Clustering Top 50 Most Populated Cities In The United States
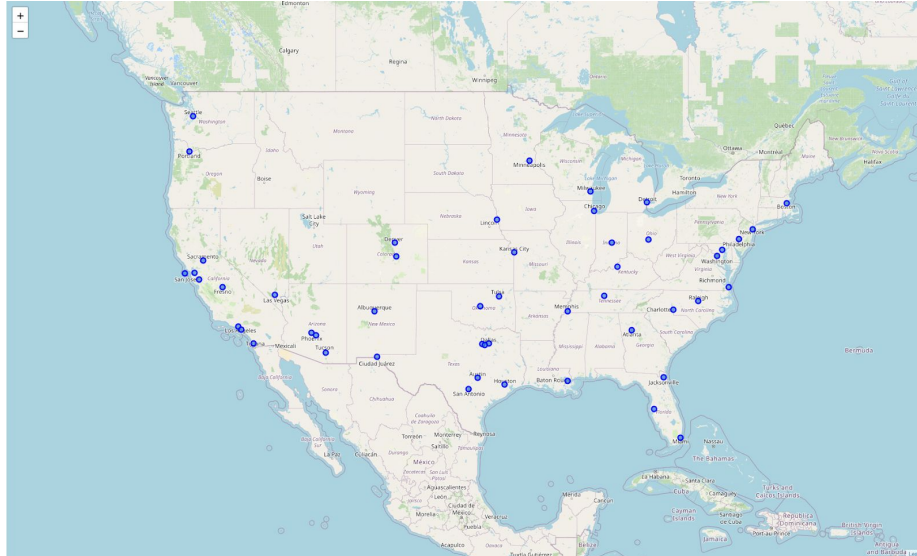
Introduction

Looking into moving to a similar city in the US? Trying to find a totally different city to live? Or trying to expand the business, open the next chain store, but don't know which city to start? This study clusters the top 50 most populated cities in the US by their most common business types. The study will help people to have ideas about which cities to explore, to move to, and for business owners to evaluate which cities to expand their stores.

Data

The population ranking of the cities in the US can be found on Wikipedia[1]. The coordinates of the cites can be located from Python module "Geopy". The Foursquare API will be used to find the venues and business types in each city. K-mean clustering will be used for clustering the cities by their most common business types. Similar cities and different cities can be identified after K-mean clustering is performed.
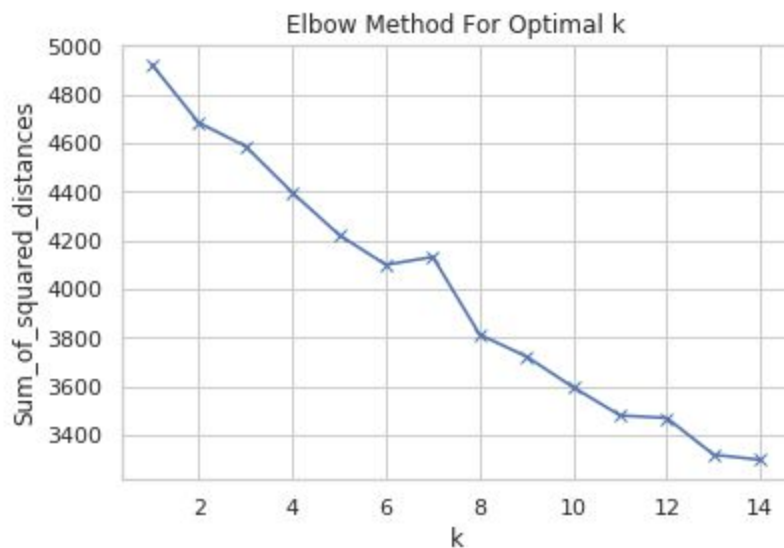
Methodology

The top 50 most populated cities in the US are first identified from the population estimated by the United States Census Bureau. The coordinates of the 50 cities are located by Geopy and plotted on the map.

Using Foursquare API, after geotagging the 50 cities, 200 venues were identified within 5000 meters of the center of the cities. The 200 venues of each city are categorized by their business types, then the most five common venue types are identified for each city.

The K-mean clustering is used to cluster the 50 cities into few groups. To find the optimal K number, the Elbow Method is used.

From the plot, K is chosen to be "8". However, there is no significant drop of Sum of Squared Distance at any K number from 1 to 15.

To understand how the population density of each city is related to its most common venue types (or the other way around). The population density information will be later merged into the table of venue information.
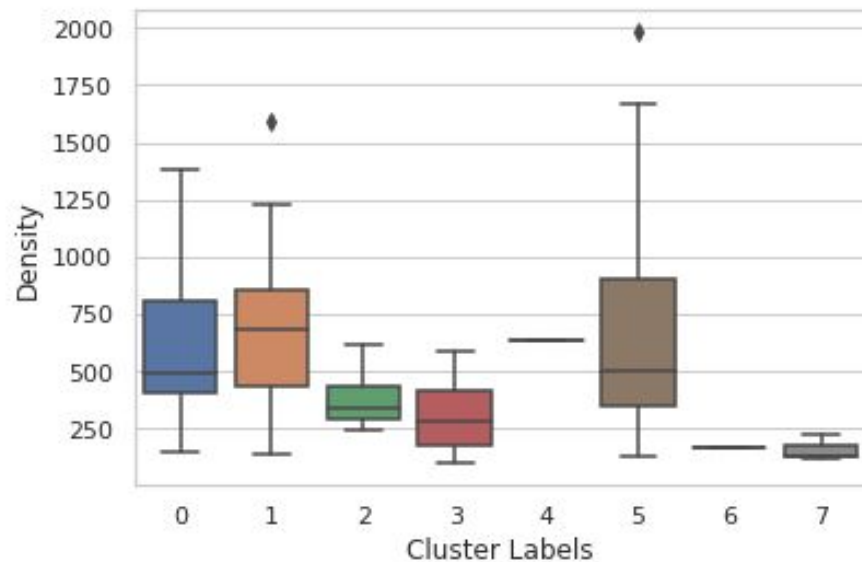
Results

50 cities are categorized into 8 groups by K-means clustering. The summary table is shown below. The table also shows the top 5 most common venues in each city.

Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Albuquerque, New Mexico	Mexican Restaurant	Brewery	Coffee Shop	Pizza Place	American Restaurant
0	Phoenix, Arizona	Coffee Shop	Mexican Restaurant	Park	Pizza Place	Taco Place
0	Las Vegas, Nevada	Cocktail Bar	Brewery	Mexican Restaurant	Coffee Shop	Lounge
0	Indianapolis, Indiana	Bar	Brewery	Mexican Restaurant	Coffee Shop	Hotel
0	Oakland, California	Coffee Shop	Bakery	Café	Beer Garden	Mexican Restaurant
0	San Jose, California	Mexican Restaurant	Sandwich Place	Coffee Shop	Grocery Store	Bar
0	El Paso, Texas	Mexican Restaurant	Pizza Place	Park	Bar	Coffee Shop
1	New York, New York	Park	Italian Restaurant	Memorial Site	Garden	Bookstore
1	New Orleans, Louisiana	Hotel	Coffee Shop	Park	Bar	Cajun / Creole Restaurant
1	Omaha, Nebraska	Pizza Place	Coffee Shop	Bar	Zoo Exhibit	American Restaurant
1	Minneapolis, Minnesota	Brewery	Park	Pizza Place	Farmers Market	Theater
1	Louisville, Kentucky	Park	Coffee Shop	Hotel	Brewery	Breakfast Spot
1	Los Angeles, California	Coffee Shop	Theater	Brewery	Hotel	Plaza
1	San Antonio, Texas	Hotel	Coffee Shop	Beer Garden	Brewery	Lounge
1	Oklahoma City, Oklahoma	Brewery	Bar	Vietnamese Restaurant	Restaurant	Pizza Place
1	Columbus, Ohio	Park	American Restaurant	Brewery	Hotel	Coffee Shop

1	Atlanta, Georgia	Trail	Park	Southern / Soul Food Restaurant	History Museum	Mexican Restaurant
1	Colorado Springs, Colorado	Coffee Shop	Café	Brewery	Bar	Italian Restaurant
1	Charlotte, North Carolina	Park	Hotel	Brewery	American Restaurant	Mexican Restaurant
1	San Diego, California	Hotel	Park	New American Restaurant	Italian Restaurant	Seafood Restaurant
2	Fresno, California	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Convenience Store	Pizza Place
2	Arlington, Texas	Mexican Restaurant	American Restaurant	Fried Chicken Joint	Vietnamese Restaurant	Asian Restaurant
2	Tucson, Arizona	Mexican Restaurant	Convenience Store	Pharmacy	Coffee Shop	Pizza Place
2	Mesa, Arizona	Mexican Restaurant	Convenience Store	Brewery	Fast Food Restaurant	Sandwich Place
3	Portland, Oregon	Coffee Shop	Hotel	Park	Grocery Store	Bookstore
3	Chicago, Illinois	Hotel	Grocery Store	Park	Italian Restaurant	New American Restaurant
3	Miami, Florida	Hotel	Seafood Restaurant	Italian Restaurant	Grocery Store	Japanese Restaurant
3	Baltimore, Maryland	Bar	Coffee Shop	Burger Joint	Hotel	Italian Restaurant
4	Virginia Beach, Virginia	Beach	Seafood Restaurant	Hotel	American Restaurant	Coffee Shop
5	Tulsa, Oklahoma	Coffee Shop	American Restaurant	Mexican Restaurant	Park	Bar
5	Raleigh, North Carolina	Beer Store	Park	Mexican Restaurant	Bakery	Coffee Shop
5	Tampa, Florida	Coffee Shop	American Restaurant	Park	Grocery Store	Smoke Shop
5	Philadelphia, Pennsylvania	Coffee Shop	Plaza	Park	Pizza Place	Grocery Store
5	Sacramento, California	Coffee Shop	American Restaurant	New American Restaurant	Sandwich Place	Park
5	Memphis, Tennessee	Park	Bar	American Restaurant	Coffee Shop	Museum
5	Milwaukee, Wisconsin	Park	Bar	Brewery	Hotel	Coffee Shop
5	Long Beach, California	Mexican Restaurant	Park	American Restaurant	Seafood Restaurant	Burger Joint
5	Kansas City, Missouri	Brewery	Coffee Shop	BBQ Joint	Park	New American Restaurant
5	Jacksonville, Florida	Park	Coffee Shop	American Restaurant	Bar	Sandwich Place

5	Houston, Texas	Coffee Shop	Mexican Restaurant	Burger Joint	Hotel	Park
5	Fort Worth, Texas	Mexican Restaurant	Brewery	Hotel	New American Restaurant	American Restaurant
5	Detroit, Michigan	Park	American Restaurant	Farmers Market	Concert Hall	Burger Joint
5	Denver, Colorado	American Restaurant	Brewery	Sandwich Place	Coffee Shop	Hotel
5	Dallas, Texas	Park	Coffee Shop	Burger Joint	American Restaurant	Hotel
5	Austin, Texas	Coffee Shop	Food Truck	Hotel	Park	Pizza Place
5	Nashville, Tennessee	Coffee Shop	Music Venue	Pizza Place	Park	Hotel
6	Washington, District of Columbia	Monument / Landmark	Art Museum	American Restaurant	Hotel	History Museum
7	San Francisco, California	Bakery	Coffee Shop	Park	Yoga Studio	Grocery Store
7	Seattle, Washington	Coffee Shop	Vietnamese Restaurant	Bakery	Sandwich Place	Sushi Restaurant
7	Boston, Massachusetts	Park	Bakery	Hotel	Seafood Restaurant	Italian Restaurant

The population density is not taken into consideration during the clustering. However, we can see from the plot below, there is some difference of population density between some groups.

Discussions

1. The coordinates of the center of the cities are used to search for venues the range of 5000 meters from the center. The venues in these areas are close to the city centers, but might not be representative of the culture of the cities.
2. There is no clear elbow point when finding the optimal K number using Elbow Method. This could imply each city has its own characteristic and is very different from other cities. This can also mean other variables, such as weather, geographical location, or city history, can be added to have more meaningful grouping.
3. Comparing the population density between the 8 groups shows some groups are different from others. This could mean some business types are favorable in more crowded areas, and some business types are preferred by people who are more distant from others.

Conclusion

Grouping the cities by their most common venues can help people to find similar cities, or to explore different cities.

For business owners trying to expand their business to other cities, it can potentially help them to find the cities where the business type is more favorable to the residents.

References
1. Wikipedia https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population
2. Foursquare API