Question

Difference between Climate Change and Global Warming? Climate change and global warming are related, but not exactly the same thing. Here's the breakdown:

Global warming: This refers specifically to the long-term increase in the average global temperature near Earth's surface. It's primarily caused by human activities that release greenhouse gases into the atmosphere. These gases trap heat like a blanket, leading to a gradual warming trend.

Climate change: This is a broader term that encompasses not just rising temperatures, but also the long-term alteration of temperature and typical weather patterns in a place. Climate change can manifest as:

More frequent and intense heat waves:

Changes in precipitation patterns (droughts in some areas, floods in others) Rising sea levels Stronger storms Changes in plant and animal life

Why it's important: Understanding the difference between climate change and global warming is important because it helps us grasp the full scope of the issue. Global warming is the root cause, but climate change is the consequence – a complex web of effects that will impact everything from weather patterns to agriculture, food security, and human health.

Here's an analogy: Imagine your house is getting warmer (global warming). Climate change would be the burst pipes, failing air conditioning, and mold growth that result from that rising heat.

By understanding the bigger picture of climate change, we can make more informed decisions about how to mitigate its effects and adapt to the changes we're already experiencing.

Data Sources Describe your data sources :where they are from,

- Metadata URL: https://berkeleyearth.org/global-temperature-report-for-2023/
- Data URL: https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data/code
- Data Type: CSV
- License NO: CC BY-NC-SA 4.0 DEED
- About Dataset: This Dataset contains information about average land temperature and for maximum and minimum land temperatures and global ocean and land temperatures.

Datasource2: Global Warming

- Data URL: https://www.kaggle.com/datasets/kkhandekar/climate-change-vs-global-warming/data?select=Breakdown_Region.csv
- Data Type: CSV

- License NO: CC0 1.0 DEED
- About Dataset : This Dataset contains information about global warning chance according to cities.

Describe your data sources: Why you have chosen them,

What is the data structure and quality of your sources? (Compare lecture D01) Describe the licenses of your data sources, why you are allowed to use the data and how you are planning to follow their obligations If your source data is under a standard open-data license just pointing out where to find that is enough information for being allowed to use it, please still describe how you plan to fulfill their obligations

Import Library

```
import pandas as pd
import matplotlib.pyplot as plt
```

Load the Dataset

```
climate_data = pd.read_csv('Project/Breakdown_Region.csv')

temperature_data =
pd.read_csv('Project/GlobalLandTemperaturesByCountry.csv')

# Display the first few rows of each dataset
climate_data.head(), temperature_data.head()
```

```
(            Country Climate change: (1/1/04 - 9/27/21)  \
 0          Kiribati                                100%
 1  Marshall Islands                                 84%
 2         Micronesia                                100%
 3   Solomon Islands                                 82%
 4           Vanuatu                                 86%

   Global Warming: (1/1/04 - 9/27/21)
 0                                NaN
 1                                16%
 2                                NaN
 3                                18%
 4                                14%  ,
            dt  AverageTemperature  AverageTemperatureUncertainty
Country
 0  1743-11-01               4.384                          2.294
Åland
 1  1743-12-01                 NaN                            NaN
Åland
 2  1744-01-01                 NaN                            NaN
Åland
 3  1744-02-01                 NaN                            NaN
```

```
Åland
 4  1744-03-01                  NaN                                  NaN
Åland)
```

Merge the Dataset

```
# Merge the datasets on the 'Country' column
merged_data = pd.merge(climate_data, temperature_data, on='Country')

# Group by 'Country' and keep the row with the highest average values
# result_data = merged_data.loc[merged_data.groupby('Country')
['AverageTemperature'].idxmax()]

# Display the result
# result_data.head()

# merged_data



# Ensure there are no NaN values in 'Country' and 'AverageTemperature'
columns
merged_data = merged_data.dropna(subset=['Country',
'AverageTemperature'])

# Reset index to avoid alignment issues
merged_data = merged_data.reset_index(drop=True)

# Group by 'Country' and keep the row with the highest
'AverageTemperature'
result_data = merged_data.loc[merged_data.groupby('Country')
['AverageTemperature'].idxmax()]

# Save the result to a new CSV file
# result_data.to_csv('filtered_data.csv', index=False)

# Display the result
result_data
```

```
               Country Climate change: (1/1/04 - 9/27/21)  \
165914     Afghanistan                                 68%
75623          Albania                                 75%
205876         Algeria                                 86%
16665   American Samoa                                100%
432419         Andorra                                 NaN
...                ...                                 ...
293286         Vietnam                                 63%
216672  Western Sahara                                100%
345914           Yemen                                100%
38902           Zambia                                 74%
```

```
25269         Zimbabwe                                    72%

       Global Warming: (1/1/04 - 9/27/21)          dt
AverageTemperature  \
165914                                    32%  1997-07-01
28.533
75623                                     25%  1757-07-01
25.843
205876                                    14%  2003-07-01
35.829
16665                                     NaN  2003-01-01
28.543
432419                                    NaN  2003-08-01
24.313
...                                       ...         ...
...
293286                                    37%  1912-06-01
28.463
216672                                    NaN  2004-08-01
30.092
345914                                    NaN  1998-06-01
32.737
38902                                     26%  2005-10-01
26.282
25269                                     28%  1995-10-01
26.601

        AverageTemperatureUncertainty
165914                          0.410
75623                           5.336
205876                          0.400
16665                           0.231
432419                          0.291
...                               ...
293286                          0.358
216672                          0.704
345914                          1.080
38902                           0.325
25269                           0.201

[196 rows x 6 columns]
```

Data Cleaning

```
# Ensure there are no NaN values in 'Country' and 'AverageTemperature'
columns
merged_data = merged_data.dropna(subset=['Country',
'AverageTemperature','Climate change: (1/1/04 - 9/27/21)', 'Global
Warming: (1/1/04 - 9/27/21)','AverageTemperatureUncertainty'])
```

```python
# Reset index to avoid alignment issues
merged_data = merged_data.reset_index(drop=True)

# Remove duplicates from the dataset
merged_data = merged_data.drop_duplicates()

# Group by 'Country' and keep the row with the highest
'AverageTemperature'
result_data = merged_data.loc[merged_data.groupby('Country')
['AverageTemperature'].idxmax()]

# Save the result to a new CSV file (uncomment the line below if you
need to save the file)
# result_data.to_csv('filtered_data.csv', index=False)

# Display the result
print(result_data)
```

```
            Country Climate change: (1/1/04 - 9/27/21)  \
147321  Afghanistan                                  68%
66911       Albania                                  75%
177845      Algeria                                  86%
332397    Argentina                                  72%
259738      Armenia                                  75%
...             ...                                  ...
203091   Uzbekistan                                  63%
322375    Venezuela                                  75%
251442      Vietnam                                  63%
32062        Zambia                                  74%
18429      Zimbabwe                                  72%

       Global Warming: (1/1/04 - 9/27/21)          dt
AverageTemperature  \
147321                                  32%  1997-07-01
28.533
66911                                   25%  1757-07-01
25.843
177845                                  14%  2003-07-01
35.829
332397                                  28%  2012-01-01
23.290
259738                                  25%  2006-08-01
25.291
...                                     ...         ...
...
203091                                  37%  1984-07-01
30.375
322375                                  25%  2010-03-01
27.807
```

```
251442                                    37%  1912-06-01
28.463
32062                                     26%  2005-10-01
26.282
18429                                     28%  1995-10-01
26.601

        AverageTemperatureUncertainty
147321                          0.410
66911                           5.336
177845                          0.400
332397                          0.333
259738                          0.254
...                               ...
203091                          0.305
322375                          0.418
251442                          0.358
32062                           0.325
18429                           0.201

[147 rows x 6 columns]
```

```
# Reset index to avoid alignment issues
merged_data = merged_data.reset_index(drop=True)
# Display the result
print(result_data)
```

```
            Country Climate change: (1/1/04 - 9/27/21)  \
147321  Afghanistan                                  68%
66911       Albania                                 75%
177845      Algeria                                 86%
332397    Argentina                                 72%
259738      Armenia                                 75%
...             ...                                 ...
203091   Uzbekistan                                 63%
322375    Venezuela                                 75%
251442      Vietnam                                 63%
32062        Zambia                                 74%
18429      Zimbabwe                                 72%

       Global Warming: (1/1/04 - 9/27/21)          dt
AverageTemperature  \
147321                                    32%  1997-07-01
28.533
66911                                     25%  1757-07-01
25.843
177845                                    14%  2003-07-01
35.829
332397                                    28%  2012-01-01
23.290
```

```
259738                                    25%  2006-08-01
25.291
...                                       ...         ...
...
203091                                    37%  1984-07-01
30.375
322375                                    25%  2010-03-01
27.807
251442                                    37%  1912-06-01
28.463
32062                                     26%  2005-10-01
26.282
18429                                     28%  1995-10-01
26.601

        AverageTemperatureUncertainty
147321                          0.410
66911                           5.336
177845                          0.400
332397                          0.333
259738                          0.254
...                               ...
203091                          0.305
322375                          0.418
251442                          0.358
32062                           0.325
18429                           0.201

[147 rows x 6 columns]
```

```
# Convert the 'Date' column to datetime format
merged_data['dt'] = pd.to_datetime(merged_data['dt'])

# Extract the month from the 'Date' column
merged_data['Month'] = merged_data['dt'].dt.month

# Reset index to avoid alignment issues
merged_data = merged_data.reset_index(drop=True)
merged_data
```

```
                Country Climate change: (1/1/04 - 9/27/21)  \
0       Solomon Islands                                 82%
1       Solomon Islands                                 82%
2       Solomon Islands                                 82%
3       Solomon Islands                                 82%
4       Solomon Islands                                 82%
...                 ...                                 ...
343221            Japan                                 66%
343222            Japan                                 66%
343223            Japan                                 66%
```

```
343224            Japan                                    66%
343225            Japan                                    66%

       Global Warming: (1/1/04 - 9/27/21)         dt
AverageTemperature  \
0                                       18% 1867-01-01
26.807
1                                       18% 1867-02-01
26.416
2                                       18% 1867-03-01
26.310
3                                       18% 1867-04-01
26.648
4                                       18% 1867-05-01
26.347
...                                     ...        ...                  .
..
343221                                  34% 2013-04-01
10.102
343222                                  34% 2013-05-01
15.256
343223                                  34% 2013-06-01
19.961
343224                                  34% 2013-07-01
24.286
343225                                  34% 2013-08-01
25.669

        AverageTemperatureUncertainty  Month
0                               1.035      1
1                               0.831      2
2                               0.802      3
3                               0.897      4
4                               0.703      5
...                               ...    ...
343221                          0.322      4
343222                          0.235      5
343223                          0.380      6
343224                          0.369      7
343225                          0.303      8

[343226 rows x 7 columns]


result_data

            Country Climate change: (1/1/04 - 9/27/21)  \
147321  Afghanistan                                  68%
66911       Albania                                  75%
177845      Algeria                                  86%
```

```
332397    Argentina                                    72%
259738      Armenia                                    75%
...             ...                                   ...
203091   Uzbekistan                                    63%
322375    Venezuela                                    75%
251442      Vietnam                                    63%
32062        Zambia                                    74%
18429      Zimbabwe                                    72%

       Global Warming: (1/1/04 - 9/27/21)          dt
AverageTemperature  \
147321                                  32%  1997-07-01
28.533
66911                                   25%  1757-07-01
25.843
177845                                  14%  2003-07-01
35.829
332397                                  28%  2012-01-01
23.290
259738                                  25%  2006-08-01
25.291
...                                     ...         ...
...
203091                                  37%  1984-07-01
30.375
322375                                  25%  2010-03-01
27.807
251442                                  37%  1912-06-01
28.463
32062                                   26%  2005-10-01
26.282
18429                                   28%  1995-10-01
26.601

        AverageTemperatureUncertainty
147321                          0.410
66911                           5.336
177845                          0.400
332397                          0.333
259738                          0.254
...                               ...
203091                          0.305
322375                          0.418
251442                          0.358
32062                           0.325
18429                           0.201

[147 rows x 6 columns]
```

```python
# Convert the Index object to a list
column_names_list = list(result_data.columns)

# Print the list of column names
print("Column names as list:", column_names_list)
```

```
Column names as list: ['Country', 'Climate change: (1/1/04 -
9/27/21)', 'Global Warming: (1/1/04 - 9/27/21)', 'dt',
'AverageTemperature', 'AverageTemperatureUncertainty']
```

```python
#Change the Columns Name
# Rename specific columns (example: 'OldName1' to 'NewName1' and
'OldName2' to 'NewName2')
columns_to_rename = {
    'Climate change: (1/1/04 - 9/27/21)': 'Climate change',
    'Global Warming: (1/1/04 - 9/27/21)': 'Global Warming',

    # Add other column renaming as needed
}
merged_data = result_data.rename(columns=columns_to_rename)
merged_data
```

```
            Country Climate change Global Warming          dt  \
147321  Afghanistan            68%            32%  1997-07-01
66911       Albania            75%            25%  1757-07-01
177845      Algeria            86%            14%  2003-07-01
332397    Argentina            72%            28%  2012-01-01
259738      Armenia            75%            25%  2006-08-01
...             ...            ...            ...         ...
203091   Uzbekistan            63%            37%  1984-07-01
322375    Venezuela            75%            25%  2010-03-01
251442      Vietnam            63%            37%  1912-06-01
32062        Zambia            74%            26%  2005-10-01
18429      Zimbabwe            72%            28%  1995-10-01

        AverageTemperature  AverageTemperatureUncertainty
147321              28.533                          0.410
66911               25.843                          5.336
177845              35.829                          0.400
332397              23.290                          0.333
259738              25.291                          0.254
...                    ...                            ...
203091              30.375                          0.305
322375              27.807                          0.418
251442              28.463                          0.358
32062               26.282                          0.325
18429               26.601                          0.201

[147 rows x 6 columns]
```

```
# Convert the Index object to a list
column_names_list = list(merged_data.columns)

# Print the list of column names
print("Column names as list:", column_names_list)
```

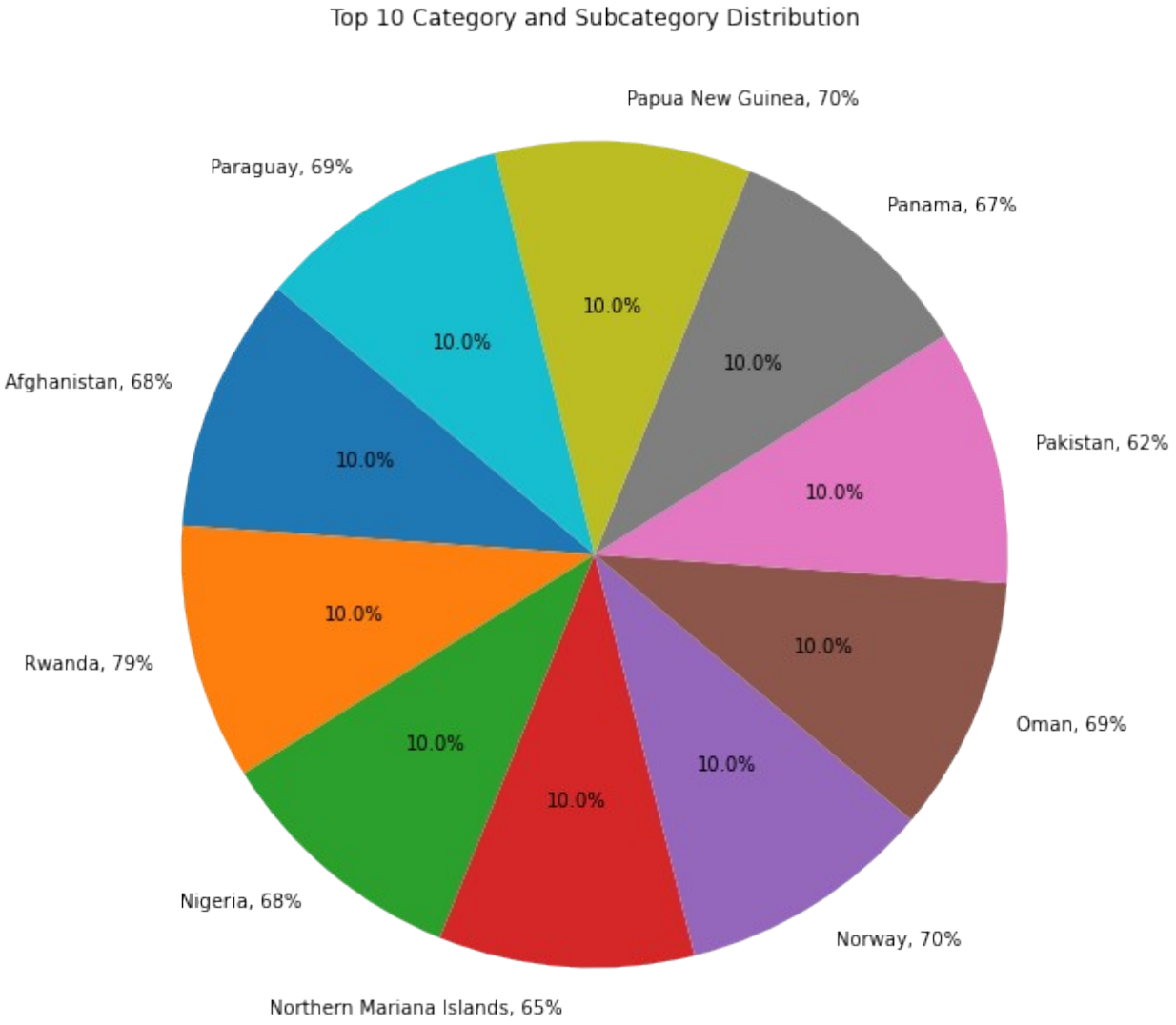
```
Column names as list: ['Country', 'Climate change', 'Global Warming',
'dt', 'AverageTemperature', 'AverageTemperatureUncertainty']
```

```
# Make sure to adjust 'Category' and 'Subcategory' to the names of
your columns
# Group by both 'Category' and 'Subcategory', and count the
occurrences
category_subcategory_counts = merged_data.groupby(['Country', 'Climate
change']).size().reset_index(name='counts')

# Create labels for the pie chart
#labels = [f'{cat}, {subcat}' for cat, subcat in
zip(category_subcategory_counts['Country'],
category_subcategory_counts['Climate change'])]
# Sort the DataFrame by counts in descending order and select the top
10
top_10_counts = category_subcategory_counts.sort_values(by='counts',
ascending=False).head(10)


# Create labels for the pie chart
labels = [f'{cat}, {subcat}' for cat, subcat in
zip(top_10_counts['Country'], top_10_counts['Climate change'])]

# Create the pie chart
plt.figure(figsize=(10, 10))
plt.pie(top_10_counts['counts'], labels=labels, autopct='%1.1f%%',
startangle=140)
plt.title('Top 10 Category and Subcategory Distribution')
plt.show()
```



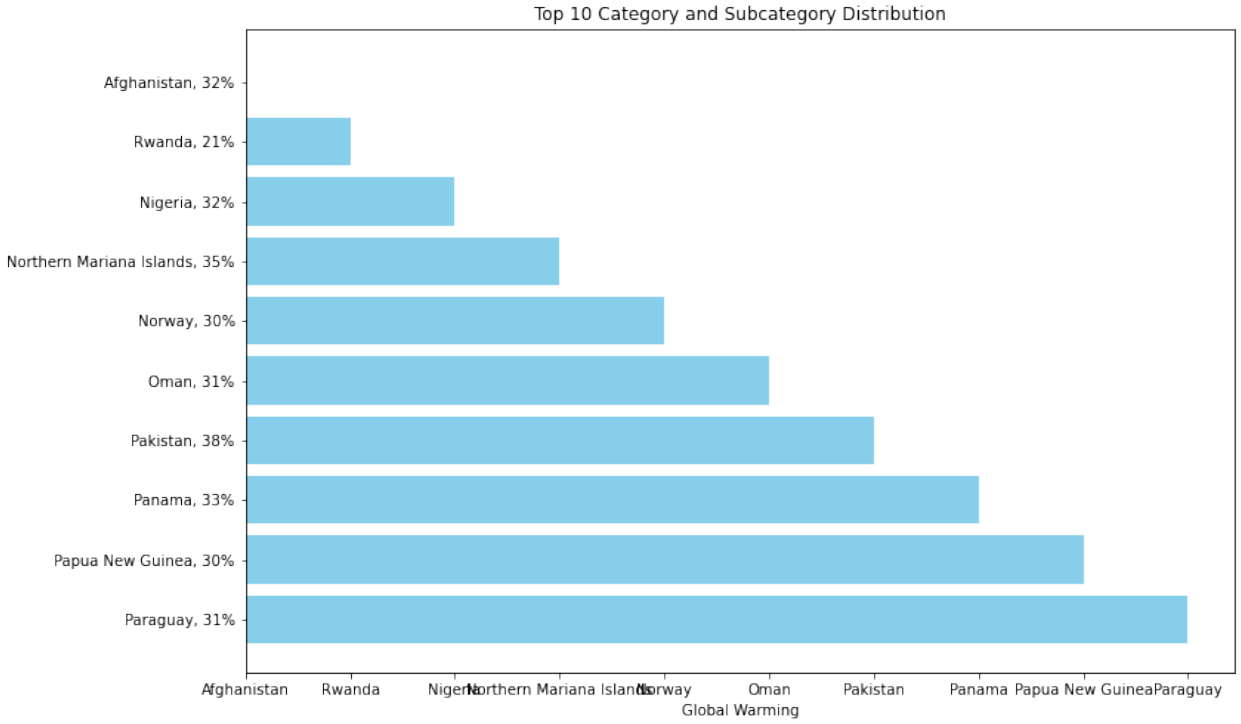

```
# Make sure to adjust 'Category' and 'Subcategory' to the names of
your columns
# Group by both 'Category' and 'Subcategory', and count the
occurrences
category_subcategory_counts = merged_data.groupby(['Country', 'Global
Warming']).size().reset_index(name='counts')

# Sort the DataFrame by counts in descending order and select the top
10
top_10_counts = category_subcategory_counts.sort_values(by='counts',
ascending=False).head(10)

# Create labels for the bar chart
```

```
labels = [f'{cat}, {subcat}' for cat, subcat in
zip(top_10_counts['Country'], top_10_counts['Global Warming'])]

# Create the bar chart
plt.figure(figsize=(12, 8))
plt.barh(labels, top_10_counts['Country'], color='skyblue')
plt.xlabel('Global Warming')
plt.title('Top 10 Category and Subcategory Distribution')
plt.gca().invert_yaxis()  # Invert y-axis to display the highest
counts at the top
plt.show()
```



Result and Limitations

Describe the output data of your data pipeline What is the data structure and quality of your result? (Compare lecture D01) What data format did you choose as the output of your pipeline and why

This dataset isn't necessarily the best dataset to definitively find the difference between global warming and climate change. Here's why:

Limited Timeframe: The data covers a period from January 1st, 2004 to September 27th, 2021 (less than 18 years). Climate change is a long-term phenomenon measured in decades or even centuries. This dataset wouldn't capture the long-term trends needed to fully distinguish climate change from natural fluctuations in weather.

Missing Data: The dataset focuses on "Climate Change" and "Global Warming" values, but it doesn't provide any specific details on what those values represent (e.g., temperature change, policy implementations). Without that context, it's difficult to understand how they differ.

Limited Scope: The dataset seems to be focused on a single country, while climate change and global warming are global issues. A broader dataset encompassing multiple countries over a longer period would be more suitable.

However, this dataset could be a starting point for further investigation if:

More context is available: If there's additional information explaining how "Climate Change" and "Global Warming" are measured in this dataset, it could provide some insights into how they differ.

Part of a larger dataset: This dataset might be a snippet of a larger study that includes more comprehensive data (e.g., covering multiple countries and longer timeframes).

Overall, a more suitable dataset to study the difference between climate change and global warming would include:

Global data: Information on average temperatures, precipitation patterns, extreme weather events, etc., collected from multiple countries over several decades.

Longitudinal data: Data measured over a long period to capture long-term trends.

Specific metrics: Clearly defined metrics for "Climate Change" and "Global Warming" within the dataset.

By analyzing these elements, scientists can compare global temperature increases (global warming) with the resulting changes in weather patterns and ecosystems (climate change).