LOCATION MATCHMAKING PLATFORM

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

INTRODUCTION

Deciding where to live is a significant and often stressful choice, yet it remains difficult to access all the necessary information in one place. Individuals are frequently left to search across multiple sources, compare details manually, and make important decisions without a clear or complete picture. This process is time-consuming and can result in frustration or disappointment after moving. We believe there should be a simpler, more reliable way to find a place to live—one that brings everything together and reflects what each person or family values most.

APPROACH

Our approach integrates a multi-criteria ranking algorithm with an interactive visualization platform to address the problem of location selection. The ranking algorithm consolidates diverse datasets and tailors results using user-defined preferences. It filters data by household size and location, fills missing values with appropriate statistical methods, encodes categorical variables numerically, and applies min-max normalization to both feature values and user inputs. When preferences are provided, it calculates the final ranking score S_i using normalized weights and proximity to user values (1); otherwise, it uses equal weights and feature directionality to rank counties (2).

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

Scenario 1

$$S_i = \sum_j w_j \left(d_j \times s_{ij}\right)$$

Scenario 2

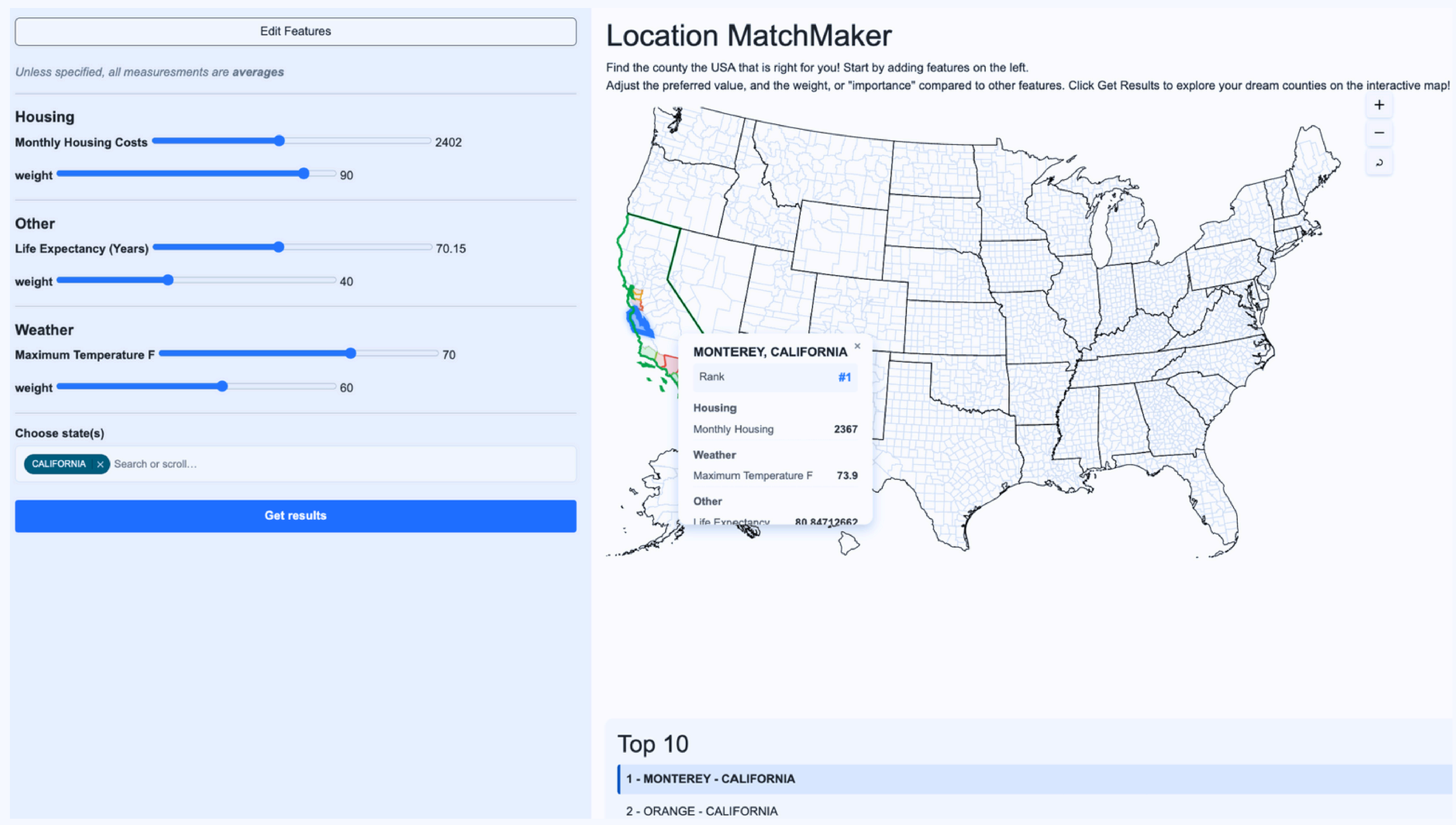
Variables such as w_j (feature weight), s_{ij} (normalized feature value), and d_j (directionality) play a key role in this computation, as illustrated in the two scenario-based algorithmic formulas above. This algorithm underpins an interactive visualization tool, where users can adjust feature preferences dynamically via sliders. The platform immediately recalculates rankings and visualizes results through heat maps and comparative charts, enabling users to explore and refine their choices in real time. Our approach stands out by combining multiple aspects of location quality—like cost, safety, and environment—into one customizable and interactive system. Unlike traditional tools that focus on just one factor or give fixed rankings, our platform lets users set their own priorities, updates results in real time, and clearly shows how scores are calculated. This gives people more control, clarity, and confidence when making decisions, turning a complex process into something more personal.

DATA

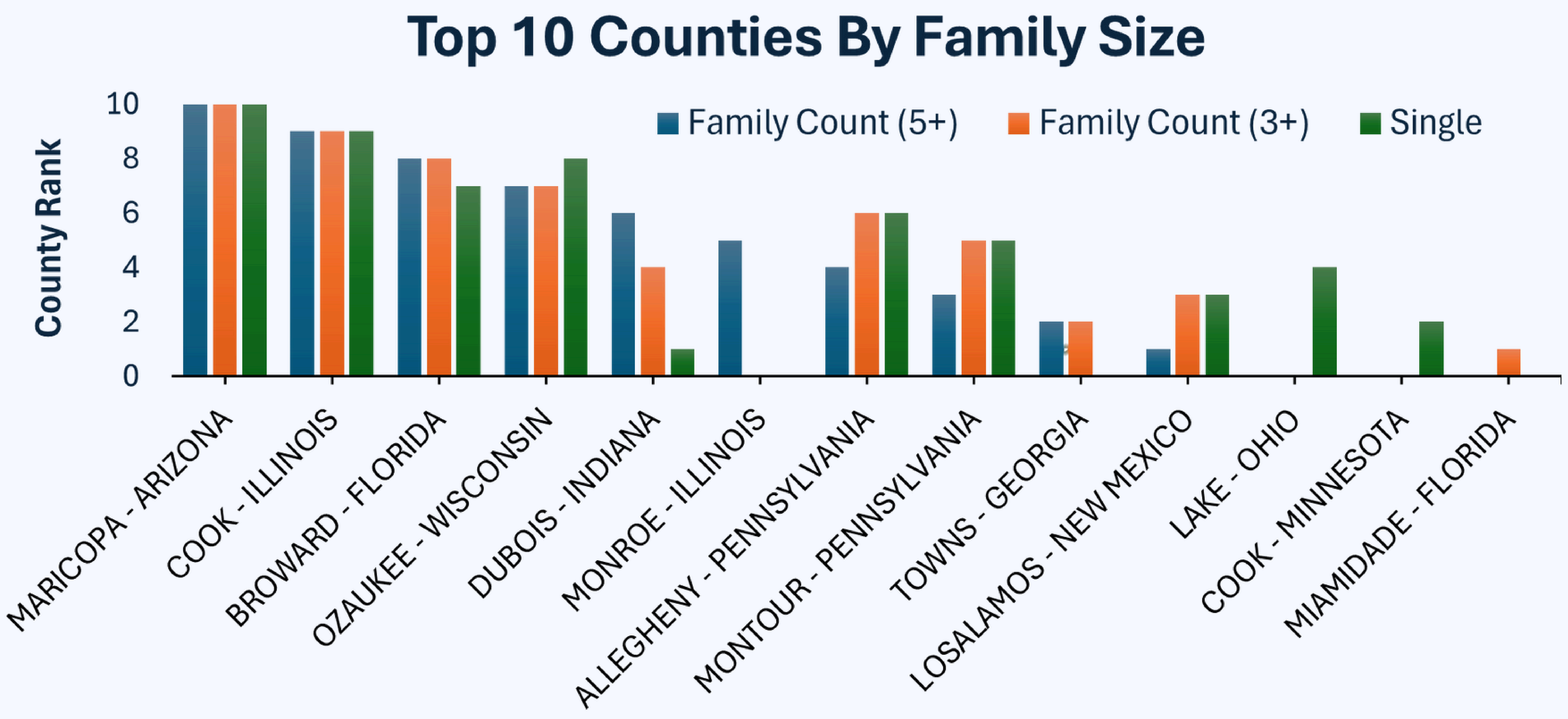
All datasets used in this project were downloaded directly from publicly available sources, including government databases, research portals, and reputable third-party data aggregators. The gathered datasets span various domains such as crime, weather, employment, housing, and cost of living—and mostly covers data from the years 2023 to 2025 to ensure relevance and comparability. Initially the datasets were a total of 1GB in size, after which extensive preprocessing was conducted to clean and unify the data, including resolving inconsistencies in county naming conventions and aligning all entries using FIPS codes as the common key. Additionally, the unemployment dataset required extra effort since only individual state files were available, which were then separately cleaned and combined together . Moreover, many of the original datasets that were temporal, were aggregated to a yearly level during preprocessing. The final merged dataset consists of approximately 3,000 county-level rows, with a total disk size of around 1.5MB. Despite the relatively modest row count, the dataset is rich in features, combining information from various sources enabling comparative and predictive analysis across U.S. counties.

EXPERIMENTS AND RESULTS

The experiments were conducted utilizing our interactive heat map, along with the relevant features in scope.



In the first evaluation, only the family size was varied in three different scenarios and the results were plotted together. As illustrated, the scoring was not significantly impacted since no other feature was included.



In the second case, different weights were assigned to a custom set of features for all three distinct 'groups'. Contrary to the first case, almost each county present was unique, demonstrating the effect and validity of manipulating weights in multiple features.

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