

# Final Project Proposal

*Talha Muhammad, Sharon Morris, Ka-Man Chan & Marco Siquera Campos*

*November 16, 2016*

## Final Project Proposal

### Overview

We would like to understand through different sources the reasons for the variation in real estate prices in New York City or Chicago. We would seek to understand how have the rental and property prices changed through out the city and the causes behind it.

### Data sources

We plan on using the following data sources:

#### Real Estate Data (API)

1. Zillow API's provide data on comparable properties and market rates and data is available at the individual property level. The API limits calls to 1000 calls per day.
2. Zillow provides neighborhood boundaries
3. Neighborhood demographic information available from the Census (available via API and CSV)

#### CSV data sources

4. Incident level data on Crime (Chicago all types by incident - available 2001 - Present, NYC only seven types of felony available for 3 Quarters of 2015)

**City of Chicago Crime data** <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>

**Zillow API data** <http://www.zillow.com/howto/api/GetDeepSearchResults.htm>

We will supplement our analysis with other data sources as needed.

### Approach

Take property level information from Zillow API. The data has a variety of interesting variables including comparables. Match it with crime data and demographic information. This can be done for Chicago or New York city.

Develop cross sectional regressions and / or tree based approaches to estimate “true value” based on the characteristics.

### Visualization

- Develop charts and visualization in ggplot;
- Resources allow develop visualization in RShiny.

## **Outputs Requirements**

- Two Different data sources (eg. csv and api)
- Includes one data transformation
- One statistical analysis & graphical analysis validates your data
- One graphic & one statistical analysis that supports your conclusions
- One feature we did not cover in class