

In [3]:

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

In [4]:

```
df1 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data21.csv')
df2 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data22.csv')
df3 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data23.csv')
df4 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data24.csv')
df5 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data25.csv')
df6 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data26.csv')
df7 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data27.csv')
df8 = pd.read_csv('https://api.covid19india.org/csv/latest/raw_data28.csv')
df1
```

Out [4] :

	Entry_ID	State Patient Number	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	State code	Num Cases	Current Status	Contracted from where Patient Suspected
	0	428033	NaN	20/12/2020	NaN	NaN	Aizawl	Mizoram	MZ	1	Hospitalized	NaN
	1	428034	NaN	20/12/2020	NaN	NaN	Lunglei	Mizoram	MZ	7	Hospitalized	NaN
	2	428035	NaN	20/12/2020	NaN	NaN	Kolasib	Mizoram	MZ	3	Hospitalized	NaN
	3	428036	NaN	20/12/2020	NaN	NaN	Saitual	Mizoram	MZ	1	Hospitalized	NaN
	4	428037	NaN	20/12/2020	NaN	NaN	Adilabad	Telangana	TG	11	Hospitalized	NaN

25491	453524	NaN	08/01/2021	NaN	NaN	NaN	Saraikele-Kharsawan	Jharkhand	JH	3	Hospitalized	NaN
25492	453525	NaN	08/01/2021	NaN	NaN	NaN	Saraikele-Kharsawan	Jharkhand	JH	2	Recovered	NaN
25493	453526	NaN	08/01/2021	NaN	NaN	NaN	Leh	Ladakh	LA	1	Hospitalized	NaN
25494	453527	NaN	08/01/2021	NaN	NaN	NaN	Leh	Ladakh	LA	11	Recovered	NaN
25495	453528	NaN	08/01/2021	NaN	NaN	NaN	Kargil	Ladakh	LA	2	Recovered	NaN

25496 rows x 20 columns

In [5]:

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25496 entries, 0 to 25495
Data columns (total 20 columns):
```

#	Column	Non-Null Count	Dtype
0	Entry_ID	25496 non-null	int64
1	State Patient Number	0 non-null	float64
2	Date Announced	25496 non-null	object
3	Age Bracket	212 non-null	float64
4	Gender	205 non-null	object
5	Detected City	11 non-null	object
6	Detected District	25161 non-null	object
7	Detected State	25496 non-null	object
8	State code	25496 non-null	object
9	Num Cases	25496 non-null	int64
10	Current Status	25496 non-null	object
11	Contracted from which Patient (Suspected)	0 non-null	float64
12	Not	221 non-null	object

```
12 Notes 201 non-null object
13 Source_1 24753 non-null object
14 Source_2 2163 non-null object
15 Source_3 63 non-null object
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: 3.9+ MB
```

In [6]:

```
df1.columns
df2.columns
df3.columns
df4.columns
df5.columns
df6.columns
df7.columns
df8.columns
```

Out[6]:

```
Index(['Entry_ID', 'State Patient Number', 'Date Announced', 'Age Bracket',
      'Gender', 'Detected City', 'Detected District', 'Detected State',
      'State code', 'Num Cases', 'Current Status',
      'Contracted from which Patient (Suspected)', 'Notes', 'Source_1',
      'Source_2', 'Source_3', 'Nationality', 'Type of transmission',
      'Status Change Date', 'Patient Number'],
      dtype='object')
```

In [7]:

```
df = df1.append([df2, df3, df4, df5, df6, df7, df8])
df
```

Out[7]:

	Entry_ID	State Patient Number	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	State code	Num Cases	Current Status	Contracte from whic Patier (Suspected)
0	428033	NaN	20/12/2020	NaN	NaN	NaN	Aizawl	Mizoram	MZ	1.0	Hospitalized	NaN
1	428034	NaN	20/12/2020	NaN	NaN	NaN	Lunglei	Mizoram	MZ	7.0	Hospitalized	NaN
2	428035	NaN	20/12/2020	NaN	NaN	NaN	Kolasib	Mizoram	MZ	3.0	Hospitalized	NaN
3	428036	NaN	20/12/2020	NaN	NaN	NaN	Saitual	Mizoram	MZ	1.0	Hospitalized	NaN
4	428037	NaN	20/12/2020	NaN	NaN	NaN	Adilabad	Telangana	TG	11.0	Hospitalized	NaN
...
17833	629709	NaN	NaN	NaN	NaN	NaN	Kargil	Ladakh	LA	1.0	Deceased	NaN
17834	629710	NaN	NaN	NaN	NaN	NaN	Leh	Ladakh	LA	196.0	Hospitalized	NaN
17835	629711	NaN	NaN	NaN	NaN	NaN	Kargil	Ladakh	LA	39.0	Hospitalized	NaN
17836	629712	NaN	NaN	NaN	NaN	NaN	Leh	Ladakh	LA	100.0	Recovered	NaN
17837	629713	NaN	NaN	NaN	NaN	NaN	Kargil	Ladakh	LA	30.0	Recovered	NaN

201296 rows × 20 columns



In [8]:

```
df = df.loc[:, ['Num Cases', 'Date Announced', 'Age Bracket',
               'Gender', 'Detected City', 'Detected District', 'Detected State',
               'Current Status']]
df
```

Out[8]:

	Num Cases	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	Current Status
0	1.0	20/12/2020	NaN	NaN	NaN	Aizawl	Mizoram	Hospitalized
1	7.0	20/12/2020	NaN	NaN	NaN	Lunglei	Mizoram	Hospitalized
2	3.0	20/12/2020	NaN	NaN	NaN	Kolasib	Mizoram	Hospitalized
3	1.0	20/12/2020	NaN	NaN	NaN	Saitual	Mizoram	Hospitalized
4	11.0	20/12/2020	NaN	NaN	NaN	Adilabad	Telangana	Hospitalized
...
17833	1.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Deceased
17834	196.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Hospitalized
17835	39.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Hospitalized
17836	100.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Recovered
17837	30.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Recovered

201296 rows × 8 columns

In [9]:

```
DATE = df['Date Announced'].str.split('/',expand = True)
DATE.columns = ['Day','Month','Year']
DATE
```

Out[9]:

	Day	Month	Year
0	20	12	2020
1	20	12	2020
2	20	12	2020
3	20	12	2020
4	20	12	2020
...
17833	NaN	NaN	NaN
17834	NaN	NaN	NaN
17835	NaN	NaN	NaN
17836	NaN	NaN	NaN
17837	NaN	NaN	NaN

201296 rows × 3 columns

In [10]:

```
df = pd.concat([df,DATE],axis=1)
df
```

Out[10]:

	Num Cases	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	Current Status	Day	Month	Year
0	1.0	20/12/2020	NaN	NaN	NaN	Aizawl	Mizoram	Hospitalized	20	12	2020
1	7.0	20/12/2020	NaN	NaN	NaN	Lunglei	Mizoram	Hospitalized	20	12	2020
2	3.0	20/12/2020	NaN	NaN	NaN	Kolasib	Mizoram	Hospitalized	20	12	2020
3	1.0	20/12/2020	NaN	NaN	NaN	Saitual	Mizoram	Hospitalized	20	12	2020
4	11.0	20/12/2020	NaN	NaN	NaN	Adilabad	Telangana	Hospitalized	20	12	2020

...	Num Cases	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	Current Status	Day	Month	Year
17833	1.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Deceased	NaN	NaN	NaN
17834	196.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Hospitalized	NaN	NaN	NaN
17835	39.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Hospitalized	NaN	NaN	NaN
17836	100.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Recovered	NaN	NaN	NaN
17837	30.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Recovered	NaN	NaN	NaN

201296 rows × 11 columns

In [9]:

```
df.to_csv('covid_data.csv')
```

In [11]:

```
data = df
data
```

Out[11]:

	Num Cases	Date Announced	Age Bracket	Gender	Detected City	Detected District	Detected State	Current Status	Day	Month	Year
0	1.0	20/12/2020	NaN	NaN	NaN	Aizawl	Mizoram	Hospitalized	20	12	2020
1	7.0	20/12/2020	NaN	NaN	NaN	Lunglei	Mizoram	Hospitalized	20	12	2020
2	3.0	20/12/2020	NaN	NaN	NaN	Kolasib	Mizoram	Hospitalized	20	12	2020
3	1.0	20/12/2020	NaN	NaN	NaN	Saitual	Mizoram	Hospitalized	20	12	2020
4	11.0	20/12/2020	NaN	NaN	NaN	Adilabad	Telangana	Hospitalized	20	12	2020
...
17833	1.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Deceased	NaN	NaN	NaN
17834	196.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Hospitalized	NaN	NaN	NaN
17835	39.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Hospitalized	NaN	NaN	NaN
17836	100.0	NaN	NaN	NaN	NaN	Leh	Ladakh	Recovered	NaN	NaN	NaN
17837	30.0	NaN	NaN	NaN	NaN	Kargil	Ladakh	Recovered	NaN	NaN	NaN

201296 rows × 11 columns

In [12]:

```
data.isnull().sum(axis = 0).sort_values(ascending = False)/len(data)*100
```

Out[12]:

Detected City	99.989568
Gender	97.552857
Age Bracket	97.548386
Detected District	1.174390
Date Announced	0.002981
Day	0.002981
Month	0.002981
Year	0.002981
Num Cases	0.000994
Detected State	0.000000
Current Status	0.000000
dtype: float64	

In [13]:

```
data.isnull().sum(axis = 1).sort_values(ascending = False)/len(data)*100
```

Out[13]:

17837	0.003477
-------	----------

```
17836    0.003477
17835    0.003477
17834    0.003477
17833    0.003477
...
6697     0.000000
4809     0.000000
4409     0.000000
16797    0.000000
8173     0.000000
Length: 201296, dtype: float64
```

In [14]:

```
data.groupby('Month')['Num Cases'].sum()
```

Out[14]:

```
Month
01    1031650.0
02     706824.0
03    1802650.0
04    11189531.0
05    17269531.0
08         2.0
12     560406.0
Name: Num Cases, dtype: float64
```

In [15]:

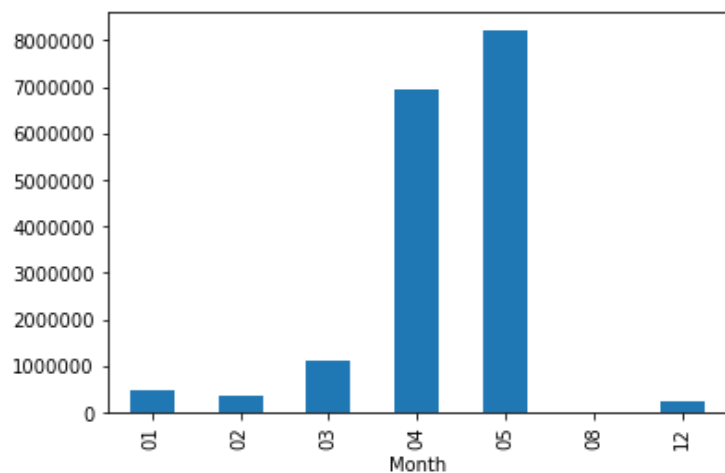
```
M = data[data['Current Status']=='Hospitalized'].groupby('Month')['Num Cases'].sum()
M
```

Out[15]:

```
Month
01    472317.0
02    353426.0
03    1108660.0
04    6936479.0
05    8210739.0
08         2.0
12    254651.0
Name: Num Cases, dtype: float64
```

In [16]:

```
M.plot.bar()
plt.show()
```



In [17]:

```
G = data.groupby('Month')['Num Cases'].sum()
G
```

Out[17]:

```

Month
01      1031650.0
02       706824.0
03      1802650.0
04      11189531.0
05      17269531.0
08           2.0
12       560406.0
Name: Num Cases, dtype: float64

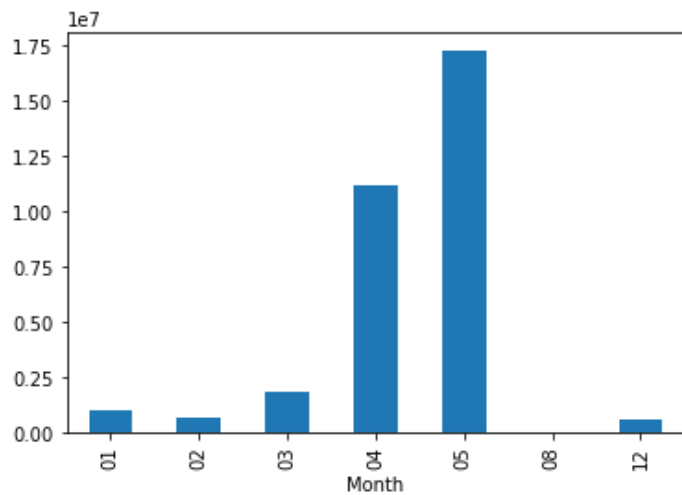
```

In [18]:

```

G.plot.bar()
plt.show()

```



In [19]:

```

G = data.groupby('Age Bracket')['Num Cases'].sum().sort_values(ascending = False)
G

```

Out[19]:

```

Age Bracket
58.0      1451.0
83.0       406.0
65.0       261.0
70.0       229.0
75.0       214.0
...
15.0         1.0
6.0          1.0
5.0          1.0
1.0          1.0
103.0         1.0
Name: Num Cases, Length: 90, dtype: float64

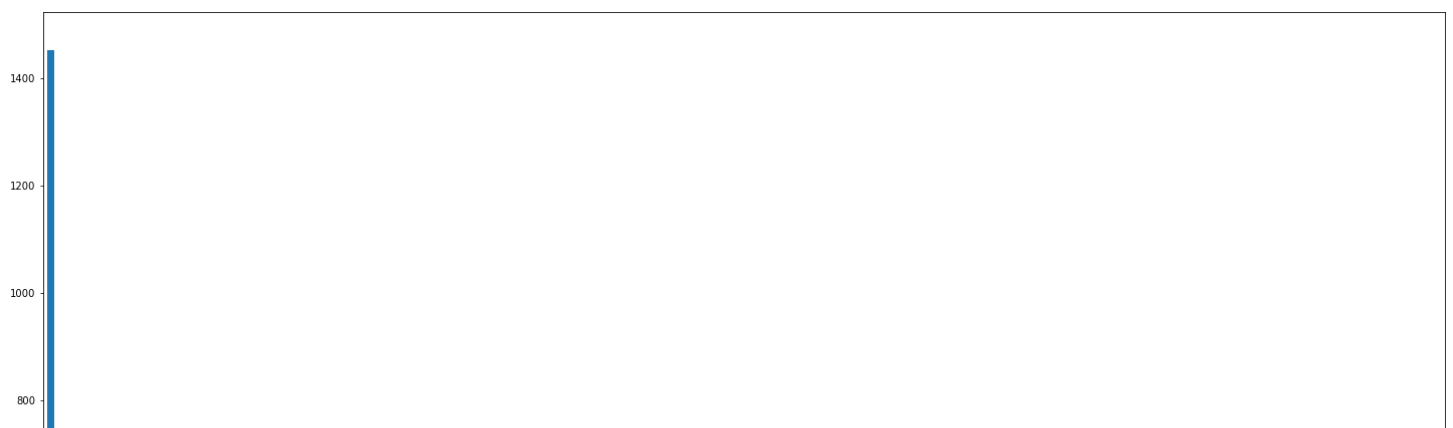
```

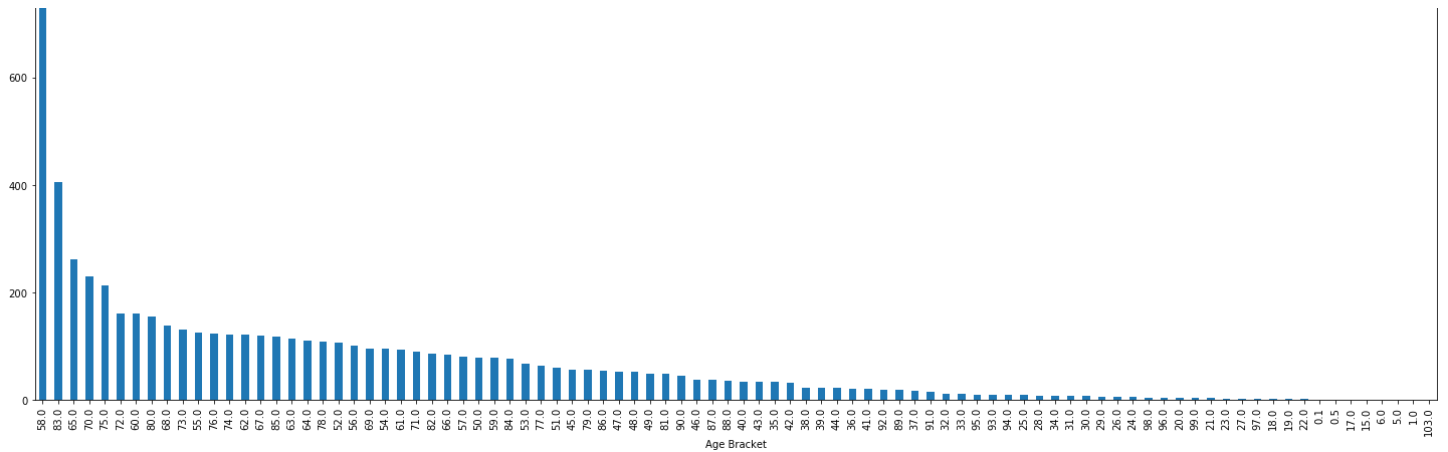
In [20]:

```

G.plot.bar(figsize = (25,15))
plt.show()

```





```
In [21]:
S = data[data['Current Status']=='Hospitalized'].groupby('Detected State')['Num Cases'].sum().sort_values(ascending = False)
S
```

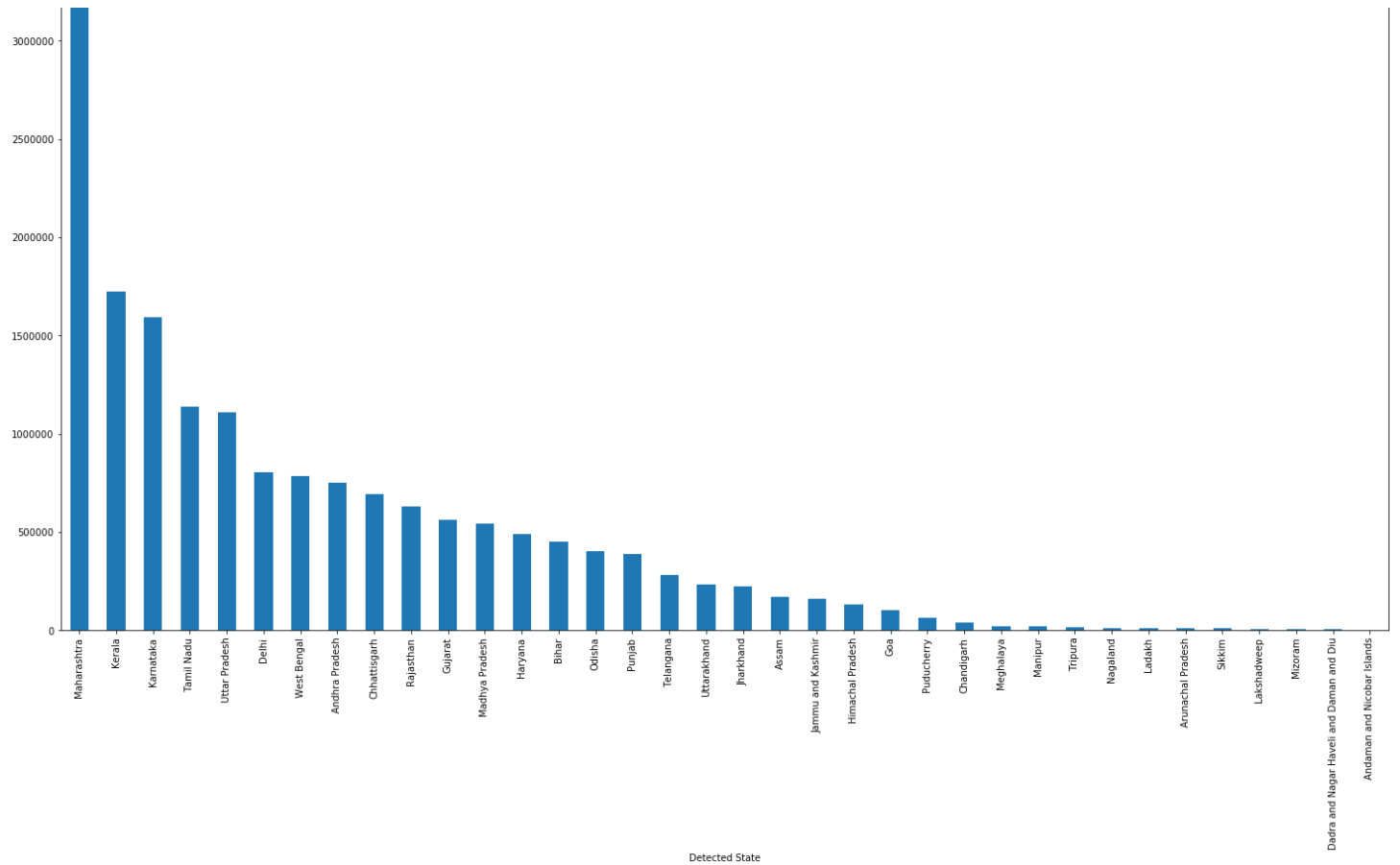
```
Out[21]:
```

Detected State	
Maharashtra	3758200.0
Kerala	1724230.0
Karnataka	1591509.0
Tamil Nadu	1139483.0
Uttar Pradesh	1107283.0
Delhi	805563.0
West Bengal	783353.0
Andhra Pradesh	749105.0
Chhattisgarh	693278.0
Rajasthan	629728.0
Gujarat	563708.0
Madhya Pradesh	541661.0
Haryana	490673.0
Bihar	451684.0
Odisha	400142.0
Punjab	389530.0
Telangana	283081.0
Uttarakhand	235484.0
Jharkhand	221182.0
Assam	171524.0
Jammu and Kashmir	160853.0
Himachal Pradesh	132337.0
Goa	100921.0
Puducherry	61825.0
Chandigarh	40172.0
Meghalaya	19074.0
Manipur	18700.0
Tripura	14820.0
Nagaland	8954.0
Ladakh	8775.0
Arunachal Pradesh	8747.0
Sikkim	8251.0
Lakshadweep	7229.0
Mizoram	6827.0
Dadra and Nagar Haveli and Daman and Diu	6597.0
Andaman and Nicobar Islands	2026.0

Name: Num Cases, dtype: float64

```
In [22]:
S.plot.bar(figsize = (25,15))
plt.show()
```





In [27]:

```
day = data[data['Current Status']=='Hospitalized'].groupby(['Month','Day'])[['Num Cases']]
      .sum()
day
```

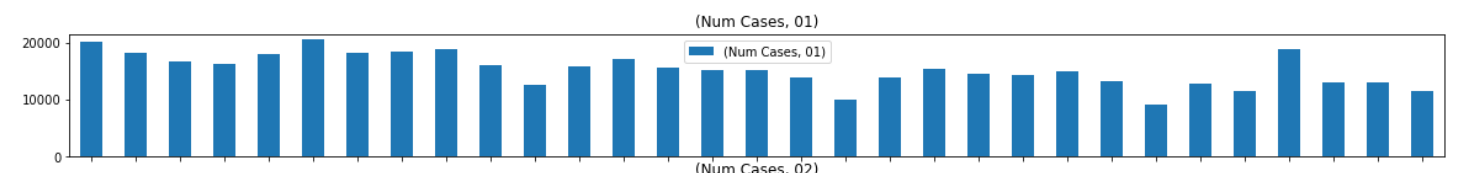
Out[27]:

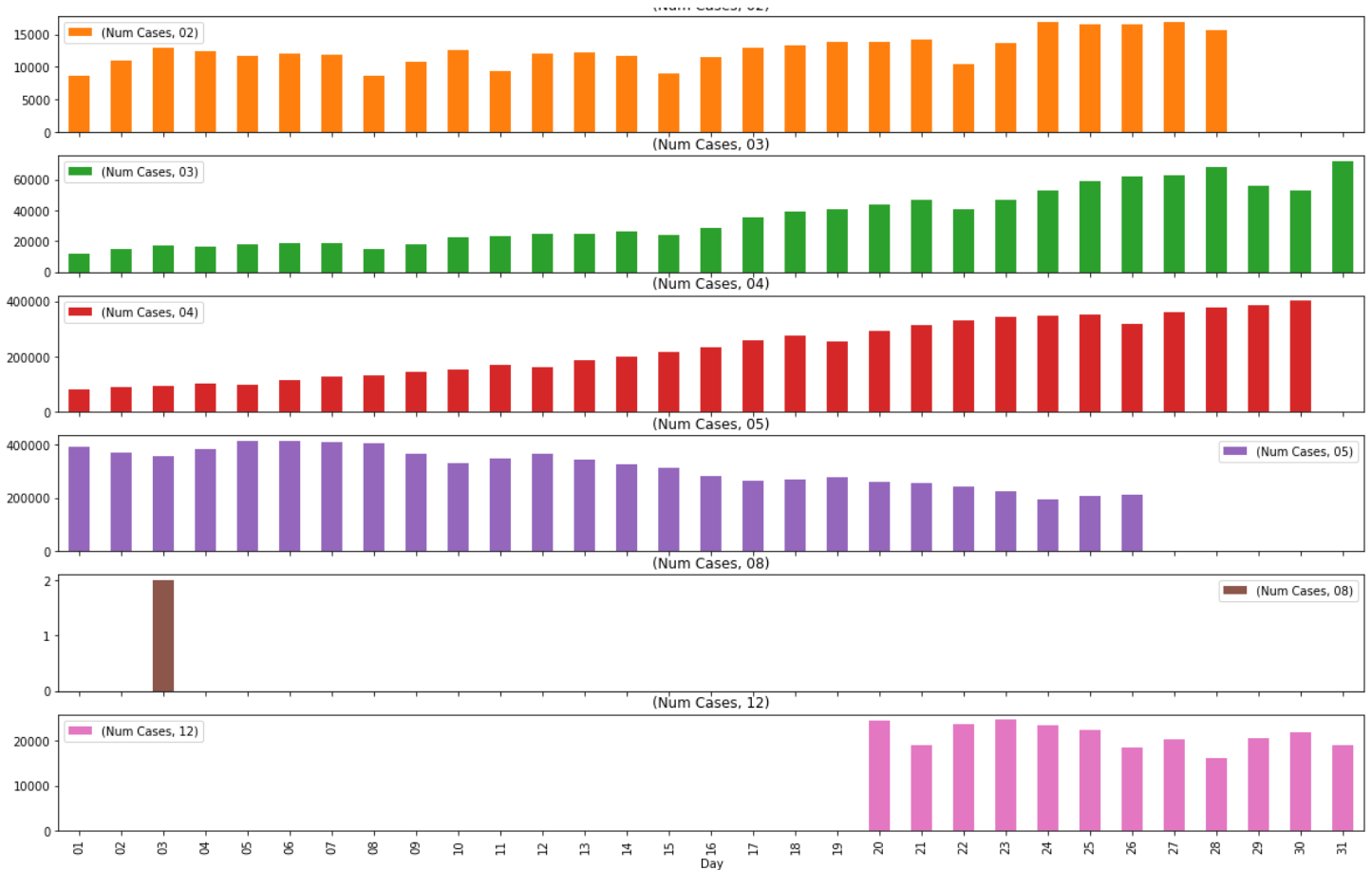
Num Cases		
Month	Day	
01	01	20159.0
	02	18144.0
	03	16678.0
	04	16278.0
	05	17909.0
...
12	27	20333.0
	28	16072.0
	29	20542.0
	30	21945.0
	31	19026.0

159 rows x 1 columns

In [35]:

```
day.unstack(level = 0).plot(kind = 'bar', subplots = True, figsize = (20,15))
plt.show()
```





In [36]:

```
data['Current Status'].unique()
```

Out[36]:

```
array(['Hospitalized', 'Recovered', 'Deceased', 'Migrated_Other',
      'hospitalized', 'Migrated_other'], dtype=object)
```

In [37]:

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

Out[37]:

169752.0

In [40]:

```
death = data[data['Current Status']=='Deceased'].groupby('Detected State')['Num Cases'].sum().sort_values(ascending = False)
death
```

Out[40]:

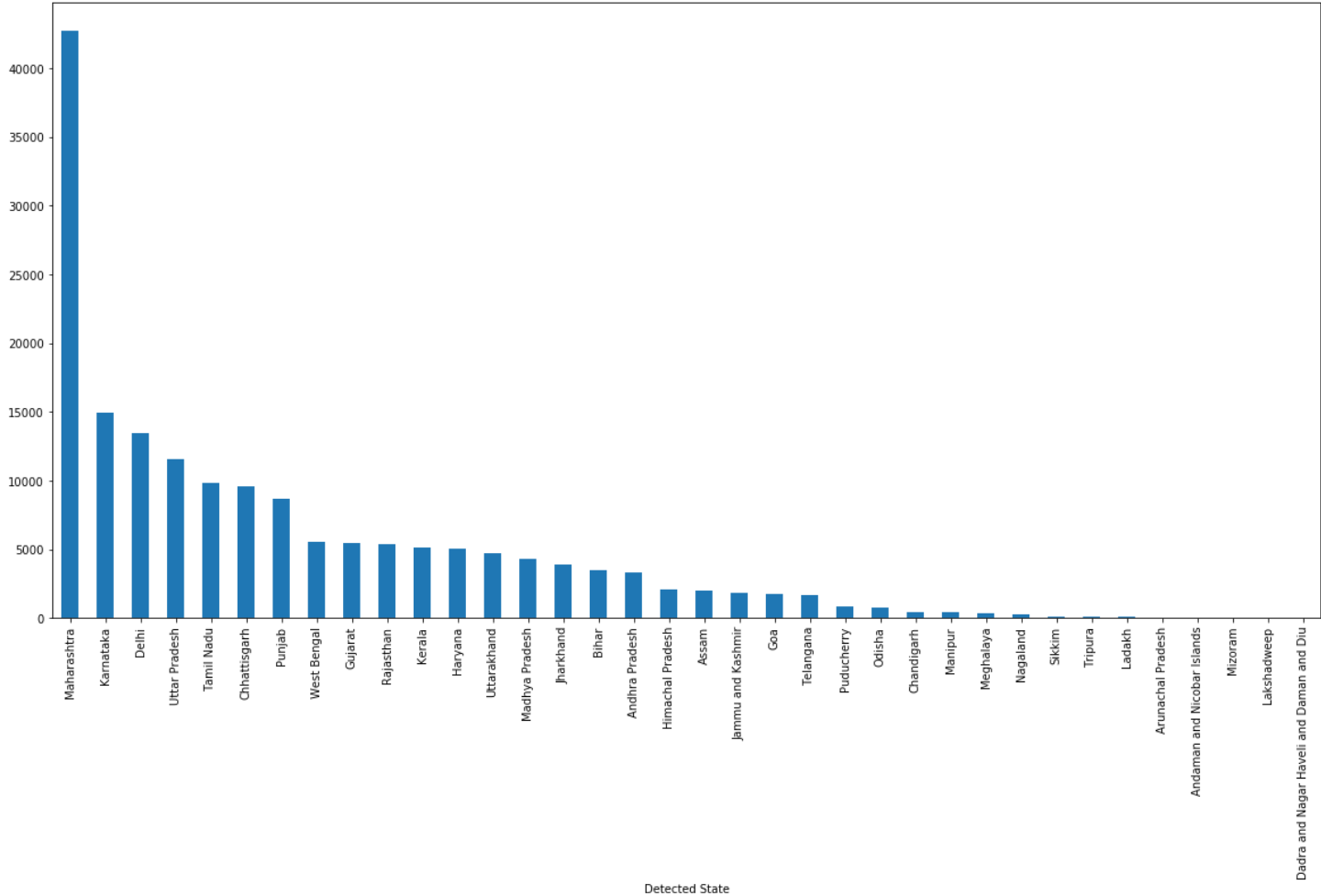
Detected State	
Maharashtra	42693.0
Karnataka	14925.0
Delhi	13444.0
Uttar Pradesh	11535.0
Tamil Nadu	9847.0
Chhattisgarh	9607.0
Punjab	8638.0
West Bengal	5507.0
Gujarat	5474.0
Rajasthan	5410.0
Kerala	5096.0
Haryana	5025.0
Uttarakhand	4705.0
Madhya Pradesh	4290.0
Jharkhand	3900.0
Bihar	3498.0
Andhra Pradesh	3353.0

Himachal Pradesh	2059.0
Assam	1993.0
Jammu and Kashmir	1865.0
Goa	1781.0
Telangana	1679.0
Puducherry	811.0
Odisha	752.0
Chandigarh	416.0
Manipur	394.0
Meghalaya	379.0
Nagaland	254.0
Sikkim	116.0
Tripura	95.0
Ladakh	57.0
Arunachal Pradesh	52.0
Andaman and Nicobar Islands	47.0
Mizoram	27.0
Lakshadweep	26.0
Dadra and Nagar Haveli and Daman and Diu	2.0

Name: Num Cases, dtype: float64

In [44]:

```
death.plot.bar(figsize = (20,10))
plt.show()
```



In []: