INDIAN COVID DATA ANALYSIS

- Name - Prathamesh Rajesh Sonar
- College - Amrutvahini College Of Engineering
- project - A Case Study on Indian covid-19 phase
- Date - 2023/02/02

In [1]:
```python
# !pip install matplotlib seaborn --upgrade --quite
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```python
df = pd.read_csv("covid_19_india.csv")
```

In [3]:
```python
df
```

Out[3]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational
0	1	2020-01-30	6:00 PM	Kerala	1	0
1	2	2020-01-31	6:00 PM	Kerala	1	0
2	3	2020-02-01	6:00 PM	Kerala	2	0
3	4	2020-02-02	6:00 PM	Kerala	3	0
4	5	2020-02-03	6:00 PM	Kerala	3	0
...	...	...	...	...	...	...
18105	18106	2021-08-11	8:00 AM	Telangana	-	-
18106	18107	2021-08-11	8:00 AM	Tripura	-	-
18107	18108	2021-08-11	8:00 AM	Uttarakhand	-	-
18108	18109	2021-08-11	8:00 AM	Uttar Pradesh	-	-
18109	18110	2021-08-11	8:00 AM	West Bengal	-	-

18110 rows × 9 columns

In [4]: `df.head(5)`

Out[4]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured
0	1	2020-01-30	6:00 PM	Kerala	1	0	0
1	2	2020-01-31	6:00 PM	Kerala	1	0	0
2	3	2020-02-01	6:00 PM	Kerala	2	0	0
3	4	2020-02-02	6:00 PM	Kerala	3	0	0
4	5	2020-02-03	6:00 PM	Kerala	3	0	0

In [5]: `df.tail(2)`

Out[5]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational
18108	18109	2021-08-11	8:00 AM	Uttar Pradesh	-	-
18109	18110	2021-08-11	8:00 AM	West Bengal	-	-

- data is from staring from 2020/01 to 2021/08
- we have 20 months of data

In [6]: `df.info();`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18110 entries, 0 to 18109
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Sno                       18110 non-null  int64
 1   Date                      18110 non-null  object
 2   Time                      18110 non-null  object
 3   State/UnionTerritory      18110 non-null  object
 4   ConfirmedIndianNational   18110 non-null  object
 5   ConfirmedForeignNational  18110 non-null  object
 6   Cured                     18110 non-null  int64
 7   Deaths                    18110 non-null  int64
 8   Confirmed                 18110 non-null  int64
dtypes: int64(4), object(5)
memory usage: 1.2+ MB
```

- We have data of 18110 rows and 9 columns

In [7]: `df.describe()`

Out[7]:

	Sno	Cured	Deaths	Confirmed
count	18110.000000	1.811000e+04	18110.000000	1.811000e+04
mean	9055.500000	2.786375e+05	4052.402264	3.010314e+05
std	5228.051023	6.148909e+05	10919.076411	6.561489e+05
min	1.000000	0.000000e+00	0.000000	0.000000e+00
25%	4528.250000	3.360250e+03	32.000000	4.376750e+03
50%	9055.500000	3.336400e+04	588.000000	3.977350e+04
75%	13582.750000	2.788698e+05	3643.750000	3.001498e+05
max	18110.000000	6.159676e+06	134201.000000	6.363442e+06

In [8]: `df['Date'] = pd.to_datetime(df['Date'])`

In [9]: `df = df.drop(['Sno','Time'],axis=1)`

In [10]: `df`

Out[10]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
0	2020-01-30	Kerala	1	0	0	
1	2020-01-31	Kerala	1	0	0	
2	2020-02-01	Kerala	2	0	0	
3	2020-02-02	Kerala	3	0	0	
4	2020-02-03	Kerala	3	0	0	
...	...	...	...	...	...	...
18105	2021-08-11	Telangana	-	-	638410	3
18106	2021-08-11	Tripura	-	-	77811	
18107	2021-08-11	Uttarakhand	-	-	334650	7
18108	2021-08-11	Uttar Pradesh	-	-	1685492	22
18109	2021-08-11	West Bengal	-	-	1506532	18

18110 rows × 7 columns

In [11]: `df.isnull().sum()`

Out[11]:
```
Date                       0
State/UnionTerritory       0
ConfirmedIndianNational    0
ConfirmedForeignNational   0
Cured                      0
Deaths                     0
Confirmed                  0
dtype: int64
```

- There is no null value in data

In [12]:
```
total_death=df.Deaths.sum()
total_cure=df.Cured.sum()
total_confirm=df.Confirmed.sum()
```

In [13]:
```
total_death,total_cure,total_confirm
print('Total Deaths are = {}'.format(total_death))
```

```
print('Total Cure are = {}'.format(total_cure))
print('Total Confirmed are = {}'.format(total_confirm))
```

```
Total Deaths are = 73389005
Total Cure are = 5046125452
Total Confirmed are = 5451678687
```

In [14]:
```
error = total_confirm - (total_death+total_cure)
print('So, there seems to have some missing data around {}, regarding the people who w
```

So, there seems to have some missing data around 332164230, regarding the people who where positive.

- Error is the data which is not known
- may be the test result are in process

In [15]:
```
death_rate=(total_death*100)/total_confirm
death_rate
print("The overall death rate reported in INDIA is = {:.2f}%".format(death_rate))
```

The overall death rate reported in INDIA is = 1.35%

In [16]:
```
df['deaths_sum']=df['Deaths'].cumsum()
df['cured_sum']=df['Cured'].cumsum()
df['confirmed_sum']=df['Confirmed'].cumsum()
df
```

Out[16]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
0	2020-01-30	Kerala	1	0	0	
1	2020-01-31	Kerala	1	0	0	
2	2020-02-01	Kerala	2	0	0	
3	2020-02-02	Kerala	3	0	0	
4	2020-02-03	Kerala	3	0	0	
...	...	...	...	...	...	...
18105	2021-08-11	Telangana	-	-	638410	3
18106	2021-08-11	Tripura	-	-	77811	
18107	2021-08-11	Uttarakhand	-	-	334650	7
18108	2021-08-11	Uttar Pradesh	-	-	1685492	22
18109	2021-08-11	West Bengal	-	-	1506532	18

18110 rows × 10 columns

In [17]:
```python
df.sort_values(by=['Deaths'], ascending=False)
```

Out[17]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
18094	2021-08-11	Maharashtra	-	-	6159676	134
18058	2021-08-10	Maharashtra	-	-	6151956	134
18022	2021-08-09	Maharashtra	-	-	6144388	133
17986	2021-08-08	Maharashtra	-	-	6139493	133
17950	2021-08-07	Maharashtra	-	-	6130137	133
...	...	...	...	...	...	
989	2020-04-15	Uttarakhand	-	-	9	
2074	2020-05-19	Dadra and Nagar Haveli and Daman and Diu	-	-	0	
5148	2020-08-13	Mizoram	-	-	330	
3610	2020-07-01	Dadra and Nagar Haveli and Daman and Diu	-	-	82	
0	2020-01-30	Kerala	1	0	0	

18110 rows × 10 columns

In [18]: ```df.sort_values('Confirmed',ascending=False).head(10)```

Out[18]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
18094	2021-08-11	Maharashtra	-	-	6159676	134
18058	2021-08-10	Maharashtra	-	-	6151956	134
18022	2021-08-09	Maharashtra	-	-	6144388	133
17986	2021-08-08	Maharashtra	-	-	6139493	133
17950	2021-08-07	Maharashtra	-	-	6130137	133
17914	2021-08-06	Maharashtra	-	-	6124278	133
17878	2021-08-05	Maharashtra	-	-	6117560	133
17842	2021-08-04	Maharashtra	-	-	6110124	133
17806	2021-08-03	Maharashtra	-	-	6103325	133
17770	2021-08-02	Maharashtra	-	-	6094896	132

In [19]:
```python
df.sort_values('Cured',ascending=False).head(100)
```

Out[19]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
18094	2021-08-11	Maharashtra	-	-	6159676	134
18058	2021-08-10	Maharashtra	-	-	6151956	134
18022	2021-08-09	Maharashtra	-	-	6144388	133
17986	2021-08-08	Maharashtra	-	-	6139493	133
17950	2021-08-07	Maharashtra	-	-	6130137	133
...	...	...	...	...	...	...
14674	2021-05-08	Maharashtra	-	-	4265326	74
14638	2021-05-07	Maharashtra	-	-	4227940	73
14602	2021-05-06	Maharashtra	-	-	4164098	72
14566	2021-05-05	Maharashtra	-	-	4107092	71
14530	2021-05-04	Maharashtra	-	-	4041158	70

100 rows × 10 columns

- Amoun all states MAHARASTRA had most deaths,cured and Confirmed cases in India
- On 2021-08-11 Date MAHARASTRA had most deaths,cured and Confirmed cases

In [20]:
```python
df['year'] = pd.DatetimeIndex(df['Date']).year
df['month'] = pd.DatetimeIndex(df['Date']).month
df['day'] = pd.DatetimeIndex(df['Date']).day
df['weekday'] = pd.DatetimeIndex(df['Date']).weekday
df
```

Out[20]:

	Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
0	2020-01-30	Kerala	1	0	0	
1	2020-01-31	Kerala	1	0	0	
2	2020-02-01	Kerala	2	0	0	
3	2020-02-02	Kerala	3	0	0	
4	2020-02-03	Kerala	3	0	0	
...	...	...	...	...	...	
18105	2021-08-11	Telangana	-	-	638410	3
18106	2021-08-11	Tripura	-	-	77811	
18107	2021-08-11	Uttarakhand	-	-	334650	7
18108	2021-08-11	Uttar Pradesh	-	-	1685492	22
18109	2021-08-11	West Bengal	-	-	1506532	18

18110 rows × 14 columns

In [21]: `df['country'] ='India'`

In [22]: `df.set_index('Date',inplace= True)`

In [23]: `df`

Out[23]:

Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Co
2020-01-30	Kerala	1	0	0	0	0
2020-01-31	Kerala	1	0	0	0	0
2020-02-01	Kerala	2	0	0	0	0
2020-02-02	Kerala	3	0	0	0	0
2020-02-03	Kerala	3	0	0	0	0
...	...	...	...	...	...	...
2021-08-11	Telangana	-	-	638410	3831	
2021-08-11	Tripura	-	-	77811	773	
2021-08-11	Uttarakhand	-	-	334650	7368	
2021-08-11	Uttar Pradesh	-	-	1685492	22775	
2021-08-11	West Bengal	-	-	1506532	18252	

18110 rows × 14 columns

In [24]:
```python
p=df.groupby('State/UnionTerritory')[['Cured','Deaths','Confirmed']].sum()
p
```

Out[24]:

State/UnionTerritory	Cured	Deaths	Confirmed
Andaman and Nicobar Islands	1848286	27136	1938498
Andhra Pradesh	370426530	2939367	392432753
Arunachal Pradesh	6588149	26799	7176907
Assam	92678680	638323	99837011
Bihar	125122902	1093466	132231166
Bihar****	1402468	18881	1430909
Cases being reassigned to states	0	0	345565
Chandigarh	10117035	147694	10858627
Chhattisgarh	151609364	2063920	163776262
Dadra and Nagar Haveli	20352	8	20722
Dadra and Nagar Haveli and Daman and Diu	1841750	1014	1938632
Daman & Diu	0	0	2
Delhi	273419887	4943294	287227765
Goa	26027201	447801	28240159
Gujarat	132487127	2219448	143420082
Haryana	126585342	1502799	134347285
Himachal Pradesh	27501110	491348	30033289
Himanchal Pradesh	200040	3507	204516
Jammu and Kashmir	53297341	839694	58117726
Jharkhand	58034506	748641	62111994
Karanataka	2821491	36197	2885238
Karnataka	441844360	6053762	485970693
Kerala	420174235	1888177	458906023
Ladakh	3758960	45804	4054293
Lakshadweep	820925	3908	915784
Madhya Pradesh	126724997	1777752	135625265
Madhya Pradesh***	780735	10506	791656
Maharashtra	1018765039	23737432	1121491467
Maharashtra***	6000911	130753	6229596
Manipur	11230568	173056	12617943
Meghalaya	6537909	101950	7355969
Mizoram	2384602	9791	2984732

State/UnionTerritory	Cured	Deaths	Confirmed
Nagaland	4519526	58460	5041742
Odisha	150923455	790814	160130533
Puducherry	18483117	312155	20065891
Punjab	91458159	2785594	99949702
Rajasthan	150356820	1473089	162369656
Sikkim	2747214	53150	3186799
Tamil Nadu	404095807	5916658	431928644
Telangana	57488245	349648	60571979
Telengana	64666267	400427	69990668
Tripura	12976846	150342	14050250
Unassigned	0	0	161
Uttar Pradesh	291479351	4143450	312625843
Uttarakhand	48362741	986001	53140414
West Bengal	247515102	3846989	263107876

In [25]:
```python
total_cure_rate = (df.Cured.sum()*100)/df.Confirmed.sum()
```

In [26]:
```python
total_cure_rate
```

Out[26]: 92.56094758542764

In [27]:
```python
total_death_rate = (df.Deaths.sum()*100)/df.Confirmed.sum()
total_death_rate
```

Out[27]: 1.3461726050547043

In [28]:
```python
100 - ( 92.56094758542764 + 1.3461726050547043 )
```

Out[28]: 6.092879809517655

- Accoring to data Out of Covid-19 Positive cases There are 92.5 % people got cured and 1.3 % people Died and 6.09 % people are still positive.

In [29]:
```python
df_new = df.groupby('month')[['Cured','Deaths','Confirmed']].sum()
df_new
```

Out[29]:

month	Cured	Deaths	Confirmed
1	315332019	4709167	326469749
2	297133802	4359434	305631889
3	342611205	4935455	356315303
4	385065633	5353568	441083113
5	646240106	8480751	754865720
6	848822379	11475067	895231838
7	955269787	13575039	993362865
8	399775889	6247991	431100375
9	118592934	2443374	149113758
10	198824412	3457615	226770312
11	246213201	3894165	264556412
12	292244085	4457379	307177353

- JULY month had reported Highest numbers of Cured,Deaths and Confirmed cases for COVID-19
- SEPTEMBER month had reported Lowest numbers of Cured,Deaths and Confirmed cases for COVID-19

In [30]:
```
df.groupby('month')[['Cured','Deaths','Confirmed']].mean()
```

Out[30]:

month	Cured	Deaths	Confirmed
1	282050.106440	4212.135063	292012.297853
2	286532.113790	4203.890068	294726.990357
3	212274.600372	3057.902726	220765.367410
4	189780.992114	2638.525382	217389.410054
5	299185.234259	3926.273611	349474.870370
6	392791.475706	5310.072652	414267.393799
7	430495.622803	6117.638125	447662.399730
8	269936.454423	4218.765024	291087.356516
9	112945.651429	2327.022857	142013.102857
10	183248.305991	3186.741935	209004.895853
11	234488.762857	3708.728571	251958.487619
12	264235.158228	4030.179928	277737.208861

*

In [31]: `df[['Cured','Deaths','Confirmed']].mean()`

Out[31]:
```
Cured          278637.518056
Deaths           4052.402264
Confirmed      301031.401822
dtype: float64
```

- On an average in INDIA per day 3 lakh cases were reported and 2.78 lakh people were cured and saddly 4 thousand people died ...

In [32]: `df.groupby('weekday')[['Cured','Deaths','Confirmed']].mean()`

Out[32]:

weekday	Cured	Deaths	Confirmed
0	279798.809799	4068.179398	302337.370756
1	281032.787447	4081.821332	303305.410859
2	281996.317204	4096.321429	304230.003840
3	274751.386019	4000.768544	297053.742136
4	276143.010861	4019.686967	298546.529480
5	277561.117396	4039.494382	300044.002712
6	279117.163376	4059.739063	301641.978320

- Amoung weekdays WEDNESDAY had reported Highest numbers of Cured,Deaths and Confirmed cases for COVID-19
- Amoung weekdays THURSDAY had reported Lowest numbers of Cured,Deaths and Confirmed cases for COVID-19

In [33]: `# df.to_csv('india_covid_data.csv',index=None)`

In [34]: `p`

Out[34]:

State/UnionTerritory	Cured	Deaths	Confirmed
Andaman and Nicobar Islands	1848286	27136	1938498
Andhra Pradesh	370426530	2939367	392432753
Arunachal Pradesh	6588149	26799	7176907
Assam	92678680	638323	99837011
Bihar	125122902	1093466	132231166
Bihar****	1402468	18881	1430909
Cases being reassigned to states	0	0	345565
Chandigarh	10117035	147694	10858627
Chhattisgarh	151609364	2063920	163776262
Dadra and Nagar Haveli	20352	8	20722
Dadra and Nagar Haveli and Daman and Diu	1841750	1014	1938632
Daman & Diu	0	0	2
Delhi	273419887	4943294	287227765
Goa	26027201	447801	28240159
Gujarat	132487127	2219448	143420082
Haryana	126585342	1502799	134347285
Himachal Pradesh	27501110	491348	30033289
Himanchal Pradesh	200040	3507	204516
Jammu and Kashmir	53297341	839694	58117726
Jharkhand	58034506	748641	62111994
Karanataka	2821491	36197	2885238
Karnataka	441844360	6053762	485970693
Kerala	420174235	1888177	458906023
Ladakh	3758960	45804	4054293
Lakshadweep	820925	3908	915784
Madhya Pradesh	126724997	1777752	135625265
Madhya Pradesh***	780735	10506	791656
Maharashtra	1018765039	23737432	1121491467
Maharashtra***	6000911	130753	6229596
Manipur	11230568	173056	12617943
Meghalaya	6537909	101950	7355969
Mizoram	2384602	9791	2984732

|	Cured	Deaths	Confirmed
State/UnionTerritory			
Nagaland	4519526	58460	5041742
Odisha	150923455	790814	160130533
Puducherry	18483117	312155	20065891
Punjab	91458159	2785594	99949702
Rajasthan	150356820	1473089	162369656
Sikkim	2747214	53150	3186799
Tamil Nadu	404095807	5916658	431928644
Telangana	57488245	349648	60571979
Telengana	64666267	400427	69990668
Tripura	12976846	150342	14050250
Unassigned	0	0	161
Uttar Pradesh	291479351	4143450	312625843
Uttarakhand	48362741	986001	53140414
West Bengal	247515102	3846989	263107876

```
In [35]:  plt.figure(figsize=(20,8))
          plt.plot(p.Deaths,'o--b')
          plt.title('Covid-Deaths')
          plt.xlabel('State')
          plt.ylabel('count')
          plt.xticks(rotation=50)
          sns.set_style('darkgrid');
```
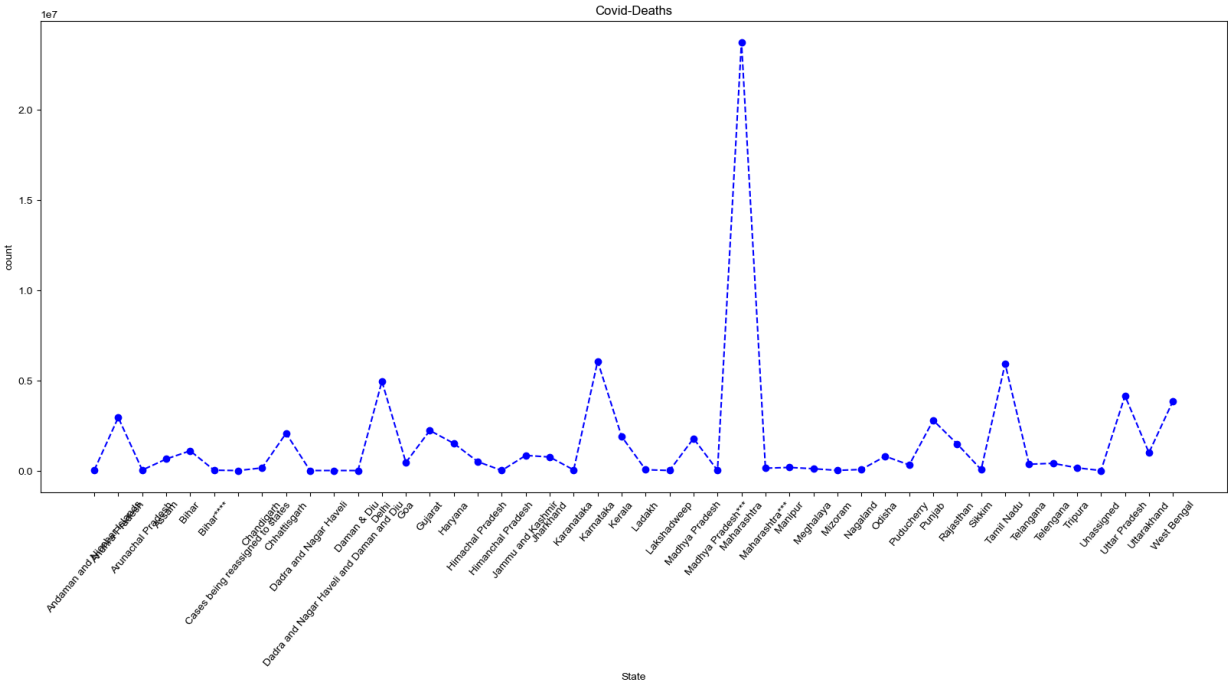
From graph, it is clear that the MAHARASHTRA has highest Deaths. followed by Karnatka, Kerala, Tamil nadu but it is not even half of Maharashtra...

```
In [36]: plt.figure(figsize=(20,8))
         plt.plot(p.Cured,'o-r',alpha=0.3)
         plt.plot(p.Deaths,'s--b',alpha=.3)
         plt.plot(p.Confirmed,'o--y')
         sns.set_style('whitegrid')

         plt.xlabel('State')
         plt.ylabel('count of people')
         plt.title('Covid-19 data')
         plt.legend(['Cured','Deaths','Confirmed'])
         plt.xticks(rotation=50);
```
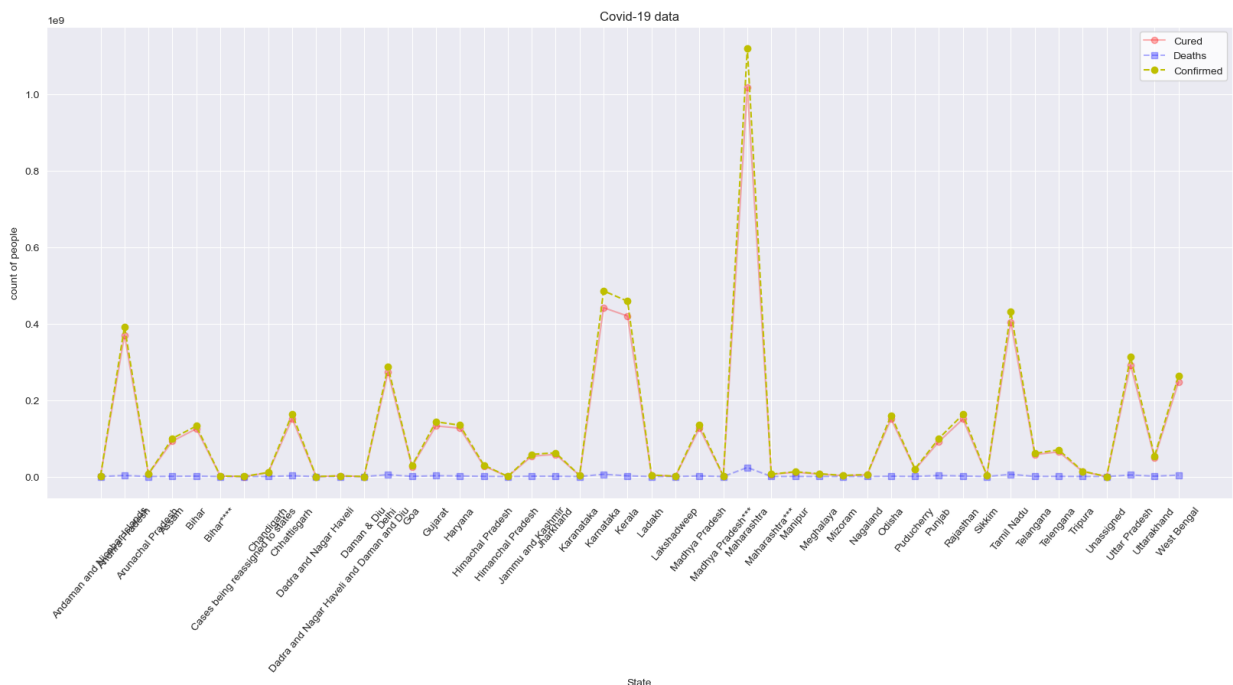

From graph,

it is clear that the MAHARASHTRA has confirmed and cured cases. followed by Karnatka, Kerala, Tamil nadu but it is half of Maharashtra...

```
In [37]: df
```

Out[37]:

Date	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Co
2020-01-30	Kerala	1	0	0	0	
2020-01-31	Kerala	1	0	0	0	
2020-02-01	Kerala	2	0	0	0	
2020-02-02	Kerala	3	0	0	0	
2020-02-03	Kerala	3	0	0	0	
...	...	...	...	...	...	...
2021-08-11	Telangana	-	-	638410	3831	
2021-08-11	Tripura	-	-	77811	773	
2021-08-11	Uttarakhand	-	-	334650	7368	
2021-08-11	Uttar Pradesh	-	-	1685492	22775	
2021-08-11	West Bengal	-	-	1506532	18252	

18110 rows × 14 columns

In [38]:
```python
plt.figure(figsize=(10,5))
plt.title('Deaths as per months')
plt.xlabel('MONTHS...')
plt.ylabel('Count of people...')
plt.bar(df.month,df.Deaths);
```

Deaths as per months
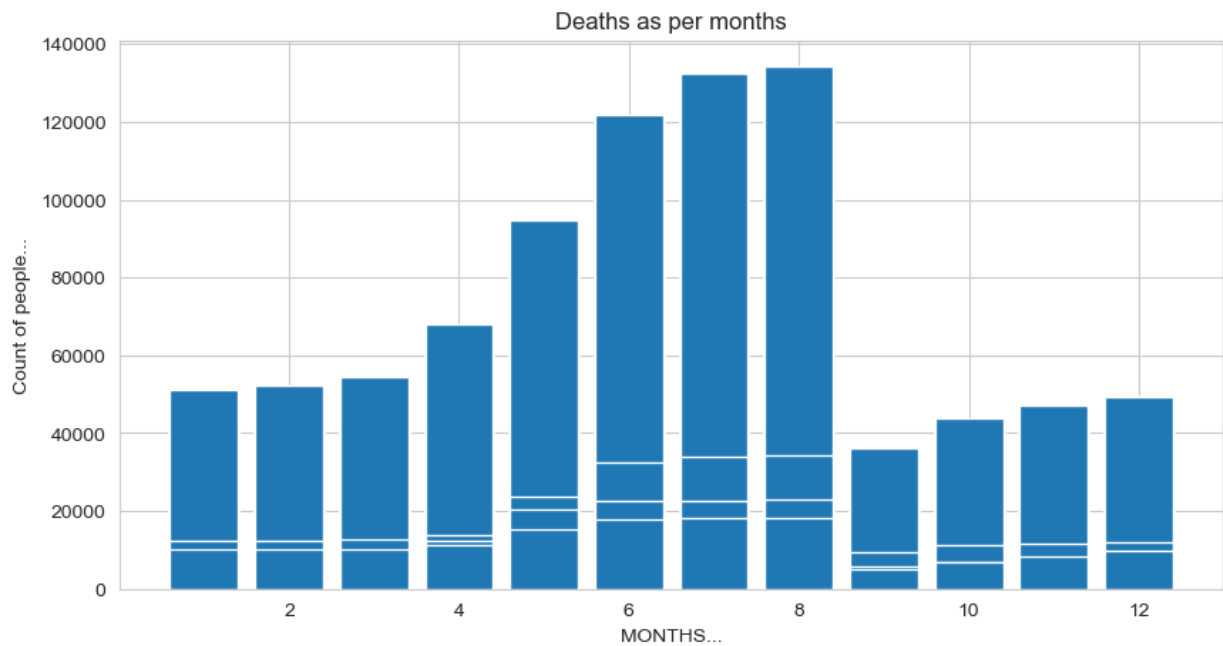
- The Deaths cases is increased in May to August and the declined rapidly.

In [39]:
```python
#plt.pie(p[['Cured','Deaths','Confirmed']],
#labels = p['State/UnionTerritory'],
#startangle = 90,
#shadow = True,
#explode =(0,0.1,0,0,0),
#autopct ='%1.1f%%')
```

In [40]:
```python
plt.figure(figsize=(10,5))
plt.title('covid data as per months')
plt.xlabel('MONTHS...')
plt.ylabel('Count of people...')
plt.bar(df.month,df.Confirmed)
plt.bar(df.month,df.Cured)
plt.legend(['Confirmed','Cured']);
```

covid data as per months
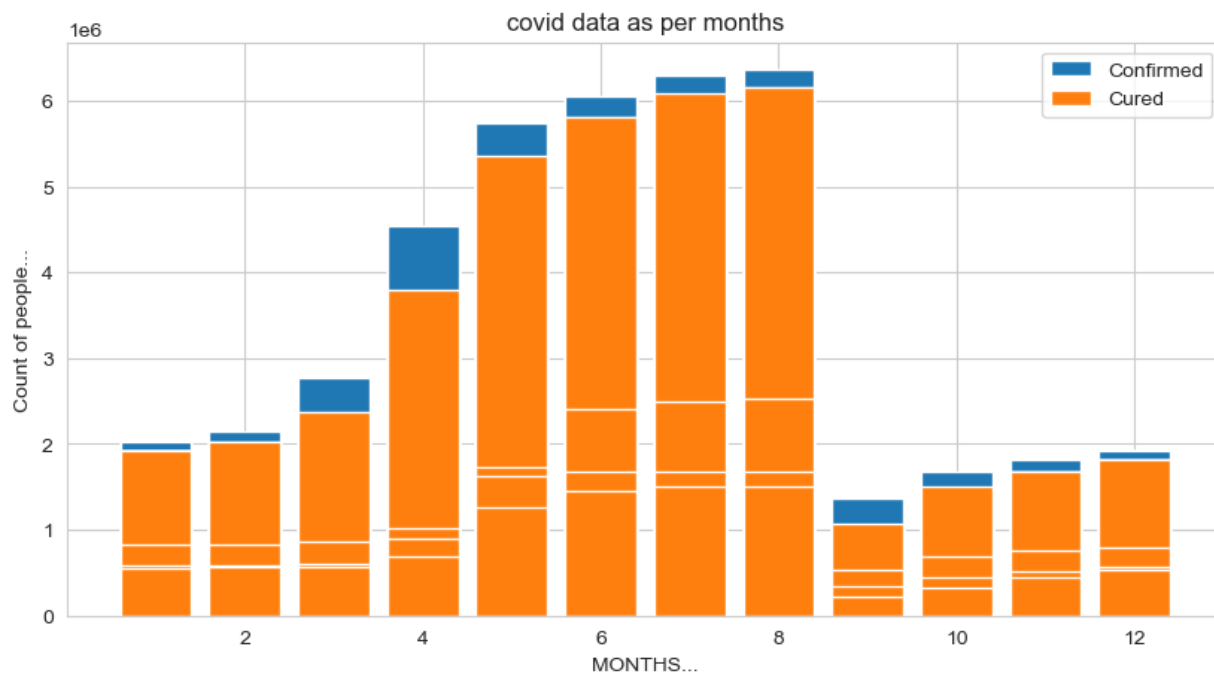
- The Confirmed and Cured cases is increased in May to August and the declined rapidly.

We can draw more insights

THANK YOU