DATA SCIENCE TASK 03

Covid-19 Data Analysis Project using Python

*# Importing Libraries*

**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** plotly.express **as** px

**import** plotly.graph\_objects **as** go

*# Load the dataset*

df **=** pd**.**read\_csv(r"C:\Users\Ishita\Downloads\country\_wise\_latest.csv")

df**.**head()

*# Checking for missing data*

missing\_data **=** df**.**isnull()**.**sum()

print("\nMissing Data Analysis:")

print(missing\_data)

*# Data Preprocessing*

*# Checking for negative values in columns like deaths, which should be impossible*

df['Deaths'] **=** df['Deaths']**.**clip(lower**=**0)

df['Confirmed'] **=** df['Confirmed']**.**clip(lower**=**0)

df['Recovered'] **=** df['Recovered']**.**clip(lower**=**0)

df['Active'] **=** df['Active']**.**clip(lower**=**0)

*# Calculate additional metrics: Case Fatality Rate (CFR) and Recovery Rate*

df['CFR'] **=** (df['Deaths'] **/** df['Confirmed']) **\*** 100 *# Case Fatality Rate (%)*

df['Recovery Rate'] **=** (df['Recovered'] **/** df['Confirmed']) **\*** 100 *# Recovery Rate (%)*

df['Recovery to Death Ratio'] **=** df['Recovered'] **/** df['Deaths'] *# Recovery to Death ratio*

*# Checking the data after preprocessing*

print("\nData after cleaning:")

print(df**.**info())

*# Function to plot the trend of 1 week change in top 10 countries*

**def** plot\_1\_week\_change\_trend(df):

plt**.**figure(figsize**=**(14, 8))

sns**.**regplot(x**=**'Index', y**=**'1 week change', data**=**df, scatter\_kws**=**{'s': 50}, line\_kws**=**{"color": "red"})

plt**.**title('Trend of 1 Week Change in Covid-19 Cases (Top 10 Countries)', fontsize**=**16)

plt**.**xlabel('Country Index', fontsize**=**14)

plt**.**ylabel('1 Week Change in Cases', fontsize**=**14)

plt**.**xticks(ticks**=**df['Index'], labels**=**df['Country/Region'], rotation**=**45)

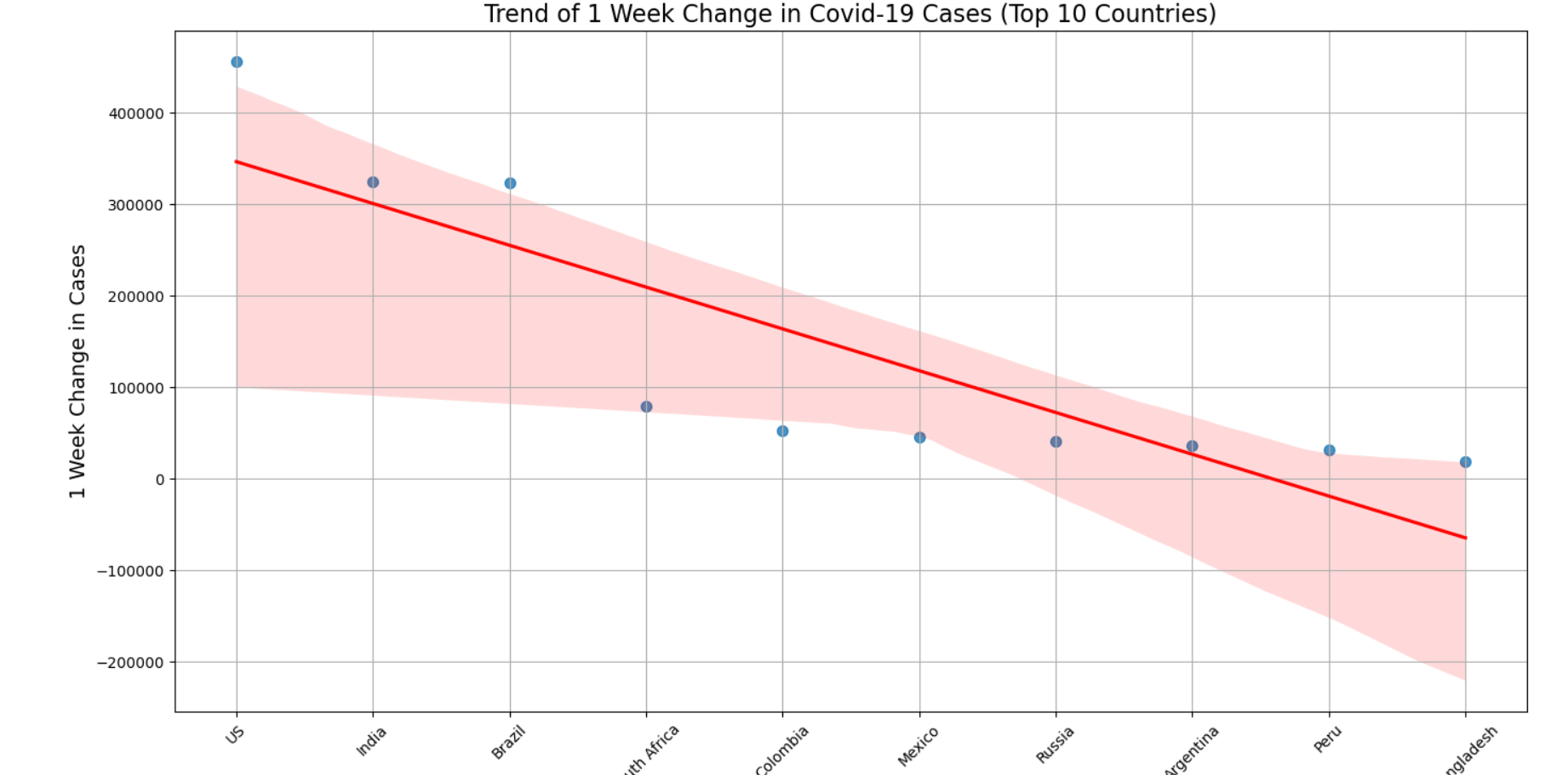
plt**.**grid(**True**)

plt**.**tight\_layout()

plt**.**show()

*# Plot the 1 Week Change Trend*

plot\_1\_week\_change\_trend(top\_10\_countries\_by\_change)



*# Plotting: Top 10 Countries by Confirmed Cases (with 'hue' to avoid warning)*

top\_countries **=** df**.**nlargest(10, 'Confirmed') *# Select the top 10 countries by confirmed cases*

plt**.**figure(figsize**=**(14, 8))

sns**.**barplot(x**=**'Confirmed', y**=**'Country/Region', data**=**top\_countries, hue**=**'WHO Region', palette**=**'magma')

plt**.**title('Top 10 Countries by Confirmed Cases', fontsize**=**16)

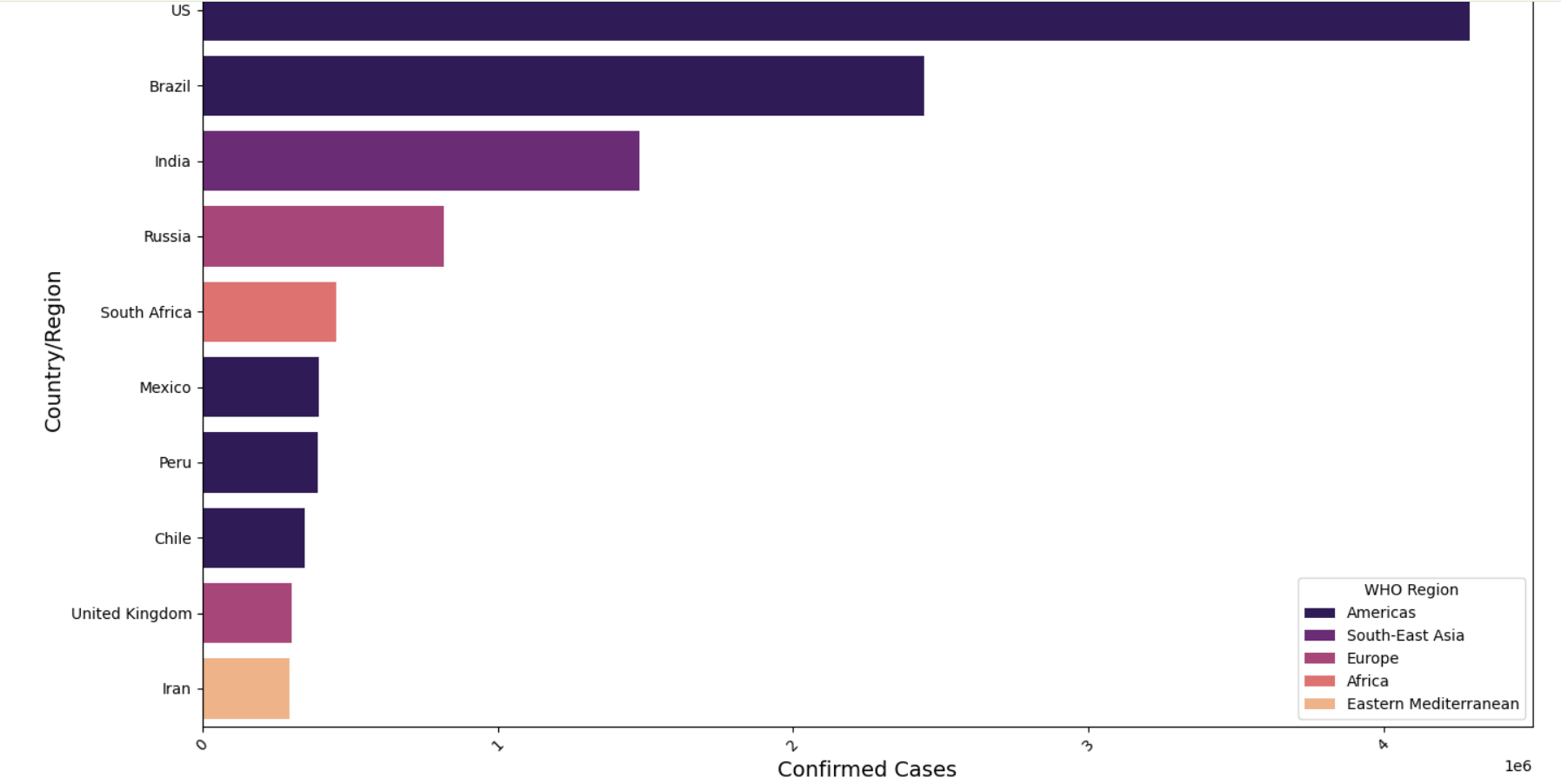
plt**.**xlabel('Confirmed Cases', fontsize**=**14)

plt**.**ylabel('Country/Region', fontsize**=**14)

plt**.**xticks(rotation**=**45)

plt**.**tight\_layout()

plt**.**show()



*# Plotting: Interactive Bar Plot using Plotly (Top 10 Countries)*

fig **=** px**.**bar(df, x**=**'Country/Region', y**=**'Confirmed', color**=**'Confirmed',

title**=**'Covid-19 Confirmed Cases by Country/Region',

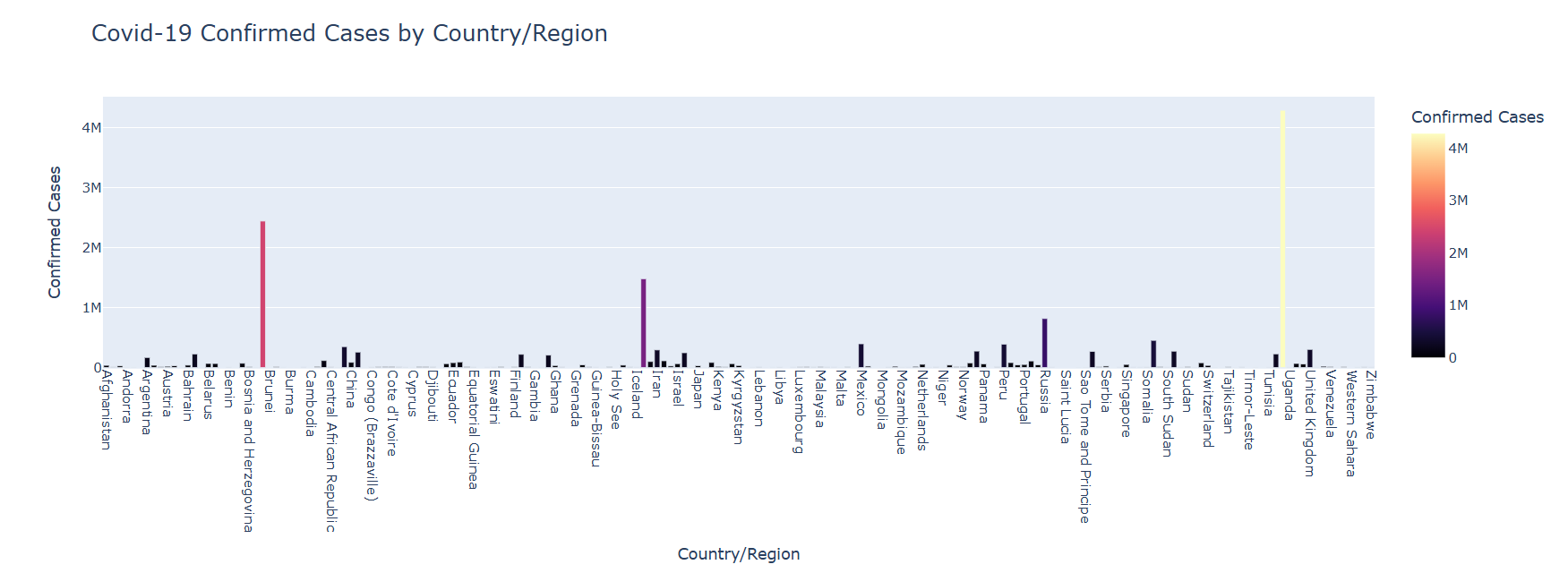
labels**=**{'Country/Region': 'Country/Region', 'Confirmed': 'Confirmed Cases'},

color\_continuous\_scale**=**"magma")

fig**.**update\_xaxes(tickangle**=**90) *# Rotate x-axis labels for readability*

fig**.**update\_layout(title\_font\_size**=**20)

fig**.**show()



*# Plotting: CFR by WHO Region (Boxplot) with Customizations*

plt**.**figure(figsize**=**(12, 6))

*# Using hue to avoid deprecation warning*

sns**.**boxplot(x**=**'WHO Region', y**=**'CFR', data**=**df, palette**=**"Set2", hue**=**'WHO Region')

plt**.**title('Case Fatality Rate (CFR) by WHO Region', fontsize**=**16)

plt**.**xticks(rotation**=**45)

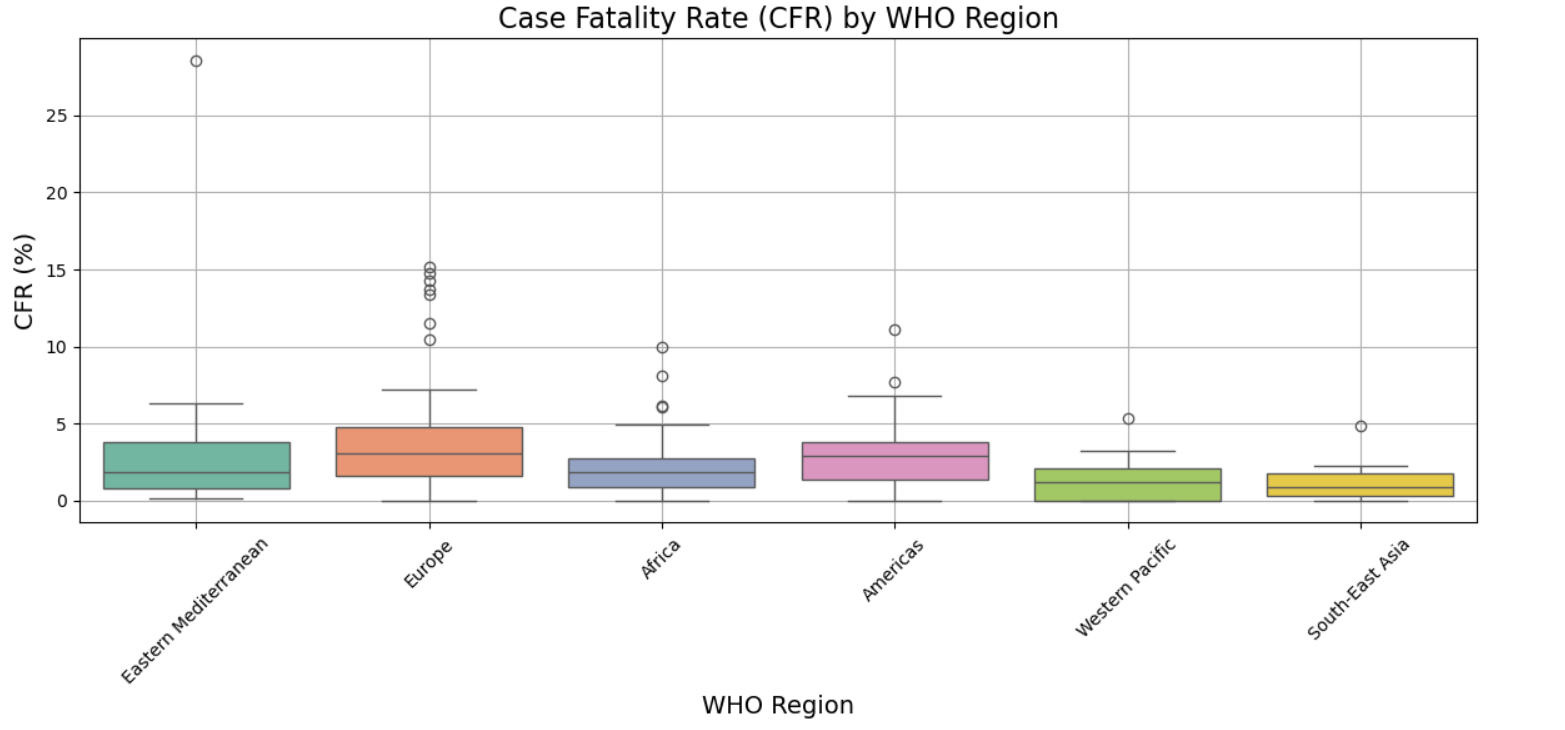
plt**.**xlabel('WHO Region', fontsize**=**14)

plt**.**ylabel('CFR (%)', fontsize**=**14)

plt**.**grid(**True**)

plt**.**tight\_layout()

plt**.**show()



*# Scatterplot: Recovery Rate vs Case Fatality Rate with Custom Marker Sizes and Hue*

plt**.**figure(figsize**=**(12, 6))

sns**.**scatterplot(x**=**'Recovery Rate', y**=**'CFR', data**=**df, hue**=**'WHO Region', palette**=**'Spectral',

size**=**'Confirmed', sizes**=**(30, 300), marker**=**'o')

plt**.**title('Recovery Rate vs Case Fatality Rate', fontsize**=**16)

plt**.**xlabel('Recovery Rate (%)', fontsize**=**14)

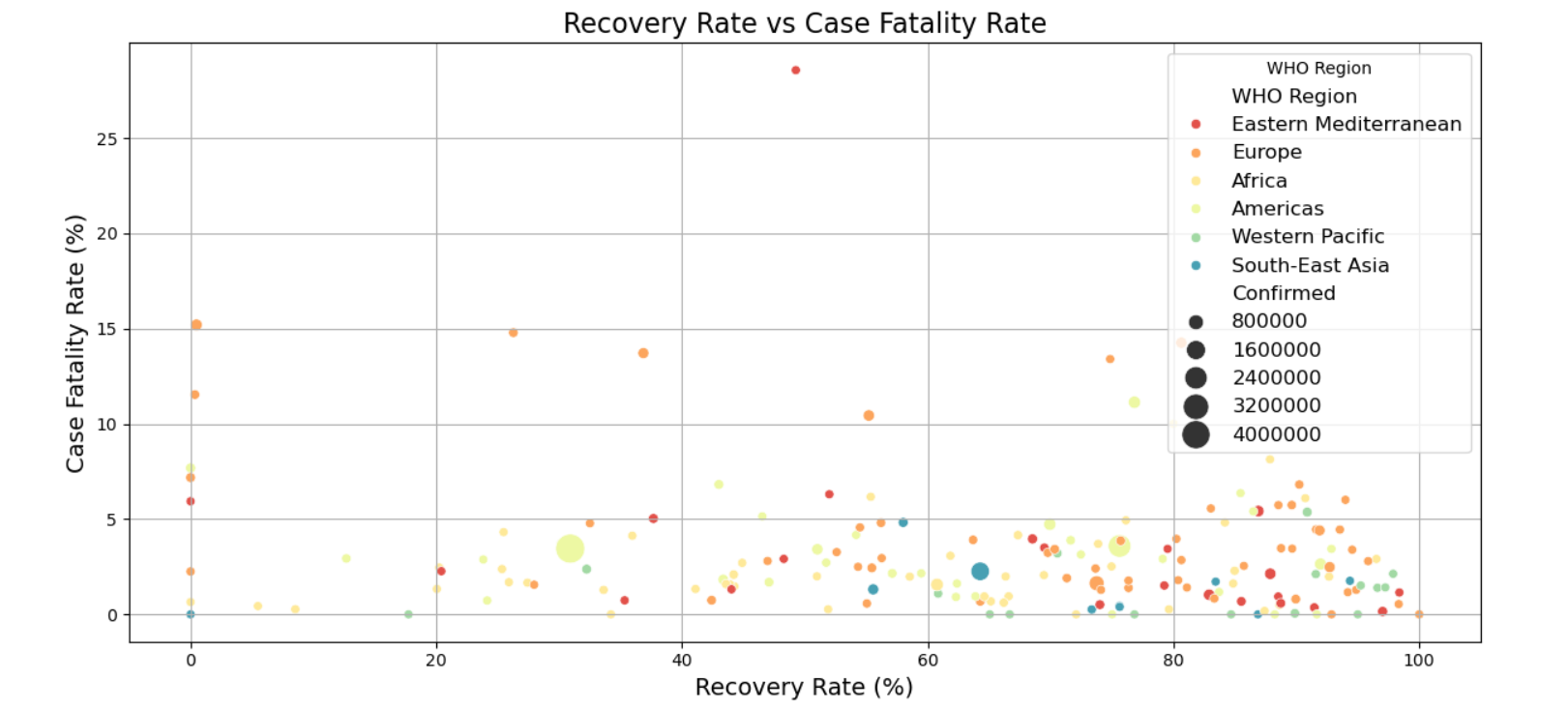
plt**.**ylabel('Case Fatality Rate (%)', fontsize**=**14)

plt**.**legend(title**=**"WHO Region", fontsize**=**12)

plt**.**grid(**True**)

plt**.**tight\_layout()

plt**.**show()



*# Plotly Choropleth Map: Covid-19 Cases by Country*

fig **=** px**.**choropleth(df, locations**=**"Country/Region", locationmode**=**'country names', color**=**"Confirmed",

hover\_name**=**"Country/Region", color\_continuous\_scale**=**"Viridis",

labels**=**{'Confirmed': 'Confirmed Cases'}, title**=**"Covid-19 Confirmed Cases by Country")

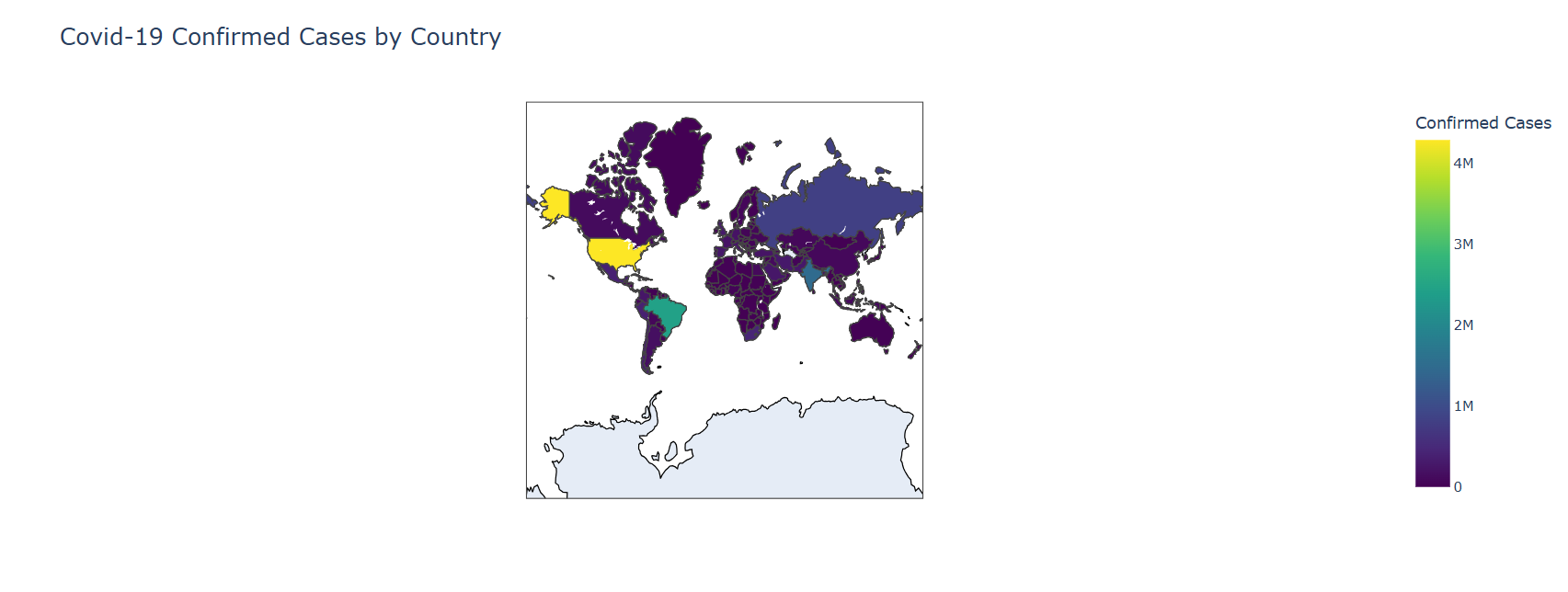
fig**.**update\_geos(showcoastlines**=True**, coastlinecolor**=**"Black", projection\_type**=**"mercator")

fig**.**update\_layout(title\_font\_size**=**20)

fig**.**show()

01M2M3M4MConfirmed CasesCovid-19 Confirmed Cases by Country

In [33]:



*# Plotting: 1 Week % Increase in Confirmed Cases for Top 10 Countries*

top\_10\_countries **=** df**.**nlargest(10, 'Confirmed')

plt**.**figure(figsize**=**(14, 8))

sns**.**lineplot(x**=**'Country/Region', y**=**'1 week % increase', data**=**top\_10\_countries, marker**=**'o', color**=**'orange')

plt**.**title('1 Week % Increase in Confirmed Cases (Top 10 Countries)', fontsize**=**16)

plt**.**xlabel('Country/Region', fontsize**=**14)

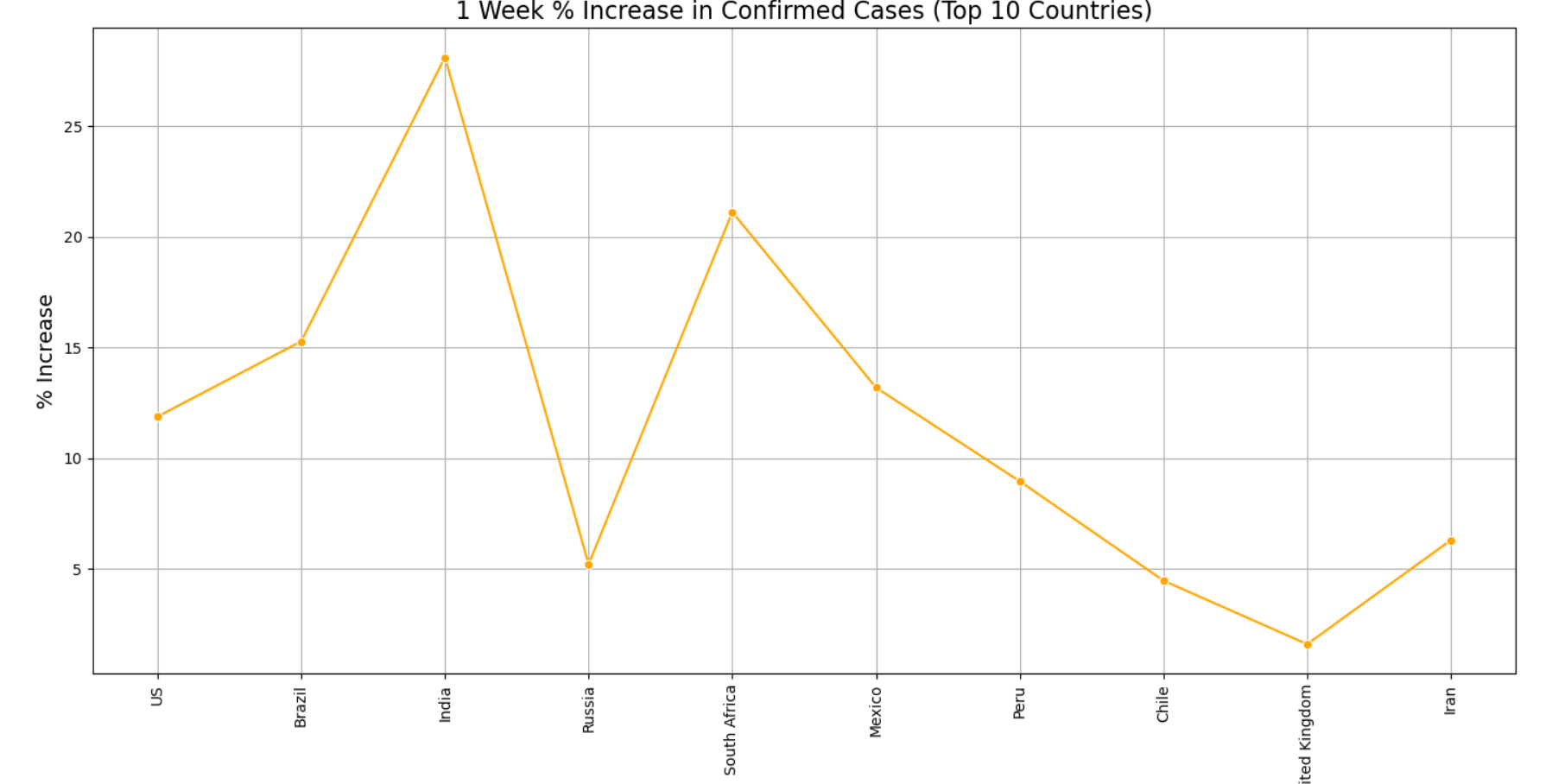
plt**.**ylabel('% Increase', fontsize**=**14)

plt**.**xticks(rotation**=**90)

plt**.**grid(**True**)

plt**.**tight\_layout()

plt**.**show()



*# Distribution of Confirmed Cases with Custom Bins and KDE*

plt**.**figure(figsize**=**(12, 6))

sns**.**histplot(df['Confirmed'], kde**=True**, color**=**'darkblue', bins**=**25)

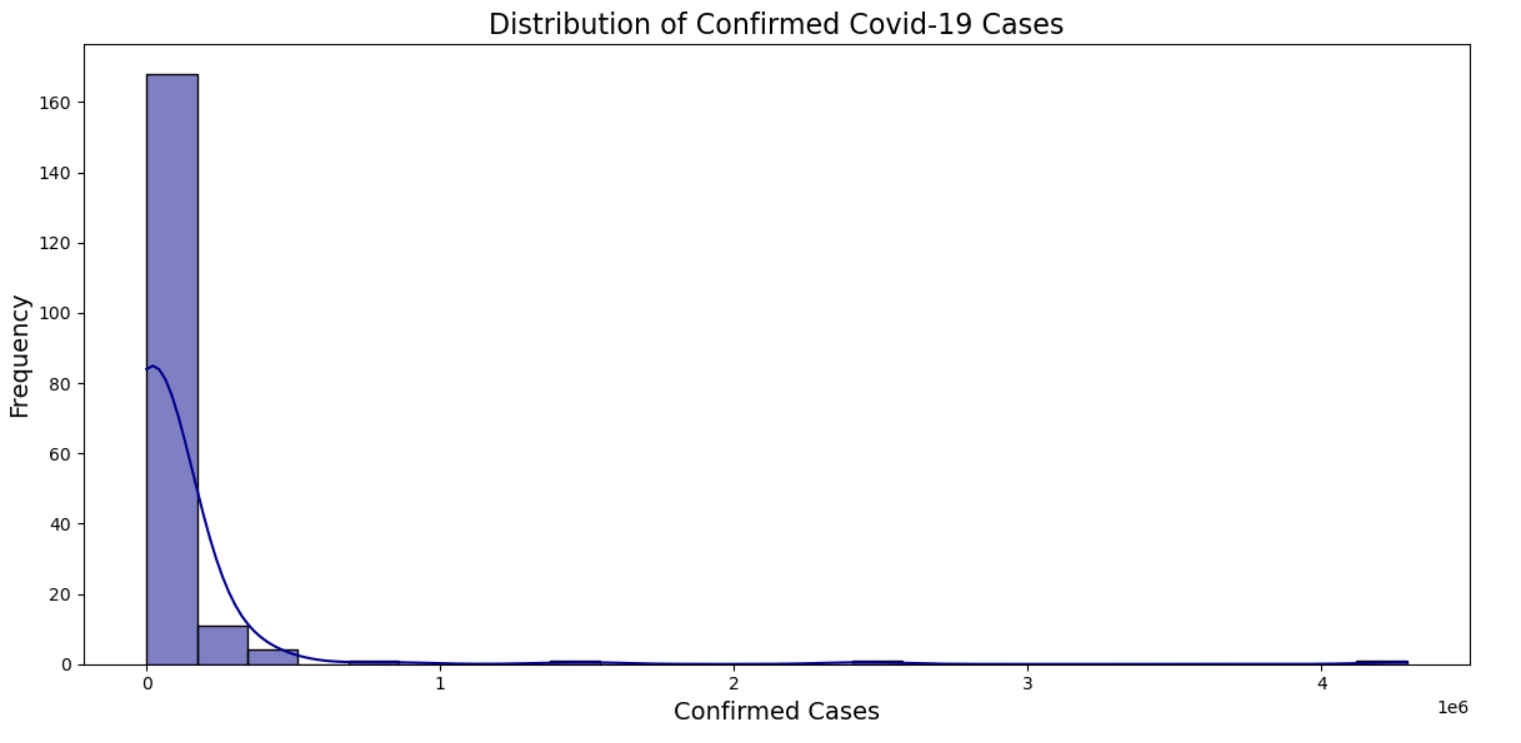
plt**.**title('Distribution of Confirmed Covid-19 Cases', fontsize**=**16)

plt**.**xlabel('Confirmed Cases', fontsize**=**14)

plt**.**ylabel('Frequency', fontsize**=**14)

plt**.**tight\_layout()

plt**.**show()



# Statistical Analysis: Correlation Heatmap

corr\_matrix = df[['Confirmed', 'Deaths', 'Recovered', 'Active']].corr()

plt.figure(figsize=(10, 6))

sns.heatmap(corr\_matrix, annot=True, cmap="Blues", fmt=".2f")

plt.title('Correlation Heatmap of Key Covid-19 Metrics', fontsize=16)

plt.tight\_layout()

plt.show()

