MEMBER TASK

Lahari Chilakuri, Graduate Student

Hospital Beds Dataset –

Description:

The enrichment dataset I chose to analyze is Hospital Beds dataset. The size of the dataset as of today (i.e., 02/15/2022) is 399863 rows × 109 columns. This dataset provides the information about the information about the hospitals, total number of beds and the available ICU units in the hospital and their average usage in all the counties across USA.

Datatype Variable Dictionary:

Name	Datatype	Description
ccn	Integer	Certification number of hospitals.
hospital_name	Object	Name of the hospital
address	Object	Address of the hospital
city	Object	City in which the hospital is located.
Zip	Object	Zip code of the hospital's physical address.
Hospital_subtype	Object	Type (CA, Children's, long term etc..) of the hospital to which the observation belongs
fips_code	Integer	Unique ID used to identify the observations from county.
total_beds_7_day _average	Float	The average number of total beds in seven days
total_icu_beds_7_ Day_average	Float	Average Total number of ICU beds in seven days
total_adult_beds_7_ Day_average	Float	Number of adult hospital beds in seven days

Total_pediatric_beds_7 _day_average	Float	Number of pediatric hospital beds
Total_staffed_beds_7 _day_average	Float	Number of beds available with staff for the patient who occupies the bed
icu_beds_used_7_day	Float	Toatal number of ICU beds used

Initial intuitions of the dataset:

This dataset can be very useful in determining the reason why the covid cases are increasing/ decreasing. This can be achieved by analyzing the relationship between the total number of covid cases and the number of available ICU or hospital beds. For example, if there are more available hospital beds in a particular city/county/state, there is a smaller number of covid cases in that city/county/state. The other observation that can be made is, if there are not many hospitals in a particular city/county/state, then the number of covid cases can be higher than the average.

How to merge:

Here, the hospital beds dataset has “FIPS_code” column which is the unique five-digit code for each county. Similarly, the large covid dataset which was obtained by merging the three datasets in the team task also has “county_FIPS” column which is unique for each county. The hospital beds dataset has been pre-processed. The rows where the values of FIPS are NaN are dropped in the enrichment hospital beds dataset. By using FIPS_code and county_FIPS from the datasets, the hospital beds dataset can be merged with the large covid dataset using the inner join.

Hypothesis:

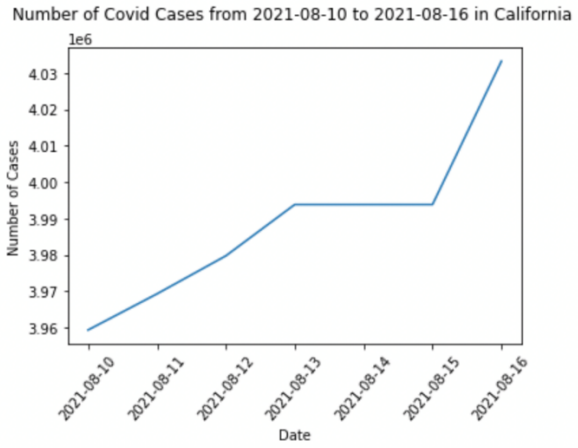
Below are some of the questions we can pose based on the information given in the dataset:

1. Counties with small population may have limited resources to the hospitals which leads to maximum utilization to the hospital beds, can this be the reason in raising the death rates?
2. Whether the counties which have the large number of hospitals eventually results the fewer confirmed covid cases?
3. Whether the number of confirmed covid cases are fewer in the areas which contains a greater number of hospitals when compared to the areas which contains a smaller number of hospitals?
4. As the hospitals are categorized as adult and pediatric, can we analyze the ratio of adults and the children who are confirmed to covid?

COVID-19 Data Trends:

To calculate the COVID-19 data trends, I chose California state to observe the total confirmed covid cases in that state in the last week i.e., from 10th Aug 2021 to 16th Aug 2021.

Below is the graph plot between the total number of covid cases in California state and the dates.



According to the graph, the number of covid cases are steadily increasing in the first four days i.e., from 10th Aug to 13th Aug. Then the number of covid cases got stabilized for the next two days. Later, on the last day, the graph sharply raised.

The graph was plotted by summing the number of covid cases on one day in the last week for every county in the California state. The total number of cases in California state in the last week is as follow:

Dates	**No. of covid cases**
08-10-2021	3959335
08-11-2021	3969262
08-12-2021	3979714
08-13-2021	3993812
08-14-2021	3993812
08-15-2021	3993812
08-16-2021	4033203