# Final Report - Predicting Effect of Venues on Mean Housing Price

### Yeranosyan Vahan

## Business Problem

This project is targeted towards real estate developers and entrepreneurs looking to open new venues in the Toronto area. Each neighbourhood is has a unique population and demographic makeup that likely have different interests based on the local population. If a property or storefront becomes available in a certain neighbourhood, we want to know, based on the traffic and interests in that neighbourhood, what type of venue will likely be the most successful by predicting check-ins at that new location.

This tool will allow developers to predict the future traffic of a new venue in a given neighbourhood to make an educated decision on both what type of venue to open, and whether or not the traffic in that new venue will be sustainable for profitable operations.

## Data

This project leverages location data from the city of Toronto for each Toronto neighbourhood, as well as location data from Foursquare including venue check-ins, likes, and details. For convenience, I uploaded data on Github, but you can download it here:

<https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/78cf2678-8cd9-442b-a8e7-ed9ae6e82e01>

and here

https://ckan0.cf.opendata.inter.prod-toronto.ca/download\_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326

## Methodology

After cleaning the data I initially ran a correlation for the presence of venues and the effect on home prices. I started by using venue types with correlation > 0.2 and built a multiple linear regression model to predict home prices based on the number of those venues in each neighbourhood. I found that the R^2 for those models was consistently <0.2 indicating relatively poor fit. I also changed the correlation minimum between 0.1 and 0.3 and found that the optimum R^2 in those ranges was still around 0.2. Since the data is relatively sparse, in that most neighbourhoods did not have the target venues, I build a new model which simply compared a sum of all Foursquare venues in a neighbourhood and used linear regression to try and predict the home price. Once again this yielded a low R^2

## Results

I found that both venues highly correlating to the home price, as well as a summation of all venues was not an adequate predictor for Home Price. While this is a negative result, it does show that simply building or purchasing venues in a given neighbourhood will not implicitly increase home prices, which is important for real estate investors.

## Conclusion

Future work should take into account not only Foursquare venues but also a local crime and other tangible data not included in this project.