Data analysis of climate change in USA

Hassan Mohamad

ID: 5716

Lebanese University-Faculty of engineering III

Hassan.mohamad@st.ul.edu.lb

AND

Mohamad Ousseily

ID: 5695

Mohamad.ousseily@st.ul.edu.lb

Abstract **- in this project, we are trying to answer the following questions: What are the key influencing factors of climate change in United States? How are those factors contributing to it? What will be the impact in next years? We will be doing our data analysis using Python.**

INTRODUCTION

Climate change refers to a shift in average weather conditions, including measures such as temperature, humidity, rainfall, cloudiness and wind patterns – and changes in the frequency or severity of these conditions. The Earth's climate has changed throughout its history, in cycles that occur over very long periods of time. This is a natural process. Today we tend to use the phrase 'climate change' to refer to the very rapid changes in the climate that we have seen over the past 50 years or so. The scientific evidence is clear that these changes are not being driven by long-term natural climate cycles. Instead their main cause is global warming and the human activities that cause it. Climate change has profound implications for people and the natural world. The temperature of air at the planet's surface has risen rapidly, especially over the last fifty years. We know this through records from thousands of weather stations across the world, satellites, and ocean data from ships and buoys. So in this small paper we aim to prove this fact, (especially in USA), and give a deeper understanding about it.

EVIDENCE OF CLIMATE CHANGE

The first Step in the evidence part where the climate change event will be demonstrated.
It will start by reading the csv file containing the global average temperature trend data from 1720 to 2020. Due to lack of data in the early years, the data is filtered by taking in consideration only the data from 1820 and above, and then we plot the global average temperature variations.

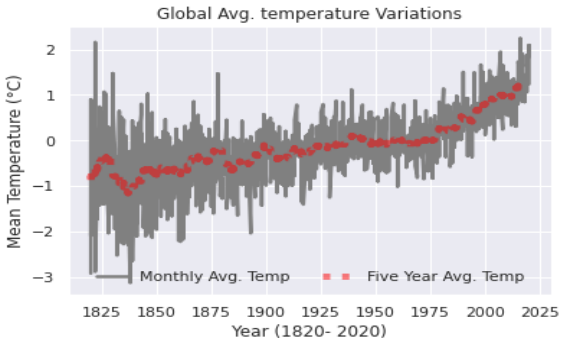

Figure 1: Monthly and five year global Avg. Temp. Variations at each year.

After that, we read the csv file containing the NA average temperature trend, and the same operations as above are done on its data and then it is plotted.

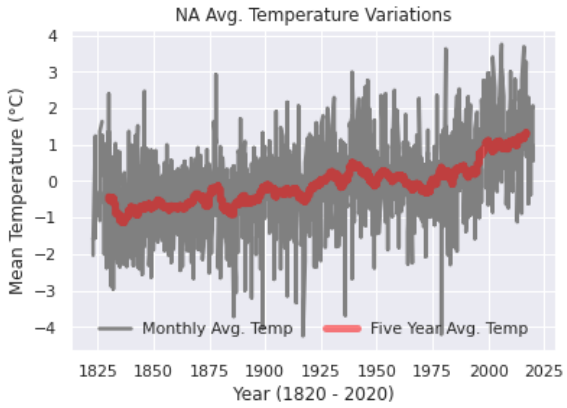

Figure 2: Monthly and five year NA Avg. Temp. Variations at each year.

It's clear here that the average global and NA temperature are increasing over years.
If we divide the interval of years in these data into ten equal quintiles where the mean of monthly Avg. temperature variation on each quintile is calculated, and the result obtained shows that the higher mean is present in the [2000,2020] quintile.
After reading the NA maximum and minimum temperature trend, and ranking its data according to higher monthly variation and lower one respectively on each month, a clear result in the data will appear that in general the top ranking high mouthy variation temperature are in the years above 2000, and the lower ones are in the years before or early 1900.
Moreover, we can see always after plotting the Tmax and Tmin trends, that over the years a higher minimum temperature and a higher maximum one is formed.

SEGMENTATION

Our aim in this part is sort the different states of USA that have similar weather conditions and similar climate change evidence over the past two centuries into different groups or clusters, so that we would be able to understand deeply where we have a big problem (so that the government starts taking actions), and where exactly we are still partially in the safe zone. This study targets mainly the data from year 2000, until 2019. And so we take the data in two parts, [2000-2009] and [2010-2019]. In each csv file we have data about every day, in every single state from 1/1/2000 until 31/12/2019. The data contains the values of five different attributes that we can call dependent variable which are:

1. PRCP (precipitation rate)
2. SNOW (snow fall rate)
3. TAVG (average daily temperature)
4. TMAX (maximum daily temperature)
5. TMIN (minimal daily temperature)

Now after analyzing the data, and cleaning it as well as fill the null values, we need to find out what is the optimal number of clusters (K) that we will perform the K-means algorithm on, and we do that using the Elbow curve algorithm. But before, let's define the necessary variables needed for this algorithm and for the K-means algorithm.

Segmentation variables:

- PRCP_MAX (Maximum precipitation in a year)
- SNOW_MAX (Maximum snowfall in a year)
- TAVG_avg_yoy(Average temperature YoY change 2010-2019 vs 2000-2009)
  TMAX_max_yoy
- Maximum temperature YoY change (2010-2019 vs 2000-2009)
- TMAX_avg_yoy(Average of maximum temperature YoY change 2010-2019 vs 2000-2009)
- TMIN_min_yoy (Miminim temperature YoY change 2010-2019 vs 2000-2009)

Profiling variables:

- PRCP_CHG (Maximum precipitation – Minimum precipitation in the same year).
- SNOW_CHG (Maximum snowfall – Minimum snowfall in the same year).
- PRCP_CNT_YoY (Number of days with precipitation change 2010-2019 vs 2000-2009)
- SNOW_CNT (number of days with snowfall change 2010-2019 vs 2000-2009**)**

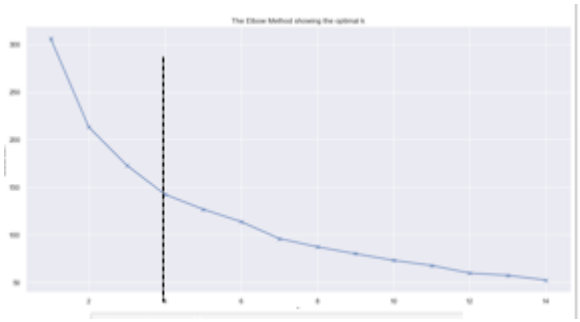

Figure 3: The Elbow Curve

We can easily see that the optimal K in our case is 4.

After performing the K-means clustering we obtain the following results:

- 24 regions got classified to cluster 0.
- 14 regions got classified to cluster 1.
- 10 regions got classified to cluster 2.
- 3 regions got classified to cluster 3.

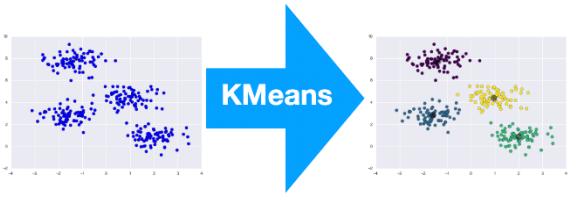

Figure 4: K-means clustering (k=4).

To identify the meaning of each cluster (i.e. what does it mean for a region X to belong to cluster Y), we perform a statistical analysis after grouping the regions by clusters. So that we can study the characteristics of every cluster, and we obtained the following results:

- Cluster 0: **Unpredictable weather states** (Snow days and quantity is reduced, and more precipitation was observed in the recent decade compared to previous, Minimum temperature went up 0.2-degree centigrade, period and overall increase in average temperature. The states are Florida, Texas, Illinois, Kansas, Ohio, Tennessee, Virginia, Los Angeles and 14 other states).

- Cluster 1: **Global Warming state** (Snow days and quantity is reduced, and more precipitation was observed in the recent decade compared to previous. Minimum temperature went up 0.2-degree centigrade, The states are California, Alaska, Nevada, Michigan, Oregon, Wisconsin, South Dakota and 8 other states).

- Cluster 2: **Wet states** (High rainfall in recent decade than before, change of 270 mm on average. Maximum temperatures increased by 0.82 degree centigrade, States are New Jersey, New York, Connecticut, Massachusetts, Maryland and other 6 states).

- Cluster 3: **Cold But Dry States** (Average, minimum and maximum temperatures have gone down by 0.4 degree centigrade. Number of days with precipitation or snow went down, the states are Arizona, New Mexico and Colorado).

SUPERVISOR INFORMATION'S

Dr. Mohamad Aoude, Professor, Department of Electrical and Computer Engineering Lebanese University, faculty of engineering III.

REFERENCES

Python Libraries Used: Numpy-Csv-Pandas-Sklearn-Matplotlib.pyplot –Seaborn

- Data-Analysis-of-Climate-Change-Using-Python
- Human Population Growth and Climate Change
- Sources of Greenhouse Gas Emissions
- Global Climate Change: What You Need to Know
- Overview: Weather, Global Warming and Climate Change
- Raw data source
- Climate Change Datasets

FIGURE I
LOGO OF THE INSTITUTE FOR ELECTRICAL AND ELECTRONICS ENGINEERS