PROJECT 1

GLOBAL WARMING – FACT OR FICTION?

Danielle Dejean, Kaidon Kennedy, Carolyn Scheese, Harpreet ‘Monty’ Singh

PROJECT OVERVIEW & GOALS

- Find and examine various environmental data to see the impact of CO2 for evidence of global warming and possible correlations by countries and continents throughout the world.
  - Explore the **temperature and CO2 changes** of different **countries** and different **continents** over a 25-year span (1995 – 2020)
  - Average **temp over time** by country and continent
  - Examine **CO2** emissions as it **relates** to increases in **surface temperature**
  - Explore possible **predication models** for increases in average global temperature over various timelines; monthly, yearly, etc.

An image that represents an exploration of data from countries and continents across the world on CO2 and Temperatures across the world that can be used as a background in a PPT presentation; modern and professional style – Dall-E

DATA SETS

Kaggle – High Quality

Examined 4 data sets

Used 3 data sets
*excluded worldbank

No copywrite issues with this data

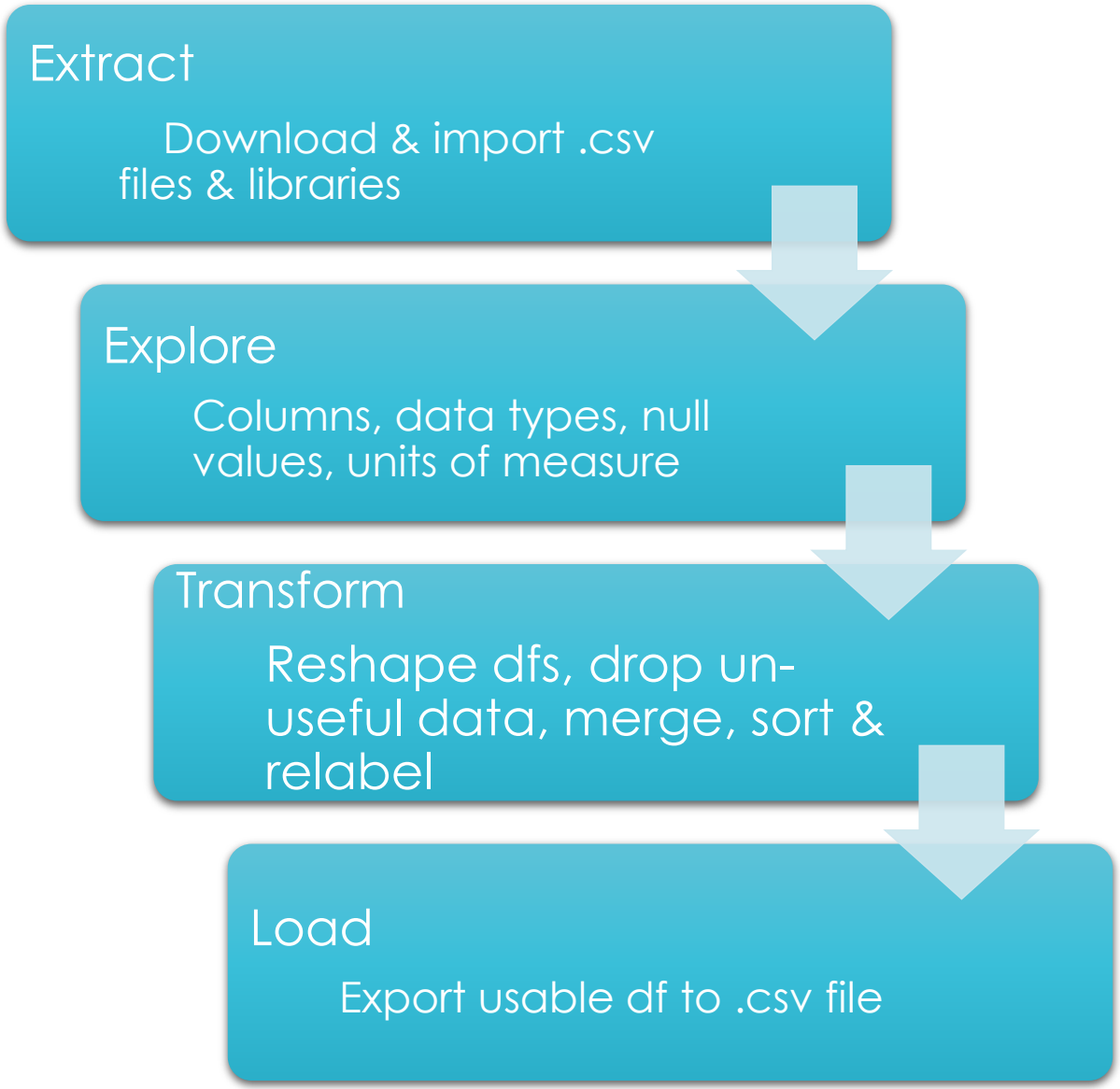
- Global Warming Trends (1961-2022)
- https://www.kaggle.com/datasets/jawadawan/global-warming-trends-1961-2022
- Country, ISO code, Year and Temperature
- Country wise surface temperature from 1961 to 2022

- Countries and their Continents
- https://www.kaggle.com/datasets/hserdaraltan/countries-by-continent

- Temperature (1960) to Temperature (2022): annual surface temp (C) from1960 to 2022
- Daily temperature of different cities with country and Region
- https://www.kaggle.com/datasets/subhamjain/temperature-of-all-countries-19952020
- Contains data of various major cities of different countries in the world.

- data.worldbank.org
- World Bank Open Data
- https://data.worldbank.org/indicator/EN.ATM.CO2E.PC?most_recent_year_desc=false
- Free and open access to global development data

DATA COLLECTION, CLEANUP & EXPLORATION



APPROACH

Extract & Explore

- Import pandas (matplotlib, datetime & prophet for charts & forecasting)
- Read in the .csv files
- Ensure data had usable values, overlapping column/years
- Check columns/data types

Challenges

- Reading in a file with a header

```
co2_df = pd.read_csv('co2_data.csv', header=2)
```

- Converting from long to wide format

```
temp_df = temp2_df.pivot(index='Country', columns='Year',
values='YearlyAvgTemperature')
```

- The first temperature file had suspect temperature units. Even though it had values, and the documentation indicated it was usable, we couldn't make sense of them. We had to disregard that data, and find another data source with known/usable units. In doing this we lost 5 years of data, and several countries, but were able to make sense of the results.

APPROACH

Transform & Load

- Drop unneeded data
- Rename columns
- Merge (inner)
- Output to .csv

Challenges

- Set ‘Country’ column to index to isolate and drop rows with NaN values

```
co2_df.set_index('Country', inplace=True)
co2_df.dropna(axis=0, how='all', inplace=True)
```

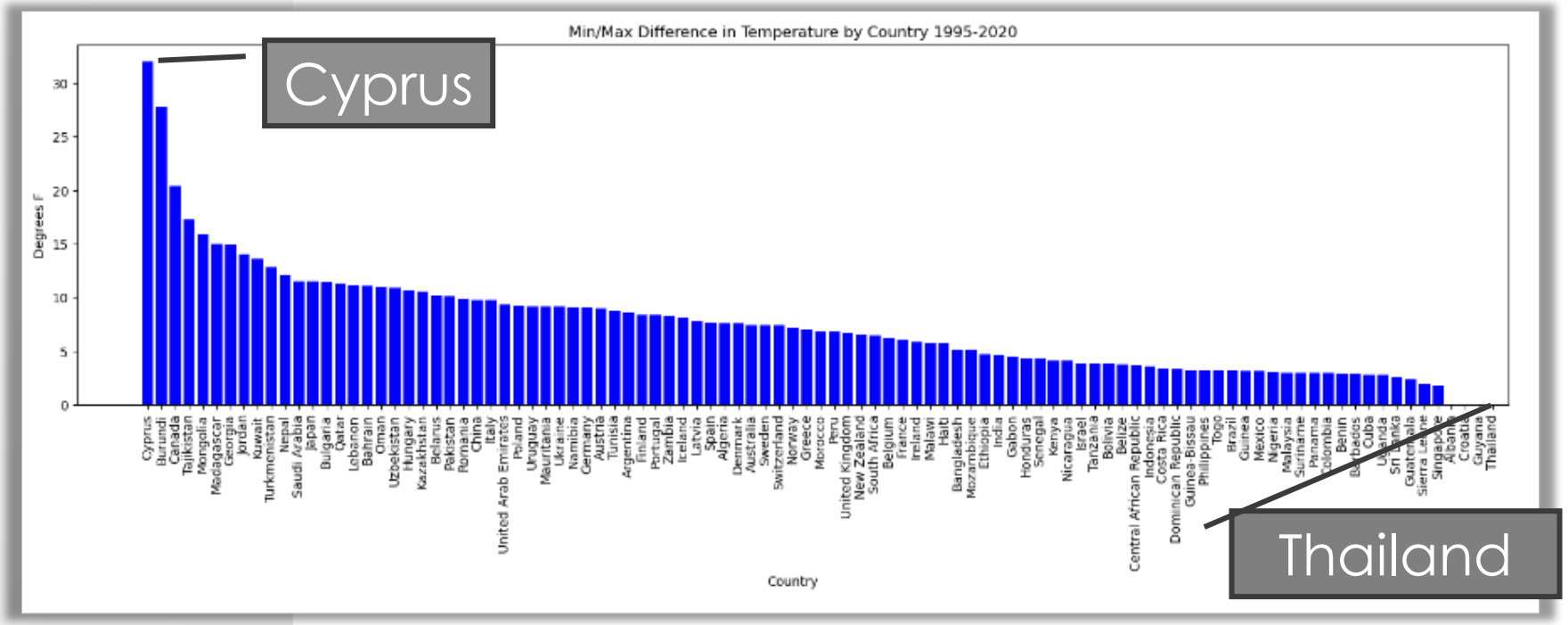
- Reset the index to make ‘Country’ a mergable column

```
temp_df = temp_df.reset_index()
```

- Loop through renaming temp/CO2 columns to differentiate year

```
for col in temp_df.columns:
    new_col = lambda x, col=col: 'Temp_' + str(col)
    temp_df.rename(columns={col: new_col(col)}, \
                   inplace=True)

for cols in co2_df.columns:
    if cols.isdigit():
        new_col = 'CO2_' + cols
        co2_df.rename(columns={cols: new_col},
inplace=True)
```

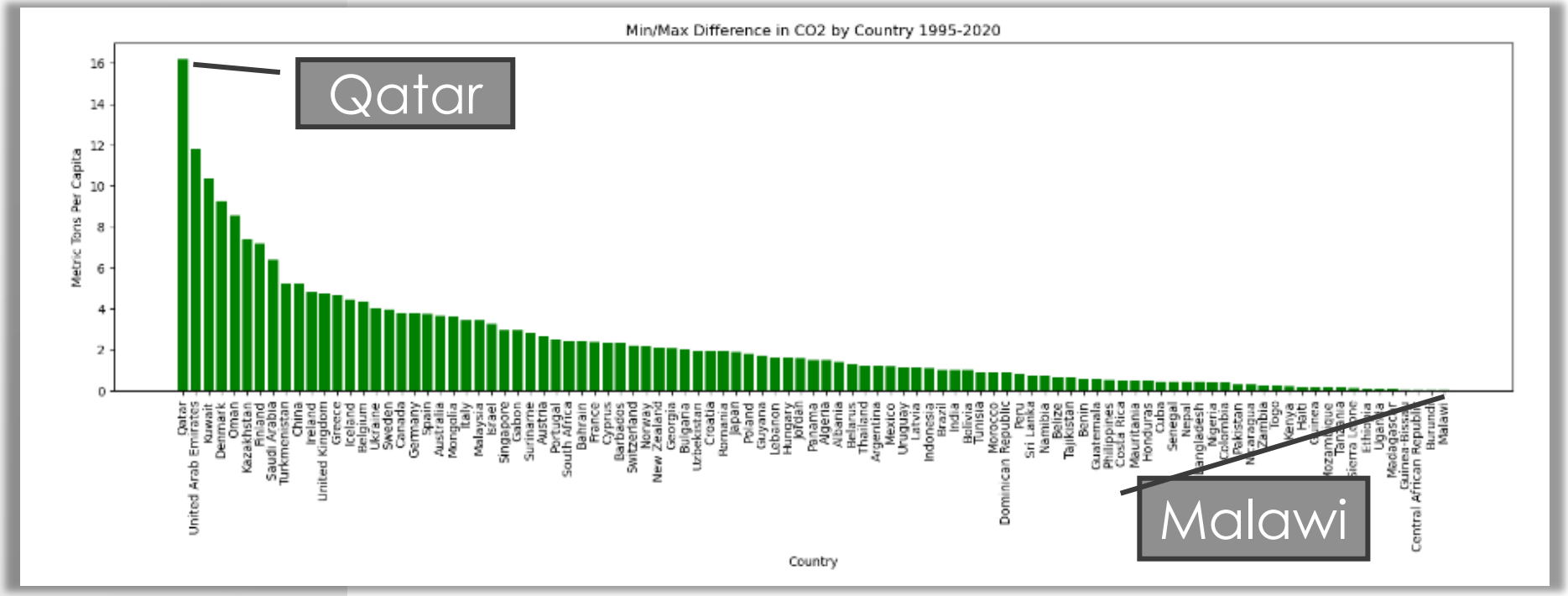
MIN/ MAX DIFFERENCE IN TEMPERATURE BY COUNTRY 1995 – 2020




```
# Which country had the highest change in Temperature
# Plot a bar chart
plt.figure(figsize=(20,5))
plt.bar(x_axis, df_sorted["Temp_difference"], color='b',
align="edge")
plt.xticks(tick_locations, df_sorted["Country"],
rotation="vertical")
plt.title("Min/Max Difference in Temperature by Country
1995-2020")
plt.xlabel("Country")
plt.ylabel("Degrees F")
plt.show()
```

Based on the temp data from 1995-2020 and chart above, we can see that Cyprus had the highest difference in Temp while Thailand remained consistent in temperatures changes over the years.

MIN/ MAX DIFFERENCE IN CO2 BY COUNTRY 1995 – 2020



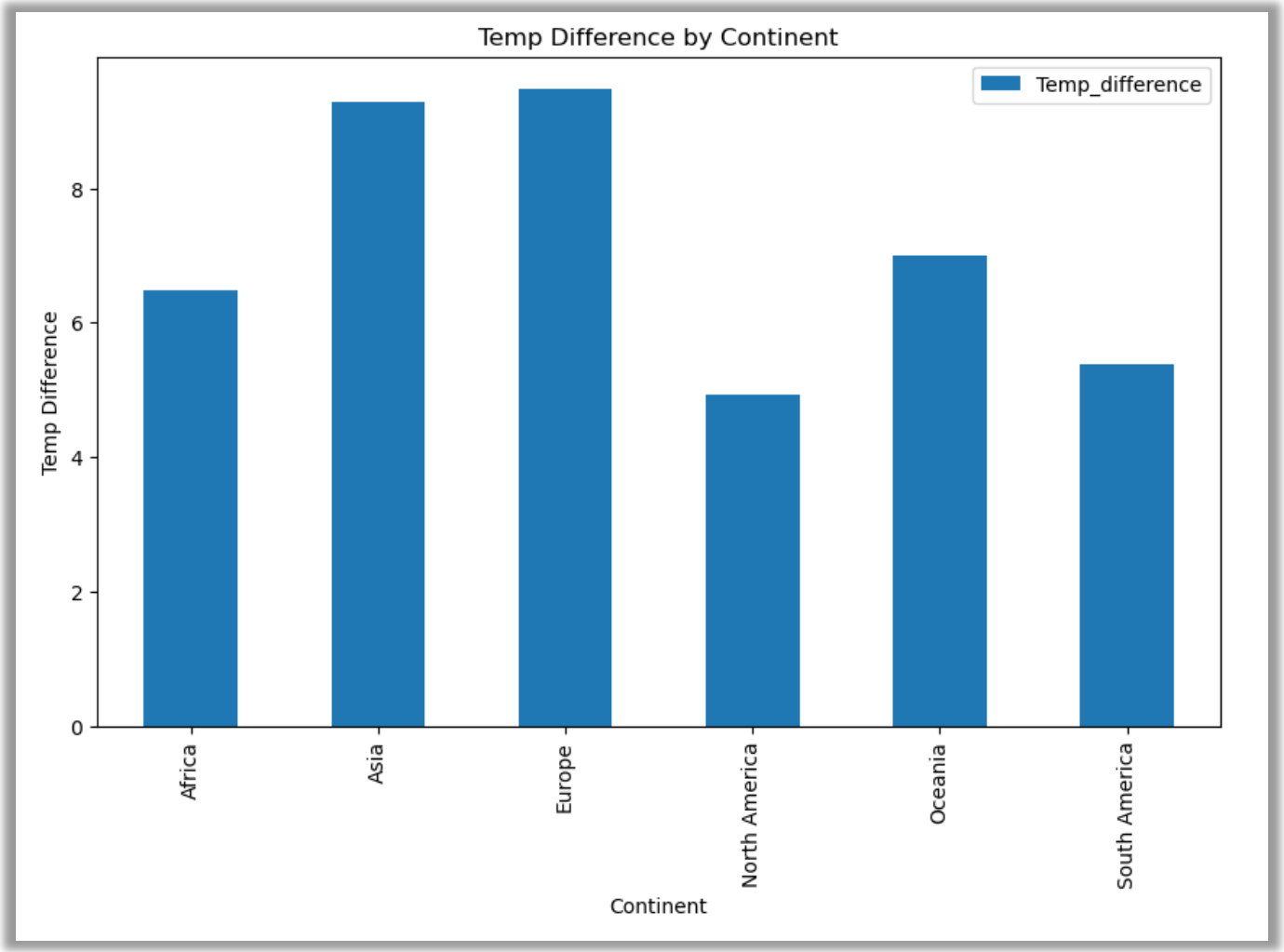

```
# Which country had the highest change in CO2
# Plot a bar chart
plt.figure(figsize=(20,5))
plt.bar(x_axis, df_sorted2["CO2_difference"], color='g',
align="edge")
plt.xticks(tick_locations, df_sorted2["Country"],
rotation="vertical")
plt.title("Min/Max Difference in CO2 by Country 1995-
2020")
plt.xlabel("Country")
plt.ylabel("Metric Tons Per Capita")
plt.show()
```

According to the data, the county with the highest CO2 difference over the years is Qatar while the African country of Malawi had the lowest difference of C02 over the same time frame.

Based on the data and chart, we can see that from 1995 to 2020 Europe saw the biggest difference temp difference while North America saw the lowest difference in temp over the same period.

```
# plt.figure(figsize=(20,5))
# plt.bar(x_axis,
avg_temp_per_continent["Temp_difference"], color='b',
align="edge")
# plt.xticks(tick_locations,
avg_temp_per_continent["Continent"], rotation="vertical")
# avg_temp_per_continent.plot(kind = 'bar', y=‘)

 bar_chart = avg_temp_per_continent.plot(kind='bar',
x='Continent', figsize=(10, 6))

 # Set the title and labels
bar_chart.set_title("Temp Difference by Continent")
bar_chart.set_xlabel("Continent")
bar_chart.set_ylabel("Temp Difference")

# Show the bar chart
bar_chart
```

MAX & MIN TEMP (1995-2020) DIFFERENCE BY CONTINENT

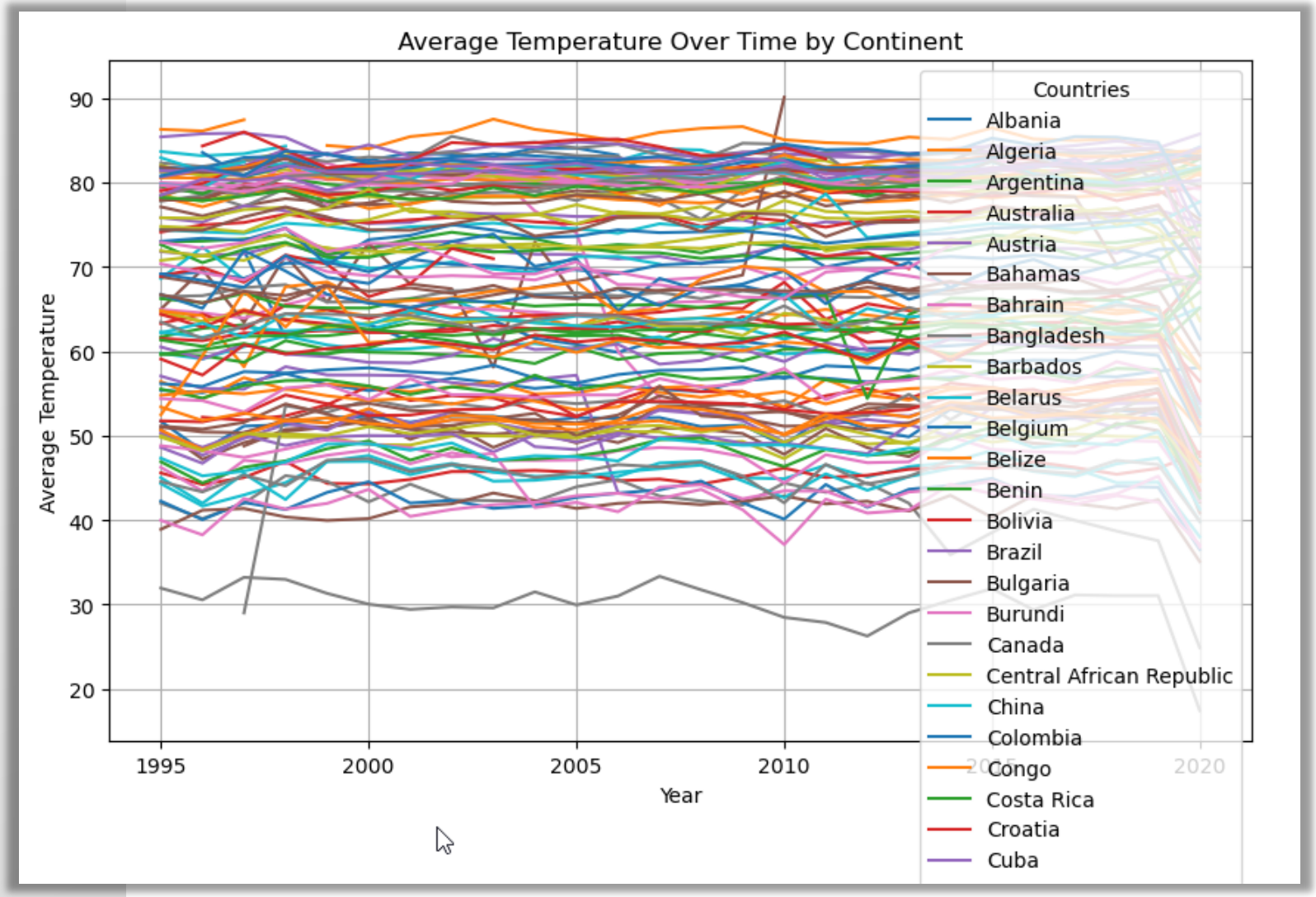


AVERAGE TEMP OVER TIME BY COUNTRY

Based on this data and chart, we can see that there is not really a general trend of all temperatures going up like we think there is.

This might be caused by our data not being as accurate as we originally thought. If we had more time, we would look for another dataset that might provide a more clear distinction of Temperature change over time.

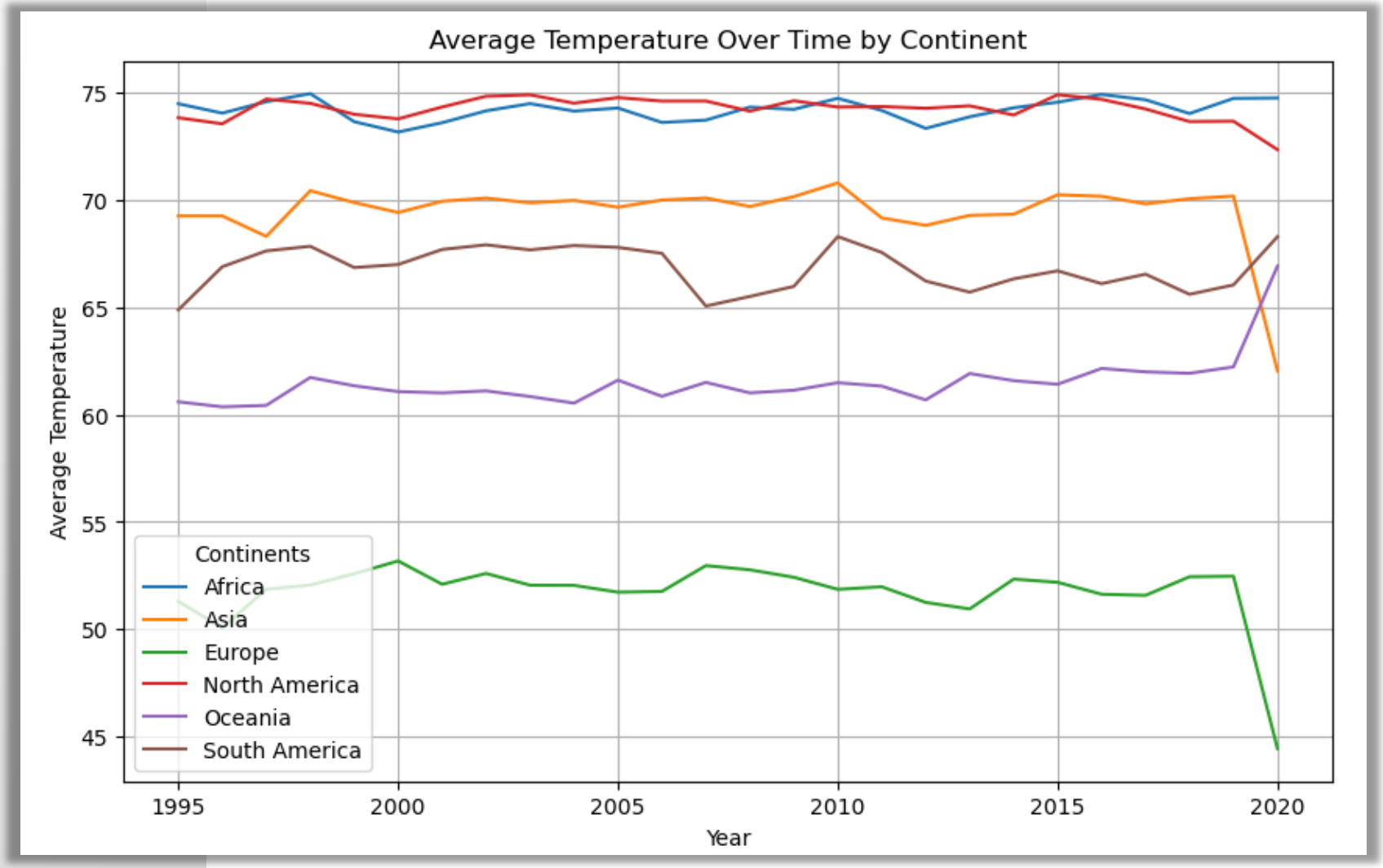
```
# plotting the change in avg temp in countries over time
pivot_df = avg_temp_continent_country.pivot(index='Year',
columns='Country', values='YearlyAvgTemperature')

# Plot
pivot_df.plot(kind='line', figsize=(10, 6))
plt.title('Average Temperature Over Time by Country')
plt.xlabel('Year')
plt.ylabel('Average Temperature')
plt.legend(title='Countries')
plt.grid(True)
plt.show()
```



AVERAGE TEMP OVER TIME BY CONTINENT

As a group we wanted to see the trends with temperatures around the globe. So to do so we grouped all country data to their respective continent. After that we just plotted each continent and its yearly average temperature over 25 years. The biggest struggle was just grouping every country to their respective continent.

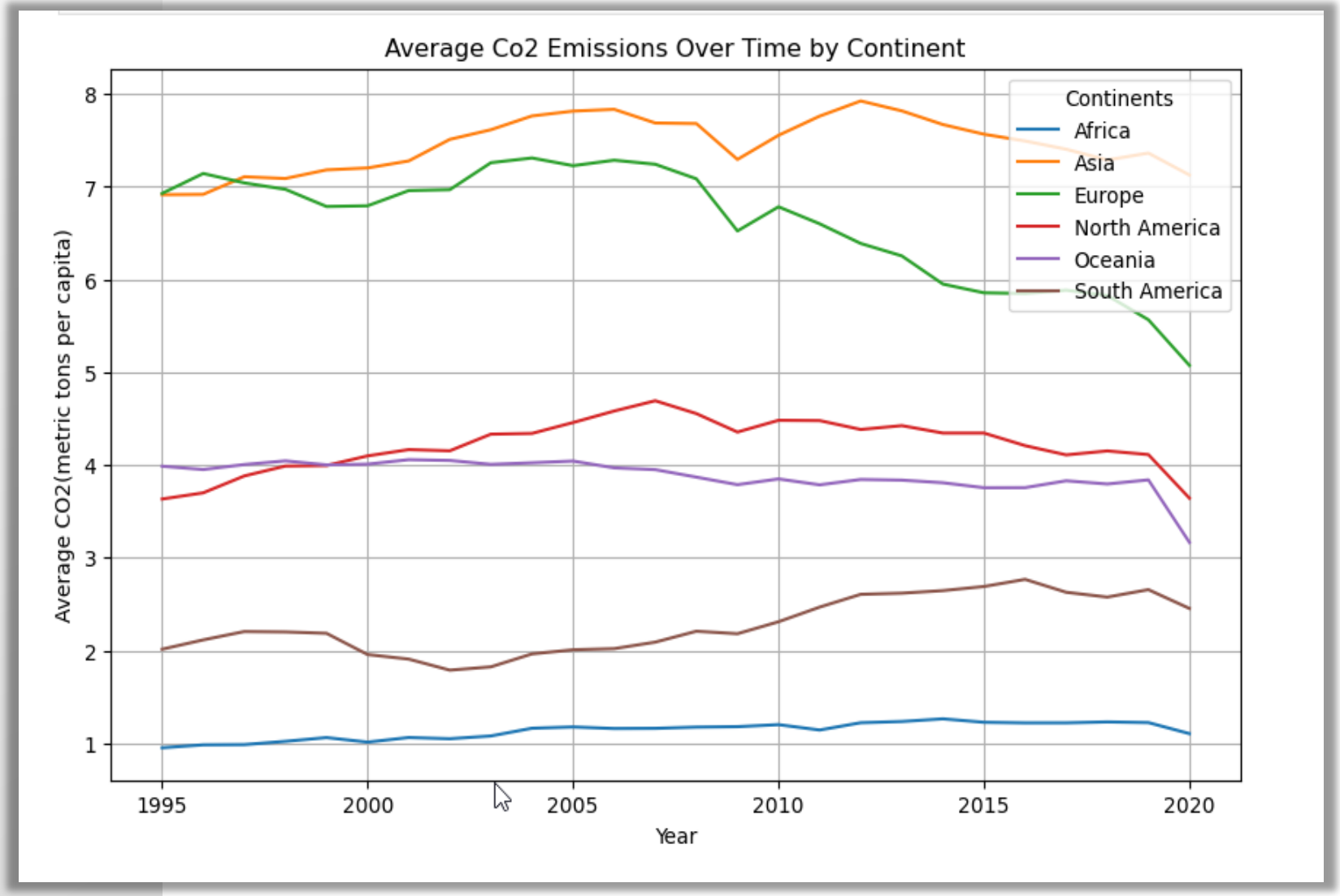
```
df_pivot =
continent_avg_temps.pivot(index='Year',
columns='Continent',
values='YearlyAvgTemperature')
df_pivot.plot(kind='line', figsize=(10, 6))

plt.title('Average Temperature Over Time by
Continent')
plt.xlabel('Year')
plt.ylabel('Average Temperature')
plt.legend(title='Continents')
plt.grid(True)

plt.show()
```



AVERAGE CO2 EMISSIONS OVER TIME BY CONTINENT

The goal was to also see the CO2 emissions per continent to compare the higher emissions and how the trend of those temperatures have changed. Which visually we didn't see as much correlation as we would think, which was confirmed **with a correlation coeficiant of -0.090863 a**s well as to see the difference between larger continents and smaller continents and their emissions. It was cool to see that Europe which is substantially smaller than Asia is closest continent in emissions. The hardest part was this was just comparing it to the average temperature plot.

```
# Plot
df_pivot_co2.plot(kind='line', figsize=(10, 6))

plt.title('Average Co2 Emissions Over Time by Continent')
plt.xlabel('Year')
plt.ylabel('Average CO2(metric tons per capita)')
plt.legend(title='Continents')
plt.grid(True)

plt.show()
```

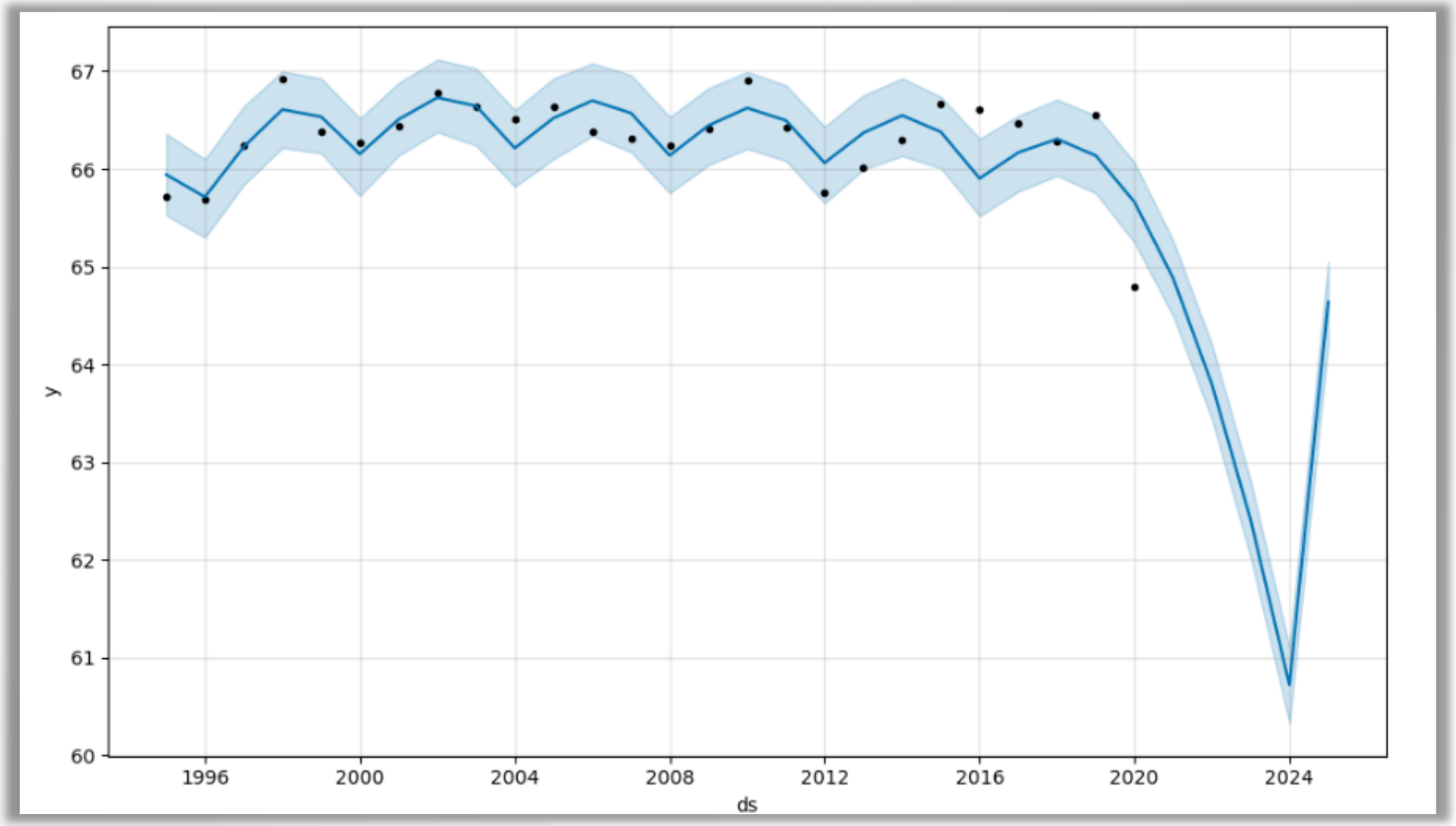


FORECASTING

To achieve the goal of potentially forecasting the average temperature of the world we needed to get an average world temp for each year. Then we just needed to make it datetime. Which to do so I had to spend sometime figuring out how to do so with the format being right. Below is the code that I had to do to specify the format.
world_avg_temperature['Year'] = pd.to_datetime(world_avg_temperature['Year'], format='%Y').

After that it was straightforward just had to change the column names and create and fit the model. Then just plot our prediction model.

Forecasted World Temperature

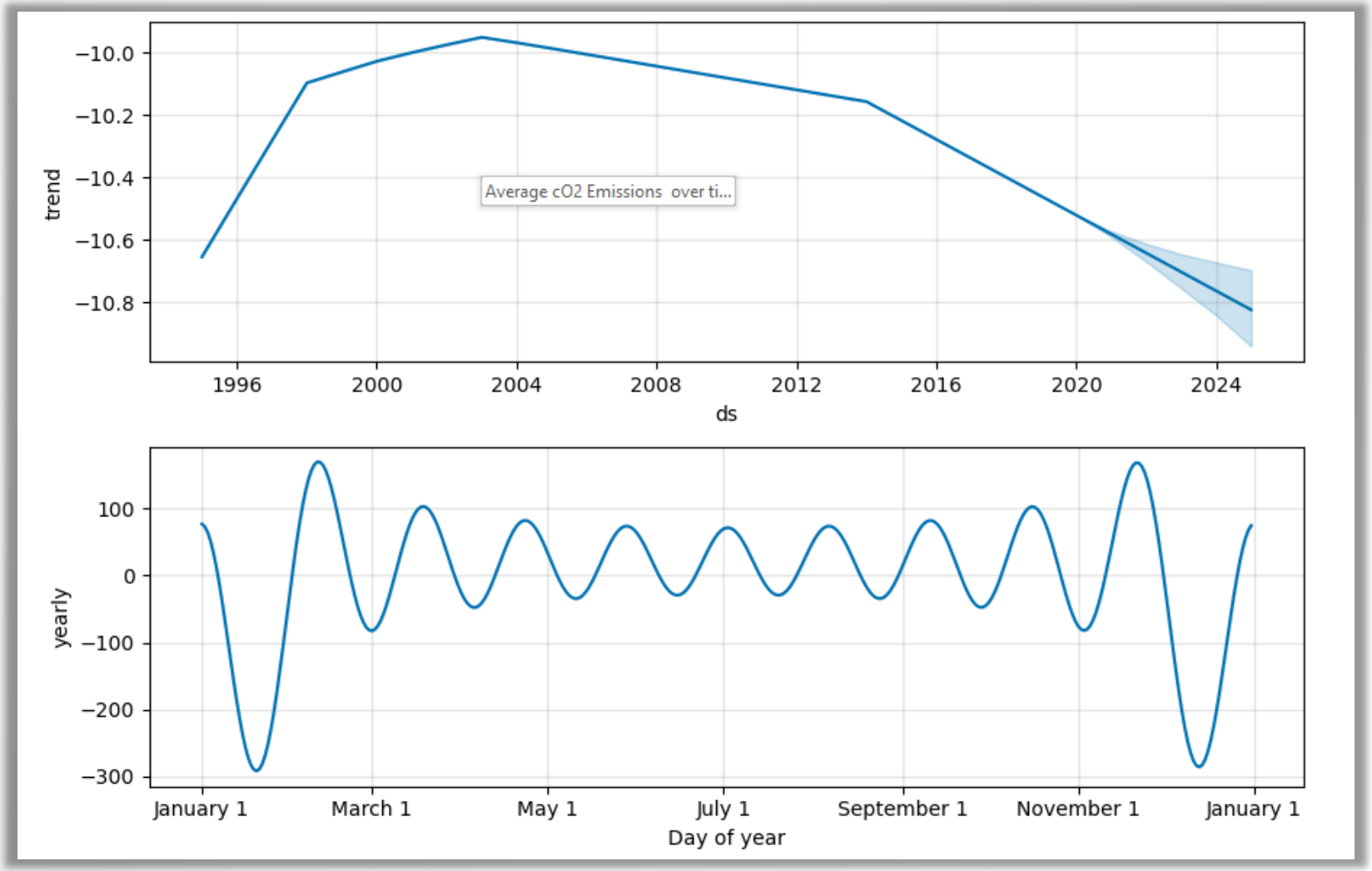


FORECASTING

After the first plot we wanted to visualize the results. With the results it only confirmed that the predictions were predicting that it would be a downward trend with the average temperature of the world. There wasn't much challenge after doing the work for the first plot.

figures = model.plot_components(forecast_trends)

Visual Results of Predictions

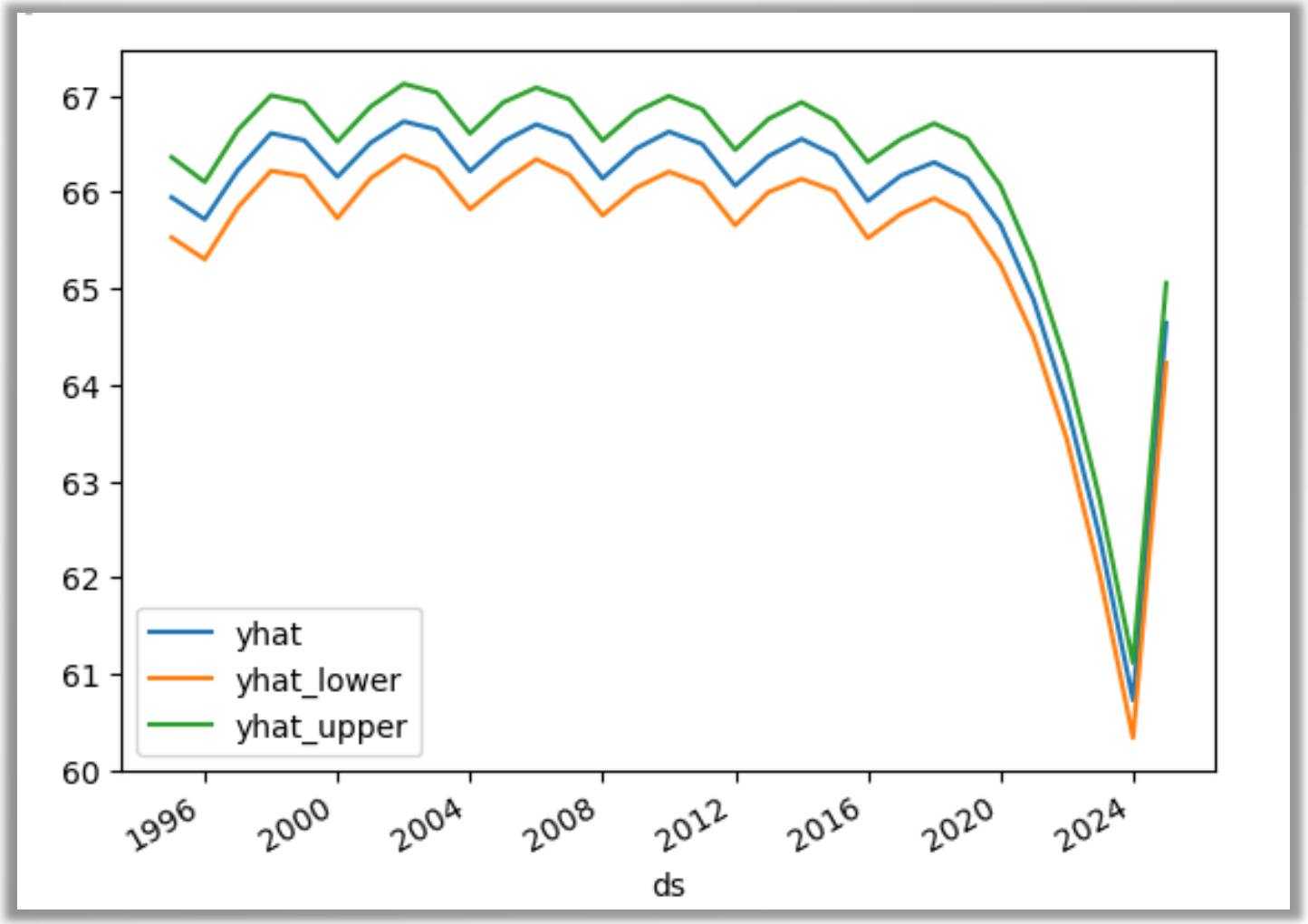


FORECASTING

After that we wanted to plot the yhat, yhat lower and upper over the last 2 years. Which just put out the same results just showing the lowest and highest prediction. This was pretty straight forward just needed to set it to go off the last 730 days(2 years)

```
forecast_trends[["yhat",
"yhat_lower",
"yhat_upper"]].iloc[-730:,
:].plot()
```

Yhat Prediction Plot



SCATTER PLOT OF TEMP & CO2 INCREASE

Findings:

- Not a direct 1:1 correlation, however...
- As CO2 increases, there is less variability / more predictability in temp

To create this chart:

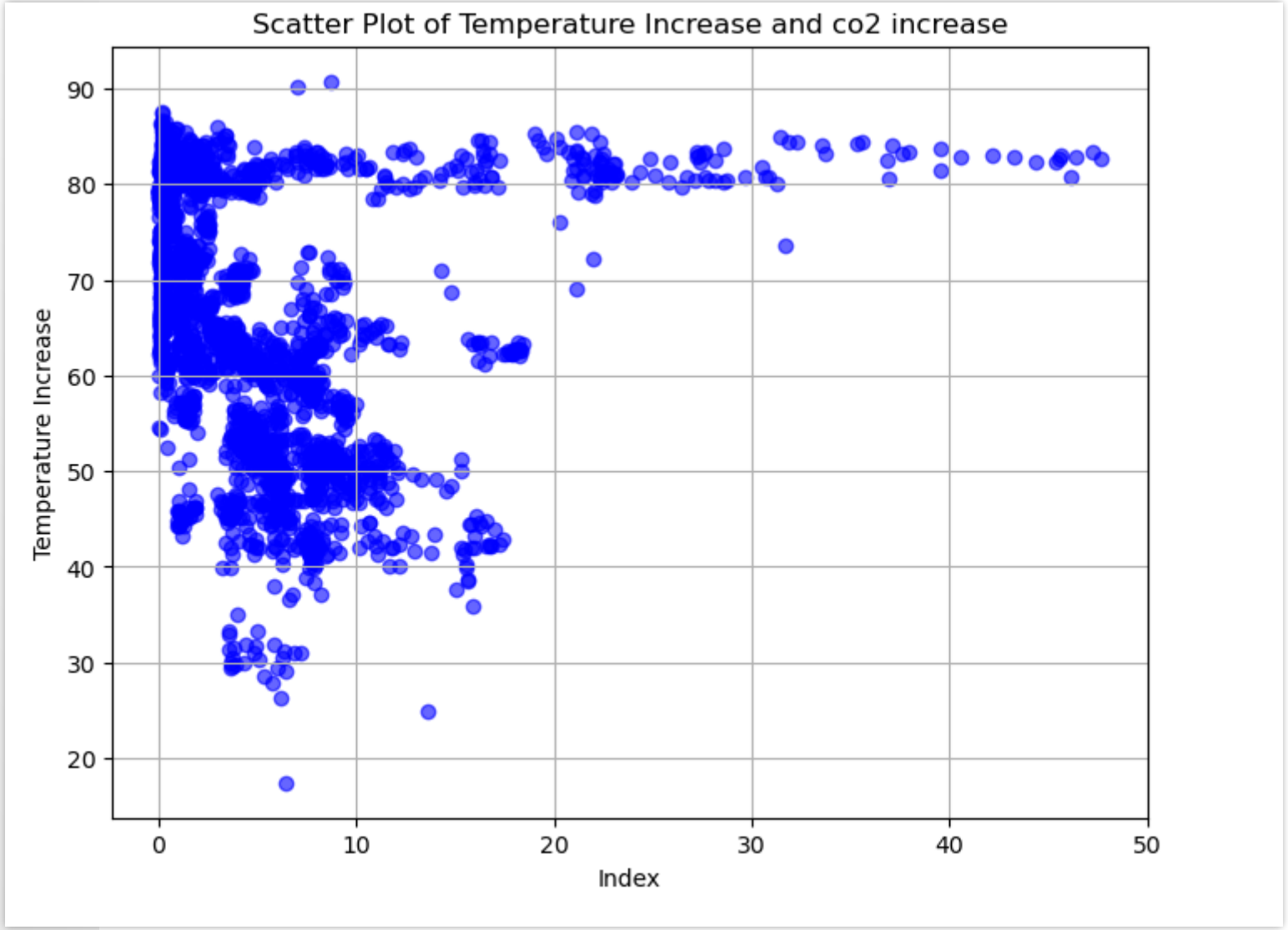
- Merged temp, country datasets
- Used plt.scatter

Further exploration:

- Attempt to fit a regression line

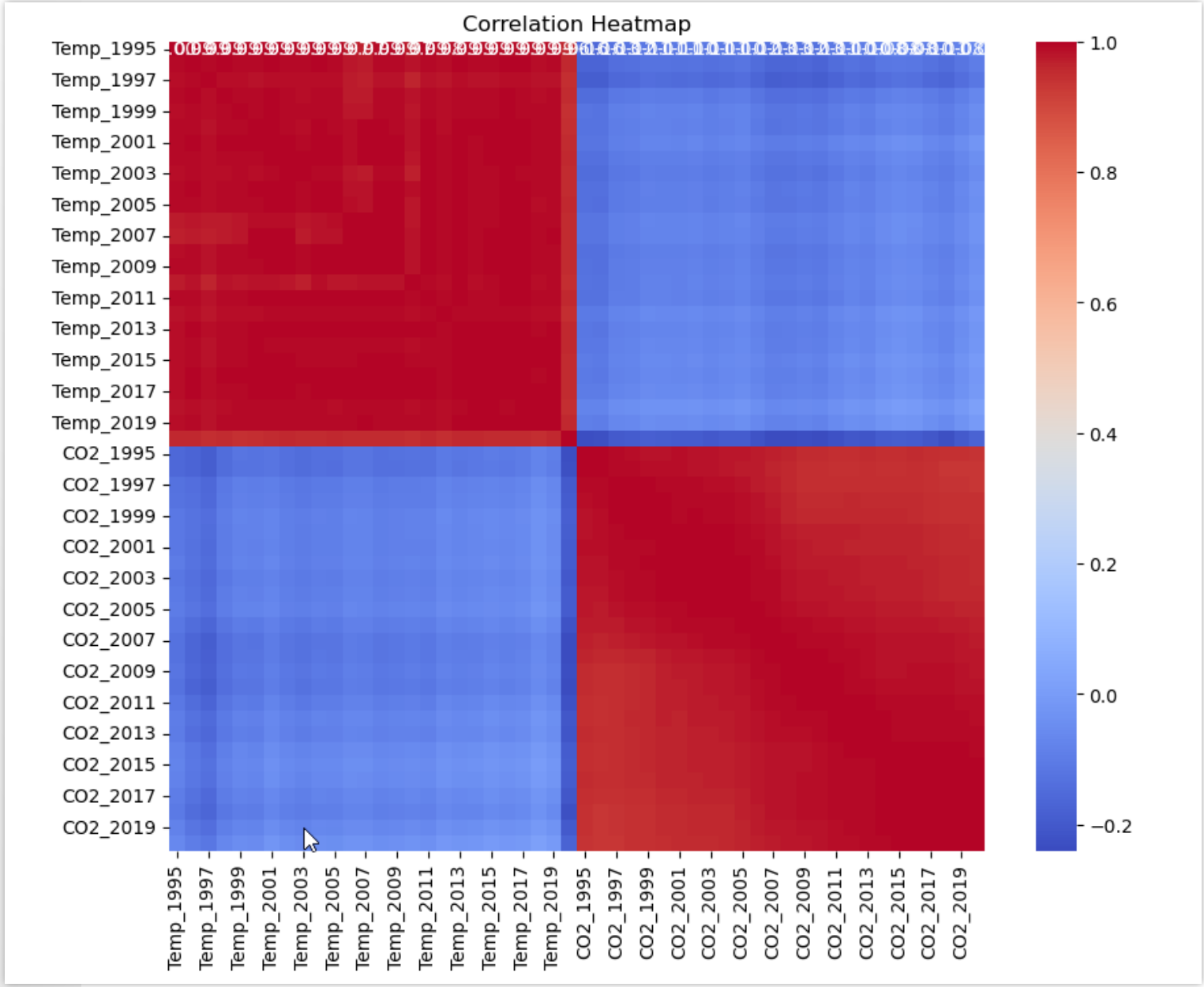
Code:

```
# Create a scatter plot
plt.figure(figsize=(8, 6))
plt.scatter(avg_co2_temp_continent_country['CO2 emissions (metric tons per capita)'],
avg_co2_temp_continent_country['YearlyAvgTemperature'],
color='blue', alpha=0.6)

# Scatter plot with index on x-axis and Temp increase on y-axis
plt.xlabel('Index')
plt.ylabel('Temperature Increase')
plt.title('Scatter Plot of Temperature Increase and co2 increase')
plt.grid(True)
plt.show()
```



CORRELATION HEATMAP

```
# see if there is a temp/CO2 correlation
Temp_CO2_Merged.drop(['Continent', 'Country', 'Temp_min',
'Temp_max', 'Temp_difference', 'CO2_min', 'CO2_max',
'CO2_difference'], axis=1, inplace=True)
display(Temp_CO2_Merged)
correlation_matrix = Temp_CO2_Merged.corr()

# Create a heatmap of the correlation matrix using Seaborn:
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True,
cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



IF WE HAD MORE TIME

- Get **additional and better data to forecast** temperature & CO2
- Explore **CO2 vs Methane**
- Explore **change in water temperature**
- **Fit a linear regression** line on the scatter plot correlation chart
- **Investigate the lag** between the drastic change in CO2 levels and Temp
- **Impact of COVID-19** on CO2 over time by continent
- Explore the temp data for **inconsistencies/ errors** – 2 continents dropped dramatically
- Explore **inequity** for which countries are most affected and which have the biggest discrepancies. **GDP** Explore **impact of population** density/per capita.
- Explore **effect of electric car policies** on production & CO2 levels; e.g., country that manufactured vs country that purchased. Did this impact 2020?

QUESTIONS?

an image or symbol to represent asking questions - Dalle