

*Finding the Best  
County in California  
State to Fight the  
Pandemic*

# Data Acquisition and Cleaning

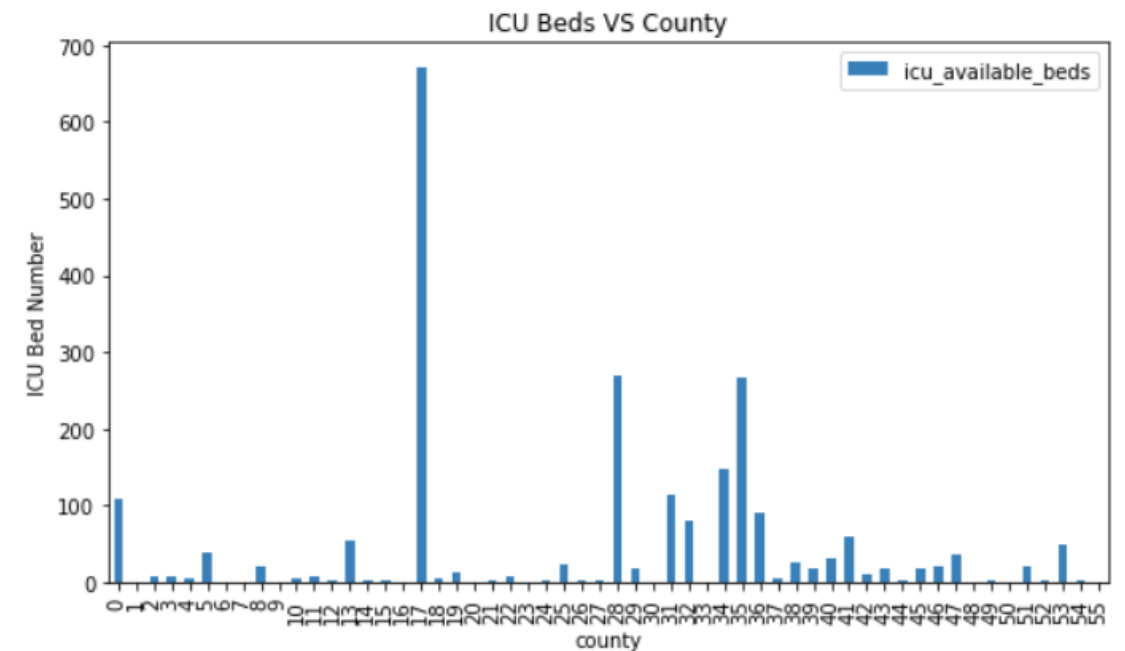
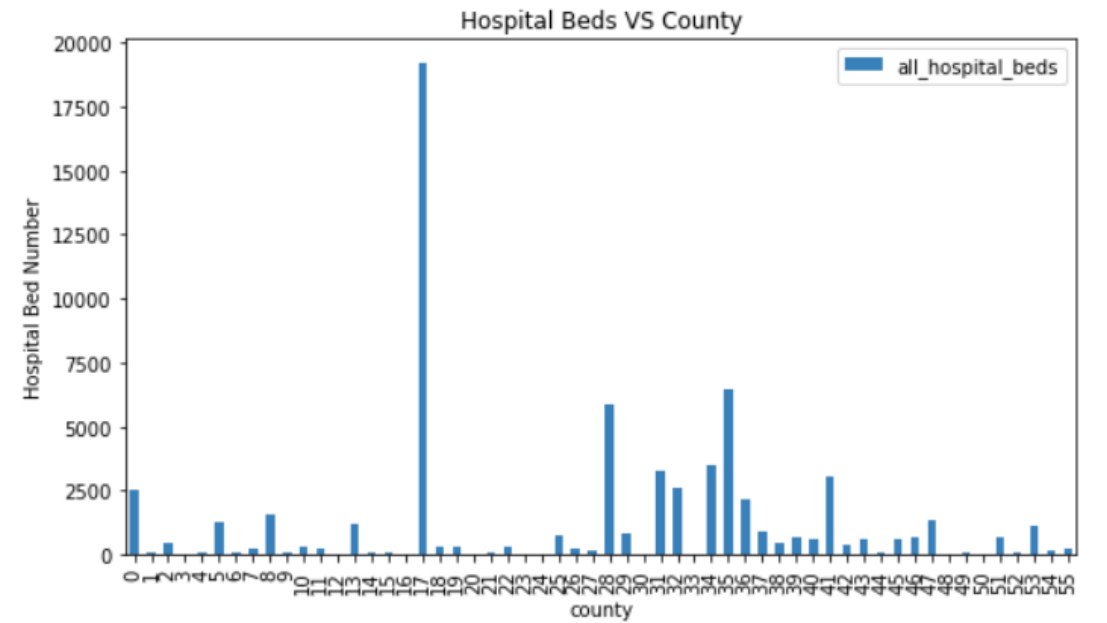
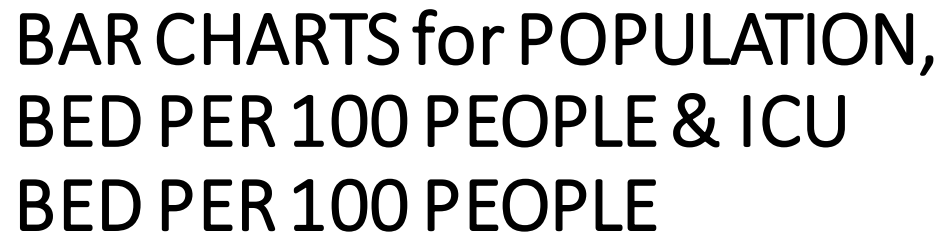
## Data Sources:

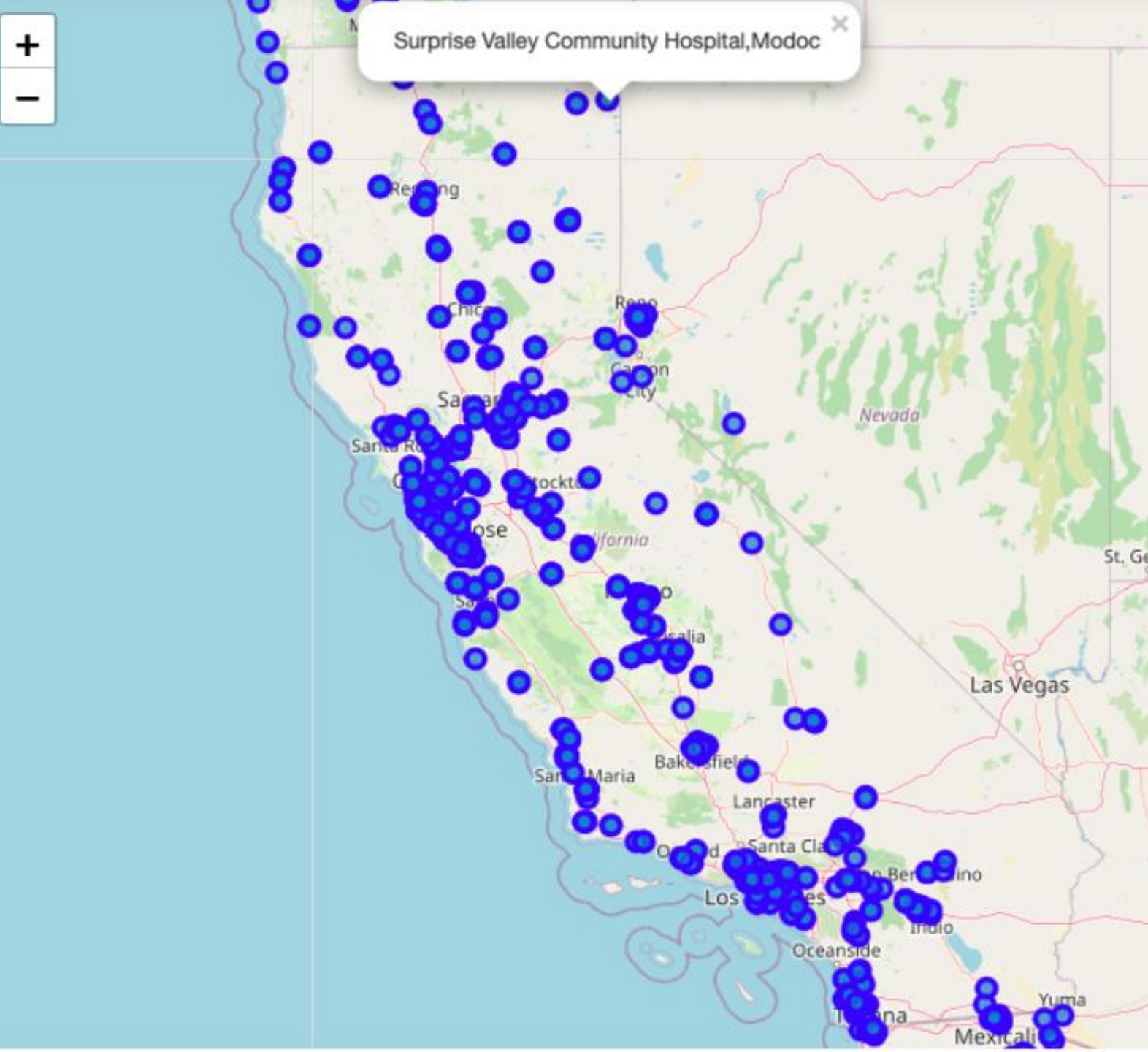
- California State Dataset with Counties, Number of Hospital Beds and ICU Beds.
  - source: [CA data set](#)
- California Counties Population data
  - source: [worldpopulationreview.com](http://worldpopulationreview.com)
- Hospitals Information (Names, Coordinates)
  - source: **FOURSQUARE API**
- Counties coordinates
  - source: [simplemaps](#)
  - Alternative source: **FOURSQUARE API** (county key: 5345731ebcbc57f1066c39b2)

## Data Cleaning:

Data downloaded or leveraged from multiple sources were combined into one table. And we sort the table alphabetically based on County names. **Alpine & Sierra counties** are excluded based on low population. So only **56 counties** are considered out of 58.

California hospital beds dataset contained data dated from **march 29, 2020**, I've cleaned most of the data and considered data only dated as of **June 1, 2021**. After merging different tables, we include two more columns namely **Bed\_per\_100\_people** and **ICU\_Bed\_per\_100\_people**. In total **56 rows** and **8 features** are considered.

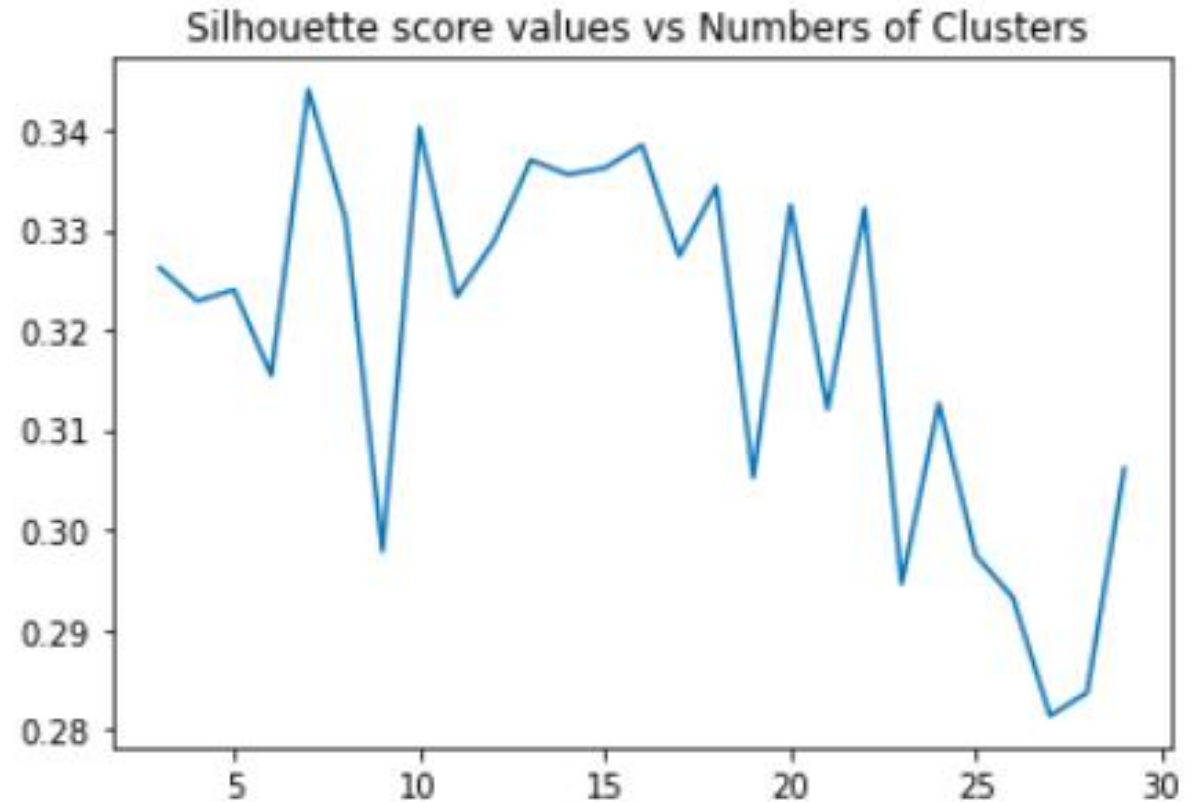




HOSPITALS  
MAPPED FROM  
FOURSQUARE  
DATA  
(total 2245  
hospitals)

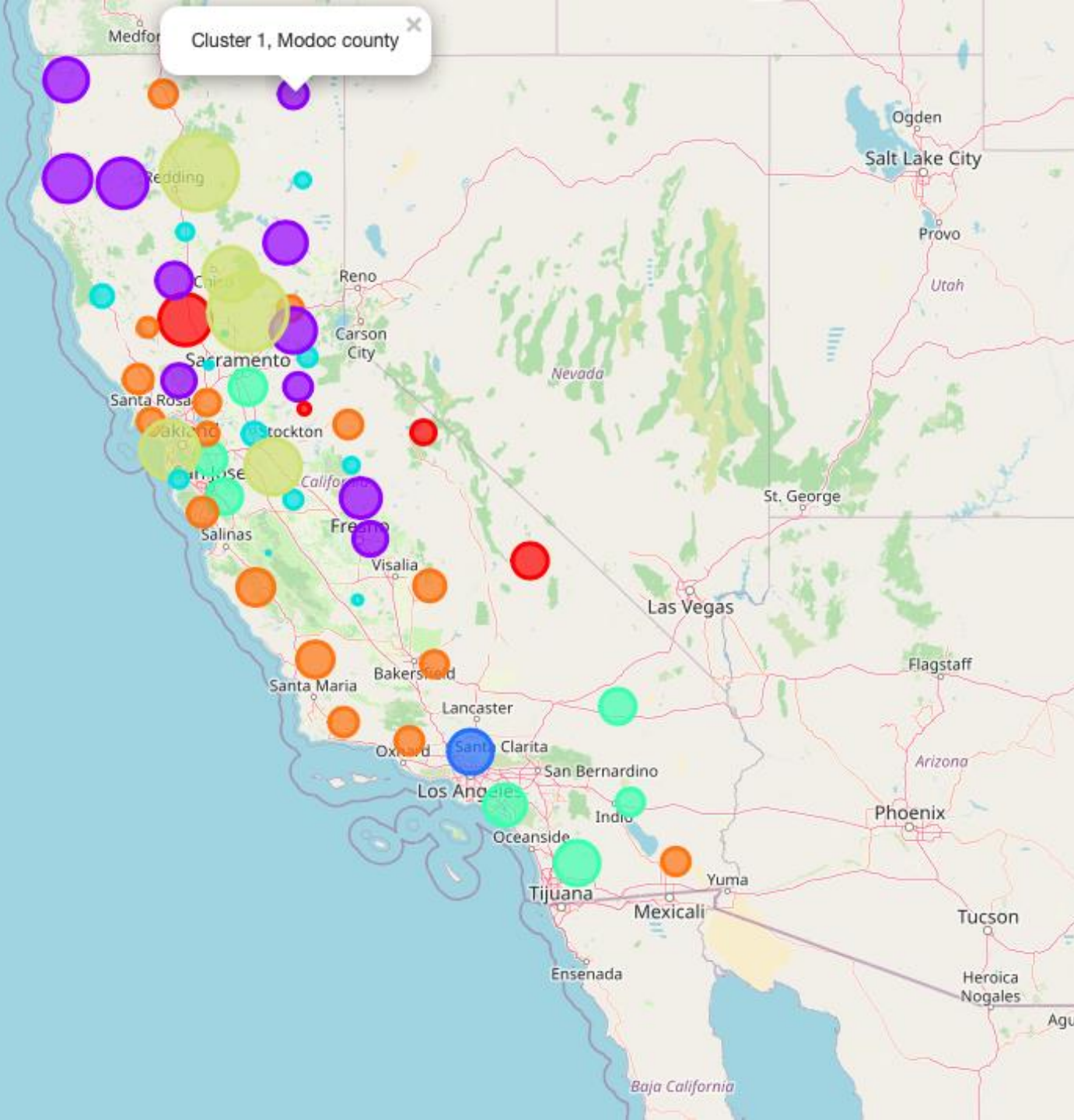


SILHOUETTE  
SCORE FOR  
IDEAL  
NUMBER OF  
CLUSTERS

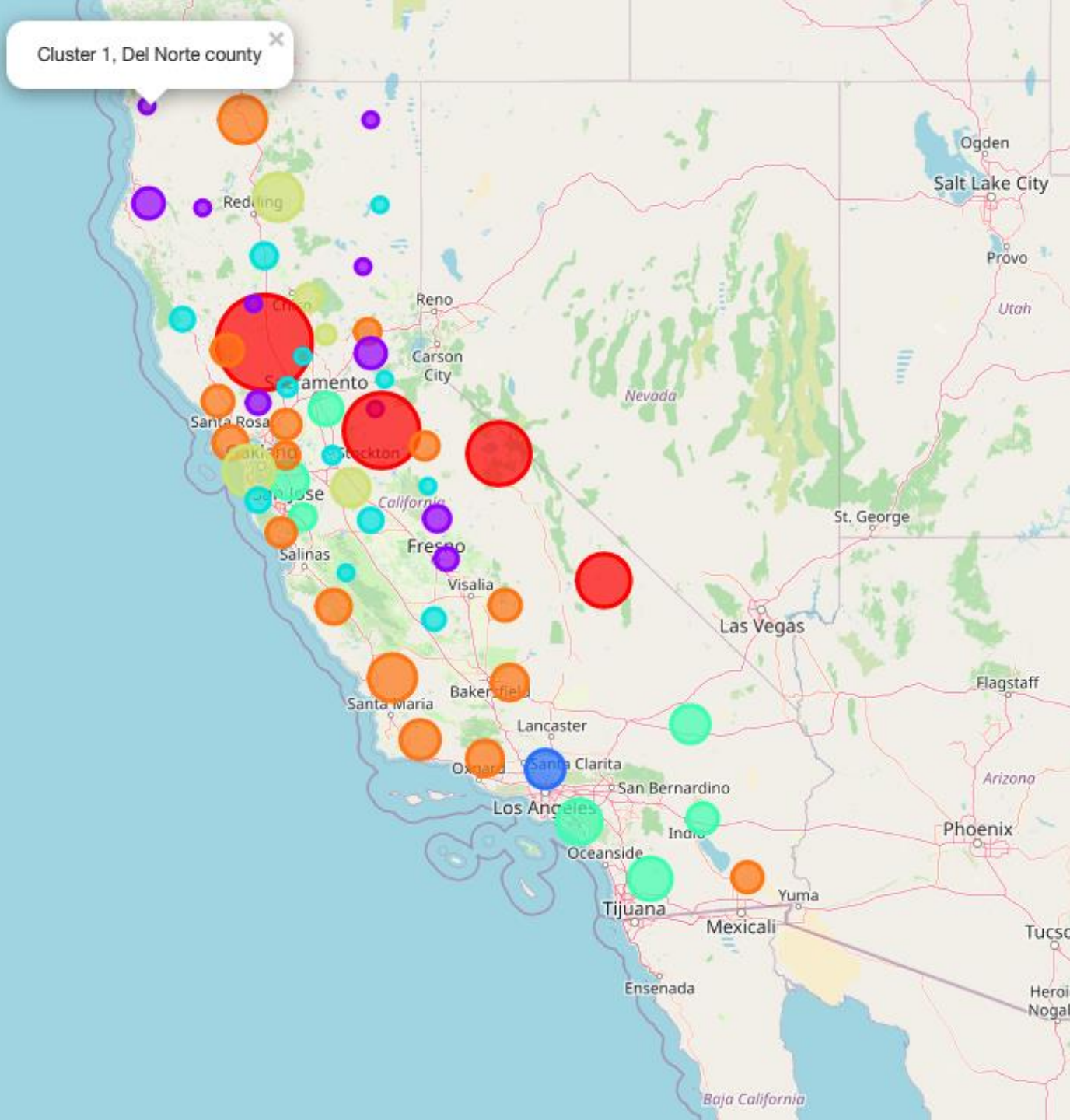


Optimal number of components is: 7

The score suggests us to have **7** clusters



CLUSTER MAP  
based on BED  
PER 100  
PEOPLE

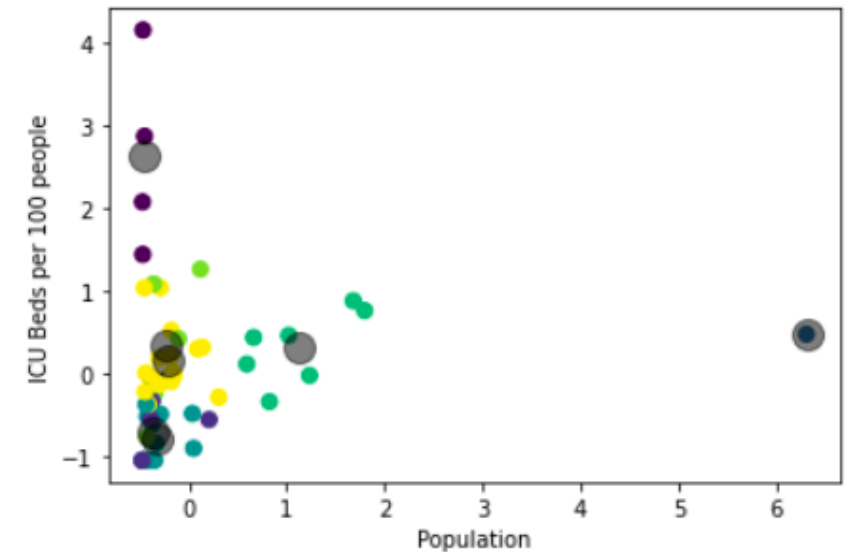
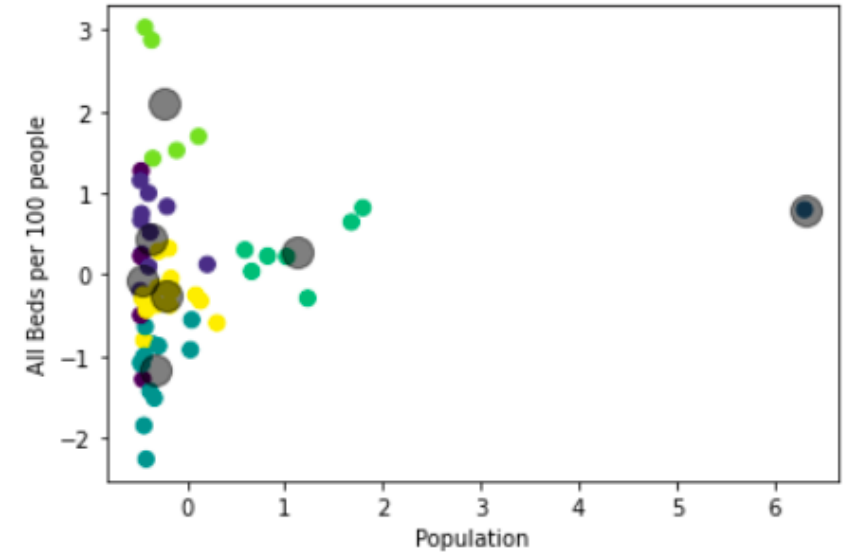


# CLUSTER MAP based on ICU BED PER 100 PEOPLE

# SCATTER PLOTS

We can observe the outliers here. In the first plot, we see the **top green circle outlier** which is **Yuba County** with **best bed per 100 people ratio** and the next green marker below Yuba County is **Shasta County** which has the **next best bed per 100 people ratio**. The other **black circle outlier** in first scatter plot is **Los Angeles County** because of its **high population**.

And coming to the second scatter plot the **purple circle outlier** is **Colusa County** which has the **best ICU bed per 100 people ratio** and the **black circle outlier** is **Los Angeles County** due to its **high population**.





# RESULTS AND CONCLUSION

---

During the analysis, a total of **7 clusters** were defined. **Single cluster** (cluster 3), which is **Los Angeles County**, is an **outlier** because of its **huge population** compared to others. **Yuba County** and **Colusa County** have the **best bed per 100 people** and **ICU bed per 100 people ratios respectively**. And they too are **also outliers** in our scatter plots for the above reason. There are **11 counties** with **no ICU beds** and they are from **clusters 2 and 4**, which means they need to concentrate on providing more emergency treatment (like ICU Beds) and also 10 out of 11 falls in the category with the least beds per 100 people (the exception being El Dorado County with 190 beds). **Los Angeles** is also an outlier because of its high population but has only **0.192447 bed per 100 people** and **0.00672 ICU bed per 100 people**, which means they need to improve a lot to provide more beds to fight the pandemic.

Finally, to conclude, basic exploratory data analysis was performed to identify the well-equipped county in California State. During the analysis, important features like **counties**, their **population** and **total beds available** and **ICU beds available** as of **June 1, 2021** were considered. And were clustered based on the above features. **Yuba County** has the best bed per 100 people ratio and **Colusa County** has the best ICU bed per 100 people ratio. **Los Angeles**, an outlier due to its high population has lot of scope for providing more beds.