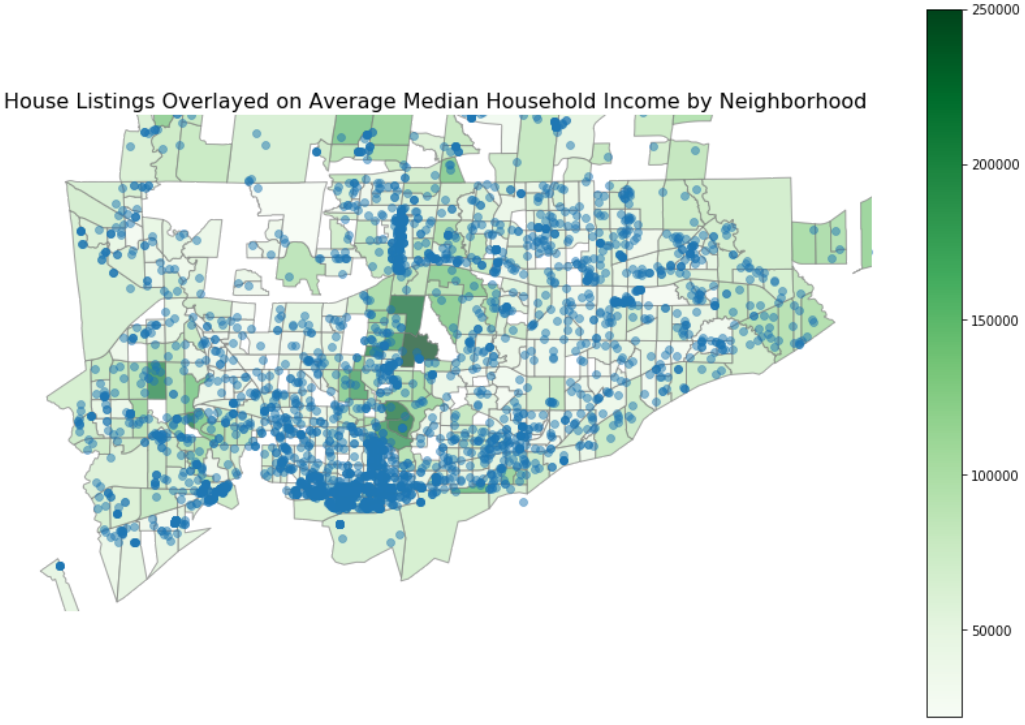
In [800]:



Listings Overlayed on Average Median Household Income by Neighbourhood

Viewing the listings overlayed on the median income show a high demand for realestate in the core despite the the lower end incomes. This could also explain why the housing prices are relatively low (under $1MM). Next we will examine the affordability of housing in Toronto based on income.

In [801]:



Mortgage Affordability of Listing by Neighbourhood based on Median Income

Based on the criteria set for affordability (i.e. Annual Mortgage Payments less than Annual Household Income, and house deposit less than 5 years worth of savings at 40%), we can see that housing, especially in the downtown core cluster, is relatively affordable. That being said, we know that the core is pre-dominately 1-2 bedroom units. The areas that tend to have more bedrooms are significantly unaffordable for the median household incomes in the neighbourhood.

In [802]:

Out[802]: Text(0.5, 1.0, 'Mortgage Deposit Affordability by Neighbourhood')

Score of Local Housing based on Median Household Income

Based on our scoring criteria, it seems that the downtown core scored pretty well in terms of affordability for the size of property, with affordabolity scores dropping as we move to the surrounding neighbourhoods. This drop in score could largely be based on the increase in price between houses and condo units. Because the score is largely dependent on the calculated affordability based on neighbourhood household income, the scoring would be largely localized by neighbourhood.

In [808]: