In [11]:
```
#importing required liberies
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline
```

In [12]:
```
#uploading the data
data_set = pd.read_csv('unemployment.csv')
```

In [13]:
```
data_set.head()
```

Out[13]:

	Region	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
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.74
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.74
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.74

In [14]:
```
data_set.tail()#Return the last 5 rows of the DataFrame.
```

Out[14]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

In [15]:
```
data_set.shape #return the number of rows and columns avalibe in dataframe.
```

Out[15]: (267, 9)

In [16]:
```
data_set.describe()
```

Out[16]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	longitude	latitude
count	267.000000	2.670000e+02	267.000000	267.000000	267.000000
mean	12.236929	1.396211e+07	41.681573	22.826048	80.532425
std	10.803283	1.336632e+07	7.845419	6.270731	5.831738
min	0.500000	1.175420e+05	16.770000	10.850500	71.192400
25%	4.845000	2.838930e+06	37.265000	18.112400	76.085600
50%	9.650000	9.732417e+06	40.390000	23.610200	79.019300
75%	16.755000	2.187869e+07	44.055000	27.278400	85.279900
max	75.850000	5.943376e+07	69.690000	33.778200	92.937600

In [17]:
```
data_set.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 9 columns):
 #   Column                                   Non-Null Count  Dtype  
---  ------                                   --------------  -----  
 0   Region                                   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
```

Data cleaning and Modifying

In [18]:
```
data_set.isnull().sum() #finding the missing value
```

Out[18]:
```
Region                                   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 [19]:
```
data_set.isnull().sum().sum() #Total missing values
```

Out[19]:
```
0
```

In [20]:
```
data_set1 = data_set.fillna(method='pad') #filling the null values with the previous value
data_set
```

Out[20]:

	Region	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	79.740
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740
...	...	...	...	...	...	...	...	...	...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

267 rows × 9 columns

In [21]:
```
data_set.isnull().sum()
```

Out[21]:
```
Region                                   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
```

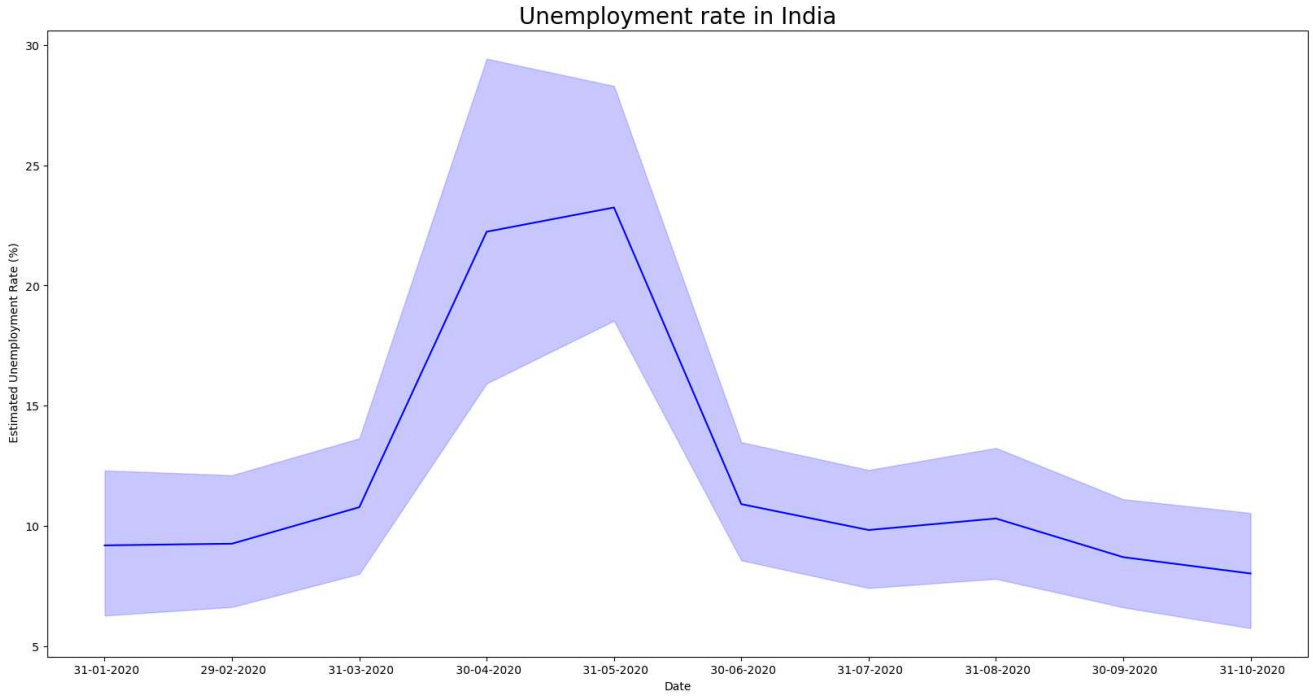
In [22]:
```
print(data_set.columns.tolist())
```

```
['Region', 'Date', ' Frequency', ' Estimated Unemployment Rate (%)', ' Estimated Employed', ' Estimated Labour Participation Ra
te (%)', 'Region.1', 'longitude', 'latitude']
```

Data Analysis Part

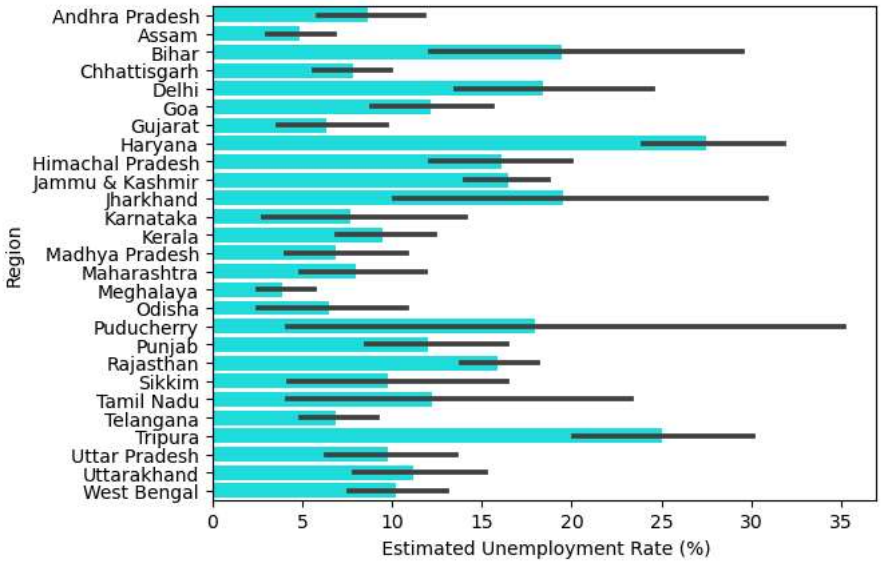
In [23]:
```
plt.figure(figsize=(20,10)) #Unemployment rate in india
sns.lineplot(data=data_set,x='Date',y=" Estimated Unemployment Rate (%)",color='blue')
plt.title("Unemployment rate in India",fontsize="20")
```

Out[23]:
```
Text(0.5, 1.0, 'Unemployment rate in India')
```



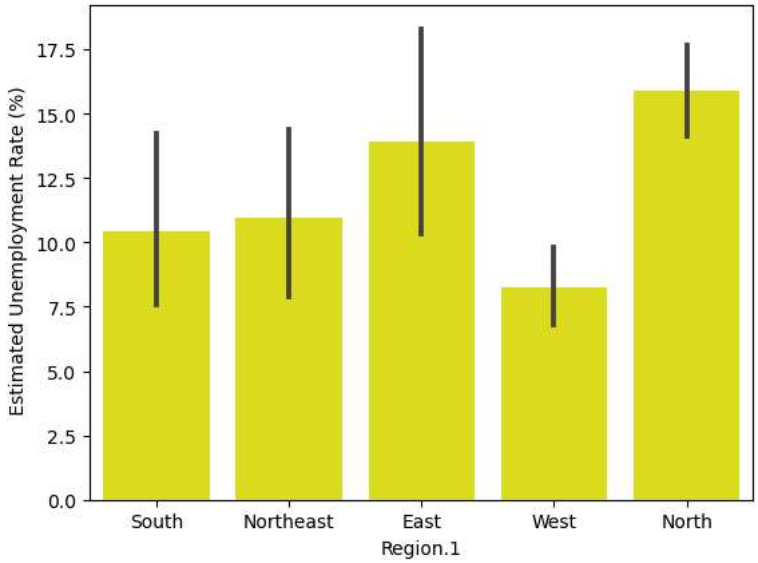
In [28]:
```
sns.barplot(data=data_set,x=" Estimated Unemployment Rate (%)",y='Region',color='cyan')
```

Out[28]:
```
<AxesSubplot:xlabel=' Estimated Unemployment Rate (%)', ylabel='Region'>
```

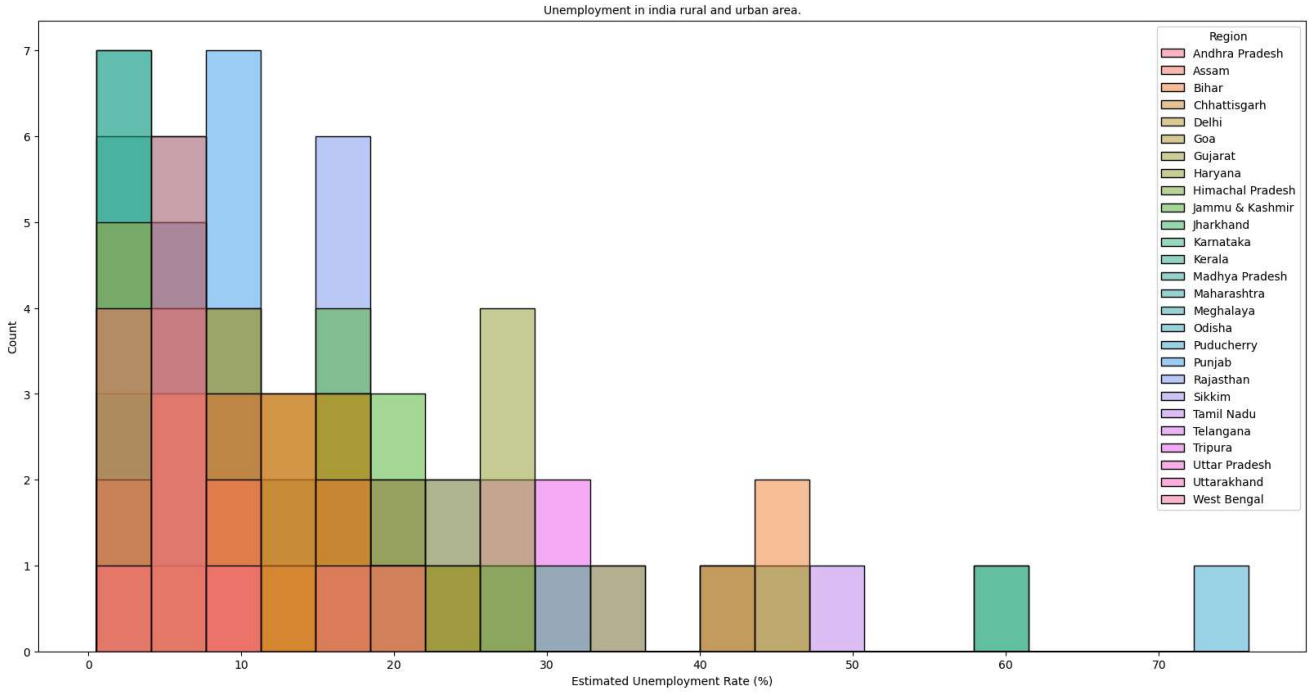


In [30]:
```
sns.barplot(data=data_set,x='Region.1',y=" Estimated Unemployment Rate (%)",color='yellow')
```

Out[30]:
```
<AxesSubplot:xlabel='Region.1', ylabel=' Estimated Unemployment Rate (%)'>
```



In [32]:
```
te','Frequency','Estimated Unemployment Rate (%)','Estimated Employment', 'Estimated Employed','Region.1 ','longitude','latitude'
ia rural and urban area.",fontsize='10')
ployment Rate (%)',hue='Region',data=data_set)
```



In [34]:
```
sns.heatmap(data_set.corr(),annot=True)
```

Out[34]: <AxesSubplot:>

In [ ]:
```
##########################################################################################################################
```