1. Introduction

1.1. Background

Foursquare provides location-based data that can be used for a variety of purposes. Two examples of this data is a list of "Top Pick" venues for a given geographical location, and a list of "Trending" venues for a given geographical location and at a specific time (i.e. when the API is called).

1.2. Problem

This project focuses on classifying the most populous US cities, based on the type of venues that are listed as "top picks," as well as venues that are "trending" on a Sunday afternoon during the summer season.

1.3. Interest

1.3.1. Categorization of Cities Based on Top Picks

This type of categorization would be of interest to many types of people. Business owners looking to expand into other large US cities would be interested to see which locations have people with similar interests. Leisure travelers could be interested in exploring similar cities based on their travel preferences.

1.3.2. Categorization of Cities Based on Trending Venues at a Given Time

This type of analysis is useful for businesses to identify peak times and off-peak times to address stock and staffing concerns. This information could also be used by customers to determine personally-ideal times for visiting particular types of venues. This project was run at one specific time – Sunday afternoon during the summer season – but the analysis could be run for all types of times for a more detailed view.

2. Data

2.1. Data Sources

All data was extracted from the following sources:

2.1.1. Most Populated Cities (Wikipedia)

The list of most populated cities, including names and geographical coordinates, was scraped from Wikipedia using the Beautiful Soup Python library.

2.1.2. Top Pick Venues (Foursquare)

The venues listed as "top picks" for each city was extracted using the Foursquare API. I extracted 100 top venues for each city, along with the venues' respective categories.

2.1.3. Trending Venues on a summer Sunday afternoon (Foursquare)

Trending venues for each city was extracted using the Foursquare API. I extracted 100 trending venues for each city on a Sunday afternoon during the summer, along with the venues' respective categories.

2.2. Data Preparation and Data Cleaning

2.2.1. Most Populated Cities (Wikipedia)

I scraped the table from Wikipedia as stated above, and removed all columns except for Rank, City, State, Population (2018 Estimate), and Location. I then separated the Latitude and Longitude from the Location column to enable easy parsing from the Foursquare API. A subset of this data is shown below:

2018 Rank	City	State	2018 Estimate	Latitude	Longitude
1	New York	New York	8,398,748	40.6635	-73.9387
2	Los Angeles	California	3,990,456	34.0194	-118.4108
3	Chicago	Illinois	2,705,994	41.8376	-87.6818
4	Houston	Texas	2,325,502	29.7866	-95.3909
5	Phoenix	Arizona	1,660,272	33.5722	-112.0901

2.2.2. Top Pick Venues (Foursquare)

Using the latitude and longitude for each city, I then called the Foursquare "explore" destination endpoint to get a list of 100 top picked venues for each city, and then sorted the categories of top picked venues for each city. A subset of this data is shown below:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Albuquerque	Pizza Place	Brewery	Mexican Restaurant	Café	American Restaurant
1	Arlington	Brewery	Coffee Shop	Art Museum	Gourmet Shop	Taco Place
2	Atlanta	Trail	Brewery	Park	American Restaurant	Ice Cream Shop
3	Austin	Coffee Shop	Ice Cream Shop	Taco Place	Pizza Place	BBQ Joint
4	Baltimore	Seafood Restaurant	Park	Ice Cream Shop	Coffee Shop	BBQ Joint

2.2.3. Trending Venues on a summer Sunday afternoon (Foursquare)

Using the latitude and longitude for each city, I then called the Foursquare "explore" destination endpoint to get a list of 100 trending venues for each city, and then sorted the categories of top trending venues for each city on a Sunday afternoon during the summer. A subset of this data is shown below:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Albuquerque	Pizza Place	American Restaurant	Coffee Shop	Brewery	Grocery Store
1	Arlington	Grocery Store	Brewery	American Restaurant	Fast Food Restaurant	Park
2	Atlanta	Trail	American Restaurant	Park	Brewery	Grocery Store
3	Austin	Pizza Place	Coffee Shop	Taco Place	Ice Cream Shop	Sandwich Place
4	Baltimore	BBQ Joint	Park	Fast Food Restaurant	Coffee Shop	Gym