Assignment -5
Topic: COVID-19

Group8 Team Members:

Name	Matriculation Number
Harshith Srinivas	6896703
Nihal Yadalam Murali Kumar	6873339
Sai Nikhil Menon	6882524
Siddhanth Janadri	6882502
Sneha Hiremath	6881479

Goal:

We are aiming to create an Interactive Data Visualization for COVID-19 data, from January 2020 to June 2020.

Introduction:

An outbreak of the coronavirus disease (COVID-19) caused by the 2019 novel coronavirus (SARS-CoV-2) has been spreading rapidly across the world since January 2020, following the diagnosis of the initial cases in Wuhan, Hubei Province, China. Through this project we are trying to present a small visualization of COVID-19 dataset, we have created an interactive dash board for World and Europe. Data provider: "Our World in Data" (OWID) dataset.

Software Specification:

Dependency	*Version*
Python	3.7.7
Plotly	4.7.1
Dash	1.4.1
Pandas	1.0.3

Division of Responsibility:

        Nihal Yadalam Murali Kumar: Data Gathering and Data pre-processing

        Sai Nikhil Menon: Data Pre-processing and Data Manipulation

        Siddhanth Janadri: Interactivity using Dash

        Sneha Hiremath: Geo-Spatial Visualisation of Choropleth Map

        Harshith Srinivas: Visualisation using line-graph for cumulative data

Exploratory Data Analysis:

From the above-mentioned dataset we have selected following features as mentioned in Concept Paper.

Features	
Numerical Data	*Categorical Data*
Date	Iso_code
new_deaths	Continent
new_cases	Location
new_tests	
hospital_beds_per_thousand	

    i)   Libraries Imported:

```
#importing libraries
import pandas as pd
import plotly.express as px
import dash                              .
import dash_core_components as dcc
import dash_html_components as html
from dash.dependencies import Input, Output
```

    ii)   Importing Dataset:

```
# reading csv file into data frame         .
df = pd.read_csv('./owid-covid-data.csv')
```

iii)   Identification of feature Data Type used:

```
#creating a new dataframe with required dimensions

df1 = df[['iso_code','continent','location','date','new_deaths','new_cases','new_tests','hospital_beds_per_thousand']].copy()

# Printing the datatypes of the dimensions extracted from the dataset file before pre-processing

print(df1.dtypes)
```

```
iso_code                    object
continent                   object
location                    object
date                        object
new_deaths                  object
new_cases                   object
new_tests                   object
hospital_beds_per_thousand  object
dtype: object
```

iv)   Size of dataset:

```
# Dataset shape: Number of rows extracted from the dataset file before pre-processing

print(df1.shape)
```

```
(26749, 8)
```

v)   Our dataset has only Non-NULL variables.

```
# Showing that there are no null/junk values

df1.apply(lambda x:sum(x.isnull()),axis=0)
```

```
iso_code                    0
continent                   0
location                    0
date                        0
new_deaths                  0
new_cases                   0
new_tests                   0
hospital_beds_per_thousand  0
dtype: int64
```

Data Pre-processing:

      We found that we need additional fields for visualization like month number and name for slider. Also, cumulative sum of deaths, tests and affected cases used in dropdown. We have created these new features based on original dataset (OWID).

Dataset after pre-processing:

```
#Dataset Shape after pre-processing

print(df1.dtypes)
```

```
iso_code                     object
continent                    object
location                     object
month_n                      object
month_num                     int64
new_deaths                    int64
new_cases                     int64
new_tests                     int64
hospital_beds_per_thousand  float64
date                         object
cumil_deaths                  int32
cumil_cases                   int32
cumil_tests                   int32
dtype: object
```

User and Task:

• User: Officials working in various government sectors around the world use this kind of visualization for COVID-19 datasets to make strategic decisions. Hospital officials can observe the trends from the visualization and make decisions regarding resource planning for the future. Common citizens use this visualization for general awareness purpose. Organizations like "European Centre for Disease Prevention and Control" (ECDPC) will be benefitted with better monitoring systems in order to effectively prevent and control the spread of the virus, as well as contribute to the assessment of social dynamics.

• Task: Deriving knowledge and observing trends about the number of cases, number of deaths, number of tests carried out and hospital beds, arising from COVID –19, for different countries in the world over period of few months, starting January 2020.

Visualization Techniques:

- We have used a geospatial visualization technique to visualize a geospatial data which describes data with respect to different locations in the real world.
- World map in the form of Equirectangular cylindrical projection is plotted using plotly library available in python. Chloropleth maps are visualized on the dashboard (dash library similar to html provided in python) which provides details regarding the new_cases related to COVID 19 for a period of 6 months (i.e., from January to June)
- Color bar is provided at the right end of the dashboard for the user to visualize the associated numerical values with colors.
- Libraries used in visualization techniques in python are plotly and dash.

Interactions:

- Various interactive elements such as dropdown and slider are provided to the user.
- Dropdown is used to filter data on the world map based on the value of the dropdown, provided by the user. Values available in the dropdown are deaths, new_cases and tests. Dropdown is also used to select the scope of geography such as World and Europe.
- Slider is provided to filter data based on the specific month provided by the user. The available months on the slider are from January to June.
- Radio button is used to delineate map with cumulative or non-cumulative data with respect to deaths, affected cases and tests.
- Combining the values of the slider, dropdowns and radio button given by the user, world map is visualized in the dashboard based on COVID 19 world data.
- Data with respect to a specific country such as location, iso code and hospital beds irrespective of selected data along with data of interest selected by the user is displayed on the specific hovered region provided by the user.
- Library used in python to create Interactive elements i.e., dropdown and slider is Dash.

Conclusion and Changes made with respect to Concept-paper:

Yes, we have met the goals defined in concept paper. We had mentioned in concept paper that we will be creating a interactive dash board for World. On implementation after sending you concept paper, we decided to compare the same features with respect to European Countries. So, by doing this we can have additional Users like "European Centre for Disease Prevention and Control" (ECDPC) which helps them to track daily COVID-19 issues in Europe and can help the governing bodies to make decisions based on this.