

Capstone Presentation

Chad Franklin

Introduction

- ▶ This project sought to compare Foursquare data within Toronto and demographic data on Toronto neighborhoods to see if Foursquare data can be a good predictor of any demographic data (and vice versa).
- ▶ The purpose of this exploration is to help citizens and governments use proxy data to understand neighborhoods when desired data is not available. For instance, can the proportion of check-ins of a particular age group for trending restaurants in a particular neighborhood predict the age range of residents in that neighborhood.

Data Sources



Data Sources, cont.

Toronto Open Data

- ▶ Launched in 2009 to meet growing demand for open data.
- ▶ Serves as a politically charged topic of discussion within civic decision-making as policymakers and the civic community navigate open data policies and address socio-political and technical barriers.
- ▶ <https://www.toronto.ca/city-government/data-research-maps/open-data/>

Foursquare API

- ▶ Local search-and-discovery for recommendations of places to go near a specific location.
- ▶ Created a Foursquare developer and made calls to API using a list of Toronto neighborhoods.
- ▶ <https://developer.foursquare.com/>

Predicting Foursquare Information

Predicting Venue Likes

We found that

- ▶ population density
- ▶ postsecondary degree, and
- ▶ unemployment rates

were the best predictors for venue likes.

Predicting Venue Price

Price was difficult to predict with the available demographic data. However, we found that

- ▶ average household size and
- ▶ unemployment rate

Account for a third of the variability in price.

Predicting Venue Rating

We found that

- ▶ percent of working age adults
- accounts for half of the variability of rating and that the higher percentage, the higher the rating will be.

The highest R-squared value we found, however, was for a model looking at

- ▶ unemployment rate,
- ▶ percent living alone,
- ▶ percentage of pre-retirement individuals, and
- ▶ percentage of people living alone.

Predicting Demographic Data

Individual predictors are useful for predicting Foursquare data, but multiple predictors achieve no significant increase in R-squared.

This is likely because the three Foursquare predictors are themselves highly correlated and don't contribute much to the model when added together.

A higher average number of likes for venues in a particular neighborhood indicates,

- ▶ a greater population density
- ▶ a higher percent of working age people
- ▶ a higher percent of people living alone
- ▶ a lower average household size
- ▶ a lower percent of people who have not completed at least a bachelor's degree
- ▶ a higher percent of people who have completed at least a bachelor's degree

A higher average price for venues in a particular neighborhood indicates

- ▶ a lower percent of people who have not completed at least a bachelor's degree
- ▶ a higher percent of people who have completed at least a bachelor's degree

A higher average rating for venues in a particular neighborhood indicates

- ▶ a greater population density
- ▶ a higher percent of working age people
- ▶ a higher percent of people living alone
- ▶ a lower average household size
- ▶ a lower percent of people who have not completed at least a bachelor's degree
- ▶ a higher percent of people who have completed at least a bachelor's degree
- ▶ a higher workforce participation rate
- ▶ a higher workforce employment rate

Recommendation

- ▶ None of the regression models we created had high enough R-squared values for us to make any meaningful recommendations. However, they do show some general trends. For instance,
 - ▶ Neighborhoods that have venues with higher than average ratings and likes, are likely populated with a higher proportion of working age people and lower household sizes. Does this indicate that working age people in smaller (or no) families are more likely to rate or like a venue? Or does this mean that they are more likely to rate it higher? Further analysis is needed.
- ▶ Other interesting, although logistical trend is that neighborhoods with more higher-priced venues (higher average price) are more likely to have more highly educated people living in them. Interestingly, though, average household income was not a good predictor for price, which may suggest that people don't patron expensive venues because they have more money, but because they are more educated.