Importing libraries

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Reading the csv File

In [2]:

```
df=pd.read_csv("Covid19_July1-July 7th.csv")
```

In [3]:

```
df
```

Out[3]:

	Entry_ID	State Patient Number	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	Stat cod
0	94849	NaN	01/07/2020	NaN	NaN	NaN	Unassigned	State Unassigned	U
1	94850	NaN	01/07/2020	NaN	NaN	NaN	NaN	Nagaland	N
2	94851	NaN	01/07/2020	NaN	NaN	NaN	Angul	Odisha	O
3	94852	NaN	01/07/2020	NaN	NaN	NaN	Balasore	Odisha	O
4	94853	NaN	01/07/2020	NaN	NaN	NaN	Bargarh	Odisha	O
...	...	...	...	...	...	...	...	...	
22790	117831	NaN	07/07/2020	NaN	NaN	NaN	Dadra and Nagar Haveli	Dadra and Nagar Haveli and Daman and Diu	D
22791	118420	NaN	07/07/2020	NaN	NaN	NaN	NaN	Sikkim	S
22792	118421	NaN	07/07/2020	NaN	NaN	NaN	NaN	Sikkim	S
22793	121323	NaN	07/07/2020	NaN	NaN	NaN	Kozhikode	Kerala	K
22794	121324	NaN	07/07/2020	NaN	NaN	NaN	Wayanad	Kerala	K

22795 rows × 20 columns

In [ ]:

```
```

In [ ]:

```
```

In [ ]:

```
```

Inspecting the dataframe

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22795 entries, 0 to 22794
Data columns (total 20 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Entry_ID                                  22795 non-null  int64  
 1   State Patient Number                      14924 non-null  object 
 2   Date Announced                            22795 non-null  object 
 3   Age Bracket                               12538 non-null  float64
 4   Gender                                    12583 non-null  object 
 5   Detected City                             772 non-null    object 
 6   Detected District                         22556 non-null  object 
 7   Detected State                            22795 non-null  object 
 8   State code                                22795 non-null  object 
 9   Num Cases                                 22795 non-null  int64  
 10  Current Status                            22795 non-null  object 
 11  Contracted from which Patient (Suspected)  0 non-null      float64
 12  Notes                                     12408 non-null  object 
 13  Source_1                                  22618 non-null  object 
 14  Source_2                                  97 non-null     object 
 15  Source_3                                  0 non-null      float64
 16  Nationality                               0 non-null      float64
 17  Type of transmission                      0 non-null      float64
 18  Status Change Date                        0 non-null      float64
 19  Patient Number                            0 non-null      float64
dtypes: float64(7), int64(2), object(11)
memory usage: 3.5+ MB
```

Inspecting null values in each columns

In [5]:

```
df.isnull().sum().sort_values(ascending=False)/len(df)*100
```

Out[5]:

```
Patient Number                                 100.000000
Status Change Date                             100.000000
Type of transmission                           100.000000
Nationality                                    100.000000
Source_3                                       100.000000
Contracted from which Patient (Suspected)      100.000000
Source_2                                        99.574468
Detected City                                   96.613292
Notes                                           45.567010
Age Bracket                                     44.996710
Gender                                          44.799298
State Patient Number                            34.529502
Detected District                                1.048476
Source_1                                         0.776486
Entry_ID                                         0.000000
Num Cases                                        0.000000
State code                                       0.000000
Detected State                                   0.000000
Date Announced                                   0.000000
Current Status                                   0.000000
dtype: float64
```

Inspecting null values in rows

In [6]:

```
df.isnull().sum(axis=1).sort_values(ascending=False)
```

Out[6]:

```
12453    14
8790     14
8789     14
16725    13
16796    13
         ..
11557     8
11558     8
11559     8
11397     8
19383     7
Length: 22795, dtype: int64
```

In [7]:

```
df_nonHosp=df[df['Current Status']=="Hospitalized"]
```

Total number of cases on average per day

In [8]:

```
D=df_nonHosp.groupby('Date Announced')['Num Cases'].sum()
D
```

Out[8]:

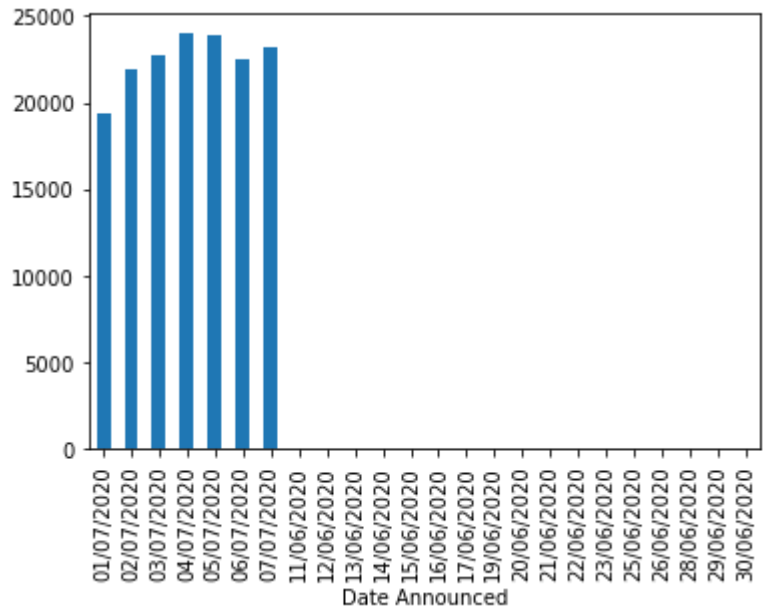
```
Date Announced
01/07/2020    19429
02/07/2020    21947
03/07/2020    22718
04/07/2020    24018
05/07/2020    23942
06/07/2020    22500
07/07/2020    23147
11/06/2020        0
12/06/2020        0
13/06/2020        0
14/06/2020        0
15/06/2020        0
16/06/2020        0
17/06/2020        0
19/06/2020        0
20/06/2020        0
21/06/2020        0
22/06/2020        0
23/06/2020        0
25/06/2020        0
26/06/2020        0
28/06/2020        0
29/06/2020        0
30/06/2020        0
Name: Num Cases, dtype: int64
```

In [9]:

```
D.plot.bar()
```

Out[9]:

```
<AxesSubplot:xlabel='Date Announced'>
```



Total Number of Male/Female Affected percent vise

In [10]:

```
df_nonHosp.groupby('Gender')['Num Cases'].sum()/len(df[df['Current Status']=="Hospitalized"
```

Out[10]:

```
Gender
F             27.064020
M             49.916008
Non-Binary     0.031108
Name: Num Cases, dtype: float64
```

This shows that 50 % of the people affected with the corona virus are men

Which age group is mostly affected

In [11]:

```
df_nonHosp.groupby('Age Bracket')['Num Cases'].sum().sort_values(ascending=False).head(10)
```

Out[11]:

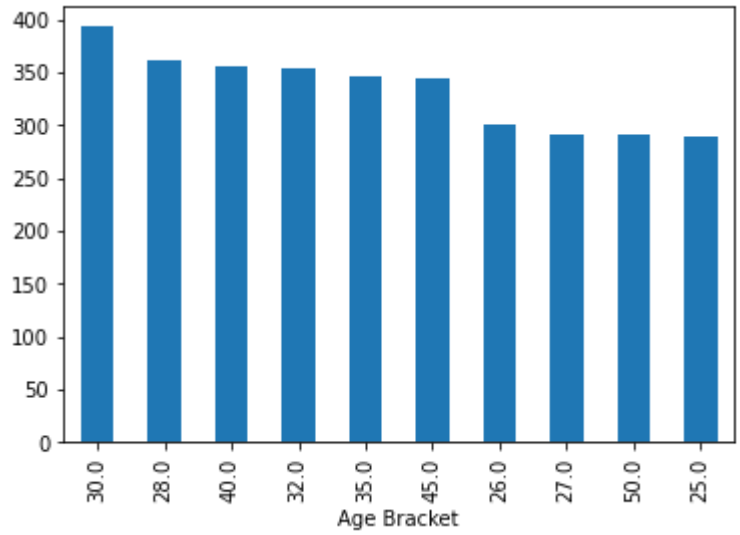
```
Age Bracket
30.0    393
28.0    362
40.0    355
32.0    354
35.0    346
45.0    344
26.0    300
27.0    292
50.0    291
25.0    290
Name: Num Cases, dtype: int64
```

In [12]:

```
df_nonHosp.groupby('Age Bracket')['Num Cases'].sum().sort_values(ascending=False).head(10).
```

Out[12]:

<AxesSubplot:xlabel='Age Bracket'>



This concludes that the age affected the most rely between 25 to 50

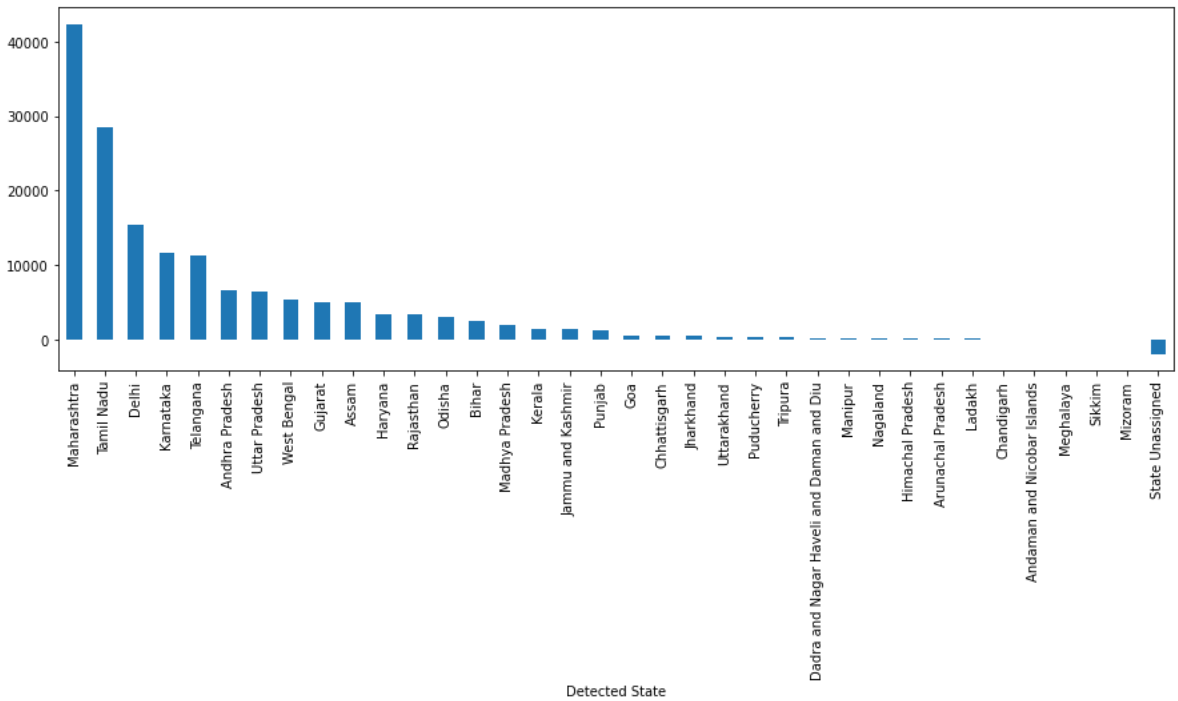
Total cases state vise

In [13]:

```
df_nonHosp.groupby('Detected State')['Num Cases'].sum().sort_values(ascending=False).plot.b
```

Out[13]:

<AxesSubplot:xlabel='Detected State'>



This shows that Maharashtra is the state that has mostly been affected

Number of people deceased

In [14]:

```
df['Num Cases'][df['Current Status']=='Deceased'].sum()
```

Out[14]:

```
3243
```

Number of People Deceased per state

In [15]:

```
df[df['Current Status']=='Deceased'].groupby('Detected State')['Num Cases'].sum().sort_valu
```

Out[15]:

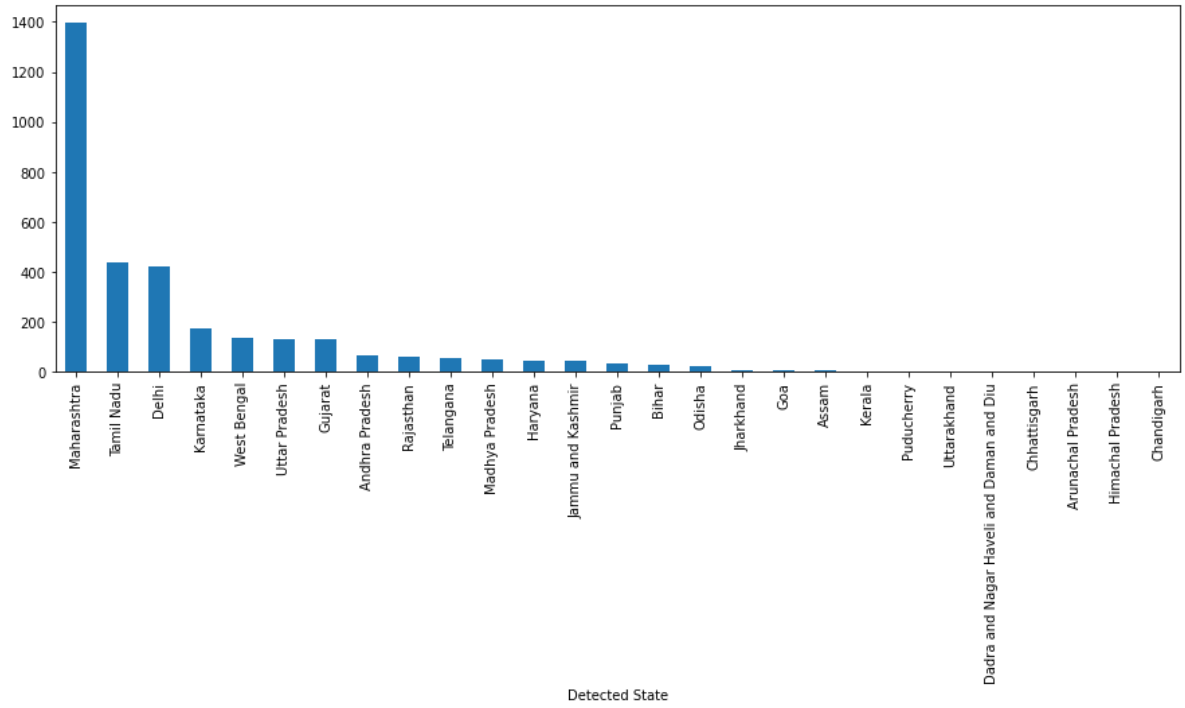
```
Detected State
Maharashtra                                1395
Tamil Nadu                                  435
Delhi                                       423
Karnataka                                   171
West Bengal                                 136
Uttar Pradesh                               130
Gujarat                                     130
Andhra Pradesh                               65
Rajasthan                                    59
Telangana                                    53
Madhya Pradesh                               50
Haryana                                      43
Jammu and Kashmir                            42
Punjab                                       31
Bihar                                        30
Odisha                                       22
Jharkhand                                     7
Goa                                           5
Assam                                         4
Kerala                                        3
Puducherry                                    2
Uttarakhand                                   2
Dadra and Nagar Haveli and Daman and Diu      1
Chhattisgarh                                  1
Arunachal Pradesh                             1
Himachal Pradesh                              1
Chandigarh                                    1
Name: Num Cases, dtype: int64
```

In [16]:

```
'Current Status']=='Deceased'].groupby('Detected State')['Num Cases'].sum().sort_values(asce
```

Out[16]:

```
<AxesSubplot:xlabel='Detected State'>
```



With this we can conclude that most of the deaths has been in Maharashtra

In [ ]: