January 2020

Customizing online real estate search engine results according to a personal profile

Introduction

- Looking for a home in a new city can be a challenging task.
- Most of us don't know the ins and outs of a city before we decide to move there, so choosing the right home will usually require an extensive research.
- Not everyone has the time or knowledge to perform such a comprehensible research.

Business Problem

- An online real estate search engines that recommends neighborhoods according to an online profile.
- The online profile will have information such as age, marital status, education, hobbies etc.
- Useful for people who don’t have the time or ability to research this themselves, and to developers of online real estate search engines that can use this approach to add an extra feature that will set them apart from the competition.

Data

- This report focuses on the city of Los Angeles, however, the methods shown here can be used for potentially any major city.
- Los Angeles, or L.A., is the most populous city in California.
- Home to more than 10 million people, Los Angeles County is comprised of nearly 300 communities that vary greatly from one another.
- For someone who is not familiar with the area choosing the right neighborhood can be a very daunting task.

Data Sources

- The Neighborhood Data for Social Change (NDSC) –

  This platform is a project of the USC Price Center for Social Innovation, a free publicly available online resource for civic actors to learn about their neighborhood.
- Foursquare API

Data Preparation

- Demographic data about L.A. neighborhoods was obtained from the NDSC portal (neighborhoods and their coordinates, age distribution, marital status, education, average household size, unemployment rate, crime rate and total number of population).
- 100 of the most popular venues within a 500 m radius of the center of each neighborhood were gathered using the foursquare API.

Methodology

- Stage I – cluster neighborhoods according to the demographic data.

  Find the cluster that fits the personal data of the user.

- Stage II – Cluster neighborhoods according to the types of venues they have.

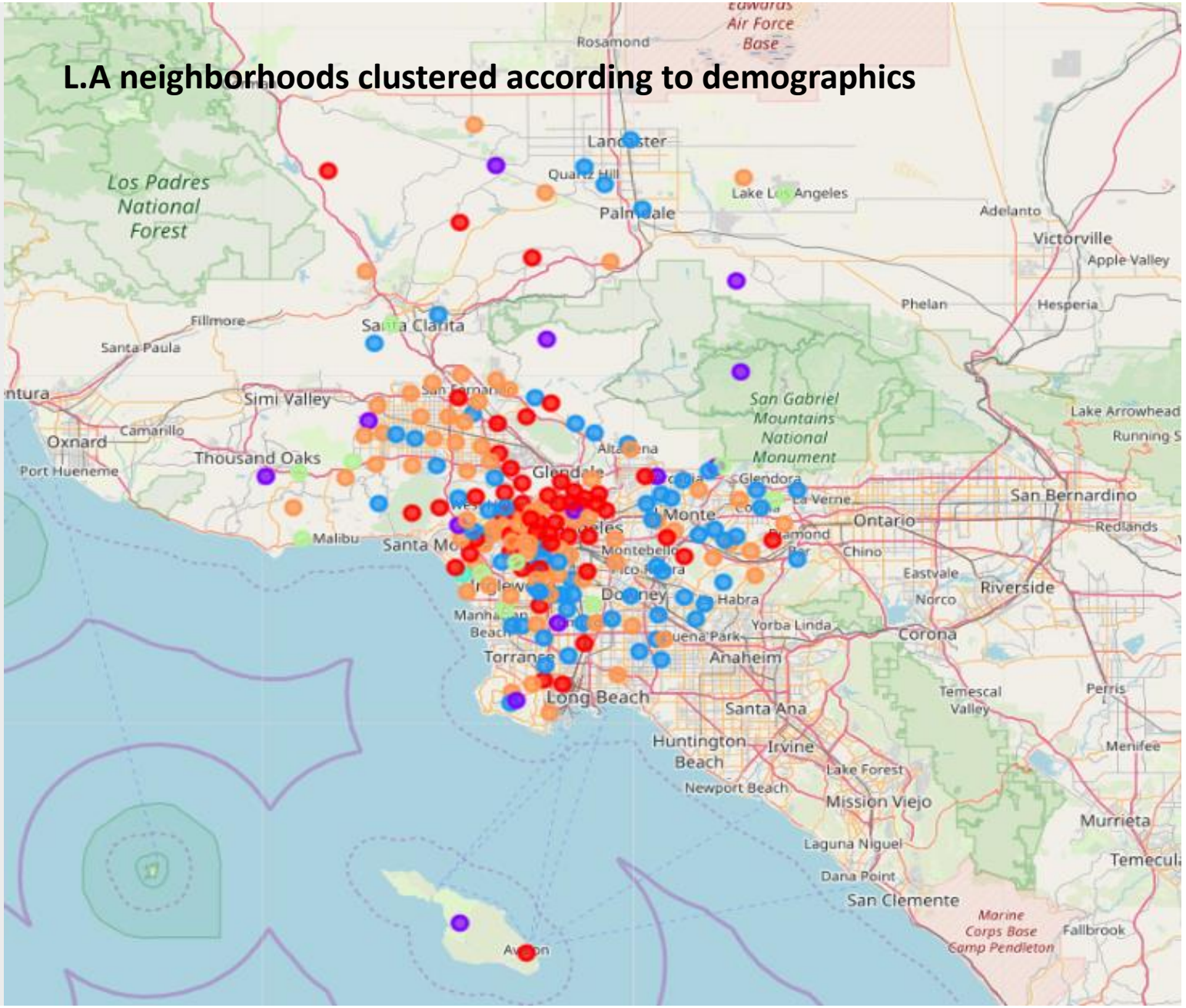
  Find the cluster that fits the user’s hobbies.

Methodology

- A data frame representing each neighborhood and its features.
- This data will be used to cluster the neighborhoods.

	Neighborhood	Latitude	Longitude	Population Ages 18-24	Population Ages 25-34	Population Ages 35-44	Population Ages 45-54	Population Ages 55-64	Population Ages 65 & Older	Population Under Age 18	Population	Divorced/Separated Population	Married Population	Never-Married Population	Widowed Population	Average Household Size	College Graduation Rate	Less than High School	Young Adults Enrolled in School	Unemployment Rate	Property Crimes Count Per 1000 People	Violent Crimes Count Per 1000 People
0	Acton	34.497355	-118.169810	11.169902	8.469121	10.291287	18.860127	20.441956	14.097019	16.670589	3885.5	8.500000	56.0	30.500000	5.000000	2.855	24.760050	10.885392	42.990420	11.454621	12.491046	1.647143
1	Adams-Normandie	34.031461	-118.300208	18.061678	15.167137	12.517903	12.066921	11.286173	8.616246	22.283943	3506.4	11.600000	34.2	54.400000	4.000000	3.250	17.561293	38.014641	54.191325	13.797029	20.134766	5.912570
2	Agoura Hills	34.146736	-118.759885	7.354860	7.939531	13.327865	19.588894	15.061215	13.090613	23.637022	6286.0	10.666667	57.0	30.666667	3.333333	2.820	54.501554	4.713622	48.622389	4.169422	13.465794	0.619836
3	Agua Dulce	34.504927	-118.317104	7.206238	10.943802	6.802904	18.822264	21.806938	16.617371	17.800484	3719.0	9.000000	68.0	21.000000	4.000000	2.960	26.909842	5.918789	46.416382	4.942166	5.646679	0.537779
4	Alondra Park	33.889617	-118.335156	7.542236	13.515688	13.334674	15.567176	10.498794	18.986324	20.555109	4972.0	14.000000	44.0	37.000000	8.000000	3.120	34.261917	9.276277	38.104839	9.914321	23.129526	3.620274

Methodology



Methodology

- 100 of the most popular venues within a 500 m radius of the center of each neighborhood were gathered using the foursquare API. The 10 most common venues in each neighborhood were collected in a data frame.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agoura Hills	Fast Food Restaurant	Sushi Restaurant	Chinese Restaurant	Breakfast Spot	American Restaurant	Hotel	Lounge	Mexican Restaurant	Multiplex	Restaurant
1	Arlington Heights	Donut Shop	Korean Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Karaoke Bar	Latin American Restaurant	Sushi Restaurant	Nightclub	Dance Studio	Music Venue
2	Bellflower	Sandwich Place	Pizza Place	Fast Food Restaurant	Video Game Store	Mexican Restaurant	Chinese Restaurant	Clothing Store	Burger Joint	Food Truck	Food
3	Beverlywood	Boutique	Food Truck	Food	Hotel	Park	Yoga Studio	Farmers Market	French Restaurant	Filipino Restaurant	Fast Food Restaurant
4	Canoga Park	Mexican Restaurant	Sports Bar	Restaurant	Sushi Restaurant	Liquor Store	Ice Cream Shop	Donburi Restaurant	Dog Run	Donut Shop	Dumpling Restaurant

Methodology

Results

- Neighborhoods were divided into clusters. In the first stage the clusters were formed using demographic data of each neighborhood. Mean values of each cluster were calculated.
- This data will allow us to match a user to the cluster that best fits his personal status.
- For instance, if the user is a 23 years old single man, with a collage degree, we can look for a cluster where a large percent of the population is single and has a collage degree.

Results

- The second part will be to cluster the neighborhoods according to the venues.
- Looking at the top number of venues in each cluster we can recommend a cluster according to the users' hobbies.
- For instance, if a user enjoys spending time outdoors, we can look for a cluster where the top venues are parks.

Conclusion

- Using simple methods it is possible to save the effort that goes into researching a new city.
- Online real estate search engines can use this method to customize the properties list they show each user.