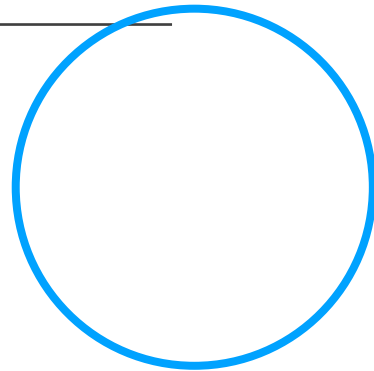# IBM Applied Data Science Capstone
(IBM Professional Data Science Certificate)

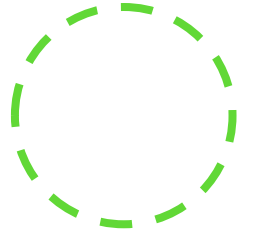ANALYSING COVID-19 IN THE US BY STATE

BUSINESS PROBLEM

- The objective of this capstone project is to analyze the state-wise cases of COVID-19 in the US to cluster them into zones, such that high-frequency zones can be identified and medical camps can be setup.

- Using data science methodology and machine learning techniques such as data wrangling and K-Means clustering, this project aims to answer the question:

Which are the spots where setting up medical camps will be most efficient in the United States of America?

- This model and analysis can be leveraged to many other conclusions than the one presented in this project.
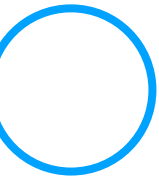
DATA

- Data Required
  ◦ List of states in the US
  ◦ Latitude and Longitude of each US state
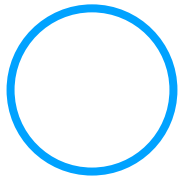 ◦ Number of active COVID-19 cases in each of the states

- Sources of Data
  ◦ Kaggle – Novel Coronavirus-19 Dataset
  ◦ Wikipedia – List of US States
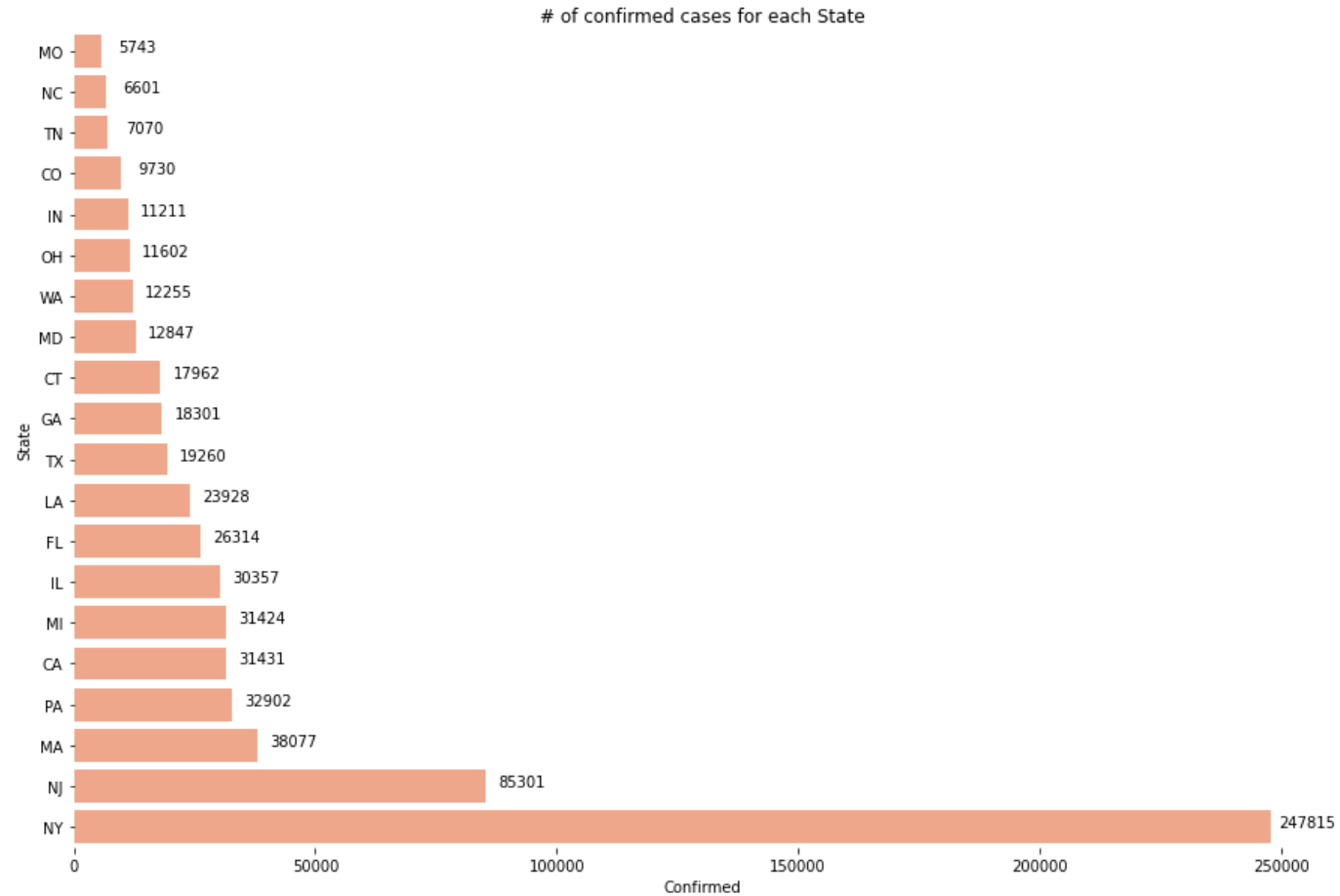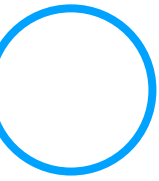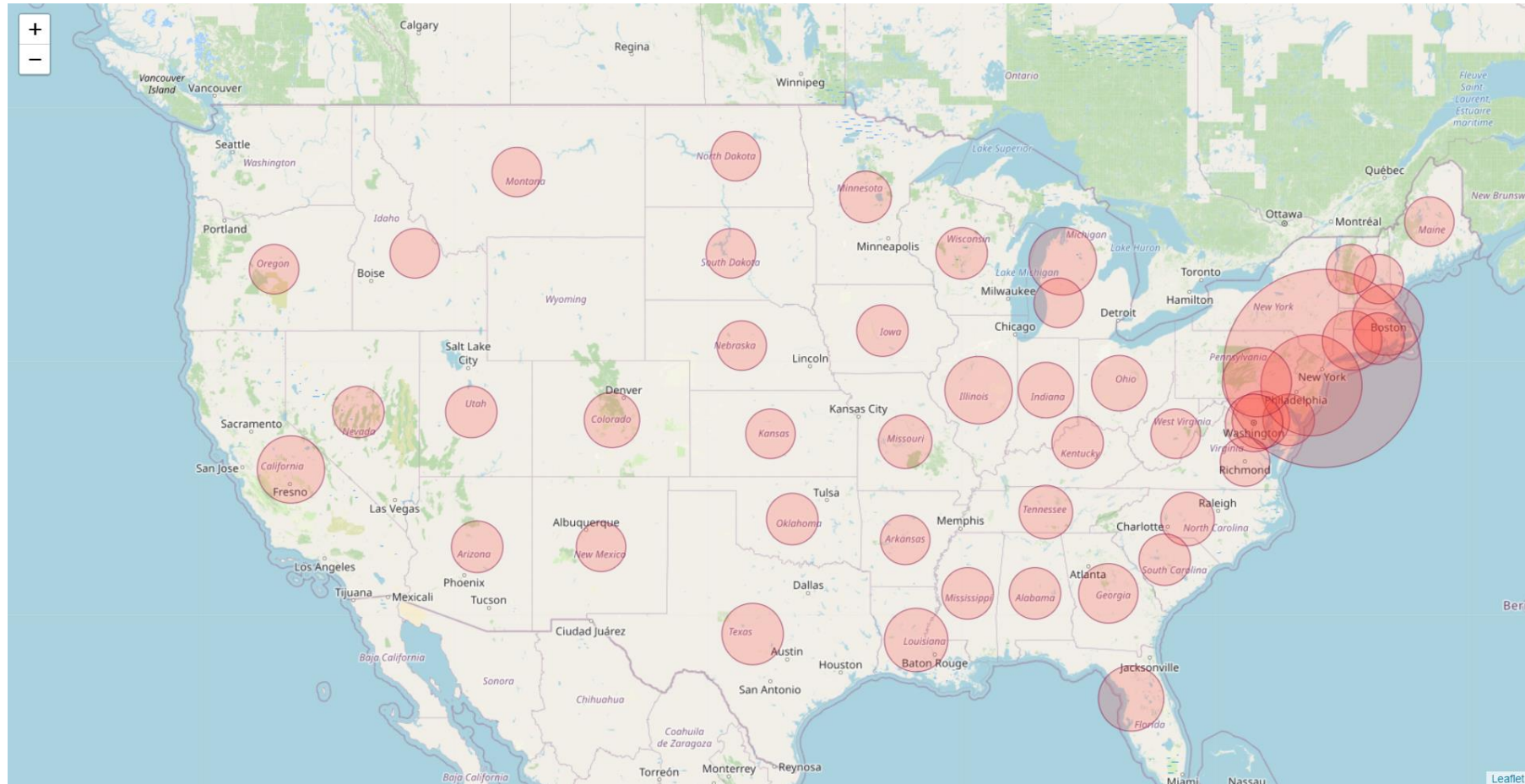 ◦ Foursquare API for Latitude and Longitude data

METHODOLOGY

- Reading, cleaning and wrangling the dataset using Pandas python library.

- Scraping the list of US States from Wikipedia using BeautifulSoup python library.

- Mapping the number of cases from each of the entries in the dataframe to each of the state of the US

- Exploratory analysis of the data acquired.

- Collecting Latitude & Longitude data for each of the states using Foursquare API.

- Clustering the number of cases of COVID-19 for each state and determining the optimal spots to set up medical camps & "red zones."

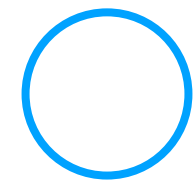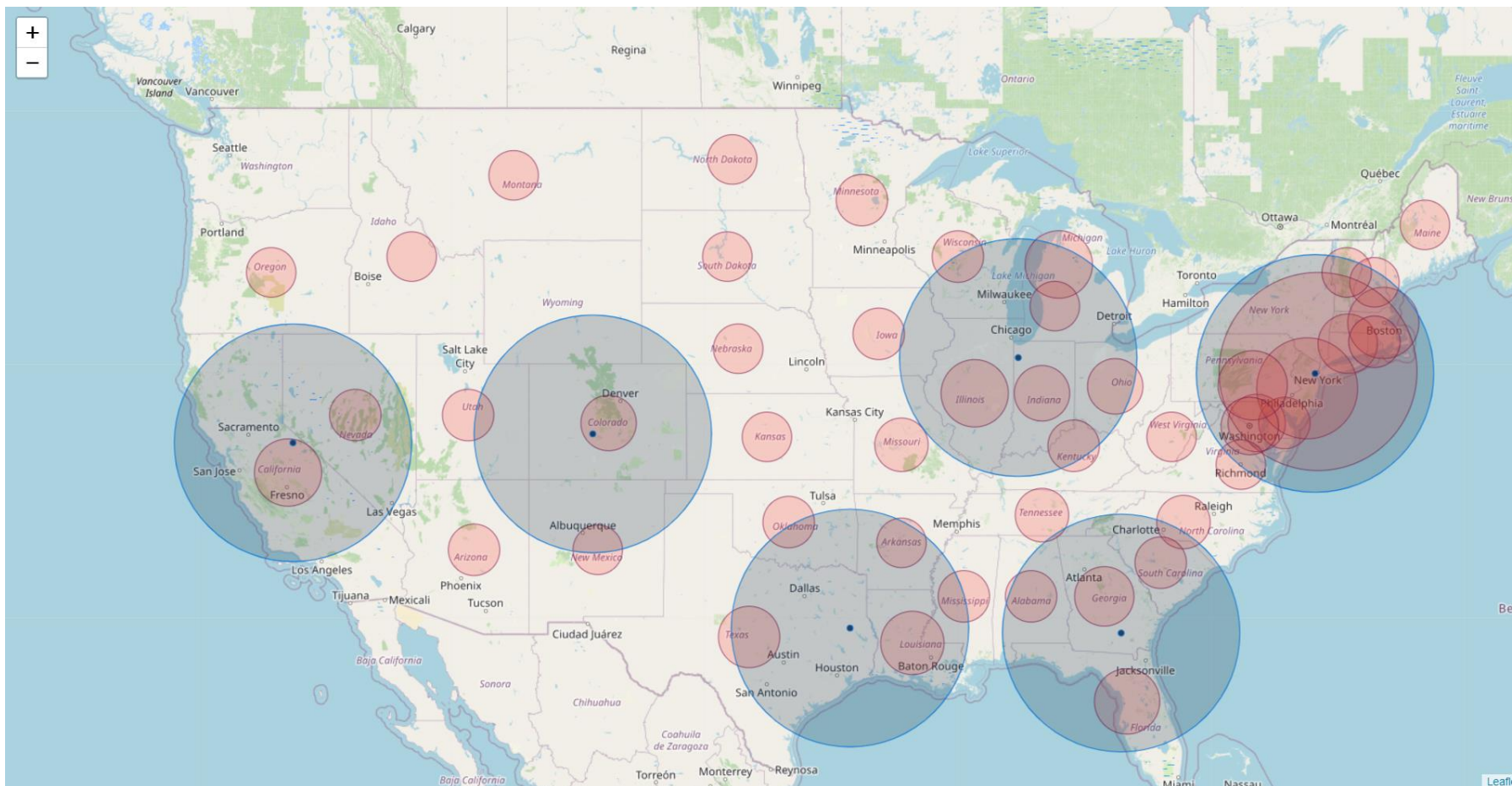# RESULTS

# of confirmed cases for each State
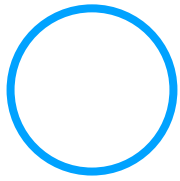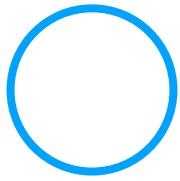
RESULTS

RESULTS

DISCUSSION & RECOMMENDATIONS

- Most of the COVID-19 cases are focussed on the east coast of the US. This could be due to the high influx of foreign population to the states of the east coast due to tourist attractions.

- Highest number of COVID-19 cases are from NY. This is explained by the high population density and high tourist influx. Therefore, to alleviate the current situation, two medical camps and "red zones" should be set up in NY.

- Other states to set up medical camps - **Indiana**, **Atlanta**, **Texas**, **Denver** and **California.**

- Following the suggested precautionary measures as mandated by WHO and by treating and preventing the spread of current cases will lead to humanity successfully fighting against the pandemic.

CONCLUSION

- Answer to the business question:

  ◦ **Which are the spots where setting up medical camps will be most efficient in the United States of America?**

  ◦ Setting up medical camps and "red zones" in the states mentioned in the previous slide will be the most efficient method of fighting against COVID-19.

- Moreover, if there are funds to set up more (k > 7) medical camps, this model can fit again and a new **k** spots can be found out where medical camps can be setup.

THANK YOU!