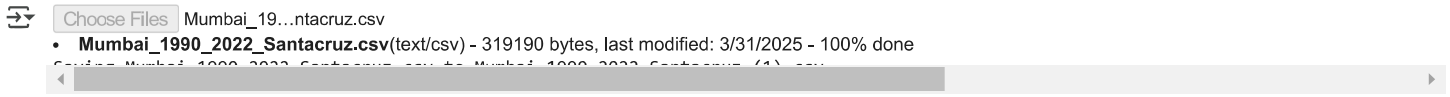
Loading Data

```
import pandas as pd
import numpy as np
```

```
from google.colab import files
data = files.upload()
```




```
weather_data = pd.read_csv("/content/Mumbai_1990_2022_Santacruz (1).csv")
weather_data.head()
```

	time	tavg	tmin	tmax	prcp
0	01-01-1990	23.2	17.0	NaN	0.0
1	02-01-1990	22.2	16.5	29.9	0.0
2	03-01-1990	21.8	16.3	30.7	0.0
3	04-01-1990	25.4	17.9	31.8	0.0
4	05-01-1990	26.5	19.3	33.7	0.0

Next steps: Generate code with weather_data | View recommended plots | New interactive sheet

Mapping Function

```
def mapper():
    file_path = "/content/Mumbai_1990_2022_Santacruz (1).csv"
    df = pd.read_csv(file_path, parse_dates=["time"], dayfirst=True)  # Parse date
    df["year"] = df["time"].dt.year  # Extract year
    df = df[["year", "tavg"]].dropna()  # Keep only year and tavg, remove NaNs

    return df
```

Reducing Function

```
def reducer(mapped_data):
    grouped = mapped_data.groupby("year")["tavg"].apply(list)  # Group by year, keep all tavg values
    avg_temps = mapped_data.groupby("year")["tavg"].mean()  # Compute mean tavg per year

    result_df = pd.DataFrame({"tavg_values": grouped, "avg_tavg": avg_temps})
    return result_df
```

Processing using MapReduce

```
if __name__ == "__main__":
    df = weather_data
    print("Weather Data loaded\n")

    mapped_data = mapper()
    print("Mapped Data loaded\n")

    reduced_result = reducer(mapped_data)
    print("Reduced Data loaded\n")
    print(reduced_result.head())

    hottest_year = reduced_result["avg_tavg"].idxmax()
```

```
    hottest_year = reduced_result["avg_tavg"].idxmax()
    coldest_year = reduced_result["avg_tavg"].idxmin()

    print(f"Hottest Year: {hottest_year} with avg temp {reduced_result.loc[hottest_year, 'avg_tavg']:.2f}°
    print(f"Coldest Year: {coldest_year} with avg temp {reduced_result.loc[coldest_year, 'avg_tavg']:.2f}°
```

```
Weather Data loaded

Mapped Data loaded

Reduced Data loaded

                                          tavg_values   avg_tavg
year
1990  [23.2, 22.2, 21.8, 25.4, 26.5, 25.1, 26.0, 26....  27.076944
1991  [18.4, 17.9, 18.8, 20.5, 22.2, 22.3, 22.0, 20....  26.933791
1992  [22.8, 24.1, 23.6, 22.2, 23.7, 22.3, 20.3, 21....  27.109836
1993  [24.9, 24.7, 24.1, 24.8, 25.7, 26.3, 23.9, 23....  27.175549
1994  [25.6, 24.7, 23.9, 26.1, 25.1, 25.6, 26.3, 25....  26.939118
Hottest Year: 2018 with avg temp 28.76°C
Coldest Year: 1991 with avg temp 26.93°C
```