**import** numpy **as** np

**import** pandas **as** pd

**from** matplotlib **import** pyplot **as** plt

**import** seaborn **as** sns

**import** datetime **as** dt

In [2]:

*#importing main dataset*

df**=** pd**.**read\_csv("C:/Users/Sakshi/Downloads/covid19.csv", parse\_dates **=** ['Date'], dayfirst **=** **True**)

In [3]:

df

Out[3]:

|  | **Date** | **Name of State / UT** | **Latitude** | **Longitude** | **Total Confirmed cases** | **Death** | **Cured/Discharged/Migrated** | **New cases** | **New deaths** | **New recovered** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 2020-01-30 | Kerala | 10.8505 | 76.2711 | 1 | 0 | 0 | 0 | 0 | 0 |
| **1** | 2020-01-31 | Kerala | 10.8505 | 76.2711 | 1 | 0 | 0 | 0 | 0 | 0 |
| **2** | 2020-02-01 | Kerala | 10.8505 | 76.2711 | 2 | 0 | 0 | 1 | 0 | 0 |
| **3** | 2020-02-02 | Kerala | 10.8505 | 76.2711 | 3 | 0 | 0 | 1 | 0 | 0 |
| **4** | 2020-02-03 | Kerala | 10.8505 | 76.2711 | 3 | 0 | 0 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **4687** | 2020-08-06 | Telangana | 18.1124 | 79.0193 | 73050 | 589 | 52103 | 2092 | 0 | 1289 |
| **4688** | 2020-08-06 | Tripura | 23.9408 | 91.9882 | 5725 | 31 | 3793 | 97 | 0 | 68 |
| **4689** | 2020-08-06 | Uttar Pradesh | 26.8467 | 80.9462 | 104388 | 1857 | 60558 | 4078 | 0 | 3287 |
| **4690** | 2020-08-06 | Uttarakhand | 30.0668 | 79.0193 | 8254 | 98 | 5233 | 246 | 0 | 386 |
| **4691** | 2020-08-06 | West Bengal | 22.9868 | 87.8550 | 83800 | 1846 | 58962 | 2816 | 0 | 2078 |

4692 rows × 10 columns

In [4]:

*#keeping required columns*

df **=** df[['Date', 'Name of State / UT', 'Cured/Discharged/Migrated', 'Death', 'Total Confirmed cases']]

*#renaming the columns*

df**.**columns **=** ['date' , 'state', 'cured', 'deaths', 'confirmed']

In [5]:

df**.**head()

Out[5]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **0** | 2020-01-30 | Kerala | 0 | 0 | 1 |
| **1** | 2020-01-31 | Kerala | 0 | 0 | 1 |
| **2** | 2020-02-01 | Kerala | 0 | 0 | 2 |
| **3** | 2020-02-02 | Kerala | 0 | 0 | 3 |
| **4** | 2020-02-03 | Kerala | 0 | 0 | 3 |

In [6]:

df**.**tail()

Out[6]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4687** | 2020-08-06 | Telangana | 52103 | 589 | 73050 |
| **4688** | 2020-08-06 | Tripura | 3793 | 31 | 5725 |
| **4689** | 2020-08-06 | Uttar Pradesh | 60558 | 1857 | 104388 |
| **4690** | 2020-08-06 | Uttarakhand | 5233 | 98 | 8254 |
| **4691** | 2020-08-06 | West Bengal | 58962 | 1846 | 83800 |

In [7]:

*#current date*

today **=** df[df['date']**==**'2020-08-06']

In [8]:

today**.**head()

Out[8]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4657** | 2020-08-06 | Andaman and Nicobar Islands | 326 | 14 | 1027 |
| **4658** | 2020-08-06 | Andhra Pradesh | 104354 | 1681 | 186461 |
| **4659** | 2020-08-06 | Arunachal Pradesh | 1210 | 3 | 1855 |
| **4660** | 2020-08-06 | Assam | 35892 | 121 | 50445 |
| **4661** | 2020-08-06 | Bihar | 42414 | 355 | 64770 |

In [9]:

*#sorting data w.r.t number of confirmed cases*

max\_confired\_cases **=** today**.**sort\_values(by **=** ['confirmed'], ascending **=** **False**)

max\_confired\_cases

Out[9]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4676** | 2020-08-06 | Maharashtra | 305521 | 16476 | 468265 |
| **4686** | 2020-08-06 | Tamil Nadu | 214815 | 4461 | 273460 |
| **4658** | 2020-08-06 | Andhra Pradesh | 104354 | 1681 | 186461 |
| **4672** | 2020-08-06 | Karnataka | 74679 | 2804 | 151449 |
| **4665** | 2020-08-06 | Delhi | 126116 | 4044 | 140232 |
| **4689** | 2020-08-06 | Uttar Pradesh | 60558 | 1857 | 104388 |
| **4691** | 2020-08-06 | West Bengal | 58962 | 1846 | 83800 |
| **4687** | 2020-08-06 | Telangana | 52103 | 589 | 73050 |
| **4667** | 2020-08-06 | Gujarat | 49433 | 2556 | 66669 |
| **4661** | 2020-08-06 | Bihar | 42414 | 355 | 64770 |
| **4660** | 2020-08-06 | Assam | 35892 | 121 | 50445 |
| **4684** | 2020-08-06 | Rajasthan | 33849 | 745 | 47272 |
| **4681** | 2020-08-06 | Odisha | 25738 | 225 | 39018 |
| **4668** | 2020-08-06 | Haryana | 31960 | 455 | 38548 |
| **4675** | 2020-08-06 | Madhya Pradesh | 26064 | 929 | 35734 |
| **4673** | 2020-08-06 | Kerala | 17533 | 94 | 29151 |
| **4670** | 2020-08-06 | Jammu and Kashmir | 15244 | 426 | 22955 |
| **4683** | 2020-08-06 | Punjab | 12943 | 491 | 19856 |
| **4671** | 2020-08-06 | Jharkhand | 5703 | 136 | 14888 |
| **4663** | 2020-08-06 | Chhattisgarh | 7871 | 71 | 10407 |
| **4690** | 2020-08-06 | Uttarakhand | 5233 | 98 | 8254 |
| **4666** | 2020-08-06 | Goa | 5287 | 64 | 7423 |
| **4688** | 2020-08-06 | Tripura | 3793 | 31 | 5725 |
| **4682** | 2020-08-06 | Puducherry | 2668 | 65 | 4433 |
| **4677** | 2020-08-06 | Manipur | 1862 | 7 | 3093 |
| **4669** | 2020-08-06 | Himachal Pradesh | 1762 | 14 | 2916 |
| **4680** | 2020-08-06 | Nagaland | 685 | 6 | 2498 |
| **4659** | 2020-08-06 | Arunachal Pradesh | 1210 | 3 | 1855 |
| **4674** | 2020-08-06 | Ladakh | 1164 | 7 | 1592 |
| **4664** | 2020-08-06 | Dadra and Nagar Haveli and Daman and Diu | 960 | 2 | 1366 |
| **4662** | 2020-08-06 | Chandigarh | 715 | 20 | 1270 |
| **4657** | 2020-08-06 | Andaman and Nicobar Islands | 326 | 14 | 1027 |
| **4678** | 2020-08-06 | Meghalaya | 330 | 5 | 929 |
| **4685** | 2020-08-06 | Sikkim | 303 | 1 | 800 |
| **4679** | 2020-08-06 | Mizoram | 286 | 0 | 537 |

In [10]:

*#getting states with maximum number of confirmed cases*

top\_state\_confirmed **=** max\_confired\_cases[0:5]

top\_state\_confirmed

Out[10]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4676** | 2020-08-06 | Maharashtra | 305521 | 16476 | 468265 |
| **4686** | 2020-08-06 | Tamil Nadu | 214815 | 4461 | 273460 |
| **4658** | 2020-08-06 | Andhra Pradesh | 104354 | 1681 | 186461 |
| **4672** | 2020-08-06 | Karnataka | 74679 | 2804 | 151449 |
| **4665** | 2020-08-06 | Delhi | 126116 | 4044 | 140232 |

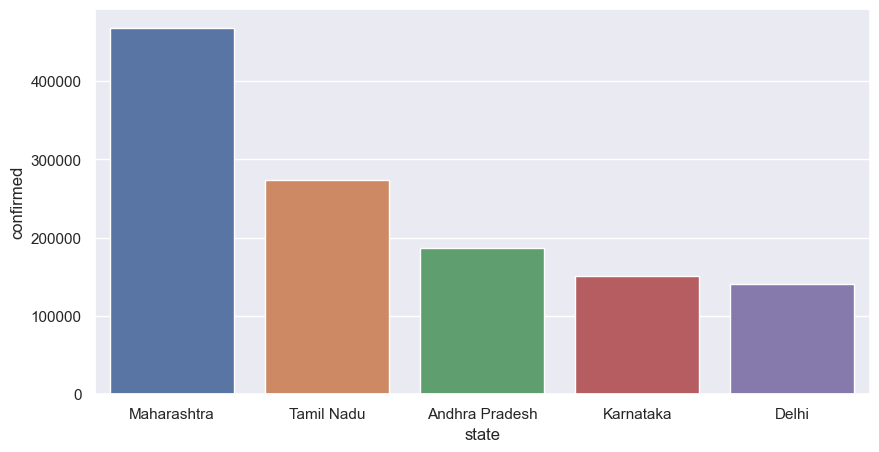
In [11]:

*# bar plot for top\_state\_confirmed*

sns**.**set(rc **=** {'figure.figsize' : (10, 5)})

sns**.**barplot(x **=** 'state', y **=** 'confirmed', data **=** top\_state\_confirmed, hue **=** 'state')

plt**.**show()



In [12]:

*#sorting data w.r.t number of death cases*

max\_death\_cases **=** today**.**sort\_values(by **=** ['deaths'], ascending **=** **False**)

max\_death\_cases

Out[12]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4690** | 2020-08-06 | Uttarakhand | 5233 | 98 | 8254 |
| **4673** | 2020-08-06 | Kerala | 17533 | 94 | 29151 |
| **4675** | 2020-08-06 | Madhya Pradesh | 26064 | 929 | 35734 |
| **4684** | 2020-08-06 | Rajasthan | 33849 | 745 | 47272 |
| **4663** | 2020-08-06 | Chhattisgarh | 7871 | 71 | 10407 |
| **4674** | 2020-08-06 | Ladakh | 1164 | 7 | 1592 |
| **4677** | 2020-08-06 | Manipur | 1862 | 7 | 3093 |
| **4682** | 2020-08-06 | Puducherry | 2668 | 65 | 4433 |
| **4666** | 2020-08-06 | Goa | 5287 | 64 | 7423 |
| **4680** | 2020-08-06 | Nagaland | 685 | 6 | 2498 |
| **4687** | 2020-08-06 | Telangana | 52103 | 589 | 73050 |
| **4678** | 2020-08-06 | Meghalaya | 330 | 5 | 929 |
| **4683** | 2020-08-06 | Punjab | 12943 | 491 | 19856 |
| **4668** | 2020-08-06 | Haryana | 31960 | 455 | 38548 |
| **4686** | 2020-08-06 | Tamil Nadu | 214815 | 4461 | 273460 |
| **4670** | 2020-08-06 | Jammu and Kashmir | 15244 | 426 | 22955 |
| **4665** | 2020-08-06 | Delhi | 126116 | 4044 | 140232 |
| **4661** | 2020-08-06 | Bihar | 42414 | 355 | 64770 |
| **4688** | 2020-08-06 | Tripura | 3793 | 31 | 5725 |
| **4659** | 2020-08-06 | Arunachal Pradesh | 1210 | 3 | 1855 |
| **4672** | 2020-08-06 | Karnataka | 74679 | 2804 | 151449 |
| **4667** | 2020-08-06 | Gujarat | 49433 | 2556 | 66669 |
| **4681** | 2020-08-06 | Odisha | 25738 | 225 | 39018 |
| **4662** | 2020-08-06 | Chandigarh | 715 | 20 | 1270 |
| **4664** | 2020-08-06 | Dadra and Nagar Haveli and Daman and Diu | 960 | 2 | 1366 |
| **4689** | 2020-08-06 | Uttar Pradesh | 60558 | 1857 | 104388 |
| **4691** | 2020-08-06 | West Bengal | 58962 | 1846 | 83800 |
| **4658** | 2020-08-06 | Andhra Pradesh | 104354 | 1681 | 186461 |
| **4676** | 2020-08-06 | Maharashtra | 305521 | 16476 | 468265 |
| **4669** | 2020-08-06 | Himachal Pradesh | 1762 | 14 | 2916 |
| **4657** | 2020-08-06 | Andaman and Nicobar Islands | 326 | 14 | 1027 |
| **4671** | 2020-08-06 | Jharkhand | 5703 | 136 | 14888 |
| **4660** | 2020-08-06 | Assam | 35892 | 121 | 50445 |
| **4685** | 2020-08-06 | Sikkim | 303 | 1 | 800 |
| **4679** | 2020-08-06 | Mizoram | 286 | 0 | 537 |

In [13]:

*#getting states with maximum number of death cases*

top\_state\_death **=** max\_death\_cases[0:5]

top\_state\_death

Out[13]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4690** | 2020-08-06 | Uttarakhand | 5233 | 98 | 8254 |
| **4673** | 2020-08-06 | Kerala | 17533 | 94 | 29151 |
| **4675** | 2020-08-06 | Madhya Pradesh | 26064 | 929 | 35734 |
| **4684** | 2020-08-06 | Rajasthan | 33849 | 745 | 47272 |
| **4663** | 2020-08-06 | Chhattisgarh | 7871 | 71 | 10407 |

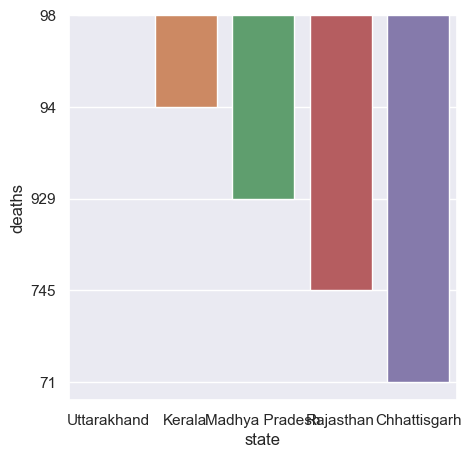
In [74]:

*# bar plot for top\_state\_death*

sns**.**set(rc **=** {'figure.figsize' : (5, 5)})

sns**.**barplot(x **=** 'state', y **=** 'deaths', data **=** top\_state\_death, hue **=** 'state')

plt**.**show()



In [70]:

*#sorting data w.r.t max cured cases*

max\_cured\_cases **=** today**.**sort\_values(by **=** ['cured'], ascending **=** **False**)

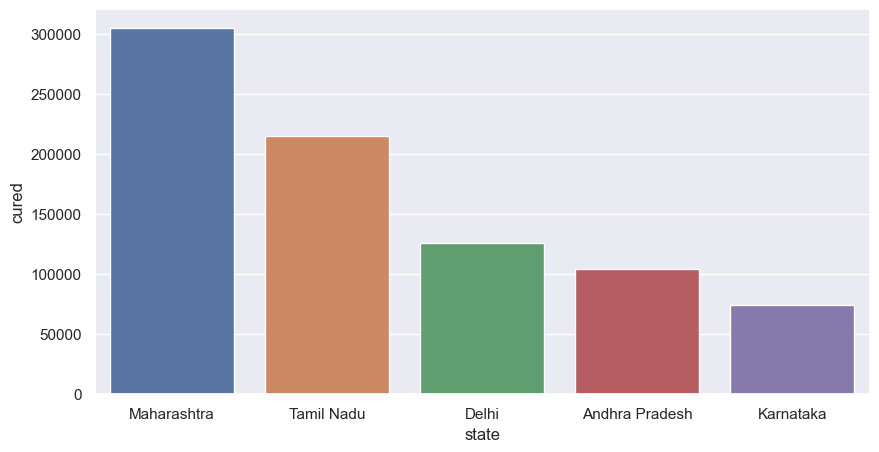
max\_cured\_cases

top\_cured\_cases **=** max\_cured\_cases[0:5]

sns**.**set(rc **=** {'figure.figsize' : (10, 5)})

sns**.**barplot(x **=** 'state', y **=** 'cured', data **=** top\_cured\_cases, hue **=** 'state')

plt**.**show()



In [16]:

*#maharashtra*

maha **=** df[df**.**state **==**'Maharashtra']

maha

Out[16]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **76** | 2020-03-09 | Maharashtra | 0 | 0 | 2 |
| **88** | 2020-03-10 | Maharashtra | 0 | 0 | 5 |
| **100** | 2020-03-11 | Maharashtra | 0 | 0 | 10 |
| **113** | 2020-03-12 | Maharashtra | 0 | 0 | 11 |
| **126** | 2020-03-13 | Maharashtra | 0 | 0 | 11 |
| **...** | ... | ... | ... | ... | ... |
| **4536** | 2020-08-02 | Maharashtra | 266883 | 15316 | 431719 |
| **4571** | 2020-08-03 | Maharashtra | 276809 | 15576 | 441228 |
| **4606** | 2020-08-04 | Maharashtra | 287030 | 15842 | 450196 |
| **4641** | 2020-08-05 | Maharashtra | 299356 | 16142 | 457956 |
| **4676** | 2020-08-06 | Maharashtra | 305521 | 16476 | 468265 |

147 rows × 5 columns

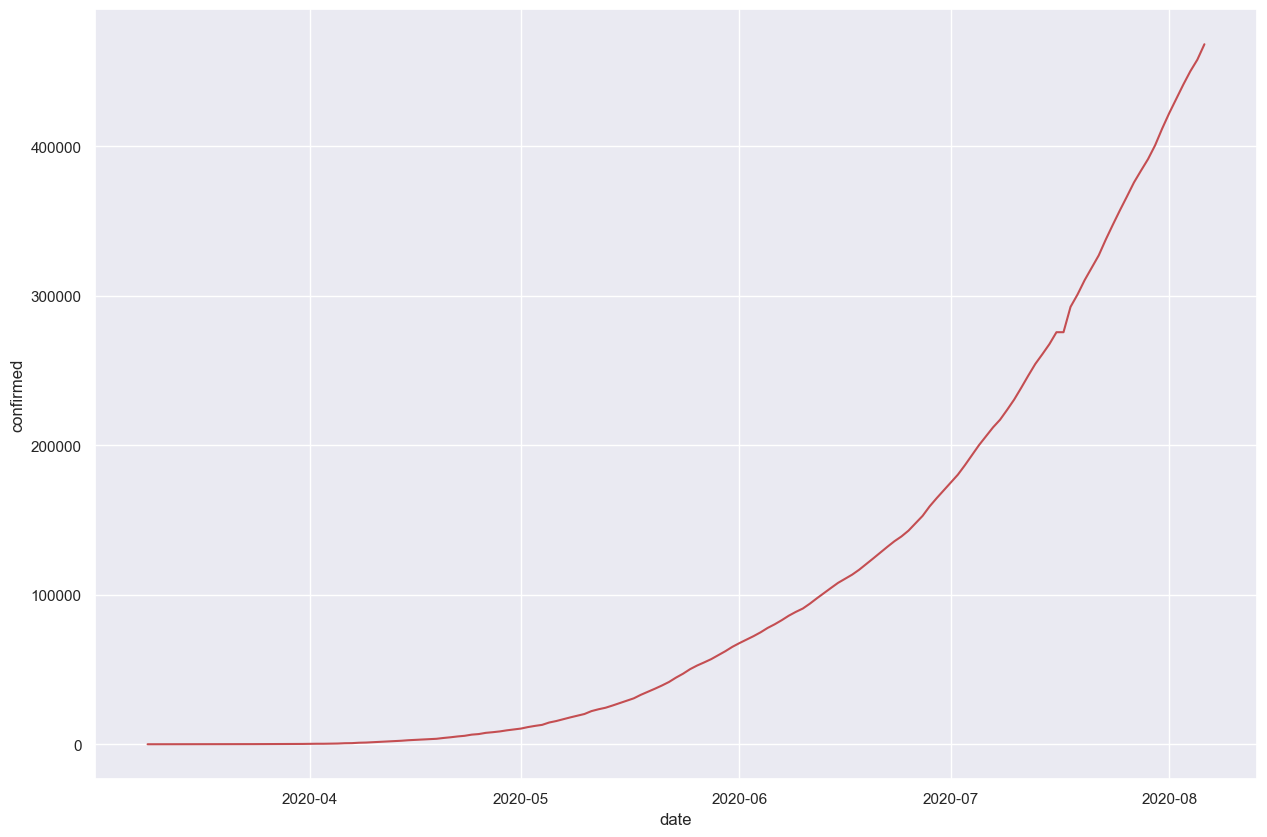
In [17]:

*#visualizing confirmed cases in maharashtra*

sns**.**set(rc **=** {'figure.figsize' : (15, 10)})

sns**.**lineplot(x **=** 'date', y **=** 'confirmed', data **=** maha, color **=** 'r')

plt**.**show()



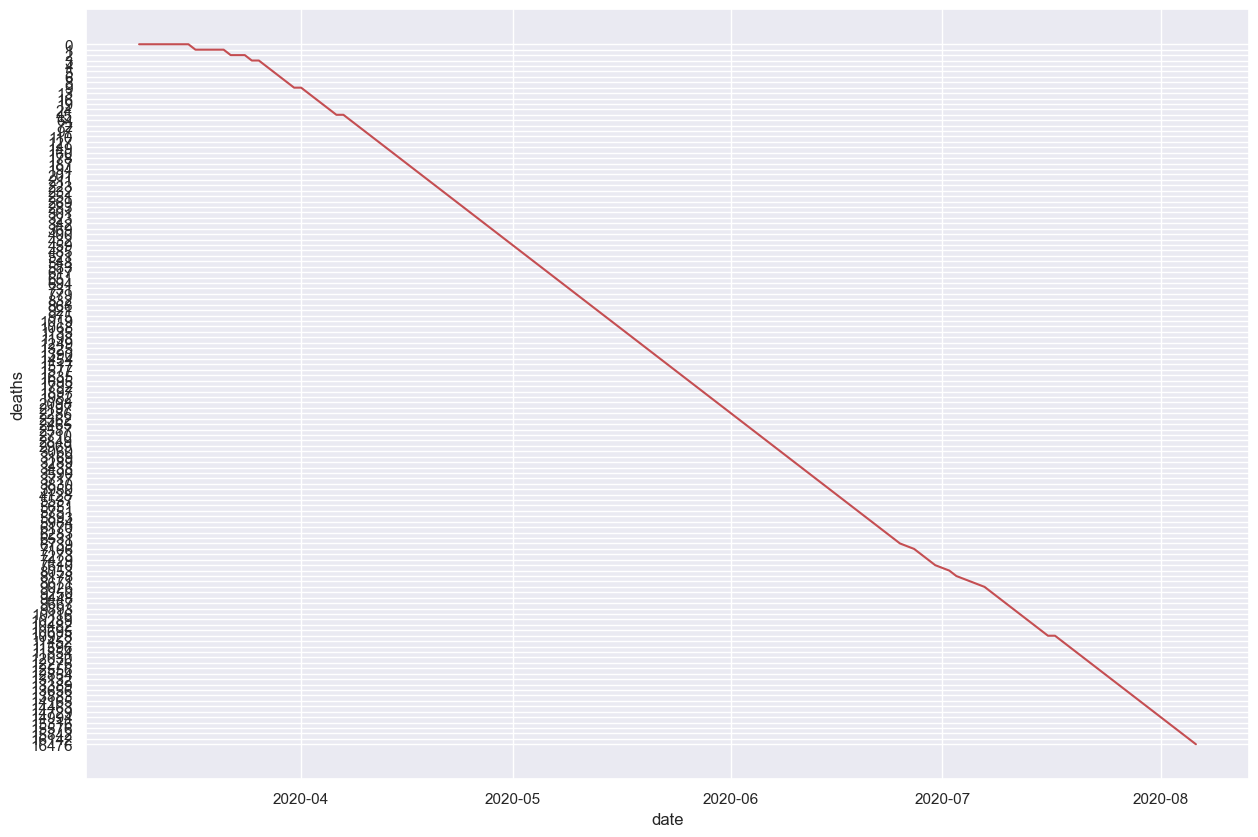
In [18]:

*#visualizing death cases in maharashtra*

sns**.**set(rc **=** {'figure.figsize' : (15, 10)})

sns**.**lineplot(x **=** 'date', y **=** 'deaths', data **=** maha, color **=** 'r')

plt**.**show()



In [19]:

*#visualizing confirmed cases in kerla*

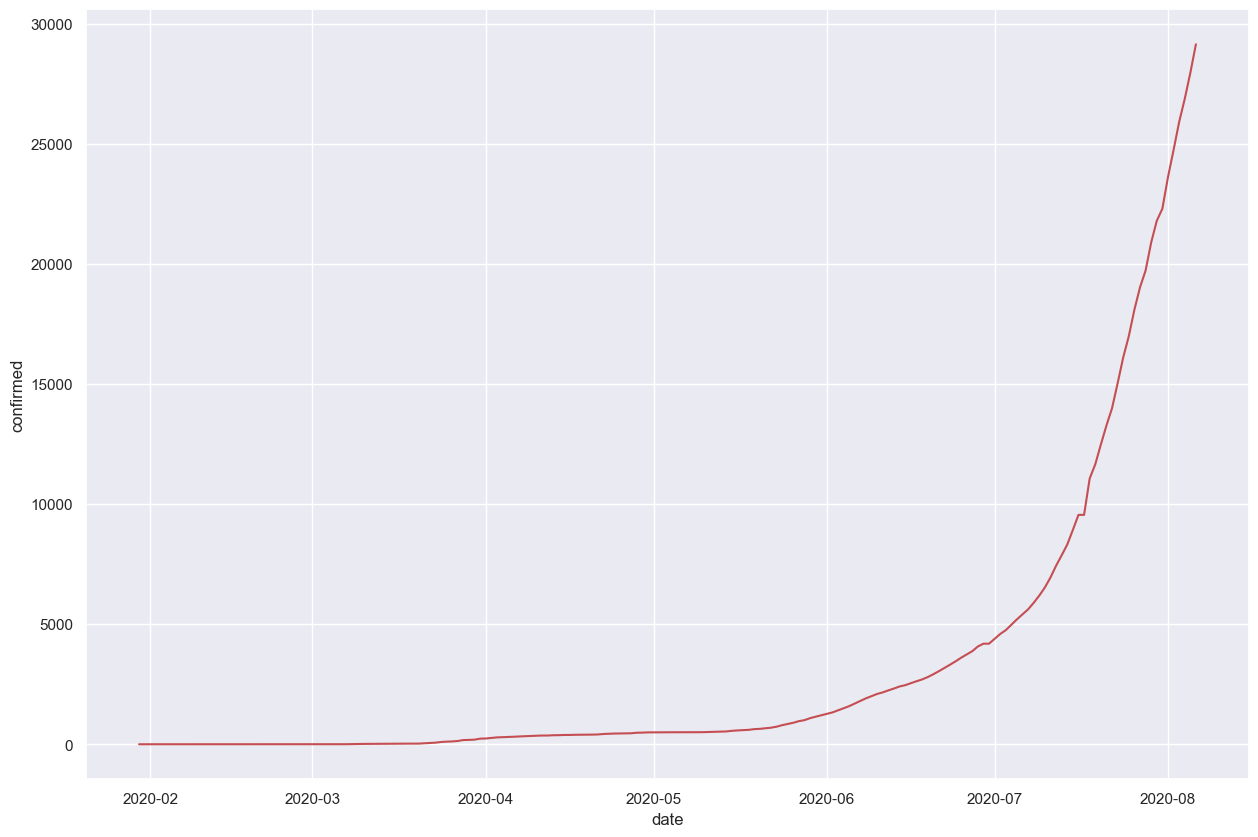
kerala **=** df[df**.**state **==**'Kerala']

kerala

sns**.**set(rc **=** {'figure.figsize' : (15, 10)})

sns**.**lineplot(x **=** 'date', y **=** 'confirmed', data **=** kerala, color **=** 'r')

plt**.**show()



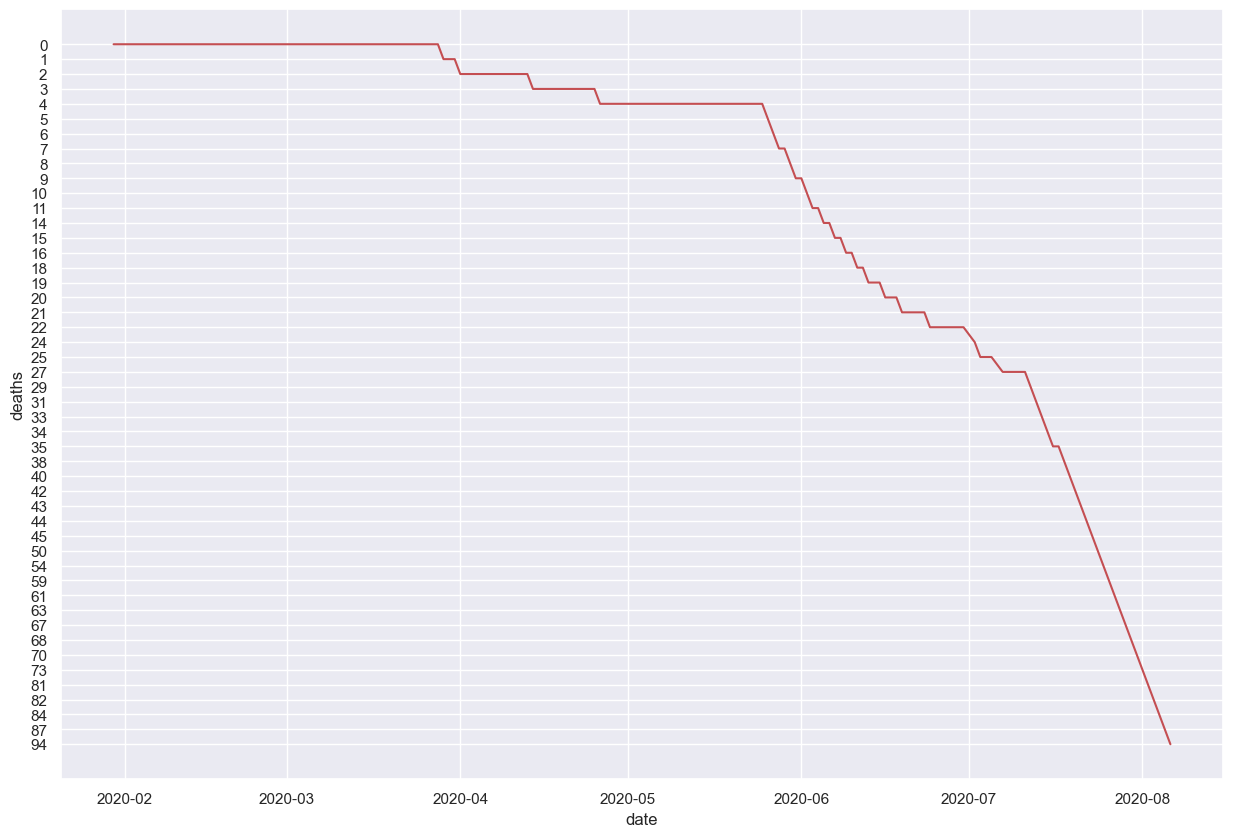
In [20]:

*#visualizing death cases in kerla*

sns**.**set(rc **=** {'figure.figsize' : (15, 10)})

sns**.**lineplot(x **=** 'date', y **=** 'deaths', data **=** kerala, color **=** 'r')

plt**.**show()



In [21]:

*#jammu and kashmir*

jk **=** df[df**.**state **==** 'Jammu and Kashmir']

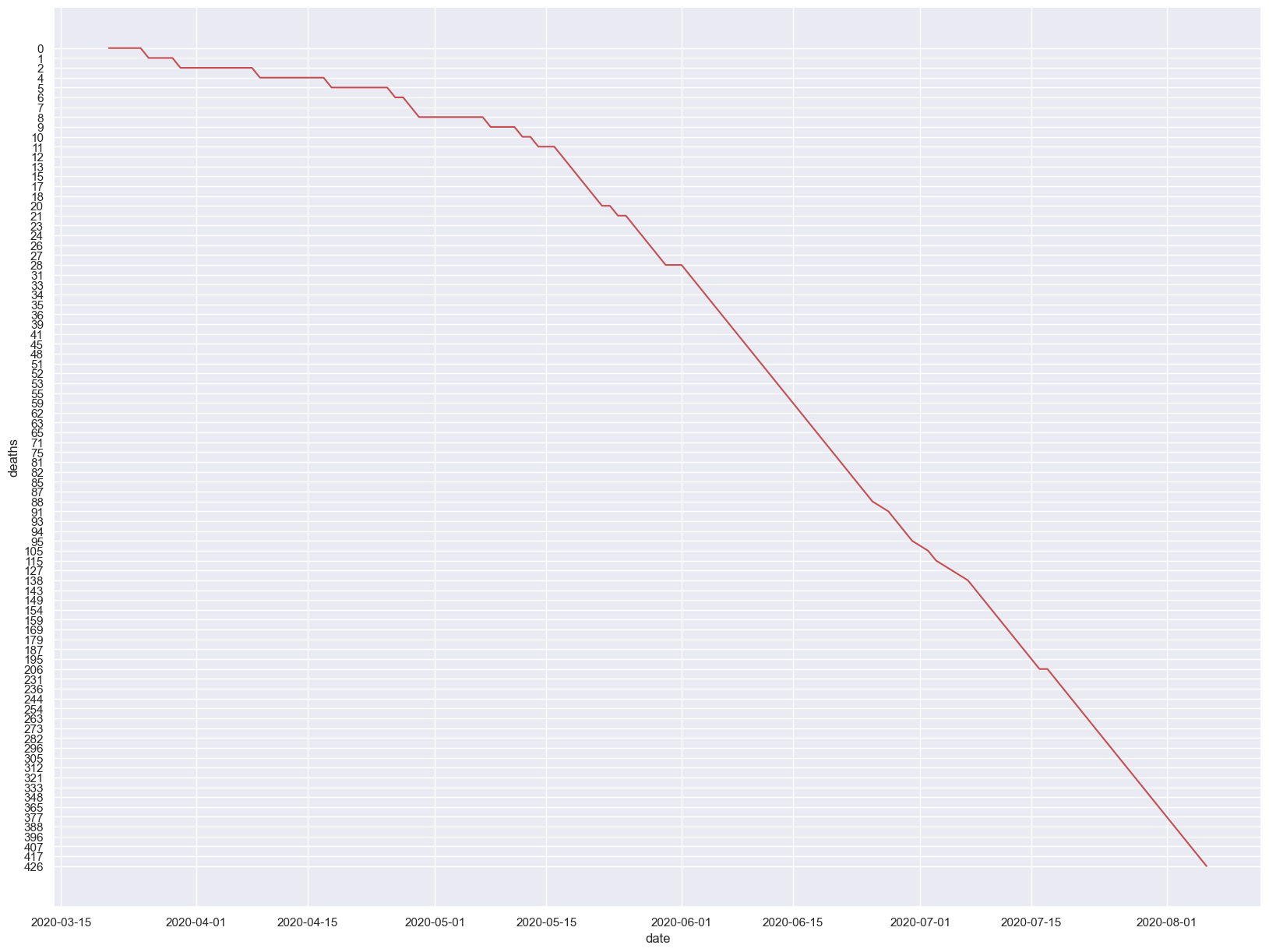
jk

*#visualizing death cases in jk*

sns**.**set(rc **=** {'figure.figsize' : (20, 15)})

sns**.**lineplot(x **=** 'date', y **=** 'deaths', data **=** jk, color **=** 'r')

plt**.**show()



In [22]:

*#linear regression*

In [23]:

**from** sklearn.model\_selection **import** train\_test\_split

In [24]:

maha

Out[24]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **76** | 2020-03-09 | Maharashtra | 0 | 0 | 2 |
| **88** | 2020-03-10 | Maharashtra | 0 | 0 | 5 |
| **100** | 2020-03-11 | Maharashtra | 0 | 0 | 10 |
| **113** | 2020-03-12 | Maharashtra | 0 | 0 | 11 |
| **126** | 2020-03-13 | Maharashtra | 0 | 0 | 11 |
| **...** | ... | ... | ... | ... | ... |
| **4536** | 2020-08-02 | Maharashtra | 266883 | 15316 | 431719 |
| **4571** | 2020-08-03 | Maharashtra | 276809 | 15576 | 441228 |
| **4606** | 2020-08-04 | Maharashtra | 287030 | 15842 | 450196 |
| **4641** | 2020-08-05 | Maharashtra | 299356 | 16142 | 457956 |
| **4676** | 2020-08-06 | Maharashtra | 305521 | 16476 | 468265 |

147 rows × 5 columns

In [25]:

*#converting date time to ordinal*

In [26]:

maha['date']**=**maha['date']**.**map(dt**.**datetime**.**toordinal)

maha**.**head()

C:\Users\Sakshi\AppData\Local\Temp\ipykernel\_12076\4140285815.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

maha['date']=maha['date'].map(dt.datetime.toordinal)

Out[26]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **76** | 737493 | Maharashtra | 0 | 0 | 2 |
| **88** | 737494 | Maharashtra | 0 | 0 | 5 |
| **100** | 737495 | Maharashtra | 0 | 0 | 10 |
| **113** | 737496 | Maharashtra | 0 | 0 | 11 |
| **126** | 737497 | Maharashtra | 0 | 0 | 11 |

In [27]:

*#getting dependent variable and independent variable*

x **=** maha['date']

y **=** maha['confirmed']

In [28]:

x\_train, x\_test, y\_train, y\_test **=** train\_test\_split(x, y, test\_size **=** 0.3)

In [29]:

**from** sklearn.linear\_model **import** LinearRegression

In [30]:

lr **=** LinearRegression()

In [31]:

y\_train

Out[31]:

3389 152765

113 11

433 180

3809 246600

3179 124331

...

3494 169883

1513 11506

1939 27524

4088 310455

2339 54758

Name: confirmed, Length: 102, dtype: int64

In [32]:

lr**.**fit(np**.**array(x\_train)**.**reshape(**-**1, 1), np**.**array(y\_train)**.**reshape(**-**1, 1))

Out[32]:

  LinearRegression[?](https://scikit-learn.org/1.4/modules/generated/sklearn.linear_model.LinearRegression.html)i

LinearRegression()

In [33]:

maha**.**tail()

Out[33]:

|  | **date** | **state** | **cured** | **deaths** | **confirmed** |
| --- | --- | --- | --- | --- | --- |
| **4536** | 737639 | Maharashtra | 266883 | 15316 | 431719 |
| **4571** | 737640 | Maharashtra | 276809 | 15576 | 441228 |
| **4606** | 737641 | Maharashtra | 287030 | 15842 | 450196 |
| **4641** | 737642 | Maharashtra | 299356 | 16142 | 457956 |
| **4676** | 737643 | Maharashtra | 305521 | 16476 | 468265 |

In [34]:

lr**.**predict(np**.**array([[737643]]))

Out[34]:

array([[309745.16043139]])

In [ ]: