

Capstone Project

The Battle of Neighborhoods

By Xueling Li

Introduction

- This project aims to find out the current status of the spread of covid-19 in the United States, particularly in the State of Michigan.
- As of today (March 31st, 2020), the State of Michigan, which is neither a populated area nor a multicultural center, are reported to have the fourth most coronavirus cases in the nation after New York, New Jersey and California.
- We are interested in finding out the reasons behind this striking phenomenon.

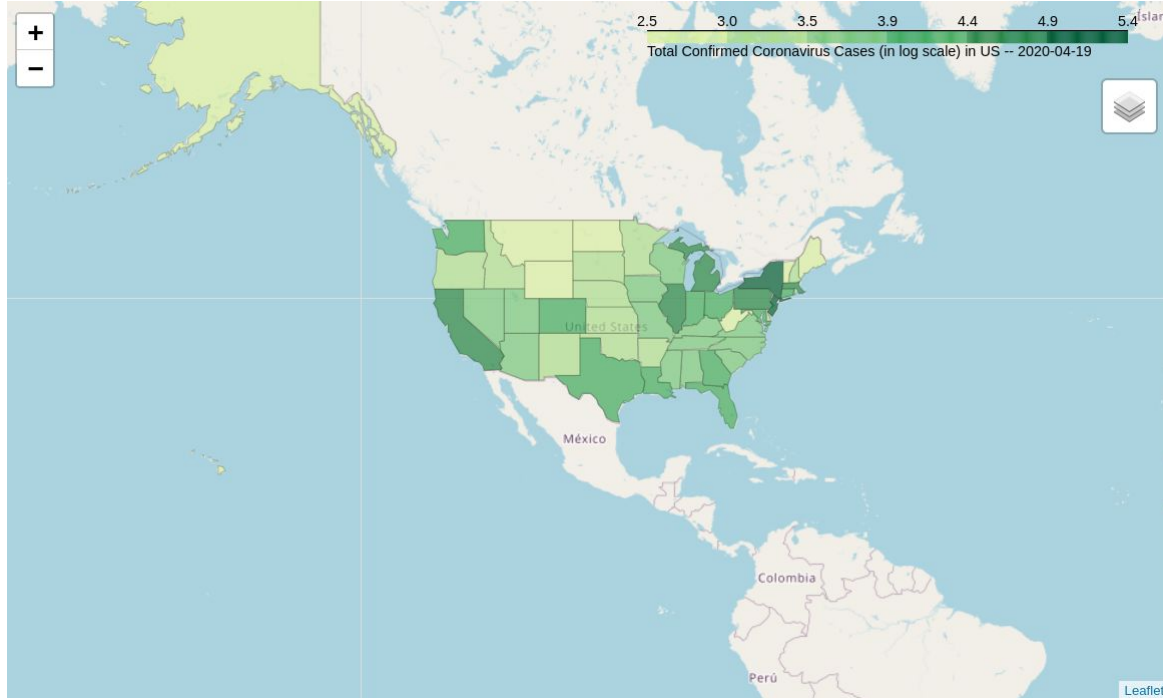
Data Sources

- Covid-19 world map from [Centers for Disease Control and Preventions \(CDC\)](#)
- Real time coronavirus status for United States from [Worldometers](#)
- Real time coronavirus status for State of Michigan from [Wikipedia](#)
- Venue information for each state in United States and each county in State of Michigan from [Foursquare API](#)

Part I: The Spread of Coronavirus in United State

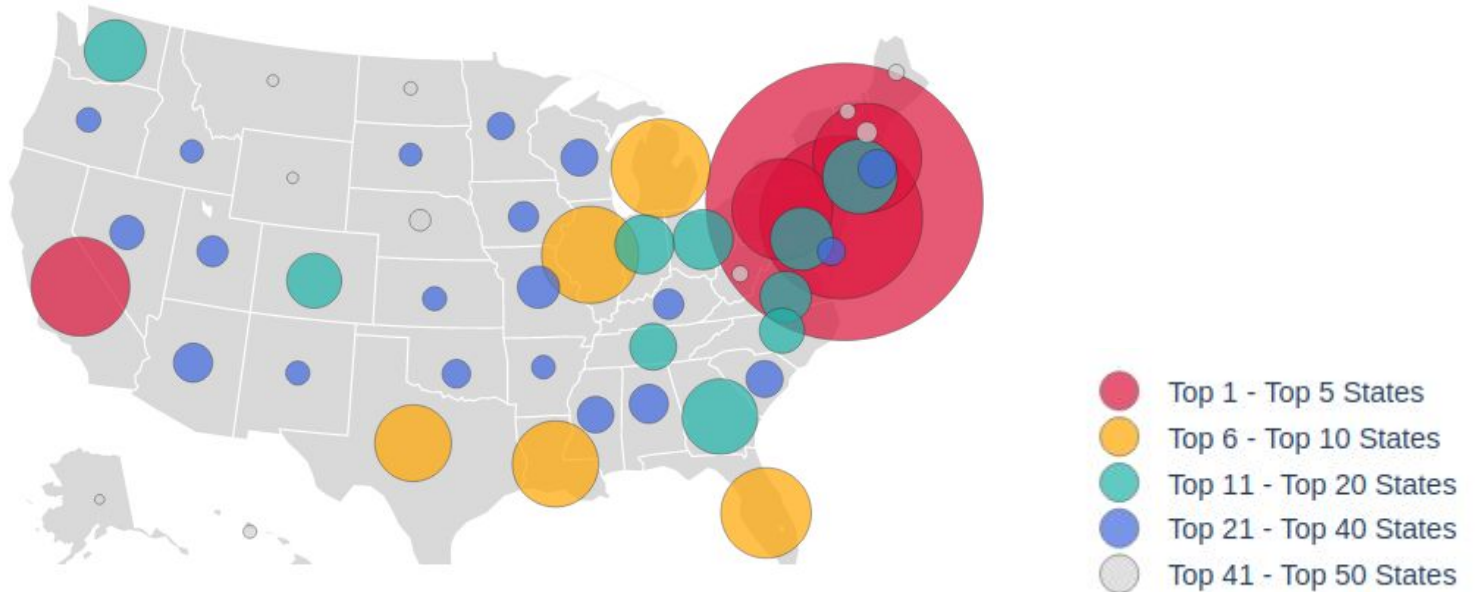


Total Confirmed Covid-19 Cases in US



- As of April, 19th, 2020, the total confirmed cases have reached above 700,000.
- Most confirmed cases occur in New York, California, Michigan, Texas, Illinois, New Jersey, and Pennsylvania.
- Michigan has 31,424 total cases, rank #5 in the United States

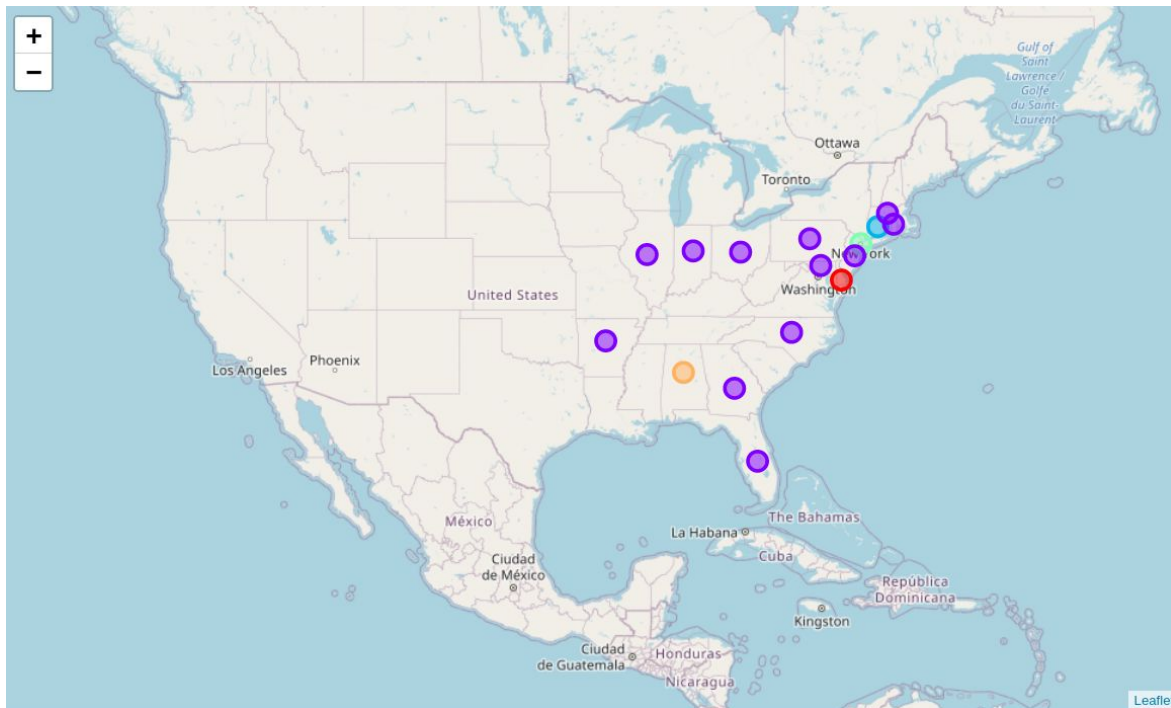
Total Confirmed Cases in US (Categorized in Groups)



Top 10 Venues for Each State in US

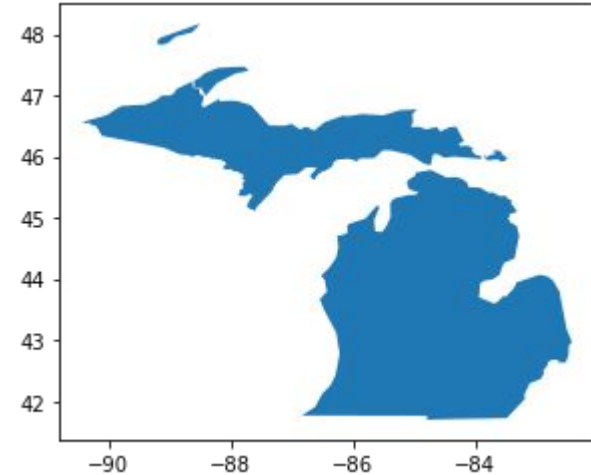
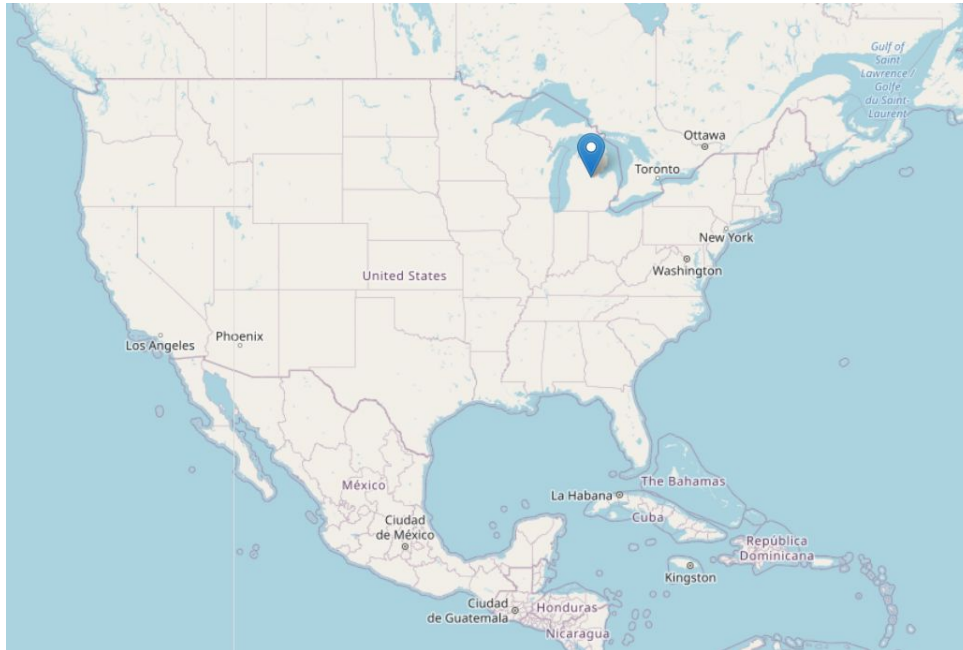
	State	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alabama	Coffee Shop	Women's Store	High School	Restaurant	Business Service	Campground	Baseball Field	Community College	Construction & Landscaping	Convenience Store
1	Arkansas	Construction & Landscaping	Discount Store	Women's Store	Diner	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop
2	Connecticut	Gas Station	Toy / Game Store	Supermarket	Sports Bar	Warehouse Store	Sandwich Place	Restaurant	Coffee Shop	Mexican Restaurant	Liquor Store
3	Delaware	American Restaurant	Diner	Gas Station	Fast Food Restaurant	Donut Shop	Community College	Discount Store	Convenience Store	Mobile Phone Shop	Pizza Place
4	Florida	Trail	Women's Store	Dance Studio	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store	Diner
5	Georgia	Business Service	Women's Store	Diner	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store
6	Illinois	Scenic Lookout	Women's Store	Dance Studio	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store	Diner
7	Indiana	Baseball Field	Women's Store	Diner	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store
8	Maryland	Home Service	Campground	Diner	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store
9	Massachusetts	Trail	River	Women's Store	Dance Studio	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store
10	New Jersey	Lawyer	Women's Store	High School	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store	Diner
11	New York	Dance Studio	Sandwich Place	French Restaurant	Hotel	Hotel Bar	Laundry Service	Park	Gym / Fitness Center	Coffee Shop	Bakery
12	North Carolina	Hotel	High School	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store	Diner
13	Ohio	Business Service	Stables	Women's Store	Dance Studio	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store
14	Pennsylvania	River	Women's Store	Dance Studio	Gas Station	French Restaurant	Fast Food Restaurant	Event Service	Donut Shop	Discount Store	Diner
15	Rhode Island	Home Service	Event Service	Warehouse Store	Gym / Fitness Center	Gas Station	French Restaurant	Fast Food Restaurant	Donut Shop	Discount Store	Diner

Clustering Results in US

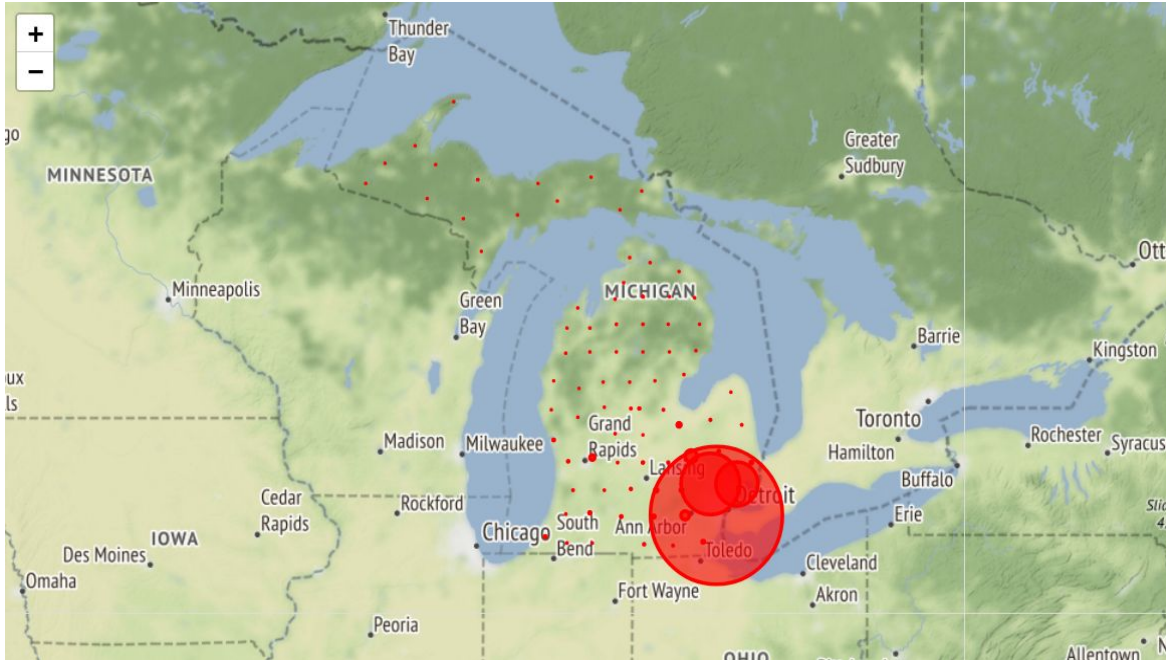


- All venues retrieved are located in East coast or central part of the US.
- Almost all States belong to 'cluster 1' (colored in purple)
- The rest 4 states each belongs to its own cluster.
- The states in 'cluster 1' are close to East coast.

Part II: The Spread of Coronavirus in State of Michigan, United State

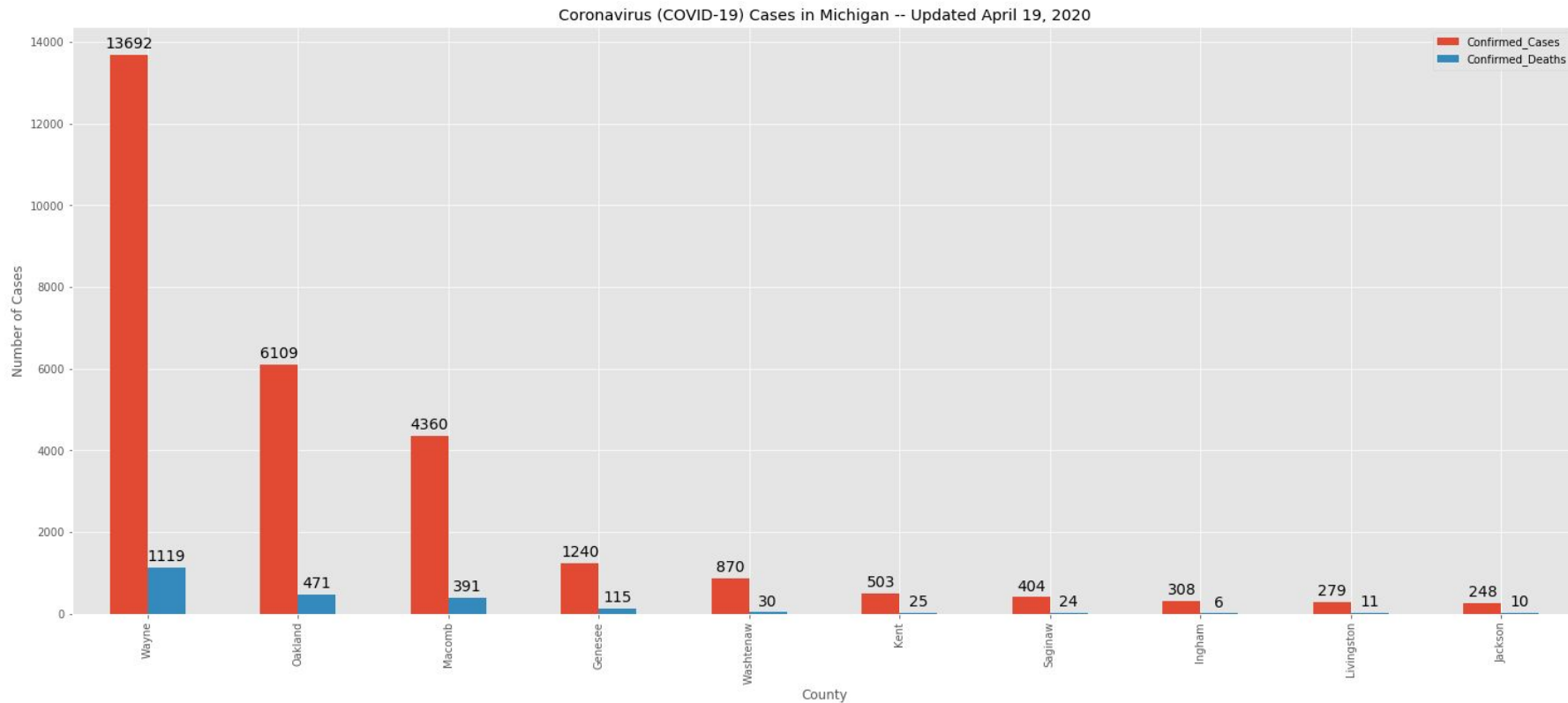


Total Confirmed Covid-19 Cases in Michigan

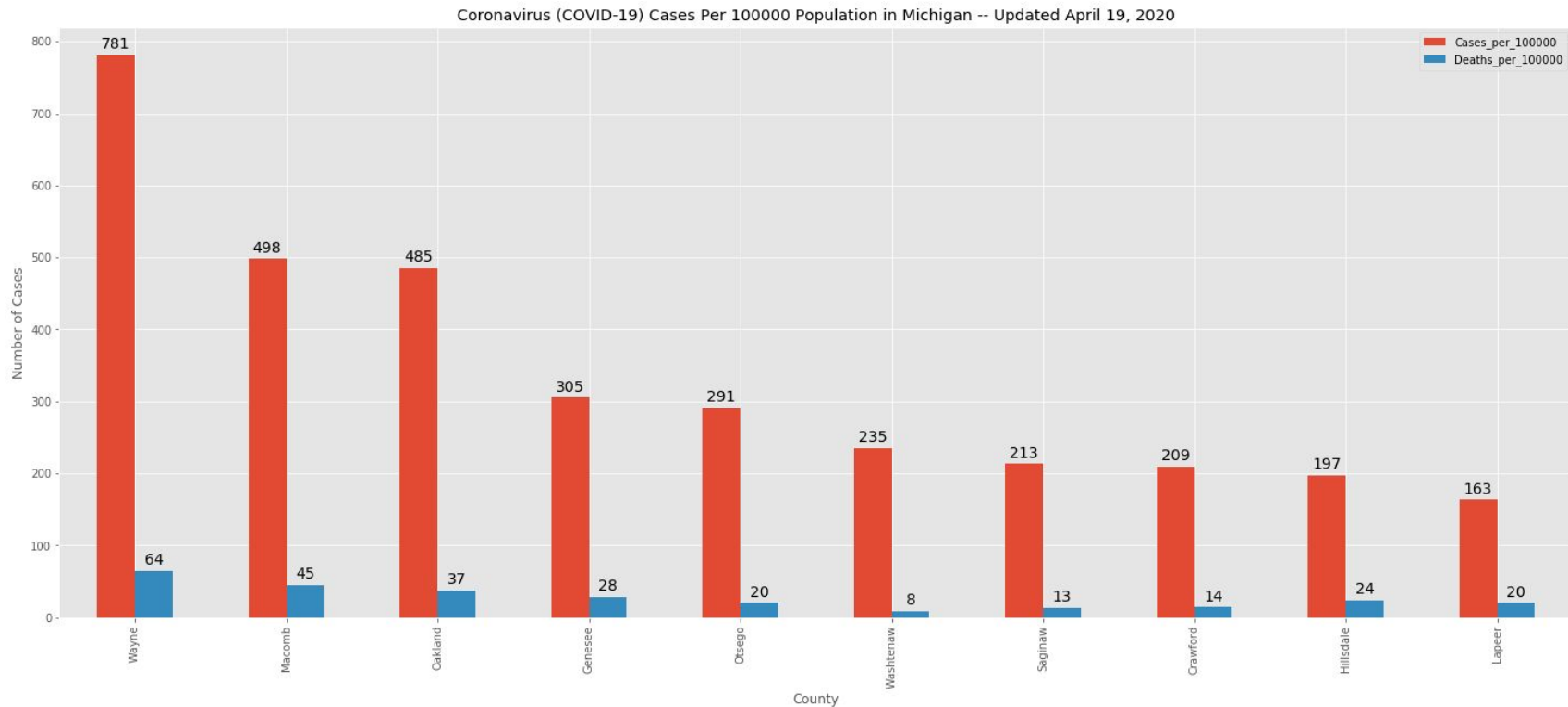


- Most confirmed cases occur in the Southeast region of Michigan
- The following 3 counties (Wayne, Oakland, Macomb) dominate the spread of the virus in the state

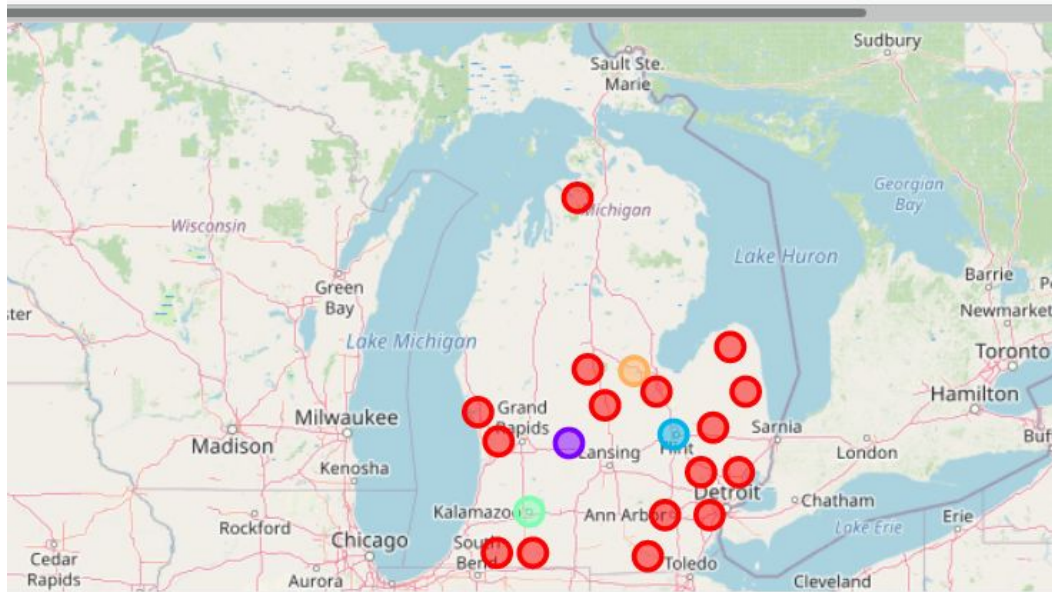
Top 10 Counties (by direct count)



Top 10 Counties (by county average)



Clustering Results in Michigan



- All venues are located in down state of Michigan.
- Almost all counties belong to 'cluster 1' (colored in red), the rest 4 counties each belongs to its own cluster.
- The counties in 'cluster 1' are relatively populated regions in the State of Michigan and close to Metro Detroit.

Discussion

- It is striking to find out that Michigan ranks top 5 for the total cases in the United States. In the venue extraction for the US, there is no venue returned for Michigan. This could be due to the fact that the coordinates of the center of the State was used, where no venues are found to be surrounding it.
- This study focuses on the techniques that were covered in the course specialization (e.g., scrape information from web, data visualization, and K-mean clustering). These methods may or may not fit best to our purposes, but it is definitely a great opportunity to practice through this capstone project.

Discussion Cont.

- Extracting GPS coordinates information purely based on state and county names could result in mistakes. (e.g., Washington and Washington State are confusing to the Python package “Nominatim” which was used for extracting GPS coordinates)
- The results for the data visualization, data analysis depend on parameters used for the analysis. (e.g., color bar settings, radius for venue extraction, and K value in the K-means clustering).