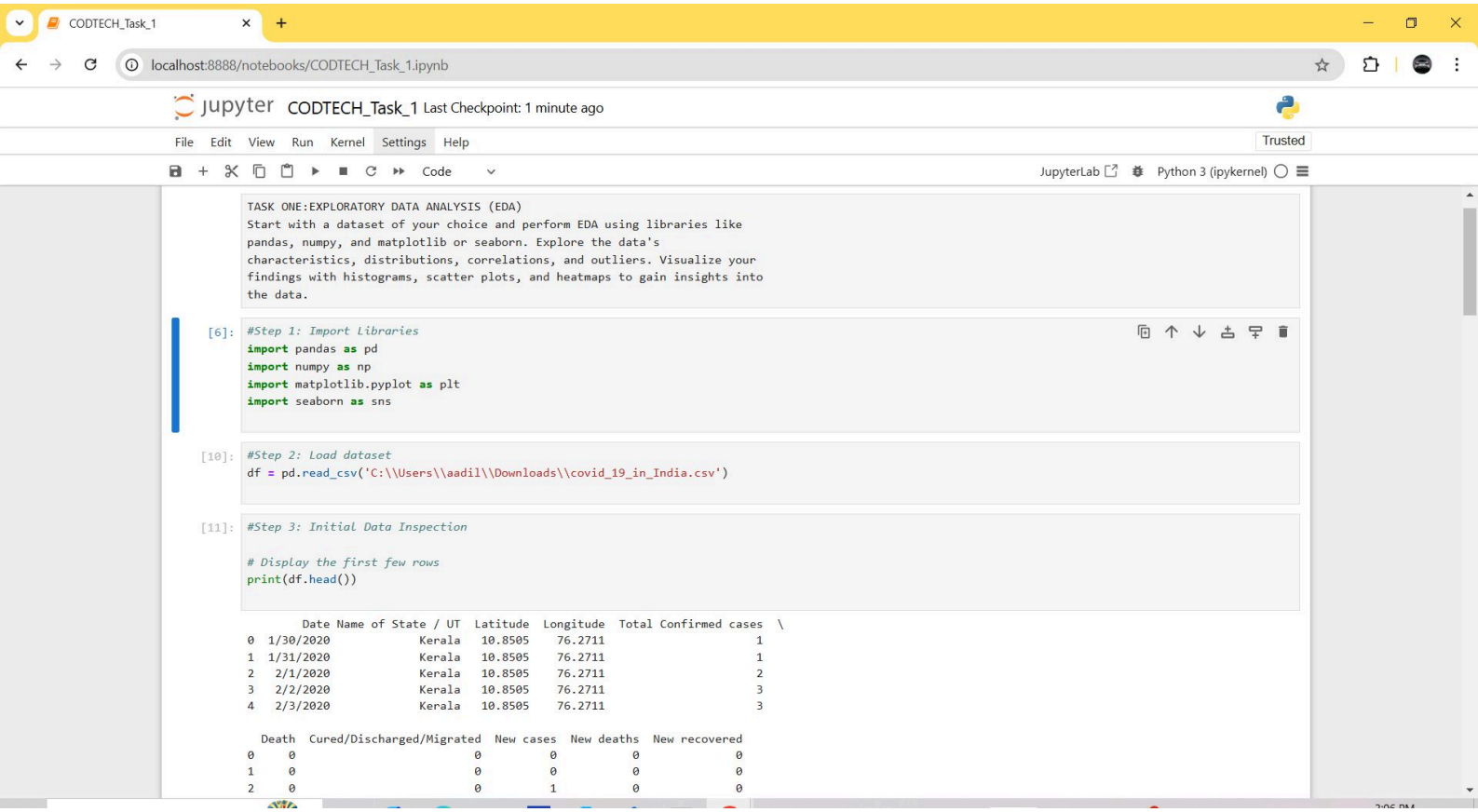

CODTECH_Task_1
localhost:8888/notebooks/CODTECH_Task_1.ipynb
jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
TASK ONE:EXPLORATORY DATA ANALYSIS (EDA)
Start with a dataset of your choice and perform EDA using libraries like
pandas, numpy, and matplotlib or seaborn. Explore the data's
characteristics, distributions, correlations, and outliers. Visualize your
findings with histograms, scatter plots, and heatmaps to gain insights into
the data.
[6]:
#Step 1: Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
[10]:
#Step 2: Load dataset
df = pd.read_csv('C:\\Users\\aadil\\Downloads\\covid_19_in_India.csv')
[11]:
#Step 3: Initial Data Inspection
# Display the first few rows
print(df.head())
Date Name of State / UT Latitude Longitude Total Confirmed cases \
0 1/30/2020 Kerala 10.8505 76.2711 1
1 1/31/2020 Kerala 10.8505 76.2711 1
2 2/1/2020 Kerala 10.8505 76.2711 2
3 2/2/2020 Kerala 10.8505 76.2711 3
4 2/3/2020 Kerala 10.8505 76.2711 3
Death Cured/Discharged/Migrated New cases New deaths New recovered
0 0 0 0 0 0
1 0 0 0 0 0
2 0 0 1 0 0

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[12]:
```
# Get info about the dataframe
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4692 entries, 0 to 4691
Data columns (total 10 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Date                       4692 non-null   object
 1   Name of State / UT         4692 non-null   object
 2   Latitude                   4692 non-null   float64
 3   Longitude                  4692 non-null   float64
 4   Total Confirmed cases      4692 non-null   int64
 5   Death                      4692 non-null   object
 6   Cured/Discharged/Migrated  4692 non-null   int64
 7   New cases                  4692 non-null   int64
 8   New deaths                 4692 non-null   int64
 9   New recovered              4692 non-null   int64
dtypes: float64(2), int64(5), object(3)
memory usage: 366.7+ KB
None
```

[13]:
```
# Get summary statistics
print(df.describe())
```

```
          Latitude    Longitude  Total Confirmed cases  \
count  4692.000000  4692.000000            4692.000000
mean     23.185327    81.451837           11393.925192
std       6.635913     6.959475           37208.600846
min       0.000000     0.000000               1.000000
25%      18.112400    76.271100              39.000000
50%      23.940800    79.019300             619.000000
75%      28.218000    85.313100            5233.000000
max      34.299600    94.727800          468265.000000

       Cured/Discharged/Migrated     New cases  New deaths  New recovered
count                4692.000000   4692.000000      4692.0    4692.000000
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
max      34.299600    94.727800          468265.000000

       Cured/Discharged/Migrated     New cases  New deaths  New recovered
count                4692.000000   4692.000000      4692.0    4692.000000
mean                 6908.130648    418.643009         0.0     283.069054
std                 23390.671258   1259.748923         0.0     947.925811
min                     0.000000      0.000000         0.0      -1.000000
25%                     9.000000      1.000000         0.0       0.000000
50%                   197.500000     26.000000         0.0       8.000000
75%                  2736.000000    210.250000         0.0     119.000000
max                305521.000000  18366.000000         0.0   13401.000000
```

[14]:
```
#Step 4: Data Cleaning

# Check for missing values
print(df.isnull().sum())
```

```
Date                          0
Name of State / UT            0
Latitude                      0
Longitude                     0
Total Confirmed cases         0
Death                         0
Cured/Discharged/Migrated     0
New cases                     0
New deaths                    0
New recovered                 0
dtype: int64
```

[15]:
```
# Fill or drop missing values (example)
df.fillna(method='ffill', inplace=True)  # Forward fill
```

```
C:\Users\aadil\AppData\Local\Temp\ipykernel_1012\3149975743.py:2: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  df.fillna(method='ffill', inplace=True)  # Forward fill
```

[39]:
```
#Step 5: Explore Data Characteristics
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

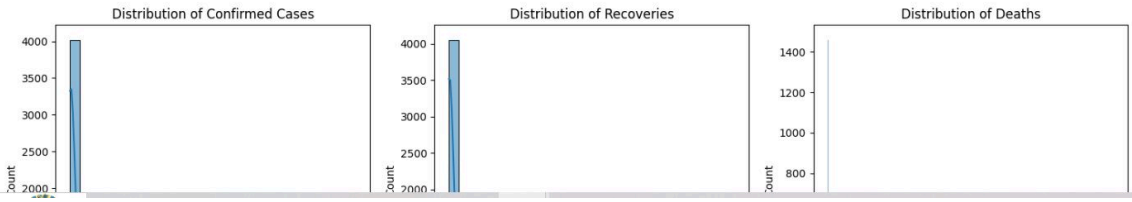
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[39]:
```
#Step 5: Explore Data Characteristics

#5.1. Distribution of Cases

plt.figure(figsize=(15, 5))

# Histogram for confirmed cases
plt.subplot(1, 3, 1)
sns.histplot(df['Total Confirmed cases'], bins=30, kde=True)
plt.title('Distribution of Confirmed Cases')

# Histogram for recoveries
plt.subplot(1, 3, 2)
sns.histplot(df['Cured/Discharged/Migrated'], bins=30, kde=True)
plt.title('Distribution of Recoveries')

# Histogram for deaths
plt.subplot(1, 3, 3)
sns.histplot(df['Death'], bins=30, kde=True)
plt.title('Distribution of Deaths')

plt.tight_layout()
plt.show()
```



CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

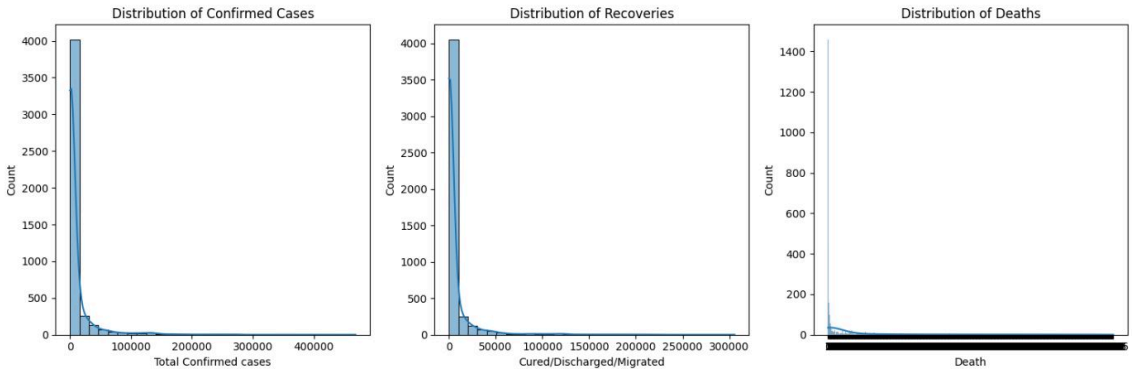
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
plt.show()
```



[41]:
```
# Check data types of the columns
print(df.dtypes)
```

```
Date                          object
Name of State / UT            object
Latitude                     float64
Longitude                    float64
Total Confirmed cases          int64
Death                         object
Cured/Discharged/Migrated      int64
New cases                      int64
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
dtype: object
```

[43]:
```
# Convert date column to datetime
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

# Optionally, drop any rows with NaT values after conversion
df.dropna(subset=['Date'], inplace=True)
```

[44]:
```
# Select only numeric columns for correlation
numeric_df = df.select_dtypes(include=[np.number])

# Check the numeric dataframe
print(numeric_df.head())

# Calculate correlation matrix
correlation_matrix = numeric_df.corr()
```

```
   Latitude  Longitude  Total Confirmed cases  Cured/Discharged/Migrated  \
0   10.8505    76.2711                      1                          0
1   10.8505    76.2711                      1                          0
2   10.8505    76.2711                      2                          0
3   10.8505    76.2711                      3                          0
4   10.8505    76.2711                      3                          0

   New cases  New deaths  New recovered
0          0           0              0
1          0           0              0
2          1           0              0
3          1           0              0
4          0           0              0
```

[46]:
```
#5.2. Correlation Analysis


# Heatmap of correlations
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 1 minute ago

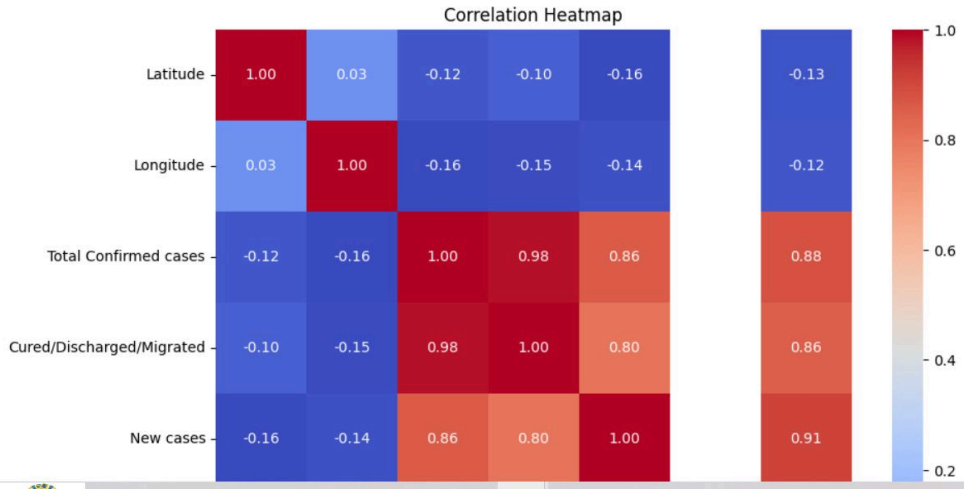
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[46]:
```
#5.2. Correlation Analysis


# Heatmap of correlations
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm', square=True)
plt.title('Correlation Heatmap')
plt.show()
```



CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

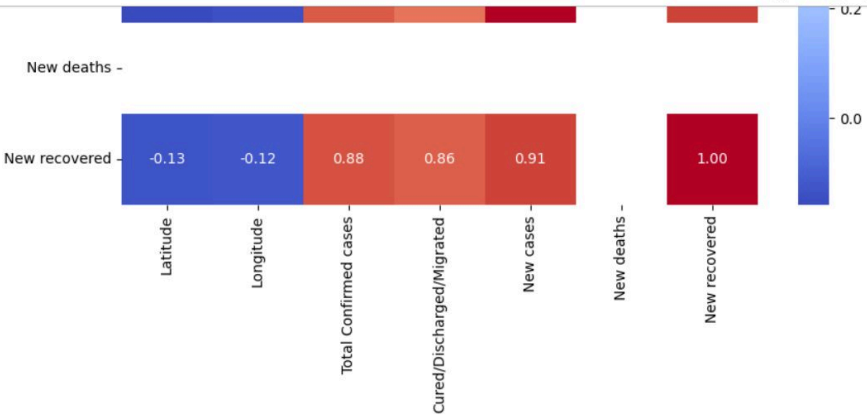
jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)



[48]:
```
#Step 6: Identify Outliers
plt.figure(figsize=(15, 5))

# Box plot for confirmed cases
plt.subplot(1, 3, 1)
sns.boxplot(y=df['Total Confirmed cases'])
plt.title('Box Plot of Confirmed Cases')

# Box plot for recoveries
plt.subplot(1, 3, 2)
sns.boxplot(y=df['Cured/Discharged/Migrated'])
plt.title('Box Plot of Recoveries')
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

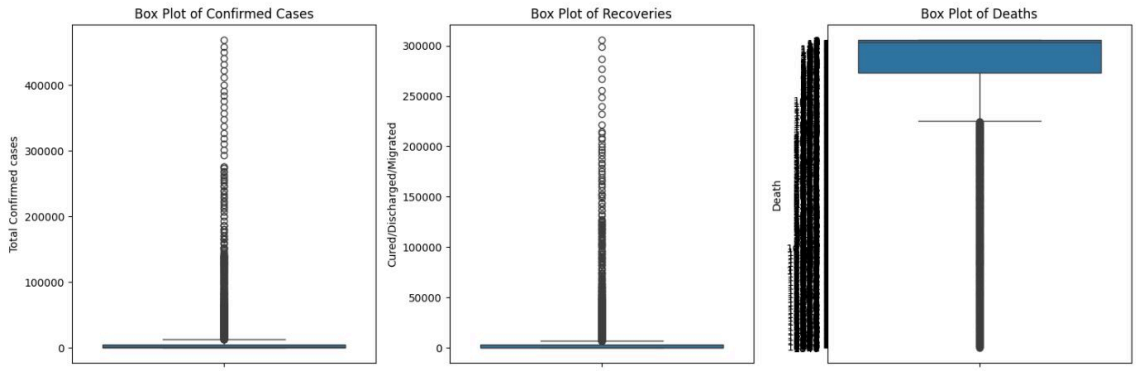
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
# Box plot for recoveries
plt.subplot(1, 3, 2)
sns.boxplot(y=df['Cured/Discharged/Migrated'])
plt.title('Box Plot of Recoveries')

# Box plot for deaths
plt.subplot(1, 3, 3)
sns.boxplot(y=df['Death'])
plt.title('Box Plot of Deaths')

plt.tight_layout()
plt.show()
```



CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

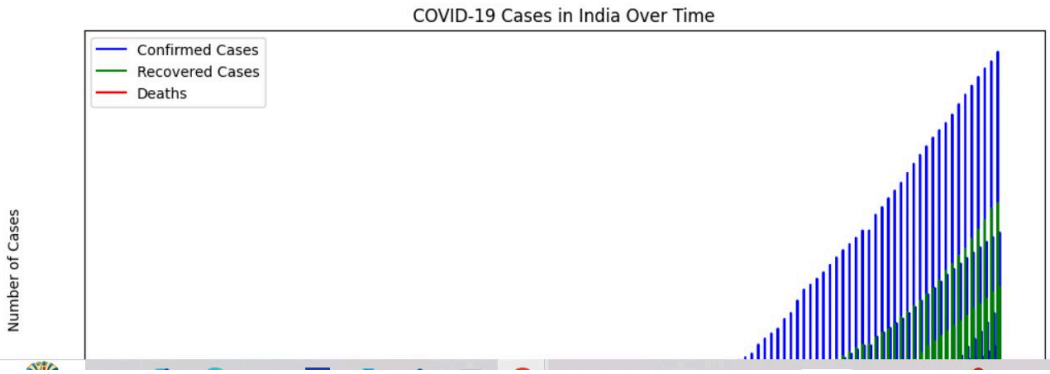
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[50]:
```
#Step 7: Time Series Analysis


# Plotting time series
plt.figure(figsize=(12, 6))
plt.plot(df['Total Confirmed cases'], label='Confirmed Cases', color='blue')
plt.plot(df['Cured/Discharged/Migrated'], label='Recovered Cases', color='green')
plt.plot(df['Death'], label='Deaths', color='red')
plt.title('COVID-19 Cases in India Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Cases')
plt.legend()
plt.show()
```



CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

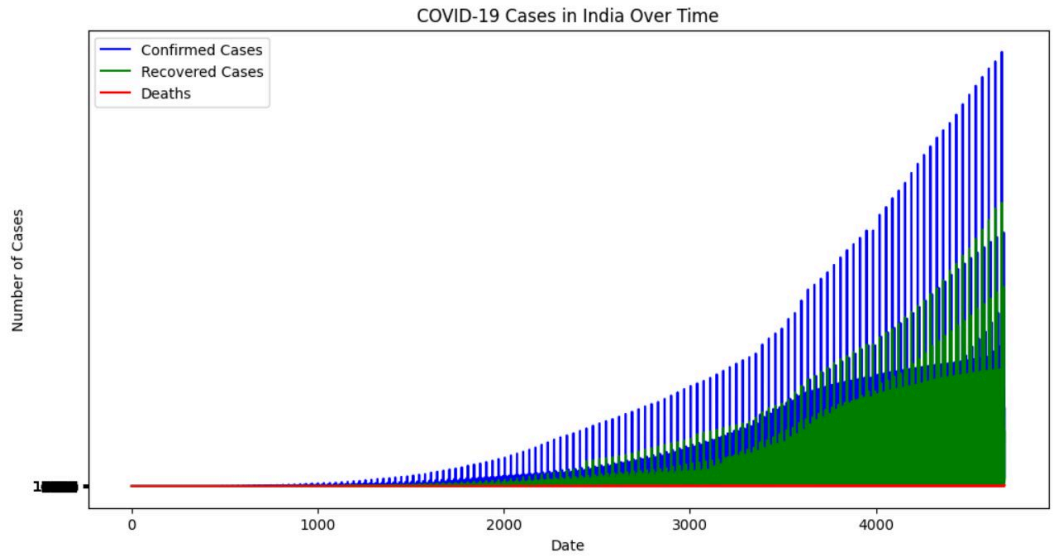
jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)



[52]:
```
#Step 8: Scatter Plots

#Confirmed Cases vs. Recoveries
```

CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

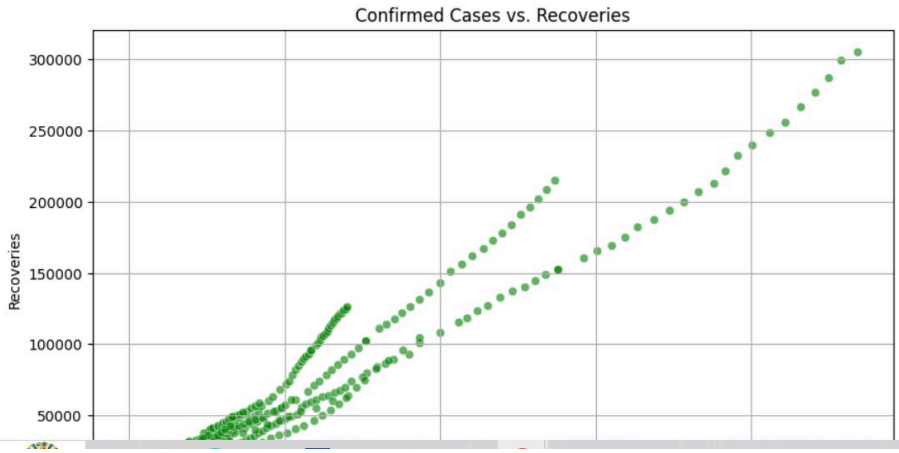
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[52]:
```
#Step 8: Scatter Plots

#Confirmed Cases vs. Recoveries
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='Total Confirmed cases', y='Cured/Discharged/Migrated', color='green', alpha=0.6)
plt.title('Confirmed Cases vs. Recoveries')
plt.xlabel('Confirmed Cases')
plt.ylabel('Recoveries')
plt.grid(True)
plt.show()
```



CODTECH_Task_1

localhost:8888/notebooks/CODTECH_Task_1.ipynb

jupyter CODTECH_Task_1 Last Checkpoint: 2 minutes ago

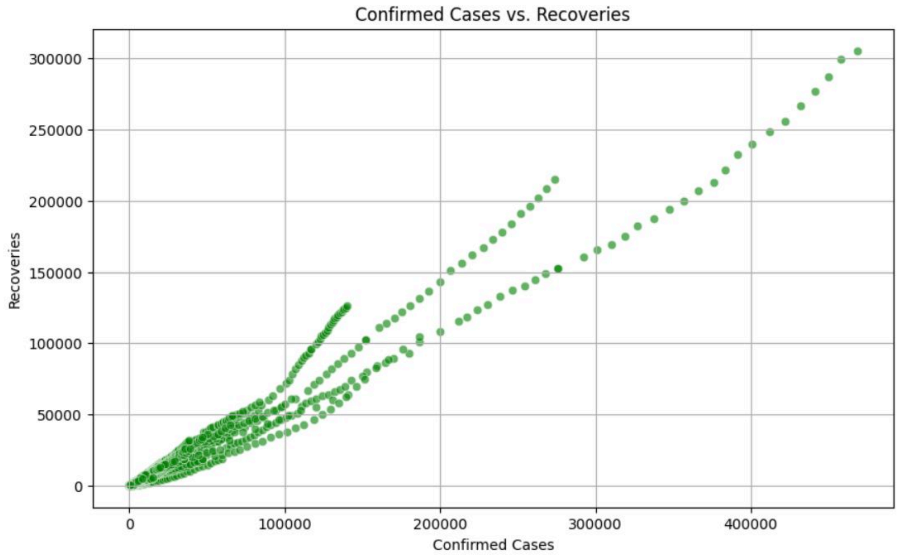
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
plt.show()
```



[ ]: