Location Matchmaking Platform

Marcellus Mohanna, Zain Khan, Sridevi Perumal, Gabriel Hernandez, Sean Nima

1 INTRODUCTION

Americans move approximately 11 times during their lifetime [11], facing complex decisions with significant life impacts. Our Location Matchmaking Platform consolidates scattered information into a single interface, highlighting optimal areas based on five key features:

- **Affordability**: Housing costs, transportation, taxes, and cost-of-living indexes.
- **Walkability**: Pedestrian-friendly scores for infrastructure and proximity to services.
- **Weather Patterns**: Temperature, precipitation, and air quality metrics.
- **Employment Opportunities**: Unemployment rates and salary benchmarks.
- **Well-being**: Access to exercise opportunities, food environment, access to parks, and primary care physicians.

This integration transforms location selection from overwhelming research into a streamlined experience aligned with users' priorities. Zhao et al. [20] also highlight that using features from multiple data sources can significantly improve model accuracy, whether those be objective or subjective parameters.

2 PROBLEM DEFINITION

Despite the critical importance of location decisions, current tools fail to provide comprehensive, personalized analysis specifically due to scattered information across various sources. These limitations force users to make suboptimal decisions based on incomplete information or invest excessive time in manual research, leading to decision fatigue and potential relocation dissatisfaction. Our goal is to create a single platform that simplifies this process by gathering all relevant data in one place and allowing users to customize their search based on what they value most.

3 LITERATURE SURVEY

Location decisions involve multiple interdependent factors such as crime rates, green space, affordability, and socioeconomic conditions. In the realm of crime analysis, Oliveira et al. [10] highlight the concentration and scaling of crime in urban areas, underscoring the need to integrate crime data into housing recommender tools. Complementary to crime-safety concerns, other studies incorporate environmental quality: for instance, Gascon et al. [12] and Reid et al. [4] emphasize green spaces' role in reducing mortality and improving self-rated health while pointing out the complexities of measuring "greenness." Hsiang et al. [18] broadened the environmental scope by estimating the economic damage from climate change, reinforcing that climate factors and extreme weather patterns also heavily affect housing choices. Furthermore, Williams and Kay [9] and Wrenn [19] demonstrate how disasters and weather volatility influences household relocation decisions, reflecting another dimension of risk in neighborhood selection.

Ennis et al. [7] show that such decisions hinge on a delicate balance between personal preference and broader economic drivers as individuals consider living and workplace locations. Okulicz-Kozaryn [14] delves into city livability rankings versus resident satisfaction, suggesting that subjective perceptions often diverge from objective metrics. Similarly, Schirmer et al. [15] review how residential location choice models incorporate neighborhood services, commute considerations, and socio-demographic factors. Gurran et al. [13] investigate how technology platforms are reshaping informal housing markets. In contrast, Chen and Haiyunlin [5] reveal the distances people are willing to travel in their housing search, influenced by household needs and the availability of information. Rae [17] illustrates how digital submarkets have emerged, changing the geography of real estate transactions.

From a methodological standpoint, several works detail multi-criteria decision-making (MCDM) or recommenders. Al-Shalabi et al. [2] present a GIS-based approach to evaluating housing site suitability, integrating spatial data with user preferences. Arif et al. [3] introduce a multi-criteria recommender system that leverages known ratings to tailor housing selection. Khodakarami et al. [8] similarly combine MCDA models with spatial analytics at the neighborhood scale to assess urban sustainability—supporting the notion that location decisions must weigh diverse and sometimes conflicting criteria. Daly et al. [6] highlight the importance of journey-aware housing recommenders.

In contrast, Truong et al. [16] demonstrate that machine learning methods significantly improve the accuracy of housing price predictions. Goldstein and Hastings [1] focus on the interplay among schools, income inequality, and housing consumption, suggesting that incorporating socioeconomic and educational data can refine housing recommender systems.

Together, these studies address Heilmeier's key questions regarding the state of current technology and the need for more holistic, user-centric solutions. Specifically, the existing literature reveals that housing and location choice is influenced by an extensive array of variables—crime, environment, weather, accessibility, socio-economics, and market trends—yet few platforms integrate them into a single, interactive decision tool. Our project builds upon this work by blending multi-criteria algorithms with comprehensive datasets to create a unified decision platform. This approach consolidates data from disparate sources and empowers users to specify their priorities, addressing gaps noted in prior research regarding the customized weighting of diverse factors. Ultimately, as these references show, a robust, multi-faceted framework is essential for capturing the real-world complexity of location decisions and mitigating the risks identified by prior work.

4 PROPOSED METHOD

We are actively cleaning up multiple large datasets to compile into one database for access by our visualizations. In the meanwhile, we have used a smaller scale of the dataset to develop "sample" visualizations that were used to verify the concept of our heatmap. Moreover, currently for our algorithm we are using a simple quartile based ranking method on the possible ranges of our focus parameters. In this case, values on the lower end of the scale will correspond to a lesser preference for that feature and vice versa. These ranges are also feature specific, since a larger magnitude for one parameter (eg. temperature) maybe a positive in one case but for another parameter (eg. crime rate) this would correspond to a negative outlook. In addition, for the user interface we have developed a front end where a user can select from any of the features from Affordability, Walkability, Weather, Employment and Well-being. Then based on their preference they can manipulate a slider and the algorithm will provide them with a couple of options that meet the specified criteria.

5 INNOVATIONS

Furthermore, the following are a list of unique innovations that our solution exhibits:

5.1 Inventive Visualization

We are exploring how our heat map visualization can empower users to visually compare multiple locations simultaneously. Our team has made substantial progress in developing an interactive visual interface where users can adjust preference settings and immediately see how different regions compare. We're investigating how color intensity variations can effectively communicate complex data relationships, making it easier for users to identify optimal locations based on their specific requirements without requiring advanced data analysis skills.

5.2 Unified Information Access

Our approach examines the potential of harmonizing diverse data categories into a single analytical framework. We've made significant advances in addressing the challenge of normalizing these disparate data types to enable meaningful comparisons. Our implementation explores how users can benefit from accessing all relevant decision factors in one cohesive hub rather than gathering and analyzing information from multiple disconnected sources.

5.3 Personalized Customization

We're investigating how a preference weighting system can deliver truly personalized results. Our implementation explores the balance between simplicity and flexibility, allowing users to express their unique priorities across multiple categories. We've made substantial progress in developing an algorithm that recalculates compatibility scores based on these weighted preferences, studying how different weighting approaches affect the relevance of recommendations for diverse user needs.

6 EVALUATION

In order to test the effectiveness of our solution in terms of the specific problems that we aim to solve for our end users, below are experiments that we look to carry out. These aim to tackle and answer questions related to model bias, scalability of the solution, accuracy, and user satisfaction.

(1) **Bias Analysis**: We want to ensure that the recommendations from our solution are fair to all kinds of input profiles and do not unintentionally favor or disadvantage specific groups or locations. Basically, checking to see if the algorithm has any tendencies to skew results for specific features. Various inputs will be analyzed for this check, putting them through statistical tests and monitoring their data distributions.
(2) **Location Accuracy**: In addition to being quick, a solution must also be accurate. To determine the accuracy of our location suggestions, we will cross-reference outputs with external resources such as real estate pricing charts and top city rankings etc.
(3) **Usability Audit**: User satisfaction is an integral part of any solution. To determine potential improvements and remedy common errors, We will conduct user surveys and incorporate feedback as necessary.
(4) **Extensibility Investigation**: Based on demand as well as necessity in the future of the project, and as we potentially incorporate larger amounts of data, we also have to be able to guarantee that the product continues to be computationally fast. To do this, multiple additional years of data will be incorporated into the original dataset and tests related to response time, computational efficiency will be carried out to certify that the program is scalable.

7 CONCLUSIONS AND DISCUSSION

Figure 1 below illustrates the preliminary activity chart that was proposed through our initial proposal document, approximately 4 weeks ago.

Phase	Time to Complete	Tools	Responsible Team Member(s)
Data Preparation	3 weeks	Python, Open Refine, Data Wrangler	ALL
Front-End Web Application Development	2 weeks	JavaScript, D3.js, React	ALL
Algorithm Development	2 weeks	Python	ALL
Testing	1 week	Manual Testing	ALL

Figure 1: Previous Activity Chart

Figure 2 below depicts the current revised activity chart in line with the submission of the progress report. As can be seen in the chart, the progress of the project is in line with what was predicted. Currently, the data preparation piece, which was a significant part of the project has been completed (albeit with a couple of challenges discussed further below). Moreover, the front-end development and algorithm development/refinement is currently ongoing. For the algorithm development, we do have a very linear quartile based ranking in place, but still need to further look into more complex machine learning algorithms that can provide better results.

Phase	Time to Complete	Tools	Status of Phase	Responsible Team Member(s)
Data Preparation	3 weeks	Python, Open Refine, Data Wrangler	Completed	ALL
Front-End Web Application Development	2 weeks	JavaScript, D3.js, React	In Progress	ALL
Algorithm Development	2 weeks	Python	In Progress	ALL
Testing	1 week	Manual Testing	Not Started	ALL

Figure 2: Revised Activity Chart

7.1 Challenges and Looking Ahead

One notable challenge we have encountered involves our datasets, which turned out to be considerably larger and more fragmented than anticipated. With data spread across multiple files, compilation has taken significantly longer than we initially planned for, creating delays in our integration timeline. We have addressed this obstacle by focusing on a more manageable subset of data covering key geographic areas while developing a more efficient processing approach. Though we may need to scale back our geographic coverage for the final presentation, we are still on track to deliver effective visualizations that showcase our platform's core functionality. Additionally, as per our timeline we are still on track to complete the deliverables that are in progress currently as well as start the testing component when the time arises. Lastly, throughout the project all team members have been active in discussions and have contributed a similar amount of effort to ensure that the best possible solution can be delivered.

REFERENCES

[1] Orestes P. Hastings Adam Goldstein. 2018. Buying In: Positional Competition, Schools, Income Inequality, Housing Consumption. *Sociological Science* (2018).

[2] Mohamed A. Al-Shalabi, Shattri Bin Mansor, Nordin Bin Ahmed, and Rashid Shiriff. 2006. GIS Based Multi-Criteria Approaches to Housing Site Suitability Assessment. In *Proceedings of the XXIII FIG Congress.* https://www.fig.net/resources/Proceedings//fig_proceedings/fig2006/papers/ts72/ts72_05_alshalabi_etal%20_0702.pdf

[3] Yunifa Miftachul Arif, Muhammad Farid Muhtarom, and Hani Nurhayati. 2023. Performance of Known Ratings-Based Multi-Criteria Recommender System for Housing Selection. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE).* 761–766. https://doi.org/10.1109/ICCoSITE57641.2023.10127720

[4] Jiayue Li Jessie L. Shmool Jane E. Clougherty Colleen E. Reid, Laura D. Kubzansky. 2018. It's not easy assessing greenness: A comparison of NDVI datasets and neighborhood types and their associations with self-rated health in New York City. *ScienceDirect* (2018).

[5] Haiyunlin Cynthia Chen. 2012. How Far Do People Search for Housing? Analyzing the Roles of Housing Supply, Intra-household Dynamics, and the Use of Information Channels. *Taylor Francis Online* (2012).

[6] Akihiro Kishimoto Radu Marinescu Elizabeth M. Daly, Adi Botea. 2014. Multi-Criteria Journey Aware Housing Recommender System. *ACM Digital Library* (2014).

[7] Alberto Porto Huberto M.Ennis, Santiago M. Pinto. 2006. Choosing a place to live and a Workplace. *Economica, La Plata* (2006).

[8] Loghman Khodakarami, Saeid Pourmanafi, Zahra Mokhtari, Ali Reza Soffianian, and Ali Lotfi. 2023. Urban sustainability assessment at the neighborhood scale: Integrating spatial modellings and multi-criteria decision making approaches. *Sustainable Cities and Society* 97 (2023), 104725. https://doi.org/10.1016/j.scs.2023.104725

[9] Daid Kay Lindy Williams. 2024. Might I have to move due to climate change? The role of exposure to risk and political partisanship in anticipation of future relocation. *Climatic Change* 177, 142 (2024).

[10] Ronaldo Menezes Marcos Oliveira, Carmelo Bastos-Filho. 2017. The scaling of crime concentration in cities. *Plos One* (2017).

[11] Alison K. Fields Matthew C. Marlay. 2004. Seasonality of Moves and the Duration and Tenure of Residence: 2004. *United States Census Bureau* (2004).

[12] David Martínez Payam Dadvand David Rojas-Rueda Antoni Plasència Mark J. Nieuwenhuijsen Mireia Gascon, Margarita Triguero-Mas. 2016. Residential green spaces and mortality: A systematic review. *ScienceDirect* (2016).

[13] Pranita Shrestha Nicole Gurran, Zahra Nasreen. 2024. Platform-Enabled Informality? *Taylor Francis Online* (2024).

[14] Adam Okulicz-Kozaryn. 2011. City Life: Rankings (Livability) Versus Perceptions (Satisfaction). *Social Indicators Research - An International and Interdisciplinary Journal for Quality-of-Life Measurement* (2011).

[15] Kay W. Axhausen Patrick M. Schirmer, Michael A.B. van Eggermond. 2014. The Role Of Location In Residential Location Choice Models A Review Of Literature. *Journal of Transport and Land Use* (2014).

[16] Hy Dang Bo Mei Quang Truong, Minh Nguyen. 2020. Housing Price Prediction via Improved Machine Learning Techniques. *ScienceDirect* (2020).

[17] Alasdair Rae. 2014. Online Housing Search and the Geography of Submarkets. *Taylor Francis Online* (2014).

[18] Amir Jina James Rising Michael Delgado Shashank Mohan D. J. Rasmussen Robert Muir-Wood Paul Wilson Michael Oppenheimer Kate Larsen Trevor Houser Solomon Hsiang, Robert Kopp. 2017. Estimating Economic Damage from Climate Change in the United States. *ScienceDirect* (2017).

[19] Douglas H. Wrenn. 2024. The Effect of Natural Disasters and Extreme Weather on Household Location Choice and Economic Welfare. *Journal of the Association of Environmental and Resource Economists* (2024).

[20] Jichang Zhao Yaping Zhao and Edmund Y. Lam. 2024. House Price Prediction - A Multi-Source Data Fusion Perspective. *Big Data Mining and Analytics* 7, 3 (2024), 603–620.