

Large City Exodus: Finding a New Home

Hanpu Liu

1. Problem Description

With COVID-19 strike the world in 2020, people's life changed dramatically around the world. In United States, the country with the most cases in the world, each family has their struggle to adapt to their new life. Staying at home and social distancing become a major part of daily life. Adult works from home and children take online class at home. People spend more and more time at home which make them realize their current home cannot fulfill all their need. People are looking for new home with more space and low costs, and some people even move away from major metropolis, such as California bay area and New York City, to places with lower living cost.

Due to Texas's no state income tax, relative low living cost, vibrant technology industry and solid infrastructure, state of Texas has become one of the most popular choice to move to, alongside with other states such as Nevada, Colorado and Florida. Even large corporations such as Tesla, Hewlett Packard and Oracle are moving their head-quarter to Texas.

In this capstone project, I'm going to use machine learning technique to find a new home in Dallas-Fort Worth (DFW) metropolis area for Asian Americans, considering several factors that is essential to their daily life



Picture 1.1 Map of DFW Metropolis Area

2. Data

Gathering geographic location data of all the boroughs in Dallas is the first step in the project. There are a lot of information on internet about DFW and borough zip code information is used as the first step to gather borough names. Utilizing Geopy package in python, latitude and

longitude can be obtained by inputting full address. In picture 2.1, latitude and longitude are listed along side with address for each borough in DFW Metropolis Area. With latitude and longitude, detailed information such as restaurants and super markets can be retrieved using FourSquare API. In picture 2.2, a brief of venue information is listed.

	Borough	Address	Latitude	Longitude
0	Addison	Addison, Texas, United States of America	32.960431	-96.830260
1	Lucas	Lucas, Texas, United States of America	33.084285	-96.576658
2	Allen	Allen, Texas, United States of America	33.103174	-96.670550
3	Anna	Anna, Texas, United States of America	33.349280	-96.549380
4	Lantana	Lantana, Texas, United States of America	33.094028	-97.124497

Picture 2.1 Example of Latitude and Longitude Data

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Addison	32.960431	-96.83026	Addison Circle Park	32.960917	-96.826488	Park
1	Addison	32.960431	-96.83026	WaterTower Theatre	32.962105	-96.828262	Performing Arts Venue
2	Addison	32.960431	-96.83026	Antonio Ristorante	32.961468	-96.826096	Italian Restaurant
3	Addison	32.960431	-96.83026	Gaia Flow Yoga	32.960921	-96.823957	Yoga Studio
4	Addison	32.960431	-96.83026	Texas de Brazil	32.954592	-96.830206	Brazilian Restaurant

Picture 2.2 Example of Venues Data

3. Methodology

Based on all the venues information, cluster is carried to divides all DFW boroughs into several groups. Generally speaking, some boroughs are not intended for living. For Example, downtown area is intended for business and corporations which lacks grocery stores, school and have high living cost. DFW airport is categorized as a borough which certainly should be excluded from our analysis

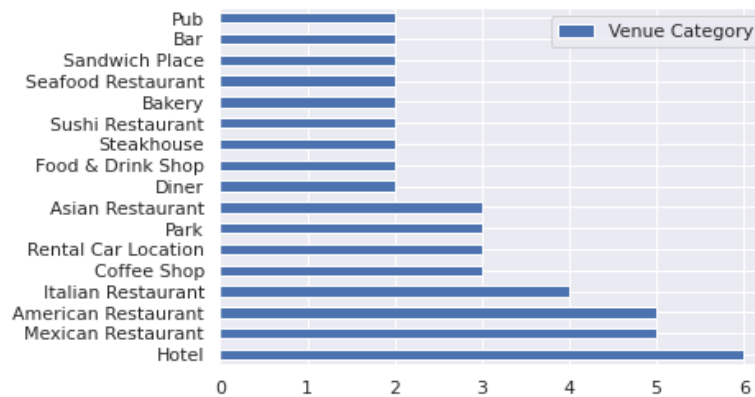
3.1 Exploratory Data Analysis

Since the type and amount of venues a borough can provide is the key of looking for a new home, the exploratory data analysis part will be centered on that. In picture 3.1.1, a top venue list is provided. This analysis provides a simple but essential overview of the borough. This analysis will also be used later in the rating section to analyze each borough in depth.

Borough	Address	Latitude	Longitude	Cluster Labels	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	Addison	Addison, Texas, United States of America	32.960431	-96.830260	0.0	Hotel	Mexican Restaurant	American Restaurant	Italian Restaurant	Coffee Shop	Asian Restaurant	Rental Car Location	Park	Steakhouse	Seafood Restaurant
1	Lucas	Lucas, Texas, United States of America	33.084265	-96.576658	0.0	Lake	Furniture / Home Store	Massage Studio	Martial Arts School	Ethiopian Restaurant	Dive Bar	Donut Shop	Dry Cleaner	Electronics Store	Yoga Studio
2	Allen	Allen, Texas, United States of America	33.103174	-96.670550	0.0	Fast Food Restaurant	Mexican Restaurant	Gym / Fitness Center	Donut Shop	Pizza Place	Restaurant	BBQ Joint	Breakfast Spot	Hotel	Asian Restaurant
3	Anna	Anna, Texas, United States of America	33.349260	-96.549380	0.0	Pizza Place	American Restaurant	Intersection	Discount Store	Italian Restaurant	Burger Joint	Gas Station	Fast Food Restaurant	Event Space	Event Service
4	Lantana	Lantana, Texas, United States of America	33.094028	-97.124497	0.0	Trail	Sandwich Place	Italian Restaurant	Tennis Court	Wine Bar	Gym	Restaurant	Golf Course	Convenience Store	Nightclub

Picture 3.1.1 Top 10 Venue of each Borough

Taking Addison as an Example, a bar chart is draw as Picture 3.1.2 to show the number of venues types in the area.



Picture 3.1.2 Addison Top Venues Categories

3.2 Clustering

Utilizing clustering, areas with few grocery stores, restaurant and schools can be easily distinguished and eliminated from our further analysis. 5 number of cluster is set before the clustering is carried out and about 20 boroughs is eliminated form further analysis. More detailed information can be found in notebook.

3.3 Rating System

Since the object is to find a new home, grocery store is a must have in the area. Specific to Asian American, Asian related venues, such as Asian restaurant, are essential part for living quality. Base on above requirement, a simple rating system is made to rate each borough with a score.

A list of grocery store is created and a list of Asian restaurant is created. Each borough starts with 0 point. By going through each venue category in the top 10 list, if any category exists in grocery store list or Asian restaurant list, 1 point can be earned by this borough. Finally, boroughs earned more than 1 point are selected as the final results.

4. Result

Using the methodology described above, Richardson, Carrollton, Chandler, Desoto, Greenville, Midlothian and Rowlett are the final boroughs to find a new home.

5. Discussion

The result in general is satisfying. Boroughs which are certain to be eliminated such as downtown and airport are not in the list, and the list is fairly small.

There are some limitations to the methodology worth improving in the future. In the data section, the radius of retrieving venue is arbitrarily set, which will cause inconsistency in the number of venues for all boroughs, because each borough has different sizes. There is no one radius fits all borough. And also the rating system is very simple. More research can be done to come up with a more sophisticated one.

Conclusion

There are various factors to be considered when relocating to new places. Commute time, education resources, housing price and neighborhood safety are some of the top reason to consider. With this project, it is proved that machine learning technique can be used to at least simplify the problem when data is available. But like many machine learning project, data quality is a key part and be the limitation to our findings.