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

1 INTRODUCTION

1.1 Motivation

Neighborhoods play a powerful role in shaping individual and familial well-being. Violence and unhealthy local conditions have an indelible impact on adults' and children's health and growth [1, 4, 14]. Studies show the majority gravitate towards properties that contribute to children's futures, with safety being a top consideration [10, 11]. House seekers tend to use the pattern of the small house market segments that differ by geography and property characteristics i.e. neighborhoods rather than individual house recommendations [17]. Therefore, we aim to rank New York city neighborhoods based on user preferences for housing affordability, accessibility, and safety and aid house seekers by providing localized information for neighborhoods, helping them search for neighborhoods that best suit their personal objectives.

1.2 Problem Definition

Current real estate applications, such as Zillow, Redfin, and Realtor, allow for search results to be filtered and sorted solely based on housing attributes (e.g. number of bedrooms, bathrooms, selling price) and nearby school rankings. This can cause the number of relevant search results to be overwhelming. And the only available information relevant to a neighborhood itself is the walk score, transit score, and bike score. However, these values are only viewable on individual property listings and buried below all the other features and amenities, making it difficult to understand how a neighborhood is. Additionally, it does not account for the recent history of crimes and the relative safety of the area.

2 LITERATURE SURVEY

The preference for safer neighborhoods is shared by low-income and high-income families regardless of their financial constraints [9]. Findings from [18] suggest a main factor affecting residential concerns about neighborhood safety was the broken window theory, thus we intend to follow this idea by ranking neighborhoods using local crime data. Crimes like aggravated assault, vandalism, and robbery have significant negative price elasticity in comparison to property crimes like motor theft and burglary [2, 5, 10, 20]. We can leverage this to assign higher importance weights for the above-mentioned crimes in comparison to other crime categories (motor thefts, burglary) when calculating safety scores.

In addition to safety, the overall quality of the properties' neighborhood is one of the top deciding factors by homebuyers [16]. This can be determined by its nearby resources and facilities. Walk Score can be used as a significant proxy for good neighborhoods in New York as neighborhoods with high walk scores usually represent good proximity to amenities (schools, parks, grocery stores, restaurants, etc.) [8]. The high walk scores exhibit a positive correlation with higher residential and commercial values [3, 19] and higher education levels and per capita GDP [15]. [13] provides models of housing prices based on individual housing attributes and Walk scores. We plan to also integrate a similar methodology into our neighborhood rankings model and then expand on the attributes considered by including safety and poverty levels as additional factors.

3 PROPOSED METHOD

Choosing a neighborhood is a multi-criteria optimization problem where every individual would have different preferences for conflicting attributes like affordability, safety, and accessibility. Neighborhood ranking has been previously done in the context of Boston based on economic and environmental factors like safety, education, environment, etc, using Multi-Criteria Decision Analysis algorithms [12]. Using this as a base, we provide prospective house buyers/renters the flexibility to use Zillow New York listings data from 2021, New York Crime Data, Poverty and Walk Score and Transit Score info, and individual preferences for affordability, safety, and convenience to generate their personal rankings for zip codes in New York.

3.1 Data Collection & Processing

The first dataset was collected on 1/20/21 and consists of 75,629 house listings on Zillow.com. Walk Scores and Transit Scores are retrieved for each property listing using Latitude and Longitude columns via Walk Score API. We assume that Walk Score when averaged for a zip code would not change drastically from 2021 to 2023. The final dataset contains 55,379 house listings after removing duplicate rows, missing values, and outliers for important columns listed in **Table 1**.

Table 1: Kaggle Zillow New York Housing data

Column	Description
Zip Code	Property zip code
Price	Current property asking price
Latitude	Latitude coordinates
Longitude	Longitude coordinates
Living Area	Property Square Footage
Walk Score	Walk Score given lat, long
Transit Score	Transit Score given lat, long

For calculating safety scores, we used the New York Crime complaint data set from NYPD from 2018 to 2022. After data processing, we ended up with 2.2 million rows. Latitude and longitude information for each complaint are reported which are eventually used to get zip codes. Details of the relevant columns are mentioned in **Table 2**.

Table 2: Filtered New York Crime compliant data

Column	Description
ComplaintID	Unique ID of reported incident
Date	Reported date of complaint
Latitude	Incident location latitude
Longitude	Incident location longitude
Category	Reported incident category

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

3.2 Algorithm

The multi-criteria decision analysis (MCDA) technique, specifically the Technique for the Order of Prioritization by Similarity to Ideal Solution (TOPSIS) was leveraged to optimize for multiple conflicting attributes. MCDA is a commonly used approach for performance ranking in decision-making contexts, with applications ranging from critical mission planning [21] to strategy evaluation [22], offering advantages in transparency, objectivity, and flexibility [12]. It defines decision-making criteria and reduces personal biases. The algorithm identifies a hypothetical ideal solution with attributes set as maximum/minimum of all values for that attribute in the data set depending on whether we want to maximize/minimize that attribute, respectively. Each option is evaluated by calculating the Euclidean distance between the ideal solution and row values. The smaller the Euclidean distance (*dp*), the closer the alternate is to the ideal solution and hence the higher the ranking. The above analysis is done for the hypothetical Worst Solution as well where higher Euclidean distance (*dn*) makes the option more desirable and higher ranked,

$$FinalScore(i) = dn(i)/(dp(i) + dn(i))$$

where i = all alternatives.

For our purposes, we use Affordability, Safety, and Accessibility as criteria for our project. The alternatives are the options to be evaluated for the criteria defined i.e. zip codes. Therefore, the final score calculated for each zip code can be described as a function of weights for each criterion along with attribute values, with weights for each criterion input using the user interface as preference for attribute:

$$Score = func(Affordability, Accessibility, Safety, UserInputWeights)$$

Where

- *Affordability* is defined for each zip code as the average Asking Price per Living Area (per sq feet) for all listings in that zip code. Lower ratios for a zip code equate to desirability on average. So we would like this ratio to be **minimized** as much as possible for the Ideal Solution.
- *Accessibility* for a zip code is the average of equal weighted metric of Walk Score and Transit Score for all listings in the zip code. We want to **maximize** accessibility for the Ideal Solution as higher accessibility should have higher corresponding rankings.
- *Safety* is calculated by the count of each crime category (Violation, Misdemeanor, Felony) for all zip codes in each crime category. We normalize

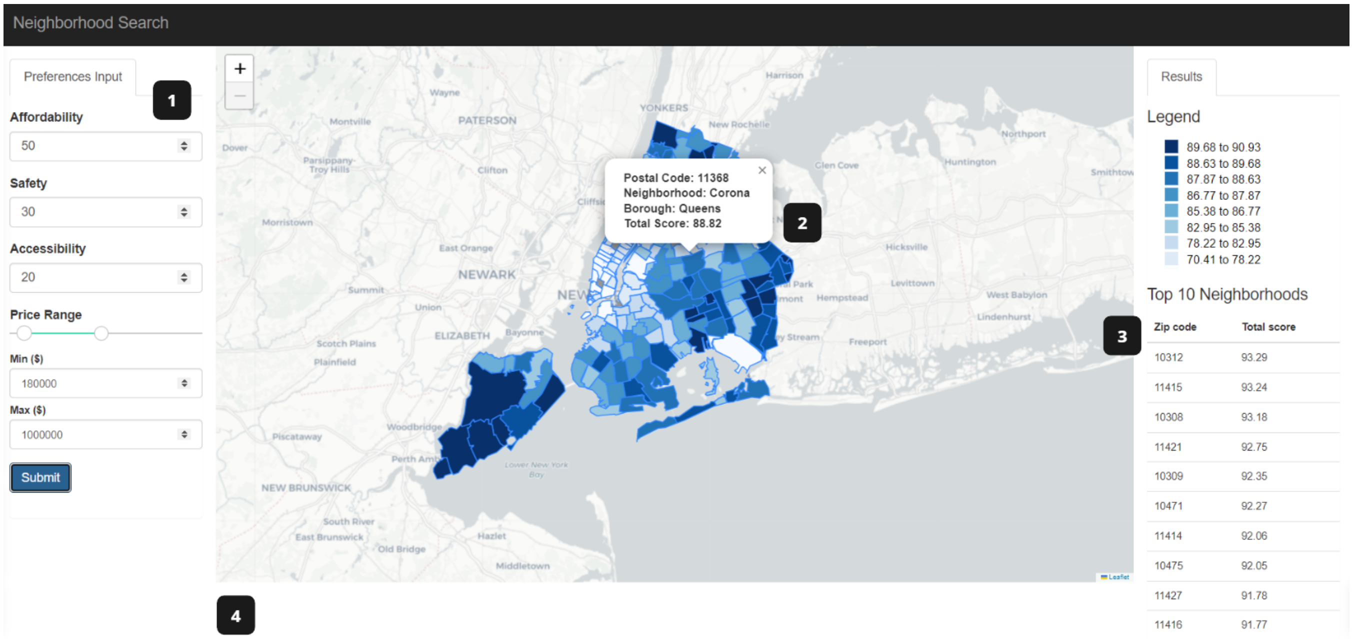

Figure 1: Illustration of interactive app with example user inputs and results. (1) Each preference is normalized so users can insert any arbitrary numerical value. (2) All zip codes are assigned scores based on the defined formula. Areas on the choropleth are also user interactive. (3) Final neighborhood 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.

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. The overall Crime Score can be calculated with $CrimeScore(i) = 0.25 \times ViolationRate(i) + 0.25 \times MisdemeanorRate(i) + 0.5 \times FelonyRate(i)$ where i = zip code. Zip codes with lower overall crime scores would be considered safer and should have higher final scores. We would want to **minimize** the overall crime rate for the Ideal solution.

Since the input scores for each zip code will be different unit scales, we need to provide normalization method to make attributes dimensionless. We will be using vector normalization as it tends to perform better in comparison to other normalization methods [23, 24].

3.3 UI Design

[6] suggests that choropleth maps prove to be most useful for visualizing aggregated patterns in and across defined boundaries. Therefore for our purposes, we implemented an interactive choropleth of NYC to display ranking results divided by zip codes. Each area on the map capable of dynamic user interaction that displays additional details such as its borough, neighborhood name, and ranking score. Users can input and adjust their preference for criteria (affordability, safety, and accessibility) to calculate the neighborhood ranks. In case of no user input, equal weights for each preference are used by default.

In addition to the ranking results, we include analytical charts for each zip code when user interacts with the zip code on choropleth to provide further contextualization about the data utilized in the ranking algorithm. This consists of percentile comparisons of each attribute as well as a count of houses available in that zip code in different selling point ranges. These additional visualizations are available for all zip codes and users can dynamically change their view of which area to display this information.

Our application package consists of two primary components: a Python Flask backend server and a single page interactive UI application built using JavaScript. The backend server uses analytical and data manipulation packages libraries like scipy, scikit-mcda, numpy etc. for performing dataset aggregations and zip code ranking. The front-end employs the following libraries: Bootstrap for building the UI, d3.js and Leaflet.js for the interactive visualization and jQuery for DOM manipulation. Communication between the frontend and the backend components is achieved using JSON-based asynchronous API calls. Furthermore, as part of our experimentation (see Section 4), we have deployed our tool to Google Cloud Platform for easier accessibility to testing participants.

3.4 Innovation

Our innovations versus other available methodologies and applications are:

- *Single criteria vs Multi criteria optimization*: Conventional real estate listing websites use one criterion at a time to recommend individual houses. We use neighborhood safety and ease of commute and accessibility as additional factors to provide a holistic view of the neighborhood. To check the model accuracy, we can compare the relative ranking of all zip codes using safety as the only attribute in our model and compare it against NYC crime map available online as detailed in Section 4.2.
- *MCDA*: Real estate listing websites present search results in either increasing or decreasing order of the one attribute selected. MCDA takes multiple attributes at once to reach a ranking for all the provided alternate options. We expect to see results sorted in increasing/decreasing order when modeling for one attribute.
- *Interactive UI*: We provide flexibility to every individual to incorporate their preferences for safety, affordability, and accessibility as the weights to be used in the MCDA ranking algorithm. For each resultant zip code, the user can access the distribution of properties available in different price ranges. We also provide the breakdown of percentiles for each ranking attribute for the selected zip code as described in **Figure 2**.

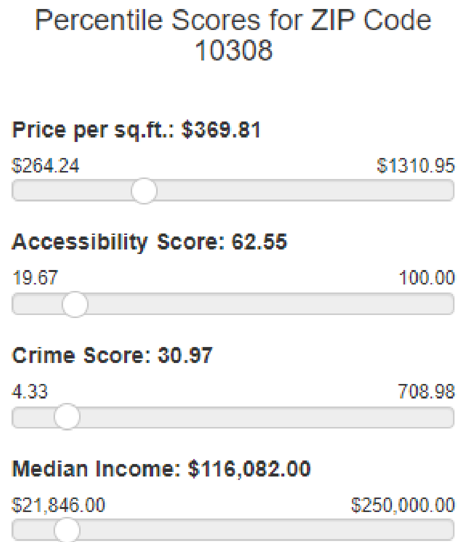

Figure 2: Comparison of sample zip code 10308 against the entirety of New York city for the three user preference attributes and median household income.

4 EXPERIMENTS & EVALUATION

Because our application aims to provide a user-friendly platform for house seekers, we will conduct a user investigation to gather subjective feedback on the features provided in our application to enhance the user experience. We also observe our application results with outside neighborhood rank lists as an objective evaluation and direct comparison with other data. We hypothesize that our application has similar results for individual criteria but offers more personalized insights to users due its multi-criteria programming to help them make informed decisions when choosing a house in their desired neighborhood.

4.1 User Experiment & Results

One of the most important standards to evaluate an application is usability. Usability is defined by three aspects: more efficient to use, easier to learn, and more user satisfaction [7]. We used laboratory experiments to test usability by letting users interact with the application through a site deployed via Google Cloud Platform. All users could independently explore the application to evaluate its utility then respond to a quick survey.

We recruited 25 volunteers for the application testing. The follow up questionnaire was designed with questions asking participants first for their basic information (age, gender, familiarity with New York, and attribute preference order) and then about the application experience (overall user experience, how easy to use, whether they would use it in the future, the most attractive feature, and suggestions for modifications and new potential features. There was almost an even split between genders, with 11 females and 13 males (1 preferred not to specify) of all ages, mainly between 25 to 44 years old. Most participants have no or only some knowledge of New York, with several participants being very familiar with New York. Of the participants, more than 50% listed safety as their first consideration among the options of affordability, accessibility, and safety while affordability came in second with 36% listed considering it as their first consideration and 48% placing it as their second choice.

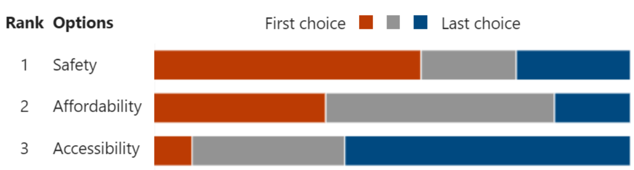

Figure 3: User survey showed safety is top priority for most.

In terms of technical usability of our application, we received an average of 8.20 out of 10 for the overall experience and 8.96 out of 10 for ease of use and how intuitive it was to figure out.

In addition, we were also able to gain valuable insights to help us improve our application from 17 of the volunteers. Users gave some interesting feedback on potential modifications – such as detailed views of safety factors, houses, prices and a better explanation for the ranking score. A few users said that it would be best used as a secondary feature on a rental site. One also mentioned that it would be useful to have a comparison view between neighborhoods. Further discussion on these points for changes and improvements are summarized in section 5.

Overall, the survey results confirmed that our application is easy to use and has relatively high user satisfaction, indicating a high level of usability.

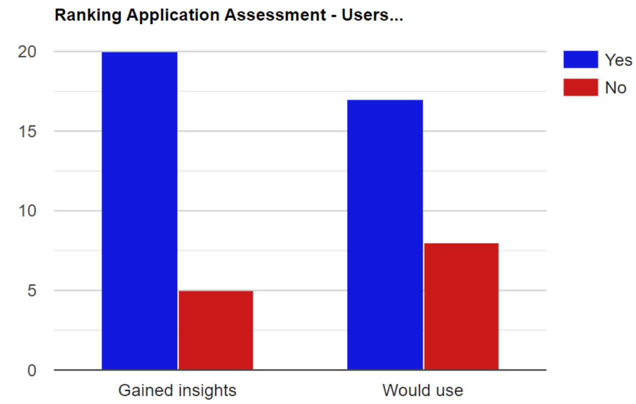

Figure 4: Majority of participants claimed that they were able to gain some valuable insights or new learnings about NYC neighborhoods and would use the app during the house search process.

4.2 Rank Comparison & Evaluation

In addition to conducting usability surveys, we conducted a quantitative assessment between results provided by our scores versus other currently available tools for certain preferences. New York Crime Map is an interactive visualization displaying precinct level crime rates developed and maintained by the City of New York. To obtain crime scores instead of safety scores, we flip MCDA signal for safety from -1 to 1 and set non-zero preference for safety (zero for affordability and accessibility. We compared the resultant crime scores for each zip code against the New York Crime Map using New York Crime data set for 2022.

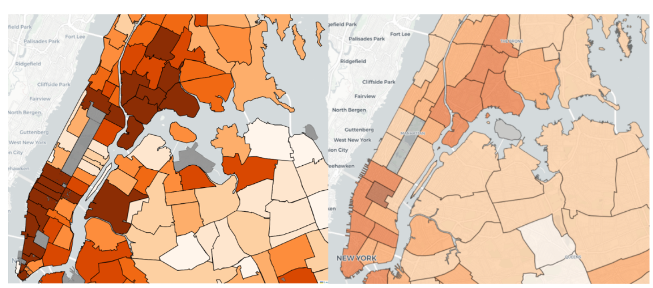

Figure 5: Zip level crime rates from MCDA algorithm (left) and precinct level crime rate from New York Crime Map (right).

Of the Top 10 ZIP Codes with highest crime rates calculated using MCDA, 8 zip codes correspond to precincts with highest crime rates as per NYC Crime Map (shown in **Figure 5**). Difference in exact ranks or scores might be due to difference in geographical boundaries of zip codes and precincts and data used in calculations i.e., Crime Map uses crime data for Year 2022 only whereas MCDA uses crime data from 2018 to 2022 (inclusive).

On the topic of affordability, we also observe that there exists a positive linear relationship between average price per square feet and median income levels for a zip code visible in **Figure 6**. One plausible explanation for this positive relationship is that higher income households tend to allocate higher budgets for their housing requirements.

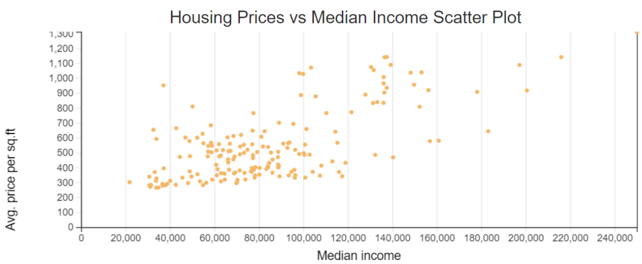

Figure 6: For average price per square feet versus median income levels, Manhattan borough has the least affordable housing of all neighborhoods as most outliers correspond to zip codes in Manhattan.

5 DISCUSSION, LIMITATIONS & FUTURE WORK

We believe our project will benefit house seekers by providing localized information for neighborhoods. The application interface we built empowers them to search for neighborhoods that best suit their personal objectives and help provide more insights to the local area. Thanks to the multi-criteria approach of this tool, users are able to better narrow down their search area for their ideal home. In addition, real estate agents and investors would also benefit from the application's neighborhood analysis as they would be able to better leverage and calculate property value based on the overall quality of the area

Due to data and time limitations, our application is restricted in scope and evaluation. To balance workload with effective results in the limited project time, we only focused on gathering and processing data from New York. Thus, the application is only useful within the New York city area. Second, the currently used data is from past years and may not accurately reflect the up-to-date conditions. To update results, we will need to manually gather updated datasets or connect to real-time incoming data. Third, we used only three criteria, safety, accessibility, and affordability, to reflect neighborhood quality and as such the neighborhood ranking results may not fully satisfy all users' needs. There are more factors that contribute to neighborhood quality.

Potential future extensions would be to expand the scope of the application with more available data and to include more areas. If neighborhood information related APIs are available or be built, the application could provide the most up-to-date neighborhood quality estimates. We could also continue adding more neighborhood related factors that concern home seekers. Several survey participants also mentioned adding renting information to expand the use of the application for renters. An even bigger picture would be getting tied-in with companies that rent apartments.

6 CONCLUSIONS

While neighborhoods play a significant role in peoples' living quality, current real estate applications barely incorporate neighborhood factors in the house seeking process. To fill this gap, our team introduced a neighborhood ranking application. The application uses integrated datasets from different resources to calculate the safety, accessibility and affordably scores of each neighborhood with an interactive UI that allows users to input their weighted preferences to these three criteria and show them the areas that best suit their needs.

From our survey results, most volunteers found our application provided valuable information and is extremely useful in helping them find their targeted neighborhoods, especially for home seekers who are not familiar with New York, the application provided a starting point for house searching. It helped users limit their effort inside the most suitable neighborhoods, instead of aimlessly searching everywhere. The application could be a great auxiliary tool in the house seeking process.

For the entire project work, all team members have contributed a similar amount of effort.

REFERENCES

[1] Elena M. Andresen, Theodore K. Malmstrom, Mario Schootman, Fredric D. Wolinsky, J. Philip Miller, and Douglas K. Miller. 2013. Observer ratings of neighborhoods: comparison of two methods. *BMC Public Health* 13, 1 (2013). https://doi.org/10.1186/1471-2458-13-1024

[2] Vania Ceccato and Mats Wilhelmsson. 2019. Do crime hot spots affect housing prices? *Nordic Journal of Criminology* 21, 1 (Sept. 2019), 84–102. https://doi.org/10.1080/2578983x.2019.1662595

[3] Joe Cortright. 2009. Walking the Walk: How Walkability Raises Home Values in U.S. Cities. https://nacto.org/docs/usdg/walking_the_walk_cortright.pdf

[4] Catherine Cubbin, Susan Egerter, Paula Braveman, and Veronica Pedregon. 2008. Where We Live Matters for Our Health: Neighborhoods and Health. http://hdl.handle.net/10244/638

[5] Jamein P. Cunningham. 2016. An evaluation of the Federal Legal Services Program: Evidence from crime rates and property values. *Journal of Urban Economics* 92 (2016), 76–90. https://doi.org/10.1016/j.jue.2015.11.008

[6] John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson. 2014. Mapping Crime: Understanding Hot Spots. https://www.ojp.gov/pdffiles1/nij/209393.pdf

[7] Nayebi Fatih, Jean-Marc Desharnais, and Alain Abran. 2012. The state of the art of mobile application usability evaluation. *25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (April 2012), 1–4. https://doi.org/10.1109/CCECE.2012.6334930

[8] C. Michael Hall and Yael Ram. 2018. Walk score® and its potential contribution to the study of active transport and walkability: A critical and systematic review. *Transportation Research Part D: Transport and Environment* 61 (June 2018), 310–324. https://doi.org/10.1016/j.trd.2017.12.018

[9] Hope Harvey, Kelley Fong, Kathryn Edin, and Stefanie DeLuca. 2020. Forever Homes and Temporary Stops: Housing Search Logics and Residential Selection. *Social Forces* 98, 4 (June 2020), 1498–1523. https://doi.org/10.1093/sf/soz110

[10] Keith Ihlanfeldt and Tom Mayock. 2010. Panel data estimates of the effects of different types of crime on housing prices. *Regional Science and Urban Economics* 40, 2-3 (May 2010), 161–172. https://doi.org/10.1016/j.regsciurbeco.2010.02.005

[11] Erdener Kaynak, Ali Kara, and Azamat Maksüdünov. 2022. An Empirical Investigation of Home Buying Behavior in a High-Context Culture: A Strategic Marketing-Oriented Approach. *International Journal of Housing Markets and Analysis* (Aug. 2022). https://doi.org/10.1108/IJHMA-07-2022-0095

[12] Despoina Kazepidou, Maria Rousi, Vasiliki Gkatziaki, and Athena Vakali. 2018. Demo: Bostonhood: A Multi-criteria Platform for Ranking City Neighborhoods. *IEEE International Conference on Smart Computing* (2018). https://doi.org/10.1109/smartcomp.2018.00060

[13] Sangwan Lee and Michael Smart. 2020. Public Transportation Accessibility, Proximity to New York City, and Residential Property Values in New Jersey. *Hatfield Graduate Journal of Public Affairs* 40, 1 (May 2020). https://doi.org/10.15760/hgjpa.2020.4.1.6

[14] Christine Leibbrand, Ryan Gabriel, Chris Hess, and Kyle Crowder. 2021. When push comes to shove: Local violence and residential mobility among families. *Journal of Marriage and Family* 85, 5 (May 2021), 1249–1267. https://doi.org/10.1111/jomf.12769

[15] Christopher B. Leinberger and Patrick Lynch. 2014. Foot Traffic Ahead: Ranking Walkable Urbanism in America's Largest Metros. https://nacto.org/wp-content/uploads/2016/02/1_Leinberger-Lynch-Foot-Traffic-Ahead_2014.pdf

[16] National Association of Realtors. 2022. 2022 Home Buyers and Sellers Generational Trends Report. https://www.nar.realtor/sites/default/files/documents/2022-home-buyers-and-sellers-generational-trends-03-23-2022.pdf

[17] Monika Piazzesi, Martin Schneider, and Johannes Stroebel. 2020. Segmented Housing Search. *The American Economic Review* 110, 3 (March 2020), 720–759. https://doi.org/10.1257/aer.20141772

[18] Ronald O. Pitner, ManSoo Yu, and Edna Brown. 2012. Making neighborhoods safer: Examining predictors of residents' concerns about neighborhood safety. *Journal of Environmental Psychology* 32, 1 (March 2012), 43–49. https://doi.org/10.1016/j.jenvp.2011.09.003

[19] Gary Pivo and Jeffrey D. Fisher. 2011. The Walkability Premium in Commercial Real Estate Investments. *Real Estate Economics* 32, 9 (March 2011), 185–219. https://doi.org/10.1111/j.1540-6229.2010.00296.x

[20] Devin G. Pope and Jaren C. Pope. 2012. Crime and property values: Evidence from the 1990s crime drop. *Regional Science and Urban Economics* 42, 1-2 (Jan. 2012), 177–188. https://doi.org/10.1016/j.regsciurbeco.2011.08.008

[21] Madjid Tavana and Adel Hatami-Marbini. 2011. A group AHP-TOPSIS framework for Human Spaceflight Mission Planning at NASA. *Expert Systems with Applications* (2011). https://doi.org/10.1016/j.eswa.2011.04.108

[22] Hasan Ture, Seyyide Dogan, and Deniz Kocak. 2018. Assessing euro 2020 strategy using multi-criteria decision making methods: Vikor and topsis. *Social Indicators Research* 142, 2 (2018), 645–665. https://doi.org/10.1007/s11205-018-1938-8

[23] Nazanin Vafaei, Rita A. Ribeiro, and Luis M. Camarinha-Matos. 2018. Data normalisation techniques in decision making: Case study with topsis method. *International Journal of Information and Decision Sciences* 10, 1 (2018), 19. https://doi.org/10.1504/ijids.2018.090667

[24] Aydın Çelen. 2014. Comparative analysis of normalization procedures in TOPSIS method: With an application to Turkish deposit banking market. *Social Indicators Research* 25, 2 (2014), 185–208. https://doi.org/10.15388/informatica.2014.10