Neighborhood Ranking Scores:
Analyzing the community quality of NYC real estate properties

(Team 53) Harsharn Kaur, Khang Tran, Katherine Tang, Ian Stalter, Xue Yu
{hkaur45, ktran322, ktang79, istalter3, xyu417} @gatech.edu

Georgia Tech

Motivation/Introduction

Neighborhoods are crucial in shaping individual and familial well-being. However current real estate applications, such as Zillow, Redfin, and Realtor, are overwhelmingly based on housing attributes (e.g. number of bedrooms/bathrooms, selling price) and nearby school rankings. The only neighborhood-relevant information is the walk score, transit score, and bike score buried under other information, and it does not account for crime history or relative safety of the area.

We aim to rank New York city neighborhoods based on user preferences for housing affordability, accessibility, and safety to aid house seekers by providing localized information for neighborhoods, helping them search for neighborhoods that best suit their objectives.

Data

Data was pulled from multiple sources

Zillow New York Housing data available on Kaggle from 1/20/21.
Size: 55,379 house listings after removing duplicate rows, missing values, and outliers for fields: ZIP Code, Price, Latitude, Longitude, Living Area, Walk Score, Transit Score

New York Crime complaint data set from NYPD 2018 - 2022.
Size: 2.2 million record.
Details of the relevant columns are mentioned in the following fields: complaint ID, Date, Latitude, Longitude, Category

In addition, we pulled data to a CSV file from the **US census** for **income by zip code**.
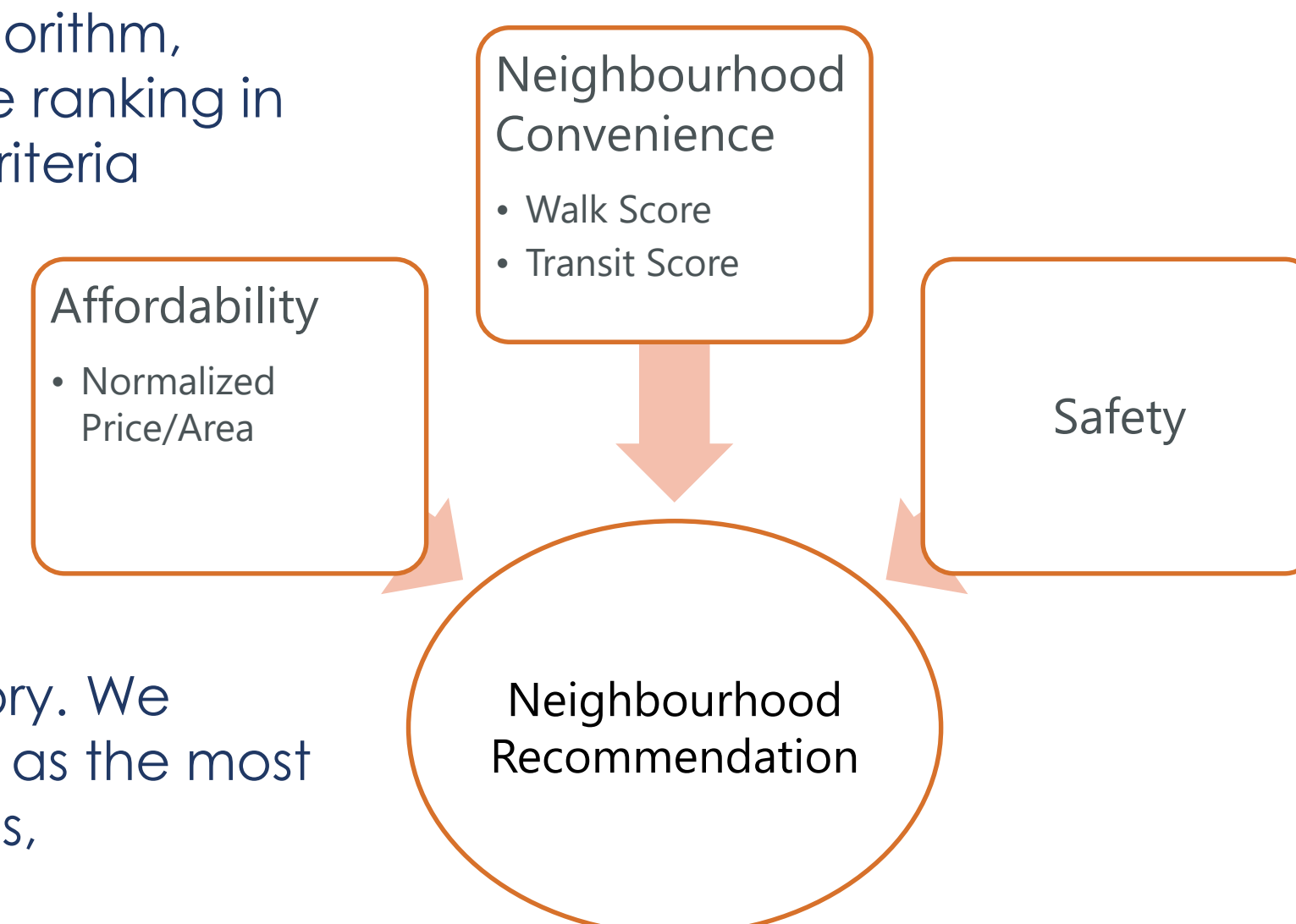
Approach

To optimize for multiple attributes for a neighborhood recommendation, we leveraged a **multi-criteria decision analysis (MCDA)** algorithm, Technique for the Order of Prioritization by Similarity to Ideal Solution (TOPSIS). MCDA is a commonly used approach for performance ranking in decision-making contexts, offering advantages in transparency, objectivity, and flexibility that allows us to define decision-making criteria and reduce personal biases.
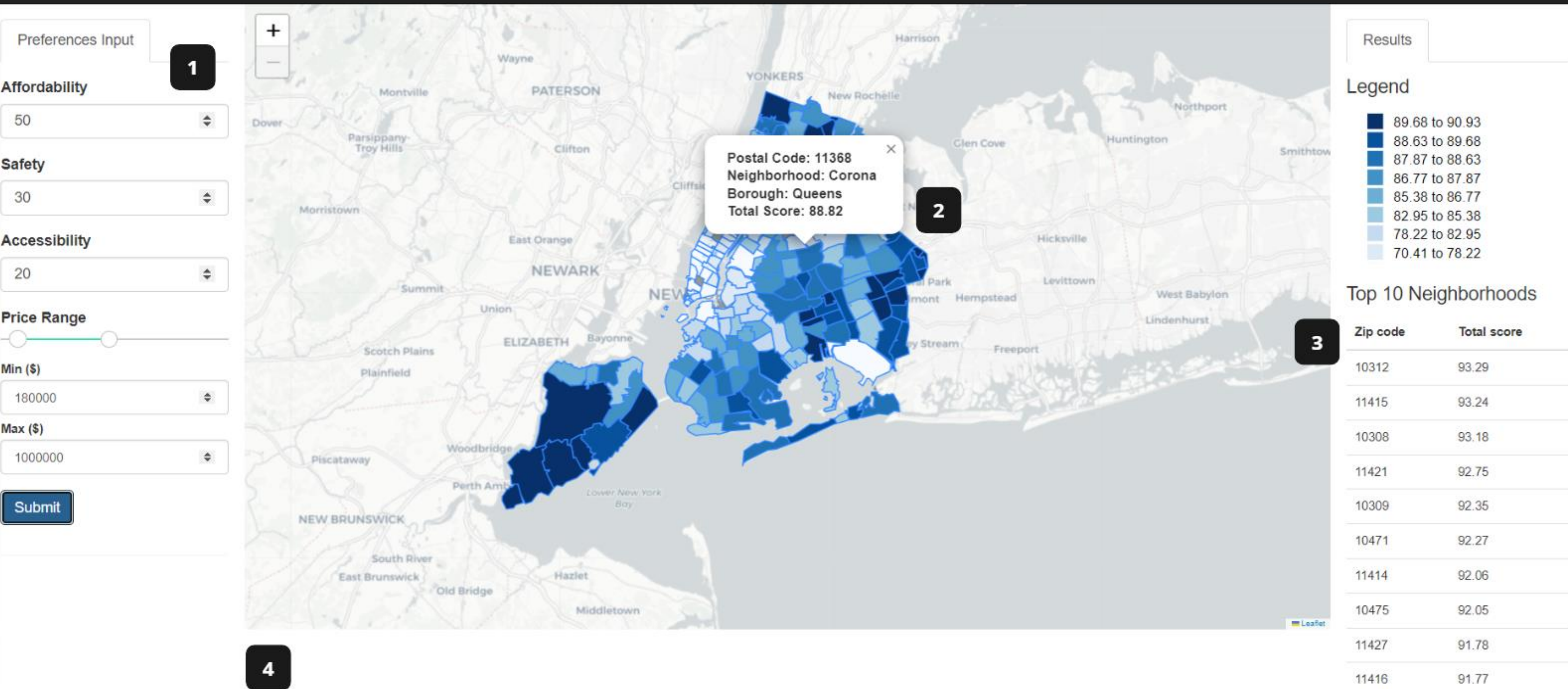
We calculate a final neighborhood ranking score for each zip code that is determined by

$$Score = functionOf(Affordability, Accessibility, Safety, User\ Input\ Weights)$$

- **Affordability** is the average Asking Price per Living area (per sq feet) for all listings in the zip code
- **Accessibility** is the average of equal weighted metric of Walk Score and Transit Score for all listings in in the zip code.
- **Safety** is calculated by the count of each crime category (Violation, Misdemeanor, Felony) for all zip codes in each crime category. We normalize the above counts per 1000 residents to get the corresponding crime rates for each zip code. Since felony is considered as the most severe of all crime categories, we assign higher weights to crime rate corresponding to felony with the final crime score defined as,

$$Crime\ Score(i) = 0.25 \times Violation\ Rate(i) + 0.25 \times Misdemeanor\ Rate(i) + 0.5 \times Felony\ Rate(i),\ where\ i = zip\ code$$

Neighbourhood Convenience
- Walk Score
- Transit Score

Affordability
- Normalized Price/Area

Safety

Neighbourhood Recommendation

The interactive UI displays the top 10 neighborhood results based on example preference weight input.

[1] Each preference is normalized so users can insert any arbitrary numerical value.

[2] All zip codes are assigned scores based on the defined formula. Areas on the choropleth are also user interactive.

**[3]** Final neighborhood recommendation ranking and scores from provided user input.

[4] Clicking on a zip code will dynamically display additional info and break downs of each preference attribute for that area. This includes comparing the three attribute percentile scores, safety incident distribution, and price distribution charts of each zip code.

Top 10 Neighborhoods

Zip code	Total score
10312	93.29
11415	93.24
10308	93.18
11421	92.75
10309	92.35
10471	92.27
11414	92.06
10475	92.05
11427	91.78
11416	91.77

Experiments and Results

We conducted a survey about the application's usability to gather user assessment on the features provided to enhance their home buying experience. In total, we had 25 volunteers who were about evenly split between male and female with the majority aged between 25-44 review our app.

Overall subjective response and feedback was quite positive, with users giving the application an average score (out of 10) of **8.20 overall** and **8.96 for ease of use** – safety was overwhelmingly a primary concern for most participants with **56%** placing it as their topmost preference. Majority of survey responses also claimed that they were able to gain some valuable insights or new learnings about NYC neighborhoods and would use the app during the house search process.

In addition, we performed a comparison evaluation between our ranking list for each attribute versus other available sources such as New York Crime Map. Of the Top 10 ZIP Codes with highest crime rates calculated using MCDA algorithm (left), 8 zip codes correspond to precincts with highest crime rates as per NYC Crime Map (right).
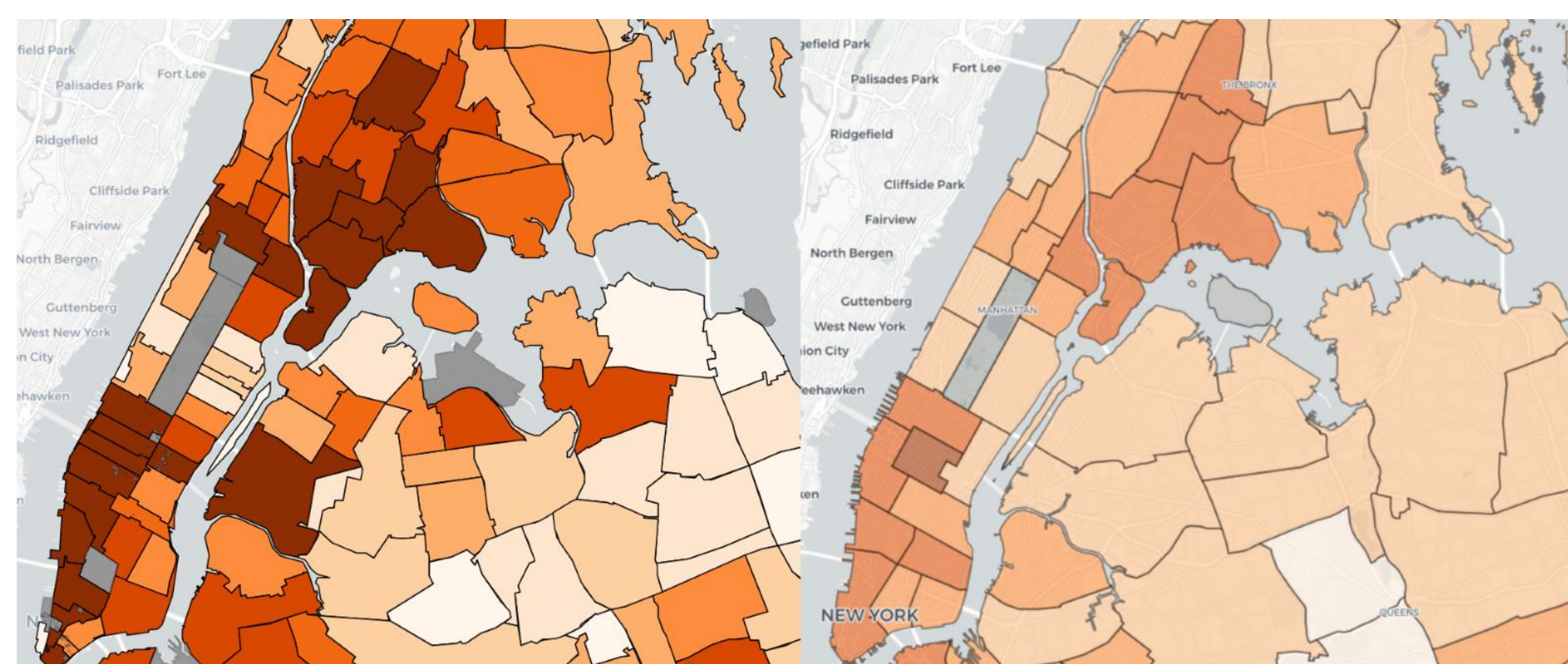
Rank	Options
1	Safety
2	Affordability
3	Accessibility

First choice — Last choice

Ranking Application Assessment - Users...
Yes / No
Gained insights / Would use