Capstone Project - The Battle of Neighborhoods

A Study of Neighborhoods in Overland Park

Huiping Qin

May 11, 2019

1. Introduction

Why Overland Park shows up frequently on the top list of best places to live in US? There are various info on the Internet, however deep neighborhoods analysis is lacking. Through data analysis and machine learning technology, we may find evidences to the answer, which may in turn help local business on recruiting marketing and attract more talents to Overland Park. It may also help people who want to find a better place to retire or to raise a family to make informed decision. I have lived in Overland Park for over 20 years and searching an answer on the question may help me to know my community better and to love it even more. I love my local community, with easy access to good education, libraries, museums, shopping, entertainment, park and trails, and so forth. But what are other neighborhoods in Overland Park like, what venues do they have to contribute to the listing of Overland Park on top places to live in the US?

Equipped with knowledge and skills just acquired from the Coursera Data Sciences classes, I hope to find the answer by using a combination of location data and machine learning to explore neighborhoods in the city of Overland Park. In this study, I am going to use the Foursquare API to explore neighborhoods in Overland Park. I will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. I will use the *k*-means clustering algorithm to complete this task. Finally, I will use the Folium library to visualize the neighborhoods in Overland Park and their emerging clusters.

2. Data

2.1 Data sources

To be able to segment and cluster neighborhoods in Overland Park, I will need the list of neighborhoods in Overland Park and their corresponding geo coordinates data. However, neither is readily available. I have searched the Internet and found neighborhoods list on the nextdoor website: https://nextdoor.com/city/overland-park--ks/. I have used the BeautifulSoup Python library to scrape the

names of the neighborhoods from the nextdoor website. To get the neighborhoods' geo coordinates, I have used the geopy library to get the latitude and longitude values of the neighborhoods. For the location data such as data describing places and venues, I have used Foursquare API to get info from their server. Please find below Fig. 1 of snapshot of neighborhoods list of Overland Park on Nextdoor and Fig. 2, a sample of data from Foursquare.

252 Overland Park neighborhoods are on Nextdoor

1	H	Q
151st/Metcalf Ave	Hampton Park	Quail Crest
	Hampton Place	Quail Valley
7	Hamptonshire	Quincy Court
75th and Metcalf	Hanover South	Quivira Falls
75th and Woodson	Harmony South	Quivira Farms
	Harwycke	
A	Hawthorne	R
Access Rd	Heatherwood	Ranchview Gardens
Adara	Heritage Farms	Regency By The Lake
Amber Meadows	Hidden Woods	Regency Park

Fig. 1 Overland Park Neighborhoods on Nextdoor Snapshot

	name	categories	lat	lng
0	Downtown Mission	Historic Site	39.014909	-94.662374
1	ARC	Gym / Fitness Center	39.013159	-94.663269
2	Henhouse	Grocery Store	39.010678	-94.667633
3	98.9 The Rock!	Rock Club	39.016930	-94.666710

Fig. 2 Sample Data from Foursquare (Both name and categories info for each venue)

2.2 Data Wrangling

Neighborhoods' names were scraped from nextdoor website. Since they are local specific neighborhood names and got erroneous coordinates back per geopy query. I have added " Overland Park, Kansas" to the neighborhood names as addresses and got back reasonable geo coordinates. The geopy package did not return geo coordinates for all neighborhoods in the list, so I have to limit the analysis on only the neighborhoods with geo coordinates returned. Eventually, I have built a dataframe of neighborhoods with their top 10 most common venues in their respective neighborhood, ready for segmentation and clustering analysis.

	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	Access Rd	Gym / Fitness Center	Rock Club	Historic Site	Grocery Store	Women's Store	Fried Chicken Joint	Food Truck	Food	Fast Food Restaurant	Fabric Shop
1	Adara	Gym / Fitness Center	Basketball Court	Salon / Barbershop	Kids Store	Sports Club	Thai Restaurant	Women's Store	Furniture / Home Store	Fried Chicken Joint	Food Truck
2	Apple Valley Estates	Gas Station	Sushi Restaurant	Pizza Place	Arts & Crafts Store	Grocery Store	Liquor Store	Garden	Furniture / Home Store	Fried Chicken Joint	Food Truck
3	Brittany Park	Playground	Park	Health & Beauty Service	Gym / Fitness Center	Women's Store	Fabric Shop	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food
4	Caenen	Gym / Fitness Center	Gas Station	Gym	Cosmetics Shop	Coffee Shop	Salon / Barbershop	Fast Food Restaurant	Smoothie Shop	Gift Shop	Sports Club

Fig. 3 First 5 rows of the dataframe with top 10 most common venues in each neighborhood