Project: Compare the Top Populated Cities in the USA

**Introduction**

I will view the top populated cities in the USA. I will cluster them into groups and visualize the result on a map. I will adjust the clusters from 2 to 5 and see the effect on the map.

One of the questions I want to see if I can answer is in the USA is there a difference or similar in certain areas of the USA? Example will we see a difference when we compare the West Coast, East Coast, and the Midwest. Also I will compare the very large cities with the smaller cities. Or is the US a “melting pot” and there is not much difference throughout the country when it comes to the types of venues located in the most populated cities in the US. I would like the compare the cities and determine how similar or dissimilar they are. Through this project I am expecting following people to benefit out of the findings.

* People moving to different cities for work
* Business Companies looking for new locations or to expand
* Restaurants to adjust their menu based on the people's likings and feedbacks

**Data**

The data I will be using will come from the following web site,

<https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population>

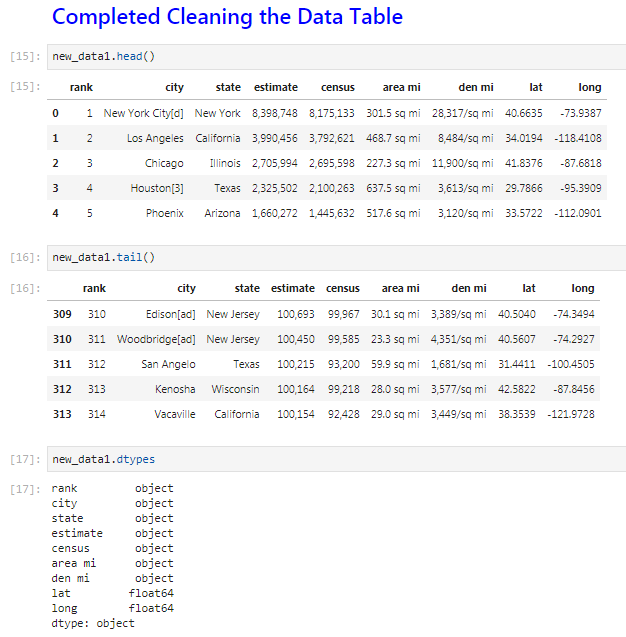


I will use beautiful soap to read the tables from this website. I will have to do some data "cleaning" in python pandas to get the data into a correct dataframe, so that the table and data will be ready to be read by the Foursquare API program section on my Python program.

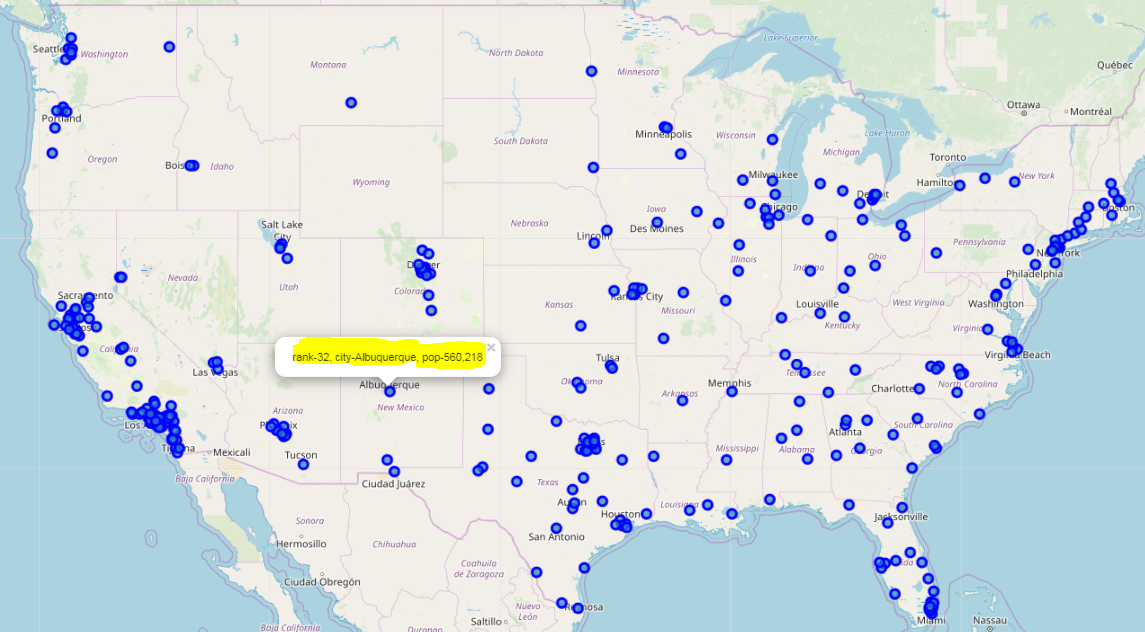
**Methodology**

After loading in the data from the website (<https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population>)

I had to use Pandas to clean the data to get the following dataframe output with Rank, City, Latitude and Longitude.



Because I am not a master at Pandas yet, it took me a number of steps to get the data into the right format. I then used the folium package to visualize the map of the US with displaying the popup label with Rank, City and Population. I did this step so that I know the data was “clean” and was displaying corredctly on the map.

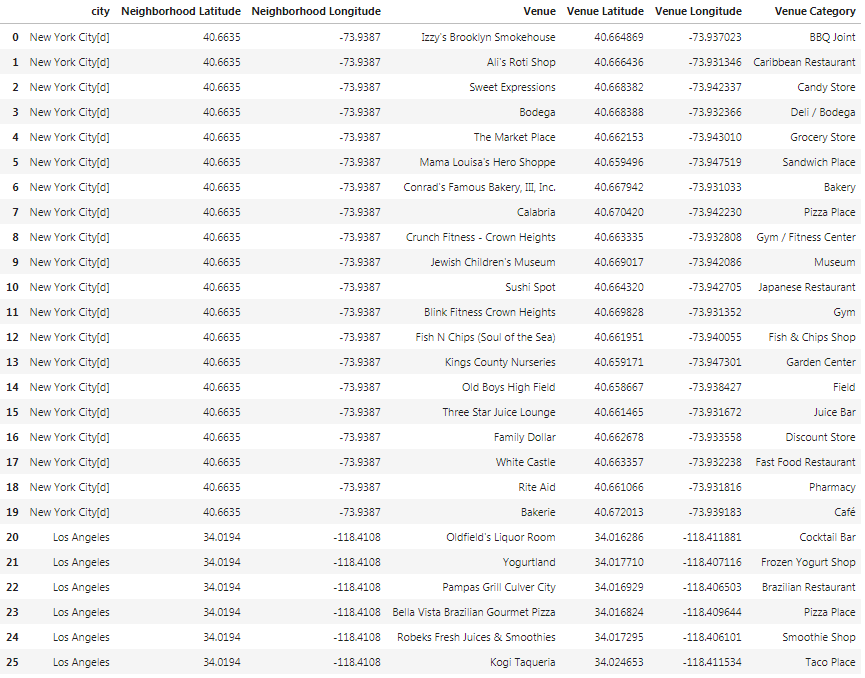


I next used the Foursquare section of the program to explore the cities and venues.

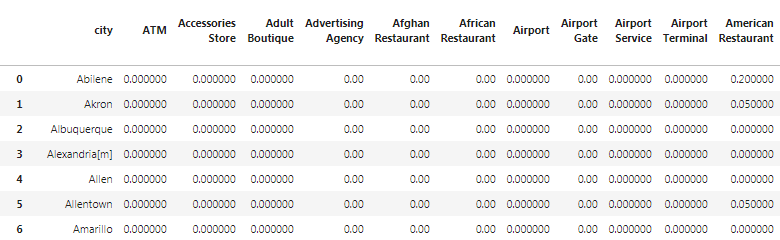
I set the limits to 20 venues and with a radius of 1000 meters. This list is 4941 lines deep. So I know I did not get 20 venues for all of the 314 cities on the list. As you can see blow.



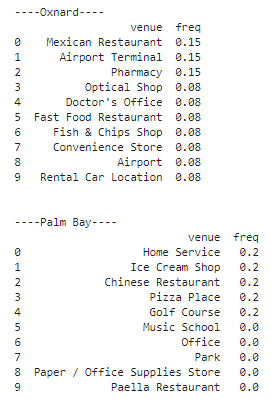
In the picture below, you can see New York with 20 venues and then Los Angeles as the next city on the list.



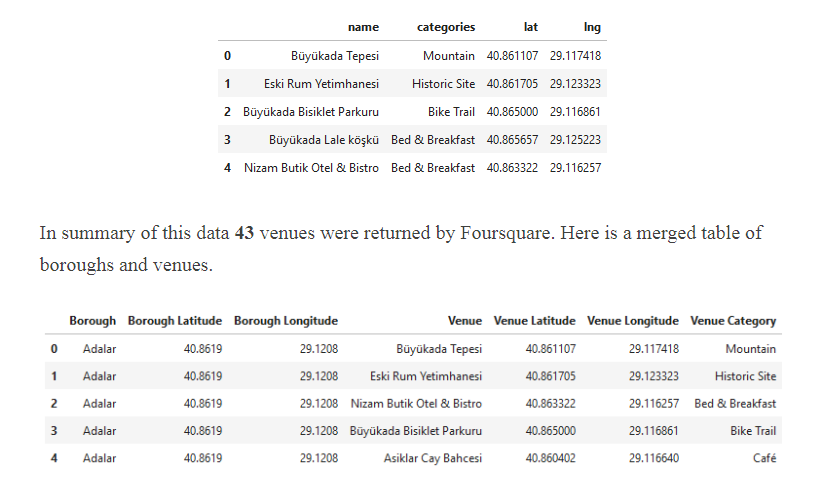
When I looked at all of the Venue Catagories, I had a total of 379.



Next I took onoy the top 10 venue catagoies for each city, see picuter below , where some cities did not have a percenagte for all 10 catagoies.



Show table here and merge table here



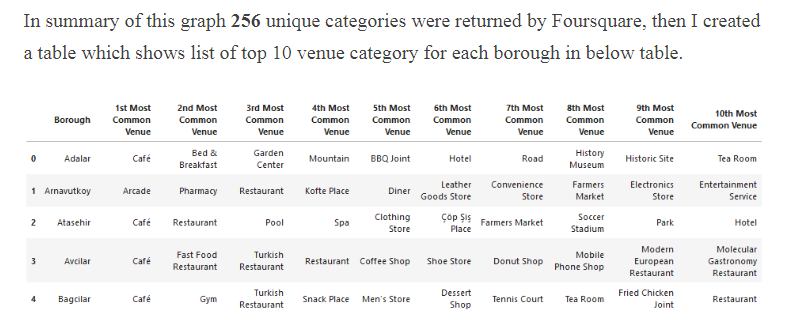
Methodology section which represents the main component of the report where you discuss and

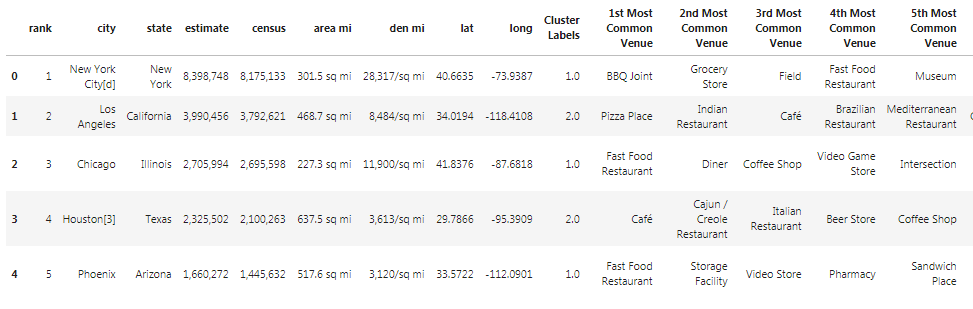
describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

**Results**

ere you discuss the results.

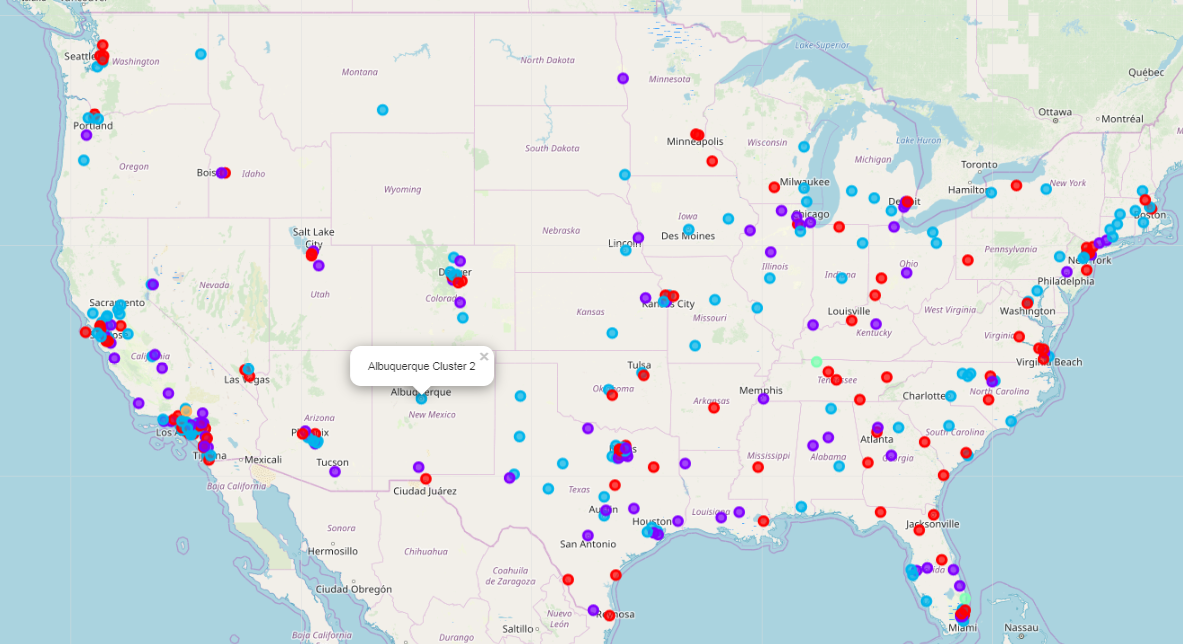






**Discussion** se

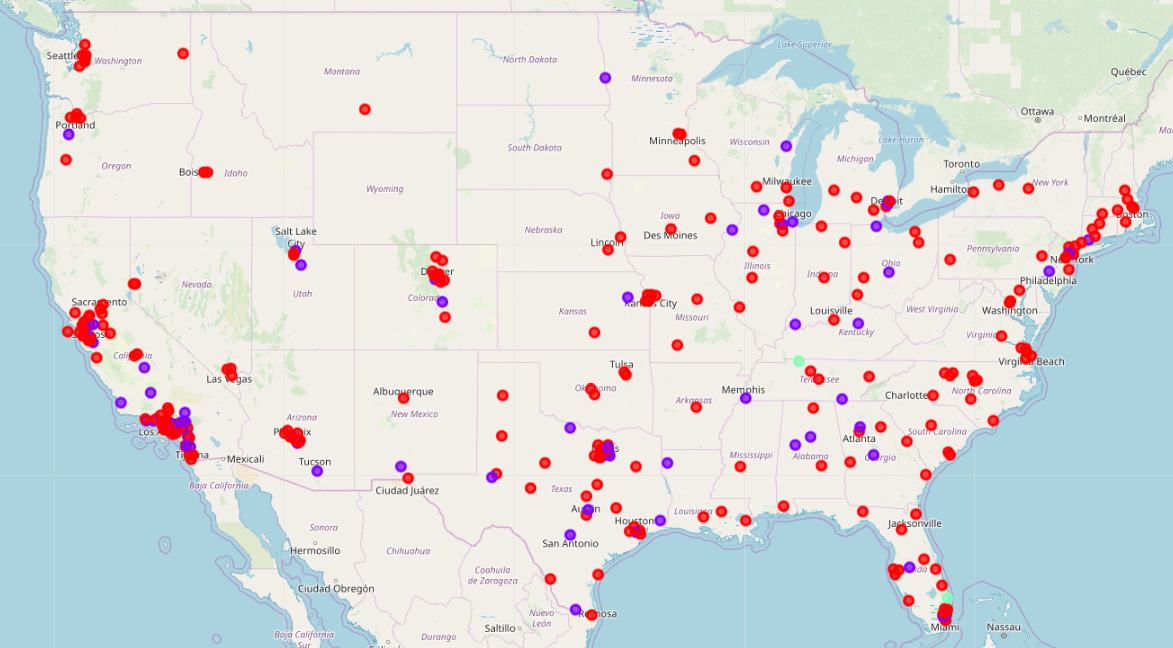
When looking at the maps of the US, I did not see a major difference in the type of venues when comparing East Coast, West Cost and the Mid West. We do see a slight diffenece in the largeste cities with the smaller cities somewhat.



When I had the clusters set to 5, there was a great drop of for cuklsters 4 and cluster 5

Cluster 4 = 2 cities

Cluster 5 = 1 ciity



Cluster 3 has 2 city

So I adjusted my cluster to 3 and rerun my program.

I think ttrying to compare over 300 cities was not a good project for a 1st project for me. I think comparing 2 to 3 cities would have been more meaningful and easy to anaylyz.

**Conclusion** section

In this study, I analyzed the lraest citues on the US to see if there is a large diffentece in the vanuea in the courty. When I cluster the groups from 2 to 5. I did not see a large difference in the map. It seems the veneues in the US are very concsistancte no matter where you go .

During my travels throughout the US, this conclusion seems correct. When we go from city to city or state to state in the very large cities in the US, it does seem very simlair.

This is my 1st project using Pyhton Pandas, FourSquare and Mapping visualization and felt a learned a great deal about the tools, Next I need to futrhter this project study with more indeft detailed analiysis.