

CAPSTONE  
COURSERA

# PENNSYLVANIA PREPAREDNESS LEVEL FOR A PANDEMIC

PREPARED BY: ROHIT RAWAL

IBM Data Science  
Certification





**INTRODUCTION**  
**DATA**  
**METHODOLOGY**  
**RESULTS**  
**DISCUSSION**  
**CONCLUSION**

# **PRESENTATION OUTLINE**

**IBM DATA SCIENCE  
CERTIFICATION**



# 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.

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.

## KEY GOALS

- Evaluating the preparedness level among the different counties and municipalities in Pennsylvania to handle a health crisis of such scale
- Analyze the hospital utilization across different counties and municipalities to understand the areas that are better equipped with adequate medical services
- 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.
- Use K-means clustering to visualize and identify the clusters having different levels of medical services (hospital beds per given population).

# 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



# METHODOLOGY



## Data Collection

PA County,  
Municipalities Data  
Population Data  
Hospital Beds Data  
for every County in  
PA  
Hospital Data for  
each Municipality  
using Foursquare  
API



## Exploratory Data Analysis

Data Preprocessing  
Data Visualization  
Statistical Analysis



## Data Classification

K-Means  
Clustering  
Parameter  
Optimization



## Graphical Visualization



## Insights and Conclusions

# DATA COLLECTION

CAPSTONE  
COURSERA

	COUNTY	Longitude	Latitude
0	ADAMS	-77.222243	39.872096

	COUNTY	FACILITY NAME	LICENSED BEDS
0	ADAMS	GETTYSBURG HOSPITAL	76

	COUNTY	MUNICIPALITY	lat	lon
0	ADAMS	ABBOTTSTOWN	39.886487	-76.984701

	COUNTY	Pop	Area[1](mi2)	Pop. dens.(p/mi2)	MUNICIPALITY
27	MONTGOMERY	34,427	3.6	9541.9	NORRISTOWN

	COUNTY	MUNICIPALITY	Pop	lat	lon
45	A				EVILLE

	COUNTY	FACILITY NAME	LICENSED BEDS	COUNTY	MUNICIPALITY
0		GETTYSBURG HOSPITAL	76	ADAMS	GETTYSBURG

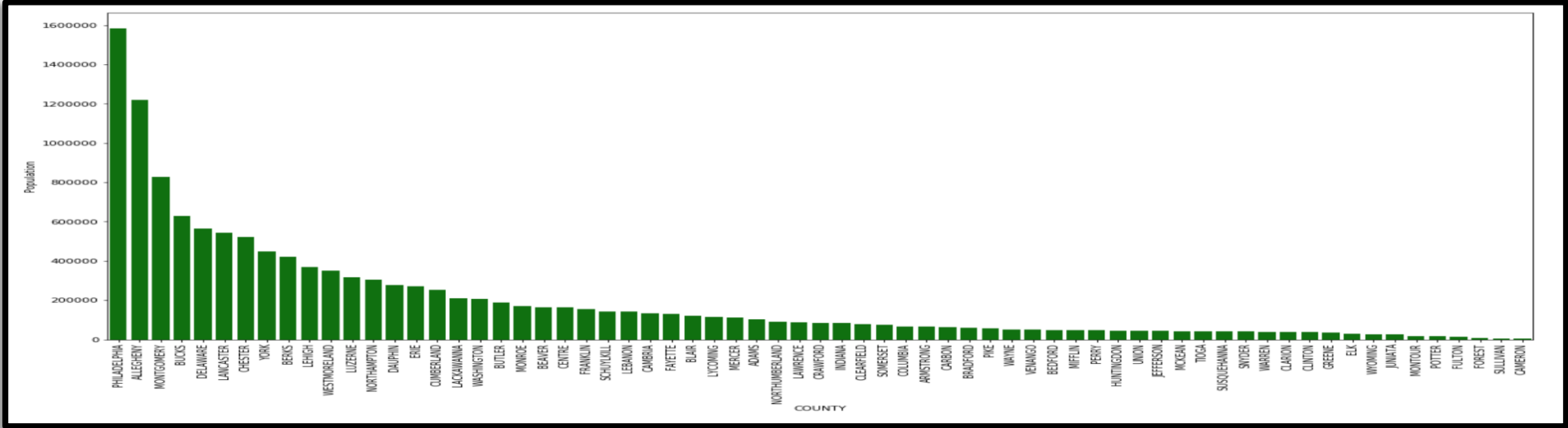
	COUNTY	MUNICIPALITY	LICENSED BEDS	Pop	lat	lon
0	ADAMS	GETTYSBURG	528	7645	39.830897	-77.231012

	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

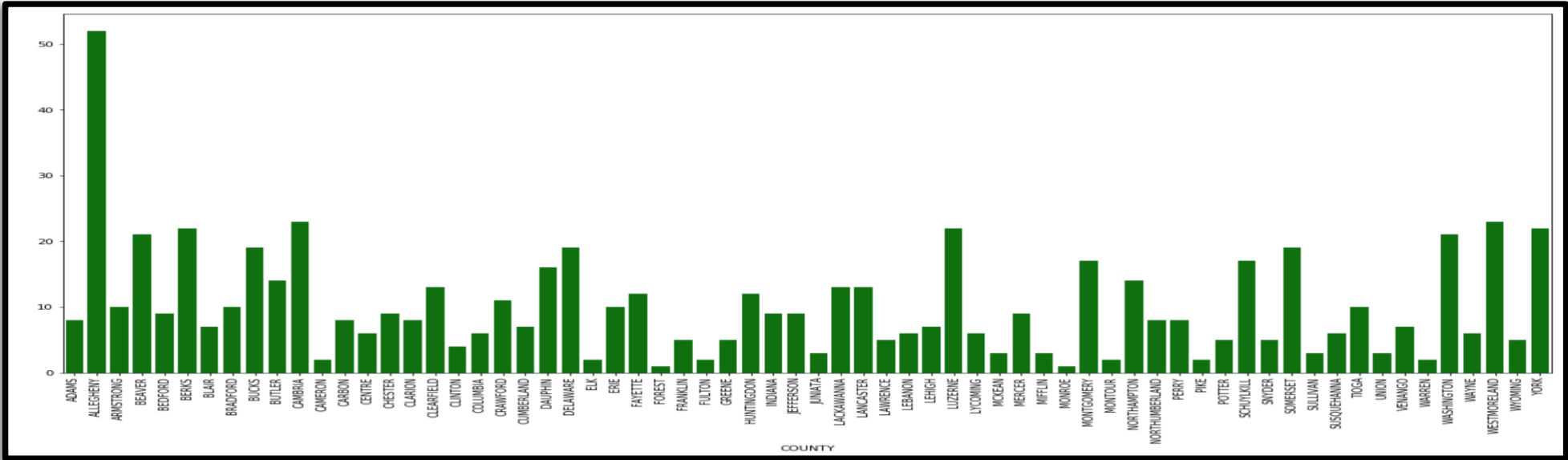
# EXPLORATORY DATA ANALYSIS

CAPSTONE  
COURSERA

County  
Population  
Count



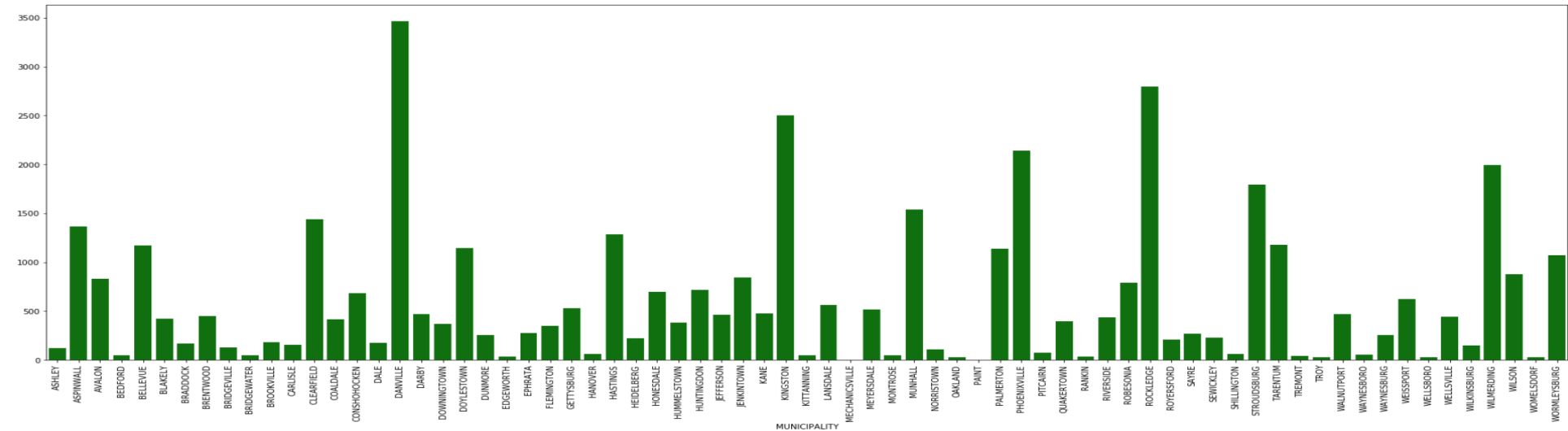
Municipalities  
Count



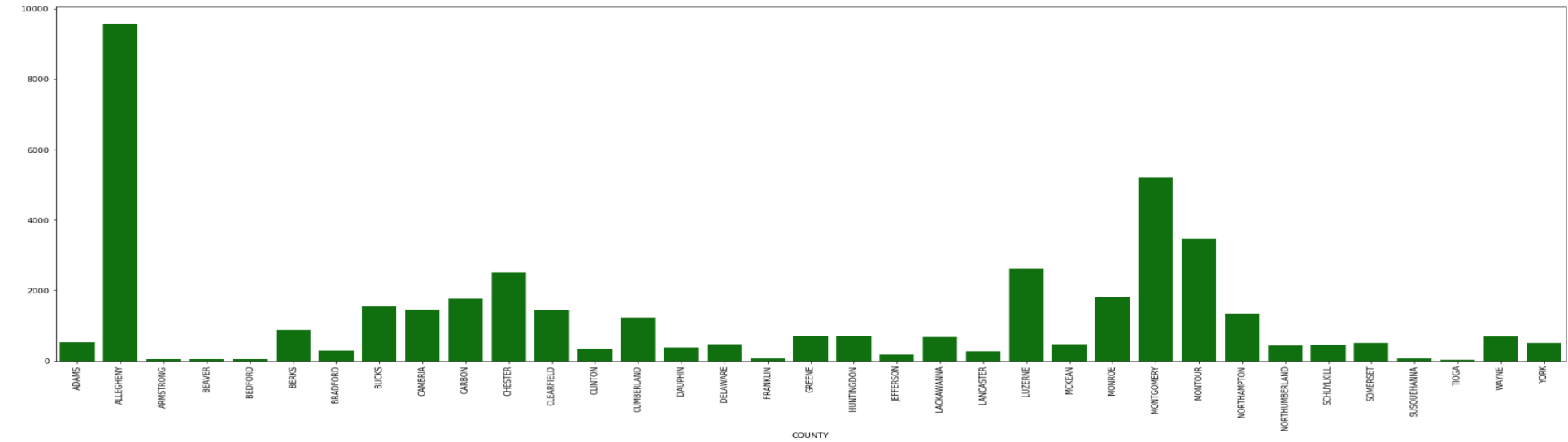
# EXPLORATORY DATA ANALYSIS

CAPSTONE  
COURSERA

Municipality  
Hospital Bed  
Count



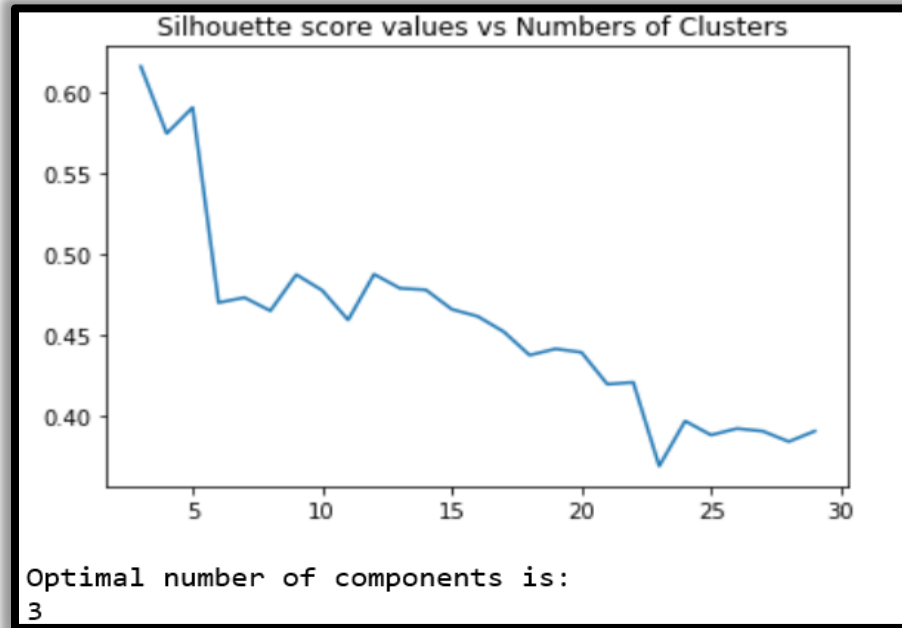
County  
Hospital Bed  
Count





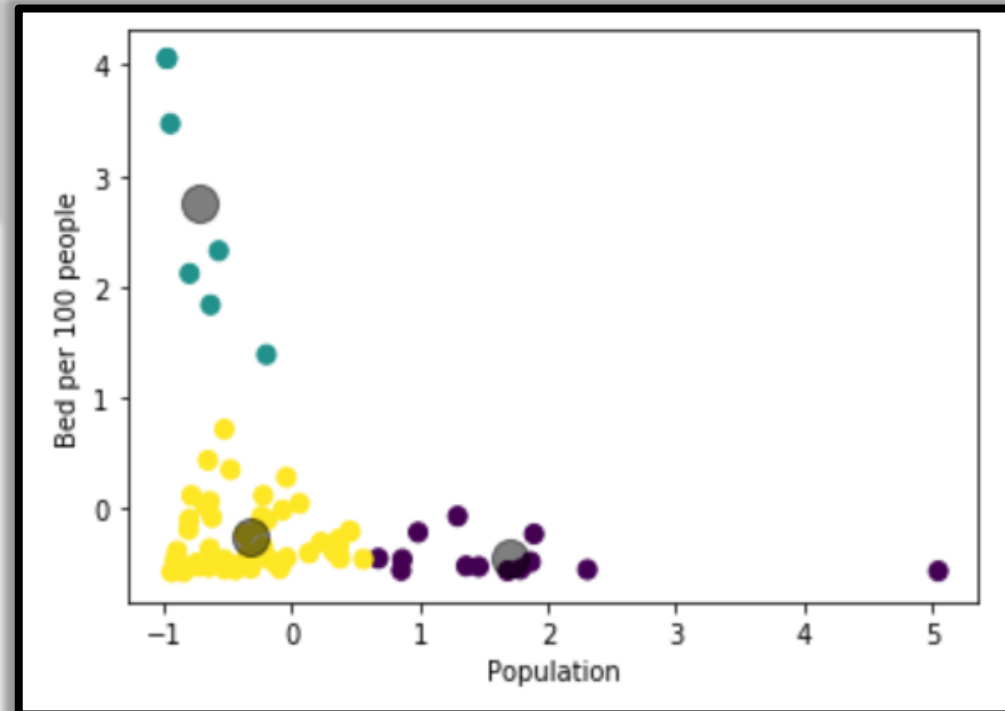
# K-MEANS CLUSTERING

CAPSTONE  
COURSERA



**Elbow Method –  
Parameter Optimization**

**Centroid  
Locations**

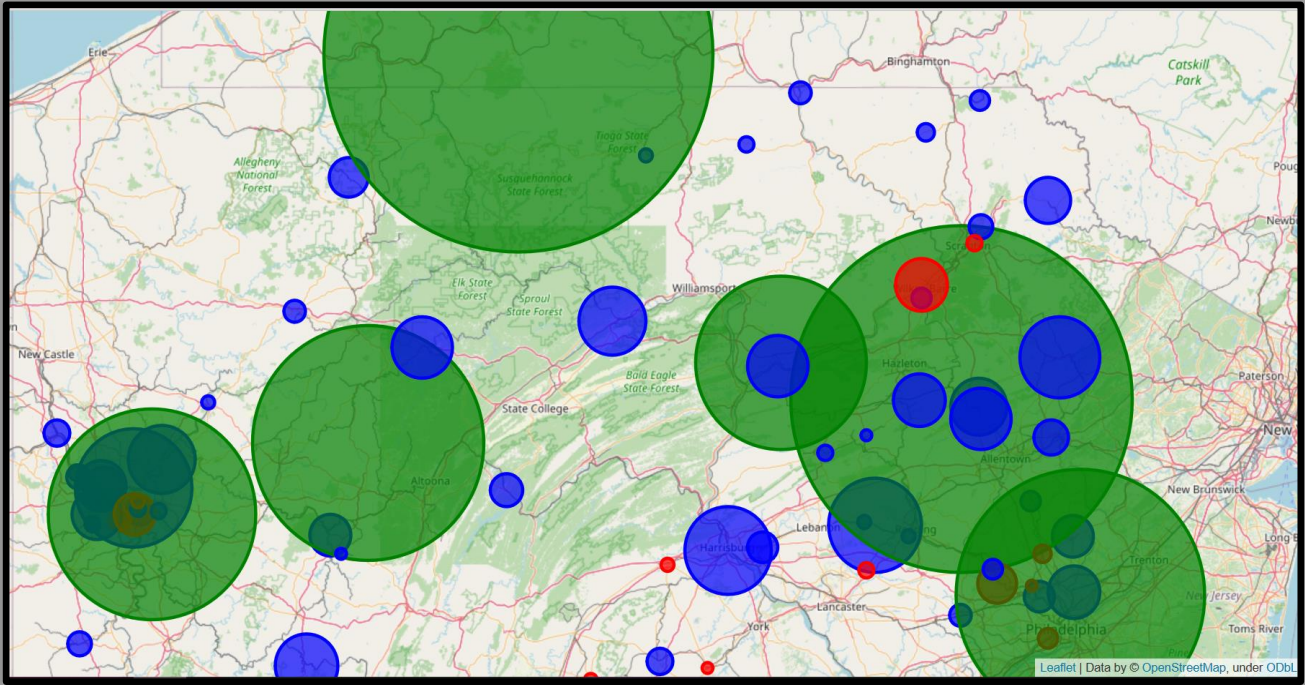


# K-MEANS CLUSTERING

CAPSTONE  
COURSERA

	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

Dataset with  
Cluster Labels



Graphical  
Cluster  
Representation

# RESULTS || CLUSTER ANALYSIS

CAPSTONE  
COURSERA

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

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

# RESULTS || CLUSTER ANALYSIS

CAPSTONE  
COURSERA

## 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



# RESULTS || WITHOUT HOSPITAL

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

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

PIKE  
PERRY  
SNYDER  
JUNIATA  
FOREST  
SULLIVAN  
CAMERON

ARENDTSVILLE  
BENDERSVILLE  
BIGLERVILLE  
BONNEAUVILLE  
FAIRFIELD  
LITTLESTOWN  
MCSHERRYSTOWN  
BALDWIN  
CARNEGIE  
CHALFANT  
CHESWICK  
CHURCHILL  
CORAOPOLIS

CAPSTONE  
COURSERA

**Counties  
without  
Hospitals**

**Municipalities  
without  
Hospitals**

# DISCUSSION

- ❑ 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.
- ❑ Cluster 3 has the maximum number of municipalities with low bed count. 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 hospitals 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 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.



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.

# CONCLUSION

**IBM DATA SCIENCE  
CERTIFICATION**



# THANK YOU

Rohit Rawal ✉ [rur281@psu.edu](mailto:rur281@psu.edu)