Location Matchmaking Platform - Team 019

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

1 INTRODUCTION

Americans move approximately 11 times during their lifetime [16], 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 general care activities

This integration transforms location selection from overwhelming research into a streamlined experience aligned with users' priorities. Zhao et al. [28] 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. [15] 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. [17] and Reid et al. [6] emphasize green spaces' role in reducing mortality and improving self-rated health while pointing out the complexities of measuring "greenness." Hsiang et al. [26] 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 [14] and Wrenn [27] demonstrate how disasters and weather volatility influences household relocation decisions, reflecting another dimension of risk in neighborhood selection.

Ennis et al. [11] 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 [20] delves into city livability rankings versus resident satisfaction, suggesting that subjective perceptions often diverge from objective metrics. Similarly, Schirmer et al. [21] review how residential location choice models incorporate neighborhood services, commute considerations, and socio-demographic factors. Gurran et al. [18] investigate how technology platforms are reshaping informal housing markets. In contrast, Chen and Haiyunlin [8] reveal the distances people are willing to travel in their housing search, influenced by household needs and the availability of information. Rae [23] 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. [4] present a GIS-based approach to evaluating housing site suitability, integrating spatial data with user preferences. Arif et al. [5] introduce a multi-criteria recommender system that leverages known ratings to tailor housing selection. Khodakarami et al. [13] 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. [9] highlight the importance of journey-aware housing recommenders. In contrast, Truong et al. [22] 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

4.1 Data

To build a platform that not only gathers relevant data but also allows the user to select features based on their preference and specify the importance levels for those features, we collected data sets for various domains such as crime, weather, employment, housing, and cost of living - and mostly cover data from the years 2023 to 2025 to ensure relevance and comparability. The team had spent a significant amount of time on deciding (1) which features would be most valuable for the user and (2) whether to use "neighborhood", "city" or "county" as the primary key. After extensive research, the team found that city-level and neighborhood-level data was sparse and inconsistent across different sources hence merging those datasets would reduce the sample size. The primary key was then chosen to be "county".

Datasets for this project came from public sources including government databases and research portals (1GB total). Pre-processing cleaned and unified data, standardizing county names and using FIPS codes as keys. The unemployment dataset required individual state downloads with separate cleaning. Temporal datasets were aggregated yearly. The final dataset contains ˜3,000 county-level rows (1.5MB) with rich features from multiple sources, enabling comparative and predictive analysis across U.S. counties. Table 1 lists the datasets used in this study along with the chosen variables.

Dataset	Field Name
National Risk Index [2]	Population, Risk Score
County Health Release [25]	Access to Exercise Opportunities, Food Environment Index, Primary Care Physicians, Air Pollution: Particulate Matter, Broadband Access, Life Expectancy, Traffic Volume, Home-ownership, Access to Parks
National Walkability Index [3]	NatWalkInd
Climate - County Mapping [10]	Average Temperature, Maximum Temperature, Minimum Temperature, Precipitation
County-level Data Sets - Unemployment [19]	Unemployment Rate
Economic Policy Institute Family Budget Calculator Data	Family Size, Monthly Food, Monthly Childcare, Monthly Healthcare, Monthly Transportation, Monthly Taxes, Monthly Other Necessities
United States crime rates by county [12]	Crime rate per 100000
Housing Market Data [24]	Median sale price, Median list price, Median ppsf, Homes sold, New listings, Inventory, Months of supply, Median dom months
Fipscode [7]	County Code, State Code, County, State

Table 1: Datasets and Their Corresponding Field Names

4.2 Algorithm

Our approach integrates a multi-criteria ranking algorithm with an interactive visualization platform to address the problem of location selection. Users can filter results based on state and specify their household size. For all other features, users will provide their preferred values along with associated importance weights, which shows the relative significance of each feature according to their individual preferences. All potential user scenarios were carefully considered and several fallback mechanisms have been incorporated to ensure the algorithm is robust in cases where the user does not provide the expected input or deviates from standard usage patterns. The ranking algorithm

- Creates the final dataset with cost-of-living data based on user-selected “family size” feature (Defaults to 2 adults if none selected)
- Filters the final dataset by user-selected state(s) to create a user-specific dataset (Defaults to all states if none selected)
- Fills in the missing values using median, mean or mode based on properties of the feature
- Converts categorical variables into a numerical format
- Applies min-max normalization to the selected feature values across all counties
- Normalizes the corresponding user-specified values using the same scaling parameters
- Normalizes the weights associated with user-selected features to ensure that their relative importance is proportionally represented in the final evaluation.

The final ranking score S_i for each county is calculated as a weighted sum of feature scores. However the score differ based on whether the user has selected features.

If the user HAS selected features, the ranking score is computed using w_j (feature weight) and $s_i j$ (normalized feature value)

$$S_i = \sum_j w_j(1 - |s_{ij} - u_j|) \tag{1}$$

where

- S_i is the ranking score of county i considering the users preference
- w_j is the normalized weight of feature j indicating the user’s importance
- $s_i j$ is the feature value j for county i
- u_j is the normalized user values for feature j

If the user HAS NOT selected any features and would like to see the top 10 counties in a particular state or the entire continent, the ranking score is computed using w_j (feature weight), $s_i j$ (normalized feature value) and d_j (directionality).

$$S_i = \sum_j w_j(d_j * s_{ij}) \tag{2}$$

where

- S_i is the ranking score of county i considering the users preference
- w_j is equal weights for all features
- $s_i j$ is the feature value j for county i
- d_j is the directionality of feature j (1 for features where lower values are more desirable, 1 for features where higher values are more desired)

This algorithm underpins an interactive visualization tool. The users can adjust feature preferences dynamically via sliders which calculates the rankings and displays the results through heat maps and tables. This enables users to explore and refine their choices in real time. This approach is user-agnostic and allows individuals to precisely express their preferences, leading to highly customized and relevant outputs. This gives people more control, clarity, and confidence when making decisions, turning a complex process into something more personal.

4.3 Innovations

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

4.3.1 Inventive Visualization. We are exploring how our heat map visualization can empower users to visually compare multiple locations simultaneously. Our team has developed an interactive visual interface where users can adjust preference settings and immediately compare different regions. The interface outputs the top 10 counties that match the user’s preferences. Each county is highlighted and colored from green to red, with green being the 1st ranked county and red 10th ranked county. This makes it easier for users to identify optimal locations based on their specific requirements without requiring advanced data analysis skills.

4.3.2 Unified Information Access. Our approach examines the potential of harmonizing diverse data categories —like cost, safety, and environment — into a single analytical framework. Unlike traditional tools that focus on just one factor or give fixed rankings, our platform lets users set their own priorities, update results

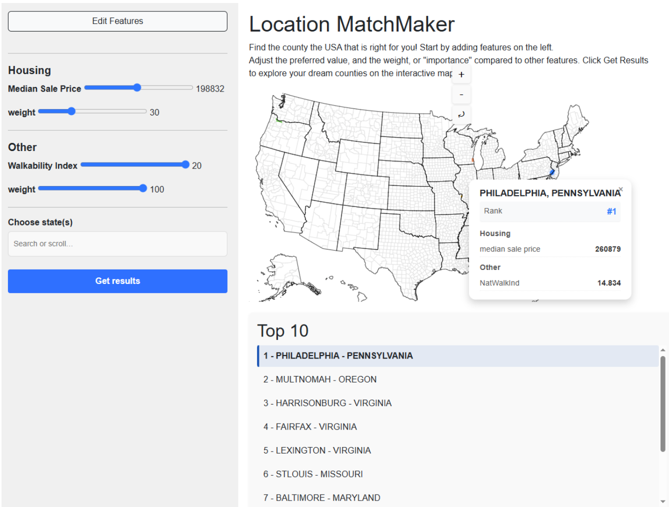

Figure 1: Image of Visualization

in real time, and shows results. We've also 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.

4.3.3 Personalized Customization. Our approach is a preference weighted ranking system that truly delivers personalized results. It strikes a balance between simplicity and flexibility, allowing users to express their specific priorities across multiple categories.

5 EVALUATION

In order to assess the performance, reliability, and fairness of our ranking algorithm, we conducted four targeted experiments designed to evaluate different aspects of the system. Each experiment focused on a specific question related to model behavior and practical outcomes. The experiments addressed parameter sensitivity, fairness across demographic profiles, and accuracy of rankings in relation to real-world expectations.

The following key questions are aimed to be answered by our experiments:

- Does household size significantly affect county rankings?
- How do different user personas influence the output rankings based on personalized preferences?
- Which parameters are most sensitive to weight changes, and how stable are the resulting rankings?
- To what extent do our top recommendations align with real-world rankings?

5.1 Experiment 1: Impact of Household Size on Rankings

This experiment aimed to test whether changes in household size significantly affect rankings. Three scenarios were evaluated: a five-member family, a three-member family, and an individual. In all cases, the features and weights assigned were kept constant. The results were visualized using a clustered bar graph comparing the top 10 county rankings across the three household sizes seen in Figure 1.

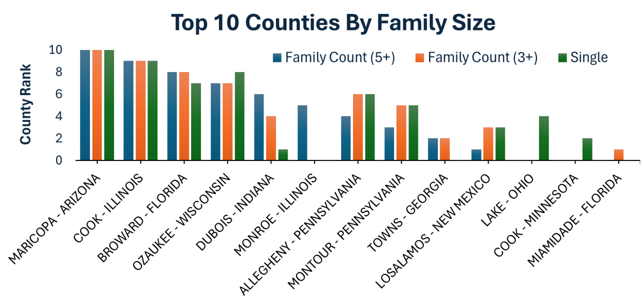

Figure 2: County Rankings Across Family Size

Aside from a few differences in the final two rankings, the outputs revealed minimal variation in rankings across the scenarios, suggesting household size alone does not substantially alter outcomes unless cost-based features are heavily weighted. This output was hypothesized and expected before the experiment was started, and also indicates a level of result stability that is required in ensuring fair treatment across household configurations. Overall, this test demonstrated the consistency of the algorithm, highlighting its robustness to small demographic changes.

5.2 Experiment 2: Persona-Based Customization and Output Variation

The second experiment investigated how different user preferences influence recommendations. Three distinct user personas were constructed: a family unit, a young professional, and a retiree. Each persona emphasized different sets of features through custom weight configurations—for instance, families prioritized affordability, access to parks, and crime rates; young professionals emphasized employment, exercise and connectivity; retirees focused on life expectancy, health related costs, environmental quality. These being such unique profiles with their own priorities, we hypothesized that the results would be dynamic in each case. Similar to as we expected, running the algorithm for each persona yielded three distinct sets of top 10 county rankings. As can be seen in Figure 2, these results were visualized using a heatmap that illustrated the relative rankings across counties for each persona.

County-State	Family	Young Professional	Retiree
SEMINOLE - GEORGIA	1		
ISSAQUENA - MISSISSIPPI		1	10
GLADES - FLORIDA			1
PANOLA - MISSISSIPPI	2		
SHARKEY - MISSISSIPPI		2	
SIERRA - CALIFORNIA			2
SUNFLOWER - MISSISSIPPI	3		
OGLALALAKOTA - SOUTH DAKOTA		3	
SHERMAN - OREGON			3
CLAIBORNE - MISSISSIPPI	4		
YUKONKOYUKUK - ALASKA		4	
NORTHSLOPE - ALASKA		5	4
LINCOLN - GEORGIA	5		
DODDRIDGE - WEST VIRGINIA			5
IRWIN - GEORGIA	6		
HANCOCK - GEORGIA		6	
JASPER - ILLINOIS			6
WASHINGTON - MISSISSIPPI	7		
APACHE - ARIZONA		7	
SCOTLAND - MISSOURI			7
HEARD - GEORGIA	8		
HUMPHREYS - MISSISSIPPI		8	
BRISTOLBAY - ALASKA			8
CLINCH - GEORGIA	9		
FOARD - TEXAS		9	
CUMBERLAND - ILLINOIS			9
WILCOX - ALABAMA		10	
FAYETTE - ILLINOIS			10

Figure 3: Persona County Ranking Heat Map

The results demonstrated that the algorithm dynamically responds to user-defined preferences as well as the system's flexibility in adapting to diverse user needs, with different regions rising to the top depending on the feature weight profile.

5.3 Experiment 3: Sensitivity Analysis of Feature Weights

This experiment focused on assessing how sensitive the rankings are to changes in the weights assigned to individual features. Using the Kendall rank correlation coefficient from the scipy.stats.kendalltau library, we conducted a series of tests in which the weights for four selected features—Crime Rate, Unemployment Rate (UR), Monthly Total Expenses (MTE), and Natural Disaster Risk were incrementally modified to evaluate how rankings responded. The constant features were kept at a magnitude of 5, while the dynamic feature was constantly changed by a magnitude of 10 from 0-100.

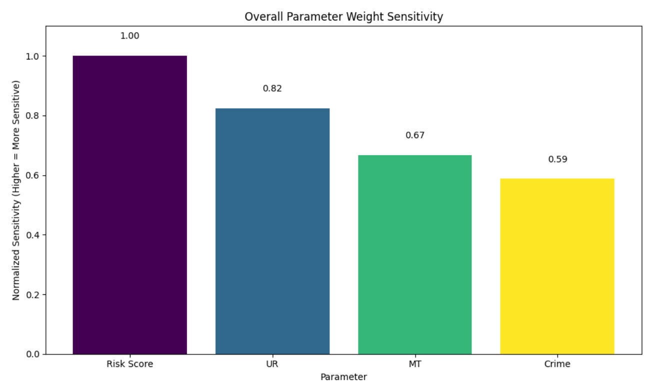

Figure 4: Parameter Sensitivity Analysis

As can be seen in Figure 3, Natural Disaster Risk Score feature exhibited the highest sensitivity, followed by UR, MTE, and finally Crime Rate. This means that small changes in the weight of Risk Score produced the most dramatic shifts in county rankings. Crime rate, in contrast, showed high stability with rankings barely shifting even with a small or large weight. Moreover, to further explore how feature influence changes with weight scaling, a parameter-weight sensitivity heatmap was generated (Figure 4).

The heatmap displayed features along the Y-axis and incremental weight levels along the X-axis, with color intensity corresponding to the Kendall tau distance from the baseline rankings. Darker shades in the heatmap signified a greater divergence from the original ranking, while lighter shades indicated minimal change. Notably, the Risk Score row darkened quickly even at low weight levels, confirming its high sensitivity. In contrast, the

Crime Rate row remained light across all weight levels, highlighting its stability. These results provide an understanding of how specific features influence outcomes and offer insight into which parameters may require more careful tuning. The observed sensitivity patterns can assist users in making more informed adjustments by communicating the potential impact of certain features on their results.

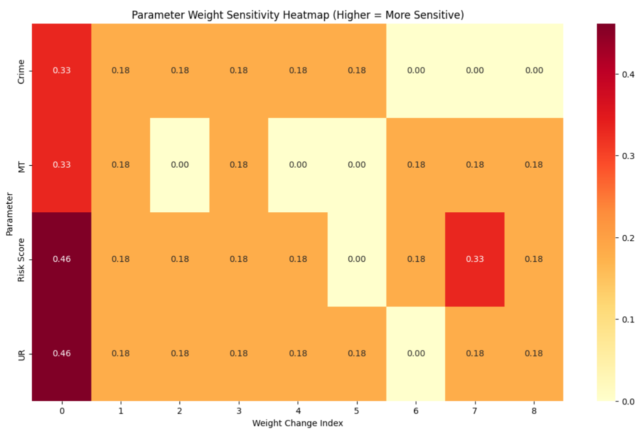

Figure 5: Parameter Sensitivity Heat Map

5.4 Experiment 4: Partial Real-world Ranking Comparison

The final experiment explored how well the algorithm's recommendations align with established online sources. Top 10 county rankings generated by the algorithm were compared to publicly available sources, i.e. the U.S. News Best States to Live In Ranking. Given that most public rankings operate at the state level rather than at the county level, exact one-to-one comparisons were not feasible. However, similarities in geographic distribution were observed. States such as Idaho, Washington, and Florida were frequently represented in both the algorithm's top results and external rankings. Although the alignment was not precise, the overlap in favorable states supports the credibility of the system's scoring approach. This experiment helped show that the algorithm follows general livability trends, even without exact county-level comparisons. It suggests that the model gives reasonable results that align with public opinion and trusted sources.

6 CONCLUSIONS AND DISCUSSION

6.1 Team and Impact

Location MatchMaker is the official name for this application, and it has proven to serve its purpose. Our original intention was to assist out of state home buyers, however we also believe this application could serve useful to investors, and researchers. There's a lot of value when a user finds an unexpected area fit their criteria. One example to note is walkability. When initially testing this feature, we expected, and did, see cities like New York City, Boston, San Francisco, etc. However, when housing price was factored (preferring lower housing price), the story changed. Philadelphia, Baltimore, and St. Louis rose to the top, revealing them as high value cities that are walkable, but also affordable. Because the the project is a web application, it can be easily hosted online for anyone to use, increasing its surface area of as a product. Consisting of a mix of software engineers and data scientists, the team had a dynamic set of technical skills to choose from. This gave the team the freedom and confidence to create a Python-based web application, and everyone contributed equal effort in their own way throughout the project.

6.2 Data Challenges

Initially, the biggest challenge was selecting and unifying the appropriate datasets. Deciding on neighborhood, county, or city-level data was an important discussion. We eventually chose county to balance granularity and practicality. It was still a struggle to find such detailed datasets that represent all 50 US states. Data cleaning took a significant portion of time spent on the project.

6.3 Looking Ahead

In future hypothetical iterations of this project, we would dedicate more effort to refine the user interface of the visualization. For example: Range sliders were used to adjust feature values and weights, but further experiments with different UI elements that reduce noise on the screen and perhaps be more fun to use. Adding additional features for advanced users could include more detailed definitions and explanations of data features, and the ability to save and share search criteria.

REFERENCES

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

[2] Federal Emergency Management Agency. 2025. National Risk Index Dataset. (2025). https://hazards.fema.gov/nri/data-resources

[3] U.S. Environmental Protection Agency. 2021. Smart Location Database. (2021). https://www.epa.gov/smartgrowth/smart-location-mapping#SLD

[4] 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

[5] 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

[6] 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).

[7] Federal Communications Commission. 2025. Federal Information Processing System (FIPS) Codes for States and Counties. (2025). https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt

[8] 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).

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

[10] National Centers for Environmental Information. 2024. Climate - County Mapping. (2024). https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping

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

[12] Kaggle. 2019. United States crime rates by county. (2019). https://www.kaggle.com/datasets/mikejohnsonjr/united-states-crime-rates-by-county/data

[13] 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

[14] 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).

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

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

[17] 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).

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

[19] U.S. Department of Agriculture. 2023. County-level Data Sets - Unemployment. (2023). https://data.ers.usda.gov/reports.aspx?ID=4038

[20] 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).

[21] 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).

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

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

[24] REDFIN. 2024. Housing Market Data. (2024). https://www.redfin.com/news/data-center/

[25] County Health Rankings Roadmaps. 2024. 2024 CHR CSV Analytic Data. (2024). https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation

[26] 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).

[27] 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).

[28] 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.

A APPENDIX

Data Field	Definition
Population	Total population of the county
Risk Score	County's annual estimated loss from natural hazards, adjusted for social vulnerability and community resilience
Access to Exercise Opportunities	Percentage of population with adequate access to locations for physical activity
Food Environment Index	Index of factors that contribute to a healthy food environment, from 0 (worst) to 10 (best)
Primary Care Physicians	Ratio of population to primary care physicians
Air Pollution: Particulate Matter	Average daily density of fine particulate matter in micrograms per cubic meter
Broadband Access	Percentage of households with broadband internet connection
Life Expectancy	Average number of years people are expected to live
Traffic Volume	Average traffic volume per meter of major roadways in the county
Home-ownership	Percentage of owner-occupied housing units
Access to Parks	Percentage of the population living within a half mile of a park
NatWalkInd	Walkability based on street connectivity, transit access, and land use diversity at block level
Average Temperature	Average temperature measured over a 12-month period
Maximum Temperature	Maximum temperature measured over a 12-month period
Minimum Temperature	Minimum temperature measured over a 12-month period
Precipitation	Precipitation(inches) measured over a 12-month period
Unemployment Rate	Percent of the labor force that is unemployed
Cost of Living	Monthly costs for food, childcare, healthcare, transportation, taxes, and other essentials
Crime rate	Number of reported crimes for every 100,000 people in a given area
Median sale price	Price for which a median house was sold
Median list price	Price for which a median house was listed
Median ppsf	Price of homes divided by square footage
Homes sold	Number of houses sold in a year
New listings	Number of new houses listed in a years
Inventory	Number of homes available for sale in a given county
Months of supply	Number of months needed to sell current inventory
Median dom months	Number of months a house is on market

Table 2: Data Definitions