A PROJECT REPORT ON

CLUSTERING APPROACH TO DATA MINING FOR IDENTIFYING LOCALITIES IN PUNE,INDIA WHERE THE HEALTH CARE INFRASTRUCTURE NEEDS TO BE IMPROVED

SUBMITTED BY

Name: VEDANT R. ASAWALE
(Data Science Enthusiast)

Under the guidance of

Alex Aklson

IBM DATA SCIENCE PROFESSIONAL CERTIFICATION

Coursera Learning Platform

ACKNOWLEDGEMENT

I sincerely thank Coursera and IBM for providing students a platform to learn new skills and enhance their knowledge.

I also sincerely convey my gratitude to the course teacher Alex Aklson and all other teachers for their support, providing all the help, motivation and knowledge from beginning till end to make this learning path a grand success.

Above all I would like to thank my parents for their wonderful support and blessings, without which I would not have been able to accomplish any of my goals.

Contents

1 INTRODUCTION **1**
- 1.1 Lack of Sufficient Health Care Infrastructure . . . 1
- 1.2 Improving Health Care Infrastructure . . . 1

2 PROBLEM STATEMENT **2**
- 2.1 Statement . . . 2
- 2.2 Target Audience . . . 2

3 MOTIVATION **3**

4 DATA **4**
- 4.1 Importance of Data . . . 4
- 4.2 Data used in the Project . . . 4
  - 4.2.1 Finding Localities in Pune . . . 4
  - 4.2.2 Getting the Geographical Coordinates . . . 5
  - 4.2.3 Fetching Hospital data . . . 5

5 METHODOLOGY **7**
- 5.1 Steps performed . . . 7
- 5.2 Importing relevant libraries . . . 7
- 5.3 Web-Scraping . . . 8
- 5.4 Extracting Geographical coordinates . . . 9
- 5.5 Preliminary Visualization . . . 10
- 5.6 Fetching Location Data . . . 11
- 5.7 Extracting Relevant Feature . . . 12
- 5.8 Finding the optimal number of Clusters . . . 13
- 5.9 Clustering . . . 14
- 5.10 Final Visualization . . . 15
- 5.11 Find clusters with highest need for new Health Care infrastructure . 15

6 RESULTS **17**

7 DISCUSSION **19**

8 FUTURE SCOPE **20**

9 CONCLUSION **21**

1 INTRODUCTION

1.1 Lack of Sufficient Health Care Infrastructure

The Corona virus outbreak in 2020 has exposed the lack of sufficient health care infrastructure throughout the World. Even the most developed economies such as the United States and Italy could not handle the unprecedented crisis.The virus outbreak not only lead to numerous deaths but also resulted in huge financial losses.Moreover it has given us a wake up call regarding how under prepared the humanity is for such kind of a crisis.There were situations wherein the Hospitals were completely filled and unfortunately no more people could be admitted for treatment eventually resulting in loss of human lives.

1.2 Improving Health Care Infrastructure

.

The society should take lessons from this crisis and prepare ourselves for any such disaster beforehand.This project will helps us in improving the Health care facilities by identifying localities where the number of hospitals are less so that a number of new hospitals can be constructed in these regions.In this project, analysis is done for regions from one city (Pune,India) only.However,the same analysis can be carried out for different regions as long as we can gather the right data.

2 PROBLEM STATEMENT

2.1 Statement

Identify the localities or regions in Pune city,India where the number of hospitals are less and hence there is urgent need to improve the Health care infrastructure to prepare for any epidemic or disaster crisis which might arise in future.

2.2 Target Audience

This project can be used by the Government or social workers to identify regions for creation of new Health Care facilities.

3 MOTIVATION

The motivation for creating such a project is driven by two main factors:
1.The damage caused by Covid-19.
2.As a Data science enthusiast its my responsibility to solve problems and make human life better using data.

The data being generated by various sources has increased multi fold over the past few years. Thus it has given rise to a completely new domain called Data Science which involves understanding and analyzing data to solve everyday as well as major problems. This technique can be used in various domains such as Business,Online Marketing,Spam Detection and many other.

The problems caused by Covid-19 were disastrous and we need to find ways to make sure that we are prepared for any such crisis.We can use data science in many different ways to do so and this project is just one such attempt.

4 DATA

4.1 Importance of Data

A model or any analysis is as good as the data which means that data is the prime and most important component of any analysis.The result produced wouldn't be on expected lines even by using the best of techniques and algorithms if the data provided is not accurate.Identifying the best sources of data,collecting it,extracting it and transforming it as per our needs are the major and highly important steps associated with Data Science.The data as well as the sources of data have increased exponentially over the years.Thus,for leveraging the value of this data,its necessary to select the right sources and thereafter select the right features for better analysis.

4.2 Data used in the Project

The current project identifies localities in the Pune city,India where the Health Care facilities are not adequate.

Thus the following data is needed for our analysis:-

1.List of all the localities in Pune city.

2.The geographical coordinates for all these localities.

3.The number of hospitals present in each locality.

4.2.1 Finding Localities in Pune

List of all localities present in Pune is readily available on the Internet.
The URL "https://www.mapsofindia.com/pune/localities/" is used to get the list of all localities in Pune using the technique of web scraping via the Beautiful Soup API.

Pune Locality Maps

Alandi Road	Hadapsar	Navi Peth
Ambegaon Budruk	Hadapsar Industrial Estate	Padmavati
Anandnagar	Hingne Khurd	Parvati Darshan
Aundh	Jangali Maharaj Road	Pashan
Aundh Road	Kalyani Nagar	Paud Road
Balaji Nagar	Karve Nagar	Pirangut
Baner	Karve Road	Prabhat Road
Baner road	Kasba Peth	Pune Railway Station
Bhandarkar Road	Katraj	Rasta Peth
Bhavani Peth	Khadaki	Raviwar Peth
Bibvewadi	Khadki	Sadashiv Peth
Bopodi	Kharadi	Sahakar Nagar
Budhwar Peth	Kondhwa	Salunke Vihar
Bund Garden Road	Kondhwa Budruk	Sasson Road
Camp	Kondhwa Khurd	Satara Road
Chandan Nagar	Koregaon Park	Senapati Bapat Road
Dapodi	Kothrud	Shaniwar Peth
Deccan Gymkhana	Law College Road	Shivaji Nagar
Dehu Road	Laxmi Road	Shukrawar Peth
Dhankawadi	Lulla Nagar	Sinhagad Road
Dhayari Phata	Mahatma Gandhi Road	Somwar Peth
Dhole Patil Road	Mangalwar peth	Swargate
Erandwane	Manik Bagh	Tilak Road
Fatima Nagar	Market yard	Uruli Devachi
Fergusson College Road	Model colony	Vadgaon Budruk
Ganesh Peth	Mukund Nagar	Wadgaon Sheri

4.2.2 Getting the Geographical Coordinates

Geographical coordinates in ready form is difficult to find.
However,many APIs are available which can fetch the latitude and longitude of a region.
The Nominatim module from Geopy.Geocoders library was used to fetch coordinates for the fetched localities.

	Locality	Latitude	Longitude
0	Alandi Road	18.5523	73.8733
1	Ambegaon Budruk	18.4511	73.8376
2	Anandnagar	18.5083	73.8152
3	Aundh	18.5619	73.8102
4	Aundh Road	18.5619	73.8102
...	...	...	...
76	Viman Nagar	18.5214	73.8545
77	Wagholi	18.5806	73.9833
78	Wanowrie	18.4884	73.8987
79	Warje	18.482	73.8002
80	Yerawada	18.5656	73.8866

81 rows × 3 columns

4.2.3 Fetching Hospital data

There are many location data providers available such as Foursquare, Gowalla, Loopt etc.
The FourSquare API was used to fetch the venue data for our localities.
The world's top companies and more than 150,000 registered developers rely on Foursquare to power geo-tagging, venue search and more in their apps.
For fetching relevant data from Foursquare,a request needs to be sent to the API through URL.The response is obtained in the JSON format.
In order to obtain Hospital data,the category Id of hospital was added to the URL.

Thus we obtained details of all the hospitals present in a particular locality.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alandi Road	18.552349	73.873259	Ruby Hall Clinic	18.533768	73.876809	Hospital
1	Alandi Road	18.552349	73.873259	Serene Hospital	18.572227	73.879019	Hospital
2	Alandi Road	18.552349	73.873259	Inlaks & Budhrani Hospital	18.535327	73.887502	Hospital
3	Alandi Road	18.552349	73.873259	Jehangir Hospital	18.530495	73.876567	Hospital
4	Alandi Road	18.552349	73.873259	Sahaydri Hospital	18.554218	73.897066	Hospital
...	...	...	...	...	...	...	...
667	Warje	18.482044	73.800170	Deoyani hospital	18.494410	73.812720	Hospital
668	Warje	18.482044	73.800170	Shashwat hospital	18.495182	73.813535	Hospital
669	Warje	18.482044	73.800170	Sahyadri Hospital	18.507457	73.805752	Hospital
670	Yerawada	18.565632	73.886576	Serene Hospital	18.572227	73.879019	Hospital
671	Yerawada	18.565632	73.886576	Sahaydri Hospital	18.554218	73.897066	Hospital

672 rows × 7 columns

5 METHODOLOGY

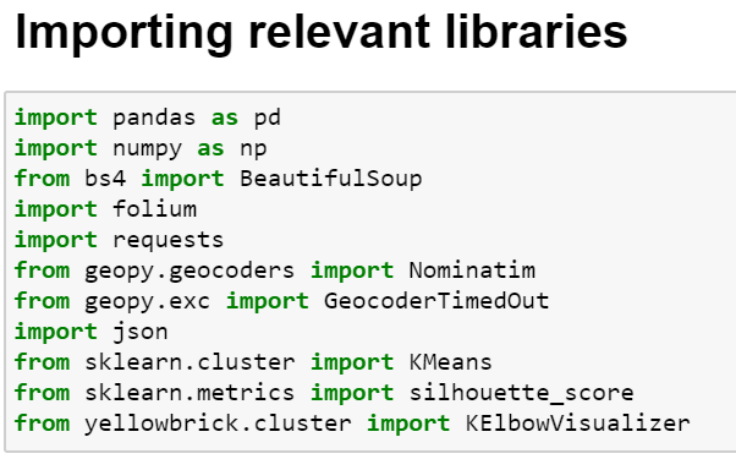
5.1 Steps performed

Data Science is always performed through a sequence of steps.The following steps were performed in the same order to find the localities with highest need for creation of new Health care infrastructure in the Pune city. :-

1.Importing relevant libraries
2.Web-Scraping.
3.Extracting Geographical coordinates.
4.Preliminary Visualization.
5.Fetching Location Data.
6.Extracting Relevant Feature.
7.Finding the optimal number of Clusters.
8.Clustering.
9.Final Visualization.
10.Find clusters with highest need for new Health Care infrastructure.

5.2 Importing relevant libraries

Inorder to perform any data analysis task,a number of libraries are required.The first step is to download all the necessary libraries and import them
The following libraries were used in this Project.
1.Pandas
2.Numpy
3.BeautifulSoup
4.Folium
5.Geopy
6.Json
7.Requests
8.Sklearn
9.Yellowbrick



5.3 Web-Scraping

The next step was to create a list of all localities in the Pune city. This was done through web scraping.
Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites,Web Scraping process will perform the same task within a fraction of the time.
The URL "https://www.mapsofindia.com/pune/localities/" was used to get the list of all localities in Pune using the technique of web scraping via the Beautiful Soup API.

Pune Locality Maps

Alandi Road	Hadapsar	Navi Peth
Ambegaon Budruk	Hadapsar Industrial Estate	Padmavati
Anandnagar	Hingne Khurd	Parvati Darshan
Aundh	Jangali Maharaj Road	Pashan
Aundh Road	Kalyani Nagar	Paud Road
Balaji Nagar	Karve Nagar	Pirangut
Baner	Karve Road	Prabhat Road
Baner road	Kasba Peth	Pune Railway Station
Bhandarkar Road	Katraj	Rasta Peth
Bhavani Peth	Khadaki	Raviwar Peth
Bibvewadi	Khadki	Sadashiv Peth
Bopodi	Kharadi	Sahakar Nagar
Budhwar Peth	Kondhwa	Salunke Vihar
Bund Garden Road	Kondhwa Budruk	Sasson Road
Camp	Kondhwa Khurd	Satara Road
Chandan Nagar	Koregaon Park	Senapati Bapat Road
Dapodi	Kothrud	Shaniwar Peth
Deccan Gymkhana	Law College Road	Shivaji Nagar
Dehu Road	Laxmi Road	Shukrawar Peth
Dhankawadi	Lulla Nagar	Sinhagad Road
Dhayari Phata	Mahatma Gandhi Road	Somwar Peth
Dhole Patil Road	Mangalwar peth	Swargate
Erandwane	Manik Bagh	Tilak Road
Fatima Nagar	Market yard	Uruli Devachi
Fergusson College Road	Model colony	Vadgaon Budruk
Ganesh Peth	Mukund Nagar	Wadgaon Sheri

Scrapping webpage to obtain names of all Localities in Pune

```
url="https://www.mapsofindia.com/pune/localities/"
page = requests.get(url).text
soup = BeautifulSoup(page, "lxml")
all_list=soup.find_all("a")
#print(soup.prettify())
#print(all_list)
flag=0
cnt=0
localities=[]
for li in all_list:
    #print(li.text)
    if(li.text.rstrip()=='Alandi Road'):
        flag=1
    if(flag==1):
        localities.append(li.text.rstrip())
    if(li.text=='Yerawada ' ):
        flag=0
```

5.4 Extracting Geographical coordinates

Thereafter the geographical coordinates ie latitude and longitude of all localities were obtained.The Geopy library along with Nominatim geocoder was used for this purpose

Geopy is a Python 2 and 3 client for several popular geocoding web services.

Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Getting geographical coordinates for all Localities

```
latitudes=[]
longitudes=[]
for x in localities:
    geolocator=Nominatim(user_agent=x)
    location=geolocator.geocode(x+',Pune',timeout=50)
    #print(location)
    if(location!=None):

        lat=location.latitude
        lon=location.longitude
        latitudes.append(lat)
        longitudes.append(lon)
    else:
        latitudes.append("na")
        longitudes.append("na")
```

	Locality	Latitude	Longitude
0	Alandi Road	18.5523	73.8733
1	Ambegaon Budruk	18.4511	73.8376
2	Anandnagar	18.5083	73.8152
3	Aundh	18.5619	73.8102
4	Aundh Road	18.5619	73.8102
...	...	...	...
76	Viman Nagar	18.5214	73.8545
77	Wagholi	18.5806	73.9833
78	Wanowrie	18.4884	73.8987
79	Warje	18.482	73.8002
80	Yerawada	18.5656	73.8866

81 rows × 3 columns

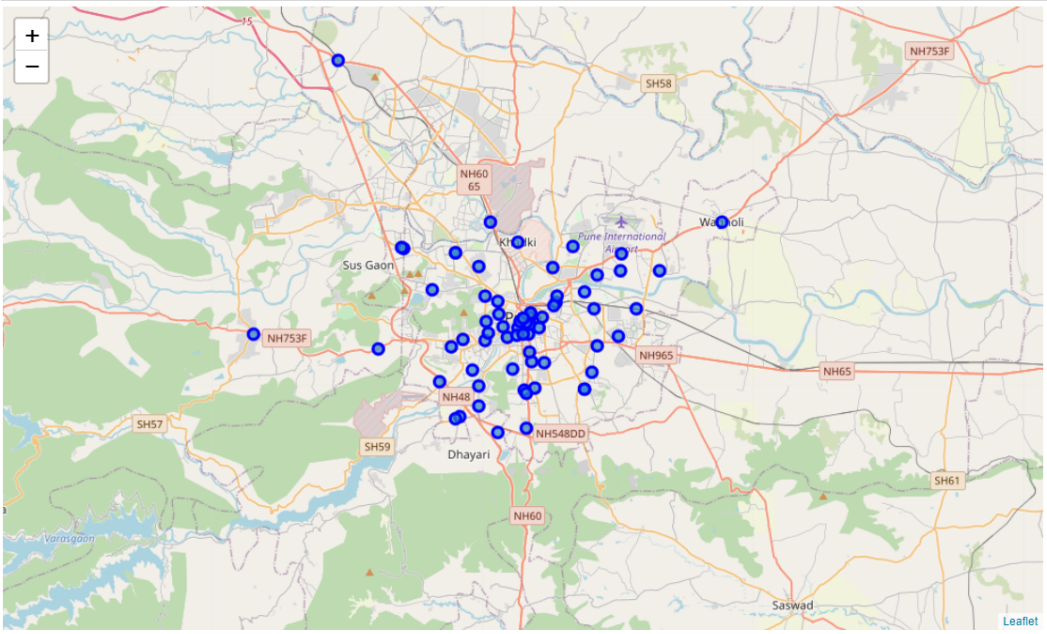
5.5 Preliminary Visualization

Getting a sense of the data before performing any analysis is always a good idea. A map of all our localities in Pune was plotted using the Folium library.
Folium makes it easy to visualize data that's been manipulated in Python on an interactive Leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing Vincent/Vega visualizations as markers on the map.

The library has a number of built-in tilesets from OpenStreetMap, MapQuest Open, MapQuest Open Aerial, Mapbox, and Stamen, and supports custom tilesets with Mapbox or Cloudmade API keys. Folium supports both GeoJSON and TopoJSON overlays, as well as the binding of data to those overlays to create choropleth maps with color-brewer color schemes.

Lets Visualize our localities before clustering

```
geolocator=Nominatim(user_agent='Pune')
location=geolocator.geocode('Pune,Maharashtra')
Pune=folium.Map(location=[location.latitude,location.longitude],zoom_start=11)
for lat, lng, label in zip(df['Latitude'], df['Longitude'], df['Locality']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(Pune)
Pune
```



5.6 Fetching Location Data

One of the core steps of our analysis was to fetch location specific data for our localities.
We needed information of all the hospitals present within 3km radius of each locality present in Pune city.
There are many location data providers available such as Foursquare, Gowalla, Loopt etc.
The FourSquare API was used to fetch the venue data for our localities.
The world's top companies and more than 150,000 registered developers rely on Foursquare to power geo-tagging, venue search and more in their apps.
For fetching relevant data from Foursquare,a request needs to be sent to the API

through URL.The response is obtained in the JSON format.
In order to obtain Hospital data,the category Id of hospital was added to the URL.
Thus we obtained details of all the hospitals present in a particular locality.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alandi Road	18.552349	73.873259	Ruby Hall Clinic	18.533768	73.876809	Hospital
1	Alandi Road	18.552349	73.873259	Serene Hospital	18.572227	73.879019	Hospital
2	Alandi Road	18.552349	73.873259	Inlaks & Budhrani Hospital	18.535327	73.887502	Hospital
3	Alandi Road	18.552349	73.873259	Jehangir Hospital	18.530495	73.876567	Hospital
4	Alandi Road	18.552349	73.873259	Sahaydri Hospital	18.554218	73.897066	Hospital
...	...	...	...	...	...	...	...
667	Warje	18.482044	73.800170	Deoyani hospital	18.494410	73.812720	Hospital
668	Warje	18.482044	73.800170	Shashwat hospital	18.495182	73.813535	Hospital
669	Warje	18.482044	73.800170	Sahyadri Hospital	18.507457	73.805752	Hospital
670	Yerawada	18.565632	73.886576	Serene Hospital	18.572227	73.879019	Hospital
671	Yerawada	18.565632	73.886576	Sahaydri Hospital	18.554218	73.897066	Hospital

672 rows × 7 columns

5.7 Extracting Relevant Feature

Feature extraction is the process of selecting a feature from a number of features and converting it to the form suitable for Algorithmic processing.
In our project,the clustering of localities was performed on the basis of number of hospitals present in those localities.
Thus,the relevant feature for our project was the count of hospitals.The count of hospitals for each locality was calculated by aggregating the data for each locality from the data which was returned by FourSquare API.

```
Finaldf.rename(columns={'Neighborhood':'Locality','Neighborhood Latitude':'Locality Latitude','Neighborhood Longitude':'Locality
Analysisdf=Finaldf.groupby(['Locality','Locality Latitude','Locality Longitude']).count()
Analysisdf.reset_index(inplace=True)
Analysisdf=Analysisdf.iloc[:,:4]
Analysisdf.rename(columns={'Venue':'Hospital Count'},inplace=True)
Analysisdf
```

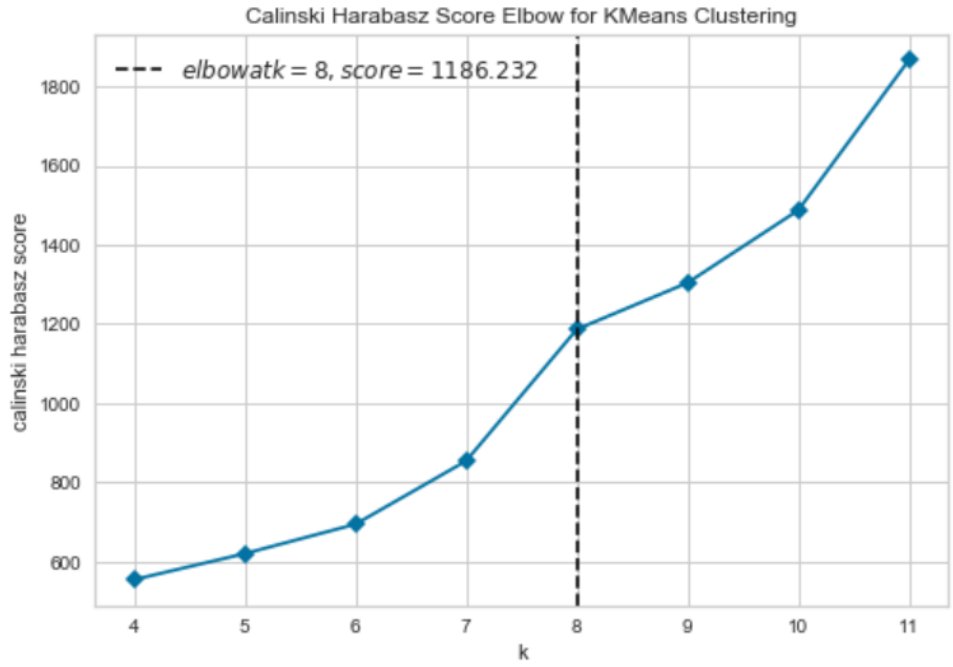
	Locality	Locality Latitude	Locality Longitude	Hospital Count
0	Alandi Road	18.552349	73.873259	6
1	Ambegaon Budruk	18.451118	73.837555	3
2	Anandnagar	18.508338	73.815208	10
3	Aundh	18.561883	73.810196	8
4	Aundh Road	18.561883	73.810196	8
...	...	...	...	...
74	Wadgaon Sheri	18.550441	73.917167	4
75	Wagholi	18.580630	73.983310	3
76	Wanowrie	18.488368	73.898667	6
77	Warje	18.482044	73.800170	4
78	Yerawada	18.565632	73.886576	2

5.8 Finding the optimal number of Clusters

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.
One of the popular clustering algorithm is Kmeans and we used the same algorithm in our project.
However,its necessary to specify the number of clusters before hand.There are a number of ways to find the optimal number of clusters
We used the elbow method along with calinski harabasz score to find the optimal number of cluster.The score is defined as ratio between the within-cluster dispersion and the between-cluster dispersion.

```
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
#!pip install yellowbrick
from yellowbrick.cluster import KElbowVisualizer
# set number of clusters
#for kclusters in range(2,20):
 #   kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Analysisdf[['Hospital Count']])
  #  print(kclusters,silhouette_score(Analysisdf[['Hospital Count']],kmeans.labels_))
model=KMeans()
visualizer = KElbowVisualizer(
    model, k=(4,12), metric='calinski_harabasz', timings=False,locate_elbow=True
)

visualizer.fit(Analysisdf[['Hospital Count']])        # Fit the data to the visualizer
visualizer.show()
```



Using the elbow method the optimal number of clusters was found to be 8.

5.9 Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning.Kmeans algorithm is easy to perform and most popular clustering algorithm and hence the same algorithm was used.

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into Kpre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.
The Sklearn library was used to perform Kmeans clustering.
The optimal number of clusters was found to be 8 and hence the localities were divided into 8 clusters.
The clustering was performed using Euclidean distance metric with number of hospitals as a feature.

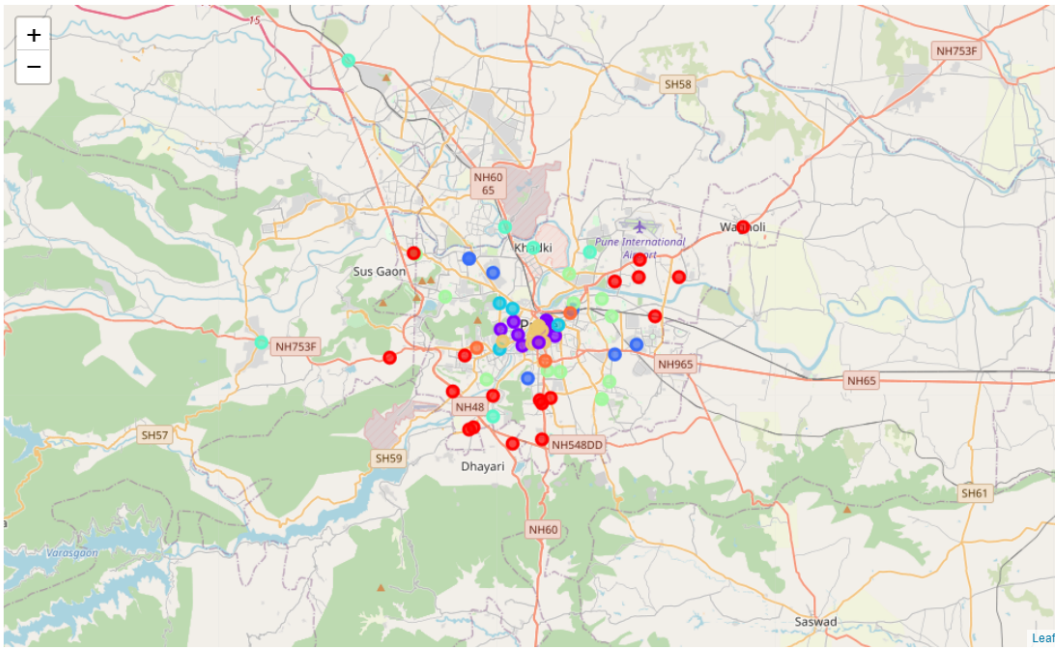
Clustering the localities into 8 clusters

```
kclusters=8
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Analysisdf[['Hospital Count']
print(kmeans.labels_,kmeans.cluster_centers_)
```

```
[5 0 7 2 2 3 5 0 7 6 0 1 2 7 0 4 1 4 0 5 2 1 1 2 1 5 3 5 6 0 2 0 0 5 3 1 0
 4 0 5 5 5 0 1 4 3 1 6 3 5 0 6 6 1 0 5 0 4 6 6 3 6 2 0 6 6 5 1 0 7 7 3 4 6
 0 0 5 0 4] [[ 3.57894737]
 [14.9       ]
 [ 7.57142857]
 [12.57142857]
 [ 1.42857143]
 [ 5.53846154]
 [16.54545455]
 [10.        ]]
```

5.10 Final Visualization

The Localities were visualized post clustering with unique color assigned to each cluster.
The Folium library was used to create this map of clustered localities.



5.11 Find clusters with highest need for new Health Care infrastructure

Once the clusters were formed,the final step was to identify clusters with highest need for new Health care infrastructure. The clusters with the lowest number of mean hospitals were identified as the Highest priority clusters.
The cluster centroids represented the mean number of hospitals value for its cluster.The centroids were sorted in ascending order and top 2 clusters were selected

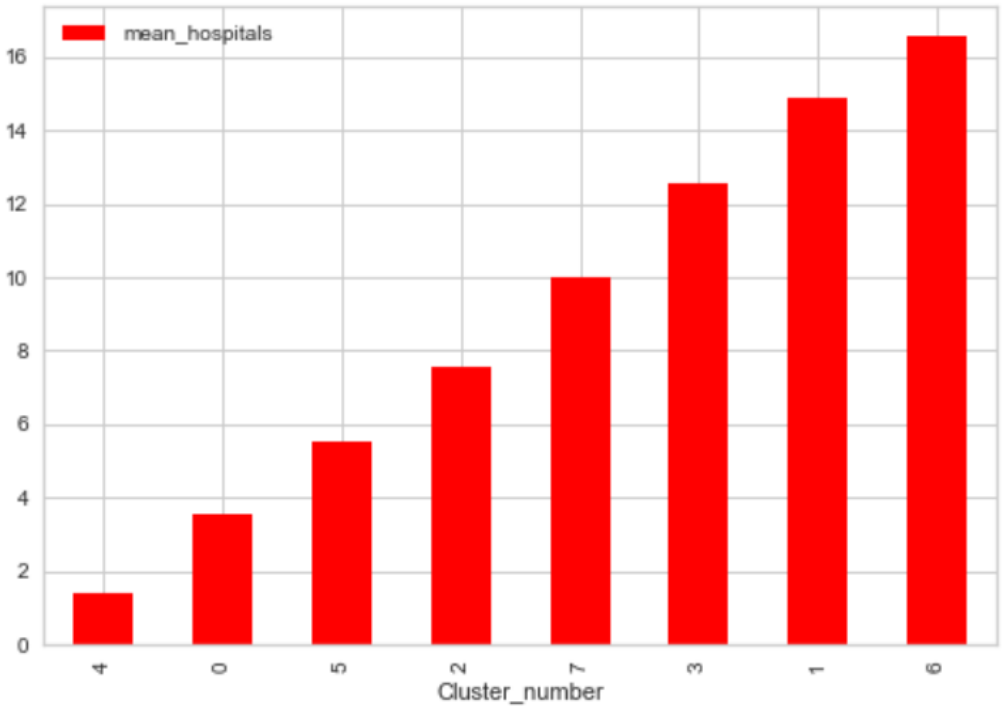
as the highest priority regions.

6 RESULTS

1)8 Clusters of localities in the Pune city based on number of hospitals were formed.
2)Cluster number 4 had the lowest mean number of hospitals with a value of 1.42 hospitals.
3)Cluster number 0 had the second lowest mean number of hospitals with a value of 3.57 hospitals.
4)Clusters 7,3,1 and 6 had mean number of hospitals greater than 10.
5)Thus, localities belonging to Cluster 4 have the highest priority followed by those belonging to Cluster 0 for creation of new Health Care infrastructure

Cluster Number	Mean Hospitals
0	3.57
1	14.9
2	7.57
3	12.57
4	1.42
5	5.53
6	16.54
7	10.00

Table 1: Clusters and Mean Hospitals



Localities belonging to Cluster 4

```
Cluster4
```

	Cluster Labels	Locality	Locality Latitude	Locality Longitude	Hospital Count
15	4	Dapodi	18.580846	73.832775	1
17	4	Dehu Road	18.680047	73.734331	1
37	4	Khadki	18.568175	73.850779	2
44	4	Laxmi Road	18.141836	74.561703	2
57	4	Pirangut	18.511282	73.679007	1
72	4	Vadgaon Budruk	18.467497	73.825365	1
78	4	Yerawada	18.565632	73.886576	2

Localities belonging to Cluster 0

```
Cluster0
```

	Cluster Labels	Locality	Locality Latitude	Locality Longitude	Hospital Count
1	0	Ambegaon Budruk	18.451118	73.837555	3
7	0	Baner road	18.564642	73.775079	4
10	0	Bibvewadi	18.478174	73.862105	3
14	0	Chandan Nagar	19.121709	72.923203	4
18	0	Dhayari Phata	18.460862	73.812755	3
29	0	Hadapsar	18.526967	73.927825	4
31	0	Hingne Khurd	18.479670	73.825099	3
32	0	Kalyani Nagar	18.548138	73.902551	4
36	0	Katraj	18.453679	73.856320	4
38	0	Kharadi	18.550518	73.942494	4
42	0	Kothrud	18.503889	73.807673	4
50	0	Nagar Road	18.561016	73.918279	4
54	0	Padmavati	18.477317	73.854770	3
56	0	Paud Road	18.502704	73.760385	4
63	0	Satara Road	18.475207	73.856165	3
68	0	Sinhagad Road	18.459608	73.809984	3
74	0	Wadgaon Sheri	18.550441	73.917167	4
75	0	Wagholi	18.580630	73.983310	3
77	0	Warje	18.482044	73.800170	4

7 DISCUSSION

This project was inspired by the suffering which humanity experienced due to the 2020 Covid-19 Pandemic.Millions of people were affected due to the lack of adequate Health Care infrastructure.
The target city in this Project was Pune,India since its the city I live in and I could relate to it better.
Apart from finding out the localities with less number of hospitals,I also found out an interesting observation.

Most of the localities belonging to High priority clusters were in the outskirts of the city whereas the localities present in the core of the city had comparatively better Health Care infrastructure.
This means that the localities belonging to the outer regions are somewhat neglected and efforts need to taken to improve the facilities and infrastructure in these regions.

8 FUTURE SCOPE

1.The project analyzed the Pune city only,however the same analysis can be repeated for other cities which will help us improve the Health Care facilties throughout the World.
2.Once we have sufficient data present for all cities,a centralized web or stand alone application can be created wherein by entering the city name one could fetch such statistics in a single click and within seconds.

9 CONCLUSION

To conclude,this project was an attempt to find out the localities in Pune,India where new Health Care infrastructure needs to be set up urgently.We found out that there are some localities where the number of hospitals are extremely less and efforts need to be taken to improve the situation.We also understood that most of these localities lie in the outskirts of the city whereas the situation is comparatively better in the cenral part.

Finally,I would like to end by stating that we need to make use of science and technology to prepare humanity for such kind of a crisis beforehand so that we can withstand any kind of storm together.