**Coursera Capstone**

IMB Applied Data Science Capstone

*Characterization of states and cities in the US*

**1. Introduction**

Moving to a new location (whether for a new job or for new business) can be a daunting task, especially if you do not know much about the different cities. In the US, there are about 20,000[[1]](#footnote-1) cities across 50 states, all unique in their own ways.

A city can be characterized in numerous ways, including but not limited to size, population, ethnic diversity, political orientation, types of venues and amenities available, etc. Here, we will explore these different aspects of cities and states in the US. Target audience would include anyone interested in learning about how cities and states in the US are similar or dissimilar.

**2. Data**

**2.1 Parler**

Following the “Capitol insurrection” on Wednesday, 1/6, tens of thousands of QAnon-linked accounts were banned on Twitter. With Donald Trump’s account banned from Twitter, Facebook, and Instagram, many of Trump’s followers migrated to Parler, the “free speech” platform known for hosting far-right content, their alternative. However, due to violent, racist posts on Parler, the app was soon removed from Google and Apple’s app stores, as well as refused hosting services from Amazon Web Services (among other cloud platforms), effectively erasing it from the internet. While this was welcomed by many, it also meant that months of posts that could be used to “understand” the attack on the Capitol would also be erased. Fortunately, all the data on Parler were accessible via public API, even posts and videos that were supposedly deleted. A self-described hacker by the Twitter handle donk\_enby downloaded more than 56.7 terabytes of data from Parler with the help of amateur data hoarders that volunteered their bandwidth.

Specifically, the data used here are GPS data of videos on Parler and were downloaded from <https://gist.github.com/kylemcdonald/8fdabd6526924012c1f5afe538d7dc09> on Jan 13, 2021. The downloaded csv contains coordinates for roughly 60,000 Parler videos and need to be annotated with state information (which can be done using geopy) in order for us to match with the election results. As such, the csv has been preprocessed to only include videos within the mainland US, annotated with the state the videos were taken in, the 2020 election result (Republican or Democratic party) for each state. The final csv consists of 54812 rows and 6 columns. The detailed processing steps can be found in the notebook.

**2.2 US states and cities**

Wikipedia contains list of states and their population (<https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population)>, 2020 election results for each state (<https://en.wikipedia.org/wiki/2020_United_States_elections)> , and list of largest cities in each state (<https://en.wikipedia.org/wiki/List_of_largest_cities_of_U.S._states_and_territories_by_population)>. We will use web scraping techniques to extract the data from the Wikipedia page.

**2.3 Foursquare**

We will use Foursquare API to retrieve different venues in each city or state.

**3. Methodology**

Geopy will be used to get coordinates of each location.

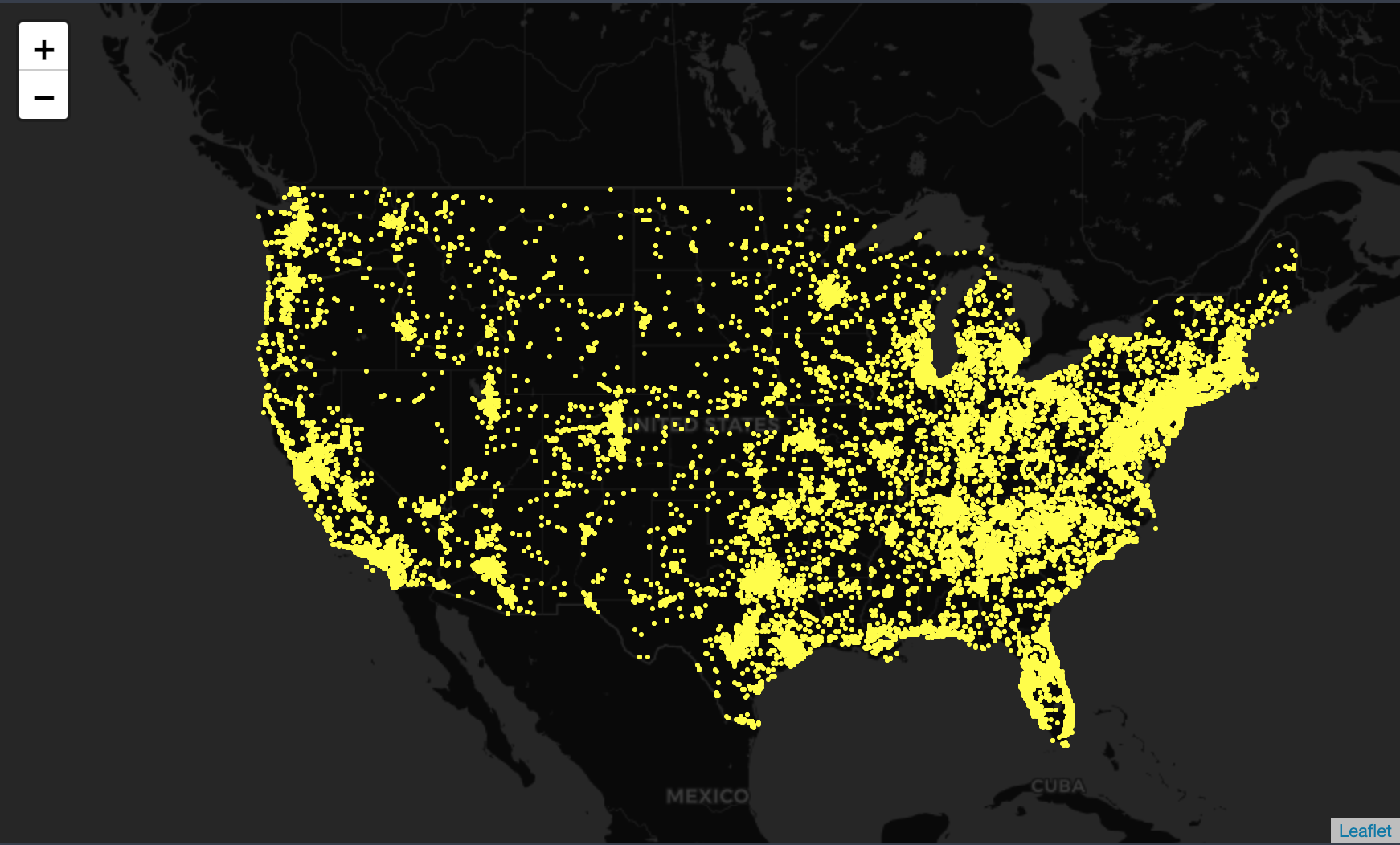
Folium will be used to map location data on the US map.

Foursquare API will be used to retrieve venues in the two cities we will compare.

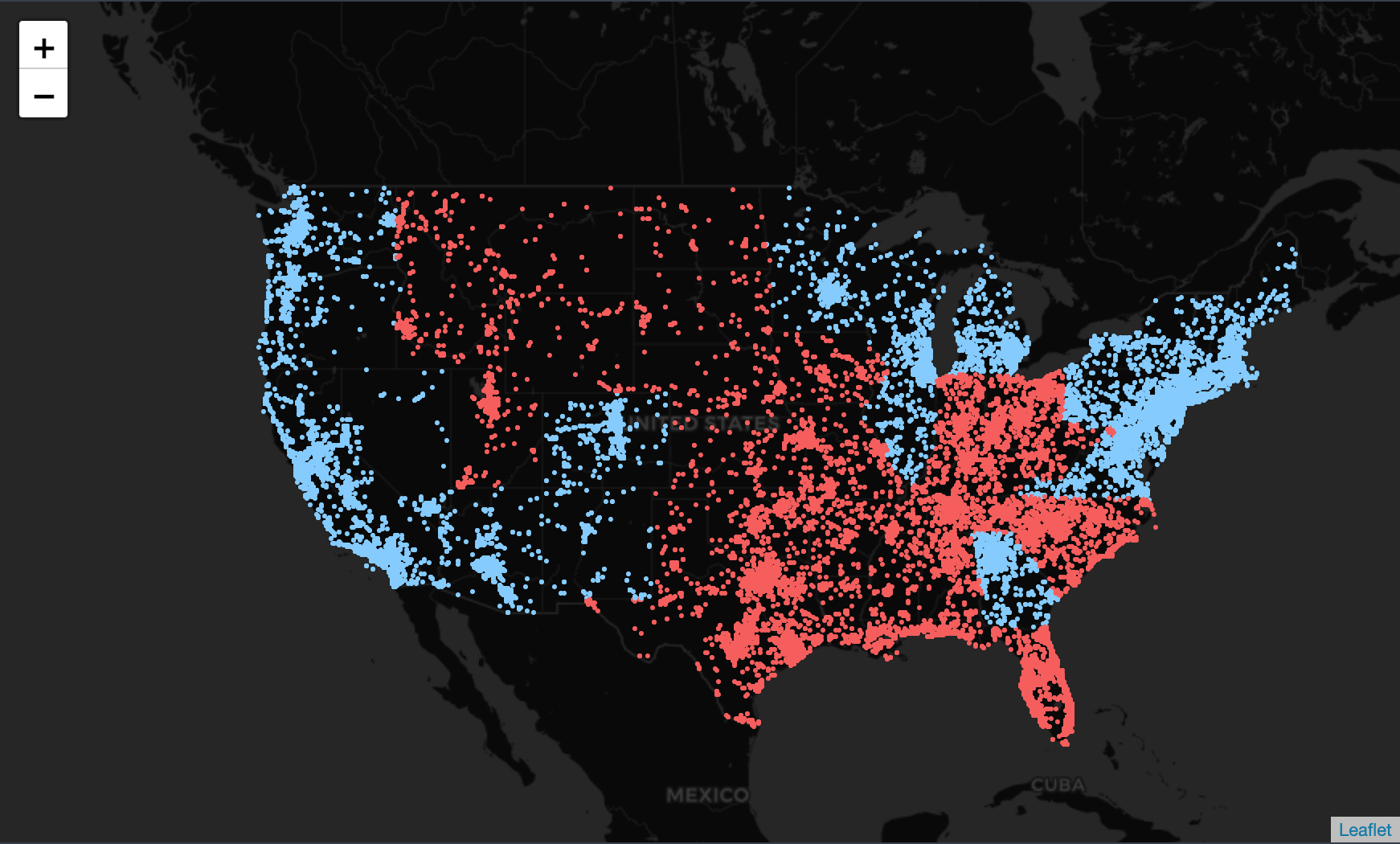
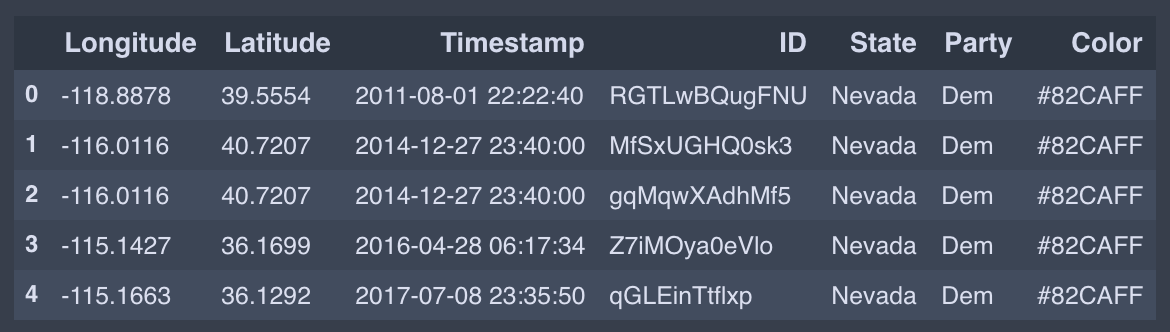
**4. Results**

**4.1 Political orientation of each state**

Below is the US map annotated with the location of each Parler video.

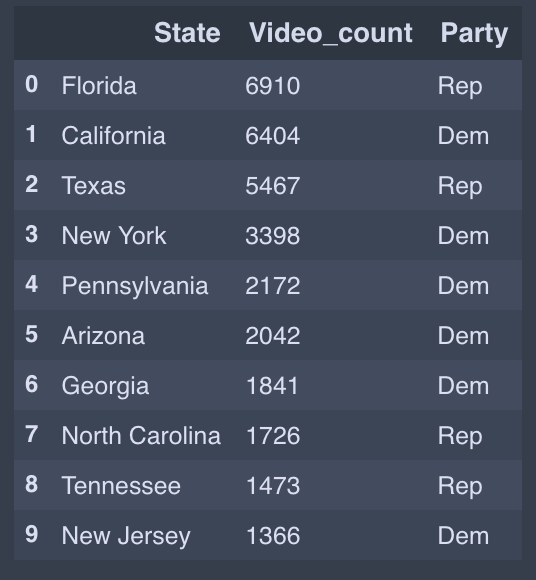


Then, each dot was colored to represent the political orientation of each state (based on the 2020 election result).



Surprisingly, there are a lot of far-right videos that were taken from blue states. In fact, among the top 10 states with the most videos, 6 were democratic states.

The top 10 states with Parler videos:



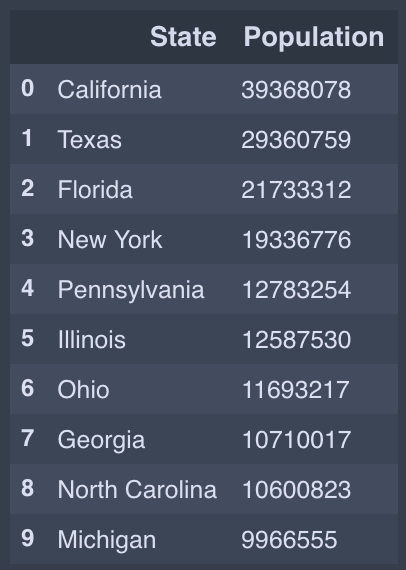
The top 10 states with Parler videos, normalized by population:



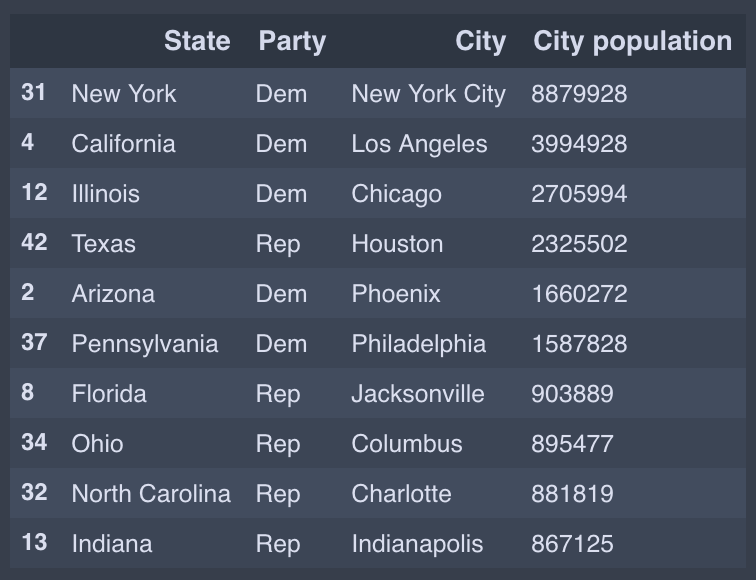
When normalized by population, however, we see that among the top 10 states, only two remain democratic states.

**4.2 Population of each state**

10 most populous states:

****

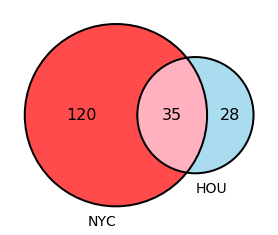
10 most populous cities:



Next, we will compare New York City (the most populous city in “Democratic” states) and Houston (the most populous city in “Republican” states).

**4.3 Common venue categories between NYC and Houston**

We will use Foursquare API to compare venues in NYC and in Houston. The five boroughs in NYC are Bronx, Brooklyn, Manhattan, Queens, and Staten Island. The five counties in Houston area are Harris County, Fort Bend County, Brazoria County, Montgomery County, and Galvestone County.



There are 155 unique venues within 2000 miles of the five boroughs in NYC, 63 unique venues within 2000 miles of the five counties in Houston. Only 35 venues overlap between the two cities.

**5. Discussion**

Surprisingly, there are a lot of far-right videos that were taken from blue states. In fact, among the top 10 states with the most videos, 6 were democratic states. When normalized by population, however, we see that among the top 10 states, only two remain democratic states. While NYC and Houston are two of the most populated cities, the two cities differ quite a bit in regard to the types of venues.

**6. Conclusion**

When considering a new location and its characteristics, it is important to gather all the information available and take into account many different aspects. For example, California’s been historically very blue, and you may move to California thinking it fits with your political view. In reality, however, you may encounter many neighbors that support far-right ideologies.

1. Statistica (https://www.statista.com/statistics/241695/number-of-us-cities-towns-villages-by-population-size/) [↑](#footnote-ref-1)