A Project Report on

Covid Data Analysis

Course Code: BCA-EC401

Course Name: Data Science using python

Submitted by:

Aman Singh (41222121)

Year & Semester: 2nd Year, 4th Semester

Session: 2022-25

Submitted to:

Mrs. Bushra Jamal

Dwarka sec-9,New Delhi

(Bachelor of Computer Application)

Certificate

This project is for the department of Bachelor of computer Application, under faculty of computer science and engineering at **Delhi skill and Entrepreneurship University**, the project entitled **Covid data analysis** in satisfactory manner as a partial fulfillment of required degree of bachelor of computer application for the academic year 2024 ,of DSEU Dwarka Campus

This is to certify that **Aman Singh(41222121)**has contributed in successfully completing the project

Acknowledgement

We would like to thank **Mrs. Bushra Jamal**, our Professor-in-charge for their support and guidance in completing our project on the topic **Covid Data Analysis.** It was a great learning experience.

We are also thankful to our parents and friend for their constant encouragement and cooperation throught out this project. Without the contribution of group members the project have not been completed so, thankful to each of group members who fairly contributed to the project.

Aman Singh (4122121)

Abstract of the Project

The COVID-19 pandemic has spurred a global effort to collect and analyze vast amounts of data. This project delves into this data to glean crucial insights into the virus's spread, treatment, and prevention.

We aim to utilize techniques like statistical analysis, data visualization , and dashboard to:

- Identify trends and patterns in confirmed cases, fatalities, and recovery rates across various regions.
- Explore correlations between COVID-19 and factors like demographics, population density, and implemented public health interventions.

By extracting knowledge from this data, this project aspires to empower policymakers, healthcare professionals, and the general public with the information needed to navigate the challenges of COVID-19 and work towards mitigating its impact.

- Depending on the specific focus of your project, you can modify the abstract to highlight additional aspects, such as:-
- Analyzing the efficacy of various vaccines or treatment protocols.
- Investigating the emergence and spread of new COVID-19 variants.
- Assessing the socioeconomic impact of the pandemic.

S.NO	Index	

1.	Introduction
	    a. About the project idea
	    b. About tools in use
	    c. About data science	

2.	Purpose/scope/Objective of the analysis	

3.	Requirements for the project
	    a. Hardware used
	    b. Software used
	    c. Technology used	

4.	Methodology
	    a. introduction
	    b. general description
	    c. specific requirements, functions and formulas
	    d. analysis results
	    e. Visualization	

5.	Coding and Output	
6.	Visualization	
7.	Future scope of project	
8.	Conclusion	
9.	Bibliography	

1. Introduction

a. About the project idea:

The COVID-19 pandemic has left an undeniable mark on the world. As we navigate a complex new normal, data analysis offers a powerful tool to understand the past, assess the present, and potentially predict future trends. This project delves into the realm of COVID-19 data analysis, aiming to extract valuable insights that can inform public health strategies, resource allocation, and our overall response to the virus.

b. About tools in use:

The project utilizes a combination of tools and technologies to achieve its objectives. This includes programming languages such as Python for data preprocessing, analysis, and dashboard development. We will use power bi and multiple python libraries like Pandas, NumPy, Matplotlib ,Seaborn and plotly for data manipulation, analysis, and visualization. The use of open-source tools provides us better flexibility, scalability, and customization,

Resources Required:

Covid order dataset

Python programming language

Jupyter Notebook or similar IDE

Power bi

Access to computational resources for model training (e.g., CPU/GPU)

Libraries: Pandas, NumPy, Matplotlib, Seaborn,Plotly

c. About data science:

Data science, along with statistical analysis, computer science, and computational biology, has been driving critical applications in epidemiology, drug discovery, and molecular design for COVID-19

Data-driven, mathematical, and predictive models have provided valuable insights into the spread of COVID-19, risk factors, and strategies for living with the endemic virus

2. Purpose/Scope/Objective of the Analysis

Purpose: The COVID-19 pandemic has spurred a global effort to collect and analyze vast amounts of data. This project delves into this data to glean crucial insights into the virus's spread, treatment, and prevention.

Scope: The scope of the report will vary depending on the specific objectives and resources available. Here are some factors to consider:

Data sources: Specify the data sources used in the analysis, such as World Health Organization (WHO) reports, national health agencies, or local case data.

Geographic focus: Define the geographic area covered in the report, whether it's global, regional, national, or local.

Timeframe: Specify the timeframe for the data analysis.

Data analysis methods: Outline the statistical methods used to analyze the data, such as time series analysis, spatial analysis, or regression modelling

Objectives: The objective of a COVID-19 data analysis report is to extract meaningful insights from the vast amount of data collected during the pandemic. This can involve understanding the spread of the virus, identifying risk factors, evaluating the effectiveness of interventions, and informing public health decisions.

3. Requirements for the Project

a. **Hardware used:**. While the hardware requirements are not overly demanding, a computer with sufficient processing power and RAM (depending on dataset size) will ensure smooth workflow.

Processor: Intel core i5

RAM: 12 GB

Storage : 512 GB SSD

b. **Software used:**

• **Python Programming Language:** Python serves as the primary programming language for data preprocessing, analysis, and model development.

• **Integrated Development Environment (IDE):** IDEs such as Jupyter Notebook or PyCharm are commonly used for coding and experimentation.

• **Data Science Libraries:** Libraries such as Pandas, NumPy, matplotlib, seaborn, and plotly are essential for data manipulation, data visualization and learning.

• **Visualization Libraries:** Matplotlib ,Plotly and Seaborn are used for data visualization to gain insights into the data and model performance.

• **Visualization tool :** Microsoft Power BI is a business intelligence tool that allows you to visualize and analyze data.

c. **Technology used:** Access to the restaurant's historical data (ideally in a well-structured format like CSV).Data science methodologies, including data preprocessing, exploratory data analysis (EDA), statistical analysis.

4. Methodology

a. Introduction:

Brief introduction to the methodology employed in the project.Overview of the steps involved in analyzing the covid order dataset and its representation. Explanation of the rationale behind the chosen approach and its relevance to achieving the project objectives.The methodology section outlines the step-by-step approach to analyzing the covid data and extracting valuable insights.

b. General Description:

The general workflow will encompass data acquisition, cleaning (handling missing values, inconsistencies), exploration (understanding overall trends and patterns), feature engineering (creating new informative features), data visualization (communicating insights through charts and graphs), and potential model building (predicting future trends, customer preferences) if deemed necessary within the project's scope.

c. Specific Requirements, Functions, and Formulas:

Data Acquisition: Methods will vary depending on data source; likely involving CSV file import or database interaction with Python libraries.

Data Cleaning: Techniques like handling missing values with techniques like mean/median imputation, removing outliers, and identifying data inconsistencies using statistical methods and visualizations.

Data Exploration: Descriptive statistics, data visualization libraries for uncovering patterns and trends, understanding customer demographics if available, and analyzing menu performance.

Data Visualization (throughout the process): Create compelling visualizations like bar charts, scatter plots, and time series plots to communicate insights effectively.

Dashboard using power bi.

d. **Analysis Results:**

- Analyze the trend of cases, recoveries, and deaths over time globally, or for specific regions/countries.
- Visualize the geographical distribution of cases, deaths, and recoveries on a map to identify hotspots and trends across different regions.
- Calculate and compare the mortality rate (deaths/cases) across different regions or demographics.

5.Coding and Output:

 # data cleaning

import pandas as pd

df = pd.read_csv('C:/Users/AMAN/Desktop/covid.csv')

df

df.columns

df.shape

df.count()

df.isnull()

df.isnull().sum()

a = df['New cases'].median()

print("median of New caese column - ",a)

b = df['New deaths'].median()

print("median of New deaths column - ",b)

c = df['New recovered'].median()

print("median of New recovered column - ",c)

df["New cases"].fillna(a, inplace = True)

df["New deaths"].fillna(b, inplace = True)

df["New recovered"].fillna(c, inplace = True)

df.isnull()

df.duplicated()

df.drop_duplicates(inplace = True)

print(df.duplicated())

df

 # insights from data

```python
df.describe()

df.info()

total = df['Confirmed'].sum()

print("Total number of covid confirmed cases in the world - ",total)

death = df['Deaths'].sum()

print("Total number of death due to covid in the world",death)

recovered = df['Recovered'].sum()

print("Total number of recovered cases from covid in the world",recovered)

active = df['Active'].sum()

print("Total number of Active cases in the world",death)

df.sum()

df['Active/Deaths Ratio']=(df['Active']/df['Deaths'])

df

df['Confirmed/Recovered Ratio']=(df['Confirmed']/df['Recovered'])

df

T1 = df.sort_values(by=['Confirmed'], ascending=False).head(10)

T1

T = df.sort_values(by=['Confirmed'])

T.head(10)

D = df.sort_values(by=['Deaths'], ascending=False).head(10)

D

T = df.sort_values(by=['Deaths'])

T.head(10)

R = df.sort_values(by=['Recovered'], ascending=False).head(10)

R

R = df.sort_values(by=['Recovered']).head(10)
```

 # Visualization

```python
import matplotlib.pyplot as plt

import seaborn as sns

g = df.groupby('WHO Region').sum()

explode = [0.03, 0.02, 0.02, 0.02, 0.1,0.03]

plt.figure(figsize=(7,7))

color=['lightsteelblue','aquamarine','orange','lightcoral','tan','plum']

plt.pie(g['Confirmed'], labels=['Africa','Americas','Eastern Mediterranean','Europe','South-East
Asia','Western Pacific']

      , colors=color, autopct='%.0f%%', explode=explode,shadow = True)

plt.title('percentage of Confirmed cases in each region',size=20,color='dimgrey')

plt.legend(loc='upper right')

plt.show()

import matplotlib.pyplot as plt

import seaborn as sns

explode = [0.03, 0.02, 0.02, 0.02, 0.1,0.03]

plt.figure(figsize=(7,7))

color=['lightsteelblue','aquamarine','orange','lightcoral','tan','plum']

plt.pie(g['Deaths'], labels=['Africa','Americas','Eastern Mediterranean','Europe','South-East
Asia','Western Pacific']

      , colors=color, autopct='%.0f%%', explode=explode,shadow = True)

plt.title('percentage of Deaths in each region',size=20,color='dimgrey')

plt.legend(loc='upper right')

plt.show()

x=df['Confirmed'].sum()

y=df['Active'].sum()

z=df['Recovered'].sum()

p=df['Deaths'].sum()

print(x,y,z,p)
```

```
import plotly.express as px

fig = px.pie(df, values=[x,y,z,p], names=['Total Confirmed cases','Active','Rcovered
cases','Deaths'], title='Covid cases Distribution')

fig.update_layout(width=700, height=550)

fig.show()

import plotly.express as px

fig = px.bar(T1, x='Country/Region', y='Confirmed'

        ,title='Top 10 country in the world with Confirmed cases',color='Confirmed')

fig.update_layout(height=600)

fig.show()

import plotly.express as px

fig = px.bar(D, x='Country/Region', y='Deaths'

        ,title='Top 10 country in the world with Deaths cases',color='Country/Region')

fig.update_layout(height=600)

fig.show()

import plotly.express as px

fig = px.bar(R, x='Country/Region', y='Recovered'

        ,title='Top 10 country in the world with Rcovered cases',color='Recovered')

fig.update_layout(height=600)

fig.show()

W = df.sort_values(by=['Confirmed'], ascending=False).head(20)

z= df.sort_values(by=['Deaths / 100 Cases'], ascending=False).head(20)

x = df.sort_values(by=['Recovered / 100 Cases'], ascending=False).head(20)

s = df.sort_values(by=['Deaths / 100 Recovered'], ascending=False).head(20)

import plotly.express as px

fig=px.line(W, x='Country/Region', y='1 week change',markers=True

        ,title='Last one week change in Top 20 Country')

fig.update_layout(height=600)

fig.show()
```

```python
import plotly.express as px
fig=px.line(z, x='Country/Region', y='Deaths / 100 Cases',markers=True
        ,title='Deaths / 100 Cases of Top 20 Country')
fig.update_layout(height=600)
fig.show()
import plotly.express as px
fig=px.line(x, x='Country/Region', y='Recovered / 100 Cases',markers=True
        ,title='Recovered / 100 Cases of Top 20 Country')
fig.update_layout(height=600)
fig.show()
import plotly.express as px
fig=px.line(s, x='Country/Region', y='Deaths / 100 Recovered',markers=True
        ,title='Deaths / 100 Recovered of Top 20 Country')
fig.update_layout(height=600)
fig.show()
import plotly.express as px
fig = px.choropleth(df, locations='Country/Region',locationmode='country names',color='Active'
            ,hover_name='Active', title='Active cases in each Country ')
fig.update_layout(geo=dict(projection_type='natural earth'))
fig.update_layout(height=650)
fig.show()
fig = px.choropleth(df, locations='Country/Region',locationmode='country
names',color='Country/Region'
            ,hover_name='New deaths', title='New deaths in each Country ')
fig.update_layout(geo=dict(projection_type='natural earth'))
fig.update_layout(height=650)
fig.show()
import plotly.express as px
```

fig = px.choropleth(df, locations='Country/Region',locationmode='country names',color='New cases'

        ,hover_name='New cases', title='New cases in each Country ')

fig.update_layout(geo=dict(projection_type='natural earth'))

fig.update_layout(height=650)

fig.show()

6. Visualization

▼ Importing data

```
[ ]:
```

```
[3]: import pandas as pd
     df = pd.read_csv('C:/Users/AMAN/Desktop/covid.csv')
     df
```

[3]:

	Country/ Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
0	Afghanistan	36263	1269	25198	9796	106.0	10.0	18.0	3.50	69.49	5.04	35526	737.0	2.07	Eastern Mediterranean
1	Albania	4880	144	2745	1991	117.0	6.0	63.0	2.95	56.25	5.25	4171	709.0	17.00	Europe
2	Algeria	27973	1163	18837	7973	616.0	8.0	749.0	4.16	67.34	6.17	23691	4282.0	18.07	Africa
3	Andorra	907	52	803	52	10.0	NaN	NaN	5.73	88.53	6.48	884	23.0	2.60	Europe
4	Angola	950	41	242	667	18.0	1.0	0.0	4.32	25.47	16.94	749	201.0	26.84	Africa
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
185	Zambia	4552	140	2815	1597	71.0	1.0	465.0	3.08	61.84	4.97	3326	1226.0	36.86	Africa
186	Zimbabwe	2704	36	542	2126	192.0	2.0	24.0	1.33	20.04	6.64	1713	991.0	57.85	Africa
187	Uzbekistan	21209	121	11674	9414	678.0	5.0	569.0	0.57	55.04	1.04	17149	4060.0	23.67	Europe
188	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa
189	Namibia	1843	8	101	1734	68.0	0.0	26.0	0.43	5.48	7.92	1344	499.0	37.13	Africa

190 rows × 15 columns

---

shape, count and columns of data

```
[ ]:
```

```
[5]: df.columns
```

```
[5]: Index(['Country/Region', 'Confirmed', 'Deaths', 'Recovered', 'Active',
            'New cases', 'New deaths', 'New recovered', 'Deaths / 100 Cases',
            'Recovered / 100 Cases', 'Deaths / 100 Recovered',
            'Confirmed last week', '1 week change', '1 week % increase',
            'WHO Region'],
           dtype='object')
```

```
[6]: df.shape
```

```
[6]: (190, 15)
```

```
[7]: df.count()
```

```
[7]: Country/Region           190
     Confirmed                190
     Deaths                   190
     Recovered                190
     Active                   190
     New cases                188
     New deaths               181
     New recovered            187
     Deaths / 100 Cases       190
     Recovered / 100 Cases    190
     Deaths / 100 Recovered   190
     Confirmed last week      190
     1 week change            189
     1 week % increase        190
     WHO Region               190
     dtype: int64
```

checking Null values

`[ ]:`

`[22]:` `df.isnull()`

`[22]:`

	Country/ Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Active/ Deaths Ratio	Confirme Recover Rat
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
185	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
186	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
187	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
188	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
189	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal

190 rows × 17 columns

```
[13]:   0       False
        1       False
        2       False
        3       False
        4       False
                ...
        185     False
        186     False
        187     True
        188     True
        189     True
        Length: 190, dtype: bool
```

```
        0       False
        1       False
        2       False
        3       False
        4       False
                ...
        182     False
        183     False
        184     False
        185     False
        186     False
        Length: 187, dtype: bool
```

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
0	Afghanistan	36263	1269	25198	9796	106.0	10.0	18.0	3.50	69.49	5.04	35526	737.0	2.07	Eastern Mediterranean
1	Albania	4880	144	2745	1991	117.0	6.0	63.0	2.95	56.25	5.25	4171	709.0	17.00	Europe
2	Algeria	27973	1163	18837	7973	616.0	8.0	749.0	4.16	67.34	6.17	23691	4282.0	18.07	Africa
3	Andorra	907	52	803	52	10.0	1.0	24.0	5.73	88.53	6.48	884	23.0	2.60	Europe
4	Angola	950	41	242	667	18.0	1.0	0.0	4.32	25.47	16.94	749	201.0	26.84	Africa
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
185	Zambia	4552	140	2815	1597	71.0	1.0	465.0	3.08	61.84	4.97	3326	1226.0	36.86	Africa
186	Zimbabwe	2704	36	542	2126	192.0	2.0	24.0	1.33	20.04	6.64	1713	991.0	57.85	Africa
187	Uzbekistan	21209	121	11674	9414	678.0	5.0	569.0	0.57	55.04	1.04	17149	4060.0	23.67	Europe
188	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa
189	Namibia	1843	8	101	1734	68.0	0.0	26.0	0.43	5.48	7.92	1344	499.0	37.13	Africa

190 rows × 15 columns

Insights from Data

[ ]:

information and description of data

[17]: df.describe()
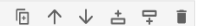
[17]:

	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 wee chang
count	1.870000e+02	187.000000	1.870000e+02	1.870000e+02	187.000000	187.000000	187.000000	187.000000	187.000000	187.00	1.870000e+02	186.000000
mean	8.813094e+04	3497.518717	5.063148e+04	3.400194e+04	1223.550802	29.005348	934.197861	3.019519	64.820535	inf	7.868248e+04	9499.25806
std	3.833187e+05	14100.002482	1.901882e+05	2.133262e+05	5710.249850	120.025692	4197.634610	3.454302	26.287694	NaN	3.382737e+05	47614.21474
min	1.000000e+01	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	1.000000e+01	-47.00000
25%	1.114000e+03	18.500000	6.265000e+02	1.415000e+02	5.000000	0.000000	0.000000	0.945000	48.770000	1.45	1.051500e+03	51.25000
50%	5.059000e+03	108.000000	2.815000e+03	1.600000e+03	55.500000	1.000000	24.000000	2.150000	71.320000	3.62	5.020000e+03	448.50000
75%	4.046050e+04	734.000000	2.260600e+04	9.149000e+03	419.500000	6.000000	221.000000	3.875000	86.885000	6.44	3.708050e+04	3248.50000
max	4.290259e+06	148011.000000	1.846641e+06	2.816444e+06	56336.000000	1076.000000	33728.000000	28.560000	100.000000	inf	3.834677e+06	455582.00000

```
<class 'pandas.core.frame.DataFrame'>
Index: 187 entries, 0 to 186
Data columns (total 15 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Country/Region         187 non-null     object
 1   Confirmed              187 non-null     int64
 2   Deaths                 187 non-null     int64
 3   Recovered              187 non-null     int64
 4   Active                 187 non-null     int64
 5   New cases              187 non-null     float64
 6   New deaths             187 non-null     float64
 7   New recovered          187 non-null     float64
 8   Deaths / 100 Cases     187 non-null     float64
 9   Recovered / 100 Cases  187 non-null     float64
 10  Deaths / 100 Recovered 187 non-null     float64
 11  Confirmed last week    187 non-null     int64
 12  1 week change          186 non-null     float64
 13  1 week % increase      187 non-null     float64
 14  WHO Region             187 non-null     object
dtypes: float64(8), int64(5), object(2)
memory usage: 23.4+ KB
```

```
Country/Region                AfghanistanAlbaniaAlgeriaAndorraAngolaAntigua ...
Confirmed                                                        16505238
Deaths                                                             654176
Recovered                                                         9479862
Active                                                            6371200
New cases                                                        229471.0
New deaths                                                         5420.0
New recovered                                                    175218.0
Deaths / 100 Cases                                                  566.3
Recovered / 100 Cases                                            12181.96
Deaths / 100 Recovered                                                inf
Confirmed last week                                              14733623
1 week change                                                   1771615.0
1 week % increase                                                 2618.03
WHO Region                    Eastern MediterraneanEuropeAfricaEuropeAfricaA...
dtype: object
```
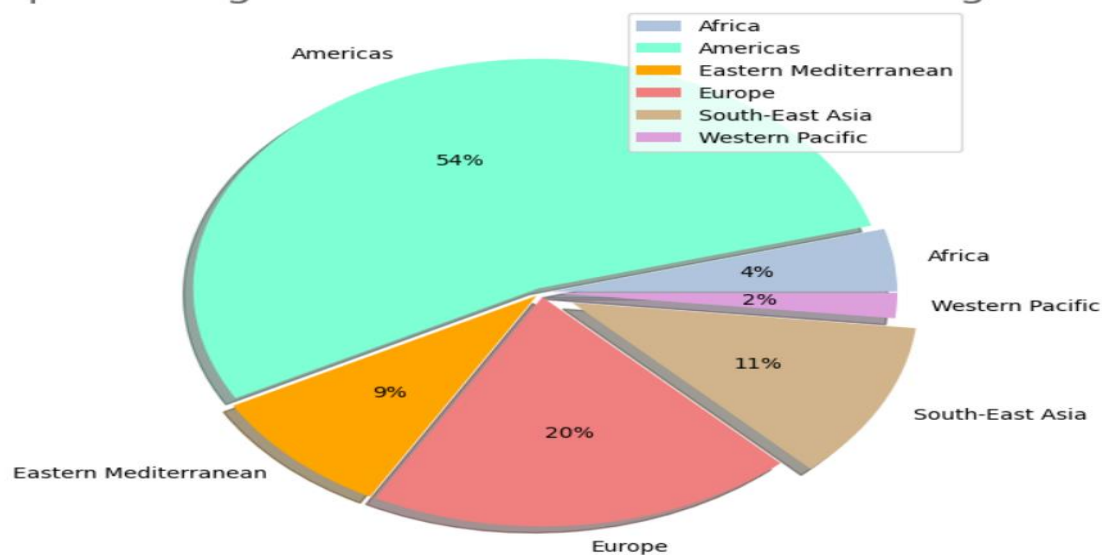
[40]:

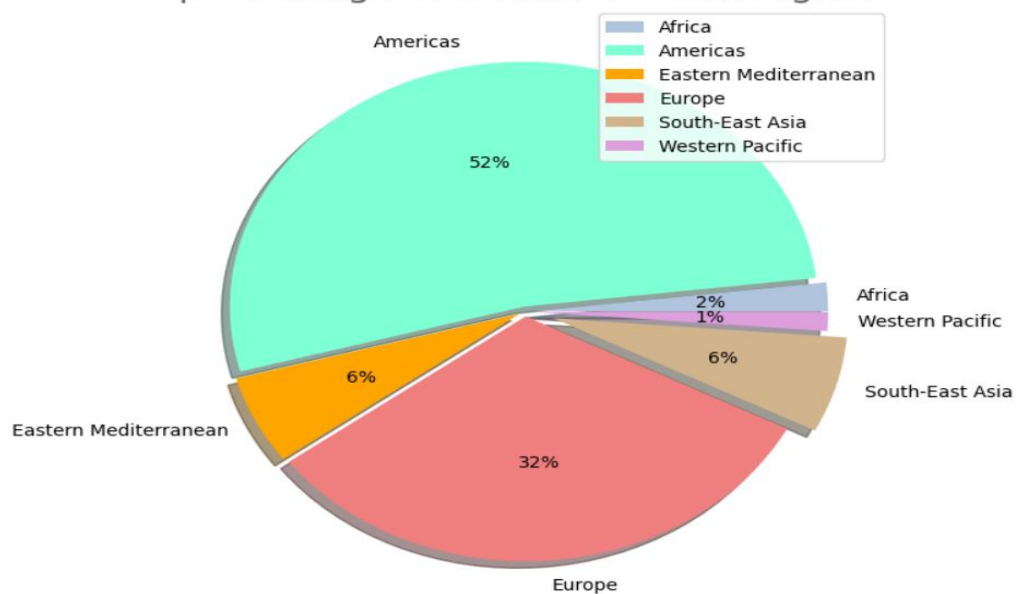|	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Confirme Recove Ra
0	Afghanistan	36263	1269	25198	9796	106.0	10.0	18.0	3.50	69.49	5.04	35526	737.0	2.07	Eastern Mediterranean	1.439'
1	Albania	4880	144	2745	1991	117.0	6.0	63.0	2.95	56.25	5.25	4171	709.0	17.00	Europe	1.777'
2	Algeria	27973	1163	18837	7973	616.0	8.0	749.0	4.16	67.34	6.17	23691	4282.0	18.07	Africa	1.485(
3	Andorra	907	52	803	52	10.0	NaN	NaN	5.73	88.53	6.48	884	23.0	2.60	Europe	1.129!
4	Angola	950	41	242	667	18.0	1.0	0.0	4.32	25.47	16.94	749	201.0	26.84	Africa	3.925(
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
185	Zambia	4552	140	2815	1597	71.0	1.0	465.0	3.08	61.84	4.97	3326	1226.0	36.86	Africa	1.617(
186	Zimbabwe	2704	36	542	2126	192.0	2.0	24.0	1.33	20.04	6.64	1713	991.0	57.85	Africa	4.988!
187	Uzbekistan	21209	121	11674	9414	678.0	5.0	569.0	0.57	55.04	1.04	17149	4060.0	23.67	Europe	1.816'
188	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa	
189	Namibia	1843	8	101	1734	68.0	0.0	26.0	0.43	5.48	7.92	1344	499.0	37.13	Africa	18.247!

190 rows × 17 columns

[39]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Confirme Recove Ra
0	Afghanistan	36263	1269	25198	9796	106.0	10.0	18.0	3.50	69.49	5.04	35526	737.0	2.07	Eastern Mediterranean	1.439'
1	Albania	4880	144	2745	1991	117.0	6.0	63.0	2.95	56.25	5.25	4171	709.0	17.00	Europe	1.777'
2	Algeria	27973	1163	18837	7973	616.0	8.0	749.0	4.16	67.34	6.17	23691	4282.0	18.07	Africa	1.485(
3	Andorra	907	52	803	52	10.0	NaN	NaN	5.73	88.53	6.48	884	23.0	2.60	Europe	1.129!
4	Angola	950	41	242	667	18.0	1.0	0.0	4.32	25.47	16.94	749	201.0	26.84	Africa	3.925(
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
185	Zambia	4552	140	2815	1597	71.0	1.0	465.0	3.08	61.84	4.97	3326	1226.0	36.86	Africa	1.617(
186	Zimbabwe	2704	36	542	2126	192.0	2.0	24.0	1.33	20.04	6.64	1713	991.0	57.85	Africa	4.988!
187	Uzbekistan	21209	121	11674	9414	678.0	5.0	569.0	0.57	55.04	1.04	17149	4060.0	23.67	Europe	1.816'
188	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa	
189	Namibia	1843	8	101	1734	68.0	0.0	26.0	0.43	5.48	7.92	1344	499.0	37.13	Africa	18.247!

190 rows × 17 columns

Top & bottom 10 country according to therir cofirmed,death,recoverd cases

```
[ ]:

[10]:  T1 = df.sort_values(by=['Confirmed'], ascending=False).head(10)
       T1
```

[10]:

	Country/ Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Acti Dea Ra
173	US	4290259	148011	1325804	2816444	56336.0	1076.0	27941.0	3.45	30.90	11.16	3834677	455582.0	11.88	Americas	19.028(
23	Brazil	2442375	87618	1846641	508116	23284.0	614.0	33728.0	3.59	75.61	4.74	2118646	323729.0	15.28	Americas	5.799;
79	India	1480073	33408	951166	495499	44457.0	637.0	33598.0	2.26	64.26	3.51	1155338	324735.0	28.11	South-East Asia	14.831;
138	Russia	816680	13334	602249	201097	5607.0	85.0	3077.0	1.63	73.74	2.21	776212	40468.0	5.21	Europe	15.081!
154	South Africa	452529	7067	274925	170537	7096.0	298.0	9848.0	1.56	60.75	2.57	373628	78901.0	21.12	Africa	24.131·
111	Mexico	395489	44022	303810	47657	4973.0	342.0	8588.0	11.13	76.82	14.49	349396	46093.0	13.19	Americas	1.082!
132	Peru	389717	18418	272547	98752	13756.0	575.0	4697.0	4.73	69.93	6.76	357681	32036.0	8.96	Americas	5.361;
35	Chile	347923	9187	319954	18782	2133.0	75.0	1859.0	2.64	91.96	2.87	333029	14894.0	4.47	Americas	2.044·
177	United Kingdom	301708	45844	1437	254427	688.0	7.0	3.0	15.19	0.48	3190.26	296944	4764.0	1.60	Europe	5.549{
81	Iran	293606	15912	255144	22550	2434.0	212.0	1931.0	5.42	86.90	6.24	276202	17404.0	6.30	Eastern Mediterranean	1.417;

	Country/ Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Active/ Deaths Ratio	Confirm Recov F
183	Western Sahara	10	1	8	1	0.0	0.0	0.0	10.0	80.00	12.5	10	0.0	0.00	Africa	1.0	1.25
75	Holy See	12	0	12	0	0.0	0.0	0.0	0.0	100.00	0.0	12	0.0	0.00	Europe	NaN	1.00
68	Greenland	14	0	13	1	1.0	0.0	0.0	0.0	92.86	0.0	13	1.0	7.69	Europe	inf	1.07
140	Saint Kitts and Nevis	17	0	15	2	0.0	0.0	0.0	0.0	88.24	0.0	17	0.0	0.00	Americas	inf	1.13
49	Dominica	18	0	18	0	0.0	0.0	0.0	0.0	100.00	0.0	18	0.0	0.00	Americas	NaN	1.00
94	Laos	20	0	19	1	0.0	0.0	0.0	0.0	95.00	0.0	19	1.0	5.26	Western Pacific	inf	1.05
69	Grenada	23	0	23	0	0.0	0.0	0.0	0.0	100.00	0.0	23	0.0	0.00	Americas	NaN	1.00
141	Saint Lucia	24	0	22	2	NaN	0.0	0.0	0.0	91.67	0.0	23	1.0	4.35	Americas	inf	1.09
168	Timor-Leste	24	0	0	24	0.0	0.0	0.0	0.0	0.00	0.0	24	0.0	0.00	South-East Asia	inf	
59	Fiji	27	0	18	9	0.0	NaN	0.0	0.0	66.67	0.0	27	0.0	0.00	Western Pacific	inf	1.50

	Country/ Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Acti Dea Ra
173	US	4290259	148011	1325804	2816444	56336.0	1076.0	27941.0	3.45	30.90	11.16	3834677	455582.0	11.88	Americas	19.028(
23	Brazil	2442375	87618	1846641	508116	23284.0	614.0	33728.0	3.59	75.61	4.74	2118646	323729.0	15.28	Americas	5.799;
177	United Kingdom	301708	45844	1437	254427	688.0	7.0	3.0	15.19	0.48	3190.26	296944	4764.0	1.60	Europe	5.549{
111	Mexico	395489	44022	303810	47657	4973.0	342.0	8588.0	11.13	76.82	14.49	349396	46093.0	13.19	Americas	1.082!
85	Italy	246286	35112	198593	12581	168.0	5.0	147.0	14.26	80.64	17.68	244624	1662.0	0.68	Europe	0.358;
79	India	1480073	33408	951166	495499	44457.0	637.0	33598.0	2.26	64.26	3.51	1155338	324735.0	28.11	South-East Asia	14.831;
61	France	220352	30212	81212	108928	2551.0	17.0	267.0	13.71	36.86	37.20	214023	6329.0	2.96	Europe	3.605·
157	Spain	272421	28432	150376	93613	0.0	0.0	0.0	10.44	55.20	18.91	264836	7585.0	2.86	Europe	3.292!
132	Peru	389717	18418	272547	98752	13756.0	575.0	4697.0	4.73	69.93	6.76	357681	32036.0	8.96	Americas	5.361;
81	Iran	293606	15912	255144	22550	2434.0	212.0	1931.0	5.42	86.90	6.24	276202	17404.0	6.30	Eastern Mediterranean	1.417;

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region	Active/ Deaths Ratio	Confirmed Recov...
94	Laos	20	0	19	1	0.0	0.0	0.0	0.0	95.00	0.0	19	1.0	5.26	Western Pacific	inf	1.05
55	Eritrea	265	0	191	74	2.0	0.0	2.0	0.0	72.08	0.0	251	14.0	5.58	Africa	inf	1.38
30	Cambodia	226	0	147	79	1.0	0.0	4.0	0.0	65.04	0.0	171	55.0	32.16	Western Pacific	inf	1.53
59	Fiji	27	0	18	9	0.0	NaN	0.0	0.0	66.67	0.0	27	0.0	0.00	Western Pacific	inf	1.50
68	Greenland	14	0	13	1	1.0	0.0	0.0	0.0	92.86	0.0	13	1.0	7.69	Europe	inf	1.07
69	Grenada	23	0	23	0	0.0	0.0	0.0	0.0	100.00	0.0	23	0.0	0.00	Americas	NaN	1.00
75	Holy See	12	0	12	0	0.0	0.0	0.0	0.0	100.00	0.0	12	0.0	0.00	Europe	NaN	1.00
19	Bhutan	99	0	86	13	4.0	0.0	1.0	0.0	86.87	0.0	90	9.0	10.00	South-East Asia	inf	1.15
49	Dominica	18	0	18	0	0.0	0.0	0.0	0.0	100.00	0.0	18	0.0	0.00	Americas	NaN	1.00
130	Papua New Guinea	62	0	11	51	0.0	0.0	0.0	0.0	17.74	0.0	19	43.0	226.32	Western Pacific	inf	5.63

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
23	Brazil	2442375	87618	1846641	508116	23284.0	614.0	33728.0	3.59	75.61	4.74	2118646	323729.0	15.28	Americas
173	US	4290259	148011	1325804	2816444	56336.0	1076.0	27941.0	3.45	30.90	11.16	3834677	455582.0	11.88	Americas
79	India	1480073	33408	951166	495499	44457.0	637.0	33598.0	2.26	64.26	3.51	1155338	324735.0	28.11	South-East Asia
138	Russia	816680	13334	602249	201097	5607.0	85.0	3077.0	1.63	73.74	2.21	776212	40468.0	5.21	Europe
35	Chile	347923	9187	319954	18782	2133.0	75.0	1859.0	2.64	91.96	2.87	333029	14894.0	4.47	Americas
111	Mexico	395489	44022	303810	47657	4973.0	342.0	8588.0	11.13	76.82	14.49	349396	46093.0	13.19	Americas
154	South Africa	452529	7067	274925	170537	7096.0	298.0	9848.0	1.56	60.75	2.57	373628	78901.0	21.12	Africa
132	Peru	389717	18418	272547	98752	13756.0	575.0	4697.0	4.73	69.93	6.76	357681	32036.0	8.96	Americas
81	Iran	293606	15912	255144	22550	2434.0	212.0	1931.0	5.42	86.90	6.24	276202	17404.0	6.30	Eastern Mediterranean
128	Pakistan	274289	5842	241026	27421	1176.0	20.0	3592.0	2.13	87.87	2.42	266096	8193.0	3.08	Eastern Mediterranean

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
32	Canada	116458	8944	0	107514	682.0	11.0	0.0	7.68	0.00	inf	112925	3533.0	3.13	Americas
168	Timor-Leste	24	0	0	24	0.0	0.0	0.0	0.00	0.00	0.0	24	0.0	0.00	South-East Asia
163	Syria	674	40	0	634	24.0	2.0	0.0	5.93	0.00	inf	522	152.0	29.12	Eastern Mediterranean
161	Sweden	79395	5700	0	73695	398.0	3.0	0.0	7.18	0.00	inf	78048	1347.0	1.73	Europe
147	Serbia	24141	543	0	23598	411.0	9.0	0.0	2.25	0.00	inf	21253	2888.0	13.59	Europe
188	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa
117	Mozambique	1701	11	0	1690	32.0	0.0	0.0	0.65	0.00	inf	1507	194.0	12.87	Africa
183	Western Sahara	10	1	8	1	0.0	0.0	0.0	10.00	80.00	12.5	10	0.0	0.00	Africa
130	Papua New Guinea	62	0	11	51	0.0	0.0	0.0	0.00	17.74	0.0	19	43.0	226.32	Western Pacific
75	Holy See	12	0	12	0	0.0	0.0	0.0	0.00	100.00	0.0	12	0.0	0.00	Europe

	Confirmed	Deaths	Recovered
WHO Region			
Africa	726751	12242	440746
Americas	8839286	342732	4468616
Eastern Mediterranean	1490744	38339	1201400
Europe	3320732	211265	2005397
South-East Asia	1835297	41349	1156933
Western Pacific	292428	8249	206770

percentage of Confirmed cases in each region

percentage of Deaths in each region

Covid cases Distribution
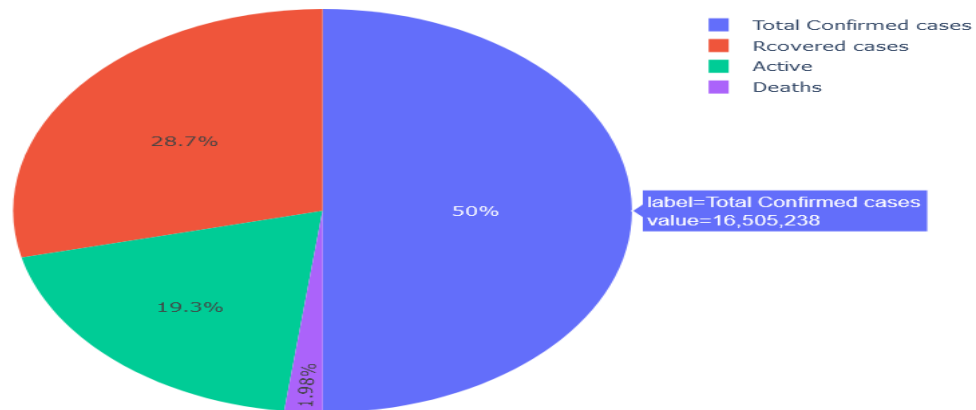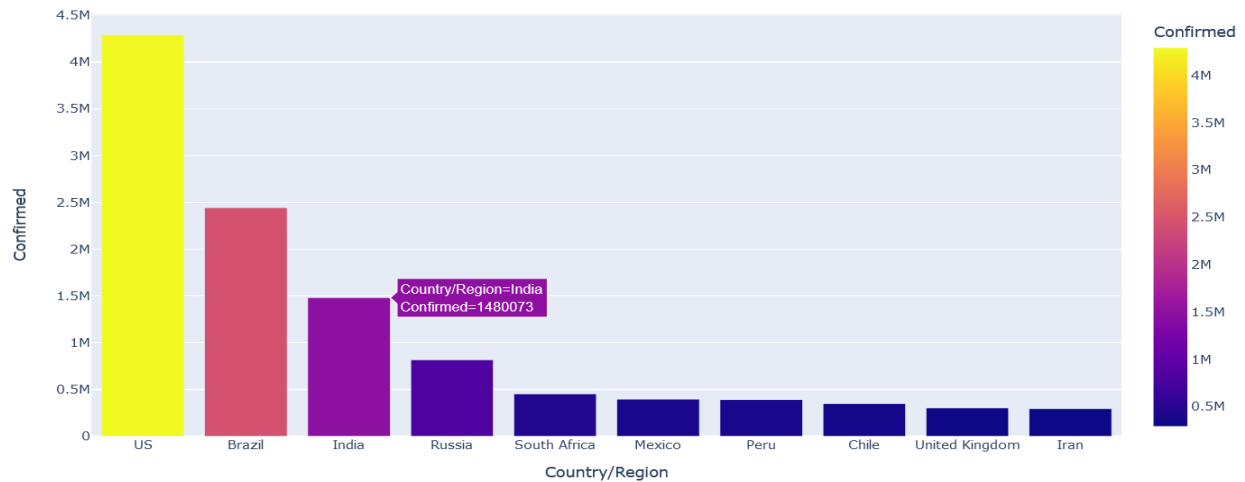
Legend:
- Total Confirmed cases
- Rcovered cases
- Active
- Deaths

- 50% — label=Total Confirmed cases value=16,505,238
- 28.7%
- 19.3%
- 1.98%

Top 10 country in the world with Confirmed cases

Country/Region=India Confirmed=1480073

Top 10 country in the world with Deaths cases

Country/Region=United Kingdom Deaths=45.844k

Legend (Country/Region):
- US
- Brazil
- United Kingdom
- Mexico
- Italy
- India
- France
- Spain
- Peru
- Iran
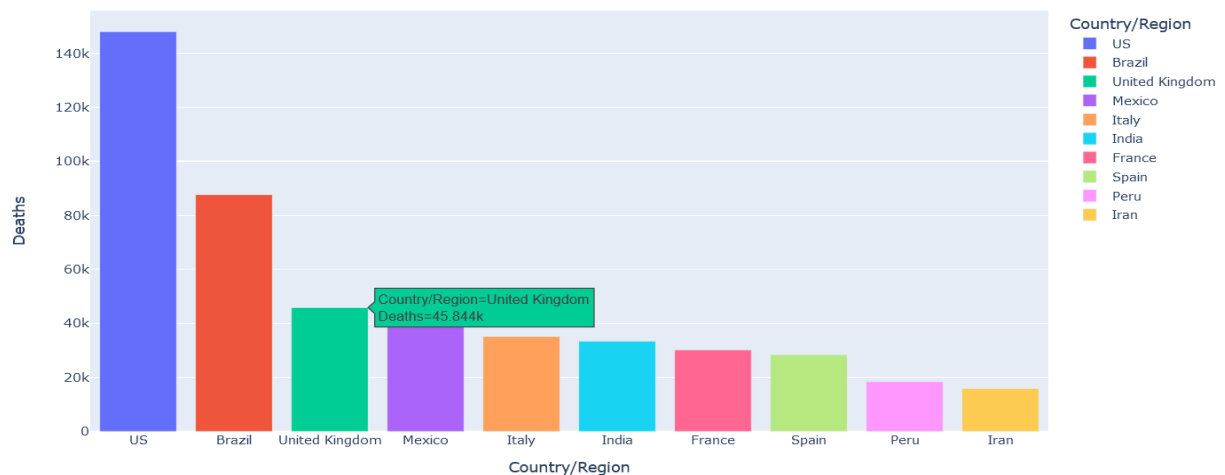
Top 10 country in the world with Rcovered cases
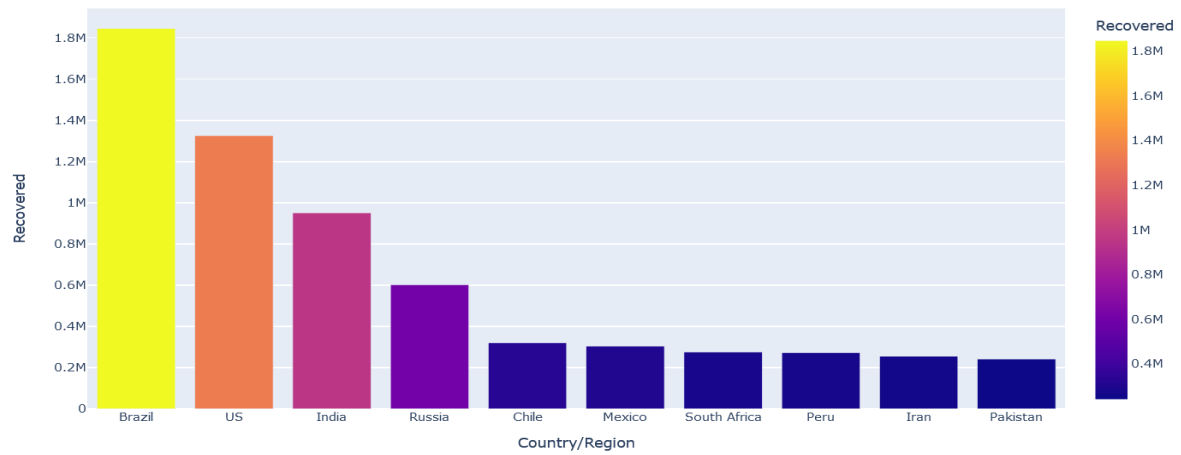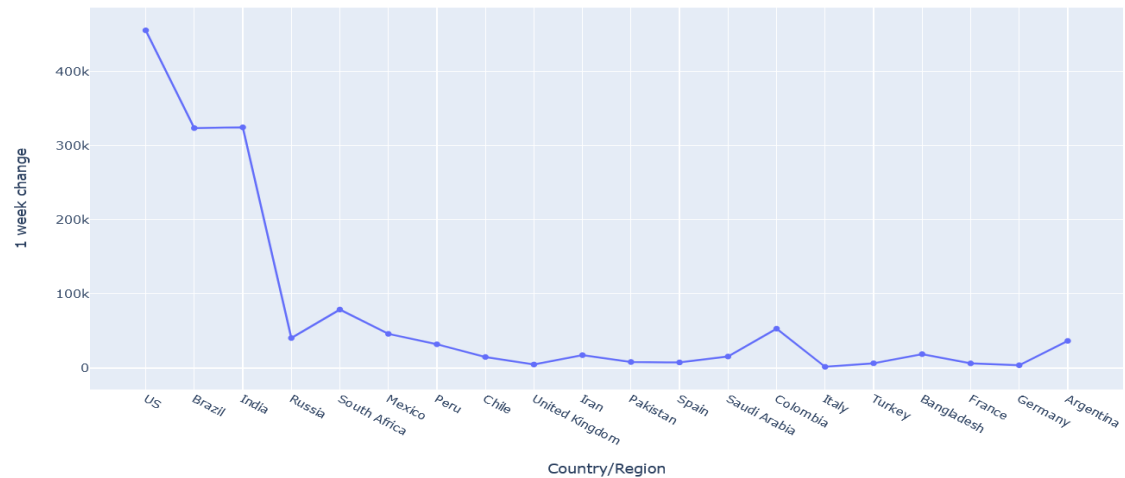
Last one week change in Top 20 Country

Deaths / 100 Cases of Top 20 Country

Country/Region=Netherlands
Deaths / 100 Cases=11.53

Recovered / 100 Cases of Top 20 Country

Country/Region=Norway
Recovered / 100 Cases=95.84

Deaths / 100 Recovered of Top 20 Country

Country/Region=France
Deaths / 100 Recovered=37.2
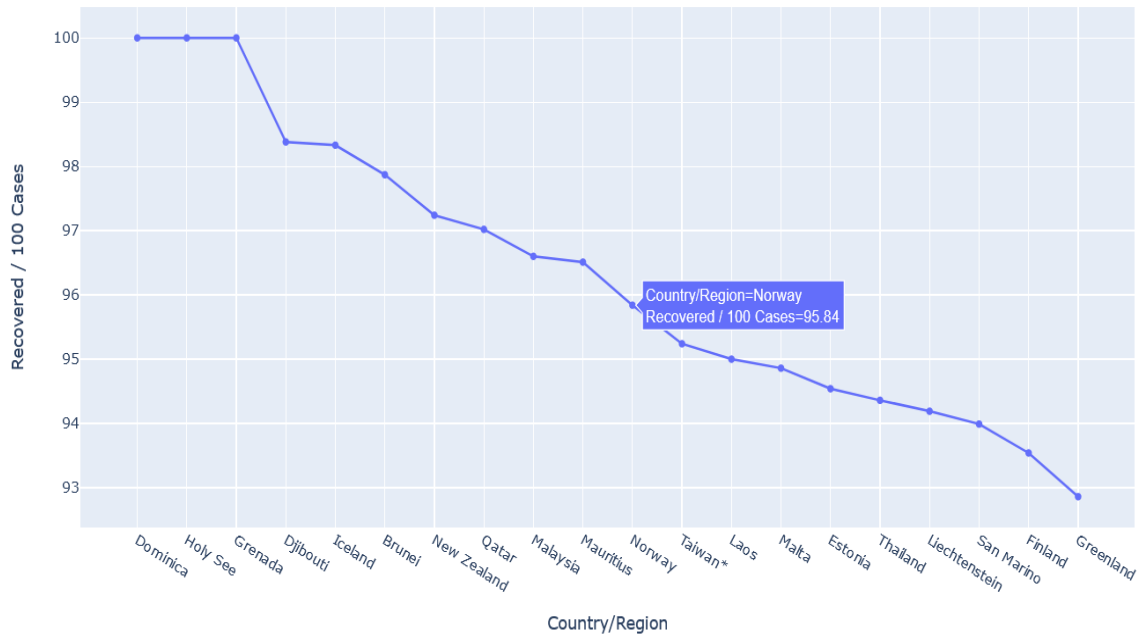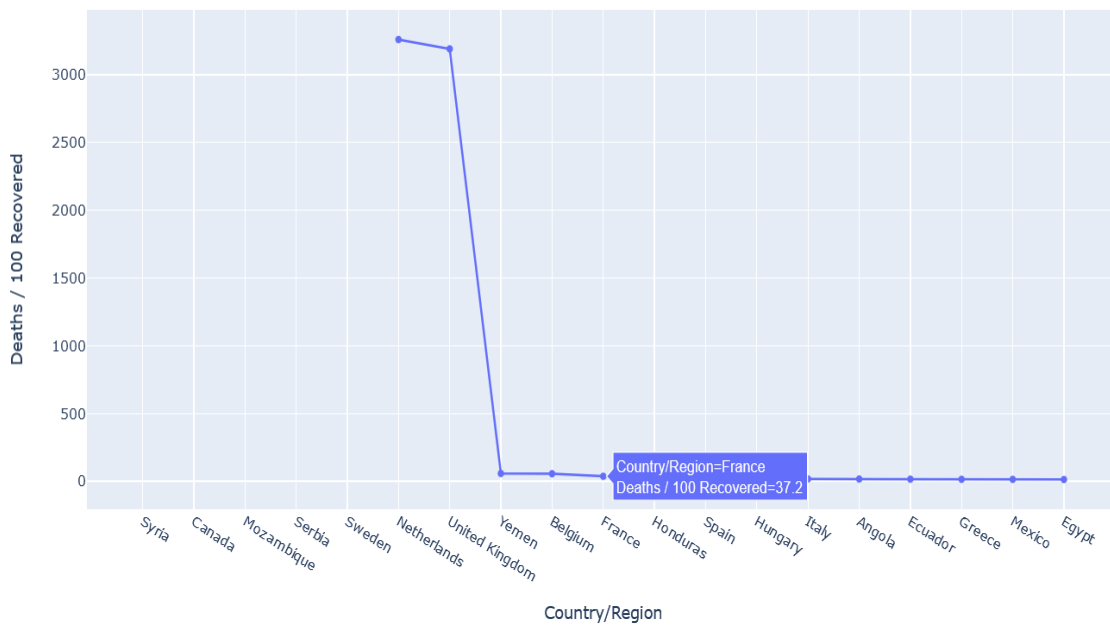
Active cases in each Country
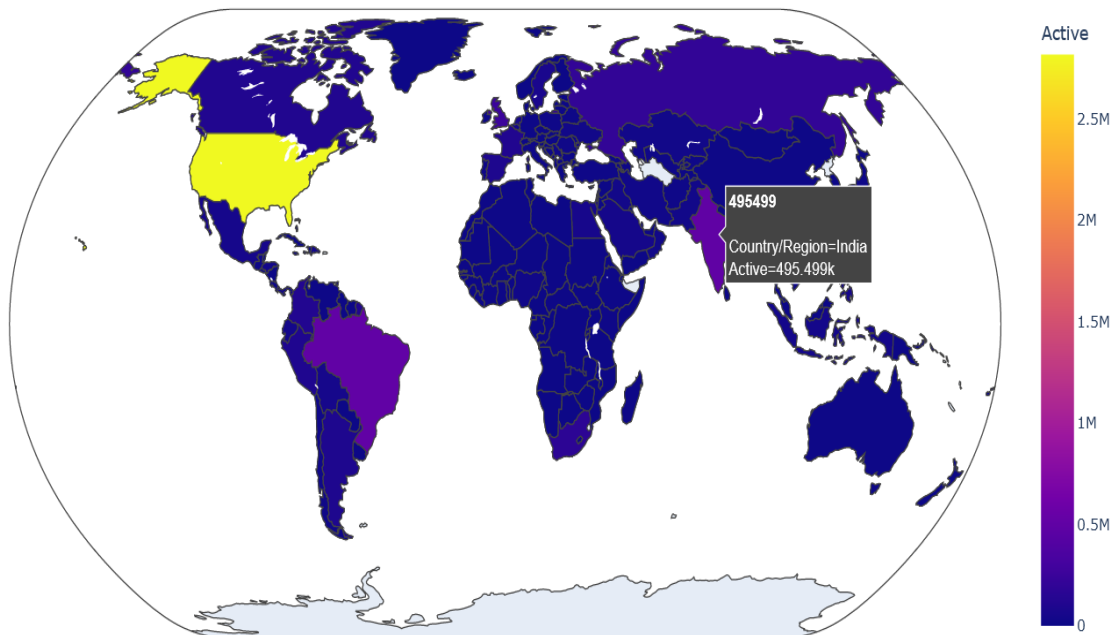
New deaths in each Country
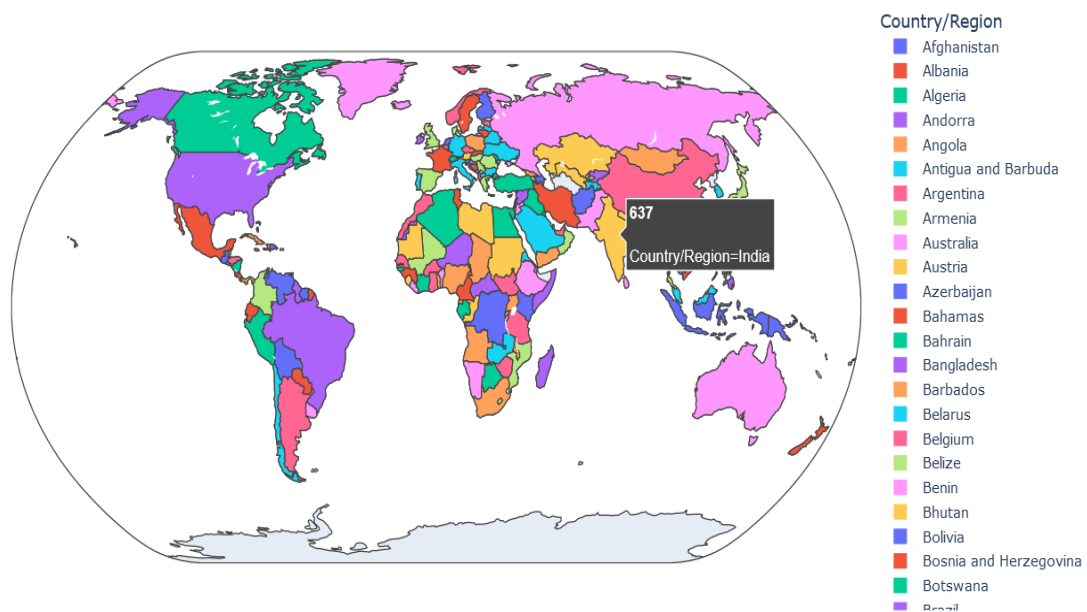
New cases in each Country

Covid Dashboard — View — Formula Bar

`= Table.Distinct(Source, {"Column1"})`

First table (columns Column6, Column7, Column8, Column9, Column10, Column11):

Column6	Column7	Column8	Column9	Column10	Column11
New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered
106	10	18	3.5	69.49	5.04
117	6	63	2.95	56.25	5.25
616	8	749	4.16	67.34	6.17
10			5.73	88.53	6.48
18	1	0	4.32	25.47	16.94
4	0	5	3.49	75.58	4.62
4890	120	2057	1.83	43.35	4.21
73	6	187	1.9	71.32	2.67
368	6	137	1.09	60.84	1.79
86	1	37	3.47	88.75	3.91
396	6	558	1.39	76.34	1.82
40	0	0	2.88	23.82	12.09
351	1	421	0.36	91.46	0.39
2772	37	1801	1.31	55.56	2.36
0		0	6.36	85.45	7.45
119	4	67	0.8	89.95	0.89
402	1	14	14.79	26.27	56.28
0	0	0	4.17	54.17	7.69
0	0	0	1.98	58.53	3.38
4	0	1	0	86.87	0
1752	64	309	3.72	30.17	12.32
731	14	375	2.8	46.96	5.96
53	1	11	0.27	8.53	3.17

15 COLUMNS, 188 ROWS   Column profiling based on top 1000 rows

APPLIED STEPS: Source, Removed Duplicates

Second screenshot:

`= Table.ReplaceValue(#"Removed Duplicates","","0",Replacer.ReplaceValue,{"Column6"})`

Replace Values dialog — "Replace one value with another in the selected columns."
Value To Find:
Replace With: 0
Advanced options
OK   Cancel

APPLIED STEPS: Source, Removed Duplicates, Replaced Value1

Column5	Column6
Active	New cases
9796	106
1991	117
7973	616
52	10
667	18
18	4
91782	4890
10014	73
5825	368
1599	86
6781	396
280	40
3231	351
97577	2772
9	0
6221	119
39154	402
20	0
699	0
13	4
47056	1752
5274	731
674	53

15 COLUMNS, 188 ROWS   Column profiling based on top 1000 rows

Covid -19 Dashboard

Confirmed	Deaths	Active	Recovered	New cases
16.48M	654.04K	6.36M	9.47M	228.69K

Confirmed cases in each region

- 0.29M (1.77%)
- 1.49M (9.05%)
- 1.84M (11.14%)
- 3.3M (20.02%)
- 8.84M (53.63%)

WHO Region

WHO Region **Americas**
Sum of Confirmed **88,39,286.00 (53.63%)**

- Eastern Mediterranean
- Africa
- Western Pacific
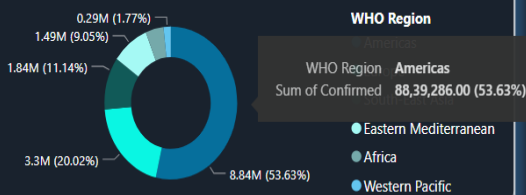
Confirmed cases in each country
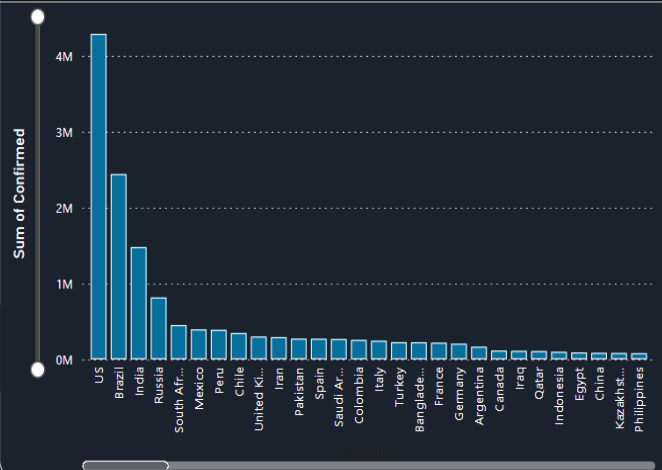
Deaths	New Deaths
654.04K	5.42K

New deaths in each country

Death cases in each region

- 38.34K (5.86%)
- 41.35K (6.32%)
- 342.73K (52.4%)
- 211.14K (32.28%)

WHO Region

WHO Region **Americas**
Sum of Deaths **3,42,732.00 (52.4%)**

- Eastern Mediterranean
- Africa
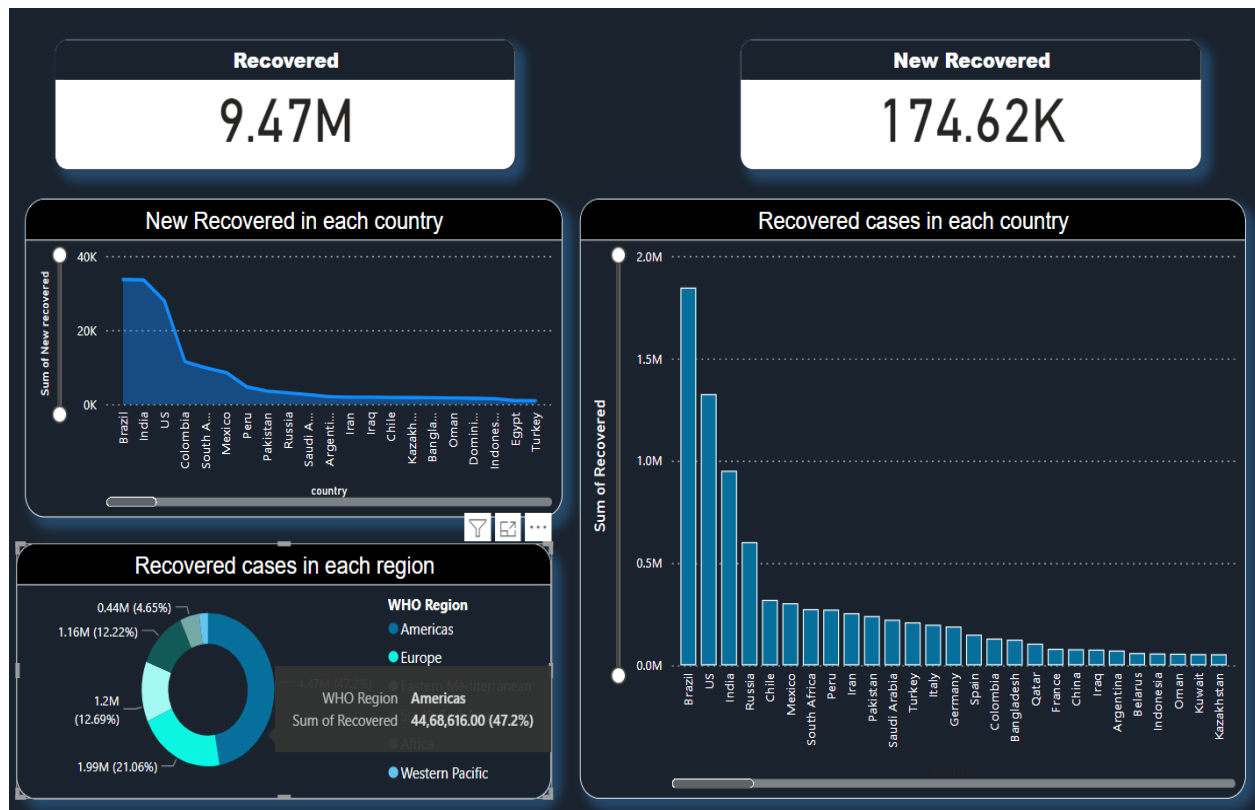- Western Pacific
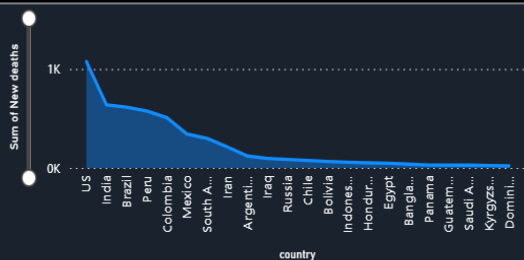
Death cases in each country
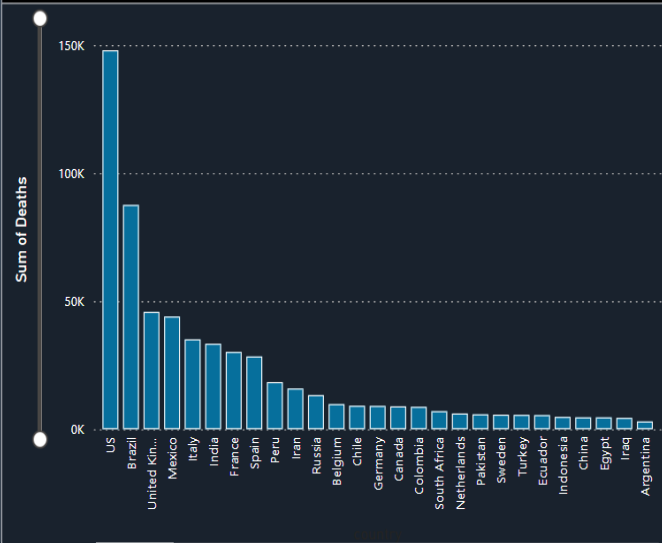
7. Future Scope of Project:

- **Epidemiological Modeling:** Develop and refine models to predict the spread of COVID-19, assess the effectiveness of public health interventions, and forecast future outbreaks. This could involve incorporating new data sources, refining model parameters, and adapting to changes in virus transmission patterns.
- **Variant Analysis:** Monitor the emergence and spread of new variants of the virus and assess their impact on transmission, severity, and vaccine effectiveness. Analyze genomic data to understand the genetic diversity of the virus and its implications for public health measures.
- **Vaccine Effectiveness:** Evaluate the real-world effectiveness of COVID-19 vaccines in different populations and settings. This could involve analyzing vaccine coverage, breakthrough infections, and the impact of vaccination on reducing transmission and severe disease.
- **Healthcare Capacity Planning:** Use data analysis to assess healthcare capacity, including hospitalizations, ICU admissions, and ventilator usage. Forecast future healthcare needs based on projected COVID-19 case counts and severity trends.
- **Policy and Decision Support:** Provide data-driven insights to policymakers and public health officials to support decision-making on mitigation measures, vaccination strategies, and resource allocation.

- **Social and Economic Impacts:** Assess the social and economic impacts of the pandemic, including employment trends, income disparities, and mental health outcomes. Use data analysis to inform policy responses and support vulnerable populations.
- **Public Health Messaging:** Evaluate the effectiveness of public health messaging and communication strategies in promoting vaccination, encouraging preventive behaviors, and countering misinformation. Use data analysis to identify effective messaging strategies and target specific populations.

8. **Conclusion:**

In this study, we set out to analyze various facets of the COVID-19 pandemic, aiming to provide insights into its spread, impacts, and response strategies. Leveraging a combination of epidemiological data, vaccination records, and genomic sequencing data, we conducted a comprehensive analysis to address the objectives outlined at the outset of the project.

We observed fluctuations in COVID-19 case counts over time, with periods of rapid spread followed by periods of relative stability. Hotspots of transmission were identified in densely populated urban areas, underscoring the importance of targeted intervention strategies.

Investments in healthcare infrastructure and workforce capacity are critical to ensure readiness for future surges in COVID-19 cases. Collaboration between public health agencies, healthcare providers, and community organizations is vital to optimize resource allocation and coordination.

In conclusion, this study contributes to our understanding of the COVID-19 pandemic and informs strategies to mitigate its impact on public health and society. By leveraging data-driven approaches and interdisciplinary collaboration, we can work towards a more resilient and equitable response to this global health crisis.

9. Bibliography :

For datasets:

https://www.kaggle.com/datasets/imdevskp/corona-virus-report

For matplotlib , numpy, pandas libraries:

https://www.w3schools.com/python/matplotlib_intro.asp

For plotly library :

https://neptune.ai/blog/plotly-python-tutorial-for-machine-learning-specialists

Other references:

https://en.m.wikipedia.org/wiki/Data_science