**Readme File**

**Assumption:**

States List :

Andhra Pradesh - ap, Arunachal Pradesh - ar, Assam- as, Bihar - br, Chhattisgarh-ct, Goa - ga, Gujarat - gj, Haryana - hr, Himachal Pradesh- hp, Jharkhand - jr, Karnataka - ka, Kerala - kl, Madhya Pradesh - mp, Maharashtra - mh, Manipur - mn, Meghalaya - ml, Mizoram - mz, Nagaland - nl, Odisha - or, Punjab - pb, Rajasthan - rj, Sikkim - sk, Tamil Nadu - tn, Telangana - tg, Tripura - tr, Uttar Pradesh - up, Uttarakhand -ut, West Bengal -wb

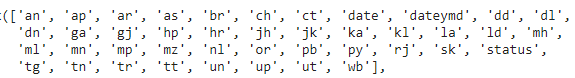
Union Territories List :

Andaman and Nicobar - an, Chandigarh - ch, Dadra Nagar Haveli - dn, Daman and Diu - dd, Delhi - dl, Jammu and Kashmir - jk, Lakshadweep- la, Ladakh - ld, Puducherry -py

Data Preprocessing :

1. Know your data:

The JSON data is stored in a dataframe. It has the following columns:



1. Feature Selection:

The “tt” and “un” are dropped. The “un” column is zero for all the rows. The Total count has been calculated without using the ‘tt’ column.

1. Handling Data Type:

The numbers in the dataset are in string form and numerical operations cannot be performed, so they are converted from ‘str’ to ‘int’.

**Q1. Data Manipulation**

1. **Count the total number of “Confirmed,” “Recovered,” and “Deceased” from 14-Mar-2020 to 16-Aug-2021 and report the numbers.**

**Solution :**

* We created three different data frames that contain rows corresponding to ‘Confirmed’, ‘Recovered’ and ‘Deceased’ status namely, df\_confirmed, df\_recovered, df\_deceased respectively.
* Then, we dropped ‘tt’, ‘date’, ‘dateymd’ ‘status’, and ‘un’ columns from these data frames.
* We calculated the total count of “Confirmed,” “Recovered,” and “Deceased” for all States and Union Territories.

**Output :**

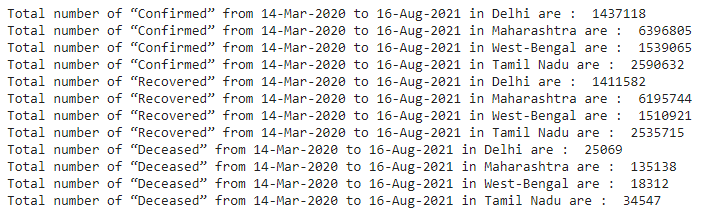


1. **Count the total number of “Confirmed,” “Recovered,” and “Deceased” from 14-Mar-2020 to 16-Aug-2021 for each state: Delhi, Maharashtra, West Bengal, and Tamil Nadu.**

**Solution :**

* The total is calculated from the above mentioned three data frames, only for 4 columns:
* We simply fetched the column name ‘dl’ (corresponding to Delhi), column name ‘mh’ (corresponding to Maharastra), column name ‘wb’ (corresponding to West Bengal), column name ‘tn’ (corresponding to Tamil Nadu) and find the sum of the entire column.

**Output :**



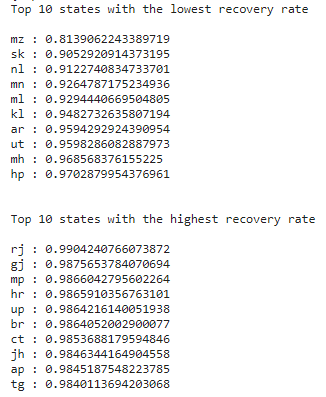
1. **Report the top 10 states with the highest recovery rate and top 10 states with the lowest recovery rate from 14-Mar-2020 to 16-Aug-2021.**

**Solution :**

* To calculate the recovery rate:

* The recovery rate of each state is calculated by dividing the cumulative recovery total of the state by the total positive cases of the state from 14-Mar-2020 to 16-Aug-2021.
* The states are sorted based on the recovery rates in ascending order and top 10 and lowest 10 are printed.

**Output :**



**Inference :**

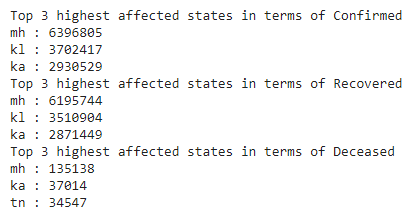
* Mizoram, Sikkim, Nagaland, Manipur, Meghalaya, Kerala, Arunachal Pradesh, Uttrakhand, Maharashtra and Himachal Pradesh have the lowest recovery rate in ascending order.
* Rajasthan, Gujrat, Madhya Pradesh, Haryana, Uttar Pradesh, Bihar, Chhattisgarh, Jharkhand, Andhra Pradesh and Telangana have the highest recovery rates in ascending order.

1. **Report the top 3 highest affected states in terms of “Confirmed,” “Recovered,” and “Deceased,” with the count from 14-Mar-2020 to 16-Aug-2021.**

**Solution :**

* A list is made for storing the total confirmed cases for each state.
* Similarly lists are made for total recovered cases and total deceased cases for each state.
* The states are sorted based on the count of the cases in the lists.

**Output :**



**Inference :**

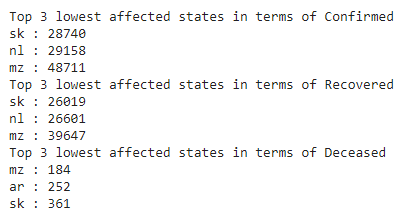
* Maharashtra, Kerala, and Karnataka have the highest affected and recovered cases.
* Maximum people deceased in Maharashtra, Karnataka, and Tamil Nadu.

1. **Report the top 3 lowest affected states in terms of “Confirmed,” “Recovered,” and “Deceased,” with the count from 14-Mar-2020 to 16-Aug-2021.**

**Solution :**

* The lowest affected states are also extracted using the same method as above.

**Output :**



**Inference :**

* Sikkim, Nagaland and Mizoram has the lowest affected and recovered count.
* The lowest deceased count is in states Mizoram, Arunachal Pradesh and Sikkim.

1. **Find the day and count with the highest spike in a day in the number of cases for each state and UTs for “Confirmed,” “Recovered” and “Deceased” between dates 14-Mar-2020 and 16-Aug-2021.**

**Solution :**

* For all the states, we find the index that has the maximum count of Confirmed, Recovered, Deceased data frames separately.
* Then we find the corresponding dates using the index retrieved.

**Output :**

+----+---------+-------------------------------------+-----------+

| | State | Highest Number of Confirmed cases | Date |

|----+---------+-------------------------------------+-----------|

| 0 | an | 149 | 14-Aug-20 |

| 1 | ap | 24171 | 16-May-21 |

| 2 | ar | 566 | 12-Jul-21 |

| 3 | as | 6573 | 20-May-21 |

| 4 | br | 15853 | 30-Apr-21 |

| 5 | ch | 895 | 09-May-21 |

| 6 | ct | 17397 | 23-Apr-21 |

| 7 | dd | 0 | 14-Mar-20 |

| 8 | dl | 28395 | 20-Apr-21 |

| 9 | dn | 359 | 22-Apr-21 |

| 10 | ga | 4195 | 07-May-21 |

| 11 | gj | 14605 | 30-Apr-21 |

| 12 | hp | 5424 | 08-May-21 |

| 13 | hr | 15786 | 04-May-21 |

| 14 | jh | 8075 | 28-Apr-21 |

| 15 | jk | 5443 | 07-May-21 |

| 16 | ka | 50112 | 05-May-21 |

| 17 | kl | 43529 | 12-May-21 |

| 18 | la | 362 | 17-Apr-21 |

| 19 | ld | 345 | 21-May-21 |

| 20 | mh | 68631 | 18-Apr-21 |

| 21 | ml | 1183 | 20-May-21 |

| 22 | mn | 1327 | 21-Jul-21 |

| 23 | mp | 13601 | 25-Apr-21 |

| 24 | mz | 1369 | 26-Jul-21 |

| 25 | nl | 366 | 13-May-21 |

| 26 | or | 12852 | 23-May-21 |

| 27 | pb | 9042 | 08-May-21 |

| 28 | py | 2049 | 11-May-21 |

| 29 | rj | 18298 | 02-May-21 |

| 30 | sk | 420 | 28-May-21 |

| 31 | tg | 11451 | 07-May-21 |

| 32 | tn | 36184 | 21-May-21 |

| 33 | tr | 879 | 19-May-21 |

| 34 | up | 37944 | 24-Apr-21 |

| 35 | ut | 9642 | 07-May-21 |

| 36 | wb | 20846 | 14-May-21 |

+----+---------+-------------------------------------+-----------+

+----+---------+-------------------------------------+-----------+

| | State | Highest Number of Recovered cases | Date |

|----+---------+-------------------------------------+-----------|

| 0 | an | 149 | 14-Aug-20 |

| 1 | ap | 24171 | 16-May-21 |

| 2 | ar | 566 | 12-Jul-21 |

| 3 | as | 6573 | 20-May-21 |

| 4 | br | 15853 | 30-Apr-21 |

| 5 | ch | 895 | 09-May-21 |

| 6 | ct | 17397 | 23-Apr-21 |

| 7 | dd | 0 | 14-Mar-20 |

| 8 | dl | 28395 | 20-Apr-21 |

| 9 | dn | 359 | 22-Apr-21 |

| 10 | ga | 4195 | 07-May-21 |

| 11 | gj | 14605 | 30-Apr-21 |

| 12 | hp | 5424 | 08-May-21 |

| 13 | hr | 15786 | 04-May-21 |

| 14 | jh | 8075 | 28-Apr-21 |

| 15 | jk | 5443 | 07-May-21 |

| 16 | ka | 50112 | 05-May-21 |

| 17 | kl | 43529 | 12-May-21 |

| 18 | la | 362 | 17-Apr-21 |

| 19 | ld | 345 | 21-May-21 |

| 20 | mh | 68631 | 18-Apr-21 |

| 21 | ml | 1183 | 20-May-21 |

| 22 | mn | 1327 | 21-Jul-21 |

| 23 | mp | 13601 | 25-Apr-21 |

| 24 | mz | 1369 | 26-Jul-21 |

| 25 | nl | 366 | 13-May-21 |

| 26 | or | 12852 | 23-May-21 |

| 27 | pb | 9042 | 08-May-21 |

| 28 | py | 2049 | 11-May-21 |

| 29 | rj | 18298 | 02-May-21 |

| 30 | sk | 420 | 28-May-21 |

| 31 | tg | 11451 | 07-May-21 |

| 32 | tn | 36184 | 21-May-21 |

| 33 | tr | 879 | 19-May-21 |

| 34 | up | 37944 | 24-Apr-21 |

| 35 | ut | 9642 | 07-May-21 |

| 36 | wb | 20846 | 14-May-21 |

+----+---------+-------------------------------------+-----------+

+----+---------+------------------------------------+-----------+

| | State | Highest Number of Deceased cases | Date |

|----+---------+------------------------------------+-----------|

| 0 | an | 149 | 14-Aug-20 |

| 1 | ap | 24171 | 16-May-21 |

| 2 | ar | 566 | 12-Jul-21 |

| 3 | as | 6573 | 20-May-21 |

| 4 | br | 15853 | 30-Apr-21 |

| 5 | ch | 895 | 09-May-21 |

| 6 | ct | 17397 | 23-Apr-21 |

| 7 | dd | 0 | 14-Mar-20 |

| 8 | dl | 28395 | 20-Apr-21 |

| 9 | dn | 359 | 22-Apr-21 |

| 10 | ga | 4195 | 07-May-21 |

| 11 | gj | 14605 | 30-Apr-21 |

| 12 | hp | 5424 | 08-May-21 |

| 13 | hr | 15786 | 04-May-21 |

| 14 | jh | 8075 | 28-Apr-21 |

| 15 | jk | 5443 | 07-May-21 |

| 16 | ka | 50112 | 05-May-21 |

| 17 | kl | 43529 | 12-May-21 |

| 18 | la | 362 | 17-Apr-21 |

| 19 | ld | 345 | 21-May-21 |

| 20 | mh | 68631 | 18-Apr-21 |

| 21 | ml | 1183 | 20-May-21 |

| 22 | mn | 1327 | 21-Jul-21 |

| 23 | mp | 13601 | 25-Apr-21 |

| 24 | mz | 1369 | 26-Jul-21 |

| 25 | nl | 366 | 13-May-21 |

| 26 | or | 12852 | 23-May-21 |

| 27 | pb | 9042 | 08-May-21 |

| 28 | py | 2049 | 11-May-21 |

| 29 | rj | 18298 | 02-May-21 |

| 30 | sk | 420 | 28-May-21 |

| 31 | tg | 11451 | 07-May-21 |

| 32 | tn | 36184 | 21-May-21 |

| 33 | tr | 879 | 19-May-21 |

| 34 | up | 37944 | 24-Apr-21 |

| 35 | ut | 9642 | 07-May-21 |

| 36 | wb | 20846 | 14-May-21 |

+----+---------+------------------------------------+-----------+

1. **Report active cases (Assume active = Confirmed - (Recovered + Deceased)) state-wise for all individual states and UTs on date 15-Aug-2021 (This date only) starting from 14-March-2020.**

**Solution :**

* Formula Used :

**Output :**

an : 6

ap : 17218

ar : 1837

as : 8947

br : 213

ch : 43

ct : 1138

dd : 0

dl : 467

dn : -18

ga : 873

gj : 183

hp : 2716

hr : 667

jh : 209

jk : 1229

ka : 22066

kl : 172769

la : 13

ld : 79

mh : 65923

ml : 3852

mn : 6263

mp : 93

mz : 8880

nl : 1958

or : 9020

pb : 557

py : 894

rj : 180

sk : 2360

tg : 6583

tn : 20370

tr : 1601

up : 419

ut : 6391

wb : 9832

**Inference :**

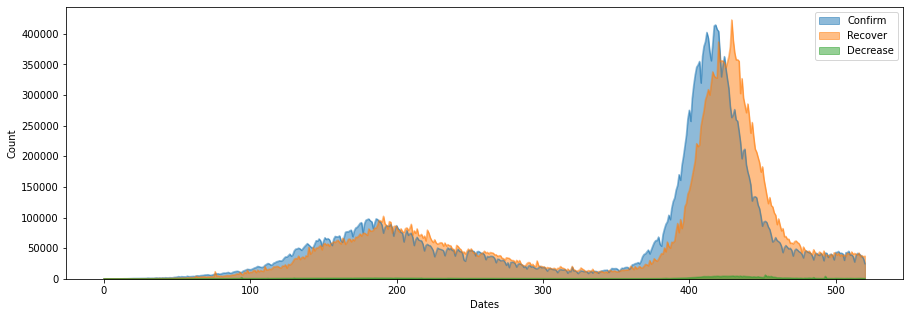
* “Dadar and Nagar Haveli - dn” has more (Recovered + deceased) cases than confirmed cases so the count is negative.
* “Daman and Diu - dn” has 0 confirmed, recovered and deceased cases.

**Q2. Plotting**

* We used matplotlib.pyplot library to plot the area trend line. We have set “stacked” to false. The legends are used to clearly demonstrate the colour code.

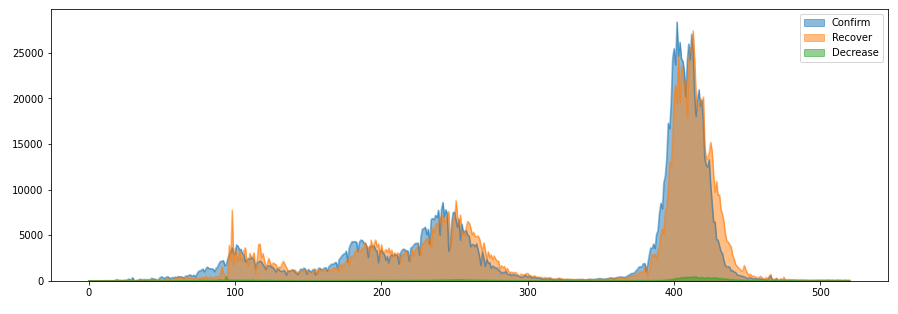
1. **Plot the area trend line for total “Confirmed,” “Recovered,” and “Deceased” cases from 14-Mar-2020 to 16-Aug-2021**.

**Output :**



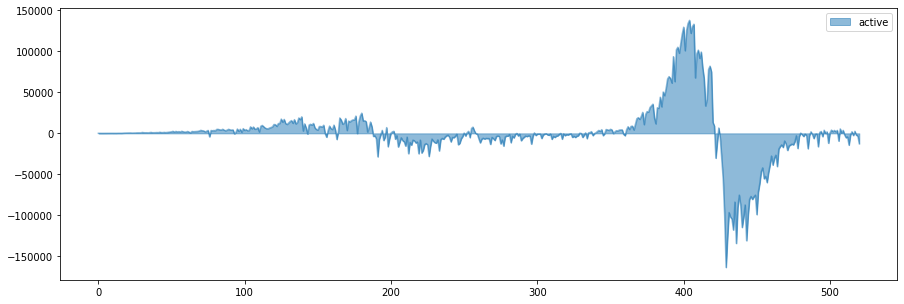
1. **Plot the area trend line for total “Confirmed,” “Recovered,” and “Deceased” cases for Delhi (dl) from 14-Mar-2020 to 16-Aug-2021.**

**Output :**



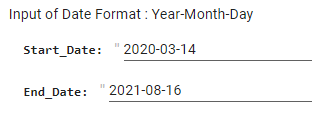
1. **Plot the area trend line for active cases. Assume active = Confirmed - (Recovered + Deceased) from 14-Mar-2020 to 16-Aug-2021.**

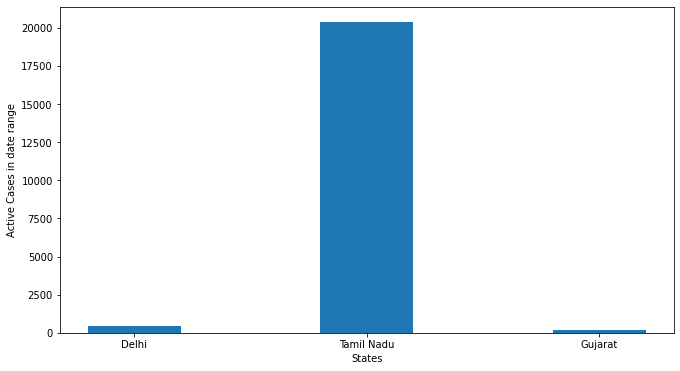
**Solution :**

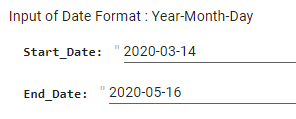


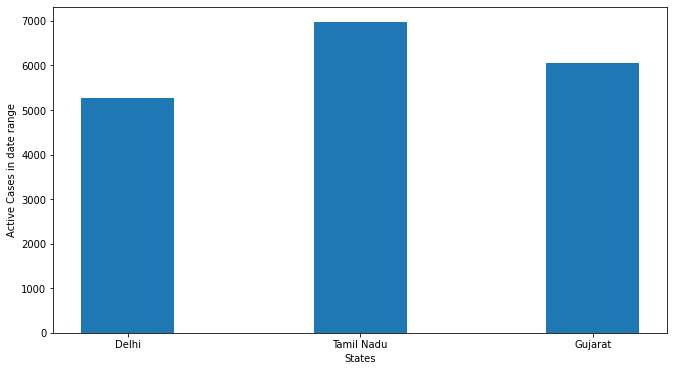
1. **Plot a bar plot of the number of active cases in Delhi, Tamil Nadu, and Gujarat for any date range of your choice.**

**Solution :**









.