In []:	1 7	#########iris data set########						
In [1]:	2 3 4 4	<pre>#importing libraries import pandas as pd import numpy as np import matplotlib.pyplot as plt %matplotlib inline</pre>						
In [2]:		<pre>import warnings warnings.filterwarnings('ignore')</pre>						
In [6]:	1	<pre>df=pd.read_csv("C:/Users/KARNARAJ RATHOD/OneDrive/Desktop/oasis/Iris.csv")</pre>						
In [8]:	1 2 3	df.head()						
	Id	SepalLeng	thCm SepalWidth	nCm PetalLengt	hCm PetalWidth	Cm Species		
	0 1	5.1	3.5	1.4	0.2	Iris-setosa		
	1 2	4.9	3.0	1.4	0.2	Iris-setosa		
	2 3	4.7	3.2	1.3	0.2	Iris-setosa		
	3 4	4.6	3.1	1.5	0.2	Iris-setosa		
	4 5	5.0	3.6	1.4	0.2	Iris-setosa		
In [9]:	1	df.shape						
In [10]:		6) df.describ Id		SepalWidthCm	PetalLengthCm	PetalWidthCm		
	count	150.000000	150.000000	150.000000	150.000000	150.000000		
	mean	75.500000	5.843333	3.054000	3.758667	1.198667		
	std	43.445368	0.828066	0.433594	1.764420	0.763161		
	min	1.000000	4.300000	2.000000	1.000000	0.100000		
	25%	38.250000	5.100000	2.800000	1.600000	0.300000		
	50%	75.500000	5.800000	3.000000	4.350000	1.300000		
	75%	112.750000	6.400000	3.300000	5.100000	1.800000		
	max	150.000000	7.900000	4.400000	6.900000	2.500000		
In [11]:	1 (df.info						

```
In [10]
                        Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
          count 150.000000 150.000000
                                             150.000000
                                                             150.000000
                                                                             150.000000
                                             3.054000
                                                            3.758667
                                                                             1.198667
          mean
                75.500000
                           5.843333
                                             0.433594
                                                             1.764420
                 43.445368
                            0.828066
                                                                             0.763161
          min
                1.000000
                            4.300000
                                             2.000000
                                                            1.000000
                                                                             0.100000
          25%
                38.250000
                           5.100000
                                             2.800000
                                                            1.600000
                                                                             0.300000
                75.500000
                           5.800000
                                             3.000000
                                                            4.350000
                                                                             1.300000
          50%
          75%
                112.750000 6.400000
                                             3.300000
                                                            5.100000
                                                                             1.800000
                                             4.400000
                                                            6.900000
                                                                             2.500000
          max
                150.000000 7.900000
In [11]:
               df.info
           <bound method DataFrame.info of</pre>
                                             Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm \
                             5.1
                                                       1.4
                                                                    0.2
                             4.9
                                         3.0
                                                       1.4
                                                                    0.2
                             4.7
                                                       1.3
                                                                    0.2
                                         3.2
                                         3.1
                                                       1.5
                                                                    0.2
                             4.6
                                                                    0.2
                             5.0
                                         3.6
                                                       1.4
           145 146
                             6.7
                                         3.0
                                                       5.2
                                                                    2.3
           146 147
                             6.3
                                         2.5
                                                       5.0
                                                                    1.9
           147 148
                             6.5
                                         3.0
                                                       5.2
                                                                    2.0
                                         3.4
                                                      5.4
                                                                    2.3
           148 149
                             6.2
           149 150
                             5.9
                                                                    1.8
                                                       5.1
                      Species
                  Iris-setosa
                  Iris-setosa
                  Iris-setosa
                  Iris-setosa
                  Iris-setosa
In [13]:
               dfispergioscavalue_counts
               Iris-virginica
           147 Iris-virginica
           <u>148</u>undrmetիզբերվeyOpsMixin.value_counts of 0
                                                           Iris-setosa
           149 Iris-Vî⊨gîA⊈t@sa
                    Iris-setosa
           150 rows Iriscottmas]>
In [12]:
               df!ri@n\irginicgum()
                 Iris-virginica
                 Iris-virginica
           147
                 Iris-virgi@ica
           SapalLemgtbCmirgi0ica
           NemelWipthims, Le@gth: 150, dtype: object>
           PetalLengthCm 0
In [14]:
           1 cits df[['Id','SepalLengthCm','SepalWidthCm','PetalLengthCm','PetalWidthCm']]
           2 y= int64 | Species']
```

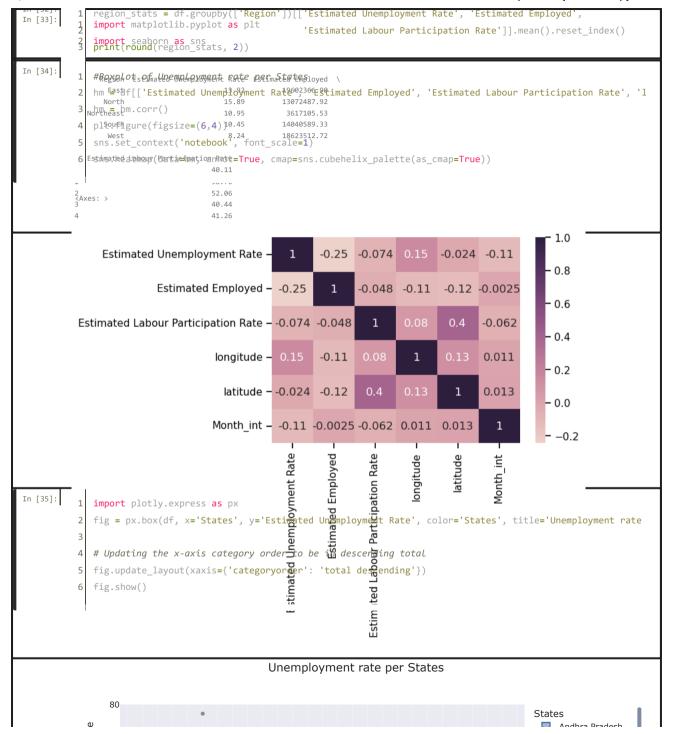
```
In [13]:
              dfiSpergiescavalue counts
               Iris-virginica
          147 Iris-virginica
           immermethode index OpsMixin.value_counts of 0
                                                          Iris-setosa
           149 Iris-Viègingtesa
                   Iris-setosa
           150 rows Iriscotumns]>
                   Iris-setosa
In [12]:
              df!risnvirg(nicgum()
                 Iris-virginica
                Iris-virginica
                Iris-virgi@ica
           S4palLemgtbCmirgi0ica
           NemelWidthims, LeOgth: 150, dtype: object>
           PetalLengthCm
In [14]:
           1 ciss df[['Idd','SepalLengthCm','SepalWidthCm','PetalLengthCm','PetalWidthCm']]
           2 /pe: int64 | Species | ]
In [15]:
                Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                  5.1
             2
                  4.9
                                   3.0
                                                   1.4
                                                                   0.2
                                   3.2
                                                   1.3
                                                                   0.2
                   4.7
                   4.6
                                   3.1
                                                   1.5
                                                                   0.2
              5
                   5.0
                                   3.6
                                                   1.4
                                                                   0.2
          145 146 6.7
                                   3.0
                                                   5.2
                                                                   2.3
          146 147 6.3
                                   2.5
                                                   5.0
                                                                   1.9
          147 148 6.5
                                   3.0
                                                   5.2
                                                                   2.0
          148 149 6.2
                                                   5.4
                                                                   2.3
                                   3.4
          149 150 5.9
In [17]:
              # Dos themtrain/test split
               from sklearn.model_selection import train_test_split
In [16]:
               X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20,random_state=42)
                   Iris-setosa
In [18]:
              # Trainsingtothe Linear Regression Model
                   Iris-setosa
                   Iris-setosa
               from IFksleatogalinear_model import LogisticRegression
In [19]:
              # Lest-Singneate an instance for the LogisticRegression model
              lr Iris virginica
L CogisticRegression()
Iris-virginica
                 Iris-virginica
               #Sprath the thodel, dhypour birdin dataset
```

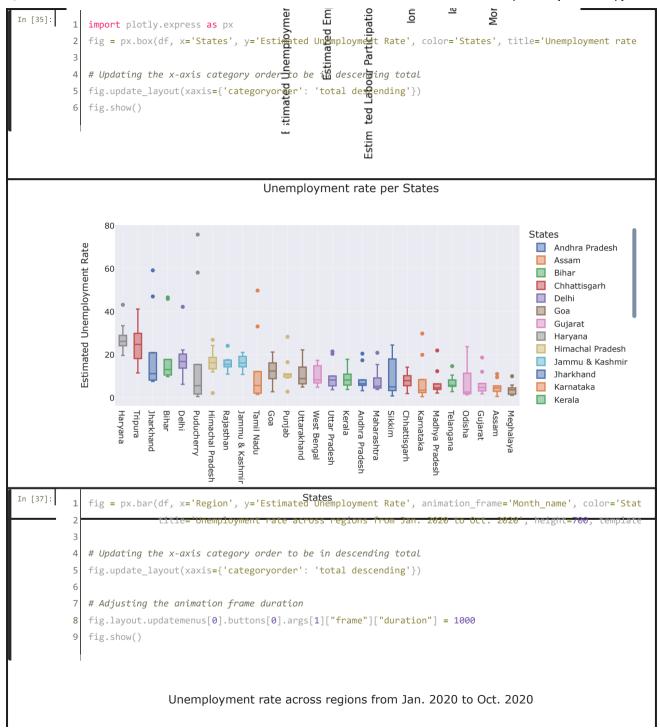
```
In [17]:
              # Dos themtrain/test split
              from sklearn.model_selection import train_test_split
 In [16]:
              X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20,random_state=42)
 In [18]:
              # Trainsingtothe Linear Regression Model
                  Iris-setosa
                  Iris-setosa
              from IFksleatosalinear_model import LogisticRegression
 In [19]:
              # Lest-Singheate an instance for the LogisticRegression model
             lr = LogisticRegression()
Iris-virginica
               Iris-virginica
             #SASSATR the thiodel, dhypour birdin dataset
              lr.fit(X,y)
              # Train the model with the training set
           9
              lr.fit(X_train,y_train)
             LogisticRegression
          LogisticRegressi
 In [20]:
              # Getting predictions from the model for the given examples.
              predictions = lr.predict(X)
              # Compare with the actual charges
 In [22]:
              from sklearn.metrics import accuracy_score
             print(accodatetions)
              Scores.head()
          100. Actual Predictions
 In [23]:
             In [24]:
             istentosat leisisetosaas pd
             isisetosa Irisisetosa np
 In [21]:
              y_test_hat=lr.predict(X_test)
In [26]:
             df = pd.read_csv("C:/Users/KARNARAJ RATHOD/OneDrive/Desktop/oasis/Unemployment_Rate_upto_11_2020.csv")
```

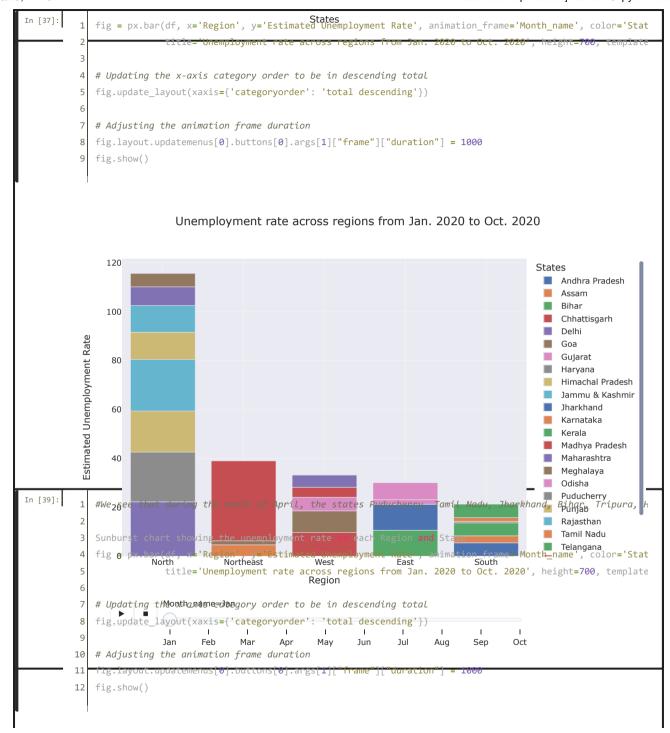
```
In [22]:
               from sklearn.metrics import accuracy_score
               FrintsaccddaPataEbame(\{testualtext_Rsed*fbjons;}predictions})
               Scores.head()
           100. Actual Predictions
In [23]:
               s##N%&#######################unemployment rate in india##############
In [24]:
              isisetosat Iris-setosa as pd
              isisetosa Irisisetosa np
In [21]:
               y test hat=lr.predict(X test)
In [26]:
               df = pd.read_csv("C:/Users/KARNARAJ RATHOD/OneDrive/Desktop/oasis/Unemployment_Rate_upto_11_2020.csv")
               df.head()
                                                       Estimated
                                                                      Estimated
                                                                                       Estimated Labour
                                                                                                         Region.1 longitude latitude
                         Date Frequency
                                             Unemployment Rate
               Region
                                                                      Employed Participation Rate (%)
                                                              (%)
             Andhra
                        31-01-
                               M
                                           5.48
                                                                   16635535
                                                                                  41.02
                                                                                                         South
                                                                                                                    15.9129
                                                                                                                              79.74
             Pradesh
                        2020
             Andhra
                        29-02-
                                           5.83
                                                                   16545652
                                                                                  40.90
                                                                                                         South
                                                                                                                    15.9129
                                                                                                                              79.74
             Pradesh
                       2020
                       31-03-
             Andhra
                                           5.79
                                                                   15881197
                                                                                  39.18
                                                                                                         South
                                                                                                                    15.9129
                                                                                                                              79.74
             Pradesh
                        2020
             Andhra
                        30-04-
                                           20.51
                                                                   11336911
                                                                                                                    15.9129
                                                                                                                              79.74
                                                                                  33.10
                                                                                                         South
             Pradesh
                       2020
                       31-05-
             Andhra
                               M
                                           17.43
                                                                   12988845
                                                                                  36.46
                                                                                                         South
                                                                                                                    15.9129
                                                                                                                              79.74
             Pradesh
                       2020
In [27]:
               df.info()
In [28]:
              df.isnull().sum()
             ass 'pandas.core.frame.DataFrame'>
           RangeIndex: 267 entries, 0 to 266
           Ragionolumns (total 9 columns):
            #ateColumn
                                                  0 Non-Null Count Dtype
           -Frequency-
                                                  0 267 non-null object
            @stiRegednUnemployment Rate (%)
                                                  0 267 non-null
            1stim2ated Employed
                                                                   object
            ≩รtim็ลเซียดีนยลิติษันา Participation Rate (%) 0 267 non-null
           RægionEstimated Unemployment Rate (%)
                                                  0 267 non-null
                                                                   float64
           lengituseimated Employed
                                                  0 267 non-null
                                                                   int64
           15titu⊞≨timated Labour Participation Rate (%) 267 non-null
                                                                   float64
           dtypeRegiot641
                                                     267 non-null
                                                                   object
                                                     267 non-null
              longitude
                                                                   float64
In [29]:
           1 peimpidptt646464)atimt64619, debject(4)
              ory usage: 18.9+ KB
# Renaming columns for better clarity
               df.columns = ['States', 'Date', 'Frequency', 'Estimated Unemployment Rate', 'Estimated Employed',
                                'Estimated Labour Participation Rate', 'Region', 'longitude', 'latitude']
```

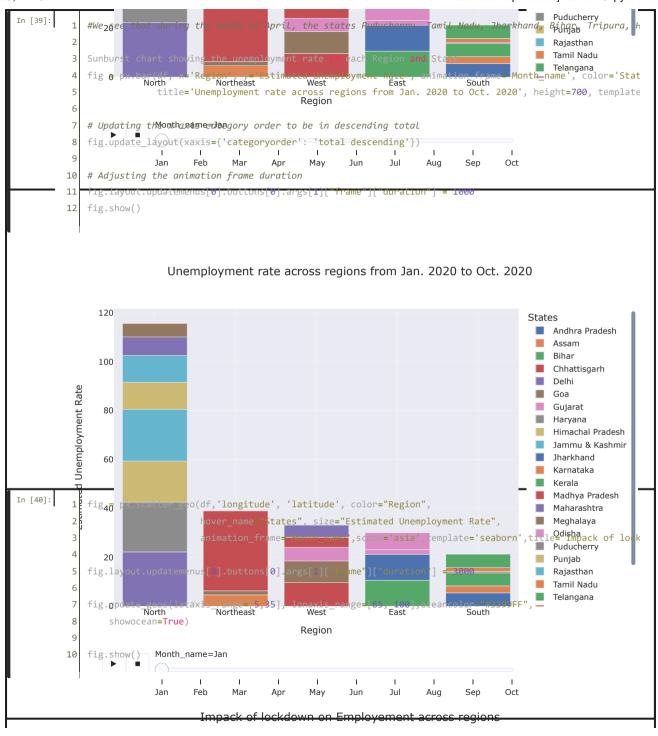
```
In [28]:
              df.isnull().sum()
             ass 'pandas.core.frame.DataFrame'>
          RangeIndex: 267 entries, 0 to 266
          Rationolumns (total 9 columns):
           #ateColumn
                                                 0 Non-Null Count Dtype
           -Frequency-
                                                 0 267 non-null object
           @stiRegednUnemployment Rate (%)
           EstimQated Employed
                                                 0 267 non-null
           ≹stimbre@ue86bur Participation Rate (%) 0 267 non-null
          RagionEstimated Unemployment Rate (%) 0 267 non-null
                                                                   float64
           lengitEsteimated Employed
                                                 0 267 non-null
                                                                   int64
          latitu@etimated Labour Participation Rate (%) 267 non-null
                                                                   float64
                                                    267 non-null
                                                                   object
          d€ypeRegiee841
           7 longitude
                                                    267 non-null
                                                                   float64
In [29]:
           1 peimp@ptt646464)atint64619, dbject(4)
              ory usage: 18.9+ KB
# Renaming columns for better clarity
               df.columns = ['States', 'Date', 'Frequency', 'Estimated Unemployment Rate', 'Estimated Employed',
                               'Estimated Labour Participation Rate', 'Region', 'longitude', 'latitude']
              # Converting 'Date' column to datetime format
               df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
              # Converting 'Frequency' and 'Region' columns to categorical data type
              df['Frequency'] = df['Frequency'].astype('category')
          11
              df['Region'] = df['Region'].astype('category')
          12
              # Extracting month from 'Date' and creating a 'Month' column
              df['Month'] = df['Date'].dt.month
          15
              # Converting 'Month' to integer format
              df['Month_int'] = df['Month'].apply(lambda x: int(x))
          18
          19
              # Mapping integer month values to abbreviated month names
In [30]:
          20
              df[head()
month_name'] = df['Month_int'].apply(lambda x: calendar.month_abbr[x])
          21
              # Dropping the original 'Mestimated Lumn Estimated
                                                                      Estimated
             States Date Frequency Unemployment Estimated Eastern Carop (Columns = Month , inplacate remission) Participation
                                                                        Labour
                                                                                 Region longitude latitude Month_int Month_name
            Andhra 2020-
Pradesh 01-31 M
                                       5.48
                                                        16635535
                                                                  41.02
                                                                                         15.9129
                                                                                                    79.74
                                                                                                                        Jan
                                                                                 South
            Andhra 2020- M
                                                        16545652
                                                                   40.90
                                                                                         15.9129
                                                                                                    79.74
                                                                                 South
                                                                                                                        Feb
            Pradesh 02-29
            Andhra 2020- M
                                                       15881197
                                                                  39.18
                                                                                 South
                                                                                         15.9129
                                                                                                    79.74
                                                                                                                        Mar
            Pradesh
                    03-31
            Andhra 2020- M
                                       20.51
                                                       11336911
                                                                  33.10
                                                                                 South
                                                                                         15.9129
                                                                                                    79.74
                                                                                                                        Apr
            Pradesh 04-30
         4 Andhra 2020-
Pradesh 05-31
                                                                                                                        May
                                       17.43
                                                        12988845
                                                                  36.46
                                                                                 South
                                                                                         15.9129
                                                                                                    79.74
```

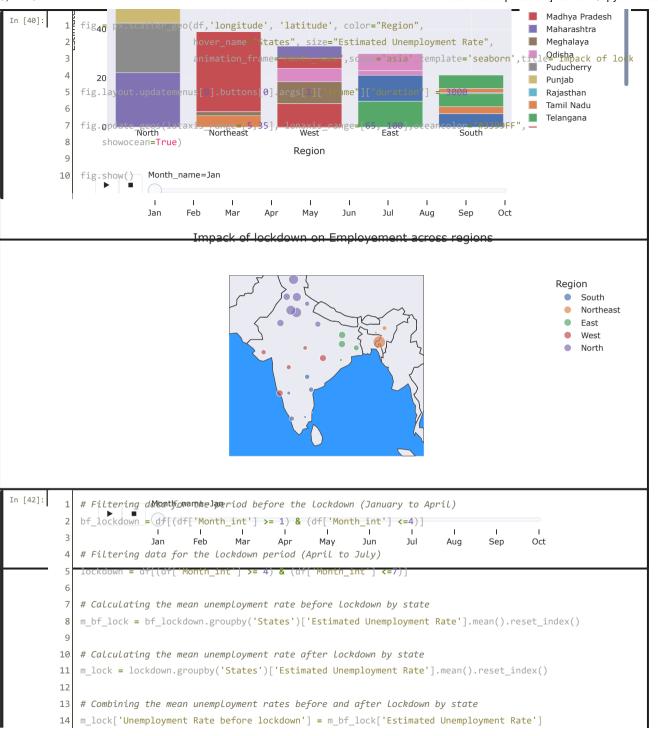
```
# Mapping integer month values to abbreviated month names
In [30]:
                              df_r^h \tilde{m} \tilde{o} dh name'] = df['Month_int'].apply(lambda x: calendar.month_abbr[x])
                                                                                                                                              Estimated
                              # Dropping the original 'Mestimated Lumn Estimated
                                                                                                                                                   Labour
                     States Date Frequency Unemployment of drop(Columns = Month, inplace Rate
                                                                                                                                                                     Region longitude latitude Month_int Month_name
                                                                                                               rutan ployed Participation
                    o Andhra 2020-
Pradesh 01-31
                                                                                                                 16635535
                                                                                                                                       41.02
                                                                                                                                                                                      15.9129
                                                                                                                                                                                                           79.74
                                                                                                                                                                    South
                                                                                                                                                                                                                                                    Jan
                          Andhra 2020- M
                                                                                                                 16545652
                                                                                                                                        40.90
                                                                                                                                                                    South
                                                                                                                                                                                      15.9129
                                                                                                                                                                                                           79.74
                                                                                                                                                                                                                                                    Feb
                          Pradesh 02-29
                          Andhra 2020-
                                                                                                                 15881197
                                                                                                                                        39.18
                                                                                                                                                                    South
                                                                                                                                                                                      15.9129
                                                                                                                                                                                                           79.74
                                                                                                                                                                                                                                                    Mar
                          Pradesh 03-31
                          Andhra 2020- M
                                                                               20.51
                                                                                                                 11336911
                                                                                                                                       33.10
                                                                                                                                                                    South
                                                                                                                                                                                      15.9129
                                                                                                                                                                                                           79.74
                                                                                                                                                                                                                                                    Apr
                          Pradesh 04-30
                         Andhra 2020- M
                                                                               17.43
                                                                                                                 12988845
                                                                                                                                       36.46
                                                                                                                                                                    South
                                                                                                                                                                                      15.9129
                                                                                                                                                                                                           79.74
                                                                                                                                                                                                                                                    May
                         Pradesh 05-31
In [31]:
                              df stat = df[['Estimated Unemployment Rate', 'Estimated Employed', 'Estimated Labour Participation Rate
                              print(round(df_stat.describe().T, 2))
                                                                                                                                            std \
                      Estimated Unemployment Rate
                                                                                                                 12.24
                      Estimated Employed
                                                                                         267.0 13962105.72 13366318.36
                      Estimated Labour Participation Rate 267.0
                                                                                                                 41.68
                                                                                                   min
                                                                                                                         25%
                                                                                                                                               50% \
                                                                                            0.50
                                                                                                                        4.84
                                                                                                                                             9.65
                      Estimated Unemployment Rate
                      Estimated Employed
                                                                                        117542.00 2838930.50 9732417.00
                      Estimated Labour Participation Rate
                                                                                                       75%
                                                                                                                               max
                      Estimated Unemployment Rate
                                                                                                   16.76
                                                                                                                           75.85
                      Estimated Employed
                                                                                         21878686.00 59433759.00
                      Estimated Labour Participation Rate
                                region_stats = df.groupby(['Region'])[['Estimated Unemployment Rate', 'Estimated Employed',
In [33]:
                              import matplotlib.pyplot as plt
                                                                                                                           'Estimated Labour Participation Rate']].mean().reset_index()
                              import seaborn as sns
print(round(region_stats, 2))
In [34]:
                              #Roxinhots eitha Une maling you ment rate perina to tensloyed \
                              hm [48] f[['Estimated Unemployment Rate 02306 Stimated Employed', 'Estimated Labour Participation Rate', '1
                                                                                       15.89
                                                                                                                13072487.92
                              udrtheastm.corr()
                                                                                       10.95
                              pllouffigure(figsize=(6,4))10.45
                                                                                                                 14040589.33
                              West
sns.set_context('notebook', font_scale=1)
                        6 Essamathed takeup (Parta in mation nation nation
                                                                                  40.11
                                                                                  52.06
                       <Axes: >
                                                                                  40.44
                                                                                  41.26
```



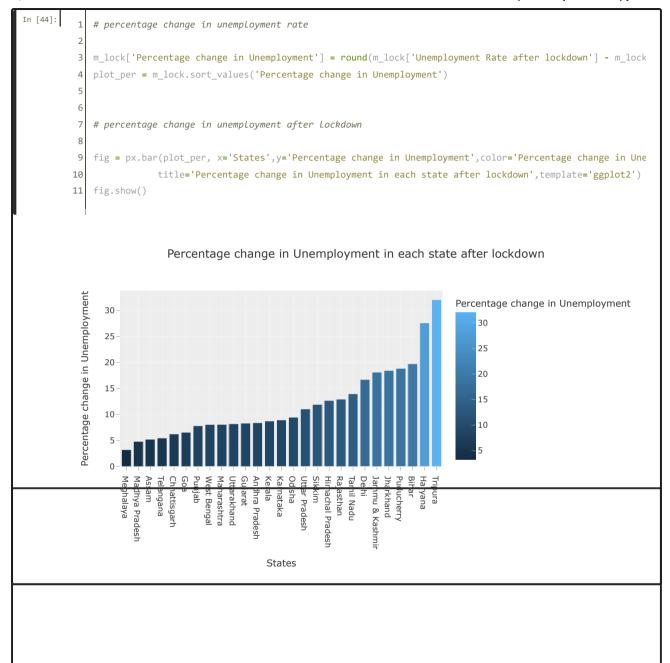








```
In [42]:
             # Filtening deMonth Comarde Jameriod before the Lockdown (January to April)
             bf lockdown = df[(df['Month int'] >= 1) & (df['Month int'] <=4)]
                          Jan
                                  Feb
                                          Mar
                                                          May
                                                                  Jun
                                                                                                    Oct
             # Filtering data for the Lockdown period (April to July)
             lockdown = df[(df[ Month_int ] >= 4) & (df[ Month_int ] <=7)]
             # Calculating the mean unemployment rate before lockdown by state
             m_bf_lock = bf_lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset_index()
             # Calculating the mean unemployment rate after lockdown by state
         10
             m lock = lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset index()
         11
         12
             # Combining the mean unemployment rates before and after Lockdown by state
             m lock['Unemployment Rate before lockdown'] = m bf lock['Estimated Unemployment Rate']
         14
         15
             m lock.columns = ['States','Unemployment Rate before lockdown','Unemployment Rate after lockdown']
         17
             m_lock.head()
                 States Unemployment Rate before lockdown Unemployment Rate after lockdown
        0 Andhra Pradesh 12.3975
                                                         9.4025
        1 Assam
                        6.2450
                                                         6.2250
                        30.8025
                                                         20.7425
        2 Bihar
        3 Chhattisgarh
                       9.6025
                                                         7.2450
        4 Delhi
                        24.3600
                                                         17.6975
```



```
States
1 # Analyzing COVID-19 Impact on Unemployment in India
2
3 # Objective:
5 # The primary aim of this analysis is to assess the repercussions of the COVID-19 pandemic on India's j
6
7 # Dataset Details:
9 # The dataset provides insights into the unemployment scenario across different Indian states:
10
11 # States: The states within India.
12 # Date: The date when the unemployment rate was recorded.
13 # Measuring Frequency: The frequency at which measurements were taken (Monthly).
14 # Estimated Unemployment Rate (%): The percentage of individuals unemployed in each state of India.
```

```
In [46]:
              # Analyzina COVID-19 Impact on Unemployment in India
           2
           3 # Objective:
           4
              # The primary aim of this analysis is to assess the repercussions of the COVID-19 pandemic on India's i
           6
           7 # Dataset Details:
           8
              # The dataset provides insights into the unemployment scenario across different Indian states:
          10
          11 # States: The states within India.
          12 # Date: The date when the unemployment rate was recorded.
          13 # Measuring Frequency: The frequency at which measurements were taken (Monthly).
          14 # Estimated Unemployment Rate (%): The percentage of individuals unemployed in each state of India.
              # Estimated Employed Individuals: The count of people currently employed.
          16 # Estimated Labour Participation Rate (%): The proportion of the working population (age group: 16-64 y
              # This dataset aids in comprehending the unemployment dynamics across India's states during the COVID-1
          17
          18
          19
              # Importing necessary libraries
          20
              # import pandas as pd
              # import numpy as np
          23 # import calendar
              # Loading the dataset into pandas dataframe
          25
          26 # df = pd.read_csv('/kaggle/input/unemployment-in-india/Unemployment_Rate_upto_11_2020.csv')
          27 # df.head()
                                     Frequency Estimated Unemployment Rate (%) Estimated Employed Estimated Labour Pa
          28 # Region Date
                                                                                        South 15.9129 79.74 object
               # 0 Andhra Pradesh
# 1 Date
                                     31-01-2020 M
                                                                 16635535 41.02
267 non-null
                                                                 16545652 40.90
267 non-null
              # 1 Andhra Pradesh
# 2 Frequency
                                     29-02-2020 M 5.83
                                                                                        South 15.9129 79.74 object
                2 Andhra Pradesh 31-03-2020 M 5.79
3 Estimated Unemployment Rate (%)
                                                                 15881197 39.18
267 non-null
                                                                                        South 15.9129 79.74 float64
                 3 Andhra Pradesh 30-04-2020 M 20.51
4 Estimated Employed
                                                                                        South 15.9129 79.74 int64
                                                                                        South 15.9129 79.74
float64
                 4 Andhra Pradesh 31-05-2020 M 17.43
5 Estimated Labour Participation Rate
                                                                 12988845 36.46
(%) 267 non-null
              # Basic information about the dataset
# 6 Region.1
                                                                       267 non-null
                                                                                         object
          35
              # 7
                      Longitude
                                                                                         float64
                                                                       267 non-null
              # df.info()
# 8 latitude
                                                                       267 non-null
                                                                                         float64
              # <class 'pandas.core.frame.DataFrame'>
# dtypes: float64(4), int64(1), object(4)
              # RangeIndex: 267 entries, 0 to 266
# memory usage: 18.9+ KB
              # Data columns (total 9 columns):
# Checking for null values
                                                                       Non-Null Count Dtype
              # df.isnull().sum()
              # 0 Region
# Region
                                                                       267 non-null
                                                                                        object
          E7 # Data
```

```
Suuln 15.9129 /9.74 object
43 # 1 Date
                                                      1003333 41.02
267 non-null
    # 1 Andhra Pradesh 29-02-2020 M
# 2 Frequency
                                                      16545652 40.96
267 non-null
                                                                            South 15.9129 79.74 object
             Frequency
                                                      15881197 39.18
267 non-null
    # 2 Andhra Pradesh 31-03-2020 M 5.79
# 3 Estimated Unemployment Rate (%)
                                                                            South 15.9129 79.74
float64
    # 3 Andhra Pradesh 30-04-2020 M 20.51
# 4 Estimated Employed
                                                      11336911 33.10
267 non-null
                                                                            South 15.9129 79.74
    # 4 Andhra Pradesh 31-05-2020 M 17.43 12988845 36.46
# 5 Estimated Labour Participation Rate (%) 267 non-null
                                                                            South 15.9129 79.74 float64
    # Basic information about the dataset # 6 Region.1
                                                            267 non-null
                                                                             object
 35
49
    # 7 Lonaitude
                                                                             float64
                                                           267 non-null
    # df.info()
# 8 latitude
                                                                             float64
                                                           267 non-null
    # <class 'pandas.core.frame.DataFrame'>
# dtypes: float64(4), int64(1), object(4)
    # RangeIndex: 267 entries, 0 to 266
# memory usage: 18.9+ KB
    # Data columns (total 9 columns):
# Checking for null values
    # # Column
                                                           Non-Null Count Dtype
                                                            ______
 41
55
    # df.isnull().sum()
    # 0 Region
# Region
                                                           267 non-null
                                                                            object
57 # Date
    # Frequency
 59 # Estimated Unemployment Rate (%)
 60 # Estimated Employed
61 # Estimated Labour Participation Rate (%)
 62 # Region.1
 63 # Longitude
 64 # Latitude
 65 # dtype: int64
 66 # Formatting the columns and their datatypes
67
68 # import datetime as dt
 69 # # Renaming columns for better clarity
 70 # df.columns = ['States', 'Date', 'Frequency', 'Estimated Unemployment Rate', 'Estimated Employed',
                       'Estimated Labour Participation Rate', 'Region', 'Longitude', 'Latitude']
71 #
 $6 # # Mapping integer month values to abbreviated month names
 93 # #f&oMoorhingmebate d&blumnthointietenel*6hameda x: calendar.month_abbr[x])
 9¼ # df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
 $9 # # Dropping the original 'Month' column
 98 # #fcdR00kfqhymmfréyg0thvy åNQLaReąłbhe)columns to categorical data type
 97 # df[headquency'] = df['Frequency'].astype('category')
 98 # ấ‡ợtRēgion<sup>D</sup>gte df[FRegtency.asξερέφαξεθεθβρηρβοyment Rate Estimated Employed Estimated Labour Partic
 98 # 0 Andhra Pradesh 2020-01-31 M 5.48 16635535
                                                                  41.02 South 15.9129 79.74 1
 86 # # ARTHGET RAD MARKET 12920 m 02 D2 to 1 Mand 5 c 8 3 at in 16 5 4 5 M 5 Ath 1 48 L 20 M 1
                                                                           South 15.9129 79.74
                                                                                                           Feb
    # afAndbnehPnadea4[ BO20 + 93d21moAth 5.79
                                                      15881197
                                                                   39.18
                                                                            South
                                                                                     15.9129 79.74
    # 3 Andhra Pradesh 2020-04-30 M 20.51
                                                    11336911
                                                                   33.10
                                                                            South
                                                                                    15.9129 79.74
    # # ABANEAtRAGOEMBnt2020005n32geM folma43
                                                     12988845
                                                                   36.46
                                                                           South
                                                                                    15.9129 79.74 5
    # ExplM6AtArintata=aaqtyMinth'].apply(lambda x: int(x))
100 # Pacie statistics
```

```
86 # # Mapping integer month values to abbreviated month names
 93 # #ffoM90F4ingmeDate dffilmmnthoinatetetemel*66mmeda x: calendar.month_abbr[x])
 98 # df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
 $9 # # Dropping the original 'Month' column
 90 # #fcdnopk€QhymmfréYgpħby; änglaReąłbwe)columns to categorical data type
 97 # df[head(wency'] = df['Frequency'].astype('category')
 98 # StateSgionBate affFREGUERCY.asEstimateGeUBEMPSoyment Rate Estimated Employed Estimated Labour Partic
 92 # 0 Andhra Pradesh 2020-01-31 M 5.48 16635535 41.02 South 15.9129 79.74
 80 # # 80 # # ቅደሞኮնዊ ተዋክያሳጠፍክቲ ከ2ዋ2 የመሰያ ነው። Mand 5 ር እያ at in 16 545 በ545 በ545 በ South 15.9129 79.74
 85 # 2-Andbrah Phadesar - 2020-93 at month 5.79
                                             15881197
                                                           39.18
                                                                   South 15.9129 79.74 3
 96 # 3 Andhra Pradesh 2020-04-30 M 20.51 11336911
                                                           33.10
                                                                   South 15.9129 79.74 4
 83 # # CBADEAtRAGGEMBntR02Q005nRegeM format3 12988845
                                                           36.46
                                                                   South 15.9129 79.74 5 May
 84 # ፱୪ምሬክሪክቲክዮጲክዊወቹወ=ወፀፍሃምራክቲከ'].apply(lambda x: int(x))
 99
100
    # Basic statistics
101
102 # df_stat = df[['Estimated Unemployment Rate', 'Estimated Employed', 'Estimated Labour Participation Ra
103 # print(round(df_stat.describe().T, 2))
104 #
                                                                      std \
                                           count
105 # Estimated Unemployment Rate
                                           267.0
                                                        12.24
                                                                     10.80
106 # Estimated Employed
                                           267.0 13962105.72 13366318.36
107 # Estimated Labour Participation Rate 267.0
                                                        41.68
                                                                     7.85
108
109 #
                                                 min
                                                            25%
                                                                        50%
110 # Estimated Unemployment Rate
                                                0.50
                                                            4.84
                                                                       9.65
111 # Estimated Employed
                                           117542.00 2838930.50 9732417.00
112 # Estimated Labour Participation Rate
                                               16.77
                                                          37.26
                                                                       40.39
113
114 #
                                                   75%
                                                               max
139 # Estimated Unemployment Rate
                                       40.11
                                                16.76
                                                             75.85
138 # Estimated Employed
                                       <sup>38.70</sup>1878686.00 59433759.00
13→ # €stimated Labour Participation Ra €€.06
                                                 44.06
                                                             69.69
# Region_stats = df.groupby(['Region 9]44['Estimated Unemployment Rate', 'Estimated Employed',
133 # 4
                                       41.26 'Estimated Labour Participation Rate']].mean().reset index()
130 # imenetroutakpetionpyetes, as) plt
135 # importReagebarnEg€imO€ed Unemployment Rate Estimated Employed \
‡ fopt/condg/fib/python3.10/site-packagqs/scipy/__initengraphic UserWarning: A NumPy version >=1.16.5
43ኛ # 1 warnings+warn(f"A NumPy version >={ቁρያซinversion}3gnd4sfng2maxversion}"
138 # Heatmarheast
                                          10.95
                                                        3617105.53
                                          10.45
                                                        14040589.33
    # Mm = df[Westimated Unemployment Rate's, 24Estimated Templeyed'7, 'Estimated Labour Participation Rate',
144 # hm = hm.corr()
142 # pltefigHG46figSb3GF(Patt)cination Rate
1/12 # cnc cot contact/!notabash! font code 1)
```

```
40.11
# 2stimated Unemployment Rate
                                                                                                                          16.76
                                                                                                                                                           75.85
130 # #stimated Employed
                                                                                                  38.701878686.00 59433759.00
137 # Estimated Labour Participation Ra 2.06
                                                                                                                          44.06
## Pegion stats = df.groupby(['Regioh944|'Estimated Unemployment Rate', 'Estimated Employed',
133 # 4
                                                                                                  41.26 'Estimated Labour Participation Rate']].mean().reset index()
130 # preaffroatakpthibnpsehet, as) plt
‡፯፮ # importRegebarnEgeimued Unemployment Rate Estimated Employed \
# fopt/condg/fib/python3.10/site-packages/scipy/__init_6028K6146: UserWarning: A NumPy version >=1.16.5
‡3፯ # 1 warnings+warn(f"A NumPy version >={ዋpgminversion}3gnd4gfnp2maxversion}"
138 # Heatmarheast
                                                                                                          10.95
                                                                                                                                              3617105.53
139 # 3
                                                                                                          10.45
                                                                                                                                            14040589.33
140 # hm = df[[hekstimated Unemployment Rate's, 24Estimated Employed'; 'Estimated Labour Participation Rate',
144 # hm = hm.corr()
143 # pltefigHGe6figSb3GF(Part)cipation Rate
143 # sns.set context('notebook', font scale=1)
144 # sns.heatmap(data=hm, annot=True, cmap=sns.cubehelix_palette(as_cmap=True))
145 # <Axes: >
146
147 # Boxplot of Unemployment rate per States
148
149 # import plotly.express as px
150 # fig = px.box(df, x='States', y='Estimated Unemployment Rate', color='States', title='Unemployment rat
151
152 # # Updating the x-axis category order to be in descending total
153 # fig.update layout(xaxis={'categoryorder': 'total descending'})
154 # fig.show()
155 # Scatter matrix cosidering the employed and unemployed rates
156
157 # fig = px.scatter_matrix(df,template='seaborn',dimensions=['Estimated Unemployment Rate', 'Estimated E
#Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart showing the unemployment rate across regions from FRA Market Bar chart shows the same across regions from FRA Market Bar chart shows the same across regions from FRA Market Bar chart shows the same across regions from FRA Market Bar chart shows the same across the same across the same across the same across regions from FRA Market Bar chart shows the same across the sam
139 # fig.show()
130 # ชิลัต กูเอช รูคอัฟส์กู ชุกิ่ยใช้สายเลือนกล่อย มกิ่ยโลยาสัตร์ และเลือนกลาย เลือนกลาย เลือน
                                                 title='Unemployment rate across regions from Jan. 2020 to Oct. 2020', height=700, templa
475 #
133 # \#f^Updemireve^{-x}pevi^{-x}pevi^{-x}unempggrøupbge^{r}states ^{-1}?^{-1}meaneg^{-1}deset^{-1}enempggrøupbge^{-1}
# fig.update_layout(xaxis={'categoryorder': 'total descending'})
169 # df_unemployed = df_unemployed.sort_values('Estimated Unemployment Rate')
188 # # Adjusting the animation frame duration
189 # fig.Laboutampdqtamemblooldobuttoselolsargs[t]ttihaqqa"Jhembyotmemt]Rattoqqqolor = 'States'.title = 'Aver
182 # fig.show() template='seaborn')
# WeaseRoth9t during the month of April, the states Puducherry, Tamil Nadu, Jharkhand, Bihar, Tripura,
190 # Haryana and Jharkhand have long been the most unemployed.
           # Sunburst chart showing the unemployment rate in each Region and State
```

100

```
# Bar chart showing the unemployment rate across regions from E349 ma2020 LtboOftpa2020 in Rate'], colo
133
           # fig.show()
           # ซีลัด อิเอิซ์ ยิศิธิผ์สีกล ซักี่ผลีดังคล่อย หกี่คลิปังพฤษัศ หลือยาโภงพละที่ ซึสซอ่า ลnimation_frame='Month_name', color='St
176
475
                                                  title='Unemployment rate across regions from Jan. 2020 to Oct. 2020', height=700, templa
176
           # plot unemp = df[['Estimated Unemployment Rate','States']]
183 # #fUndeminovible=xpexisumembgaryuobderstates:inmedaeegndinatetenaex()
           # fig.update Layout(xaxis={'categoryorder': 'total descending'})
            # df_unemployed = df_unemployed.sort_values('Estimated Unemployment Rate')
           # # Adjusting the animation frame duration
           # fig.LaxxubayraftememutbolabuxiostiQlsargs[filtimates"]himhutstinent]rātl@@Color = 'States'.title = 'Aver
182 # fig.show() template='seaborn')
           # មុខ្មែនទូក្រុស្សា during the month of April, the states Puducherry, Tamil Nadu, Jharkhand, Bihar, Tripura,
            # Haryana and Jharkhand have long been the most unemployed.
            # Sunburst chart showing the unemployment rate in each Region and State
186
187
            # # Creating a DataFrame with relevant columns
            # unemployed df = df[['States', 'Region', 'Estimated Unemployment Rate', 'Estimated Employed', 'Estimat
189
            # unemployed = unemployed_df.groupby(['Region', 'States'])['Estimated Unemployment Rate'].mean().reset_
191
            # # Creating a Sunburst chart
           # fig = px.sunburst(unemployed, path=['Region', 'States'], values='Estimated Unemployment Rate', color
193
194
                                                              title='Unemployment rate in each Region and State', height=550, template='presentat
195
           # fig.show()
196
197
            # Impact of Lockdown on States Estimated Employed
198
            # fig = px.scatter geo(df, 'longitude', 'latitude', color="Region",
199
                                                                      hover_name="States", size="Estimated Unemployment Rate",
200
            # lockdown = df[(df['Monthmint'] + \bar{r}ame] = Month_int' + \bar{r}ame] = Month_int' + \bar{r}ame = Month_int' + \bar{r}ame = Monthmint' + \bar{r}am
305
305
            ##iG?laybatingathemeras[uncentermerto]rafashaforqrhaekdowauhutetate = 3000
            # m_bf_lock = bf_lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset_index()
           # fig.update_geos(lataxis_range=[5,35], lonaxis_range=[65, 100],oceancolor="#3399FF",
            # # Casebbationg=thgemean unemployment rate after lockdown by state
           # m Lock = Lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset index()
208
           # fig.show()
200 # #heombitheantheateng urethurumeetmsateshbefonerendnefittovkackdewreby state
# m_lock['Unemployment Rate before lockdown'] = m_bf_lock['Estimated Unemployment Rate']
           # # Filtering data for the period before the lockdown (January to April)
          # B_{f}_06&k\d0\Hm\subseteq d\f[{\dagger}[\dagger]\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\f
           # m_Lock.head()
           # #t#ttterinUnameloxmenthBatBckafQneplatbdowApriUnemplayment Rate after Lockdown
           # 0 Andhna Daadach 12 2075 0 4025
```

```
# Lockdown = df[(df['Monthmant', rame=&M6A£h'Mameh_sitbpe= {asia', template='seaborn', title='Impack of Lo
                206
                         # #iGalaybatingathemenas/byenalexment/ratas/afforerbackdownubxtitate = 3000
                         # m_bf_lock = bf_lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset_index()
                209
                        # fig.update_geos(lataxis_range=[5,35], lonaxis_range=[65, 100],oceancolor="#3399FF",
                        ## Calaybating the mean unemployment rate after lockdown by state
                        # m Lock = Lockdown.groupby('States')['Estimated Unemployment Rate'].mean().reset index()
                327
                        # fig.show()
                         # #h@ombiriegntheareng urernurymenemsateshbreomerendnemtebvbackdowneby state
                         # m_lock['Unemployment Rate before lockdown'] = m_bf_lock['Estimated Unemployment Rate']
                        # # Filtering data for the period before the lockdown (January to April)
                        # \mathfrak{B}_{f}Leekkalkum\underline{n}saf[{dfamasin_'Unemgley\underline{m}anter[basaee, faekdev\underline{n}4) jUnemployment Rate after Lockdown']
                236
                         # m_lock.head()
                223
                         # $t#teterinUnameLoxmentnBateckafaneplatedowApriUnemplayment Rate after Lockdown
                        # 0 Andhra Pradesh 12.3975 9.4025
                230
                        # 1 Assam 6.2450 6.2250
                         # 2 Bihar 30.8025 20.7425
                         # 3 Chhattisaarh 9.6025 7.2450
                232
                         # 4 Delhi 24.3600 17.6975
                233
                         # # percentage change in unemployment rate
                235
                         \# \ m\_lock['Percentage \ change \ in \ Unemployment'] = round(m\_lock['Unemployment \ Rate \ after \ lockdown'] - m\_lock['Percentage \ change \ in \ Unemployment'] + m\_lock['Unemployment \ Rate \ after \ lockdown'] - m\_lock['Nemployment \ Rate \ after \ lockdown'] + m\_lock['Unemployment \ Rate \ After \ Aft
                236
                         # plot_per = m_lock.sort_values('Percentage change in Unemployment')
                237
                238
                239
                240
                         # # percentage change in unemployment after lockdown
                241
                         # fig = px.bar(plot_per, x='States',y='Percentage change in Unemployment',color='Percentage change in U
                243
                                                    title='Percentage change in Unemployment in each state after lockdown',template='ggplot2'
In [50]:
               244
                         # fig.show()
                245
                         # The most affected states/territories in India during the lockdown in case of unemployment were:
                246
In [51]:
                         import pandas as pd
                        # Tripura import matplotlib.pyplot as plt
                        # Haryana
import seaborn as sns
                         from sklearn.model_selection import train_test_split
                         # Puducherry
from sklearn.linear_model import LinearRegression
                        # Jharkhand
from sklearn.linear_model import Lasso
                        # Jammu & Kashmir
from sklearn import metrics
                253
                        # Delhi
In [53]:
In [47]:
                         # inspecting the first 5 rows of the dataframe
```

```
In [50]:
                     244
                                  # fig.show()
                      245
                                  # The most affected states/territories in India during the lockdown in case of unemployment were:
                      246
In [51]:
                                   import pandas as pd
                                 # Tripura
import matplotlib.pyplot as plt
                      247
                     248
                                  # Haryana
import seaborn as sns
                      249
4
                                  # Bihar
from sklearn.model_selection import train_test_split
                      250
                                  # Puducherry from sklearn.linear_model import LinearRegression
                                 # Jharkhand from sklearn.linear_model import Lasso
                      251
                                 # Jammu & Kashmir
from sklearn import metrics
                      253
                                  # Delhi
Tn [53]:
In [47]:
                                   In [54]:
                                  # inspecting the first 5 rows of the dataframe
                                  car dataset.head()
                              Car_Name Year Selling_Price Present_Price Driven_kms Fuel_Type Selling_type Transmission Owner
                       0 ritz
                                                        2014 3.35
                                                                                                      5.59
                                                                                                                                        27000
                                                                                                                                                                      Petrol
                                                                                                                                                                                                Dealer
                                                                                                                                                                                                                                                               0
                                                                                                                                                                                                                              Manual
                                                        2013 4.75
                                                                                                     9.54
                                                                                                                                        43000
                                                                                                                                                                                               Dealer
                       1 sx4
                                                                                                                                                                     Diesel
                                                                                                                                                                                                                              Manual
                                                                                                                                                                                                                                                               0
                       2 ciaz
                                                        2017 7.25
                                                                                                     9.85
                                                                                                                                        6900
                                                                                                                                                                     Petrol
                                                                                                                                                                                               Dealer
                                                                                                                                                                                                                              Manual
                                                                                                                                                                                                                                                               0
                                                        2011 2.85
                                                                                                     4.15
                                                                                                                                        5200
                                                                                                                                                                     Petrol
                                                                                                                                                                                               Dealer
                                                                                                                                                                                                                                                               0
                       3 wagon r
                                                                                                                                                                                                                              Manual
                       4 swift
                                                        2014 4.60
                                                                                                     6.87
                                                                                                                                        42450
                                                                                                                                                                     Diesel
                                                                                                                                                                                                Dealer
                                                                                                                                                                                                                              Manual
                                                                                                                                                                                                                                                               0
In [55]:
                                  # checking the number of rows and columns
                                  car dataset.shape
                         (301, 9)
                                 # checking the number of missing values taset
                                 car_dataset.isnull().sum()
car_dataset.info()
                       Car Name
(clars)
Vears
(clars)
                                                                 301 non-null object
                           Selling Type 0
                                                                 301 non-null int64
                          Transmission 0 301 non-null float64
                         Owner Price 301 non-null float64 dtype: int64 4 Driven_kms 301 non-null int64
                           6 Selling_type 301 non-null
                                                                                                 object
                           7 Transmission 301 non-null
                                                                 301 non-null
                         dtypes: float64(2), int64(3), object(4)
                         memory usage: 21.3+ KB
```

```
# checking the number of missing values
# getting some information about the dataset
                car_dataset.isnull().sum()
            Year
RangeIndex: 301 entries, 0 to 300
            Rangerhoex: 50 entries, 0 to 50 Selling Price 0
Data columns (total 9 columns):
Present Price 0
# Column Non-Null Cou
                               Non-Null Count Dtype
                               301 non-null object
                               301 non-null int64
            Transmission 0
2 Selling Price 301 non-null float64
            Owner 0
3 Present Price 301 non-null float64
            dtype: int64
4 Driven_kms
                               301 non-null int64
             6 Selling type 301 non-null
                                               object
             7 Transmission 301 non-null
                                               object
                               301 non-null
            dtypes: float64(2), int64(3), object(4)
            memory usage: 21.3+ KB
In [58]:
                # checking the distribution of categorical data
                print(car_dataset.Fuel_Type.value_counts())
                print(car_dataset.Selling_type.value_counts())
                print(car_dataset.Transmission.value_counts())
            Fuel_Type
            Petrol
            Diesel
            CNG
            Name: count, dtype: int64
            Selling type
                         195
            Dealer
            Individual 106
            Name: count, dtype: int64
            Transmission
            Manual
                     261
               le: count, dtype: int64
  car_dataset.head()
In [60]:
In [59]:
              # encoding "Fuel Type" Column

Car_Name Year Selling_Price Present_Price Driven_kms Fuel_Type Selling_type Transmission Owner
                x4 2013 4.75 9.54
# encoding "Seller_Type" Column
az 2017 7.25 9.85
                                                                 43000
                                                                                           0
                                                                                                          0
                                                                                                                          0
                                                                 6900
                2014 4.60
                                                                 42450
                # encoding "Transmission" Column
                car_dataset.replace({'Transmission':{'Manual':0,'Automatic':1}},inplace=True)
X = car_dataset.drop(['Car_Name','Selling_Price'],axis=1)
In [61]:
                Y = car_dataset['Selling_Price']
In [62]:
                print(X)
```

```
In [60]:
              e: count, dtype: int64
car dataset.head()
In [59]:
             # encoding "Fuel_Type" Column

Car_Name Year Selling_Price Present_Price Driven_kms Fuel_Type Selling_type Transmission Owner
              x4 2013 4.75 9.54
# encoding "Seller_Type" Column
az 2017 7.25 9.85
                                                            43000
                                                                                     0
                                                                                                  0
                                                                                                                 0
                                                            6900
                                                                         0
                                                                                                  0
              car_dataset.replace({'Selling_type':{'Dealer':0,'Individual':1}},inplace=True)
                        2014 4.60
               # encoding "Transmission" Column
               car_dataset.replace({'Transmission':{'Manual':0,'Automatic':1}},inplace=True)
In [61]:
               X = car_dataset.drop(['Car_Name', 'Selling_Price'], axis=1)
               Y = car_dataset['Selling_Price']
In [62]:
               print(X)
               Year Present Price Driven kms Fuel Type Selling type Transmission \
               2014
                             5.59
                                       27000
               2013
                             9.54
                                       43000
                             9.85
               2011
                             4.15
                                       5200
               2014
                             6.87
                                       42450
           296 2016
                           11.60
                                      33988
           297 2015
                            5.90
                                       60000
           298 2009
                           11.00
           299 2017
                           12.50
                                       9000
                            5.90
           300 2016
                                       5464
               Owner
In [63]:
               print(Y)
           299
                  3035
           000
                  4.75
           2301 row₹.257 columns]
                  2.85
           296
                  9.50
           297
                  4.00
           298
                  3.35
           299
                 11.50
           300
           Name: Selling_Price, Length: 301, dtype: float64
In [64]:
               X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.1, random_state=2)
```

