ANALYZING GLOBAL CLIMATE CHANGE DATA FOR POLICY INSIGHTS

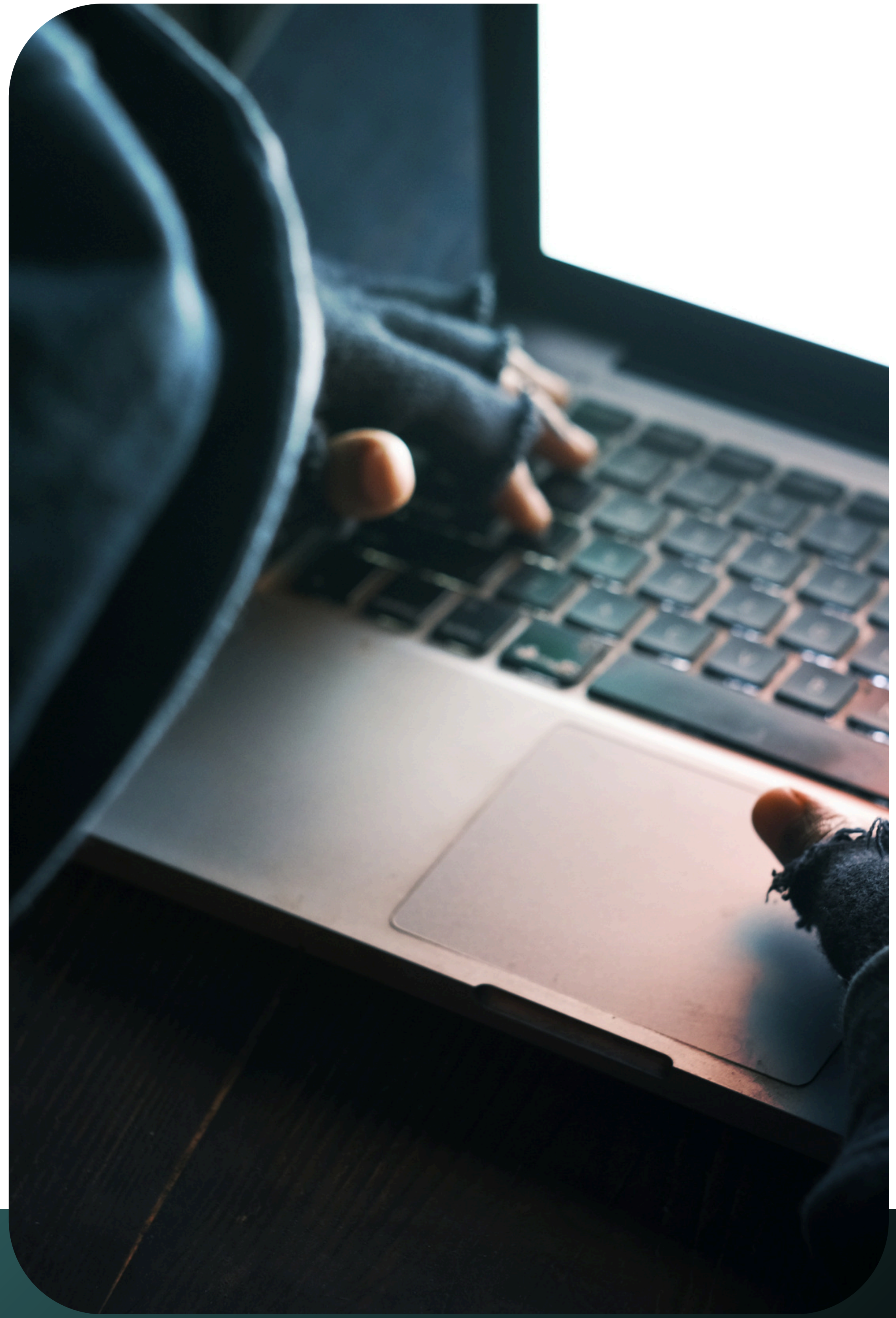
BY TEAM BUG BUSTERS

Analyzing Global Climate Change Data for Policy Insights ..

in this track, you'll work with a dataset on global climate change indicators from Kaggle.
This task simulates how data analysts in environmental organiza ons or government agencies explore data to inform policies that combat climate change and build resilience.
Your role: Act as a data consultant for an interna onal environmental agency (e.g., like the UN Environment Programme). The agency needs ac onable insights from this dataset to recommend policies that reduce climate risks and promote sustainability. Focus on exploratory data analysis (EDA) to uncover pa erns, trends, and rela onships and no machine learning models required. This keeps the emphasis on understanding the data deeply and crea vely.

Problem Statement

- Climate change is accelerating 🌍
- Rising temperatures, extreme weather, and environmental degradation affect billions
- Policymakers need evidence-based insights to guide interventions
- Our Goal: Analyze climate datasets to uncover patterns, risks, and solutions



Objectives

01 Perform EDA to explore climate indicators

02 Develop 10+ guiding EDA questions

03 Generate 5–7 actionable insights

04 Translate into 3–5 policy recommendations

05 Present findings for decision-making

Dataset Overview

The dataset provides climate-related metrics across countries and years.

Key columns include: □ Year: The year of the data point. □ Country: The country or region. □ Average Temperature (°C): Average annual temperature. □ CO_2 Emissions (Tons/Capita): Per-person carbon dioxide emissions. □ Sea Level Rise (mm): Annual Sea level increase. □ Rainfall (mm): Total annual rainfall. □ Popula on: Total popula on. □ Renewable Energy (%): Percentage of energy from renewable sources. □ Extreme Weather Events: Number of extreme events (e.g., storms, floods). □ Forest Area (%): Percentage of land covered by forests.

Dataset Link - h ps://www.kaggle.com/datasets/bhadramohit/climate-change-dataset The dataset may contain inconsistencies (e.g., outliers in temperatures or popula ons). treat this as part of real-world data challenge.

Analysis Framework: Adapted EDA Roadmap

To structure your work like a professional data project, follow this simple framework (inspired by real-world analy cs processes like those used in consul ng firms). It provides a clear path but leaves room for your innova ve approaches:

1. Data Understanding: Load the dataset, review its structure (e.g., data types, missing values), and summarize basic sta s cs (e.g., means, distribu ons). Iden fy any anomalies or cleaning needs.

2. Data Prepara on: Handle issues like duplicates, outliers, or inconsistencies. For example, normalize units if needed or group data by regions/years for be er analysis. Keep it lightweight—focus on enabling explora on.

Analysis Framework: Adapted EDA Roadmap

3. Ques on Formula on and Explora on: Brainstorm and select at least 10 EDA ques ons. These should cover univariate (single variable), bivariate (rela onships between two), and mul variate (mul ple variables) analyses. Use sta s cal methods (e.g., correla ons, trends) and visualiza ons (e.g., histograms, sca er plots, heatmaps) to answer them. Be crea ve: Think about geographic or temporal angles.

4. Insight Genera on: From your EDA, extract 5-7 key insights. Link them to policy implications, e.g., "Countries with higher renewable energy adop on show lower emission growth - recommend incen ves for solar/wind investments."

5.Policy Recommenda ons and Presenta on: Translate insights into 3-5 policy proposals. Prepare a presenta on that tells a story or if using PowerBI or tableau, share the dashboard file or link highligh ng how your analysis supports resilience-building.

Key Visualizations

Component	Content	Policy Implication
Insight 1: CO2 Drives Extreme Weather Risk	There is a **strong positive correlation ($r = +0.54$)** between **CO2 Emissions** and the occurrence of **Extreme Weather Events.**	**Focus Policy:** Direct and immediate emission reduction must be prioritized in high-risk zones to lower the frequency of climate shocks.
Insight 2: The Value of Natural Resilience	There is a measurable **negative correlation ($r = -0.26$)** between **Forest Area (%)** and **Extreme Weather Events.**	**Focus Policy:** Incentivize forest area expansion as a critical, natural buffer against climate shocks. This offers dual benefits: mitigation and protection.

Key Visualizations

Component	Content	Policy Implication
Insight 3: The Climate Equity Paradox	A counter-intuitive **negative correlation ($r = -0.57$)** exists between **Avg Temperature** and **CO_2 Emissions**, meaning high-temperature, vulnerable countries are often low-emission contributors.	**Focus Policy:** Adaptation funding must prioritize vulnerable, high-temperature, low-emission countries (Climate Justice).
Insight 4: Ineffective Renewable Transition	A weak **positive correlation ($r = +0.27$)** exists between **Renewable Energy (%)** and **CO_2 Emissions**, suggesting new renewables are *not* reliably displacing fossil fuels.	**Focus Policy:** Policy must shift from *installation targets* to *mandates* that actively displace fossil fuels to achieve net reduction.

Policy Recommendations

Proposal 1: The 'Emissions Displacement Mandate' (Addresses Insight 1 & 2)
Action: Impose a Global Carbon Efficiency Tax on the Top 10 Polluting Countries (from your Q6 Bar Plot). This tax is triggered only if their increase in Renewable Energy (%) over a 5-year period fails to reduce their CO2 Emissions (Tons/Capita).

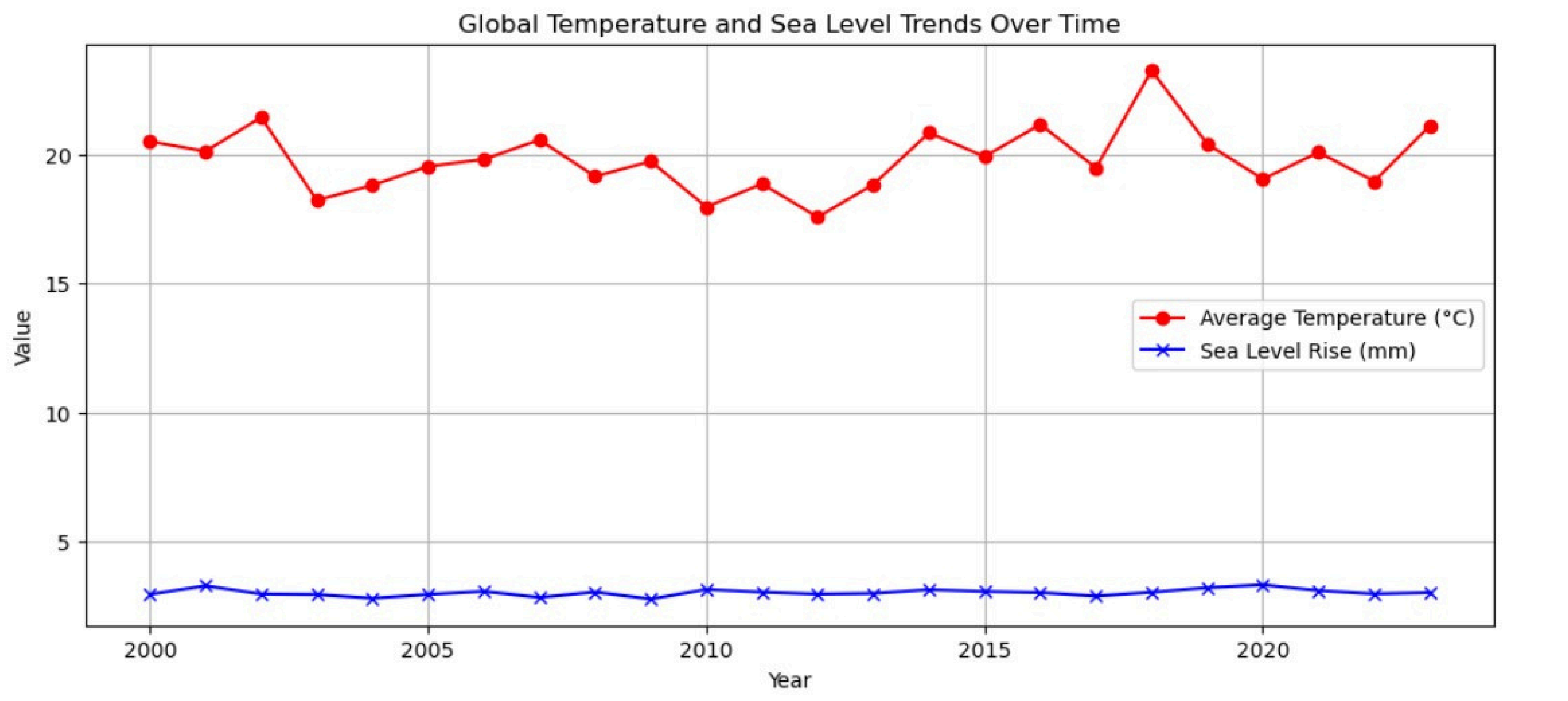
Rationale: Uses the weak correlation (r=+0.27) to force renewables to actively displace, rather than simply add to, the energy supply, directly addressing the concentrated risk (r=+0.54).

Proposal 2: The 'Resilience for Conservation' Fund (Addresses Insight 3)
Action: Create a dedicated Natural Capital Resilience Fund where access to adaptation aid is strictly conditioned on a nation’s commitment to increasing its Forest Area (%) by a minimum of 1% per year.

Rationale: Levers the measurable negative correlation (r=−0.26) to ensure resilience funding delivers dual benefits: ecosystem health and protection from extreme weather events.

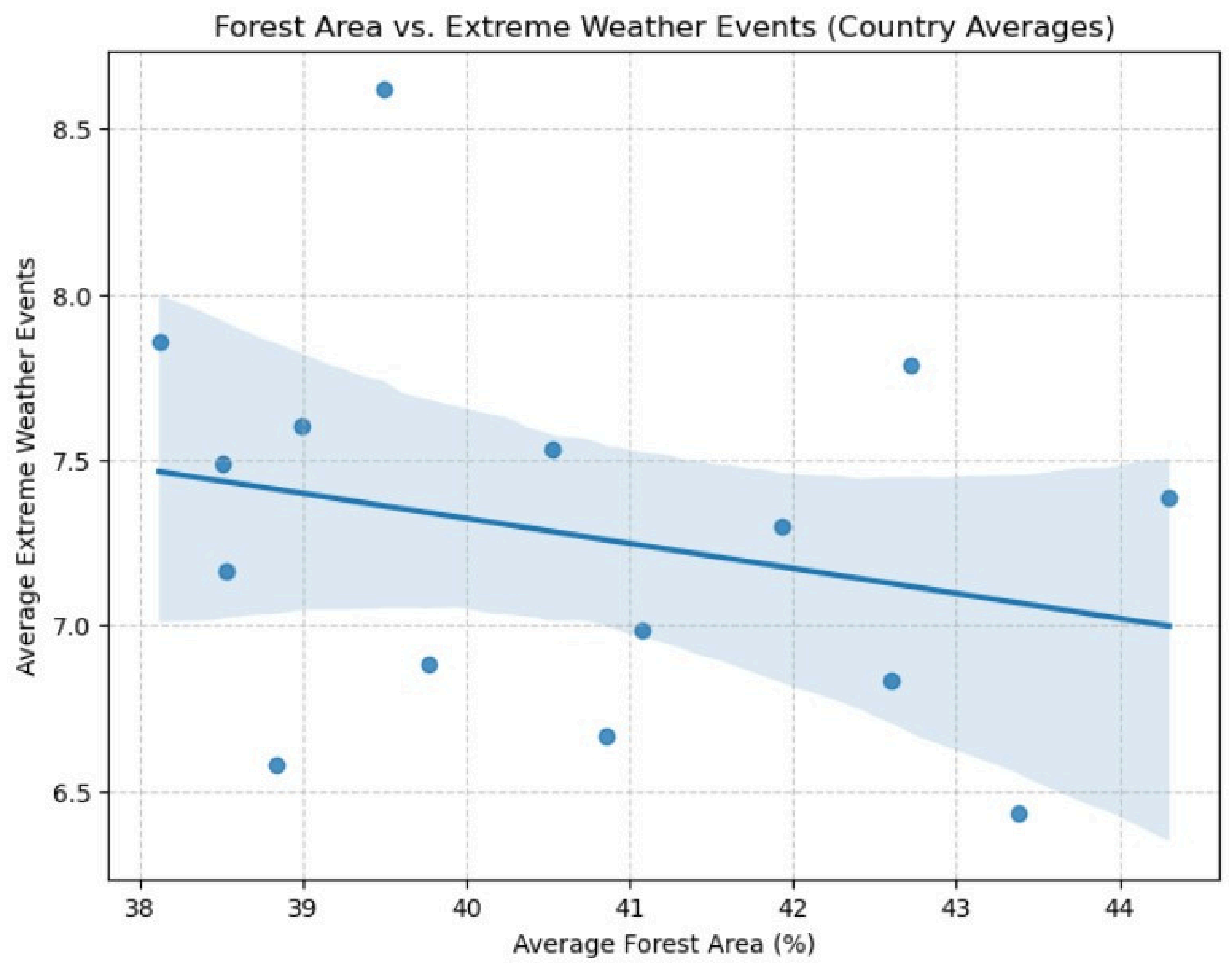
EDA QUESTIONS FORMULATION..

```
# Question 1: Global Warming Trajectory .

# 1. Global Warming Trajectory
plt.figure(figsize=(12, 5))
plt.plot(yearly_data['Year'], yearly_data['Avg Temperature (°C)'],
         label='Average Temperature (°C)', marker='o', color='red')
plt.plot(yearly_data['Year'], yearly_data['Sea Level Rise (mm)'],
         label='Sea Level Rise (mm)', marker='x', color='blue')
plt.title('Global Temperature and Sea Level Trends Over Time')
plt.xlabel('Year')
plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.savefig('1_Global_Warming_Trajectory.png')
plt.show()
```



EDA QUESTIONS FORMULATION..

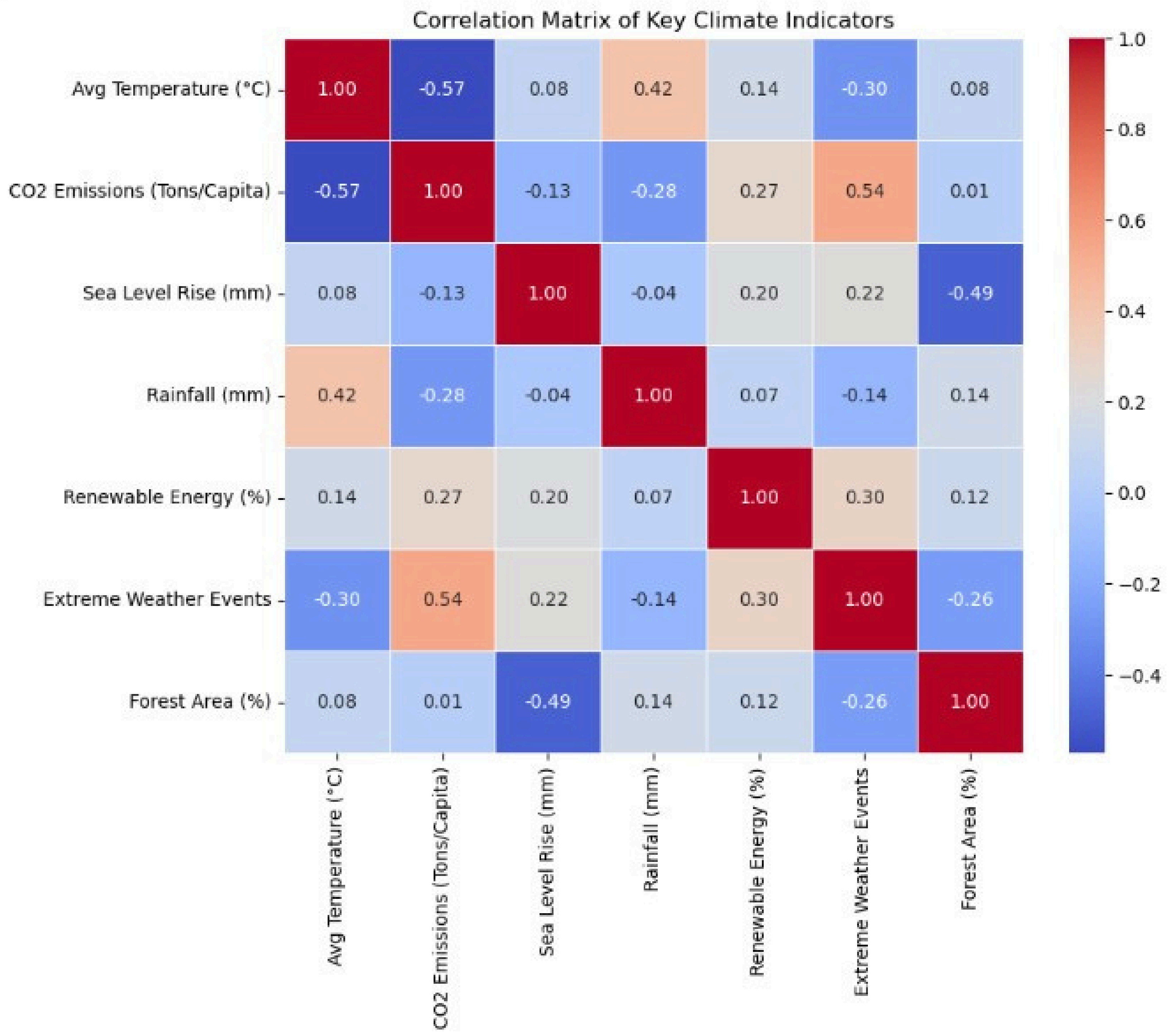
```python
#  Question 4: Ecosystem Resilience

# 4. Ecosystem Resilience (Bivariate Regression Plot)
plt.figure(figsize=(8, 6))
sns.regplot(
    x='Forest Area (%)',
    y='Extreme Weather Events',
    data=country_data
)
plt.title('Forest Area vs. Extreme Weather Events (Country Averages)')
plt.xlabel('Average Forest Area (%)')
plt.ylabel('Average Extreme Weather Events')
plt.grid(True, linestyle='--', alpha=0.6)
plt.savefig('4_Forest_vs_Extreme_Events_Scatter.png')
plt.show()
```



EDA QUESTIONS FORMULATION..

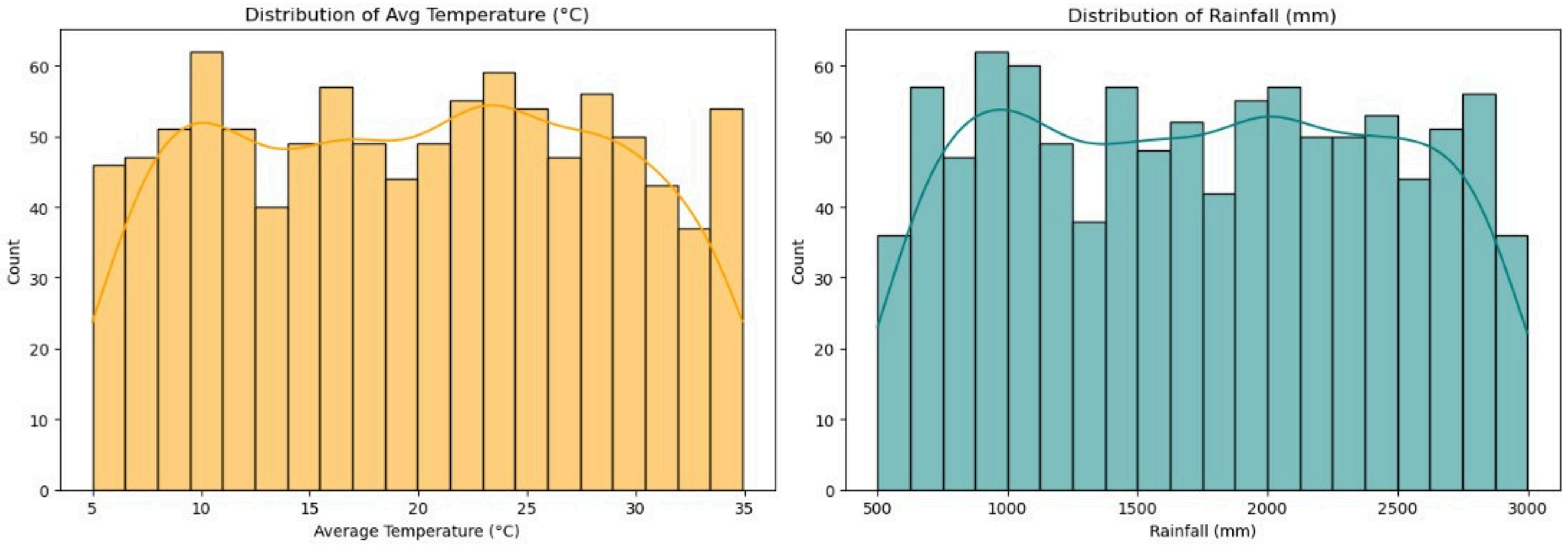
```
# Question 5: Overall Risk Factors

# 5. Overall Risk Factors (Correlation Heatmap)
# Calculate the correlation matrix
# Question 5: Overall Risk Factors

# 5. Overall Risk Factors (Correlation Heatmap)
cols = ['Avg Temperature (°C)', 'CO2 Emissions (Tons/Capita)',
        'Sea Level Rise (mm)', 'Rainfall (mm)',
        'Renewable Energy (%)', 'Extreme Weather Events',
        'Forest Area (%)']

# Calculate the correlation matrix
correlation_matrix = country_data[cols].corr()

plt.figure(figsize=(9, 7))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
plt.title('Correlation Matrix of Key Climate Indicators')
plt.savefig('5_Correlation_Heatmap.png')
plt.show()
```

```
plt.show()
```



EDA QUESTIONS FORMULATION..

```
# Question 10: Temperature & Rainfall Distribution

fig, axes = plt.subplots(1, 2, figsize=(14, 5))

# Histogram 1: Temperature
sns.histplot(df['Avg_Temp_C'], bins=20, kde=True, ax=axes[0], color='orange')
axes[0].set_title('Distribution of Avg Temperature (°C)')
axes[0].set_xlabel('Average Temperature (°C)')
axes[0].set_ylabel('Count')

# Histogram 2: Rainfall
sns.histplot(df['Rainfall (mm)'], bins=20, kde=True, ax=axes[1], color='teal')
axes[1].set_title('Distribution of Rainfall (mm)')
axes[1].set_xlabel('Rainfall (mm)')
axes[1].set_ylabel('Count')

plt.tight_layout()
plt.savefig('10_Temp_Rainfall_Distributions.png')
plt.show()
```



[ ]:

Conclusion :- Policy for a Resilient Future

Key Takeaway :- Our analysis shows that policies must be targeted and multi-functional: linking emission reduction directly to extreme weather risk, and tying adaptation aid to natural conservation.

Call to Action :- We urge the UN Environment Programme to integrate the following principles into its next 5-year climate strategy: 1. Mandate Displacement over mere adoption of renewables. 2. Condition Aid on conservation and resilience efforts. 3. Prioritize Equity by transferring resources to the most vulnerable populations.