Retirement Recommender

NAVEEN RAM

Problem

- Retirees have so many options to choose from when it comes to where to retire.
  - There are so many cities in the US
    - The dataset used in this project contains more than 300 US cities.
  - Retirees want an easy and painless transition to retirement
- Moving to a new city is hard
  - Having amenities and attractions similar to your previous city is useful
  - Choosing a new place is a hassle

Question

- Can we make it easy for those people who are choosing a new place to live by allowing them to enter a city of their choice, and provide multiple suggestion of similar cities that might also be a good idea for retirement?
- Can we reduce the number of options they have?

Dataset

- US top cities by population data from Wikipedia 314 cities
- From this data, we used three data sources
  - Population data from Wikipedia
  - Geographical Latitude and Longitude data from geocoders
  - Amenity and attraction data from Foursquare API
- 517 features
  - 516 amenity and attraction categories

Data transformation

- Start with top 314 cities in US
- Add population data
- Add geographic data
- Create new data frame with venues for each city listed
- Convert the venue category into a numerical value
- Get the mean for each category for each city
- Make the location the index of the table

Data Analysis and Modeling

KMeans Clustering algorithm

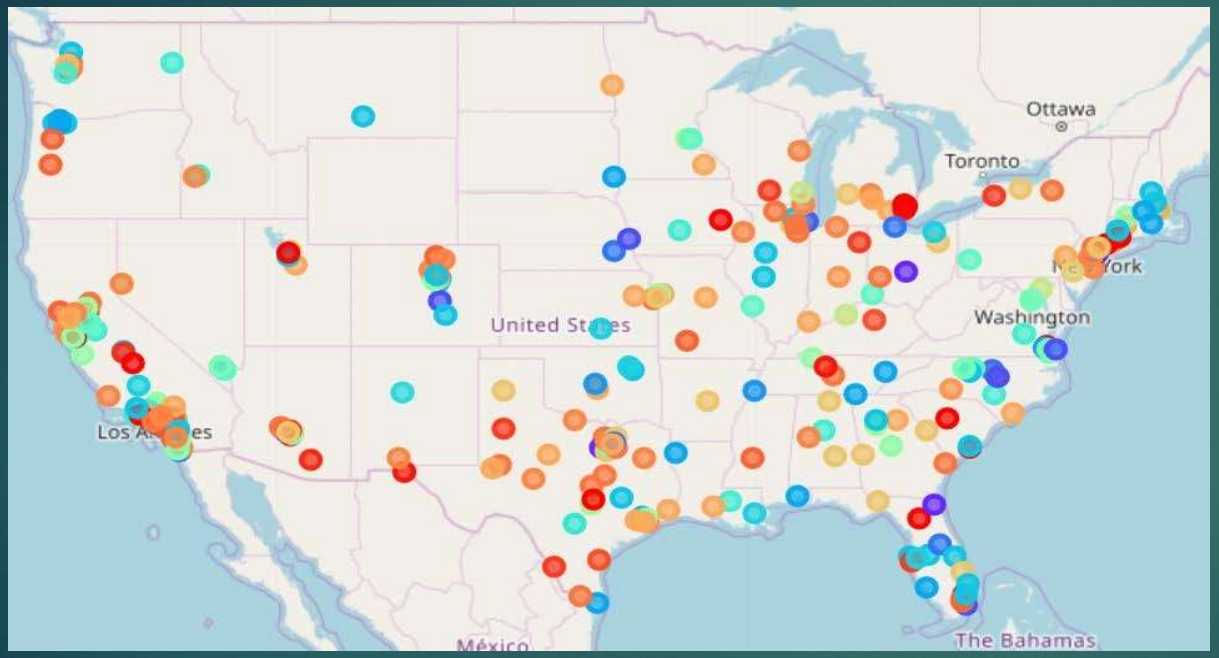
K = 40 clusters centers

N = 312 cities

Features = 517 (population data + Foursquare data)

Output clusters were filtered to only contain the groups that have 5+ cities

Results



N = 266 cities labeled

Filtered by cluster label size > 5 and available Foursquare data

Results (Cont)

Cluster Labels			
0	19	19	19
5	6	6	6
7	9	9	9
10	14	14	14
12	36	36	36
13	8	8	8
16	13	13	13
18	7	7	7
20	11	11	11
22	15	15	15
29	20	20	20
31	28	28	28
33	32	32	32
34	25	25	25
35	14	14	14
37	9	9	9

40 clusters total

Filtered where cluster size > 5

Providing enough recommendations

	Cluster Labels	Location
0	31	Abilene,TX
1	29	Akron,OH
2	14	Albuquerque,NM
3	22	Alexandria,VA
4	33	Allen,TX
5	31	Allentown,PA
6	29	Amarillo,TX
7	27	Anaheim,CA
8	7	Anchorage,AK
9	31	Ann Arbor,MI
10	12	Antioch,CA
11	13	Arlington,TX
12	31	Arvada,CO
13	31	Athens,GA
14	23	Atlanta,GA

Example cluster labels

Output

```
cities_final2.loc[cities_final2['Location'] == 'Miami,FL']
```

	Cluster Labels	Location	City Latitude	City Longitude
15064	5	Miami,FL	25.774266	-80.193659

```
cities_final2.loc[cities_final2['Cluster Labels'] ==5]
```

	Cluster Labels	Location	City Latitude	City Longitude
4894	5	Colorado Springs,CO	38.833958	-104.825349
13663	5	Long Beach,CA	33.769016	-118.191605
15064	5	Miami,FL	25.774266	-80.193659
17519	5	Omaha,NE	41.258746	-95.938376
19791	5	Raleigh,NC	35.780398	-78.639099
26167	5	Virginia Beach,VA	36.852984	-75.977418

Input: city in dataset
Ex: Miami, FL

Output: similar cities in the US with the same cluster label