Carbon Emissions Worldwide data analysis project

Business Problem

The aim of this exploratory analysis is to investigate historical carbon dioxide (CO_2) emission levels and their relationship with global temperature changes using reliable datasets from 1958 to 2023. The insights will inform future simulations and policy recommendations.

- How much the Earth is warming?
- What could happen in the future?
- How can we stop or slow it?

Summary Statistics

	count	mean	std	min	25%	50%	75%	max
ObjectId	1570.0	785.500000	453.364276	1.0	393.25	785.500	1177.7500	1570.0
Value	1570.0	180.716153	180.554706	-0.1	0.45	313.835	355.0725	424.0

	count	mean	std	min	25%	50%	75%	max
ObjectId	225.0	113.000000	65.096083	1.000	57.00000	113.0000	169.00000	225.000
F1961	188.0	0.163053	0.405080	-0.694	-0.09700	0.0645	0.31850	1.892
F1962	189.0	-0.013476	0.341812	-0.908	-0.16400	-0.0560	0.11400	0.998
F1963	188.0	-0.006043	0.387348	-1.270	-0.20550	-0.0030	0.23050	1.202
F1964	188.0	-0.070059	0.309305	-0.877	-0.23650	-0.0560	0.13250	1.097
...	...	...	...	...	...	...	...	...
F2018	213.0	1.302113	0.596786	0.238	0.86500	1.1250	1.83400	2.772
F2019	213.0	1.443061	0.467510	0.050	1.16900	1.4120	1.69800	2.689
F2020	212.0	1.552038	0.621930	0.229	1.16175	1.4770	1.82625	3.691
F2021	213.0	1.343531	0.484692	-0.425	1.01900	1.3270	1.62900	2.676
F2022	213.0	1.382113	0.669279	-1.305	0.87800	1.3150	1.91800	3.243

63 rows × 8 columns

Data Preparation

- Reshaped df_temp from wide to long format for year-wise analysis.
- Converted df_co2['Date'] from "1958M03" to proper datetime format.
- Aggregated CO_2 monthly data into yearly averages for comparability with temperature data.
- Handled missing values and filtered relevant rows like 'World' for global trend lines.

2: Assumptions

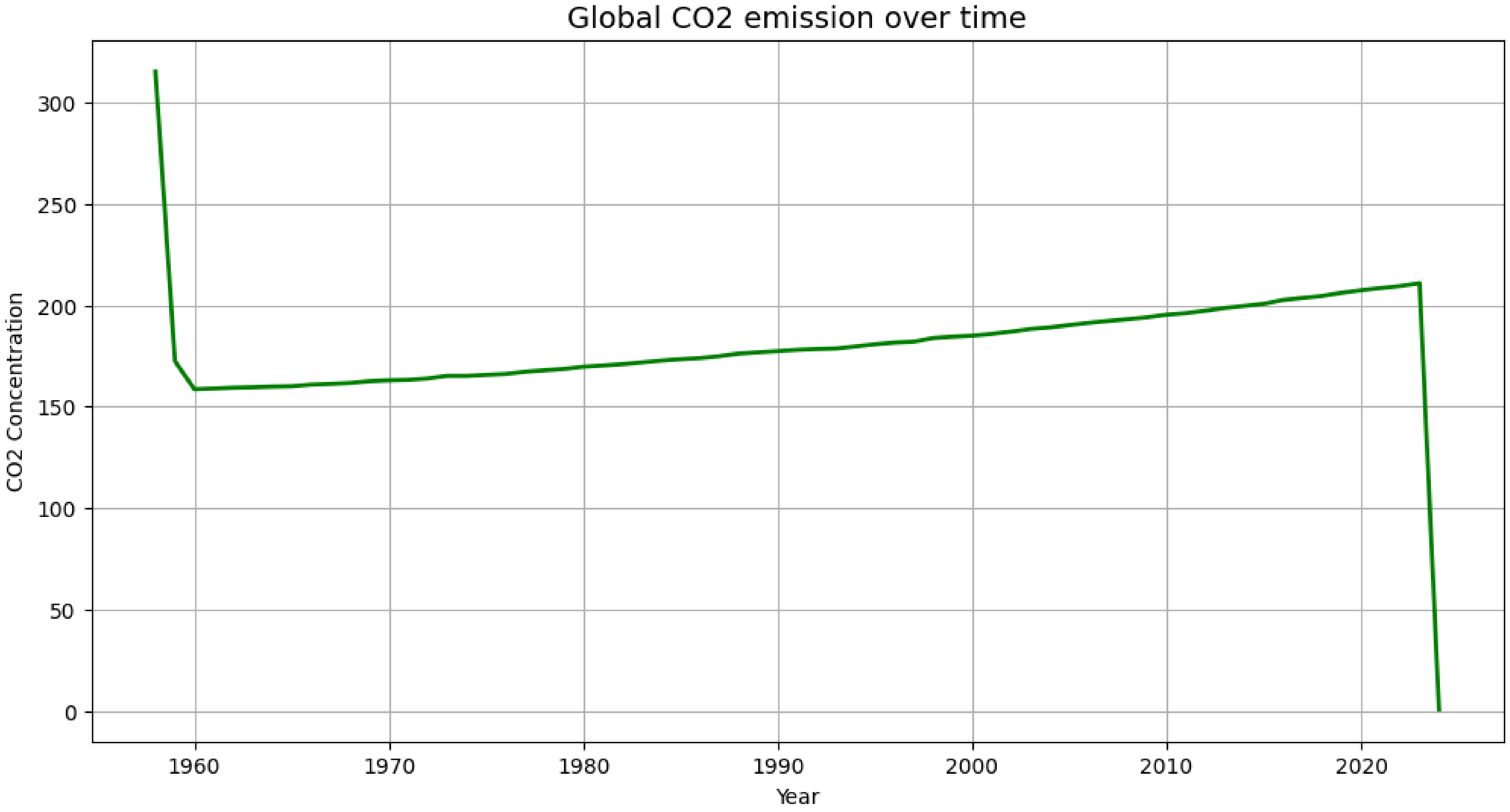
- Global CO_2 levels and temperature are connected.
- Past trends can help us understand the future.
- Data is correct and clean.

3. Research Questions

- How have CO_2 emissions changed from 1958 to 2023?
- How has global temperature changed in the same period?
- Is there a strong relationship between CO_2 and temperature?
- Which countries are heating the most?
- How did temperature patterns change after 1990?
- Is there a statistical correlation between CO_2 and temperature?
- Has warming accelerated after 1990?
- What if CO_2 is reduced by 30% by 2050?
- What if CO_2 keeps rising at current rate till 2050?
- What if CO_2 increases twice as fast?

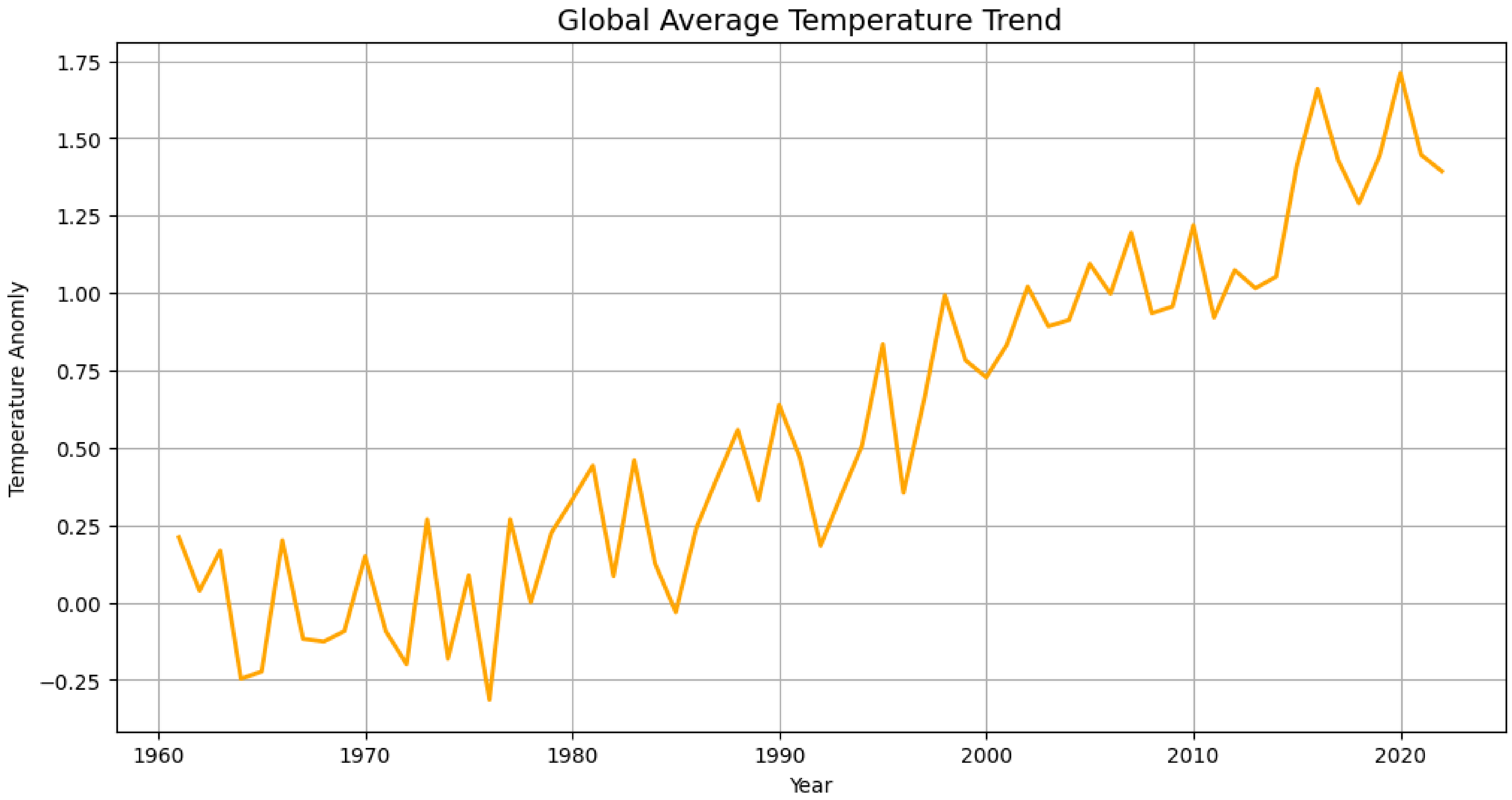
Step 3: Research Questions

1. **How have CO_2 emissions changed from 1958 to 2023?**



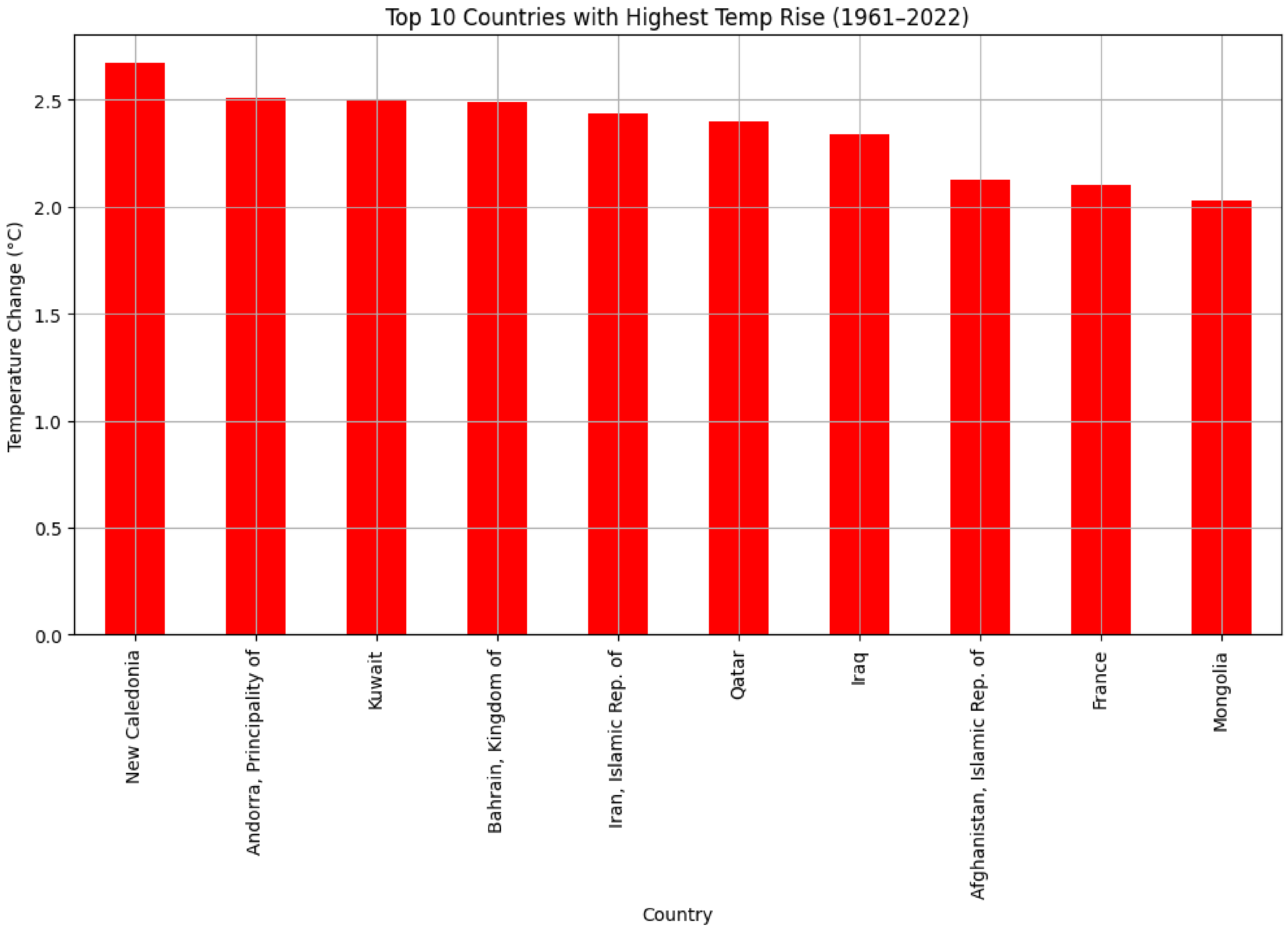
The green line shows that CO_2 levels have gone up steadily from 1960 to 2023. There was a small dip around 2020–2021, likely because of COVID-19 lockdowns, but emissions went back up afterward.

2. How has global temperature changed in the same period?



The orange line shows that the Earth is getting hotter over time. Even though there are ups and downs in some years, the overall trend is clearly rising, especially after the year 2000.

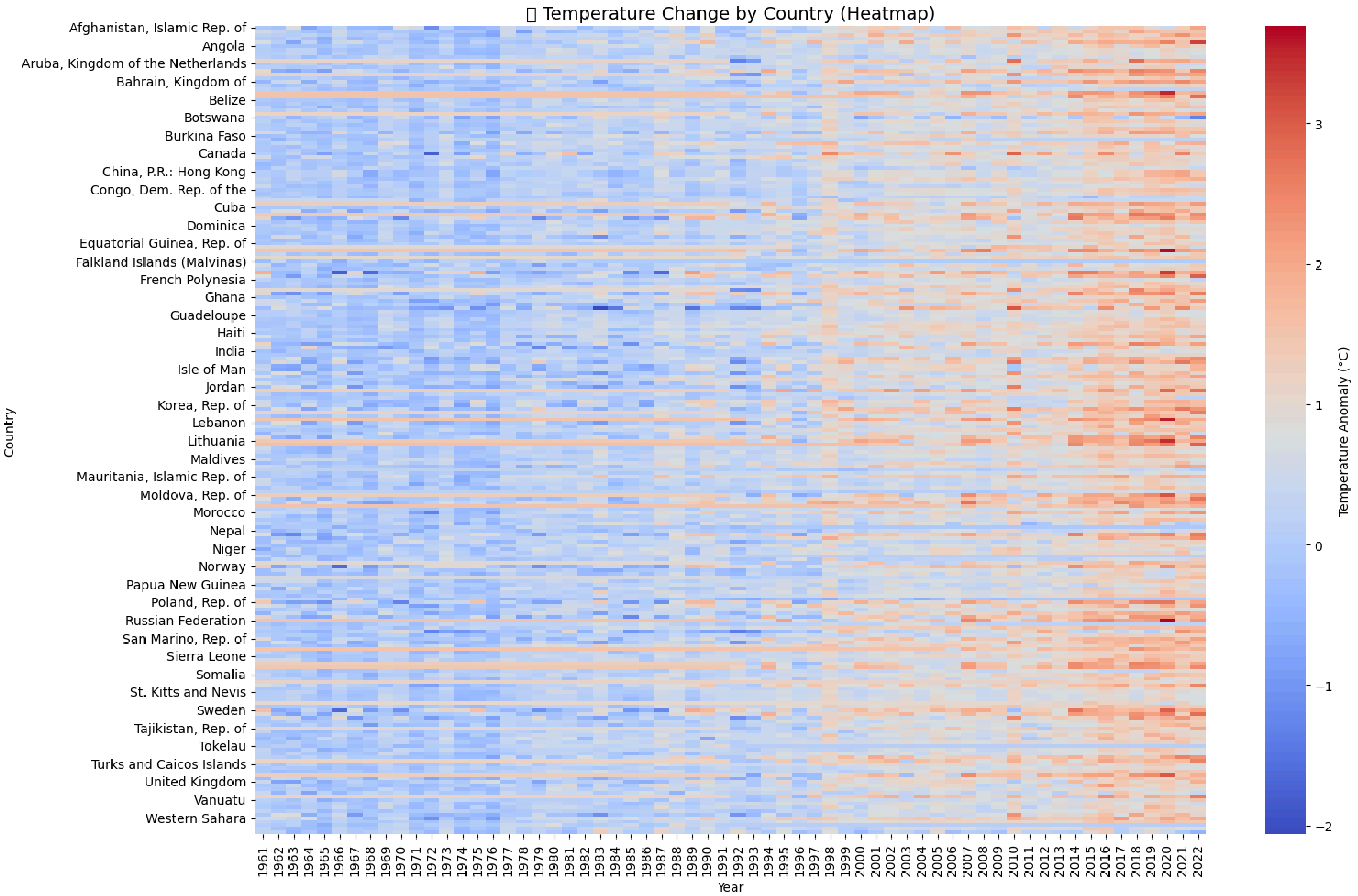
3. Which countries are heating the most?



Based on bar chart, New Caledonia showed the highest temperature rise from 1961 - 2022 and Other top countries also showed significant increases, like parts of Europe and Oceania.

4. How did temperature patterns change after 1990?

This heatmap shows how temperatures have changed in different countries from 1961 to 2022.
Blue means cooler years,
Red means hotter years
Most countries are turning red over time, showing that global warming is affecting many parts of the world.



5.Is there a statistical correlation between CO_2 and temperature?

```
# Pearson Correlation
corr, p_value = pearsonr(df_merged['Avg_CO2_ppm'], df_merged['Temperature'])
print(f"Pearson Correlation: {corr:.4f}, P-value: {p_value:.4e}")
```

```
Pearson Correlation: 0.9430, P-value: 2.2883e-30
```

Pearson correlation coefficient between global CO_2 concentration and average temperature anomaly is 0.943, with a p-value < 0.001, indicating a strong and statistically significant positive relationship between CO_2 emissions and global warming.

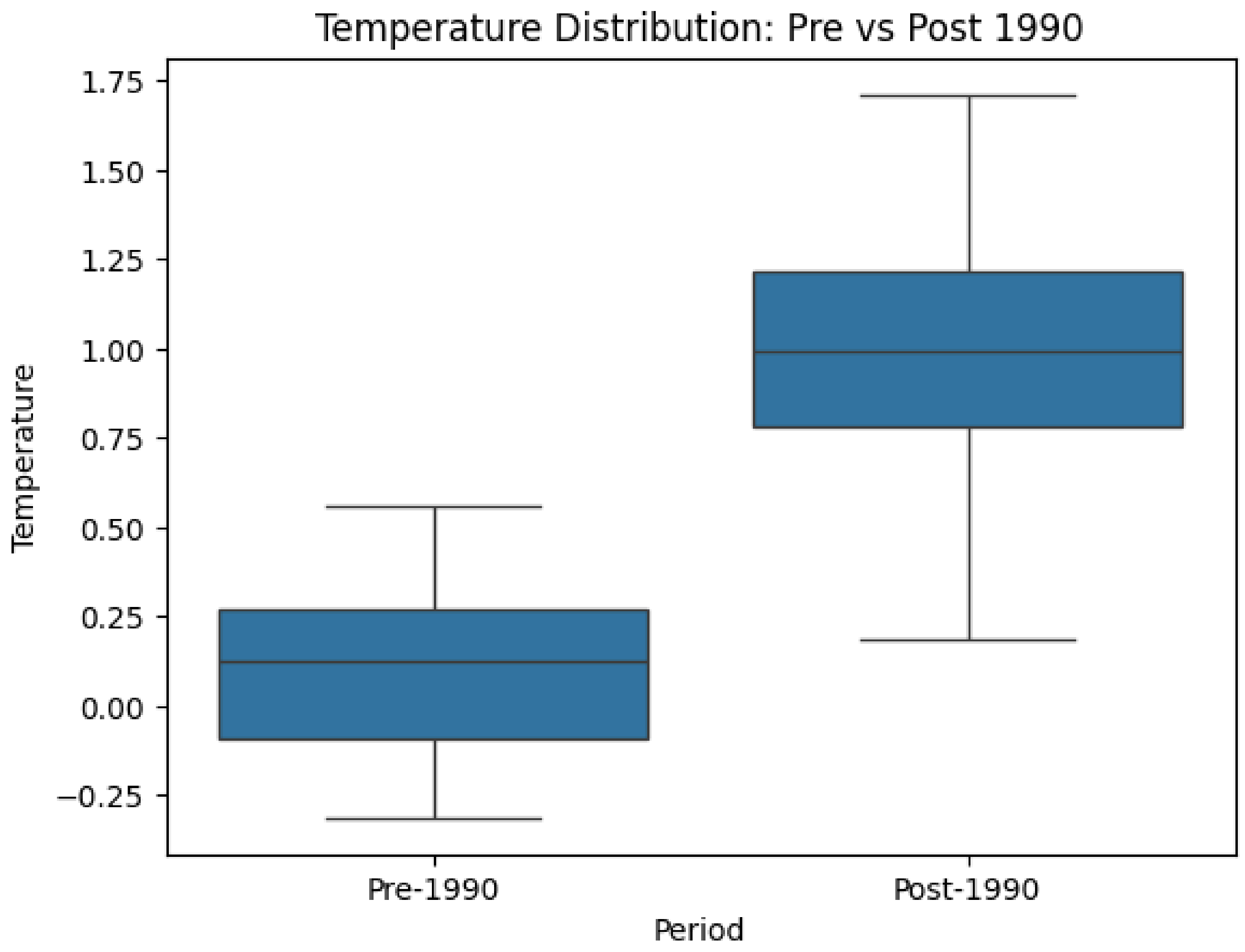
This is a very strong positive correlation — as CO_2 increases, so does temperature.

6. Has warming accelerated after 1990?

```
# Split temp data pre and post 1990
pre_1990 = df_temp_world[df_temp_world['Year'] < 1990]['Temperature']
post_1990 = df_temp_world[df_temp_world['Year'] >= 1990]['Temperature']

# Perform t-test
t_stat, p_val = ttest_ind(post_1990, pre_1990, equal_var=False)
print(f"T-Statistic: {t_stat:.4f}, P-value: {p_val:.4e}")
```

```
T-Statistic: 11.1855, P-value: 9.8266e-16
```



A two-sample t-test was conducted to compare global temperatures before and after 1990.

- T-Statistic: 11.1855
- P-Value: 9.8266e-16

These results indicate a highly significant difference between the two periods. The extremely low p-value (much less than 0.05) strongly suggests that the observed increase in temperature after 1990 is not due to random chance, but reflects a real shift—likely due to climate change.

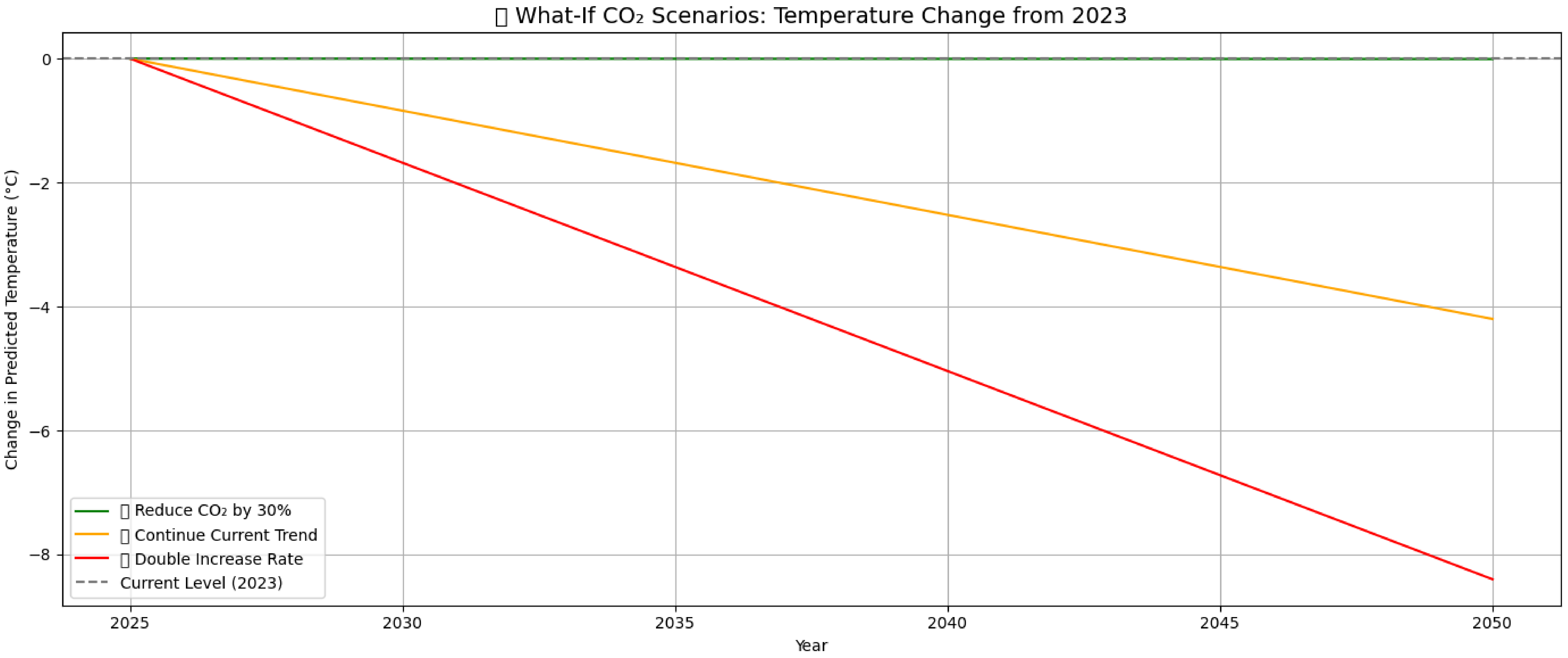
The boxplot visualization supports this finding, showing that:

Temperatures before 1990 were centered around 0 with low variation.

Post-1990 temperatures are consistently higher, with a clear upward shift in both the median and range.

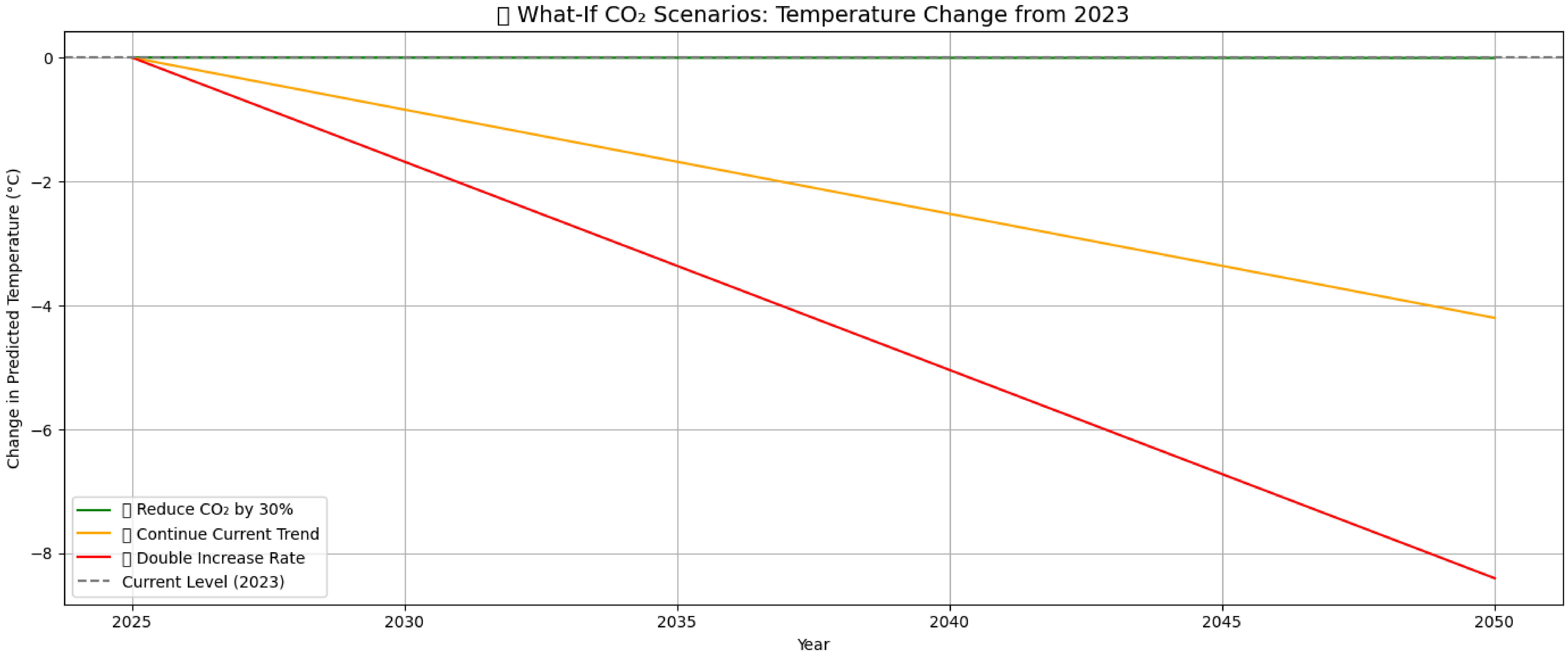
- This linear regression model analyzes the relationship between atmospheric CO_2 levels and global temperature. The model's R-squared value of 0.889 indicates that approximately 88.9% of the variation in global temperature can be explained by CO_2 concentration alone — reflecting a very strong relationship.
- The P-value for CO_2 is 0.000, confirming that the relationship is highly statistically significant and not due to random chance.
- The T-statistic of 21.955 for the CO_2 coefficient further supports the strong impact of CO_2 on temperature.
- Additionally, the model has a high F-statistic of 482.0, indicating that the regression as a whole is highly significant.
- The Durbin-Watson statistic is 1.937, suggesting that there is no autocorrelation in the residuals, which validates the model's reliability.
- The AIC (-33.22) and BIC (-28.96) scores show that this model performs well compared to alternative models.

7. What if CO_2 is reduced by 30% by 2050?



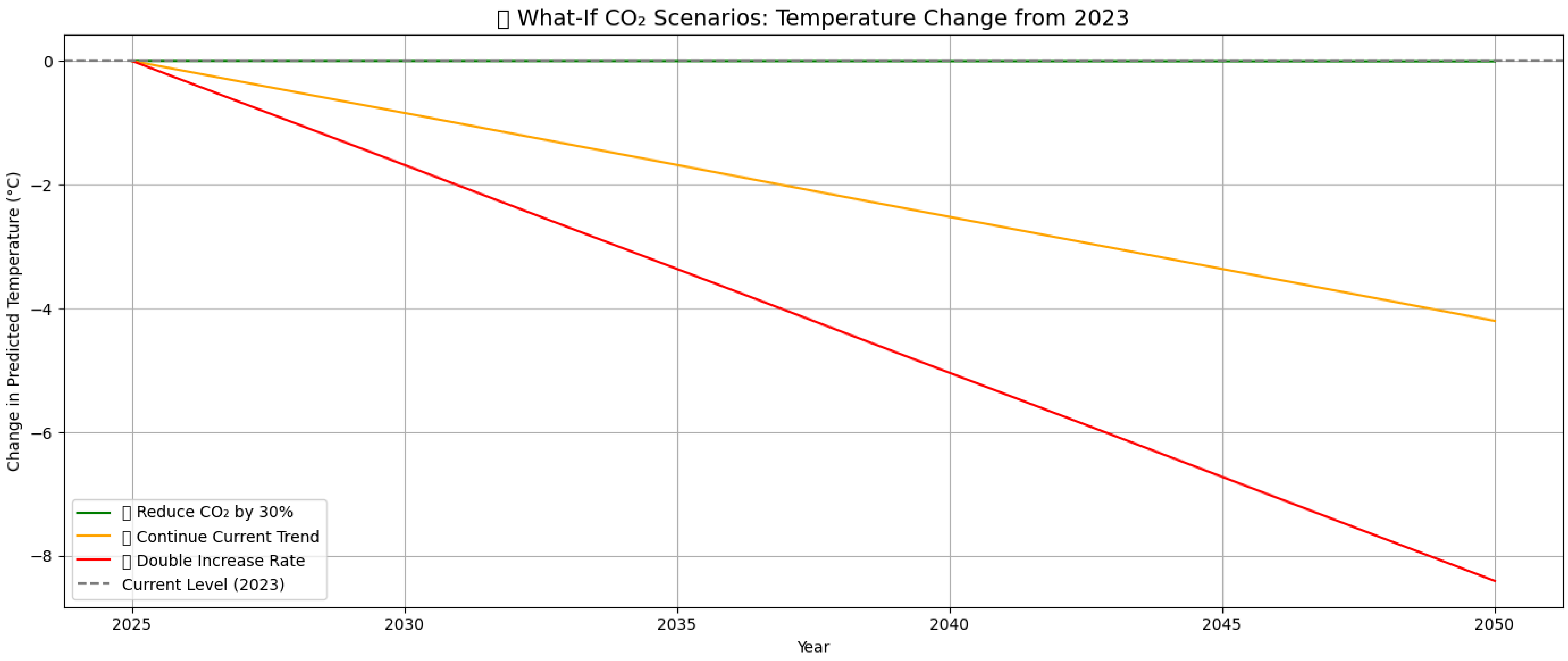
If we reduce CO_2 emissions by 30% by the year 2050, the rise in global temperature will almost stop.
This means the Earth's climate will become more stable, and we can avoid dangerous warming.
It is the best and safest option for the planet's future.

What if CO_2 keeps rising at current rate till 2050?



- If CO_2 keeps increasing at the same rate until 2050, the global temperature will continue to rise.
- By 2050, the Earth could become much hotter, leading to more heatwaves, rising sea levels, and extreme weath

What if CO_2 increases twice as fast?



- If CO_2 emissions grow twice as fast, the Earth's temperature will rise very quickly.
- By 2050, global temperatures could increase by up to 8°C, which is extremely dangerous.
- This could cause serious problems like ice melting, sea levels rising fast, crop failures, water shortages, and more deadly heatwaves.

Final Recommendations

- Reduce CO_2 emissions by at least 30% by 2050 to stabilize global temperatures and avoid dangerous climate thresholds.
- Adopt global climate policies such as carbon pricing, emissions caps, and international agreements like the Paris Accord.
- Invest in renewable energy sources (solar, wind, hydro) and phase out fossil fuels to accelerate the transition to a green economy.
- Enhance forest conservation and afforestation efforts to naturally absorb atmospheric CO_2.
- Promote global collaboration between governments, industries, and communities to ensure a united climate response.
- Encourage climate education and awareness, empowering individuals to adopt sustainable practices.
- Support continuous monitoring and research through data-driven models and scientific analysis to guide future decisions.