APPLIED DATA SCIENCE CAPSTONE FINAL PROJECT REPORT

USING VENUE INFORMATION TO CLUSTER ZIP CODES AREAS IN THE CITY OF CHICAGO

Julio Guerrero

agust 2021

Objectives

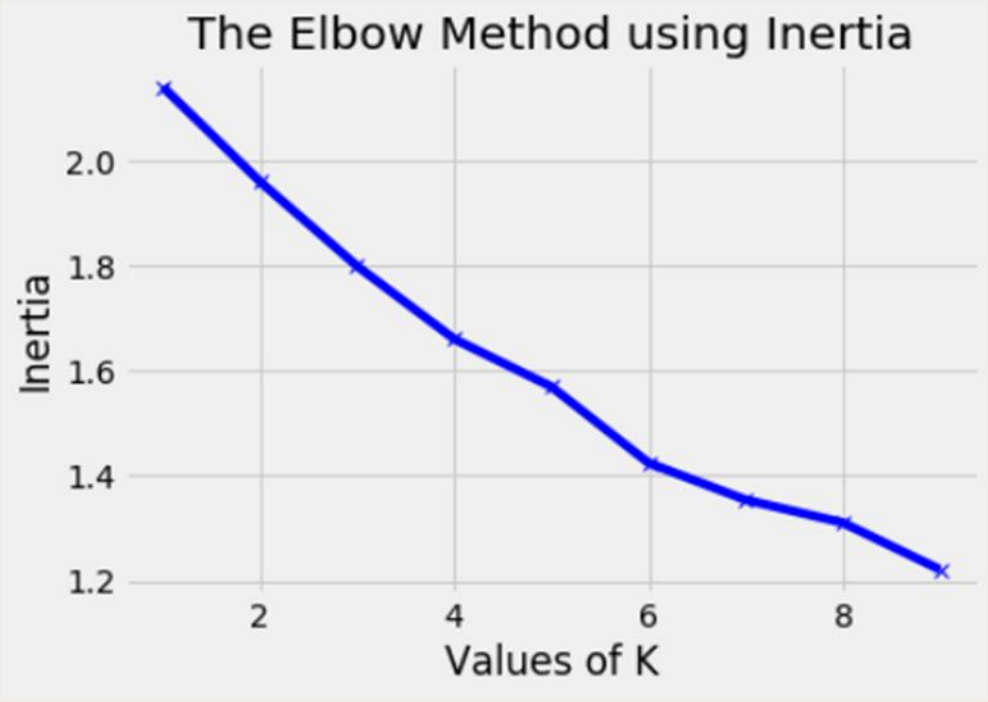
- Cluster the zip codes areas of the city of Chicago based on their nearby venues.
- Make a map to show the distribution of the percentage of people fully vaccinated in the city
- Show the locations of each cluster in the city

Data

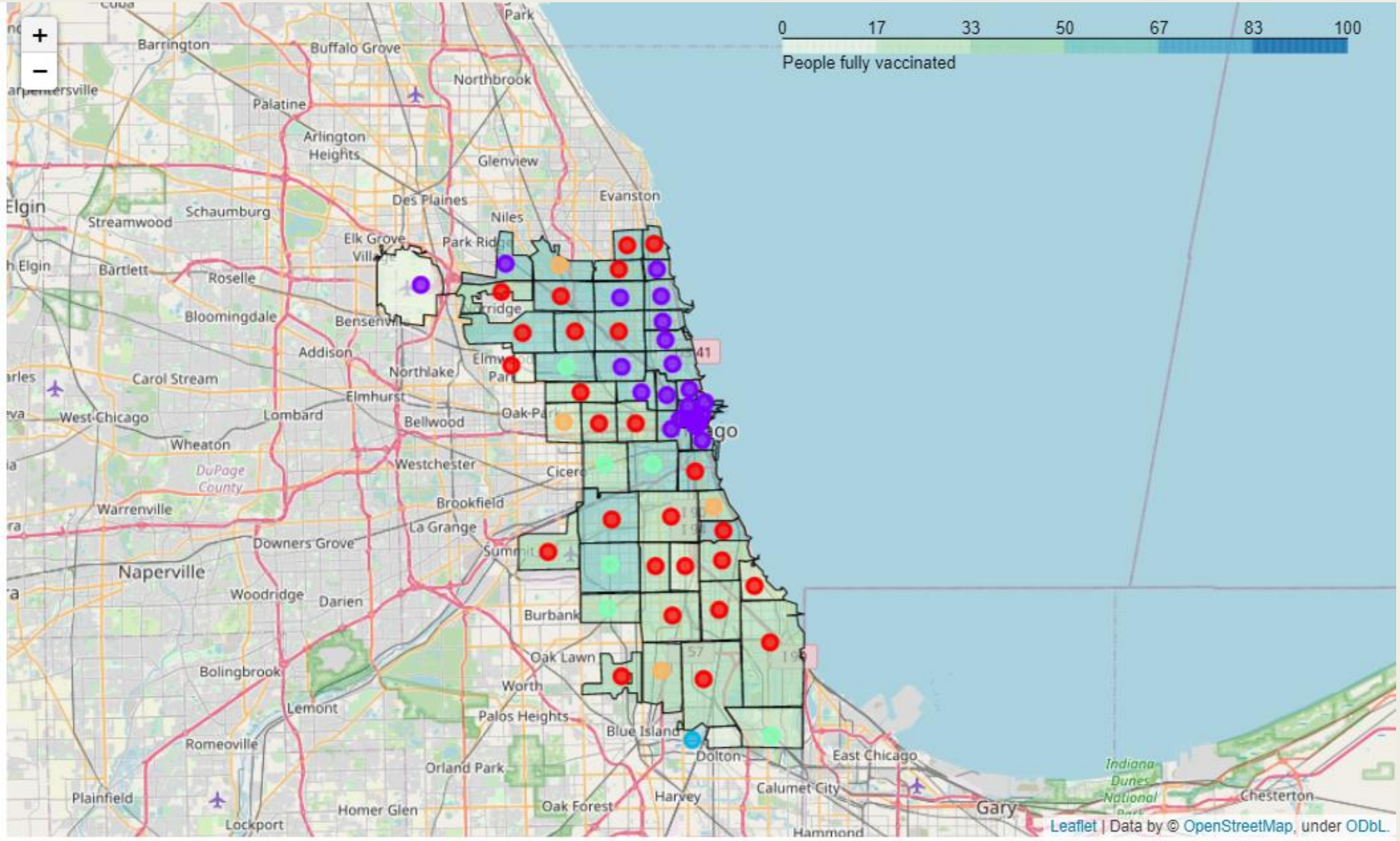
1. A GeoJSON file with the Zip Codes boundaries of the City of Chicago. This will be used for visualization of the data collected.(*)
2. A Dataset of the City of Chicago Data Portal call “COVID-19 Vaccine Doses by ZIP Code - Series Completed”. We are interested in the percentage of the population fully vaccinated by zip code and the coordinates of the area. (*)
3. We will use the Foursquare API to get the most common venues of each Zip Code.

Methodology

- We used K means for clustering the zip codes areas
- We used the elbow method to choose the number of clusters



Results


0
17
33
50
67
83
100
People fully vaccinated
Barrington
Buffalo Grove
Palatine
Northbrook
Arlington Heights
Glenview
Des Plaines
Evanston
Niles
Elgin
Streamwood
Schaumburg
Bartlett
Roselle
Bloomingdale
Addison
Carol Stream
Elmhurst
Northlake
West Chicago
Lombard
Bellwood
Wheaton
DuPage County
Westchester
Cicero
Brookfield
La Grange
Warrenville
Downers Grove
Naperville
Woodridge
Darien
Burbank
Oak Lawn
Bolingbrook
Worth
Lemont
Palos Heights
Blue Island
Romeoville
Orland Park
Dolton
East Chicago
Plainfield
Homer Glen
Oak Forest
Harvey
Calumet City
Gary
Chesterton
Indiana Dunes National
Hammond
Leaflet | Data by © OpenStreetMap, under ODbL.

Conclusions

- We were able to group each Zip Code of the city in 5 clusters based on their venues. With that information, we made a couple of data frames that can tell future investors which are the most common venues and their frequency. We also created a Map to give an idea of the distribution of our clusters and the percentages of people fully vaccinated in the city of Chicago.