In [1]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
import seaborn as sns

import datetime as dt
import calendar


import warnings
warnings.filterwarnings('ignore')
```

In [5]:
```
ds=pd.read_csv("C:/Users/HP/Downloads/Unemployment_Rate_upto_11_2020.csv")
```

In [6]:
```
df=pd.read_csv("C:/Users/HP/Downloads/Unemployment in India.csv")
```

In [7]:
```
ds.head()
```

Out[7]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitu
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.

In [8]:
```
df.head()
```

Out[8]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural

In [9]:
```
ds.columns
```

Out[9]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

In [10]:
```
df.columns
```

Out[10]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

In [13]:
```
df.columns=["state","date","frequency","estimated unemployment rate","estimated emp
```

In [15]:
```
ds.columns=["state","date","frequency","estimated unemployment rate","estimated emp
```

In [16]:
```
ds.head(2)
```

Out[16]:

	state	date	frequency	estimated unemployment rate	estimated employed	estimated labour participation rate	region 1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.74
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.74

In [17]:
```
df.head(2)
```

Out[17]:

	state	date	frequency	estimated unemployment rate	estimated employed	estimated labour participation rate	area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural

In [18]:
```
ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 9 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   state                                 267 non-null    object 
 1   date                                  267 non-null    object 
 2   frequency                             267 non-null    object 
 3   estimated unemployment rate           267 non-null    float64
 4   estimated employed                    267 non-null    int64  
 5   estimated labour participation rate   267 non-null    float64
 6   region 1                              267 non-null    object 
 7   longitude                             267 non-null    float64
 8   latitude                              267 non-null    float64
dtypes: float64(4), int64(1), object(4)
memory usage: 18.9+ KB
```

In [19]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 7 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   state                                 740 non-null    object 
 1   date                                  740 non-null    object 
 2   frequency                             740 non-null    object 
 3   estimated unemployment rate           740 non-null    float64
 4   estimated employed                    740 non-null    float64
 5   estimated labour participation rate   740 non-null    float64
 6   area                                  740 non-null    object 
dtypes: float64(3), object(4)
memory usage: 42.1+ KB
```

In [20]:
```
df.dtypes
```

Out[20]:
```
state                                  object
date                                   object
frequency                              object
estimated unemployment rate           float64
estimated employed                    float64
estimated labour participation rate   float64
area                                   object
dtype: object
```

In [21]:
```
ds.dtypes
```

Out[21]:
```
state                                 object
date                                  object
frequency                             object
estimated unemployment rate          float64
estimated employed                     int64
estimated labour participation rate  float64
region 1                              object
longitude                            float64
latitude                             float64
dtype: object
```

In [22]:
```
ds.shape
```

Out[22]:
```
(267, 9)
```

In [23]:
```
df.shape
```

Out[23]:
```
(768, 7)
```

In [24]:
```
ds.size
```

Out[24]:
```
2403
```

In [25]:
```
df.size
```

Out[25]:
```
5376
```

In [26]:
```
df.corr()
```

Out[26]:

	estimated unemployment rate	estimated employed	estimated labour participation rate
estimated unemployment rate	1.000000	-0.222876	0.002558
estimated employed	-0.222876	1.000000	0.011300
estimated labour participation rate	0.002558	0.011300	1.000000

In [27]:
```
ds.corr()
```

Out[27]:

	estimated unemployment rate	estimated employed	estimated labour participation rate	longitude	latitude
estimated unemployment rate	1.000000	-0.245176	-0.073540	0.149976	-0.023976
estimated employed	-0.245176	1.000000	-0.047948	-0.113664	-0.119321
estimated labour participation rate	-0.073540	-0.047948	1.000000	0.080372	0.397836
longitude	0.149976	-0.113664	0.080372	1.000000	0.125895
latitude	-0.023976	-0.119321	0.397836	0.125895	1.000000

In [28]:
```
ds.isnull().sum()
```

Out[28]:
```
state                                  0
date                                   0
frequency                              0
estimated unemployment rate            0
estimated employed                     0
estimated labour participation rate    0
region 1                               0
longitude                              0
latitude                               0
dtype: int64
```

In [29]:
```
df.isnull().sum()
```

Out[29]:
```
state                                  28
date                                   28
frequency                              28
estimated unemployment rate            28
estimated employed                     28
estimated labour participation rate    28
area                                   28
dtype: int64
```

In [30]:
```
df1=df.dropna()
```

In [31]:
```
df1.isnull().sum()
```

Out[31]:
```
state                                  0
date                                   0
frequency                              0
estimated unemployment rate            0
estimated employed                     0
estimated labour participation rate    0
area                                   0
dtype: int64
```

In [32]:
```
df1.shape
```

Out[32]:
```
(740, 7)
```

In [33]:
```
duplicate_count = ds.duplicated().sum()
print(duplicate_count)
```

```
0
```

In [34]:
```
duplicate_count = df1.duplicated().sum()
print(duplicate_count)
```

```
0
```

In [35]:
```
ds.state.value_counts()
```

Out[35]:
```
Andhra Pradesh      10
Assam               10
Uttarakhand         10
Uttar Pradesh       10
Tripura             10
Telangana           10
Tamil Nadu          10
Rajasthan           10
Punjab              10
Puducherry          10
Odisha              10
Meghalaya           10
Maharashtra         10
Madhya Pradesh      10
Kerala              10
Karnataka           10
Jharkhand           10
Himachal Pradesh    10
Haryana             10
Gujarat             10
Goa                 10
Delhi               10
Chhattisgarh        10
Bihar               10
West Bengal         10
Jammu & Kashmir      9
Sikkim               8
Name: state, dtype: int64
```

In [36]:
```
df1.state.value_counts()
```

Out[36]:
```
Andhra Pradesh      28
Kerala              28
West Bengal         28
Uttar Pradesh       28
Tripura             28
Telangana           28
Tamil Nadu          28
Rajasthan           28
Punjab              28
Odisha              28
Madhya Pradesh      28
Maharashtra         28
Karnataka           28
Jharkhand           28
Himachal Pradesh    28
Haryana             28
Gujarat             28
Delhi               28
Chhattisgarh        28
Bihar               28
Meghalaya           27
Uttarakhand         27
Assam               26
Puducherry          26
Goa                 24
Jammu & Kashmir     21
Sikkim              17
Chandigarh          12
Name: state, dtype: int64
```

In [37]:
```
df1['state'].value_counts().idxmin()
```

Out[37]:
```
'Chandigarh'
```

In [38]:
```
df1['state'].value_counts().idxmax()
```

Out[38]:
```
'Andhra Pradesh'
```

In [39]:
```
ds['state'].value_counts().idxmax()
```

Out[39]:
```
'Andhra Pradesh'
```

In [40]:
```
ds['state'].value_counts().idxmin()
```

Out[40]:
```
'Sikkim'
```

In [41]:
```
import datetime as dt
import calendar
```

In [42]:
```
ds['date'] = pd.to_datetime(ds['date'], dayfirst=True)    # here, date column is de
ds['month_int'] =  ds['date'].dt.month                    # here,month component is
ds['month'] =  ds['month_int'].apply(lambda x: calendar.month_abbr[x])      #here,
```

In [43]:
```
ds.head(2)
```

Out[43]:

	state	date	frequency	estimated unemployment rate	estimated employed	estimated labour participation rate	region 1	longitude	latitude
0	Andhra Pradesh	2020-01-31	M	5.48	16635535	41.02	South	15.9129	79.74
1	Andhra Pradesh	2020-02-29	M	5.83	16545652	40.90	South	15.9129	79.74

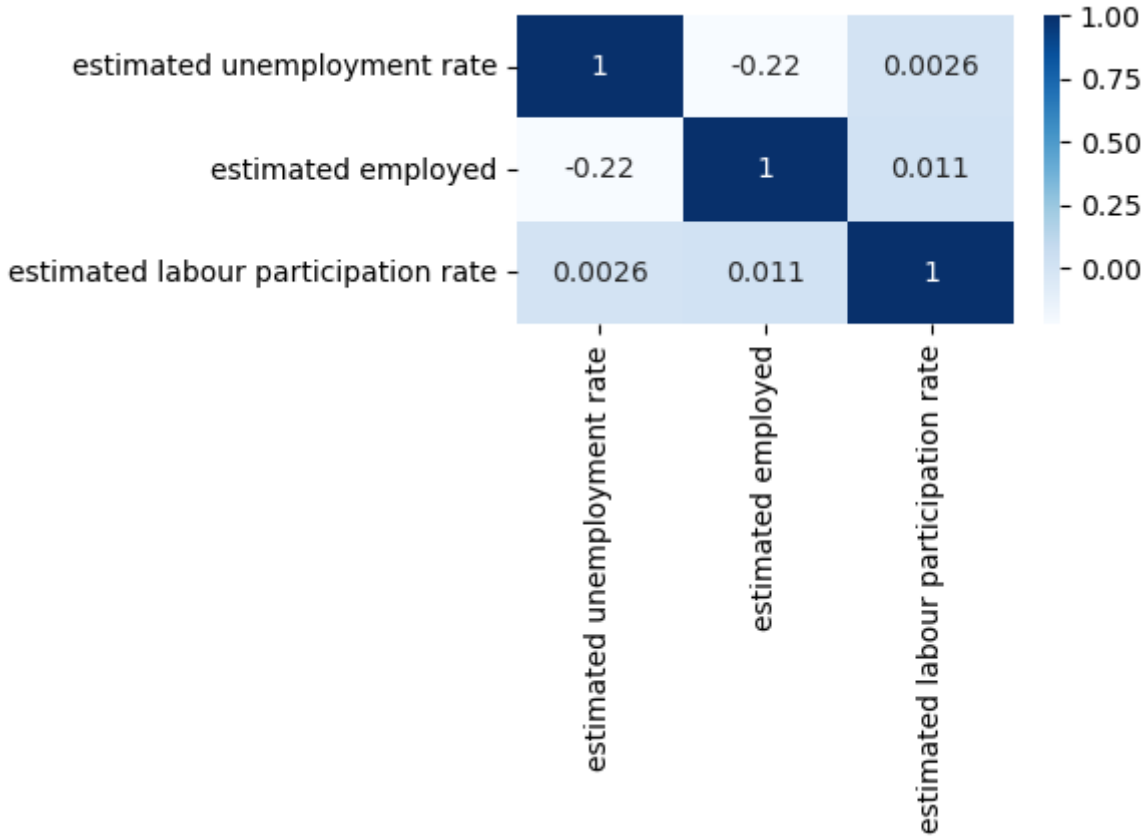
In [44]:
```
ds['month'].value_counts().idxmax()
```

Out[44]:
```
'Mar'
```

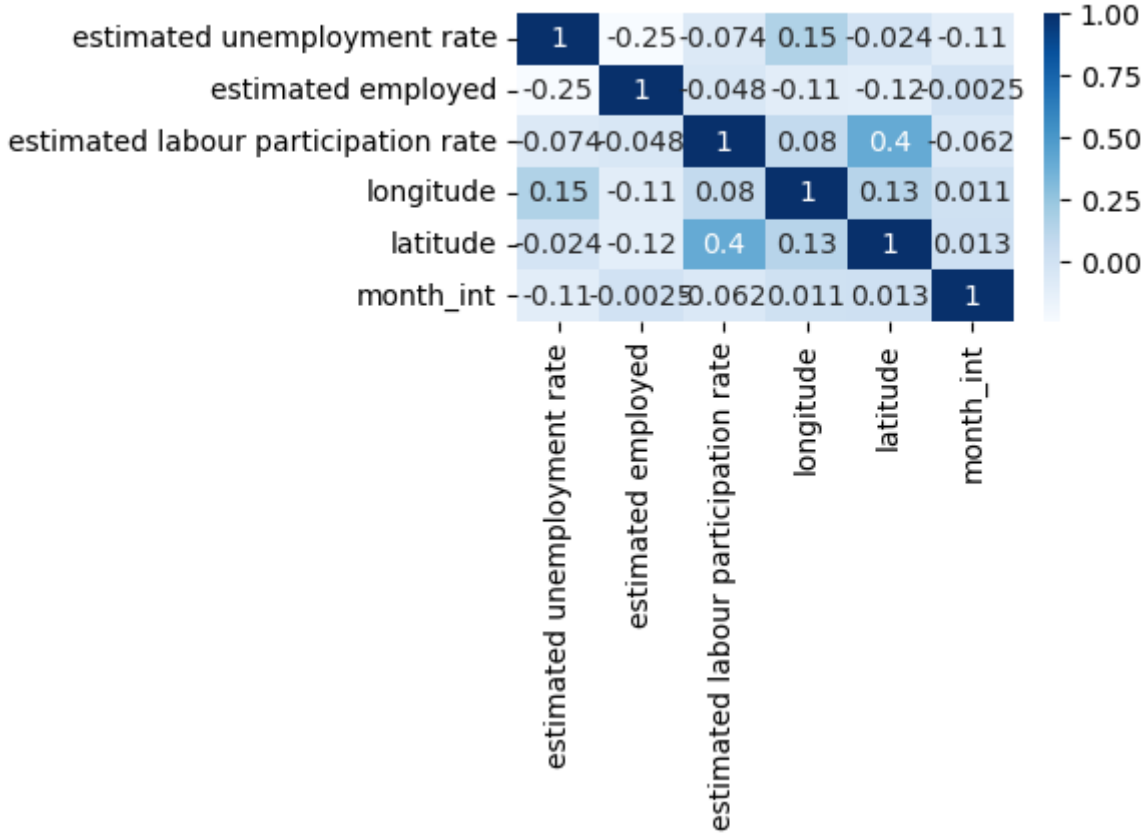
In [45]:
```
ds['month'].value_counts().idxmin()
```

Out[45]:
```
'Jan'
```

In [46]:
```
plt.figure(figsize=(4,2))
sns.heatmap(df1.corr(),annot=True,cmap='Blues')
plt.show()
```



In [47]:
```
plt.figure(figsize=(4,2))
sns.heatmap(ds.corr(),annot=True,cmap='Blues')
plt.show()
```



In [50]:
```
ds1=ds[['state','estimated unemployment rate']].groupby('state').sum().sort_values(
```

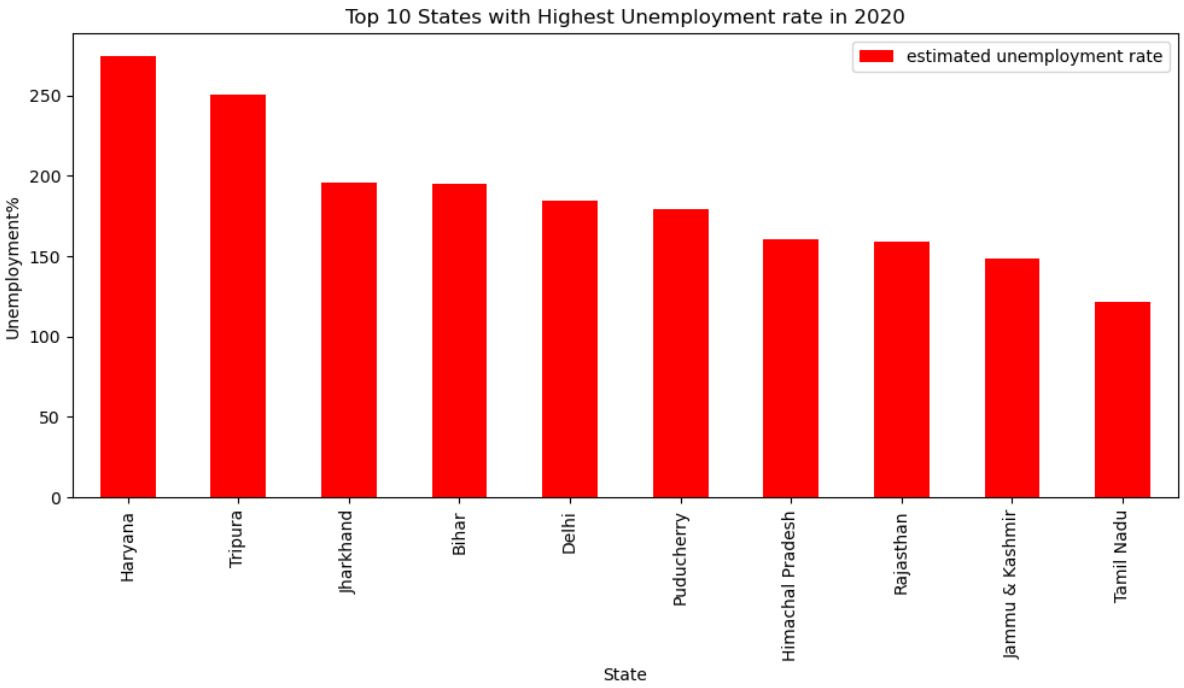
In [51]:
```
ds1.head(10) #2020 data
```

Out[51]:

state	estimated unemployment rate
Haryana	274.77
Tripura	250.55
Jharkhand	195.39
Bihar	194.71
Delhi	184.14
Puducherry	179.42
Himachal Pradesh	160.65
Rajasthan	158.68
Jammu & Kashmir	148.30
Tamil Nadu	121.87

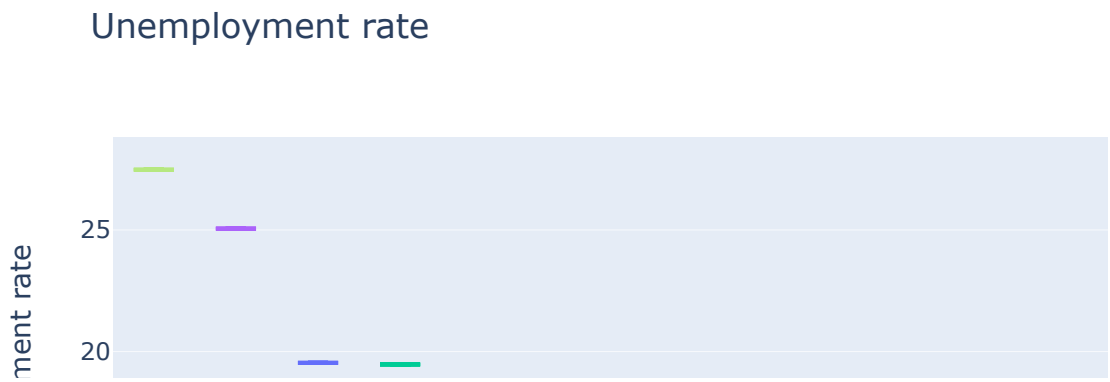
In [52]:
```
fig = plt.figure()
ax0 = fig.add_subplot(1,2,1)
ds1[:10].plot(kind= 'bar', color='red', figsize=(26,5),ax=ax0)
ax0.set_title("Top 10 States with Highest Unemployment rate in 2020")
ax0.set_xlabel('State')
ax0.set_ylabel('Unemployment%')
```

Out[52]:
```
Text(0, 0.5, 'Unemployment%')
```



In [53]:
```
ds2 = ds.groupby(["state"])[["estimated unemployment rate", "estimated employed", '
ds2 = pd.DataFrame(ds2).reset_index()
```

In [54]:
```
fig = px.box(ds2,x='state',y='estimated unemployment rate',color='state',title='Une
fig.update_layout(xaxis={'categoryorder':'total descending'})
fig.show()
```



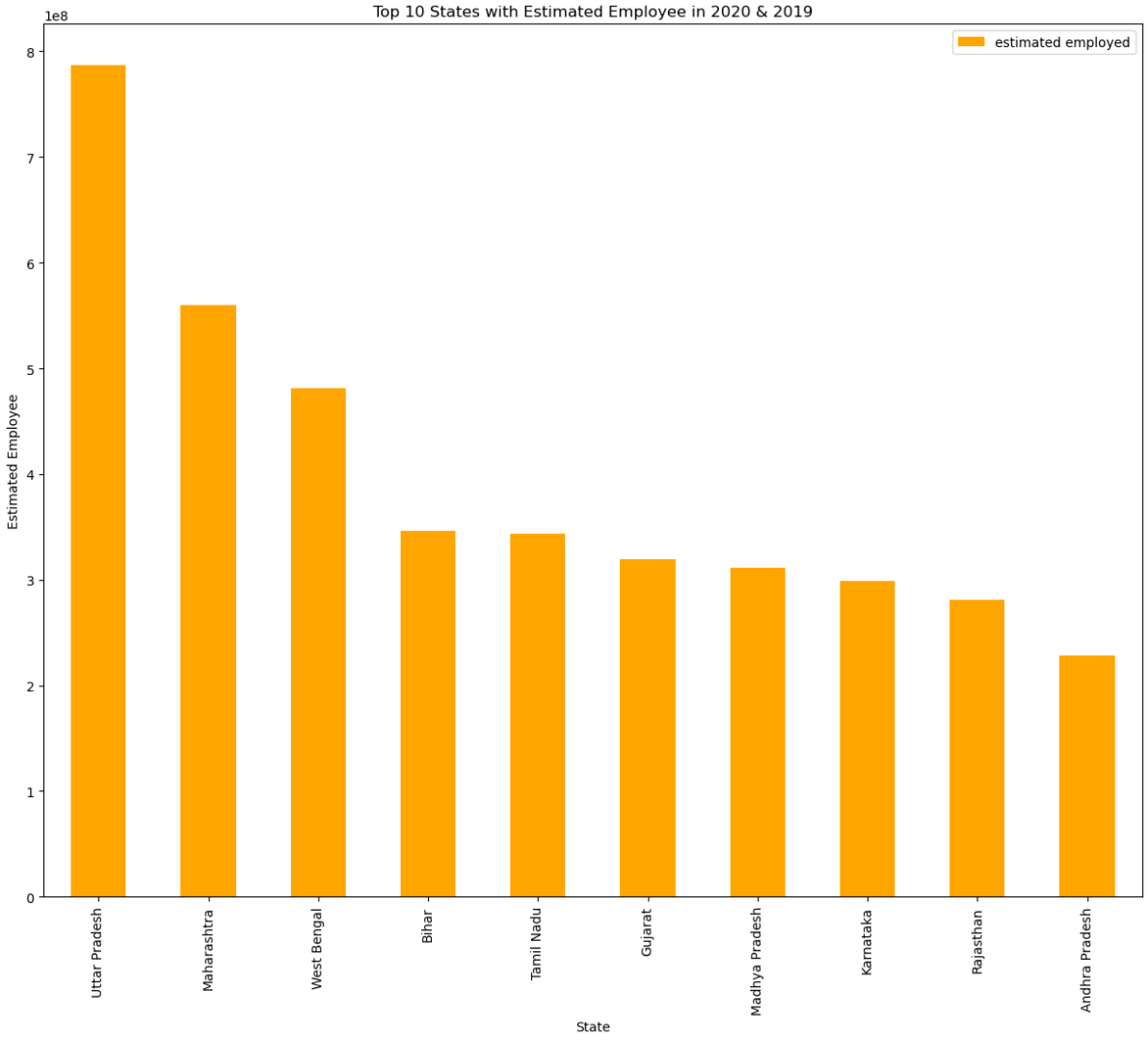
In [55]:
```
df1 = df[["state","estimated employed"]].groupby('state').sum().sort_values("estima
df1
```

Out[55]:

state	estimated employed
Uttar Pradesh	786655301.0
Maharashtra	559725484.0
West Bengal	481559064.0
Bihar	346253296.0
Tamil Nadu	343547309.0
Gujarat	319256358.0
Madhya Pradesh	311233561.0
Karnataka	298679340.0
Rajasthan	281149813.0
Andhra Pradesh	228314609.0
Telangana	222310557.0
Odisha	183280915.0
Assam	139224076.0
Punjab	127102136.0
Jharkhand	125138732.0
Kerala	123925186.0
Chhattisgarh	120497960.0
Haryana	99598029.0
Delhi	73570360.0
Jammu & Kashmir	37798565.0
Uttarakhand	37536159.0
Himachal Pradesh	29675064.0
Tripura	20076074.0
Meghalaya	18622894.0
Puducherry	5519230.0
Goa	5431400.0
Chandigarh	3801975.0
Sikkim	1816972.0

In [56]:
```
df1[:10].plot(kind='bar',color='orange', figsize=(15,12))
plt.title("Top 10 States with Estimated Employee in 2020 & 2019")
plt.xlabel('State')
plt.ylabel('Estimated Employee')
```

Out[56]:
```
Text(0, 0.5, 'Estimated Employee')
```



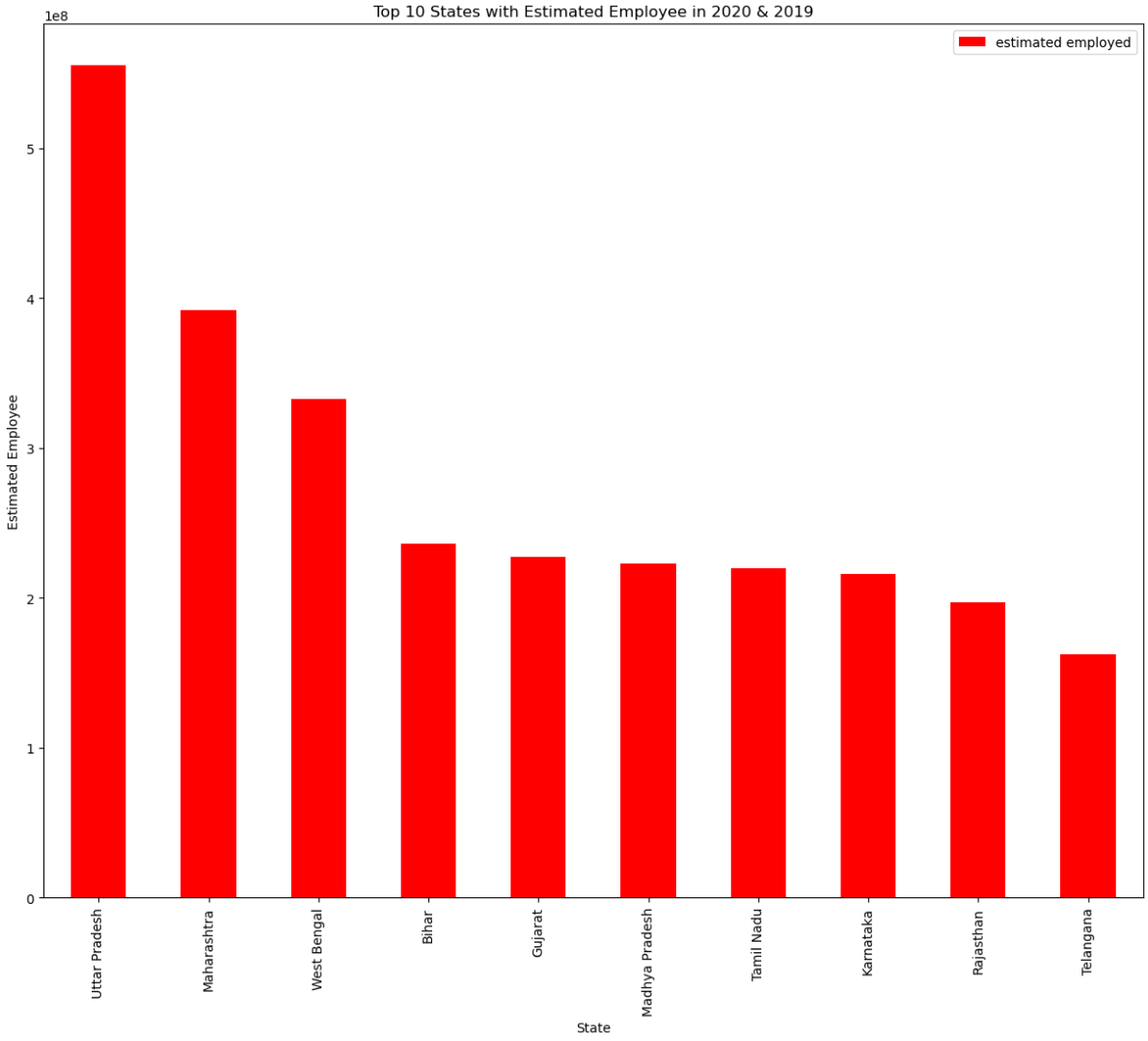
In [57]:
```
ds3 = ds[["state","estimated employed"]].groupby('state').sum().sort_values("estima
ds3
```

Out[57]:

state	estimated employed
Uttar Pradesh	555247990
Maharashtra	392047582
West Bengal	333051643
Bihar	236068280
Gujarat	227307461
Madhya Pradesh	223183353
Tamil Nadu	219878981
Karnataka	216240176
Rajasthan	197317522
Telangana	162440825
Andhra Pradesh	154254800
Odisha	127268329
Assam	108102755
Punjab	87830342
Jharkhand	87706424
Kerala	85967949
Chhattisgarh	84213492
Haryana	68440590
Delhi	46328219
Jammu & Kashmir	29790285
Uttarakhand	27432749
Himachal Pradesh	20338849
Tripura	13972916
Meghalaya	13498153
Goa	4423748
Puducherry	3652629
Sikkim	1876186

In [58]:
```
ds3[:10].plot(kind='bar',color='red', figsize=(15,12))
plt.title("Top 10 States with Estimated Employee in 2020 & 2019")
plt.xlabel('State')
plt.ylabel('Estimated Employee')
```

Out[58]:
```
Text(0, 0.5, 'Estimated Employee')
```

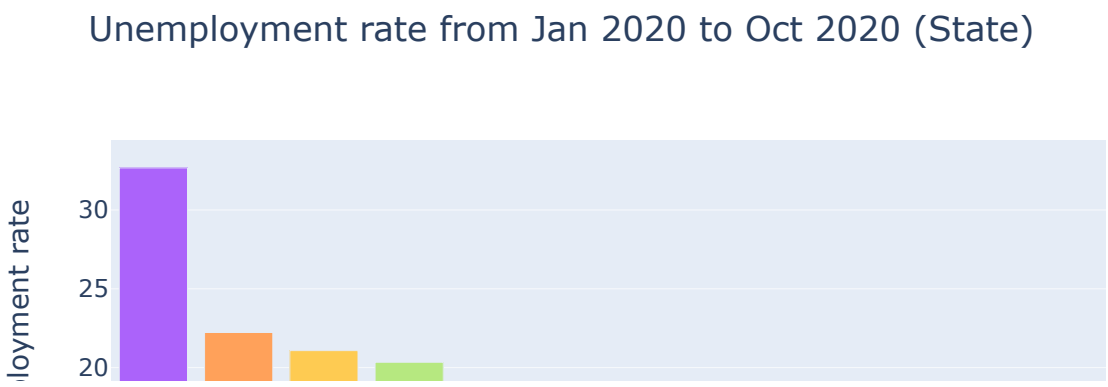


In [59]:
```
fig = px.bar(ds, x='state',y='estimated unemployment rate', animation_frame = 'mont
             title='Unemployment rate from Jan 2020 to Oct 2020 (State)')

fig.update_layout(xaxis={'categoryorder':'total descending'})

fig.layout.updatemenus[0].buttons[0].args[1]["frame"]["duration"]=2000

fig.show()
```


Unemployment rate from Jan 2020 to Oct 2020 (State)
loyment rate
30
25
20