

# Long Term Analysis and Visualization of Temperature Changes in the USA

DTSC 630 Data Visualization

Arvin Jadonath, Zhishan Liu, Jungmin Sung



# 0 Presentation Overview

1. Project Purpose / Objective
  2. About the Dataset
  3. Data Cleaning
  4. Demo - Plots / Analysis
  5. Conclusion / Possible Future Work
- References

# 1 Project Purpose and Objective

- **Project Objective:**

- Obtain and analyze temperature data over a set period of time in the United States of America.
- Visualize the data using *plotly*, *seaborn*, *matplotlib* to show the levels of temperature in different states.

## 2 About the Dataset: Original Source

- Kaggle: **Daily Temperature of Major Cities** (*1995 Jan ~ 2020 May*)
- Original Dataset Source: (*1995 Jan ~ Present*)
  - **University of Dayton, Ohio** - Environmental Protection Agency Average Daily Temperature Archive
  - Daily average temperature for 324 international cities.
  - Data collected from *Global Summary of the Day (GSOD)* database archived by the *National Centers for Environmental Information (NCEI)*

## 2 About the Dataset: Kaggle

- Kaggle: **Daily Temperature of Major Cities**
  - [city\\_temperature.csv](#) (140.6MB) - About 3 million rows and 8 columns.
  - **Columns:** [Region, Country, State, City, Month, Day, Year, AvgTemperature]
    - $\text{AvgTemperature} = (\text{Daily Min Temp} + \text{Daily Max Temp}) / 2$
  - Our project is focused on the U.S. data - About 1.5 million rows.
    - 50 States + Washington D.C.
    - 153 US Cities.
    - Time Period: About 25.5 years 1/1/1995 ~ 5/13/2020

### 3 Data Cleaning

- Dataset contains no-data flags for unavailable data:
  - **AvgTemperature** == -99 (4,083 in US data) *Global Reigion*  
79,672
- Some invalid values:
  - **Day** == 0 (none in the US data) 8
  - **Year** == 200 or 201 (none in the US data) 440
- Excluding the Year 2020 when calculating Yearly-Average:
  - Year 2020 contains only first 5 months.

## 4 Plots and Analysis - Demo

- Load & Clean Dataset
- 25-Year-Average Temperature of US & of States
- Yearly-Average Temperature of US & of States
- Decade-Average Temperature of US

# Loading the Data into the Data Frame & Exploring The Data

1. All of the US time series data from 1995 to 2019.
2. The mean of average temperatures (from 1995 to 2019) statewise (this is not a time series data).
3. The Maximum average temperatures (from 1995 to 2019) statewise (this is not a time series data).

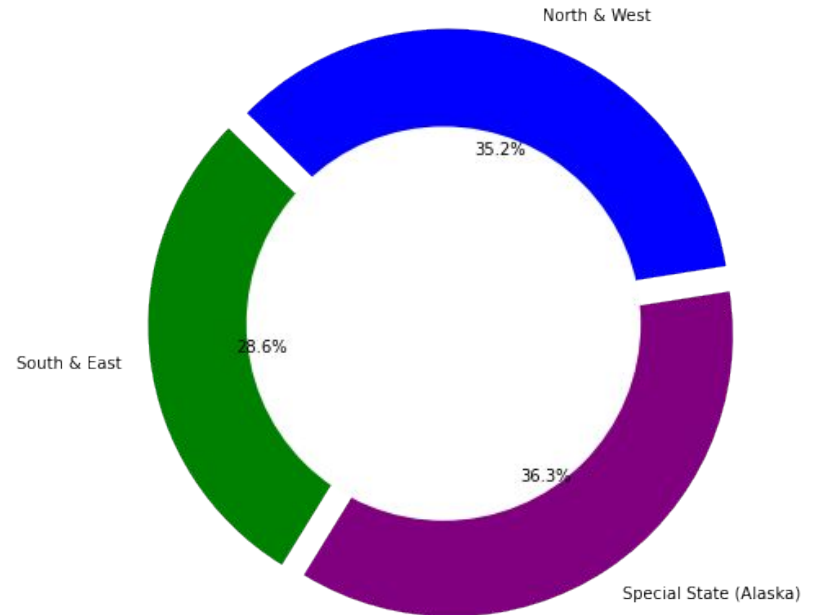


The North, West and Alaska parts of the US have seen the most temperature increase while the South East parts of the US have experienced a little change.

For this analysis we compared the states from all categories.

Region wise Difference (based on selected states)

- North , West parts of USA:  
Kansas, Illinois, California, Utah, South Dakota
- South, East part of USA:  
Alabama, Florida, New York, Georgia, Maine
- Special Case:  
Alaska



# Visualize the Warmest Decade in the US

Since we have data from 1995 to 2020 we are going to divide it as 3 decades

- 1995–2005
- 2006–2015
- 2016–2019 (Incomplete decade)

## 5 Conclusion and Possible Future Work

- 25 Years of Data showed that:
  - Yearly-Average Temperatures are not increasing each year.
  - In Decade-Average, Temperatures are increasing.
- For Possible Future Work:
  - Expand the time periods & Use global data.
  - Compare with the environmental variables such as CO2 emission:
    - *CO2 Emissions \_ 1960 - 2018* available on Kaggle  
<https://www.kaggle.com/datasets/kkhandekar/co2-emissions-1960-2018>
  - Implement Dash interface to provide interactive visualization.

# Reference: Data Source

- **Kaggle Dataset:**  
<https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities>
- **University of Dayton Original Dataset Source:**  
<https://academic.udayton.edu/kissock/http/Weather/default.htm>
- **National Centers for Environmental Information:**  
<https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00516>

# Reference: Library Documentation

- **Plotly:** <https://plotly.com/python/>
- **Pandas:** <https://pandas.pydata.org/pandas-docs/stable/>
- **Matplotlib:** <https://matplotlib.org/stable/index.html>
- **NumPy:** <https://numpy.org/devdocs/user/index.html#user>
- **Seaborn:** <https://seaborn.pydata.org/>