Date - 17/10/2023

Team ID - 3881

Project Title - Covid 19 Vaccines Analysis

Importing Dependencies

In [32]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Loading Dataset

In [33]:
```
dataset = pd.read_csv("C:\\Users\\Lenovo\\Downloads\\country_vaccinations.c
```

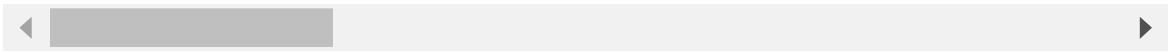
Data Exploration

In [34]:
```
dataset
```

Out[34]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vacci
0	Afghanistan	AFG	2021-02-22	0.0	0.0	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	
...	...	...	...	...	...	
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0	3473
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0	3487
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0	3493
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0	3501
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0	3510

86512 rows × 15 columns



In [35]:
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   country                               86512 non-null  object 
 1   iso_code                              86512 non-null  object 
 2   date                                  86512 non-null  object 
 3   total_vaccinations                    43607 non-null  float64
 4   people_vaccinated                     41294 non-null  float64
 5   people_fully_vaccinated               38802 non-null  float64
 6   daily_vaccinations_raw                35362 non-null  float64
 7   daily_vaccinations                    86213 non-null  float64
 8   total_vaccinations_per_hundred        43607 non-null  float64
 9   people_vaccinated_per_hundred         41294 non-null  float64
 10  people_fully_vaccinated_per_hundred   38802 non-null  float64
 11  daily_vaccinations_per_million        86213 non-null  float64
 12  vaccines                              86512 non-null  object 
 13  source_name                           86512 non-null  object 
 14  source_website                        86512 non-null  object 
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
```

In [36]:
```
dataset.describe()
```

Out[36]:

	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw
count	4.360700e+04	4.129400e+04	3.880200e+04	3.536200e+04
mean	4.592964e+07	1.770508e+07	1.413830e+07	2.705996e+05
std	2.246004e+08	7.078731e+07	5.713920e+07	1.212427e+06
min	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00
25%	5.264100e+05	3.494642e+05	2.439622e+05	4.668000e+03
50%	3.590096e+06	2.187310e+06	1.722140e+06	2.530900e+04
75%	1.701230e+07	9.152520e+06	7.559870e+06	1.234925e+05
max	3.263129e+09	1.275541e+09	1.240777e+09	2.474100e+07

In [37]:
```
dataset.columns
```

Out[37]:
```
Index(['country', 'iso_code', 'date', 'total_vaccinations',
       'people_vaccinated', 'people_fully_vaccinated',
       'daily_vaccinations_raw', 'daily_vaccinations',
       'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred',
       'people_fully_vaccinated_per_hundred', 'daily_vaccinations_per_mill
ion',
       'vaccines', 'source_name', 'source_website'],
      dtype='object')
```

Data Pre-Processing

Check for missing values

In [38]:
```
dataset.isnull()
```

Out[38]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinate
0	False	False	False	False	False	Tr
1	False	False	False	True	True	Tr
2	False	False	False	True	True	Tr
3	False	False	False	True	True	Tr
4	False	False	False	True	True	Tr
...	...	...	...	...	...	
86507	False	False	False	False	False	Fal
86508	False	False	False	False	False	Fal
86509	False	False	False	False	False	Fal
86510	False	False	False	False	False	Fal
86511	False	False	False	False	False	Fal

86512 rows × 15 columns

In [39]:
```
dataset.isnull().sum()
```

Out[39]:
```
country                                0
iso_code                               0
date                                   0
total_vaccinations                 42905
people_vaccinated                  45218
people_fully_vaccinated            47710
daily_vaccinations_raw             51150
daily_vaccinations                   299
total_vaccinations_per_hundred     42905
people_vaccinated_per_hundred      45218
people_fully_vaccinated_per_hundred    47710
daily_vaccinations_per_million       299
vaccines                               0
source_name                            0
source_website                         0
dtype: int64
```

In [40]:
```
dataset.dropna(inplace=True)
```

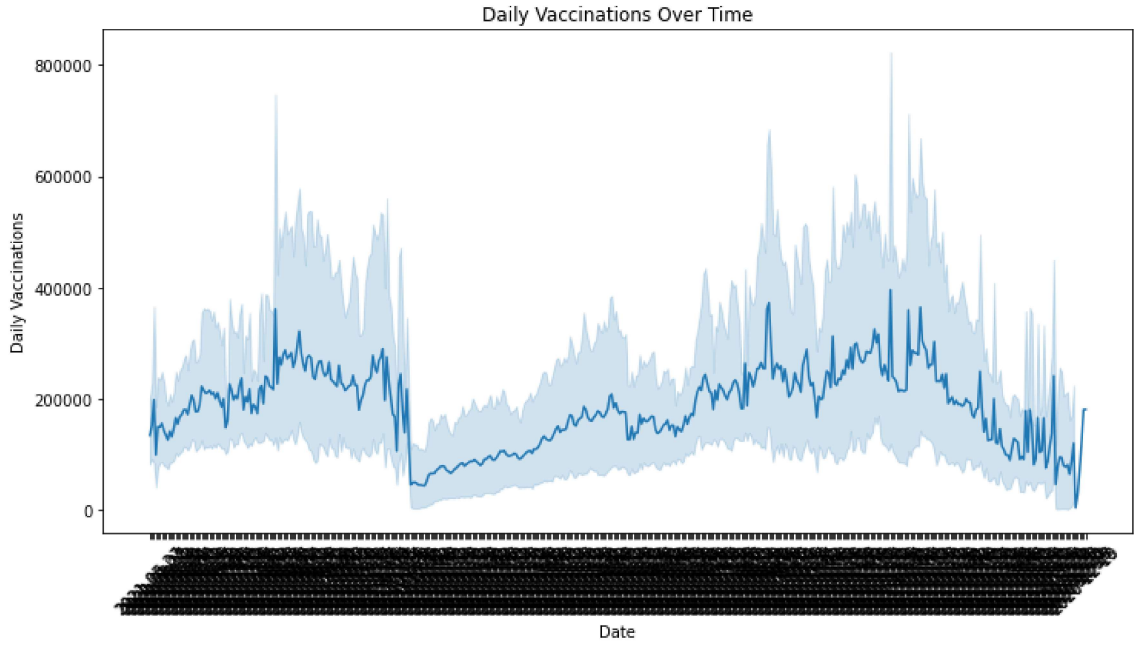
In [41]:
```
dataset.isnull().sum()
```

Out[41]:
```
country                                0
iso_code                               0
date                                   0
total_vaccinations                     0
people_vaccinated                      0
people_fully_vaccinated                0
daily_vaccinations_raw                 0
daily_vaccinations                     0
total_vaccinations_per_hundred         0
people_vaccinated_per_hundred          0
people_fully_vaccinated_per_hundred    0
daily_vaccinations_per_million         0
vaccines                               0
source_name                            0
source_website                         0
dtype: int64
```

Data Visualization

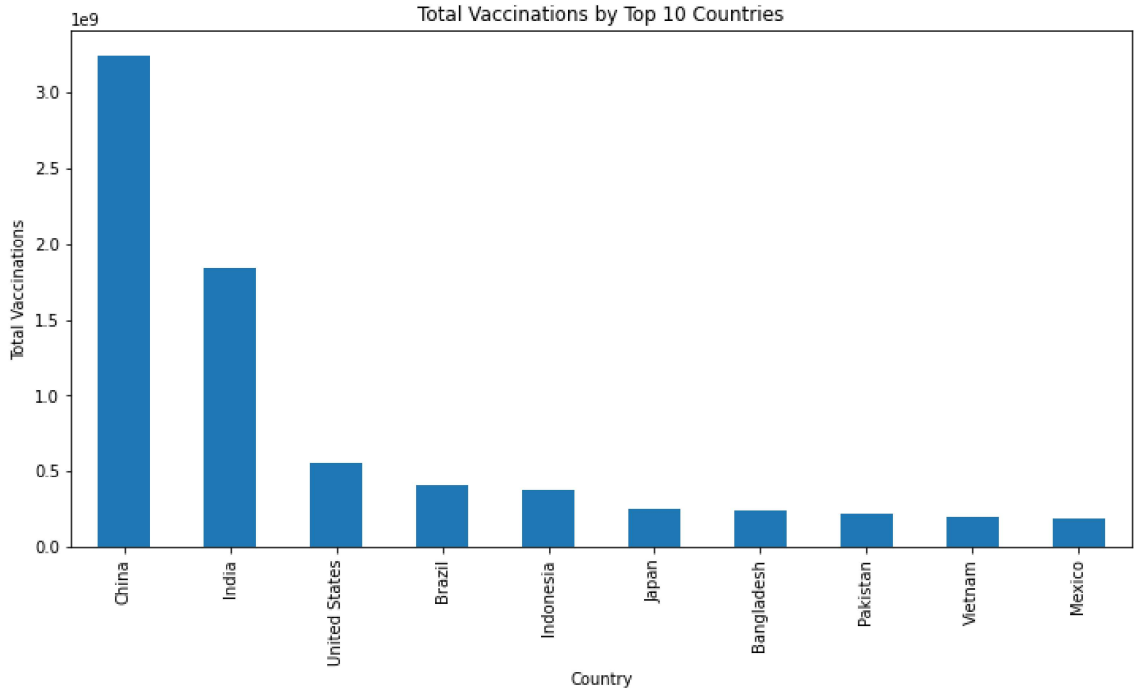
Plot a line chart for daily vaccinations over time

In [42]:
```
plt.figure(figsize=(12, 6))
sns.lineplot(x='date', y='daily_vaccinations', data=dataset)
plt.title('Daily Vaccinations Over Time')
plt.xticks(rotation=45)
plt.xlabel('Date')
plt.ylabel('Daily Vaccinations')
plt.show()
```



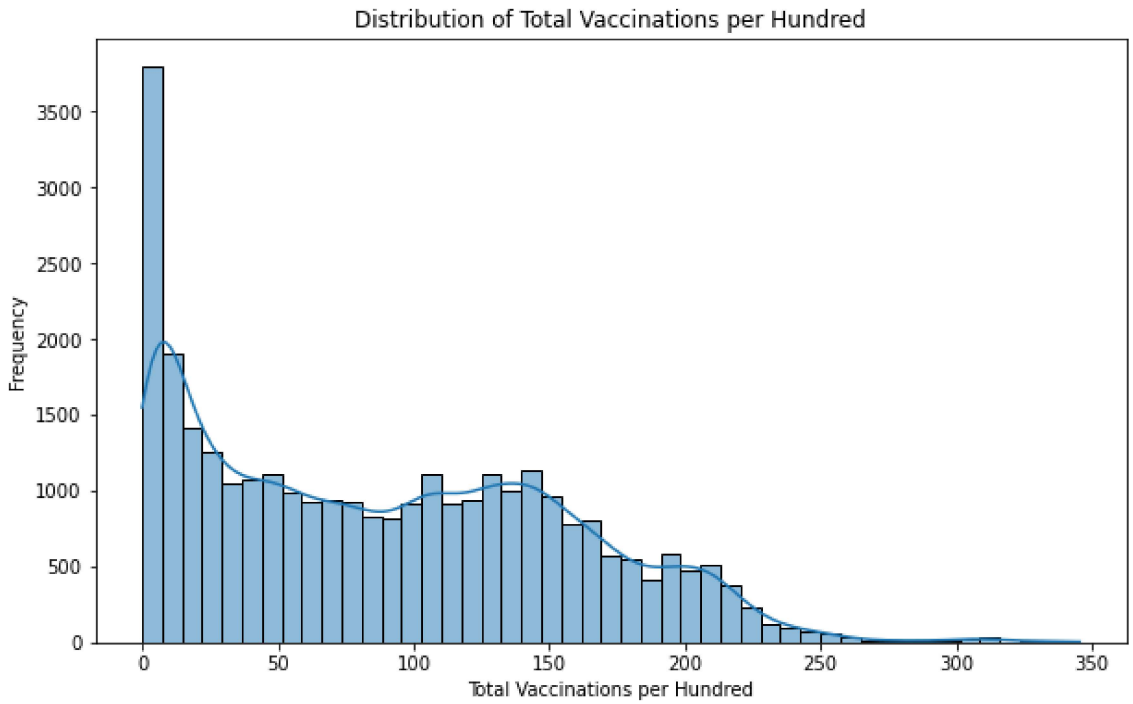
Create a bar chart to show total vaccinations by country

In [43]:
```
plt.figure(figsize=(12, 6))
total_vaccinations_by_country = dataset.groupby('country')['total_vaccinati
total_vaccinations_by_country[:10].plot(kind='bar')
plt.title('Total Vaccinations by Top 10 Countries')
plt.xlabel('Country')
plt.ylabel('Total Vaccinations')
plt.show()
```



Create a histogram for total vaccinations per hundred

In [44]:
```
plt.figure(figsize=(10, 6))
sns.histplot(dataset['total_vaccinations_per_hundred'], kde=True)
plt.title('Distribution of Total Vaccinations per Hundred')
plt.xlabel('Total Vaccinations per Hundred')
plt.ylabel('Frequency')
plt.show()
```



In [ ]: