

A Tale of Two Cities

Introduction & Business Problem

Background and Problem:

United States is one of the most popular countries for tourist to visit. It hosts millions of people every year and it does not matter the season; it is always full of tourists. Whether it is New York or Los Angeles, these two cities are the most popular destinations for tourists all over the world. People often believe that these two states are the center of the 'East Coast vs West Coast war' to determine which one is the best place to be. They offer all the tourist attractions every city has, but what makes them outstand from other places is the different cultures that live there. These two sates are the most culturally diverse among the U.S and this is why they are so special for tourists.

The objective of this study is to use data science tools to help a tourist agency to determine a new travel package for their customers. Our job as data scientists is to give the agency multiple travel package options so they can sell them to their customers; these options will be presented by comparing New York and Los Angeles (what makes them different) and what different activities you can do in each city.

Target Audience:

Our target audience consists on every tourist that is interested to travel to the U.S. Since our objective is to give multiple package options, we can have different options for people that travel alone, in groups or as a family.

Applicability in the real world:

For travel agencies, it is sometimes difficult to predict what kind of attractions their clients want and often they only offer a general travel package so they can later choose by themselves where they want to go. But, what would happen if you could offer your clients a experience package based on their tastes or by the number of people going on a trip? This is the question I want to answer by analyzing the venues of these two cities in order to offer them the best experience possible.

Data acquisition

Looking for the information about each borough and neighborhood in Los Angeles and New York was relatively easy since these are two of the most popular cities in United States and, also, the information is of public domain. The files used for this project consist of two kinds; the first one, was acquired from the “New York neighborhoods” example from the course and the data for Los Angeles was acquired from a ‘.csv’ file in an online site that provides general information about all the cities in United States.

Data Cleaning

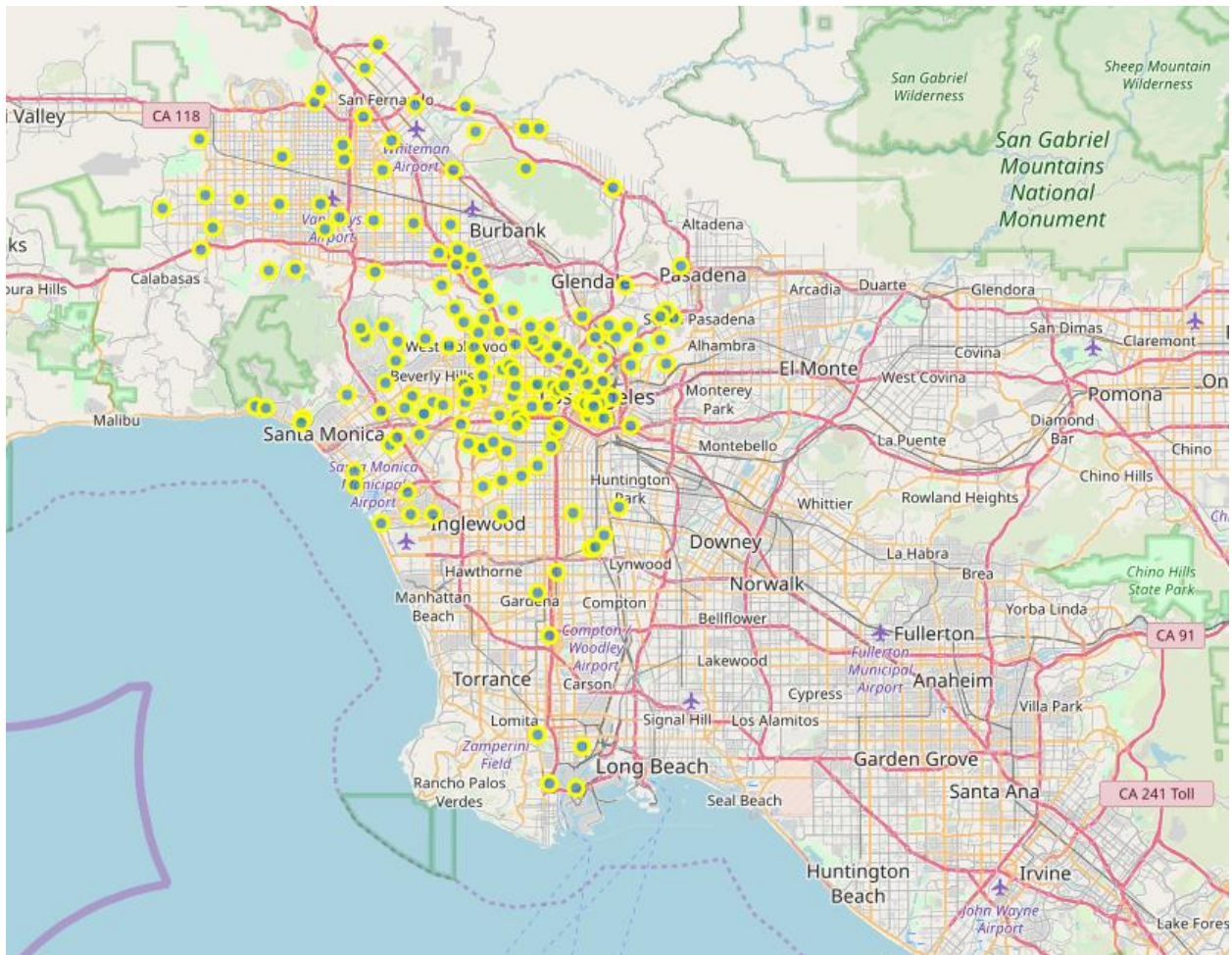
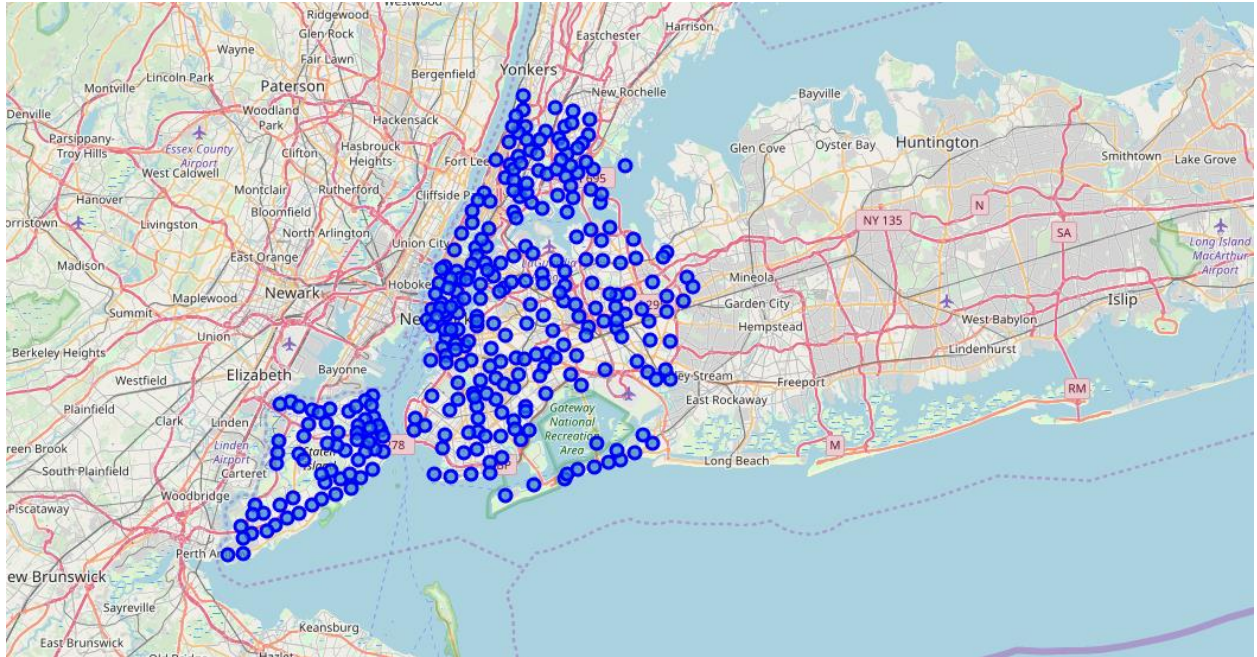
In order to use the Los Angeles data, I had to clean the dataset first since it had unnecessary information (native vegetation, native wildlife, number of houses, etc.). Once all the ‘junk’ information was removed from the dataset, the process on analyzing the data and using Foursquare to rank venues began.

Data Analysis

The first dataset to analyze was the New York one, since it was already analyzed for one city (Manhattan) in the exercise, I proceeded to analyze the data of all the State of New York. The data collected was bigger than the one in the example so I had to reduce the Foursquare search. Instead of using a limit of a 100 for the nearest venues, I changed to 50 because tourists tend to move around very often and do not want to be in the same place for too much time. The next thing I did was to search for the top 3 most frequented venues instead of 5 because of the same reason I mentioned above. Tourists only visit a maximum of 3 venues in a determined location and then they move along. The process mentioned above was realized for the Los Angeles dataset, it was important to match the limit and ranked venues to get a more exact result and comparison between these two cities.

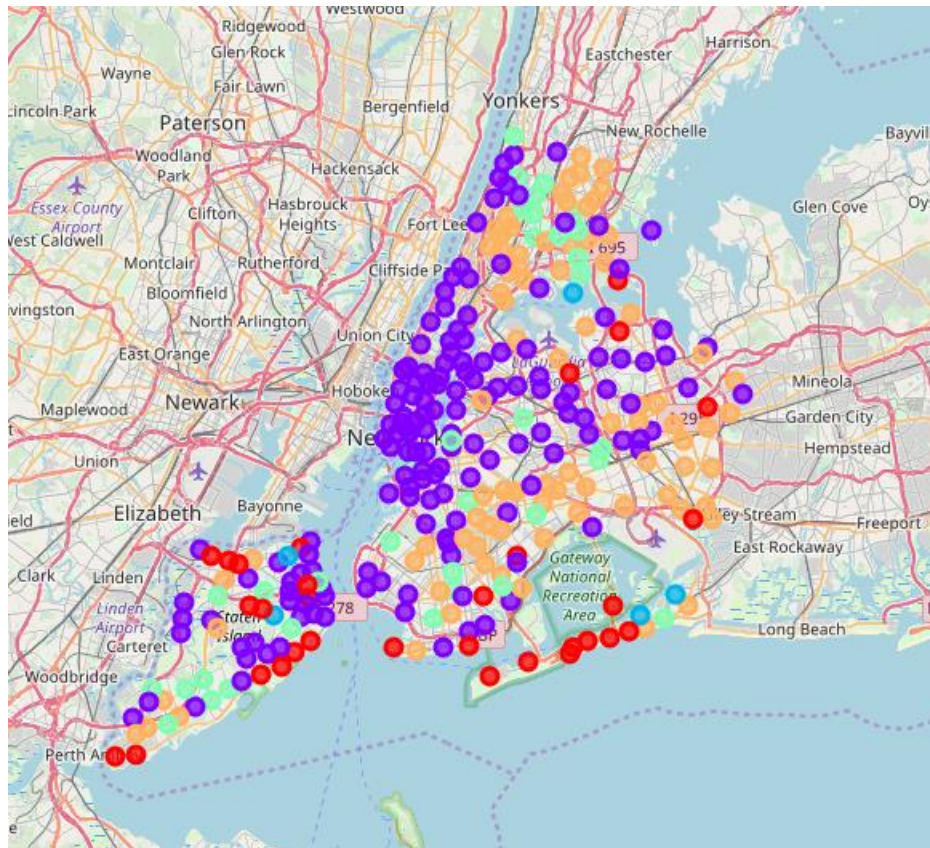
Results

To analyze the results the program gave, it is important to make a first comparison between New York and Los Angeles. First of all, we will compare how the neighborhoods are distributed. In the images below, we can see the difference between New York (blue) and Los Angeles (yellow) in their neighborhoods distribution.



As you can see, New York is a more centric location with neighborhoods interacting with each other by being close by. On the other hand, Los Angeles also present near by neighborhoods but we can also see there are various places close to the sea and mountains, this is mainly because people want to have a quieter lifestyle in California.

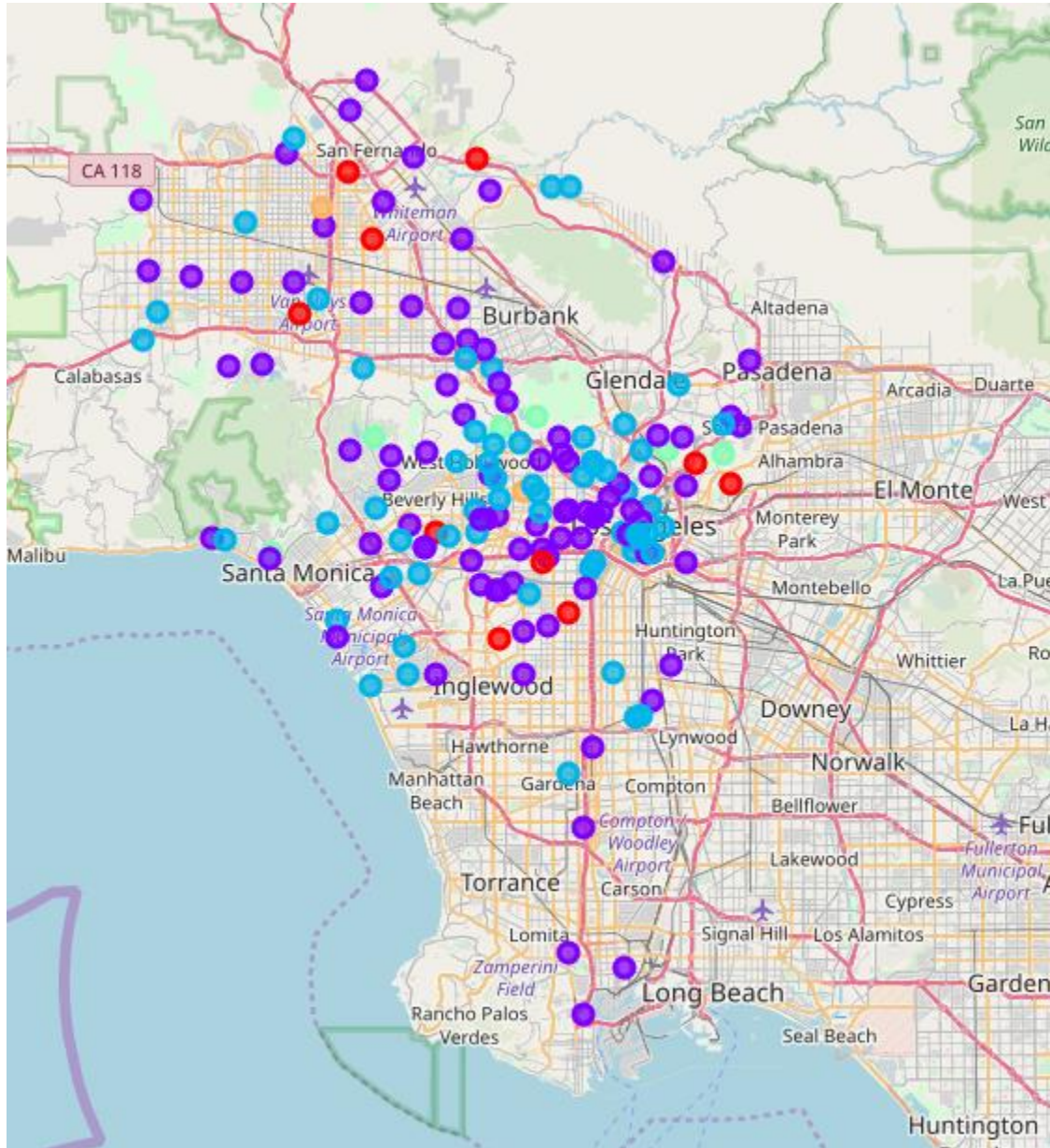
With this first comparison understood, we can now see how the neighborhoods are cluster together in order to determine how can we influence potential tourists on picking a destination. For New York, the map below represents the neighborhoods in each cluster



With this information I proceeded to describe each cluster for future customers:

Cluster #	Color on map	Cluster Name	Description
1	Red	<u>Main Attraction</u> Beaches, Bodegas, Pubs and more	For tourists that come either alone or as a couple
2	Purple	<u>Family friendly</u> Parks, Playgrounds, Restaurants, Spas, etc.	Different venues for the likes of all family members. This group is ideal as an ‘all around experience’
3	Sky blue	<u>Kids only</u> Parks, Fast Food restaurants	Perfect destination for families with small children.
4	Light green	<u>What you want to eat?</u> Italian, Chinese, Mexican, Latin American restaurants.	For people who consider themselves ‘foodies’, this group offers different food experiences.
5	Orange	<u>NY and chill</u> Ice cream Shops, Bars, Lounges, Theaters, Supermarkets	This package is for all kind of tourists since it offers different activities.

Now we can see the map for Los Angeles and their packages options.



Cluster #	Color on map	Cluster Name	Description
1	Red	<u>For the family</u> Parks, Playgrounds, Museums	A perfect package option for the entire family

2	Purple	<u>Some 'Me' time</u> Theaters, Art galleries, Spas, Restaurants, Wine Bars	This group is represented with places for people who want to have some alone time while relaxing in different places
3	Sky blue	<u>Fit and Delicious</u> Bakeries, Gyms, Yoga Studios, Restaurants, Coffee Shops	This group is filled with food options and different gym and sports options for the people that like 'that healthy lifestyle' but also enjoy good food
4	Light Green	<u>Deep breath of mother nature</u> Mountain trails	This is perfect for more adventurous tourists who want to take a break from all the noise of the city
5	Orange	<u>For the classic ones</u> Golf Courses	This small group is for tourist who want to feel a little bit classy and relax while playing some golf

Conclusion

We can see that New York and Los Angeles are not that different, they both offer interesting choices for every type of tourists. The main difference between these cities is that New York offers a more night time venues while Los Angeles offers options where you can explore the country side by taking a hike in the mountains. So, depending on whether you are a type of tourist who wants to experience the night life of a multi-cultural city, then New York is definitely your next destination. But if you are the kind of person who wants to explore different landscapes or try different types of physical activities then Los Angeles is where you need to be. We also created different packages for each city, so you have a clearer idea on where you want stay so you can enjoy the activities you like and don't have to deal with activities or places you are not comfortable with.