# The Battle of Neighborhoods | Data Description

## Data sources

1.

2021 Medicare Advantage ratebook and Prescription Drug rate information

<https://www.cms.gov/medicarehealth-plansmedicareadvtgspecratestatsratebooks-and-supporting-data/2021>

2.

MA State/County Penetration

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/MA-State-County-Penetration>

3.

Medicare Enrollment Dashboard

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMSProgramStatistics/Dashboard>

4.

Inpatient Charge Data FY 2017

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient2017>

5.

<https://www.aha.org/system/files/2018-07/2018-aha-chartbook.pdf>

6.

Fast Facts on U.S. Hospitals, 2020

<https://www.aha.org/statistics/fast-facts-us-hospitals>

7.

<https://www.census.gov/topics/health/data.html>

8.

Population density dataset:

<https://web.archive.org/web/20150807220054/http://quickfacts.census.gov/qfd/download_data.html>

9.

Population distribution by age

<https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>

10.

<https://blog.bismart.com/en/10-aging-population-statistics-big-data>

## Work Flow

To choose the best location, I decided to explore every Medical center, segment them and group them into clusters to find similar medical center in the USA. To be able to do that, we need to use Python web scrawling to crawl locations both XXX healthcare and its competitors. Then, we need to cluster data which is a form of unsupervised learning: k-means clustering algorithm.

## Libraries Which are Used to Develope the Project:

Numpy

Panas

Beautiful Soup

Matplotlib

SK-Learn

Geocoder

JSON

XML

Request