Covid19India

April 16, 2020

1 Covid19India - EDA

Data Description The dataset consists of the information about Covid19India cases taken from Covid19India API.

Below is a table showing names of all the columns and their description.

Attributes	Dtype
agebracket	object
backupnotes	object
contractedfromwhichpatientsuspected	object
currentstatus	object
dateannounced	object
detectedcity	object
detecteddistrict	object
detectedstate	object
estimatedonsetdate	object
gender	object
nationality	object
notes	object
patientnumber	object
source1	object
source2	object
source3	object
statecode	object
statepatientnumber	object
statuschangedate	object
typeoftransmission	object

1.1 Import Libraries

[1]:

```
import os
from requests import request
import urllib.request
import json
from pandas.io.json import json_normalize
```

```
import numpy as np
import pandas as pd
import pandas_profiling
import seaborn as sns
import matplotlib.pyplot as plt
import plotly
import plotly.graph_objects as go
import plotly.express as px

%matplotlib inline
```

2 Read Data from Covid19India API

[2]:

```
def read_from_api(URL):
    response = request(url=URL, method='get')
    x = URL.split('/').pop(-1)
    x = x[:-5]
    elevations = response.json()
    rec = elevations[x]
    return json_normalize(rec)
```

[3]:

```
df_raw_data = read_from_api('https://api.covid19india.org/raw_data.json')
df_raw_data.head()
```

[3]:

```
  agebracket                                 backupnotes  \
0         20                           Student from Wuhan
1                                      Student from Wuhan
2                                      Student from Wuhan
3         45         Travel history to Italy and Austria
4         24  Travel history to Dubai, Singapore contact

  contractedfromwhichpatientsuspected currentstatus dateannounced  \
0                                         Recovered    30/01/2020
1                                         Recovered    02/02/2020
2                                         Recovered    03/02/2020
3                                         Recovered    02/03/2020
4                                         Recovered    02/03/2020

               detectedcity detecteddistrict detectedstate estimatedonsetdate  \
0                  Thrissur         Thrissur        Kerala
1                 Alappuzha        Alappuzha        Kerala
2                 Kasaragod        Kasaragod        Kerala
3  East Delhi (Mayur Vihar)       East Delhi         Delhi
4                 Hyderabad        Hyderabad     Telangana

  gender nationality                                              notes  \
0      F       India                            Travelled from Wuhan
1              India                            Travelled from Wuhan
2              India                            Travelled from Wuhan
3      M       India                   Travelled from Austria, Italy
4      M       India  Travelled from Dubai to Bangalore on 20th Feb,...

  patientnumber                                            source1  \
0             1  https://twitter.com/vijayanpinarayi/status/122...
1             2  https://www.indiatoday.in/india/story/kerala-r...
2             3  https://www.indiatoday.in/india/story/kerala-n...
3             4  https://www.indiatoday.in/india/story/not-a-ja...
4             5  https://www.deccanherald.com/national/south/qu...

                                             source2  \
0  https://weather.com/en-IN/india/news/news/2020...
1  https://weather.com/en-IN/india/news/news/2020...
2  https://twitter.com/ANI/status/122422148580539...
3  https://economictimes.indiatimes.com/news/poli...
4  https://www.indiatoday.in/india/story/coronavi...

                                             source3 statecode  \
0                                                           KL
1                                                           KL
2  https://weather.com/en-IN/india/news/news/2020...        KL
3                                                           DL
4  https://www.thehindu.com/news/national/coronav...        TG

  statepatientnumber statuschangedate typeoftransmission
0           KL-TS-P1       14/02/2020           Imported
1           KL-AL-P1       14/02/2020           Imported
2           KL-KS-P1       14/02/2020           Imported
3              DL-P1       15/03/2020           Imported
4              TS-P1       02/03/2020           Imported
```

[4]:

```
df_death_and_recoveries = read_from_api('https://api.covid19india.org/
 ↪deaths_recoveries.json')
df_death_and_recoveries.head()
```

[4]:

```
  agebracket       city        date         district gender nationality  \
0         85     Mumbai  29/03/2020           Mumbai      M
1         80     Mumbai  29/03/2020           Mumbai      M
2         86  Ghatkopar  29/03/2020  Mumbai Suburban      F
3                        29/03/2020           Mumbai
4                        29/03/2020           Mumbai

                                               notes  \
0  Suffering from Diabetes, had a pacemaker, no t...
1  patient passed away at the Fortis Hospital, Mu...
```

```
2
3
4

  patientnumbercouldbemappedlater patientstatus slno  \
0                                      Deceased    1
1                                      Deceased    2
2                                      Deceased    3
3                                      Deceased    4
4                                      Deceased    5

                                             source1  \
0  https://arogya.maharashtra.gov.in/pdf/epressno...
1  https://arogya.maharashtra.gov.in/pdf/epressno...
2  https://arogya.maharashtra.gov.in/pdf/epressno...
3  https://arogya.maharashtra.gov.in/pdf/epressno...
4  https://arogya.maharashtra.gov.in/pdf/epressno...

                                             source2 source3        state  \
0  https://www.deccanherald.com/national/west/dea...          Maharashtra
1  https://www.indiatoday.in/india/story/coronavi...          Maharashtra
2                                                             Maharashtra
3                                                             Maharashtra
4                                                             Maharashtra

  statecode
0        MH
1        MH
2        MH
3        MH
4        MH
```

[5]:
```
df_raw_data.head()
```

[5]:
```
  agebracket                                backupnotes  \
0         20                         Student from Wuhan
1                                    Student from Wuhan
2                                    Student from Wuhan
3         45         Travel history to Italy and Austria
4         24  Travel history to Dubai, Singapore contact

  contractedfromwhichpatientsuspected currentstatus dateannounced  \
0                                         Recovered    30/01/2020
1                                         Recovered    02/02/2020
2                                         Recovered    03/02/2020
3                                         Recovered    02/03/2020
4                                         Recovered    02/03/2020
```

```
               detectedcity detecteddistrict detectedstate estimatedonsetdate  \
0                  Thrissur         Thrissur        Kerala
1                 Alappuzha        Alappuzha        Kerala
2                 Kasaragod        Kasaragod        Kerala
3  East Delhi (Mayur Vihar)       East Delhi         Delhi
4                 Hyderabad        Hyderabad     Telangana

  gender nationality                                               notes  \
0      F       India                                Travelled from Wuhan
1              India                                Travelled from Wuhan
2              India                                Travelled from Wuhan
3      M       India                       Travelled from Austria, Italy
4      M       India  Travelled from Dubai to Bangalore on 20th Feb,...

  patientnumber                                            source1  \
0             1  https://twitter.com/vijayanpinarayi/status/122...
1             2  https://www.indiatoday.in/india/story/kerala-r...
2             3  https://www.indiatoday.in/india/story/kerala-n...
3             4  https://www.indiatoday.in/india/story/not-a-ja...
4             5  https://www.deccanherald.com/national/south/qu...

                                             source2  \
0  https://weather.com/en-IN/india/news/news/2020...
1  https://weather.com/en-IN/india/news/news/2020...
2  https://twitter.com/ANI/status/12242214858053...
3  https://economictimes.indiatimes.com/news/poli...
4  https://www.indiatoday.in/india/story/coronavi...

                                             source3 statecode  \
0                                                           KL
1                                                           KL
2  https://weather.com/en-IN/india/news/news/2020...        KL
3                                                           DL
4  https://www.thehindu.com/news/national/coronav...        TG

  statepatientnumber statuschangedate typeoftransmission
0           KL-TS-P1       14/02/2020           Imported
1           KL-AL-P1       14/02/2020           Imported
2           KL-KS-P1       14/02/2020           Imported
3              DL-P1       15/03/2020           Imported
4              TS-P1       02/03/2020           Imported
```

[6]:

```
df_raw_data.columns
```

[6]:

```
Index(['agebracket', 'backupnotes', 'contractedfromwhichpatientsuspected',
       'currentstatus', 'dateannounced', 'detectedcity', 'detecteddistrict',
       'detectedstate', 'estimatedonsetdate', 'gender', 'nationality', 'notes',
       'patientnumber', 'source1', 'source2', 'source3', 'statecode',
       'statepatientnumber', 'statuschangedate', 'typeoftransmission'],
      dtype='object')
```

[7]:
```
df_raw_data.shape
```

[7]:
```
(13060, 20)
```

[8]:
```
data=df_raw_data.copy()
data.head()
```

[8]:
```
  agebracket                                 backupnotes  \
0         20                          Student from Wuhan
1                                     Student from Wuhan
2                                     Student from Wuhan
3         45         Travel history to Italy and Austria
4         24  Travel history to Dubai, Singapore contact

  contractedfromwhichpatientsuspected currentstatus dateannounced  \
0                                         Recovered    30/01/2020
1                                         Recovered    02/02/2020
2                                         Recovered    03/02/2020
3                                         Recovered    02/03/2020
4                                         Recovered    02/03/2020

               detectedcity detecteddistrict detectedstate estimatedonsetdate  \
0                  Thrissur         Thrissur        Kerala
1                 Alappuzha        Alappuzha        Kerala
2                 Kasaragod        Kasaragod        Kerala
3  East Delhi (Mayur Vihar)       East Delhi         Delhi
4                 Hyderabad        Hyderabad     Telangana

  gender nationality                                              notes  \
0      F       India                               Travelled from Wuhan
1              India                               Travelled from Wuhan
2              India                               Travelled from Wuhan
3      M       India                      Travelled from Austria, Italy
4      M       India  Travelled from Dubai to Bangalore on 20th Feb,...

  patientnumber                                            source1  \
0             1  https://twitter.com/vijayanpinarayi/status/122...
1             2  https://www.indiatoday.in/india/story/kerala-r...
2             3  https://www.indiatoday.in/india/story/kerala-n...
3             4  https://www.indiatoday.in/india/story/not-a-ja...
4             5  https://www.deccanherald.com/national/south/qu...

                                             source2  \
0  https://weather.com/en-IN/india/news/news/2020...
1  https://weather.com/en-IN/india/news/news/2020...
2  https://twitter.com/ANI/status/122422148580539...
```

```
3  https://economictimes.indiatimes.com/news/poli...
4  https://www.indiatoday.in/india/story/coronavi...

                                              source3 statecode  \
0                                                            KL
1                                                            KL
2  https://weather.com/en-IN/india/news/news/2020...        KL
3                                                            DL
4  https://www.thehindu.com/news/national/coronav...        TG

  statepatientnumber statuschangedate typeoftransmission
0          KL-TS-P1       14/02/2020           Imported
1          KL-AL-P1       14/02/2020           Imported
2          KL-KS-P1       14/02/2020           Imported
3             DL-P1       15/03/2020           Imported
4             TS-P1       02/03/2020           Imported
```

[9]:
```
profile = pandas_profiling.ProfileReport(df_raw_data)
profile.to_file(output_file="covid19_data_before_preprocessing.html")
```

[10]:
```
#pandas_profiling.ProfileReport(df)
```

Observations - `agebracket` has a high cardinality: 86 distinct values - `backupnotes` has a high cardinality: 223 distinct values - `contractedfromwhichpatientsuspected` has a high cardinality: 144 distinct values - `detectedcity` has a high cardinality: 313 distinct values
- `detecteddistrict` has a high cardinality: 349 distinct values
- `estimatedonsetdate` has constant value as NULL NEEDS TO BE Rejected - `notes` has a high cardinality: 709 distinct values
- `source1` has a high cardinality: 785 distinct values - `source2` has a high cardinality: 338 distinct values - `source3` has a high cardinality: 102 distinct values - `statepatientnumber` has a high cardinality: 1463 distinct values

[11]:
```
print("Data Shape : Rows = {} , Columns = {}".format(df_raw_data.
 ↪shape[0],df_raw_data.shape[1]))
```

```
Data Shape : Rows = 13060 , Columns = 20
```

[12]:
```
print("Column Names are : \n", df_raw_data.columns)
```

```
Column Names are :
 Index(['agebracket', 'backupnotes', 'contractedfromwhichpatientsuspected',
       'currentstatus', 'dateannounced', 'detectedcity', 'detecteddistrict',
       'detectedstate', 'estimatedonsetdate', 'gender', 'nationality', 'notes',
       'patientnumber', 'source1', 'source2', 'source3', 'statecode',
       'statepatientnumber', 'statuschangedate', 'typeoftransmission'],
      dtype='object')
```

[13]:

```
df_raw_data.drop(['estimatedonsetdate', 'notes',␣
 ↪'contractedfromwhichpatientsuspected', 'source1', 'source2', 'source3',␣
 ↪'backupnotes' ], axis = 1, inplace = True)
df_raw_data.sample(10)
```

[13]:

```
      agebracket currentstatus dateannounced  detectedcity  \
1887              Hospitalized    01/04/2020
6229              Hospitalized    09/04/2020  Shastrinagar
12903             Hospitalized    16/04/2020
1971              Hospitalized    01/04/2020
6947              Hospitalized    10/04/2020
2552              Hospitalized    03/04/2020
3522              Hospitalized    04/04/2020
9729              Hospitalized    13/04/2020
7663              Hospitalized    11/04/2020
11139             Hospitalized    14/04/2020

          detecteddistrict   detectedstate gender nationality patientnumber  \
1887              Thrissur          Kerala                             1888
6229                Jaipur       Rajasthan                             6230
12903                       Madhya Pradesh                            12904
1971                Jorhat           Assam                             1972
6947            Coimbatore      Tamil Nadu      F                      6948
2552                  Tonk       Rajasthan                             2553
3522                Bhopal  Madhya Pradesh                             3523
9729   Gautam Buddha Nagar   Uttar Pradesh                             9730
7663              Vadodara         Gujarat                             7664
11139              Barwani  Madhya Pradesh                            11140

      statecode statepatientnumber statuschangedate typeoftransmission
1887         KL                          01/04/2020
6229         RJ                          09/04/2020
12903        MP                          16/04/2020
1971         AS                          01/04/2020              Local
6947         TN            TN-P868       10/04/2020
2552         RJ                          03/04/2020              Local
3522         MP                          04/04/2020
9729         UP                          13/04/2020
7663         GJ                          11/04/2020
11139        MP                          14/04/2020
```

[14]:

```
df_raw_data['agebracket'] = pd.to_numeric(df_raw_data['agebracket'],␣
 ↪errors='coerce')
df_raw_data['agebracket'] = df_raw_data['agebracket'].astype('float')
#df['patientnumber'] = df['patientnumber'].astype('float')
```

[15]:

```
df_raw_data['statuschangedate'] = pd.
 ↪to_datetime(df_raw_data['statuschangedate'])
```

```
df_raw_data['dateannounced'] = pd.to_datetime(df_raw_data['dateannounced'])

df_raw_data['durationOfAnyStatus'] = df_raw_data['statuschangedate'] -␣
 ↪df_raw_data['dateannounced']
df_raw_data['durationOfAnyStatus'] = df_raw_data['durationOfAnyStatus'].dt.days

df_raw_data['statuschangedate'] = df_raw_data['statuschangedate'].dt.
 ↪strftime('%Y-%m-%d')
df_raw_data['dateannounced'] = df_raw_data['dateannounced'].dt.
 ↪strftime('%Y-%m-%d')
```

[16]:
```
df_raw_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13060 entries, 0 to 13059
Data columns (total 14 columns):
agebracket             1551 non-null float64
currentstatus          13060 non-null object
dateannounced          13060 non-null object
detectedcity           13060 non-null object
detecteddistrict       13060 non-null object
detectedstate          13060 non-null object
gender                 13060 non-null object
nationality            13060 non-null object
patientnumber          13060 non-null object
statecode              13060 non-null object
statepatientnumber     13060 non-null object
statuschangedate       13060 non-null object
typeoftransmission     13060 non-null object
durationOfAnyStatus    12858 non-null float64
dtypes: float64(2), object(12)
memory usage: 1.4+ MB
```

[17]:
```
df_raw_data.sample(10)
```

[17]:
```
       agebracket currentstatus dateannounced detectedcity detecteddistrict  \
7979          NaN  Hospitalized    2020-11-04                   Thiruvallur
3290          NaN  Hospitalized    2020-04-04
9343          NaN  Hospitalized    2020-04-13                        Mumbai
1414         46.0  Hospitalized    2020-03-31    Kandukuru         Prakasam
5226          NaN  Hospitalized    2020-07-04
969           NaN  Hospitalized    2020-03-28                        Mumbai
10431         NaN  Hospitalized    2020-04-13
7630          NaN  Hospitalized    2020-11-04                     Ahmadabad
7918         37.0  Hospitalized    2020-11-04  Arundelpeta           Guntur
7004          NaN  Hospitalized    2020-10-04                  S.A.S. Nagar

        detectedstate gender nationality patientnumber statecode  \
7979       Tamil Nadu                             7980         TN
3290            Delhi                             3291         DL
9343      Maharashtra                             9344         MH
1414   Andhra Pradesh      M                      1415         AP
5226            Delhi                             5227         DL
969       Maharashtra                              970         MH
10431           Delhi                            10432         DL
7630          Gujarat                             7631         GJ
7918   Andhra Pradesh      M                      7919         AP
7004           Punjab                             7005         PB

      statepatientnumber statuschangedate typeoftransmission  \
7979             TN-P954       2020-11-04
3290                           2020-04-04
9343                           2020-04-13
1414              AP-P38       2020-03-31              Local
5226                           2020-07-04
969                            2020-03-28                TBD
10431                          2020-04-13
7630                           2020-11-04
7918             AP-P387       2020-11-04
7004                           2020-10-04

       durationOfAnyStatus
7979                   0.0
3290                   0.0
9343                   0.0
1414                   0.0
5226                   0.0
969                    0.0
10431                  0.0
7630                   0.0
7918                   0.0
7004                   0.0
```

[18]:

```
profile = pandas_profiling.ProfileReport(df_raw_data)
profile.to_file(output_file="covid19_data_after_preprocessing.html")
```

Observations

- Dataset info

Data	Info
Number of variables	14
Number of observations	8067
Missing cells	301 (0.3%)
Duplicate rows	0 (0.0%)
Total size in memory	882.4 KiB

- Variables types

Varibale	Count
Numeric	2
Categorical	12

- agebracket has a high cardinality: 86 distinct values
- detectedcity has a high cardinality: 314 distinct values
- detecteddistrict has a high cardinality: 349 distinct values
- durationOfAnyStatus has 7579 (94.0%) zeros
- durationOfAnyStatus has 301 (3.7%) missing values
- statepatientnumber has a high cardinality: 1463 distinct values
- currentstatus distribution

Value	Count	Frequency (%)
Hospitalized	7706	95.5%
Unknown	192	2.4%
Recovered	137	1.7%
Deceased	31	0.4%
Migrated	1	< 0.1%

- typeoftransmission distribution

Value	Count	Frequency (%)
Unknown	5233	64.9%
Local	1606	19.9%
TBD	630	7.8%
Imported	596	7.4%

[19]:

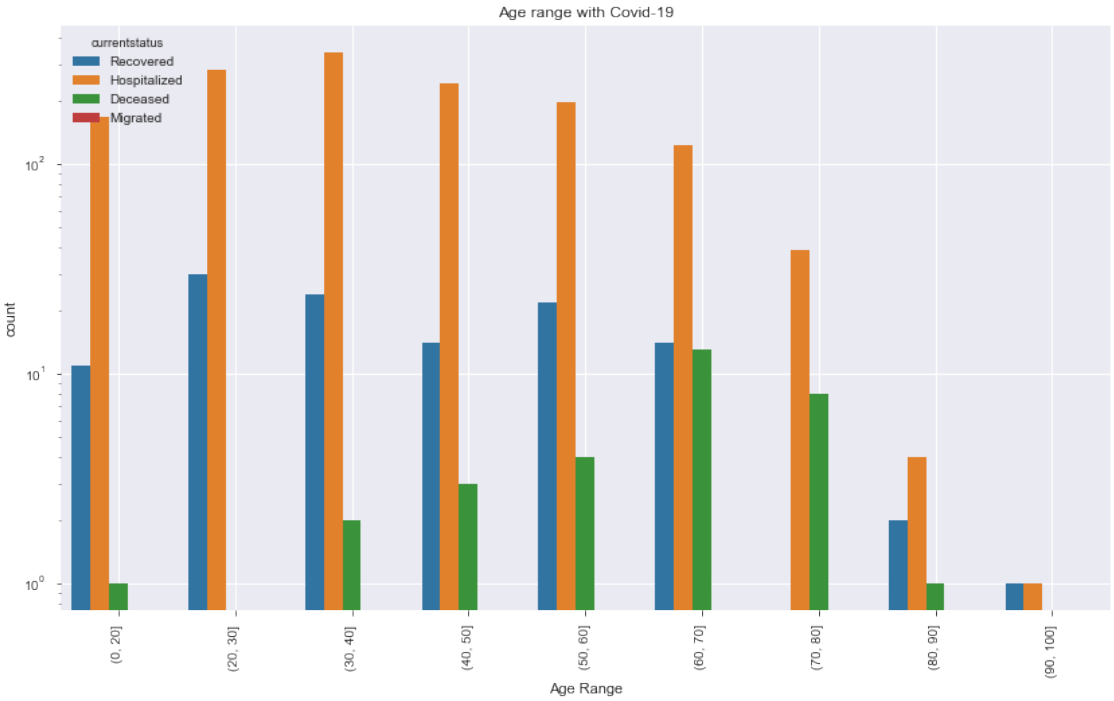
```
df_raw_data['agebracket'] = pd.to_numeric(df_raw_data['agebracket'],␣
 ↪errors='coerce')
```

2.1 Age range distribution with Covid-19

[20]:

```
age = df_raw_data['agebracket']
status = df_raw_data['currentstatus']
age_bins = [0,20,30,40,50,60,70,80,90,100]
plt.figure(figsize=(14,8))
```

```
sns.countplot(x=pd.cut(age, age_bins), hue=status)
plt.xticks(rotation=90)
plt.xlabel("Age Range")
plt.yscale('log')
plt.title("Age range with Covid-19")
plt.grid(True)
plt.show()
```



2.2 Covid-19 Cases Distribution across States

[21]:

```
state = df_raw_data.groupby('detectedstate').count()
fig = px.pie(state, values='currentstatus', names=state.index
             ,color_discrete_sequence=px.colors.sequential.
 ↪Plasma_r,title='Covid19 cases based on State')
fig.update_traces(textposition='outside', textinfo='value+label')
fig.show()
```

2.3 Covid-19 cases distribution based on Nationality

[22]:

```
nationality = df_raw_data.groupby('nationality').count()
fig = px.pie(nationality, values='currentstatus', names=nationality.index
             ,color_discrete_sequence=px.colors.qualitative.G10,title='Covid19␣
 ↪cases based on Nationality in India')
```

```
fig.update_traces(textposition='outside', textinfo='value+label')
fig.show()
```

2.4 No. of foreign citizens affected by Covid-19 in India

[23]:
```
temp = df_raw_data.groupby('nationality')['patientnumber'].count().reset_index()
temp = temp.sort_values('patientnumber')
temp = temp[temp['nationality']!='']
temp = temp[temp['nationality']!='India']
fig = px.bar(temp, x='patientnumber', y='nationality', orientation='h',␣
 ↪text='patientnumber', width=600,
       color_discrete_sequence = ['#35495e'], title='No. of foreign citizens')
fig.update_xaxes(title='')
fig.update_yaxes(title='')
fig.show()
```

2.5 Covid-19 distribution based on Type of Transmission

[24]:
```
temp = pd.DataFrame(df_raw_data[['typeoftransmission']].
 ↪groupby('typeoftransmission')['typeoftransmission'].count())
temp = temp.dropna()
temp.columns = ['count']
temp = temp.reset_index().sort_values(by='count')

fig = px.bar(temp, x='count', y='typeoftransmission', orientation='h',␣
 ↪text='count', width=600, height=300,
       color_discrete_sequence = ['#35495e'], title='Type of transmission')
fig.update_xaxes(title='')
fig.update_yaxes(title='')
fig.show()
```

2.6 Covid-19 cases Vs Age Brackets along with current status

[25]:
```
fig = plotly.subplots.make_subplots(
    rows=1, cols=2, column_widths=[0.8, 0.2],
    subplot_titles = ['Cases vs Age', ''],
    specs=[[{"type": "histogram"}, {"type": "pie"}]]
)

temp = df_raw_data[['agebracket', 'currentstatus']].dropna()
print('Total no. of values :', df_raw_data.shape[0], '\nNo. of missing values :
 ↪', df_raw_data.shape[0]-temp.shape[0], '\nNo. of available values :',␣
 ↪df_raw_data.shape[0]-(df_raw_data.shape[0]-temp.shape[0]))
gen_grp = temp.groupby('currentstatus').count()
```

```
fig.add_trace(go.Pie(values=gen_grp.values.reshape(-1).tolist(),␣
 ↪labels=['Deceased', 'Hospitalized', 'Recovered'],
                     marker_colors = ['#fd0054', '#393e46', '#40a798'], hole=.
 ↪3),1, 2)

fig.add_trace(go.
 ↪Histogram(x=temp[temp['currentstatus']=='Deceased']['agebracket'],␣
 ↪nbinsx=50, name='Deceased', marker_color='#fd0054'), 1, 1)
fig.add_trace(go.
 ↪Histogram(x=temp[temp['currentstatus']=='Recovered']['agebracket'],␣
 ↪nbinsx=50, name='Recovered', marker_color='#40a798'), 1, 1)
fig.add_trace(go.
 ↪Histogram(x=temp[temp['currentstatus']=='Hospitalized']['agebracket'],␣
 ↪nbinsx=50, name='Hospitalized', marker_color='#393e46'), 1, 1)

fig.update_layout(showlegend=False)
fig.update_layout(barmode='stack')
fig.data[0].textinfo = 'label+text+value+percent'

fig.show()
```

```
Total no. of values : 13060
No. of missing values : 11509
No. of available values : 1551
```

2.7 Covid-19 cases Gender Vs Age Brackets along with gender distribution

[26]:

```
fig = plotly.subplots.make_subplots(
    rows=1, cols=2, column_widths=[0.8, 0.2],
    subplot_titles = ['Gender vs Age', ''],
    specs=[[{"type": "histogram"}, {"type": "pie"}]]
)

temp = df_raw_data[['agebracket', 'gender']].dropna()
print('Total no. of values :', df_raw_data.shape[0], '\nNo. of missing values :
 ↪', df_raw_data.shape[0]-temp.shape[0], '\nNo. of available values :',␣
 ↪df_raw_data.shape[0]-(df_raw_data.shape[0]-temp.shape[0]))
gen_grp = temp.groupby('gender').count()

fig.add_trace(go.Histogram(x=temp[temp['gender']=='F']['agebracket'],␣
 ↪nbinsx=50, name='Female', marker_color='#6a0572'), 1, 1)
fig.add_trace(go.Histogram(x=temp[temp['gender']=='M']['agebracket'],␣
 ↪nbinsx=50, name='Male', marker_color='#39065a'), 1, 1)

fig.add_trace(go.Pie(values=gen_grp.values.reshape(-1).tolist(),␣
 ↪labels=['Female', 'Male'], marker_colors = ['#6a0572', '#39065a']),1, 2)
```

```
fig.update_layout(showlegend=False)
fig.update_layout(barmode='stack')
fig.data[2].textinfo = 'label+text+value+percent'

fig.show()
```

```
Total no. of values : 13060
No. of missing values : 11509
No. of available values : 1551
```

2.8 Covid-19 cases Age distribution of confirmed patients

[27]:

```
print('Total no. of values :', df_raw_data.shape[0], '\nNo. of missing values :
 ↪', df_raw_data.shape[0]-df_raw_data[['agebracket']].dropna().shape[0],
      '\nNo. of available values :', df_raw_data.shape[0]-(df_raw_data.
 ↪shape[0]-df_raw_data[['agebracket']].dropna().shape[0]))
px.histogram(df_raw_data, x='agebracket', color_discrete_sequence =␣
 ↪['#35495e'], nbins=50,
             title='Distribution of ages of confirmed patients')
```

```
Total no. of values : 13060
No. of missing values : 11509
No. of available values : 1551
```

2.9 Covid-19 cases distribution across states

[28]:

```
dist = df_raw_data.groupby(['detectedstate',␣
 ↪'detecteddistrict'])['patientnumber'].count().reset_index()
dist.head()
fig = px.treemap(dist, path=['detectedstate', 'detecteddistrict'],␣
 ↪values='patientnumber', height=700,
           title='Number of Confirmed Cases', color_discrete_sequence = px.
 ↪colors.qualitative.Prism)
fig.data[0].textinfo = 'label+text+value'
fig.show()
```

[29]:

```
df_raw_data['statuschangedate'] = pd.
 ↪to_datetime(df_raw_data['statuschangedate'])
df_raw_data['dateannounced'] = pd.to_datetime(df_raw_data['dateannounced'])
```

[30]:

```
temp = df_raw_data[['dateannounced', 'statuschangedate', 'currentstatus']].
 ↪dropna()
temp = temp[temp['statuschangedate']!=temp['dateannounced']]
temp['no_of_days'] = temp['statuschangedate'] - temp['dateannounced']
temp['no_of_days'] = temp['no_of_days'].dt.days
temp = temp[temp['no_of_days']>0]
```

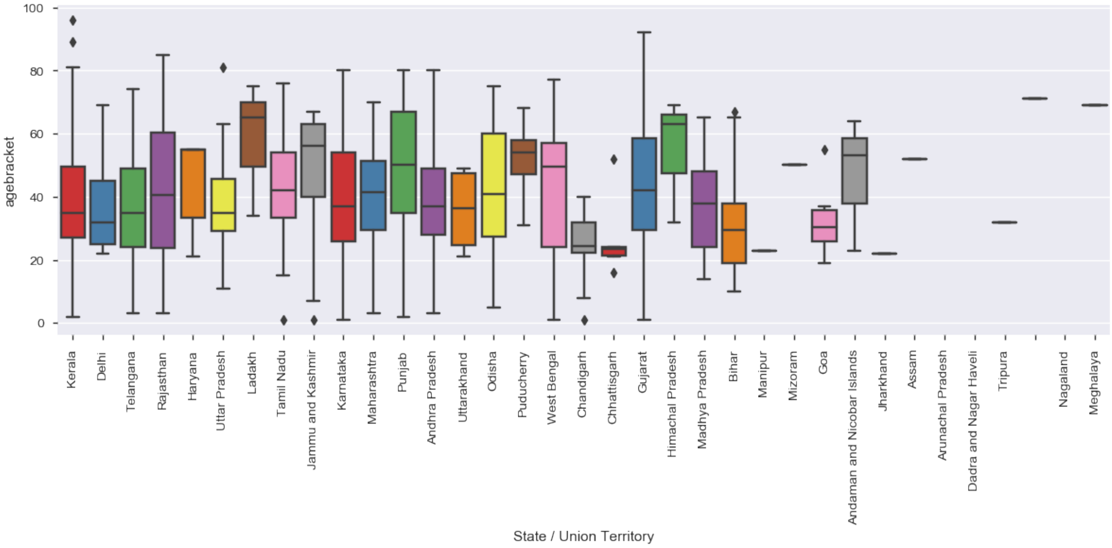
[31]:

```
print('Total no. of values :', df_raw_data.shape[0], '\nNo. of missing values :
↪', df_raw_data.shape[0]-temp.shape[0], '\nNo. of available values :',␣
↪df_raw_data.shape[0]-(df_raw_data.shape[0]-temp.shape[0]))
px.box(temp, x="currentstatus", y="dateannounced", color='currentstatus')
```

```
Total no. of values : 13060
No. of missing values : 12948
No. of available values : 112
```

[32]:

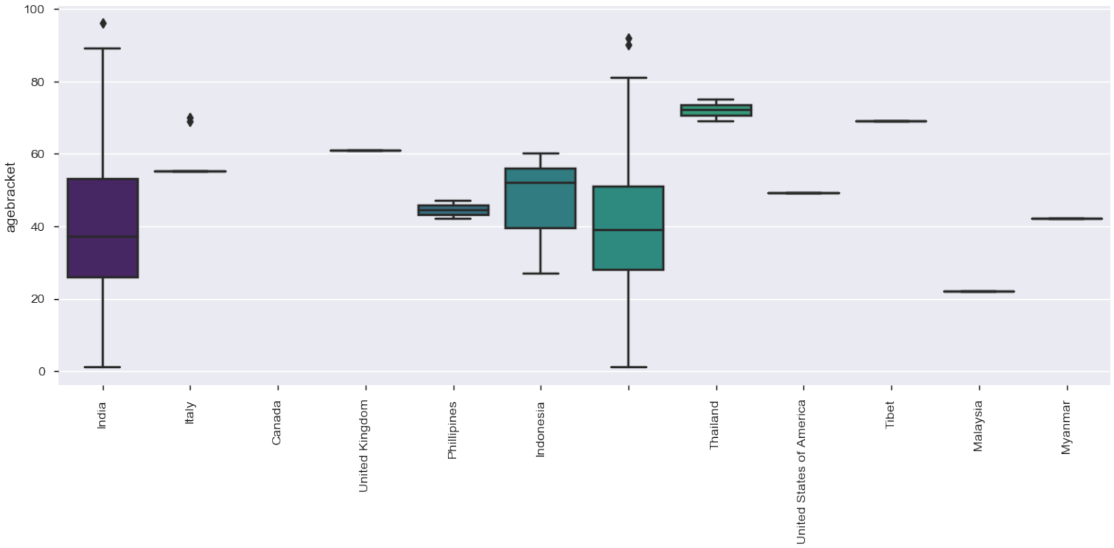
```
plt.figure(figsize=(12, 6), dpi = 100)
sns.boxplot(x = 'detectedstate', y = 'agebracket', data = df_raw_data, palette␣
↪= 'Set1')
plt.xlabel('State / Union Territory')
plt.ylabel('agebracket')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```



2.9.1 Nationality AgeBracket Distribution

[33]:

```
plt.figure(figsize=(12, 6), dpi = 100)
sns.boxplot(x = 'nationality', y = 'agebracket', data = df_raw_data, palette =␣
↪'viridis')
plt.xlabel('')
plt.xticks(rotation=90)
plt.ylabel('agebracket')
plt.tight_layout()
plt.show()
```



2.9.2 Age Distribution of COVID-19 Recovered Patients

[34]:

```
dist = df_raw_data.groupby(['agebracket','currentstatus'])['patientnumber'].
 ↪count().reset_index()
dist = dist[dist['currentstatus']=='Recovered']
dist
fig = px.bar(dist, x='agebracket', y='patientnumber', orientation='v',␣
 ↪text='patientnumber', width=1200,
       color_discrete_sequence = ['#00CC96'], title='Age distribution of␣
 ↪Recovered COVID Patient')

fig.update_xaxes(title='Age')
fig.update_yaxes(title='# Patient')
fig.show()
```

2.9.3 Gender Distribution of COVID-19 Recovered Patients

[35]:

```
dist = df_raw_data.groupby(['gender','currentstatus'])['patientnumber'].count().
 ↪reset_index()
dist = dist[dist['currentstatus']=='Recovered']
dist
fig = px.pie(dist, values=dist['patientnumber'], names=dist.gender
         ,color_discrete_sequence=["#636EFA"],title='Gender distribution of␣
 ↪COVID19 Recovered Patients')
fig.update_traces(textposition='outside', textinfo='value+label')
fig.show()
```

[36]:
```
df_raw_data.head()
```

[36]:
```
   agebracket currentstatus dateannounced              detectedcity  \
0        20.0     Recovered    2020-01-30                  Thrissur
1         NaN     Recovered    2020-02-02                 Alappuzha
2         NaN     Recovered    2020-03-02                 Kasaragod
3        45.0     Recovered    2020-02-03  East Delhi (Mayur Vihar)
4        24.0     Recovered    2020-02-03                 Hyderabad

  detecteddistrict detectedstate gender nationality patientnumber statecode  \
0         Thrissur        Kerala      F       India             1        KL
1        Alappuzha        Kerala              India             2        KL
2        Kasaragod        Kerala              India             3        KL
3       East Delhi         Delhi      M       India             4        DL
4        Hyderabad     Telangana      M       India             5        TG

  statepatientnumber statuschangedate typeoftransmission  durationOfAnyStatus
0           KL-TS-P1       2020-02-14           Imported                 15.0
1           KL-AL-P1       2020-02-14           Imported                 12.0
2           KL-KS-P1       2020-02-14           Imported                -17.0
3              DL-P1       2020-03-15           Imported                 41.0
4              TS-P1       2020-02-03           Imported                  0.0
```

[ ]:
```

```