Optimal Locations to Build Additional Hospitals in New Jersey

IBM Data Science Capstone Presentation

By: Michael Manian

Adding New Hospitals Helps Saves Lives

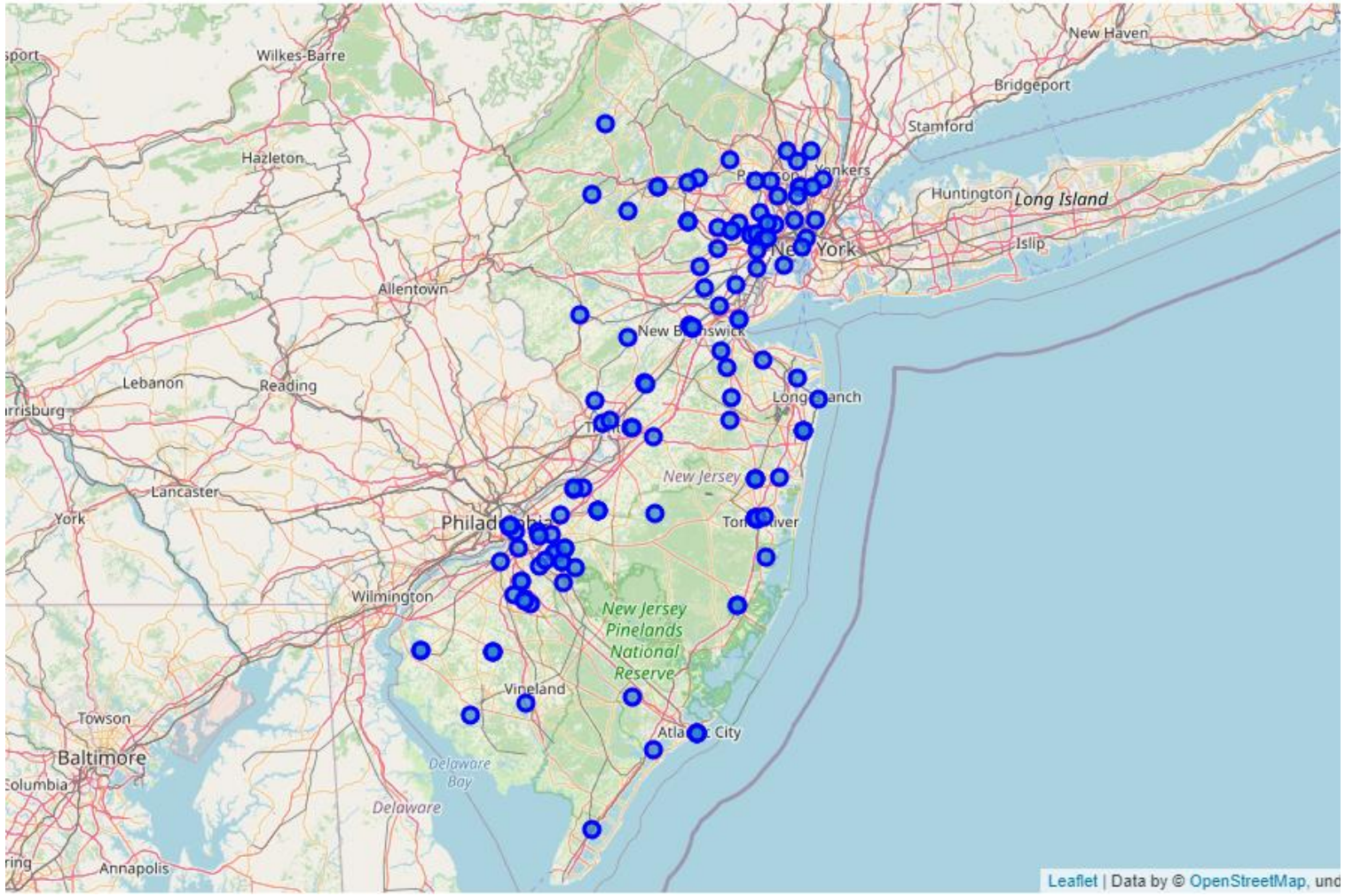
- Hospitals are an integral part to society. They serve to provide treatment and help cure illnesses all around the world.
- Over the years, technology and science have discovered over 10,000 diseases that could affect humans and most of those diseases do not have a cure but has many treatments.
- In order to survive global pandemics, there must be enough medical supplies and hospitals to take care of those affected. The current COVID-19 pandemic has proved that the United States does not have enough hospitals to handle severe pandemics.
- Adding new hospitals in New Jersey would be of interest to the Department of Health and Human Services, considering they are the ones in charge of enhancing the health and well-being of all Americans through providing effective health and human services. It would also be of interest to any fresh doctors and nurses looking for a job.

Data Used

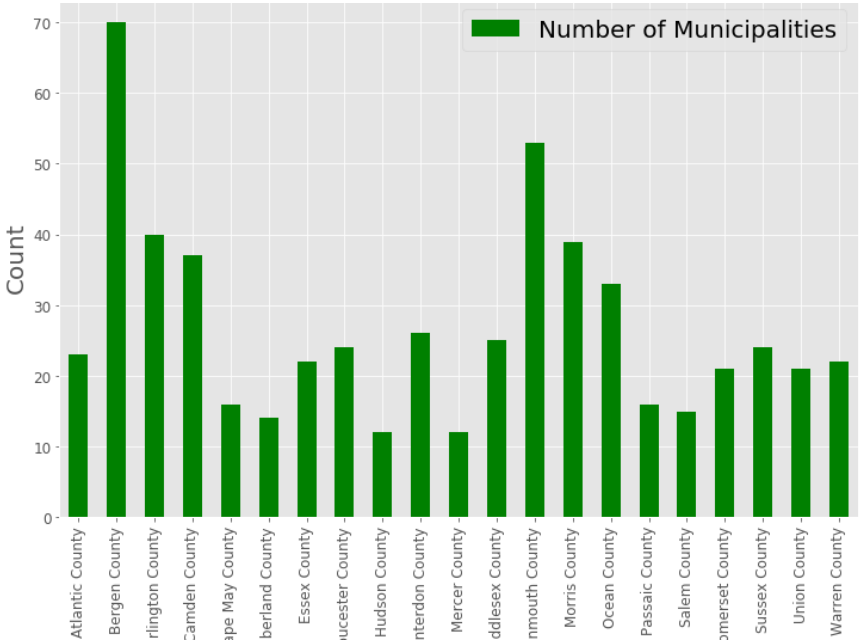
- Basic data about New Jersey, including counties and municipalities was scrapped from **Wikipedia**
- Population data per county as of early 2020 and median income data per county as of 2018 was also scrapped from **new jersey demographics**, and **Indexmundi**, respectively.
- **Python's Geopy library** was used to locate the geographical coordinates of each county.
- **Foursquare API** was used to gather all the hospitals located near each county.
- Data was stored in a Pandas dataframe.
- Duplicate rows and columns were dropped along with excess information which will not help formulate insight.

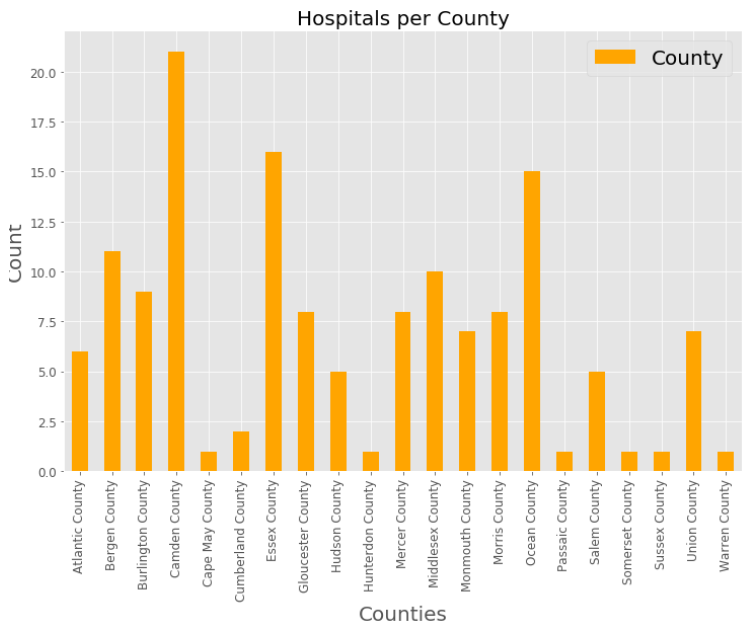
Machine Learning Techniques

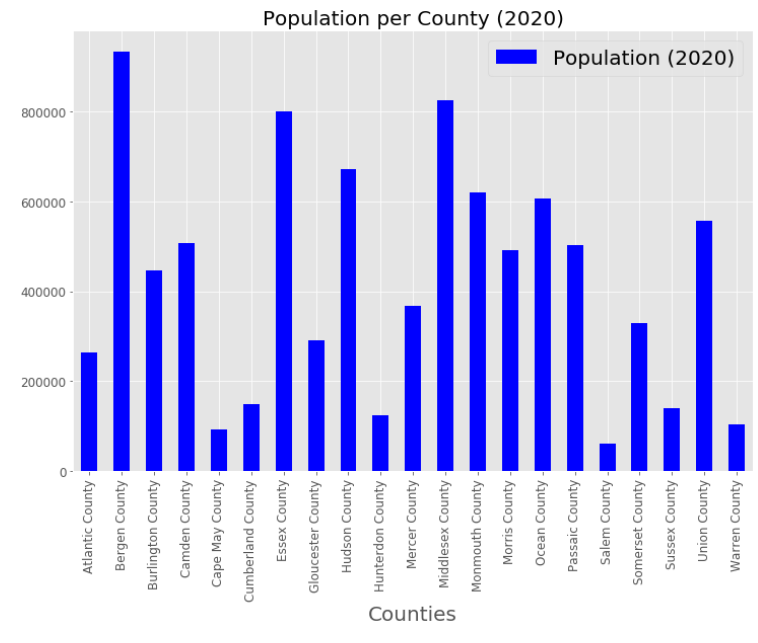
- Normalize features by using preprocessing and StandardScalar libraries in order to scale, fit, and transform the data.
- K-means unsupervised machine learning algorithm was used to cluster data sets.
- Silhouette score method was used to find the optimal k clusters to ensure model is as accurate as it should be.

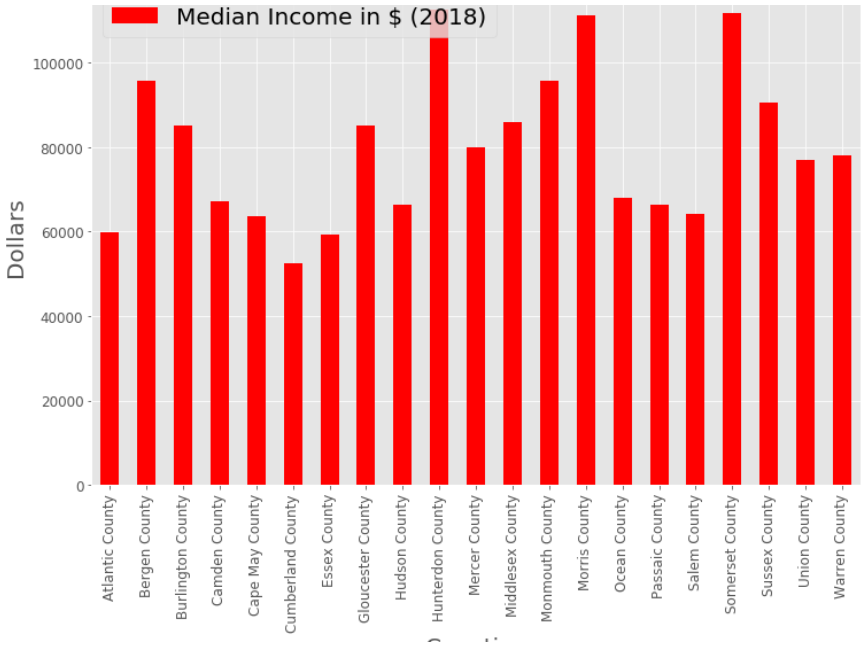


Map of the Hospitals in NJ



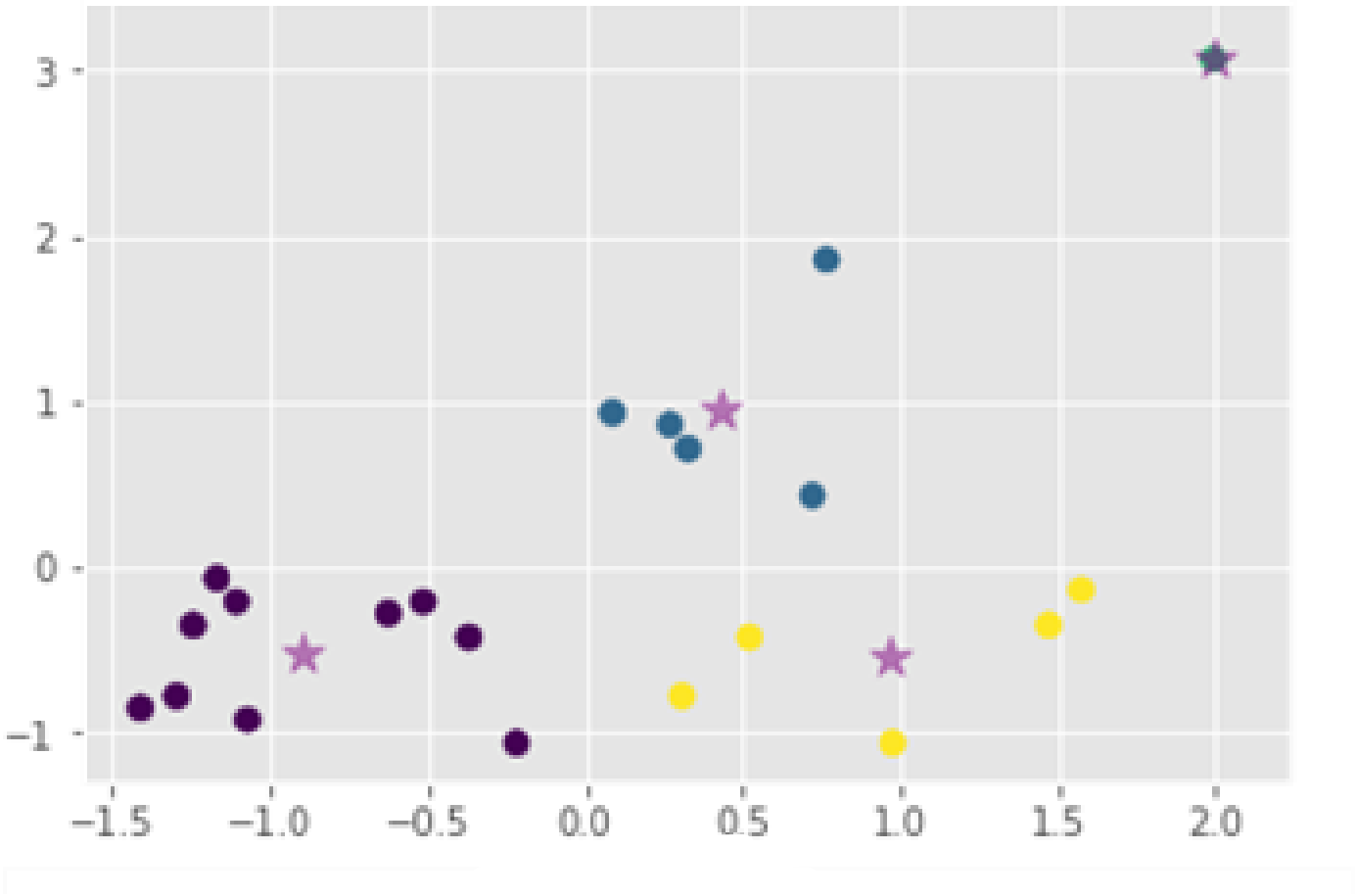






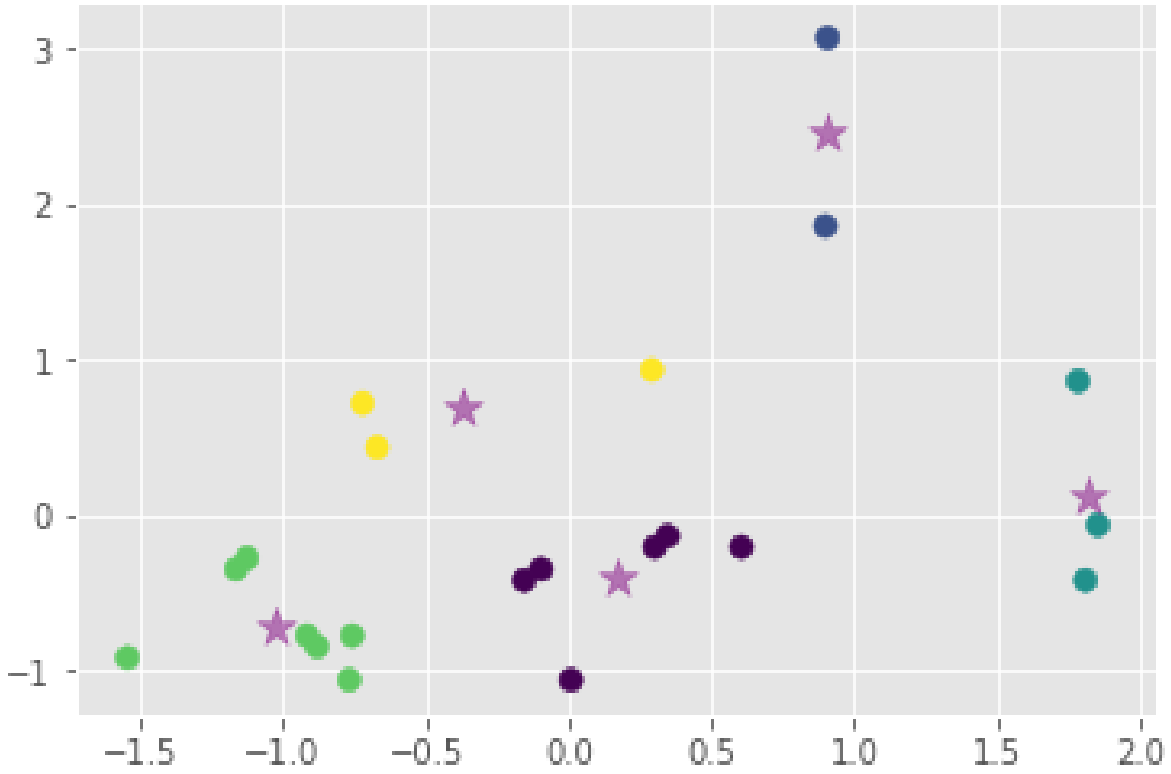
Quantitative Data Visual

Population vs # of Municipalities



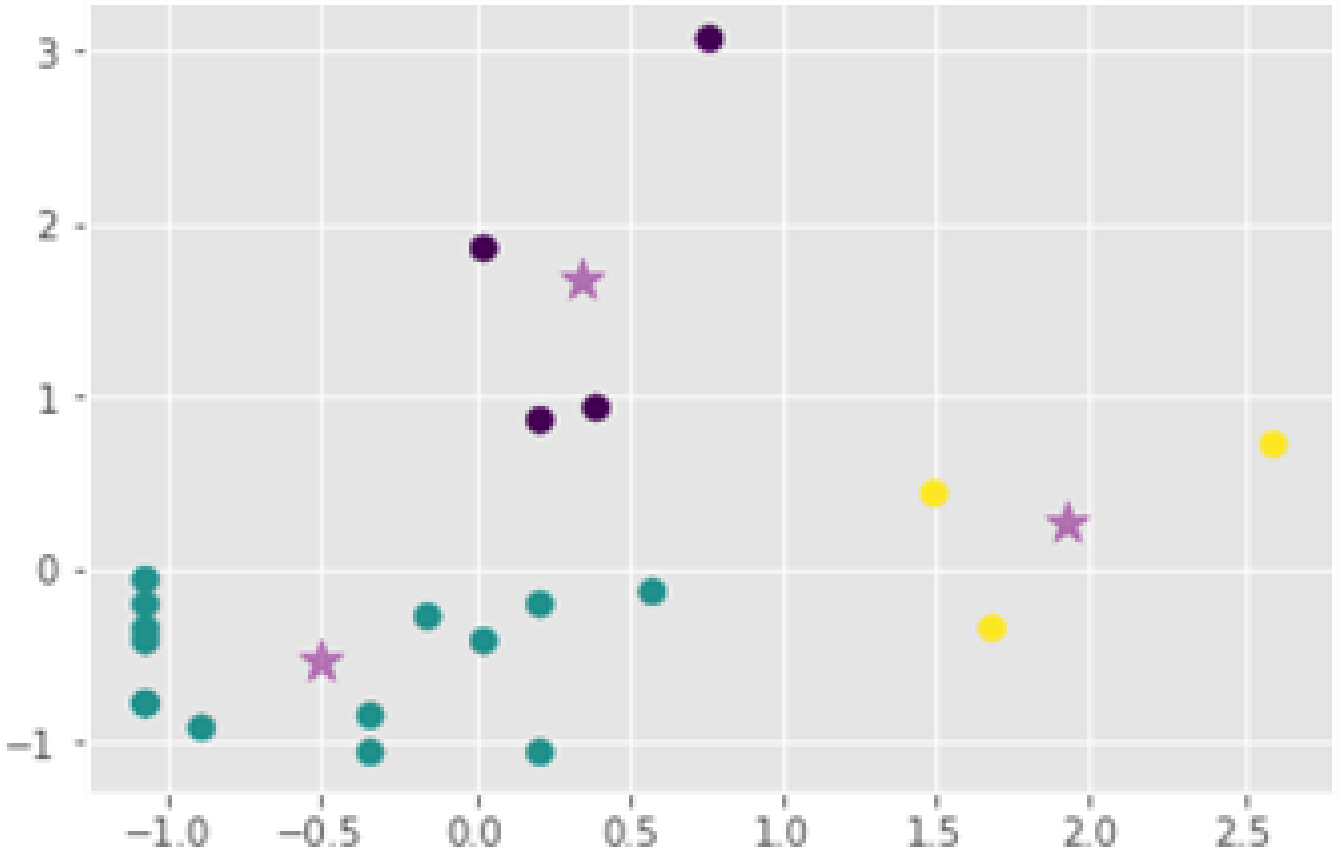
- Strong positive linear relationship.
- Higher the population, the higher the number of towns.
- Lower the population, the higher the lower of towns.
- Yellow cluster is the outlier- high population, low number of towns (Essex, Hudson, Middlesex, Passaic, Union).

Median Income vs # of Municipalities



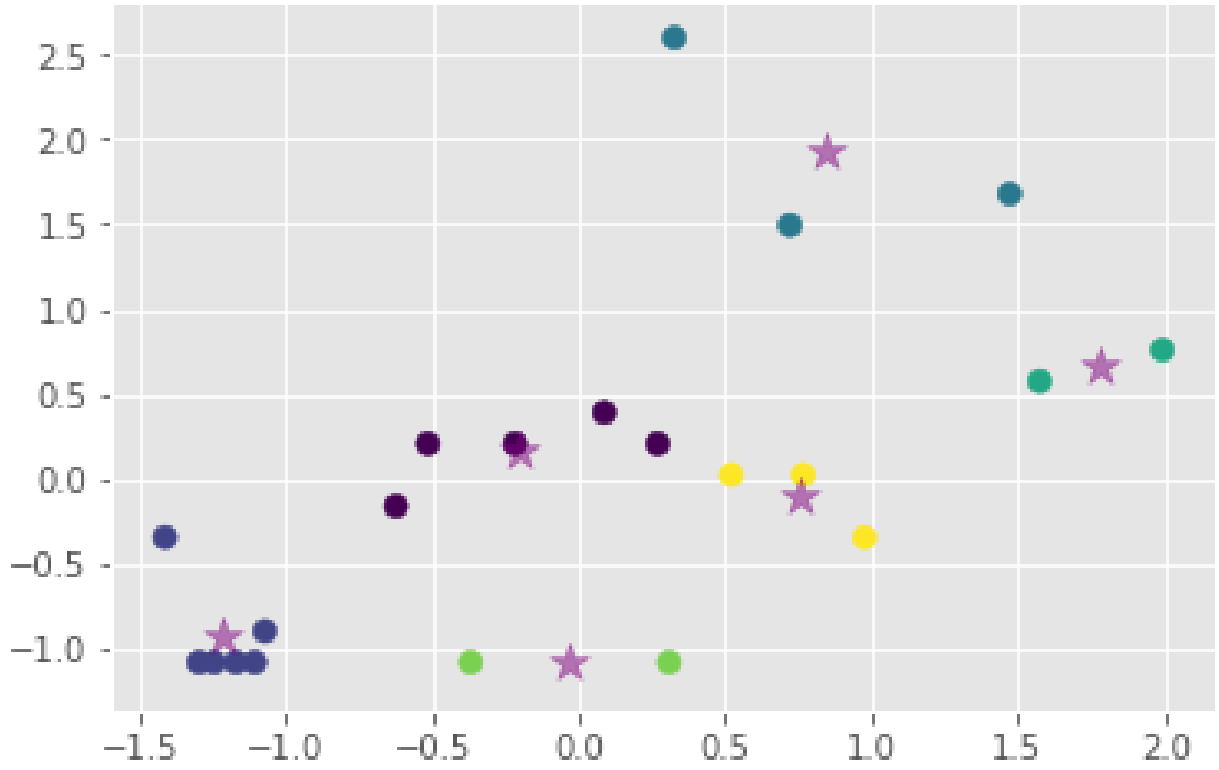
- Somewhat positive linear relationship.
- People with lower income resides in counties with not as many towns.
- Light blue cluster is the outlier- high income residing in counties with an average/low amount of towns (Hunterdon, Morris, Somerset).

of Hospitals vs # of Municipalities



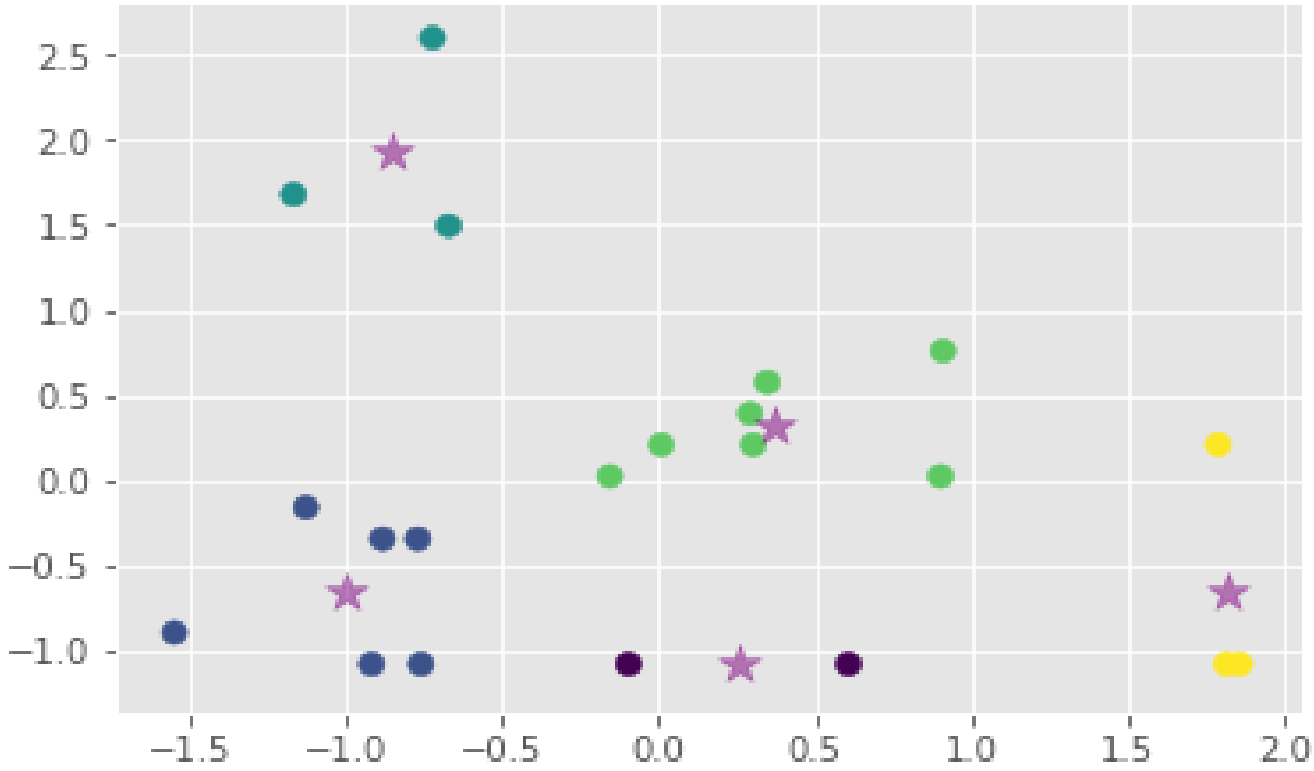
- Strong positive relationship.
- Less hospitals reside in counties with less towns.
- Purple cluster is the outlier- towns to hospitals ratio doesn't add up compared to the other counties in the other clusters (Bergen, Burlington, Monmouth, Morris).

Population vs # of Hospitals



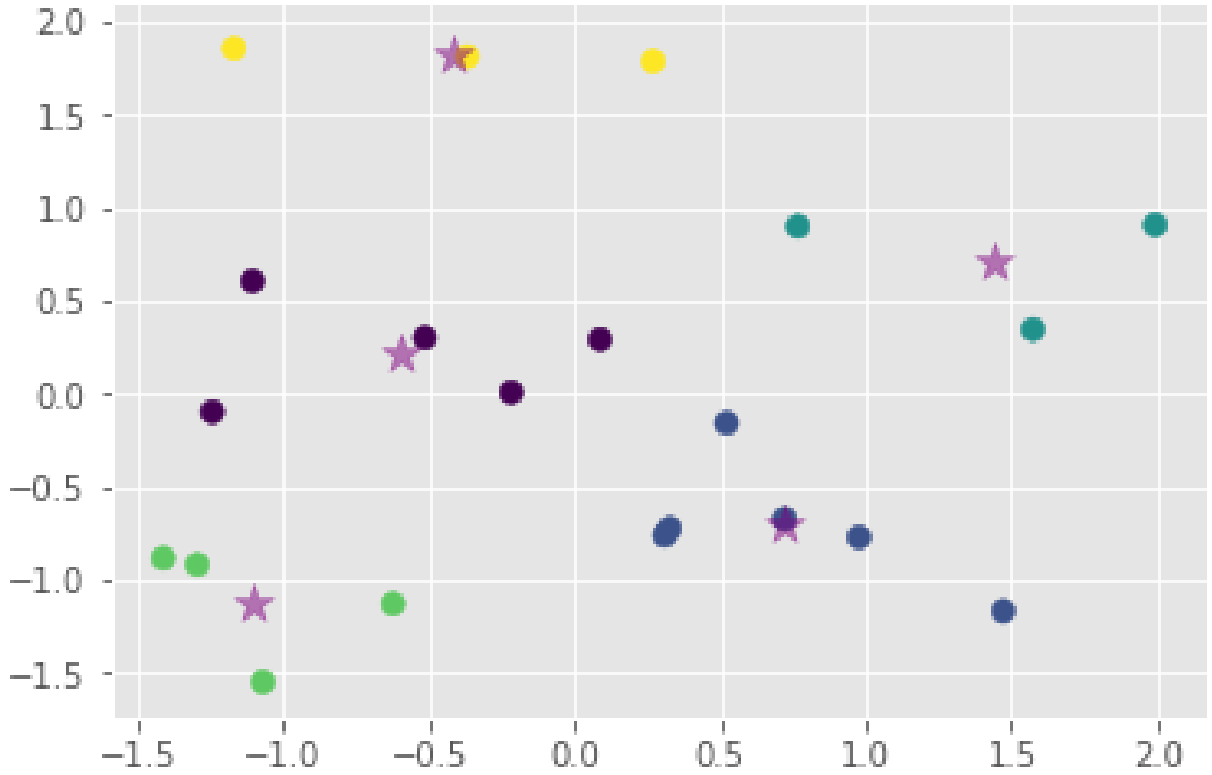
- Towns with lower population do not have as many hospitals.
- The green cluster is an example of average population but not as many hospitals- we can build additional hospitals here (Passaic and Somerset).

Median Income vs # of Hospitals



- Hard to agree on a correlation here.
- Light blue cluster and yellow cluster potential outliers- both extremes of the spectrum (Camden, Essex, Ocean) and (Hunterdon, Morris, Somerset).

Population vs Median Income



- Hard to agree on correlation here as well.
- Yellow cluster potential outlier- low population yet the highest median income (Hunterdon, Morris, Somerset).

Conclusion

- Summing up all the data and gathering all the outlier clusters and comparing them with each other, I have produced these five counties that would benefit from an addition of a new hospital:
  - Morris County
  - Somerset County
  - Passaic County
  - Middlesex County
  - Hunterdon County

Full in dept report can be found here:
https://github.com/mrmanian/Coursera_Capstone/blob/master/Capstone%20Project%20Report.pdf

Model Improvement

- This model can be improved by implementing some socioeconomic data as well as data from the COVID-19 pandemic, but they were more difficult to extract.
- Also analyzing the towns rather than the counties would have provided a better model however there are 565 towns in New Jersey which would have been difficult to implement especially with the limitations of Foursquare and Geopy.
- With this additional data and more accurate hospital readings/coordinates, we would be able to visualize a regression line on the scatter plots much easier since the points would be much closer together.