

## The Data

This dataset contains latest COVID-19 India state-wise data as on August 25, 2021. This dataset can be used to analyze COVID conditions in India. This dataset is great for **Exploratory Data Analysis** (source: [kaggle](#); [mygov.in](#))

## Columns

1. state\_uts - Names of Indian States and Union Territories.
2. total\_cases - Total number of confirmed cases
3. active - Total number of active cases
4. discharged - Total number of discharged cases
5. deaths - Total number of deaths
6. active\_ratio - Ratio of number of active cases to total cases
7. discharge\_ratio - Ratio of number of discharged cases to total cases
8. death\_ratio - Ratio of number of deaths to total cases

## Questions

1. Load your data in; ensure that none of the columns utilize the “factor” data type after loading in the CSV file.
2. Create a new column “entity\_type” that identifies whether the territory is a state or Union Territory. The new column should have the character data type where a state should have the value “State” and a Union Territory should have the value “Union Territory”. The union territories are:
  - a. Andaman and Nicobar
  - b. Chandigarh
  - c. Dadra and Nagar Haveli and Daman and Diu
  - d. Delhi
  - e. Jammu and Kashmir
  - f. Ladakh
  - g. Lakshadweep
  - h. Puducherry
3. Are states/UTs more likely to have a high death rate if their discharge rate is low? Use the correct visualization(s) to justify your answer.

4. The Indian Health Services (IHS) would like to prioritize which states/UTs to provide the most COVID-19 assistance and support.
  - a. Find all states/UTs that have higher than the mean number of total deaths. This variable should be defined as “entities\_\_high\_deaths”.
  - b. Find all states/UTs that have lower than the mean death rate across India. This variable should be defined as “entities\_\_low\_death\_rate”.
  - c. Create a new column “special\_state\_ut” in your dataframe that signifies whether a state/UT has both higher than the mean number of total deaths and lower than the mean death rate. The column’s data type should be logical (True if the state falls within the criteria, False otherwise).
  - d. Use the new information to provide guidance on which states should have prioritized COVID-19 assistance.