**Analysis of the Economic Impact of Cultural Diversity in the**

**City of Toronto**

Melissa Siqueira

July 30, 2021

1. **Introduction**

There are many reasons people move to large cities. They might move because they have a better job opportunity, or because there are more options for professionals in areas with a greater concentration of offices and people. If you live and work in the city, you generally enjoy lower commute times. Perhaps you may just be a people person and there is something about being surrounded by people 24/7 that attracts you to high-density areas.

A reason often cited for going to cities is the proximity to good entertainment, food and attractions. In a higher population area, it makes sense that there would be more businesses open to service the population. However, some people go further and paint the increase in attractions as a reflection of the cultural diversity within an area. Since a higher number of businesses increases tourism and traffic to an area (and tourism and traffic increase the profitability of opening new businesses), this goes to say that cultural diversity can be beneficial to a city's economy.

This study will examine Toronto location and demographic data to determine if the number of businesses open in an area are in fact a result of demographic diversity. We will do this by seeing if there is an impact of population characteristics (i.e. population, age, income and language as a measure of cultural diversity) on the number of restaurants in an area. If, after controlling for population density, there are still significant impacts of age, income and cultural diversity, then this could indicate that there is some economic value to be found in incentivizing certain groups to move to cities.

1. **Data**

In order to conduct our analysis, we will be looking at a cross-section of Toronto neighbourhoods where we have information on the number of restaurants in each area and various demographic characteristics of each area including age groups, income groups, native languages and population density.

We will be using four main data sources: a list of Toronto neighbourhoods and postal codes, web-scraped from Wikipedia ([link](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)); the latitude and longitude coordinates for Toronto neighbourhoods, provided by IBM; Foursquare location data identifying restaurants in each neighbourhood; and Toronto Census data provided by the City of Toronto ([link](https://open.toronto.ca/dataset/neighbourhood-profiles/)).

The Wikipedia dataset is a table containing all Toronto postal codes starting with M (hereby known as boroughs) and their underlying neighbourhoods. These data were originally retrieved from the Canada Post website and compiled by the authors of the Wikipedia page. This dataset was merged with the latitude and longitude coordinates based on postal code, which is unique for each neighbourhood.

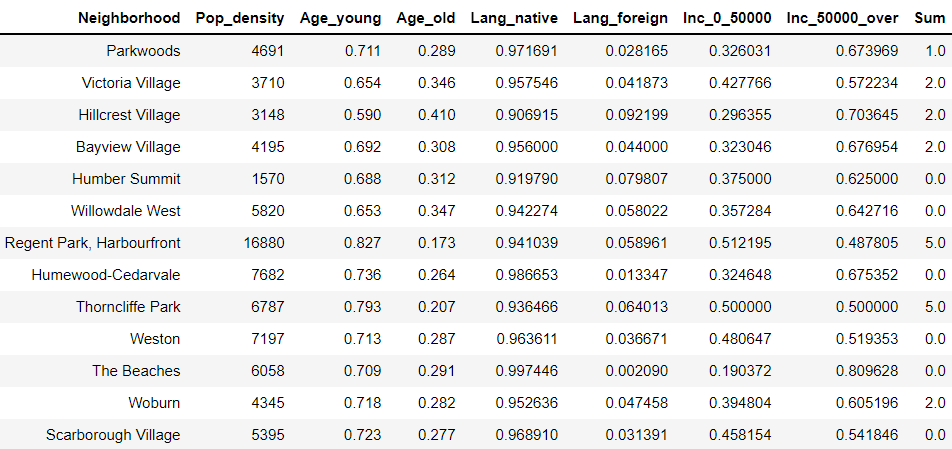
Next, we obtained Foursquare location data on venues based on the latitude and longitude coordinates of each postal code. These data were retrieved through an API and merged based on neighbourhood with the master dataset. We did not keep the names of all venues; rather we limited the data to venues with “Restaurant” in their name.

Finally, the City of Toronto dataset was downloaded off the City of Toronto website and contains aggregated demographic data on each of the city's neighbourhoods. Collected in 2016, this data was collected at the last nationwide census and includes information such as age, gender, languages spoken, income and population. Age, languages spoken, and income are represented as proportion of the total respondents. We will be using this dataset for the population density per square km, proportion of individuals from each age and income group and the proportion of non-native English/French speakers.

The overall dataset is a cross-section of 20 neighbourhoods that appear both in the list of Toronto neighbourhoods and postal codes and in the census data.

1. **Exploratory Data Analysis**
   1. **Summary Statistics**

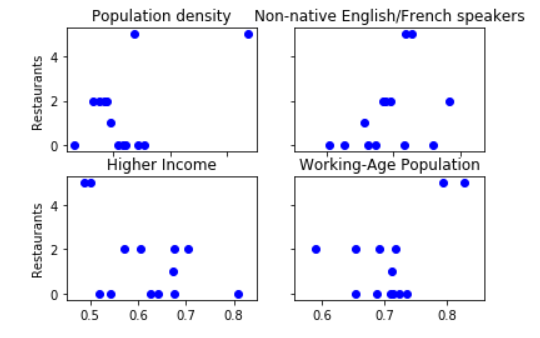
We generate some summary statistics to get a higher-level overview of how the neighbourhoods differ. As seen in Table 1, we can see that there are some neighbourhoods that are a lot denser such as Regent Park/Harbourfront which is in the Downtown Toronto borough. Thorncliffe Park is the lowest-income neighbourhood where 50% of the population's households earn less than $50,000. There are also a few neighbourhoods that have zero restaurants (represented by the column “Sum”). These neighbourhoods may affect our analysis, however, their demographic characteristics seem to be different enough that there is no systematic bias.



*Table 1: Mean Summary Statistics by Neighbourhood*

* 1. **Correlation between Variables of Interest**

In Figure 1, we plot some relationships between the dependent variable, number of restaurants, and the various independent variables (population density, income proportion, age proportion and non-native English/French speaker proportion). Aside from the zero observations, there seems to be a positive correlation between the number of restaurants and population density and proportion of non-native English/French speakers. There is also a weak negative correlation between the number of restaurants and the proportion of high-income households. We will see how these relationships change when we include them all in one regression. It is possible that these demographic variables are correlated (younger people may have more income, for example) so that the correlation, after controlling for confounding variables, may switch.

 *Figure 1: Plots of Correlation between Number of Restaurants and Variables of Interest*

1. **Methodology**

In this project, we are attempting to determine the impact of cultural diversity on the number of restaurants in an area. Due to data limitations, we proxy cultural diversity using the proportion of non-native English or French speakers. Inevitably, there are other factors that may affect the number of restaurants. Based on the data we have, we are adding proportion of high-income households, proportion of working-age population and population density as covariates. The greater the proportion of high-income households, the wealthier a neighbourhood is. We may expect wealthier neighbourhoods to have more restaurants as people in these areas have more disposable income to spend on eating out. Neighbourhoods with a larger working-age population (ages 15-55) may also have more disposable income to spend. Younger individuals also tend to go out more than their older counterparts. Finally, a higher population density means that there is a larger customer base and higher demand for restaurants. In order to meet demand, we would expect a higher supply of restaurants as well. For these reasons, we are including income, age and population density as controls in our linear regression.

Once we have collected the number of restaurants within each neighbourhood, we run an OLS regression to determine the impact of demographic factors on the number of restaurants in an area. This is shown in the below equation:



where *j* is a neighbourhood in Toronto, *popdensity* is the population density per square km, *ageyoung* is the proportion of individuals between 15 and 55 years old, *inchigh* is the proportion of households making over $50000/year and *langforeign* is the proportion of individuals not speaking English or French as a first language.

The coefficient we are particularly interested in is which indicates the impact of cultural diversity on the number of restaurants. Here, we use not speaking English or French as a first language as a proxy of cultural diversity since we do not have immigration or ethnic background information in our dataset.

It is important to note that we do not run the regression on a testing and training set as our combined dataset only has 20 different observations. If we were to split our dataset further, this would further increase the standard errors of our estimates, thus decreasing the precision. This will hurt interpretation of the statistical validity of our results. In the end, our goal is not to predict the number of restaurants in the area, it is just to test whether there is an impact of the proportion of non-native English/French speakers on the number of restaurants.

1. **Results & Discussion**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Coefficients** | **Standard Errors** | **T-Values** |
| *Population Density* | 0.0001 | 0.000 | 0.636 |
| *Proportion of individuals between 15-55 years old* | 16.1465 | 12.055 | 1.339 |
| *Proportion of households with income over $50000* | 1.0873 | 6.187 | 0.176 |
| *Proportion of non-native English/French speakers* | 42.6274 | 20.105 | 2.120 |
| *Constant* | -13.2330 | 11.068 | -1.196 |
| *R-squared* | 0.59 | | |
| *Number of observations* | 20 | | |

The results can be found in Table 2. Somewhat unexpectedly, there is a small, albeit positive, coefficient on population density. This means that after controlling for other demographic factors, there is almost no correlation between having a higher amount of people living in a given area does not necessarily lead to a higher number of restaurants. We had previously hypothesized that this coefficient would be positive and significant due to higher demand for restaurants. One possible explanation for this is that higher density areas have more land dedicated to residential buildings and less to commercial buildings (where restaurants are). Some neighbourhoods will have more restaurants to meet the demand, but in other neighbourhoods, the supply is limited.

*Table 2: Results*

There are also positive coefficients on the working-age proportion and higher-income proportions. Although these coefficients are not significant at the 5% level, they have larger magnitudes than the coefficient on population density. This aligns with our hypothesis that these two populations would have larger disposable incomes and would be able to spend more on restaurants. However, this is a two-way correlation as younger and higher-income people might tend to move into areas with more restaurants.

The largest and only significant coefficient belonged to our proxy of cultural diversity: the proportion of non-native English/French speakers. This coefficient was significant at the 5% level, meaning that having a higher number of ethnically diverse constituents within the neighbourhood has a positive impact on the number of restaurants in a neighbourhood. This could be explained by those who do not speak an official language well choosing to open up their own businesses instead of pursuing the job market. Those with different cultural backgrounds also tend to fare better in the restaurant business due to them having a unique cuisine and selling point.

For an economic model, the R-squared is not unreasonably low - it sits at 0.59. However, except for the language variable, none of the variables were significant at the 5% level which means that statistically, they did not have a lot of explanatory power. One possible explanation is that the cross-section only consists of 20 neighbourhoods. When the number of observations is that low, standard errors tend to go up which can deflate the statistical significance of estimates. It is also possible that with a larger sample, the coefficients themselves might change.

Given our limited data, it also would have been helpful to be able to control for other variables. For example, places with many offices or easy access to transit may have more businesses and restaurants due to a larger customer base so that we may want to control for these type of venues. As well, crime rates, tax policies and rents may also affect the number of restaurants in an area. Including these covariates would make the findings of this study more robust.

1. **Conclusion**

Through measuring cultural diversity through its proxy of non-native speakers of English and French, we see a positive impact of cultural diversity on the number of restaurants within an area. Increasing the number of restaurants has economic and social benefits for a neighbourhood. It not only is a place where inhabitants can spend money and boost the economy, it can also encourage people (and tourists) to visit the area. Moreover, economically speaking, more restaurants lead to more choices for consumers which allow consumers to choose what benefits them the most (whether it be having a type of food they like or getting a lower cost). This increases the social welfare of people living and visiting a particular neighbourhood.

For immigration policy-makers, this study helps in forming an economic argument why immigration and the resulting multiculturalism are good for Toronto (and cities in general). In order to make the findings more robust, policy-makers will have to include more controls and experiment with various regression forms, however it is a starting point for more study in this area.