In [1]:
```
%matplotlib inline
import pandas as pd
from pandas.api.types import is_numeric_dtype
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.metrics import roc_curve, roc_auc_score, accuracy_score, precision_score, recall_score, classification_report
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler, LabelEncoder

# Metrics
from sklearn.metrics import precision_score, recall_score, log_loss, accuracy_score, f1_score, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objs as go
import plotly.figure_factory as ff
from plotly.subplots import make_subplots
# import calmap
#import folium

# color pallette
cnf, dth, rec, act = '#393e46', '#ff2e63', '#21bf73', '#fe9801'

# converter
#from pandas.plotting import register_matplotlib_converters
#register_matplotlib_converters()

# hide warnings
import warnings
warnings.filterwarnings('ignore')

from datetime import timedelta
```

```
Bad key "text.kerning_factor" on line 4 in
C:\Users\cyine\Anaconda3\lib\site-packages\matplotlib\mpl-data\stylelib\_classic_test_patch.mplstyle.
You probably need to get an updated matplotlibrc file from
http://github.com/matplotlib/matplotlib/blob/master/matplotlibrc.template
or from the matplotlib source distribution
```

In [52]:
```
# Offline mode
from plotly.offline import init_notebook_mode, iplot
init_notebook_mode(connected=True)
```

In [2]:
```
#Read in data
path='https://raw.githubusercontent.com/charlenelau-github/COVID-19-Analysis/master/preprocessed_COVID19.csv'
df = pd.read_csv(path, parse_dates=['Date'])
df.head(n=10)
```

Out[2]:

	Unnamed: 0	Province/State	Country/Region	Date	Confirmed_Cases	Recoveries	Deaths	Active_Cases
0	0	NaN	Afghanistan	2020-01-22	0	0.0	0	0.0
1	1	NaN	Albania	2020-01-22	0	0.0	0	0.0
2	2	NaN	Algeria	2020-01-22	0	0.0	0	0.0
3	3	NaN	Andorra	2020-01-22	0	0.0	0	0.0
4	4	NaN	Angola	2020-01-22	0	0.0	0	0.0
5	5	NaN	Antigua and Barbuda	2020-01-22	0	0.0	0	0.0
6	6	NaN	Argentina	2020-01-22	0	0.0	0	0.0
7	7	NaN	Armenia	2020-01-22	0	0.0	0	0.0
8	8	Australian Capital Territory	Australia	2020-01-22	0	0.0	0	0.0
9	9	New South Wales	Australia	2020-01-22	0	0.0	0	0.0

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28462 entries, 0 to 28461
Data columns (total 8 columns):
Unnamed: 0         28462 non-null int64
Province/State     8774 non-null object
Country/Region     28462 non-null object
Date               28462 non-null datetime64[ns]
Confirmed_Cases    28462 non-null int64
Recoveries         26857 non-null float64
Deaths             28462 non-null int64
Active_Cases       26857 non-null float64
dtypes: datetime64[ns](1), float64(2), int64(3), object(2)
memory usage: 1.7+ MB
```

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
Unnamed: 0             0
Province/State     19688
Country/Region         0
Date                   0
Confirmed_Cases        0
Recoveries          1605
Deaths                 0
Active_Cases        1605
dtype: int64
```

In [5]:
```
#Keep rows with recoveries not null
df=df.dropna(subset=['Recoveries'])
```

In [6]:
```
#Change all counts to integers
cols = ['Recoveries', 'Active_Cases']
df[cols] = df[cols].applymap(np.int64)
df.head()
```

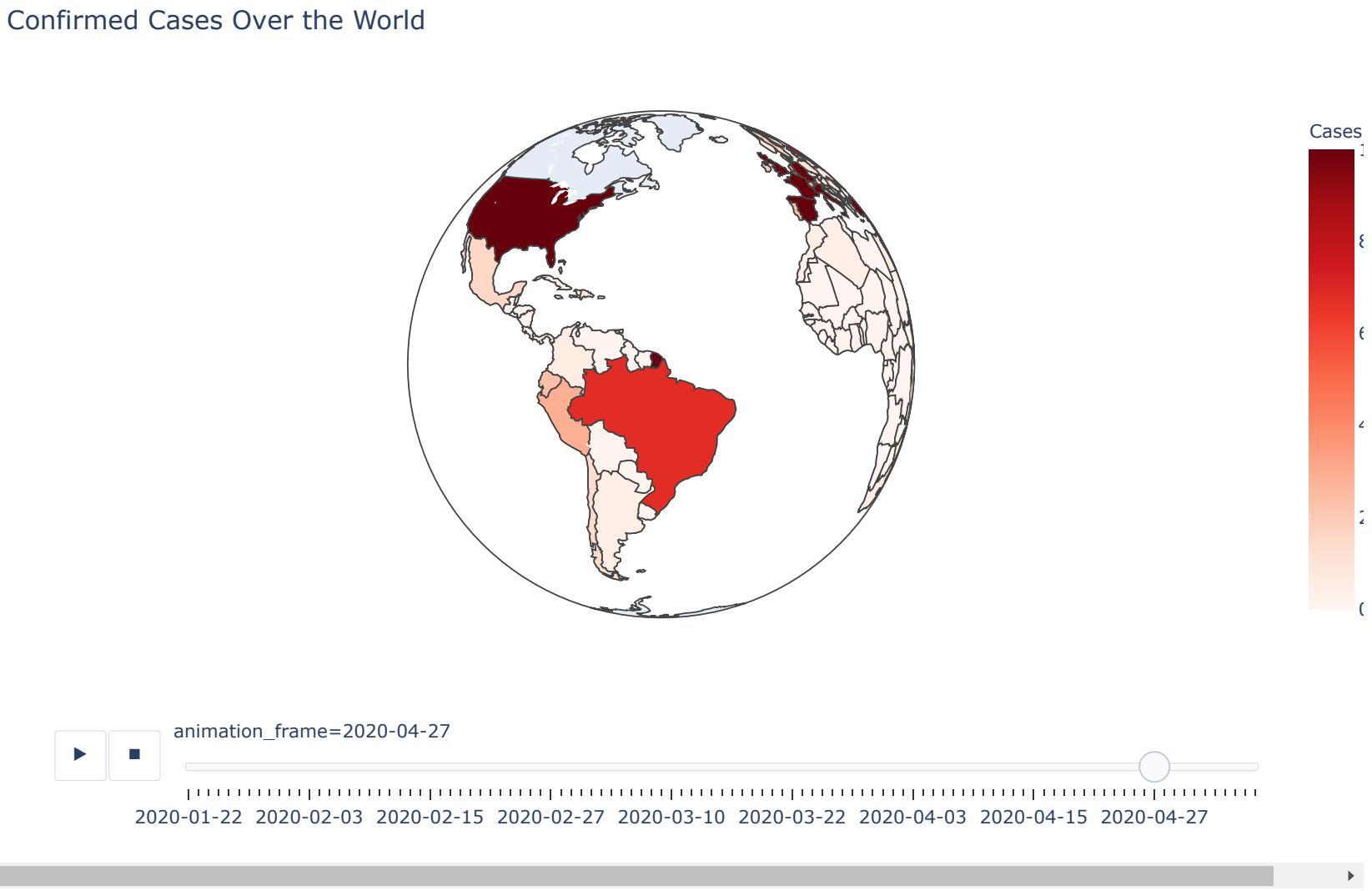
Out[6]:

	Unnamed: 0	Province/State	Country/Region	Date	Confirmed_Cases	Recoveries	Deaths	Active_Cases
0	0	NaN	Afghanistan	2020-01-22	0	0	0	0
1	1	NaN	Albania	2020-01-22	0	0	0	0
2	2	NaN	Algeria	2020-01-22	0	0	0	0
3	3	NaN	Andorra	2020-01-22	0	0	0	0
4	4	NaN	Angola	2020-01-22	0	0	0	0

In [11]:
```
df_grouped = df.groupby(['Date', 'Country/Region'])['Confirmed_Cases', 'Deaths', 'Recoveries', 'Active_Cases'].sum().reset_
index()
```

Chloropleth Map with Plotly

In [50]:
```
fig = px.choropleth(df_grouped,
                    locations="Country/Region",
                    locationmode='country names',
                    color="Confirmed_Cases",
                    color_continuous_scale="Reds",
                    animation_frame=df_grouped["Date"].dt.strftime('%Y-%m-%d'),
                    title='Confirmed Cases Over the World',
                    labels={'Confirmed_Cases':'Cases'},
                    range_color=(0, 100000),
                    width=1000,
                    height=600
                    )
fig.update(layout_coloraxis_showscale=True)
fig.update_geos(projection_type="orthographic")
fig.show()
```



In [ ]:
```
#jupyter notebook --NotebookApp.iopub_data_rate_limit=1.0e10
```

In [44]:
```
fig = px.choropleth(df_grouped,
                    locations="Country/Region",
                    locationmode='country names',
                    color="Deaths",
                    color_continuous_scale="Greens",
                    hover_name="Date",
                    animation_frame=df_grouped["Date"].dt.strftime('%Y-%m-%d'),
                    title='Global Deaths',
                    labels={'Deaths':'Cases'},
                    range_color=(0, 100000),
                    width=1000,
                    height=800
                    )
fig.update(layout_coloraxis_showscale=True)
fig.show()
```

Global Deaths