**Project Proposal – Eric Kim, Ryan Liu, Pedro Juarez**

**Goal of the project**

When people are searching for new places to live, they are limited in the range of factors they can consider. Apart from the basic information about the house itself, such as rent or real estate value, the lack of data on the neighborhood that is available online prevents people from making well-informed decisions. Thus, they tend to rely on a narrow scope of knowledge, including the house’s proximity to workplace/school and their basic perceptions on the neighborhood.

However, there are much more factors to consider when people consider moving into a new house. For example, people may be interested in whether the community is safe, or whether there are good restaurants nearby. They may also want to have enough parking spaces, while avoiding traffic jams or frequent road closures during rush hours. Some of these questions can be answered by visiting houses and exploring the neighborhood, but this will be time-consuming and costly. We are trying to build a software that would help people make better decisions by providing them with relevant data on the communities that they are considering moving in.

**Data Sources**

1. City of Chicago Data Portal (<https://data.cityofchicago.org>)

This would be our main source of public data on the various neighborhoods in Chicago.

2. Zillow(<http://www.zillow.com/howto/api/GetSearchResults.htm>)

Zillow is an online real estate database company that provides open API for getting search results. By typing in address or state+city / zip code, the API would return estimates for rent or real estate and other relevant information in an xml format.

3. Yelp Dataset(<https://www.yelp.com/dataset_challenge/dataset>)

We will use the dataset provided by Yelp to show what types of restaurants and other businesses are present in each neighborhood.

**New Tools to Use**

1. Display in map: We will use a tool that will allow us to display the houses that meet the search conditions on the map as dots.

2. Web interface: In order to allow the users to use our software via web browsers, we will use Django to create a web interface.

3. Algorithm for weighting search conditions: We want our software to weight the search conditions based on how a user sets the priority.

4. Algorithm for extracting data on neighborhoods, houses, amenities, etc. that are within a certain distance from a point specified by the search query.

**Timeline and Basic sketch**

Week 4: Scrape the data form each dataset and understand how each dataset is structured.

Find additional useful data sources if available

If real-time data is needed (Zillow’s price estimate), determine how to update the existing data.

Week 5: Format the dataset so that each dataset can be linked via geographical information.

Determine where to store the processed data in which structure.

Determine what search conditions the users can use and what information would be provided as an output.

Week 6: Work on algorithm for weighting the search conditions and ranking/filtering housing options.

Make the software to extract data on neighborhoods, houses, amenities, etc. that are within a certain distance from a point specified by the search query. (To avoid unnecessary tasks)

Week 7: Find other ways to improve the processing speed.

Create a web interface based on the python program we created.

Week 8: Work on displaying the ranked/filtered results in a map.

Week 9: Check if the search results make sense and fix bugs.

Week 10: Finalize the project and prepare for the final presentation.