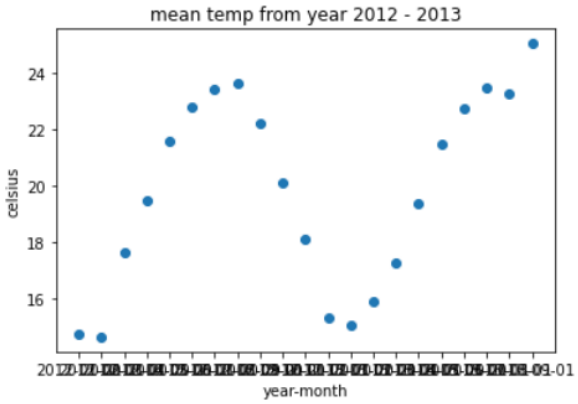
Qirui Zheng

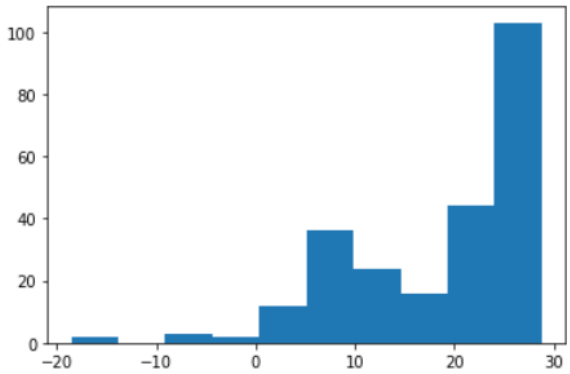
March 15, 2023

Summary of Regression on Earth Surface Temperature

Time is an abstract ideology, yet it is simple enough to be widely used is daily life. From Bastian's article, "Fatally Confused: Telling the Time in the Midst of Ecological Crises", the author suggested to add on to the definition of time to raise awareness in climate change. For example, the Doomsday Clock is time that ticks backwards to midnight. If the clock is ever moved pass midnight, then it will be too late to take action. With the new technological innovations occurring every day and at this speed, the clock is ticking closer and closer to midnight. From another perspective, this event can also be seen from the perpetually increasing Earth surface temperature. From a dataset found on Kaggle titled, "Climate Change: Earth Surface Temperature Data", with techniques of EDA and regression, a prediction rule can be found with the given information in the dataset. Thus, I am hoping to give a prediction about the given raise of temperature for the future years to come and add on to the role that time plays in the current climate crisis.

EDA, also known as Exploratory Data Analysis, is a process where from the given dataset, the observer first cleans the data, and then implement graph to understand the way that the current dataset is structured. Out of the four datasets that gives the earth temperature the dataset of temperature by country is chosen with the purpose of giving an overall prediction to the earth temperature. From the graph on the right, with the oscillation of the temperature due to different months and seasons, when running regression, it is important to keep in mind the differences in seasons. Next, to understand the data even more to see the actual distribution of where all the data cluster, a histogram is used in visualizing such changes.

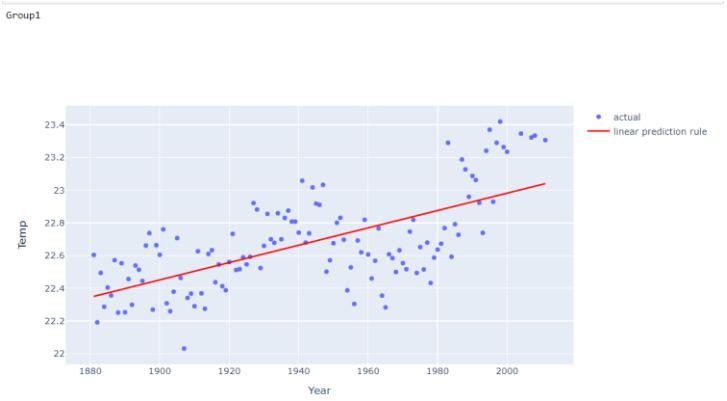


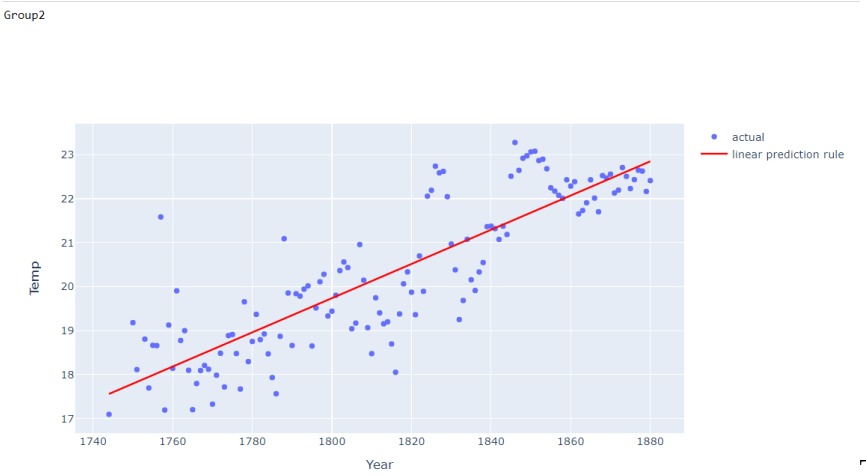


From the histogram there is two clear groups of data that can be seen, and thus a single prediction line might not be the best for all the data. Which then the data is split into two different groups using the method of K-means and iterating over the dataset for 30 epochs.

After all data cleaning steps are performed, we obtained two different groups group1 from 1881 to 2013, and group2 1744 to 1880. Since this is designed to analyze Earth Surface temperature increase, I've decide to take the maximum temperature of each year and predict for

the maximum increased in temperature. Below both graphs for Group1 and Group2 are shown.





Thus, to predict the highest temperature in 2023, we can use the first prediction rule that ranges from 1880 to 2013. All the steps can be found on at this link. However, there exists a big issue in the regression above such that it is only giving a trend for temperature increase. Which cannot be considered a prediction for the temperature when considering technological improvements that can affect the environment and cause a rapid increase in temperature. Thus, a new parameter needs to be introduced because in predicting temperature increase, time (year) does not weigh as much as other influential matters. This then leads on to further work needed to be continued.