

2. Data

The region of our interest is the United States. In this work, three datasets have been used.

Dataset 1: The county-wise COVID-19 cases in the US

Source: <https://github.com/nytimes/covid-19-data/blob/master/us-counties.csv>

For the USA, New York Times has released and has been daily updating state-wise and county-wise cases of COVID-19 in [github](#) repository. The NY Times is compelling the record of registered cases due to this ongoing pandemic from local government, health department, and hospitals. The dataset is released in the public interest to better understand the outbreak. The county-wise dataset is considered for this analysis, which includes COVID-19 cases from 21st Jan to 18th April 2020. As listed in Table 1, this dataset consists of 6 features - date, county, state, FIPS id, COVID 19 cases, and number of deaths.

Table 1: County-wise COVID-19 cases in the US

	date	county	state	fips	cases	deaths
0	2020-01-21	Snohomish	Washington	53061.0	1	0
1	2020-01-22	Snohomish	Washington	53061.0	1	0
2	2020-01-23	Snohomish	Washington	53061.0	1	0
3	2020-01-24	Cook	Illinois	17031.0	1	0
4	2020-01-24	Snohomish	Washington	53061.0	1	0
...
70213	2020-04-18	Sublette	Wyoming	56035.0	1	0
70214	2020-04-18	Sweetwater	Wyoming	56037.0	10	0
70215	2020-04-18	Teton	Wyoming	56039.0	62	0
70216	2020-04-18	Uinta	Wyoming	56041.0	6	0
70217	2020-04-18	Washakie	Wyoming	56043.0	4	0
70218 rows × 6 columns						

The total number of counties collected for COVID-19 cases is 2704. As per 2016, there are 3007 counties in the US. It seems COVID-19 data is not available for nearly 300 counties. Therefore, these counties have been excluded from this study. There are 590123 cases are reported by April 18th.

Dataset 2: County-wise geo JSON data file for visualization

Source: https://raw.githubusercontent.com/python-visualization/folium/master/examples/data/us_counties_20m_topo.json

The visualization of counties in the US has been carried out using **folium** library with 'county_geo JSON' data file. For the analysis and visualization, these counties are characterized into the following 5 categories as shown in Figure 1:

1. *Black county*: No COVID-19 data is available
2. *Safe county*: COVID-19 cases are less than or equal to 10
3. *Yellow county*: Number of cases greater than 10 and less than or equal to 1000
4. *Orange county*: Number of cases greater than 1000 and less than or equal to 10000
5. *Red county*: Number of cases greater than 10000

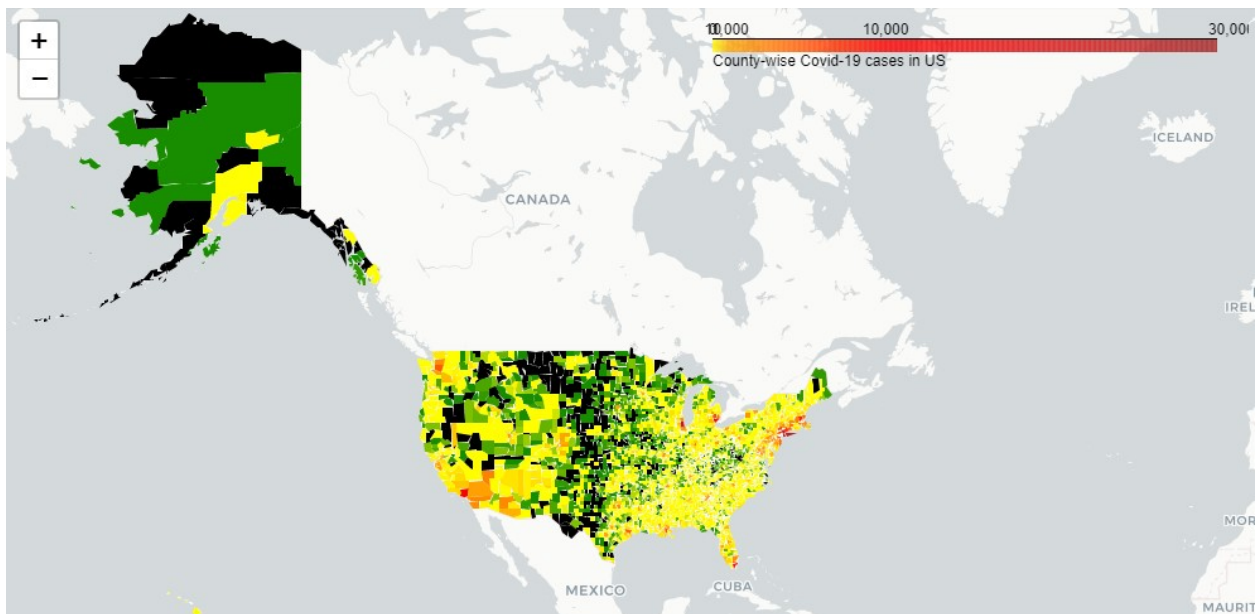


Figure 1: County-wise distribution of COVID-19 cases in the US.

Dataset 3: Starbuck venue location data using Foursquare API

Source: <https://developer.foursquare.com/places>

Foursquare is a very popular location-based networking website. It has been used by several users to explore nearby places and also been actively using venue-based data analyses. In this work, the venue locations (latitude and longitude) of Starbucks stores within the radius of 15 km of each safe county are obtained using Foursquare APIs. Since the free account of Foursquare is used in this work, the number of venues for a specific query (i.e. Starbuck) for

each county is limited (max 50). The number of calls to be made in a day is also limited. Thus the program has to be run multiple times since the number of safe counties is 1177, and these results merged into a new data frame.

Furthermore, it is important to note that only one location request is made for each county despite the size of the county. Therefore the number of Starbucks stores in each county may be higher than the numbers represented in this work. The accuracy of these numbers can be improved by using a grid-based location search approach for each county; however, this requires a significantly higher number of calls to make in Foursquare API. This can be done by using a premium membership account with Foursquare. For this analysis, a small group of Starbucks venue points is sufficient to get insights into which counties are allowed to reopen the businesses.

For a few counties, some of the Starbucks venues from the Foursquare are located in neighboring counties as for 598, 599 and 600 cases shown in Table 2. This problem has solved by obtaining and verifying the FIPS codes of those Starbucks's latitude and longitude coordinates with the FIPS codes from the COVID-19 data set. The results have been verified to make sure that each one of Starbuck venues is located into corresponding county used for Foursquare API.

Table 2: Sample data of Starbucks venue location from Foursquare

	Requested county	Requested FIPS	Category ID	Venue name	Venue latitude	Venue longitude	FIPS	State
0	Bullock	1011	5793e354498e922ae570bdf3	Starbucks	32.014877	-85.746056	1011	Alabama
1	Graham	4009	4fc9457fd4f24895b4467c83	Starbucks	32.835233	-109.734477	4009	Arizona
2	Graham	4009	4c8fd1c590ab1f7ac93e27d	Starbucks In Safeway	32.835672	-109.734042	4009	Arizona
3	Baxter	5005	5637f89dcd104b1868f1d572	Starbucks	36.349273	-92.371507	5005	Arkansas
4	Baxter	5005	5637f82ecd1099bc2507a5e5	Starbucks	36.347948	-92.371883	5005	Arkansas
...
598	Vilas	55125	549db741498ea7de5014f63e	Starbucks	43.075622	-89.386640	55025	Wisconsin
599	Vilas	55125	5a4a7a59bcbf7a68d9d7e169	Starbucks	43.025453	-89.417466	55025	Wisconsin
600	Vilas	55125	4cc96d214650a35ddb358e1e	Starbucks (inside The Sheraton)	43.047460	-89.373318	55025	Wisconsin
601	Washburn	55129	4b8187f2f964a520deac30e3	starbucks	45.896496	-91.827999	55129	Wisconsin
602	Wood	55141	57fd457b498e08976645bd16	Starbucks	44.017366	-90.508795	55081	Wisconsin