

# Finding Home: A Zip Code Based Recommender

By: Monica London

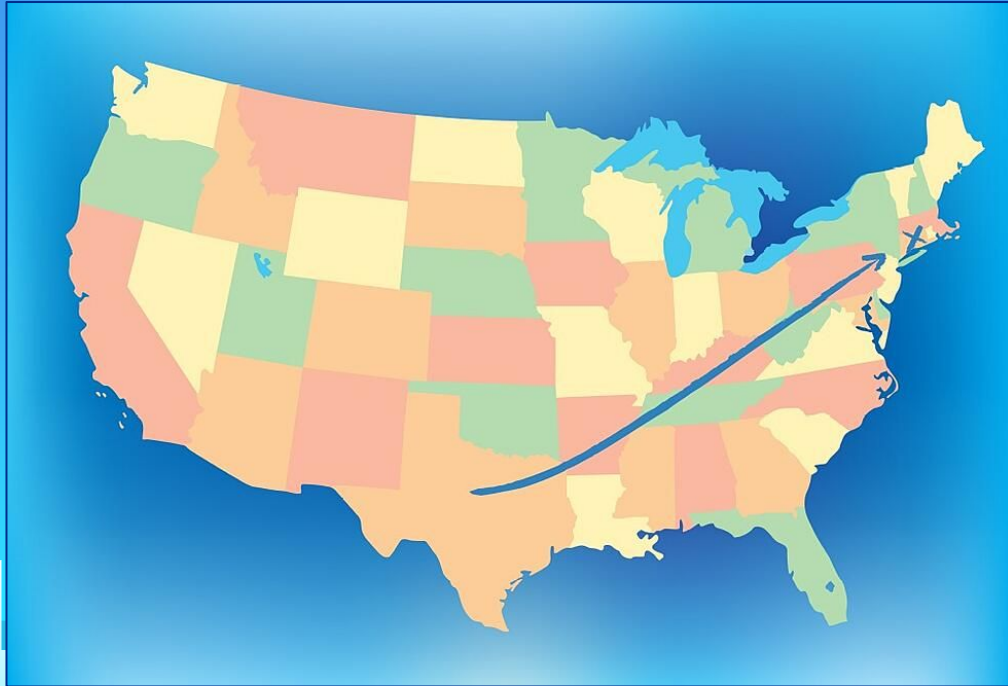


# Problem

- 1.5% of the US population moves to a new state each year (Source: moving.com)
- Many people have a general idea of where they are going, but how do they narrow it down?
- Some people have limited time with job transfers and must act quickly

# Value Proposition

**Find a location that fits your vibe!**



# Commercial Use Case

**Realtors who work with relocation buyers**



# Tech Stack

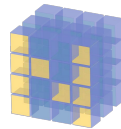
## Programming & Workflow Tools



Visual Studio Code



## Data Wrangling, EDA, & Building the Recommender System



NumPy

matplotlib



# Feature Categories

## Demographics

- Income
- Age breakdown
- Marital statistics
- Veterans
- Language spoken
- Employment
- Education
- Families with children

**Zip codes: 14,633**  
**Features: 92**



## Location & Housing Data

- Total population
- Commute type and time
- Density
- Home Price Index
- Home type
- Age of home
- Years in current home



# Approach

1. Develop a similarity matrix using cosine similarity with all selected features
2. Judged results based on location knowledge
3. Allow filtering between state or city/state
4. Determine how removing various features/feature categories impacts the recommender

# Dallas, TX, 75205

## Dallas (University Park), TX, 75205 > MA

- |    |                        |
|----|------------------------|
| 1  | Wellesley Hills, 02481 |
| 2  | Wellesley, 02482       |
| 3  | Andover, 01810         |
| 4  | Dover, 02030           |
| 5  | Westborough, 01581     |
| 6  | Chestnut Hill, 02467   |
| 7  | North Andover, 01845   |
| 8  | Lexington, 02420       |
| 9  | Auburndale, 02466      |
| 10 | Needham, 02494         |





# Feature Comparisons

	<b>Dallas (University Park) TX 75205</b>	<b>Wellesley Hills, MA 02481</b>
<b>Has Graduate Degree</b>	38.1%	50.3%
<b>% Commuting Via Public Transportation</b>	0.8%	11.0%
<b>% Owner Occupied</b>	64.5%	84.0%
<b>Median Owner Household Income</b>	\$230,395	\$233,266

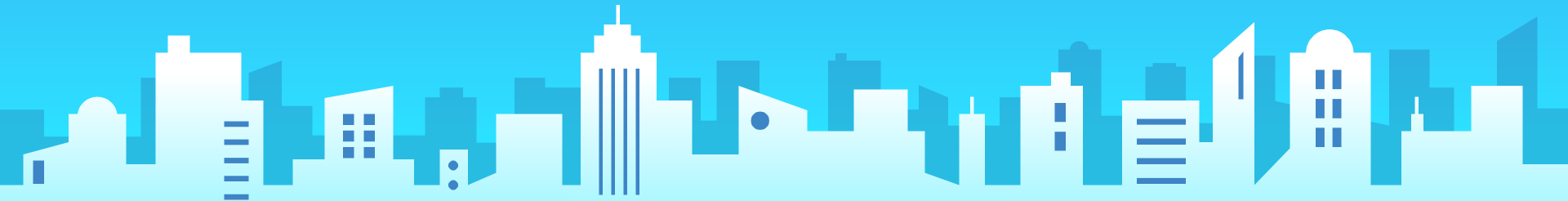
# Iterations

**Method** -> Removed a single feature/feature category

**Result** -> Total population, median income, density, or language had little to no effect on the recommendations. Removing employment breakdowns had the largest effect although it wasn't substantial

# Results

**It became clear that the features all work together and that none are significant in every instance**



# Future Steps

**School Data**



**Coffee & Restaurants**



**Flask App**





# Thank You!

**Monica London**

[www.linkedin.com/in/monicamercerlondon/](http://www.linkedin.com/in/monicamercerlondon/)

<http://github.com/MLondon1027>

