

DATA ACQUISITION and CLEANING:

We will be collecting data from the following sources:

Data Sources:

1. California State Dataset with Counties, Number of Hospital Beds and ICU Beds.
 - source: [CA data set](#)
2. California Counties Population data
 - source: [worldpopulationreview.com](#)
3. Hospitals Information (Names, Coordinates)
 - source: **FOURSQUARE API**
4. 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 columns namely **Bed_per_100_people** and **ICU_Bed_per_100_people** and it resulted in the following final data frame:

```
3 | cal_df.head()
```

county	all_hospital_beds	icu_available_beds	population	lat	lng	Bed_per_100_people	ICU_Bed_per_100_people
Alameda	2493.0	110.0	1680480	37.6469	-121.8889	0.148350	0.006546
Amador	53.0	0.0	40446	38.4464	-120.6511	0.131039	0.000000
Butte	451.0	7.0	196880	39.6669	-121.6007	0.229074	0.003555
Calaveras	33.0	8.0	46319	38.2046	-120.5541	0.071245	0.017272
Colusa	48.0	5.0	21805	39.1775	-122.2370	0.220133	0.022931