

# Project Proposal

Jie Chen - 3816600200    Jason Cheves - 7562936534    Jieun Choi - 3387252157  
Jiaying Gu - 2134487579    Yiqiong Zhu - 1062031378

## Problem of Interest

Los Angeles is the second largest city in the United States. Each year, thousands of people move to Los Angeles looking for ideal places to call their next home. Although price is a major factor influencing their decision, neighborhood condition, community attributes and other factors also play a significant role in their decision process. Our project would like to create an overall view of each neighborhood with these factors to help future residents locate their ideal living situation. If the project is successful, we can easily expand our analysis to other major cities utilizing similar methodologies and data.

## Methodology

We will combine normal factors seen on typical real estate website from schools, tax information, etc. along with other important factors like demographics, safety, vitality, and culture. This will be unsupervised research since we currently do not have this information.

## Dataset Description

Our datasets cover various aspects we would like to explore. Since our focus is on neighborhoods and communities, we find these variables the most important for our research:

1. Housing information: housing units, council district, addresses
2. Demographics: population, race, household size, gender distribution, age distribution, employment distribution, average annual wage, education level
3. Safety: crimes, building and safety inspections, building permits
4. Vitality: events, new commercial buildings, new construction permits, walkability

Majority of our data comes from Los Angeles Open Data (<https://data.lacity.org>). To ensure the quality of our project, we may include other datasets for further research.

The following are a few sample datasets we consider most important to our project:

### 2010 Census Populations by Zip Code

- 1) Source: <https://data.lacity.org/dataset/2010-Census-Populations-by-Zip-Code/nxs9-385f>
- 2) Dimension: 319\*7
- 3) Variables: Zip Code, Total Population, Median Age, Total Males, Total Females, Total Households, Average Household Size

### Addresses in the City of Los Angeles

- 1) Source: <https://data.lacity.org/A-Well-Run-City/Addresses-in-the-City-of-Los-Angeles/4ca8-mxuh>
- 2) Dimension: 996K\*18 (real time data, updated every day)
- 3) Variables: House number ID, Street Name, Latitude, Longitude, Council District etc.

### Crime 2012-2016

1. Source: <https://data.lacity.org/A-Safe-City/Crimes-2012-2016/s9rj-h3s6>
2. Dimension: 1.14M\*14
3. Variables: Date reported, Date Occurred, Time Occurred, Area, Crime Code, Crime Code Description, etc.