

PENNSYLVANIA PREPAREDNESS LEVEL FOR COVID-19 PANDEMIC

ROHIT RAWAL

The Pennsylvania State University, State College PA 16801, USA
rur281@psu.edu

Keywords: Pennsylvania Municipalities, Foursquare API, K-means clustering

1 INTRODUCTION

The COVID-19 Pandemic is one of the biggest global health crisis that has gripped the world currently and has spread across the globe with alarmingly high number of confirmed cases in the United States of America. At the time of writing this report, the global number of confirmed COVID-19 cases were over 2,400,000 with around 165,000 deaths and Pennsylvania had over 31000 confirmed cases and more than 1000 deaths. Governments across the world are working hard to impede the spread of the virus by providing robust testing and medical services to patients, deploying travel restrictions, stay at home orders, and cancelling large gatherings such as sporting events, concerts, and school sessions.

Through this project, we will be evaluating the preparedness level among the different counties and municipalities in Pennsylvania to handle a health crisis of such scale. We will analyze the hospital utilization across different counties and municipalities to understand the areas that are better equipped with adequate medical services. Also, we will use K-means clustering to visualize and identify the clusters having different levels of medical services (hospital beds per given population). This analysis would help the audience to gauge the level of preparedness in each PA county and would help the government to identify the regions that require critical attention in terms of medical services.

2 DATA

In order to perform the required analysis, we would need to use four different datasets. The analysis would require identifying the hospital utilization across the counties in Pennsylvania. For this, we would need the population & location data of all the PA Counties, and the number of available hospital beds across PA.

- Pennsylvania data that contains **counties** and **Municipalities** along with their **latitudes** and **longitudes**.
Data source: PA County Location, PA County Municipalities
- Pennsylvania County and Municipality **population** data from online database and Web Scraping (Wikipedia).
Data source: PA County Population, PA Municipality Population
- Information on available hospitals in PA Municipalities from Foursquare API.
Data source: Foursquare API
- Hospital utilization information from **PA Health Data**.
Data source: PA Hospital Utilization

3 METHODOLOGY

The illustration below outlines the methodology adopted in order to address the given problem:



3.1 Data Collection

The data is captured through various sources in the form of excel spreadsheets (csv format) and through web scraping (using Beautiful Soup Python Package). Let us elaborate on the captured Datasets in detail.

County Location Data:

Here is the data of each PA County with their latitude and longitude coordinates extracted from the PA Government data repository. There are a total of 67 counties in Pennsylvania.

	COUNTY	Longitude	Latitude
0	ADAMS	-77.222243	39.872096
1	BERKS	-75.930773	40.419396
2	CAMBRIA	-78.718942	40.491275
3	SULLIVAN	-76.514792	41.448099
4	MERCER	-80.260094	41.302378

Figure 1 PA County Data

PA Municipalities Data (Boroughs) :

Here is the list of municipalities within each County. There are a total of 956 municipalities (Boroughs).

	COUNTY	MUNICIPALITY	CLASS
0	ADAMS	Abbottstown	Borough
1	ADAMS	Arendtsville	Borough
2	ADAMS	Bendersville	Borough
4	ADAMS	Biglerville	Borough
5	ADAMS	Bonneauville	Borough
7	ADAMS	Carroll Valley	Borough
10	ADAMS	East Berlin	Borough
11	ADAMS	Fairfield	Borough
15	ADAMS	Gettysburg	Borough

Figure 2 PA Borough Data

PA Hospital Data:

Here is the list of Hospitals within each county of Pennsylvania along with their available licensed beds. This data is taken through PA Health Data repository. This data is based in the year 2018 and hence the numbers may vary for future analysis.

	COUNTY	FACILITY NAME	LICENSED BEDS
0	ADAMS	GETTYSBURG HOSPITAL	76
1	ALLEGHENY	ALLEGHENY GENERAL HOSPITAL	552
2	ALLEGHENY	ALLEGHENY VALLEY HOSPITAL	190
3	ALLEGHENY	FORBES HOSPITAL	315
4	ALLEGHENY	HERITAGE VALLEY SEWICKLEY	176

Figure 3 PA Hospital Data

PA County Population Data:

Below is the list of Pennsylvania Counties along with their Population Count. The database was available in csv format on PA Government Data Repository.

	COUNTY	Population
0	PHILADELPHIA	1584138
1	ALLEGHENY	1218452
2	MONTGOMERY	828604
3	BUCKS	628195
4	DELAWARE	564751

Figure 4 County Population Data

Municipality Location Data:

We have used OpenCageGeocode API to determine the location coordinates for each municipality within the counties of Pennsylvania.

	COUNTY	MUNICIPALITY	lat	lon
0	ADAMS	ABBOTTSTOWN	39.886487	-76.984701
1	ADAMS	ARENDTSVILLE	39.923149	-77.298596
2	ADAMS	BENDERSVILLE	39.982593	-77.249428
4	ADAMS	BIGLERVILLE	39.930378	-77.248271
5	ADAMS	BONNEAUVILLE	39.812042	-77.137205

Figure 5 Municipality Location Data

Municipality Population Data

With the use of Web Scraping, we have extracted the population data for each municipality with the Pennsylvania Counties from Wikipedia Page.

	COUNTY	Pop	Area[1](mi2)	Pop. dens.(p/mi2)	MUNICIPALITY
27	MONTGOMERY	34,427	3.6	9541.9	NORRISTOWN
45	ALLEGHENY	28,386	19.7	1437.6	MONROEVILLE
48	ALLEGHENY	27,395	29.0	946.0	PLUM
73	MONTGOMERY	22,480	5.0	4528.6	POTTSTOWN
82	FRANKLIN	20,360	6.9	2940.1	CHAMBERSBURG

Figure 6 Municipality Population Data

Foursquare Hospital Data:

Using the Foursquare API, we will collect the hospital data within each municipality.

	ID	FACILITY NAME	Latitude	Longitude	MUNICIPALITY	COUNTY
0	59b1545b79f6c733be83a7c7	GETTYSBURG HOSPITAL	39.825022	-77.234436	GETTYSBURG	ADAMS
1	4ba6494bf964a520904239e3	WELLSPAN GETTYSBURG HOSPITAL: EMERGENCY ROOM	39.824420	-77.234929	GETTYSBURG	ADAMS
2	4e568e3c7d8b60849c50fe36	TRANSITIONS HEALTH CARE	39.840539	-77.229855	GETTYSBURG	ADAMS
3	50ec5f65e4b0d785bc62494a	C113	39.824442	-77.235076	GETTYSBURG	ADAMS
4	4bfbdc84d0382d7f9e28ca0a	MEDEXPRESS URGENT CARE	40.491290	-79.907848	ASPINWALL	ALLEGHENY

Figure 7 Foursquare Hospital Data

3.2 Exploratory Data Analysis

3.2.1 Data Preprocessing

Let us combine the Population data with the location data to get the below dataframe.

	COUNTY	MUNICIPALITY	Pop	lat	lon
0	ADAMS	ARENDTSVILLE	941	39.923149	-77.298596
1	ADAMS	BENDERSVILLE	640	39.982593	-77.249428
2	ADAMS	BIGLERVILLE	1,207	39.930378	-77.248271
3	ADAMS	BONNEAUVILLE	1,804	39.812042	-77.137205
4	ADAMS	FAIRFIELD	506	39.787317	-77.368598

Figure 8 Municipality Location Combined Data

We will now combine the foursquare data with the hospital data to get the number of beds attributed to each hospital in a given municipality. In order to match the hospital names we have used Fuzzywuzzy module:

```
# Now Let us combine hospital ber neighborhood data with hospital bed data
def combine_hospital_beds_with_COUNTY_municipality(hospital_df, hospital_boro_nei_df):
    # Uses fuzzywuzzy to match hospital name from Foursquare to PA Hospital Data
    data = []
    column_names = ["FACILITY NAME", "LICENSED BEDS"]
    boro_neig_column_names = ["COUNTY", "MUNICIPALITY"]
    for i, row in hospital_df.iterrows():
        data_per_hospital = None
        max_ratio = 0
        for x, hrow in hospital_boro_nei_df.iterrows():
            ratio = fuzz.token_sort_ratio(row["FACILITY NAME"], hrow["FACILITY NAME"])
            if ratio > max_ratio:
                max_ratio = ratio
                data_per_hospital = [
                    row[column] for column in column_names + \
                    [hrow[column] for column in boro_neig_column_names]
                ]
        if data_per_hospital:
            data.append(data_per_hospital)

    df = pd.DataFrame(data, columns=column_names+boro_neig_column_names)
    df.to_csv('cleaned_hospital_data.csv')
    return df
```

Figure 9 Fuzzywuzzy Code

Here is the final dataset after merging on the hospital name:

	FACILITY NAME	LICENSED BEDS	COUNTY	MUNICIPALITY
0	GETTYSBURG HOSPITAL	76	ADAMS	GETTYSBURG
1	ALLEGHENY GENERAL HOSPITAL	552	ALLEGHENY	AVALON
2	ALLEGHENY VALLEY HOSPITAL	190	ALLEGHENY	TARENTUM
3	FORBES HOSPITAL	315	LUZERNE	KINGSTON
4	HERITAGE VALLEY SEWICKLEY	176	ALLEGHENY	SEWICKLEY

Figure 10 Municipality Hospital Data

After cleaning the data, we have the following dataset highlighting the total number of hospitals beds in each municipality of a given county:

	COUNTY	MUNICIPALITY	LICENSED BEDS
0	ADAMS	GETTYSBURG	76
1	ALLEGHENY	AVALON	552
2	ALLEGHENY	TARENTUM	190
3	LUZERNE	KINGSTON	315
4	ALLEGHENY	SEWICKLEY	176

Figure 11 Updated Dataframe

Now let us combine the above data with the population and the location dataset:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon
0	ADAMS	GETTYSBURG	528	7645	39.830897	-77.231012
1	ALLEGHENY	ASPINWALL	1365	2797	40.491242	-79.903510
2	ALLEGHENY	AVALON	833	4695	40.500902	-80.067555
3	ALLEGHENY	BELLEVUE	1172	8352	40.493958	-80.051722
4	ALLEGHENY	BRADDOCK	171	2153	40.403402	-79.868382

Figure 12 Hospital Data with Population

Let us add another attribute capturing the bed count per hundred people in each municipality:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon	BEDS PER 100 PEOPLE
0	ADAMS	GETTYSBURG	528	7645	39.830897	-77.231012	6.906475
1	ALLEGHENY	ASPINWALL	1365	2797	40.491242	-79.903510	48.802288
2	ALLEGHENY	AVALON	833	4695	40.500902	-80.067555	17.742279
3	ALLEGHENY	BELLEVUE	1172	8352	40.493958	-80.051722	14.032567
4	ALLEGHENY	BRADDOCK	171	2153	40.403402	-79.868382	7.942406

Figure 13 Beds per 100 People

3.2.2 Statistical Analysis and Data Visualization

Municipalities Count:

The number of municipalities in each Pennsylvania County is displayed in the bar chart shown below:

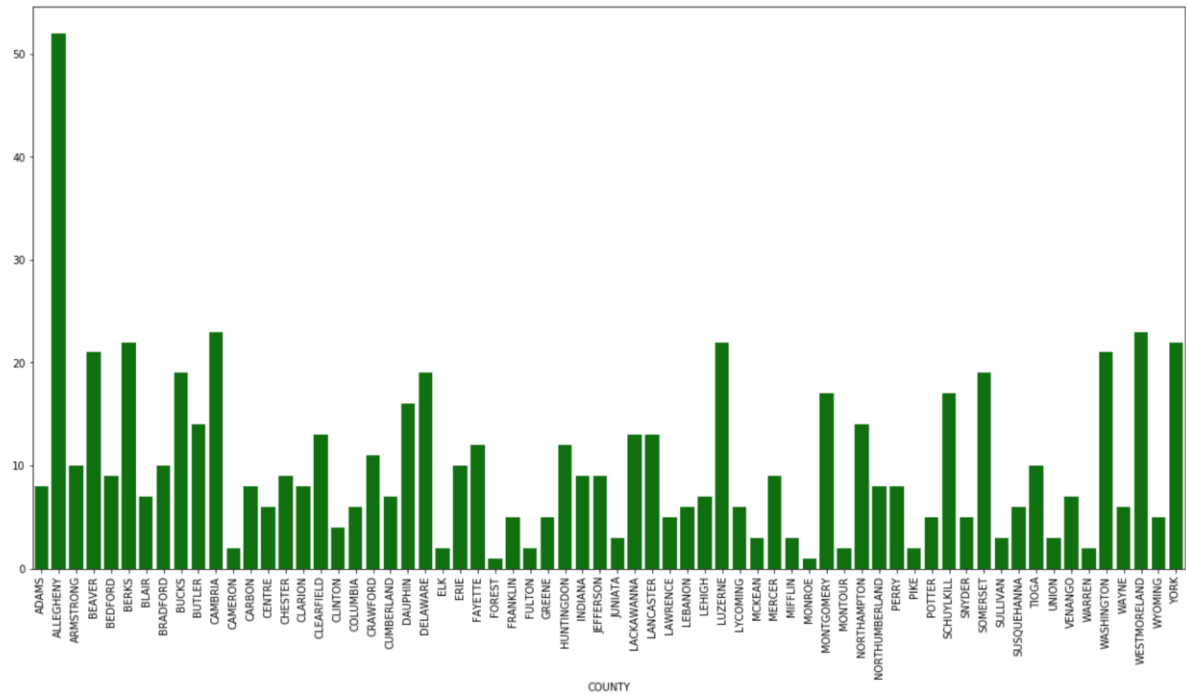


Figure 14 Municipalities Count

County Population Count:

Population for each county is shown below:

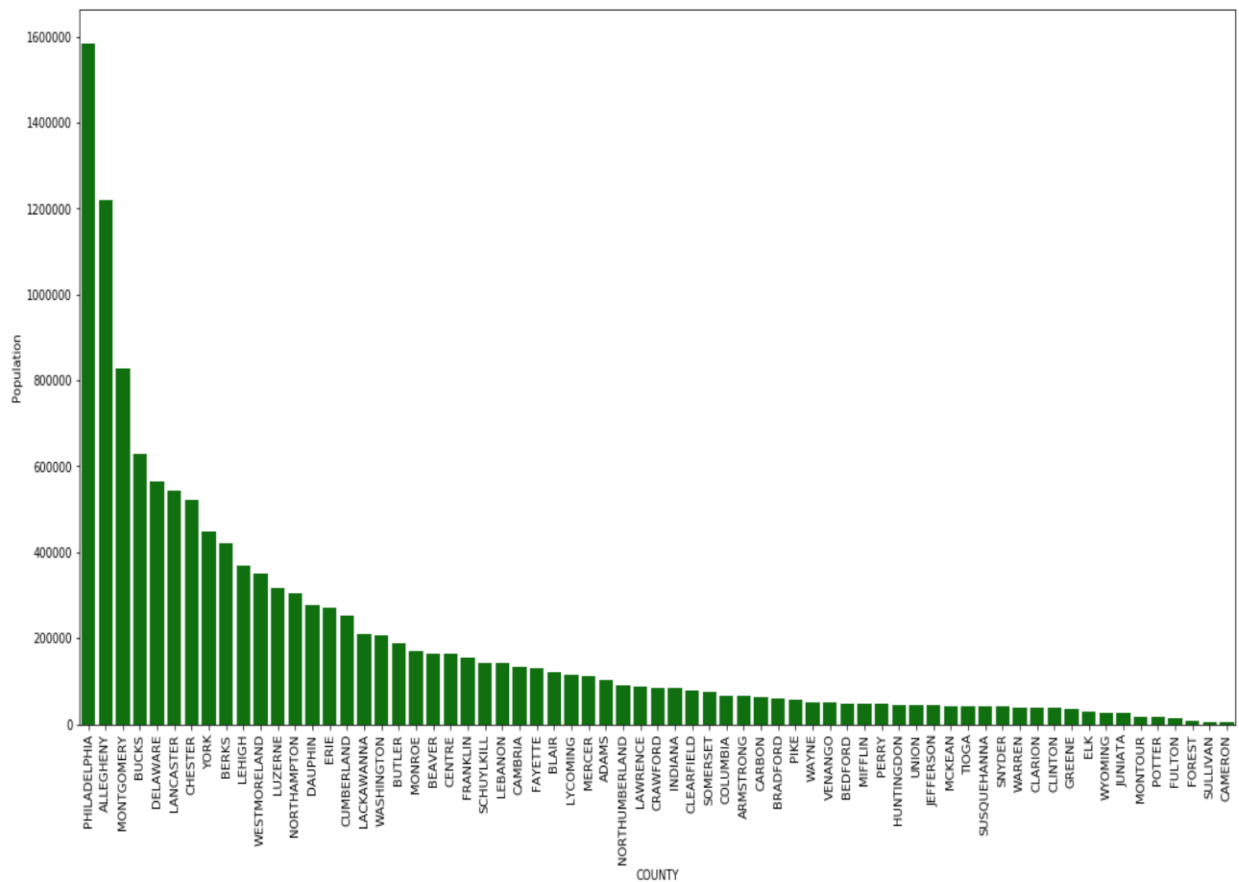


Figure 15 County Population

Municipality Hospital Bed Count:

The bar chart below represents the number of licensed beds available in each municipality:

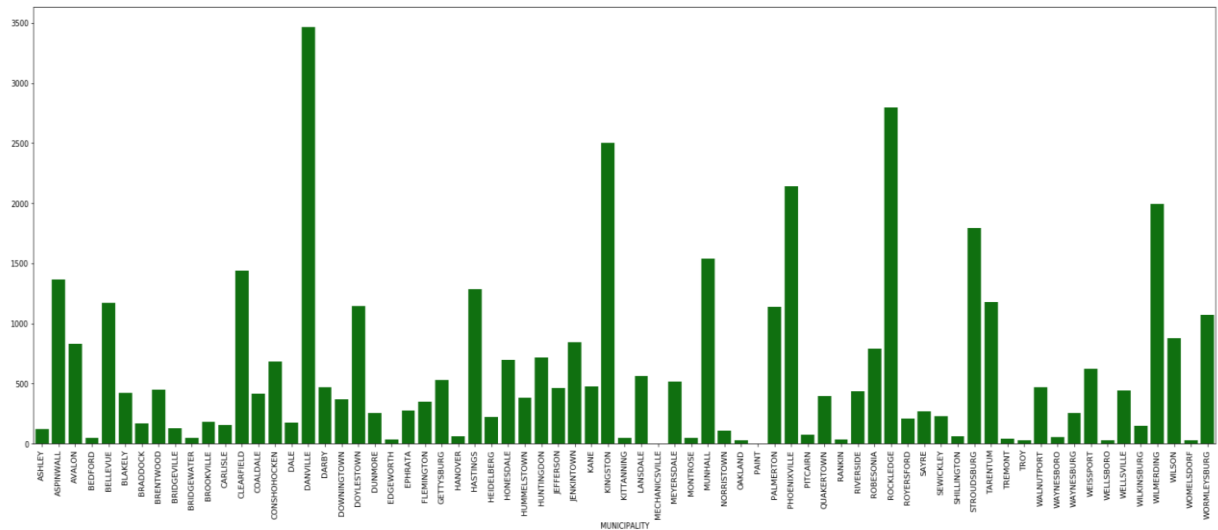


Figure 16 Municipality Hospital Bed Count

County Hospital Bed Count:

The bar chart below represents the number of licensed beds available in each county:

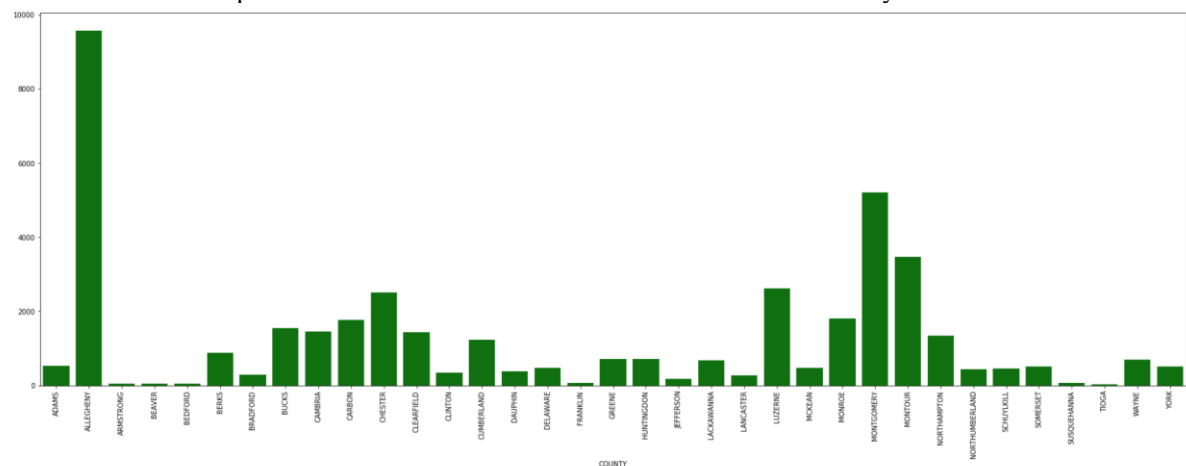


Figure 17 County Hospital Bed Count

The hospital bed counts are very inconsistent among counties with few counties having very low bed count and few having more than 60% of the total hospital beds attributed to it

3.3 Data Classification:

Let us now use K- means clustering to identify clusters within the dataset and partition the dataset. We will also be optimizing the parameter k using elbow method.

3.3.1 K-Means Clustering

For this we will use the Bed Count and Population Data attributes. We will also normalize the data first as the Population data is very large as compared to the Bed Counts and hence it may skew the results.


```
#Optimum k = 3
k_clusters = 3
# run k-means clustering
kmeans = KMeans(n_clusters=k_clusters, random_state=0).fit(cl_data)
# check cluster labels generated for each row in the dataframe
kmeans.labels_

array([2, 2, 2, 2, 2, 0, 2, 2, 2, 0, 2, 2, 2, 2, 0, 1, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2, 2, 0, 2, 2, 0, 0, 1, 2, 2, 2, 2,
       0, 0, 2, 0, 2, 2, 2, 2, 0, 0, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 0, 1])
```

Figure 18 Cluster Labels

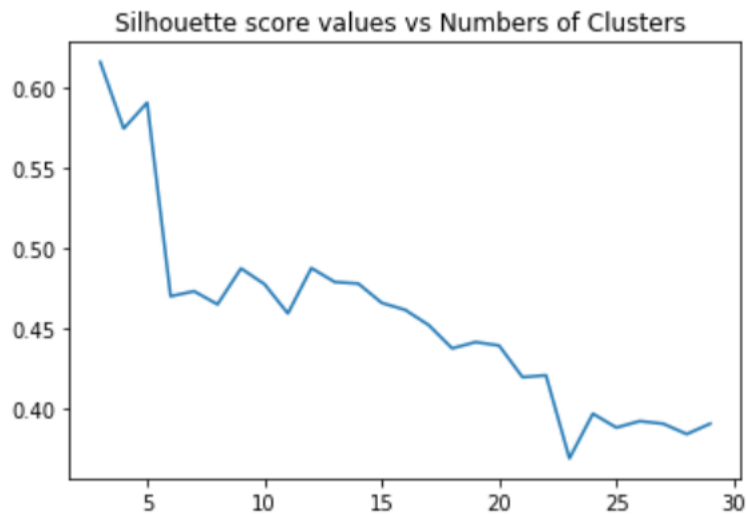
Adding the cluster labels to the original dataset we get the following data:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon	BEDS PER 100 PEOPLE	Cluster
0	ADAMS	GETTYSBURG	528	7645	39.830897	-77.231012	6.906475	2
1	ALLEGHENY	ASPINWALL	1365	2797	40.491242	-79.903510	48.802288	2
2	ALLEGHENY	AVALON	833	4695	40.500902	-80.067555	17.742279	2
3	ALLEGHENY	BELLEVUE	1172	8352	40.493958	-80.051722	14.032567	2
4	ALLEGHENY	BRADDOCK	171	2153	40.403402	-79.868382	7.942406	2

Figure 19 Updated Dataset with Cluster Labels

3.3.2 Parameter Optimization:

Using the elbow method, we determine the optimal number of k to be 3.



Optimal number of components is:
3

Figure 20 Elbow Method: Parameter Optimization

3.4 Graphical Analysis:

Let us now visualize the hospital bed count per 100 people in each municipality on the Pennsylvania Map using Folium:

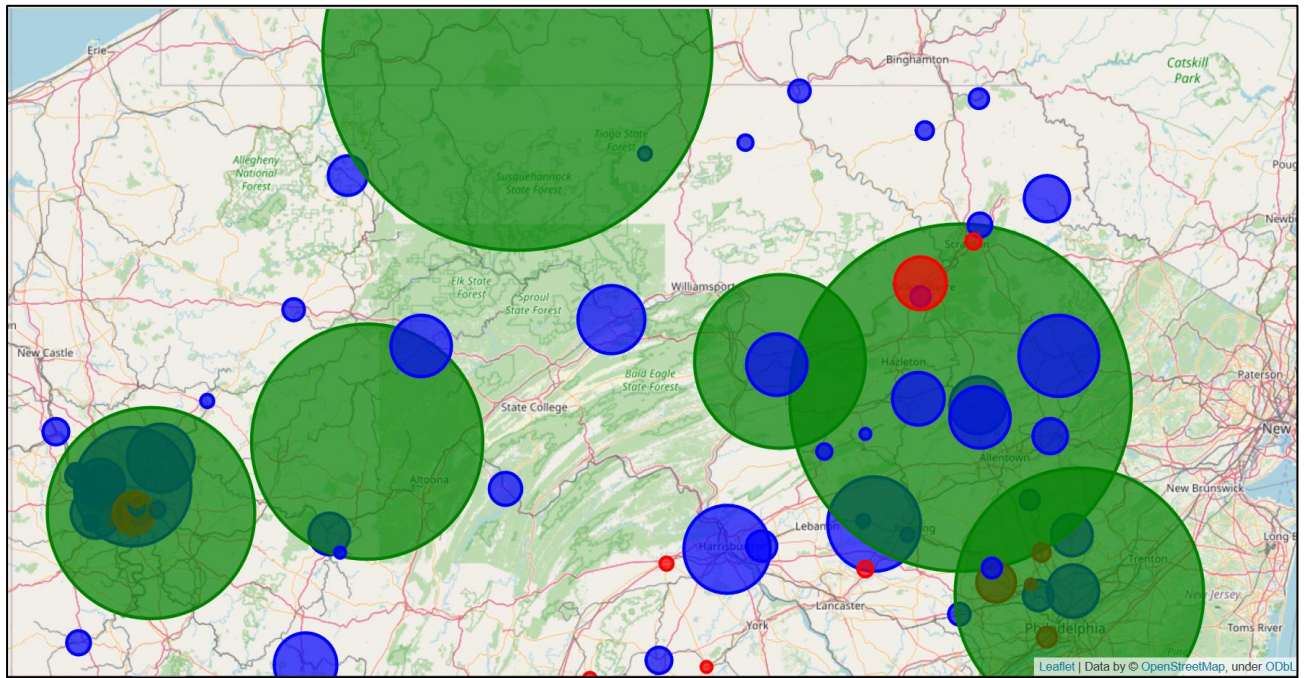
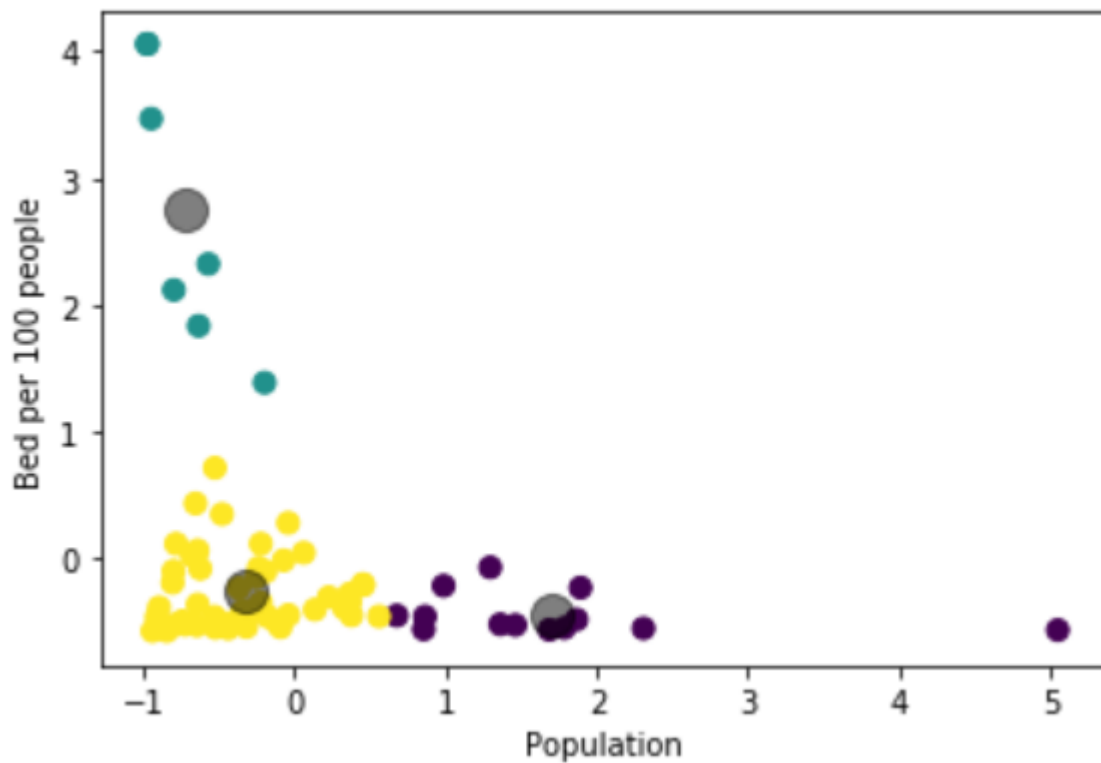


Figure 21 Cluster Representation

Here each color represents a different cluster and as we can see there are many municipalities that do not have any hospitals. Also, there are many regions which have a large count of hospitals available. This inconsistent distribution of the hospitals among municipalities makes it difficult for the state government to stay prepared and equipped to fight a global pandemic.

Now we will use a scatter plot to demonstrate the centroid positions based on the clusters formed:



4 RESULTS

Below are the datasets attributed to each cluster identified:

Cluster 1:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon	BEDS PER 100 PEOPLE	Cluster
5	ALLEGHENY	BRENTWOOD	446	9620	40.373269	-79.975621	4.636175	0
9	ALLEGHENY	MUNHALL	1542	11380	40.392291	-79.900050	13.550088	0
14	ALLEGHENY	WILKINSBURG	147	15906	40.441736	-79.881994	0.924180	0
31	CHESTER	PHOENIXVILLE	2143	16518	40.130382	-75.514913	12.973726	0
34	CUMBERLAND	CARLISLE	153	18880	40.201499	-77.189078	0.810381	0
37	DELAWARE	DARBY	466	10682	39.918446	-75.259072	4.362479	0
38	FRANKLIN	WAYNESBORO	57	10633	39.755783	-77.577651	0.536067	0
44	LACKAWANNA	DUNMORE	254	14069	41.419803	-75.632411	1.805388	0
45	LANCASTER	EPHRATA	273	13506	40.179911	-76.178924	2.021324	0
47	LUZERNE	KINGSTON	2502	13131	41.261748	-75.896863	19.054147	0
52	MONTGOMERY	LANSDALE	564	16367	40.241495	-75.283786	3.445958	0
53	MONTGOMERY	NORRISTOWN	106	34427	40.121497	-75.339905	0.307898	0
69	YORK	HANOVER	60	15349	39.806325	-76.984274	0.390905	0

Figure 22 Data for Cluster 1

Cluster 2:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon	BEDS PER 100 PEOPLE	Cluster
15	ALLEGHENY	WILMERDING	1994	2185	40.390902	-79.810047	91.258581	1
27	CAMBRIA	HASTINGS	1283	1258	40.665067	-78.712248	101.987281	1
29	CARBON	WEISSPORT	622	406	40.828992	-75.698547	153.201970	1
39	GREENE	JEFFERSON	460	262	42.015261	-94.377468	175.572519	1
54	MONTGOMERY	ROCKLEDGE	2799	2550	40.081222	-75.089618	109.764706	1
56	MONTOUR	DANVILLE	3462	4661	40.966481	-76.613147	74.275906	1
70	YORK	WELLSVILLE	444	253	42.122012	-77.948058	175.494071	1

Figure 23 Data for Cluster 2

Cluster 3:

	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon	BEDS PER 100 PEOPLE	Cluster
0	ADAMS	GETTYSBURG	528	7645	39.830897	-77.231012	6.906475	2
1	ALLEGHENY	ASPINWALL	1365	2797	40.491242	-79.903510	48.802288	2
2	ALLEGHENY	AVALON	833	4695	40.500902	-80.067555	17.742279	2
3	ALLEGHENY	BELLEVUE	1172	8352	40.493958	-80.051722	14.032567	2
4	ALLEGHENY	BRADDOCK	171	2153	40.403402	-79.868382	7.942406	2
6	ALLEGHENY	BRIDGEVILLE	126	5142	40.356181	-80.110056	2.450408	2
7	ALLEGHENY	EDGEWORTH	32	1681	40.551179	-80.192837	1.903629	2
8	ALLEGHENY	HEIDELBERG	224	1245	40.392291	-80.090889	17.991968	2
10	ALLEGHENY	PITCAIRN	76	3286	40.403124	-79.778102	2.312842	2
11	ALLEGHENY	RANKIN	35	2119	40.412569	-79.879216	1.651723	2
12	ALLEGHENY	SEWICKLEY	226	3821	40.536457	-80.184503	5.914682	2
13	ALLEGHENY	TARENTUM	1181	4521	40.601456	-79.759771	26.122539	2
16	ARMSTRONG	KITTANNING	49	3980	40.816453	-79.521989	1.231156	2
17	BEAVER	BRIDGEWATER	50	704	40.701020	-80.294370	7.102273	2
18	BEDFORD	BEDFORD	50	2792	52.136381	-0.467504	1.790831	2
19	BERKS	ROBESONIA	790	2070	40.351759	-76.134389	38.164251	2
20	BERKS	SHILLINGTON	60	5261	40.307870	-75.965490	1.140468	2

Figure 24 Data for Cluster 3

Let us now examine the counties and the municipalities without any hospitals.

Counties without Hospital:

Number of Counties : 67
 Number of Counties with Hospital: 60
 Number of Counties without Hospital: 7

PIKE
 PERRY
 SNYDER
 JUNIATA
 FOREST
 SULLIVAN
 CAMERON

Figure 25 Counties without Hospital

Municipalities without hospitals:

Number of Municipalities : 662
 Number of Municipalities with Hospital: 136
 Number of Municipalities without Hospital: 526

ARENDSVILLE
BENDERSVILLE
BIGLERVILLE
BONNEAUVILLE
FAIRFIELD
LITTLESTOWN
MCSHERRYSTOWN
BALDWIN
CARNEGIE
CHALFANT
CHESWICK
CHURCHILL
CORAOPOLIS
CRAFTON
DORMONT
DRAVOSBURG
EDGEWOOD
ELIZABETH
EMSWORTH
GLASSPORT
GLENFIELD
HAYSVILLE
INGRAM
LEETSDALE
LINCOLN
MILLVALE
OAKDALE
PLUM

Figure 26 Municipalities without Dataset (Few Entries)

5 DISCUSSION

As per our analysis, we identified various hospitals in different municipalities across Pennsylvania and the number of licensed beds available in each of those hospitals. We also used K-means clustering to identify 3 clusters. Out of the 3 clusters, Cluster 2 has least number of areas, which means it is currently better equipped to fight against such global pandemic. Similarly, Cluster 3 has the maximum number of municipalities with low bed counts, hence the PA Government should target improving the preparedness level of those facilities, Also, there are around 7 Counties (10%) and 526 Municipalities (80%) with no medical facility with licensed beds available. This is a major concern and needs to be addressed urgently. Although the areas without medical facilities are low in population, but in a scenario of global health crisis we need to ensure proper medical facilities within the proximity of each municipality.

Improvement Scope

- **Foursquare API** does not contain the entire list of hospitals available in Pennsylvania and hence misses on many municipalities.

- The **PA Hospital Report** is based in 2018 and needs to be updated with current bed counts based on any new renovation/expansions.
- Out of 136 Municipalities only 71 Municipalities matched with the hospital data **using Fuzzywuzzy**. This may be incorrect as Fuzzywuzzy uses similar words to match and may lead to incorrect entries.
- **Only Hospital data is considered** in the current scope of analysis and it misses on other medical facilities like small clinics and nursing homes.
- The population data is also based on 2018 figures and needs to be updated as per the latest census.

6 CONCLUSION

Exploratory Data Analysis was performed on the collected County, Municipality and Hospital Data to identify the preparedness levels of each municipality for a global health crisis or a pandemic. Several statistical analysis were performed to highlight the availability of hospital beds in each county and municipalities.

Clustering Techniques were used to classify the data into different groups of optimal areas and identified regions having the highest bed counts per 100 people. The clusters were also graphically visualized to understand the locations of different clusters and which municipalities are in which cluster. The region Jefferson, Greene is observed to have the highest bed counts per 100 people in the Pennsylvania State and hence is the most well equipped to battle global health crisis.