Analyzing Climate Change of Capital Cities: Asia's Perspective

Syed Mahbubul Huq
Msc in Data Science
City, University of London
syed.huq@city.ac.uk

***Abstract*—This paper analyses climate change in all of the capital cities of Asia. Here everyday weather data for 32 years for 47 capital cities are considered, which makes 564,940 rows. From the analysis, weather trends and the most affected regions of climate change are indicated. Besides, the paper also studies if reducing human activity for some months can have an effect on present and future climate or not. Decision Tree is applied to the given dataset to predict and compare future climate effects.**

I. Introduction

The long-term pattern of weather experienced in a particular region or place is known as climate. The factors that consist of climate have important effects on almost all aspects of our lives, starting from food, health, agriculture, economic growth, transportation, and so on [1]. Changes in climate thus can have extensive impacts in our lives.

Among the continents in the globe, the change in climate is severe in Asia. Temperatures are rising two times faster here compared to the world [2]. Besides, the contribution of Asia in global warming is regarded as one of the key sources of the contribution. Half of the world's CO2 emits from Asia, which is also the home to the frontline countries that emits greenhouse gasses [2]. Though factors like poverty, geographical location, etc. made Asia vulnerable to climate change risks, it is considered, human activities which produce greenhouse gasses are the leading cause of climate change. It can easily be understood the vulnerability and extremity of climate change in Asia, and humans contributing a large portion to it in making it worse.

The conducted analysis focuses vastly on finding interesting insights and trends of climate changes in Asian regions which would help to create alertness among the frontiers affected nations and make them take necessary precautionary measures. This analysis would warn countries that are at risk of extreme climate change and would give us a good overview of climate change in Asia for further research. The study also analyzes if pausing human activity for some months can have an effect on climate change or not.

II. Analytical questions

The main goal of the study is to extract interesting insights into weather trends and climate changes in regions of Asia and to compare a certain part of weather data where human activity pausation occured to other data to see if there are interesting statistics and facts to consider.

Following this motivation, this analysis would answer the given questions:
1. What is the current climate trend in Asia and its region?
2. Which regions are affected most because of climate change?
3. Was there any change in climate during the Covid lockdown phase?
4. Would the climate in the future be affected considering human activity pausation like the lockdown happened during Covid period?

The question that is set to answer in the analysis would be very interesting observations, it would not only give an overview of the trend of weather in Asia but also would give an insights into the weather changes in various Asian regions and locations. Besides, research and analysis have been done in the past stating that human activities are the prime reason for global warming [3]. As the covid lockdown period is also included in the studied dataset time zone, It would be interesting in answering whether pausing human activities for some months can have an effect on present and future climate change or not.

III. Data

For this analysis, over 32 years of everyday weather data, from 1 January 1990 to 1 December 2022 for 47 capital cities of Asia are considered. Everyday weather data of all the capital cities collected from database of NASA known as POWER in csv format [4]. Data for 47 cities are collected separately from the website and merged together to form the primary dataset consisting of 565,081 rows and 11 columns.

For getting insights and more details about the data, two additional datasets are collected from stitisticstime.com and whereig.com consisting of the corresponding regions and countries which are later added to the previously merged primary weather dataset. The final dataset is renamed, cleaned and made which consists of 565,081 entries and 13 different columns. The main features are temperature and precipitation columns for analysis, rest of the features and columns would act as a supporting reference for the analysis.

Details of the main dataset:

Column Name	Details	Data type
city	Name of the city	object
country	Associated name of the country	object
region	Asian Sub-Region	object
month	Denotes month	int64
day	Denotes day	int64
year	Denotes year	int64
temp_min	Minimum temperature (C)	object

temp_max	Maximum temperature (C)	object
temp_range	Temperature range (C)	object
precipitation	Amount of rain (mm)	object
atmospheric_pressure	Surface pressure (kPa)	object

For observing climate change a vast amount of weather data is required. A suitable and large amount of historical weather entries are there in the dataset which would be suitable for efficiently taking out insights and also comparing certain periods to other periods. The key element of the data is its diversity and a vast collection of data. Besides, weather data from the lockdown period is also reflected in the dataset, which is one of the targeted analytical questions. As the dataset is manually scrapped, it gives opportunity to have the latest and up to date data which would give more reliability to the findings and results.

IV. Analysis

A. *Preparing and Transforming Data*

1. Data for 47 corresponding capital cities in csv format are loaded in python and merged to form a single dataset.
2. Two other datasets, country and region containing information of the corresponding country and sub-region of Asia are loaded and merged using left join with the main dataset. Function str.rstrip() is called to clean the non-numeric columns of the merged dataset.
3. Columns of the merged and final dataset are lower-cased, renamed and sorted to make the dataset robust and understandable.
4. Among the 565,080 entries, only 141 rows were missing which is only 0.02% of the data. Missing values are removed as the quantity is negligible.
5. A new column, temp_avg is calculated and also added to the main dataset which is the mean of per day temperature.
6. And lastly, data types of some numeric values are in object format which is cleaned and transformed into numeric format. Figure 1 shows the final formed dataset information.

```
Int64Index: 564940 entries, 0 to 565080
Data columns (total 13 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   city                  564940 non-null  object
 1   year                  564940 non-null  int64
 2   month                 564940 non-null  int64
 3   day                   564940 non-null  int64
 4   atmospheric_pressure  564940 non-null  float64
 5   relative_humidity     564940 non-null  float64
 6   precipitation         564940 non-null  float64
 7   temp_max              564940 non-null  float64
 8   temp_min              564940 non-null  float64
 9   temp_range            564940 non-null  float64
 10  surface_temp          564940 non-null  float64
 11  country               564940 non-null  object
 12  region                564940 non-null  object
dtypes: float64(7), int64(3), object(3)
memory usage: 60.3+ MB
```

Figure 1

B. *Data Derivation*

1. Creating Lag Features: For finding the temperature changes for a city over the years, a lag feature is created. Dataset is grouped by city and region to find the aggregated mean of each cities, each year's mean temperature. A new column temp_change_diff is then produced which is the difference of mean temperature and says how increased or decreased the mean temperature of city has been compared to previous year which is shown in Figure 2.

	city	region	year	temp_change_year	temp_change_dif
1	Abu Dhabi	Western Asia	1991	23.641178	0.111397
2	Abu Dhabi	Western Asia	1992	23.232240	-0.408938
3	Abu Dhabi	Western Asia	1993	23.271027	0.038787
4	Abu Dhabi	Western Asia	1994	23.139260	-0.131767
5	Abu Dhabi	Western Asia	1995	23.517055	0.377795
...	...	...	...	...	...
1545	Yerevan	Western Asia	2017	14.344274	0.647662
1546	Yerevan	Western Asia	2018	15.392945	1.048671
1547	Yerevan	Western Asia	2019	14.686164	-0.706781
1548	Yerevan	Western Asia	2020	14.271216	-0.414949
1549	Yerevan	Western Asia	2021	15.279014	1.007798

1457 rows × 5 columns

Figure 2

2. For rest of the analysis task, date is formatted and a new column Date is made merging, day, month and year column. A new column rain_y_n is also created which classifies if it rained on that individual day or not.

C. *Data Analysis*

Trend of Temperature in Asia

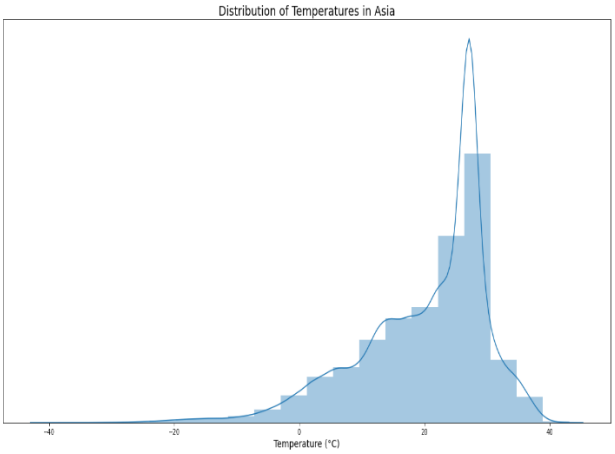


Figure 3

Overall, the distribution of mean temperature in Asia is left-skewed. Here, it is also observed that the mean temperature around of 30-degree Celsius has the highest peak among the other compared bars.

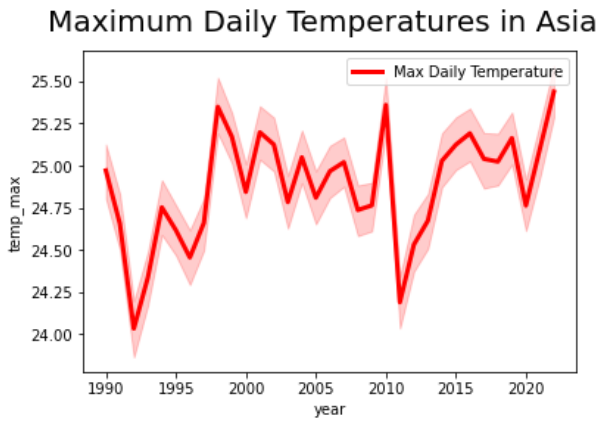


Figure 4

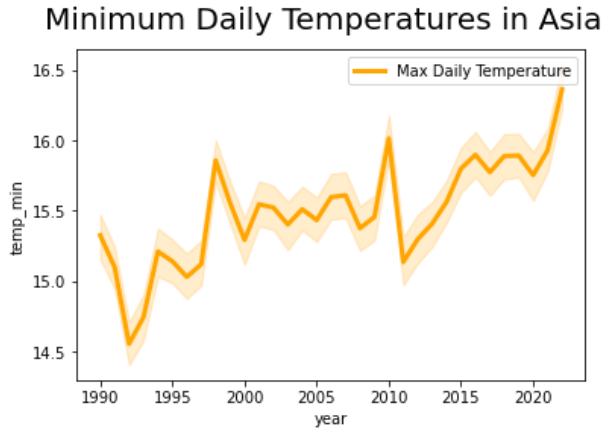


Figure 5

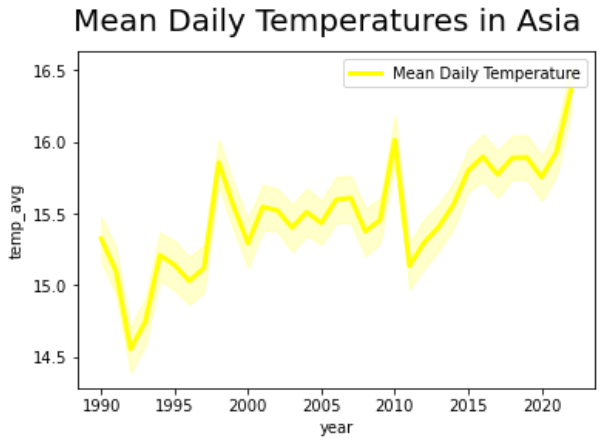


Figure 6

The overview of the growth of the temperature of Asia gives us the indication that by whatever direction we look or do our analysis, the temperature of Asia is seen to be increasing for all the aspects.

Trend of Temperature and Relative Humidity

For finding the trend for all months of the year, a linear regression model from sklearn library is used. Linear Equation, Y =mX + c would give trend and relationship.

	temp_avg	relative_humidity
1	0.008168	0.062107
2	0.020590	0.058394
3	0.031154	0.080464
4	0.018925	0.086406
5	0.021885	0.090533
6	0.016255	0.084125
7	0.021877	0.077406
8	0.020883	0.058496
9	0.030172	0.055580
10	0.013178	0.099472
11	0.010005	0.122091
12	0.004390	0.060997

Figure 7

It is observed that for all 12 months in the year, the average temperature and relative humidity has a positive correlation and upward trend. A positive correlation is seen with temperature and humidity increasing 0.018 deg Celsius peryear and Relative Humidity increasing 0.078 (unit) per year because of global warming.

Temperature changes over the years

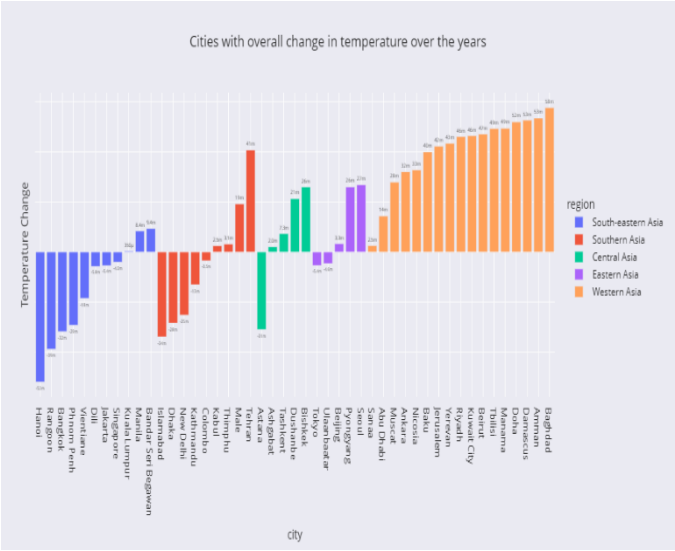


Figure 8

It shows 32 out of 47 total cities temperatures increased over the years. South Eastern Asia showed a reduction in

temperature over the years, but all other region's temperatures went on increasing over the course of 32 years of data.

Comparing lockdown period temperature and rainfall to historic average

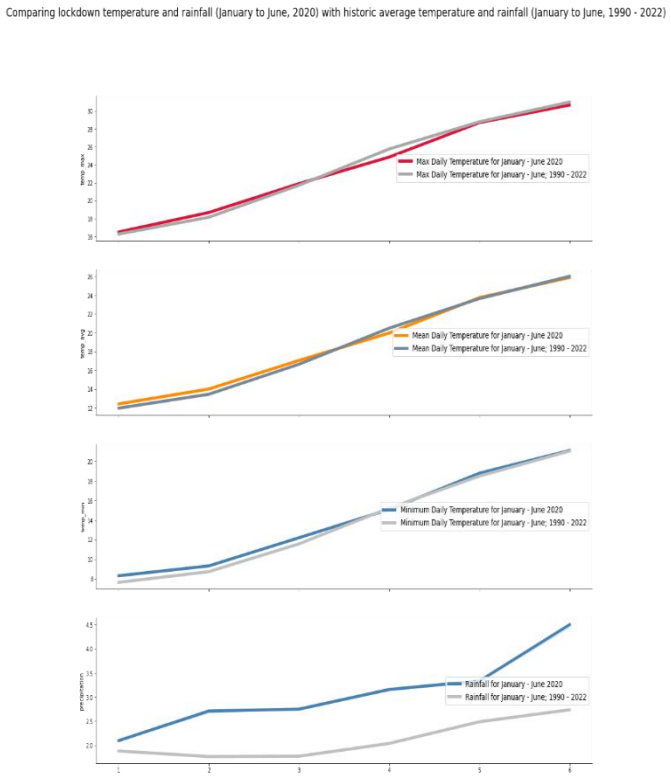


Figure9

Comparing temperatures and rainfall data of six-month time period of covid lockdown period, January to June, 2020 with historical average data of six months, January to June, 1990 to 2022, excluding January to June 2020, it can be observed maximum daily temperature slightly reduced in lockdown period. A drastic growth in rainfall can be observed in this period, the average rainfall during lockdown is very high compared to the average rainfall occurrence in Asia during those 6 months. Minimum and mean temperature remained almost similar in both cases.

D. Construction of Model

For the analysis, two similar Decision Tree models are constructed and applied on the original dataset and a new slightly modified dataset to predict rainfall.

The modified dataset contains all the data for 32 years excluding the covid lockdown period data, which is considered for first six months of 2020, from January to June. Comparing the accuracy score of the same model for the original and the modified dataset would give the idea of the impact of human pausation on climate change. Here rainfall occurrence is classified as 1 and no rainfall is classified as 0.

For training and testing, data are divided in both cases as 80% for training the model and 20% for testing the model with out making random shuffles. For timeseries datasets random shuffles would lower the efficiency in many cases for the model.

0.7116507947746663

	precision	recall	f1-score	support
0	0.55	0.56	0.55	36126
1	0.79	0.78	0.79	76862
accuracy			0.71	112988
macro avg	0.67	0.67	0.67	112988
weighted avg	0.71	0.71	0.71	112988

Figure 10

An accuracy score of detecting rainfall of 71.16% is achieved when the entire dataset including covid period is considered.

0.7050180628695699

	precision	recall	f1-score	support
0	0.54	0.56	0.55	35667
1	0.79	0.77	0.78	75611
accuracy			0.71	111278
macro avg	0.66	0.67	0.66	111278
weighted avg	0.71	0.71	0.71	111278

Figure 11

On the contrast, while testing the model with similar features on the modified dataset which excludes 6 months of lockdown period data, an accuracy in detecting rainfall falls down to 70.50%

E. Validation

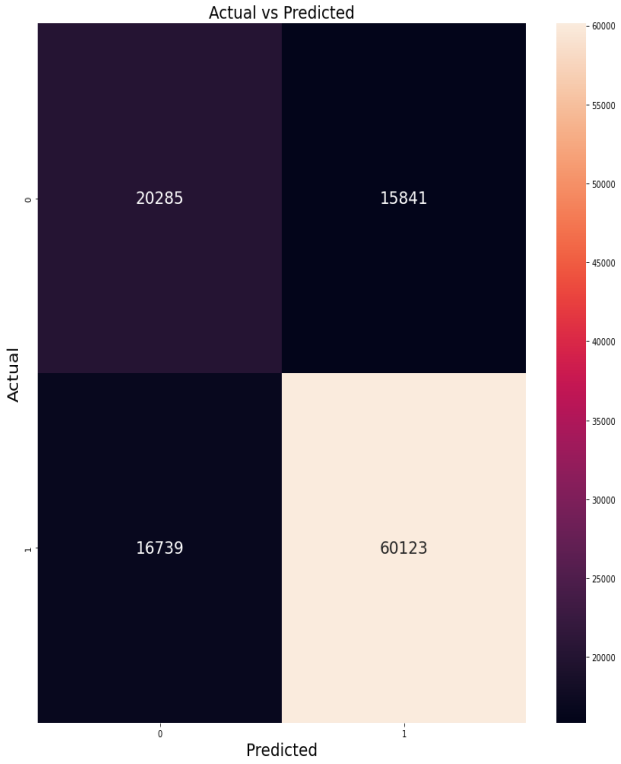


Figure 12 Confusion Matrix for Model applied on original dataset

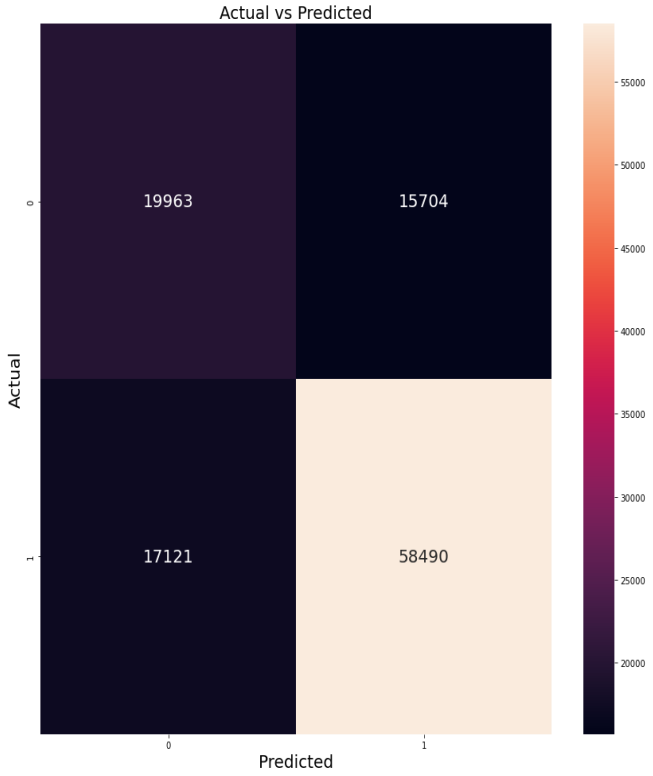


Figure 13 Confusion Matrix for Model applied on modified dataset

Though in both cases for the validation result almost similar accuracy and score is achieved, but the recall and f1-score for detecting true positive is slightly higher considering the original dataset compared to the modified one. As a whole in both cases the validation result looks promising considering the decision tree model and the feature used for modeling.

V. FINDINGS, REFLECTIONS AND FURTHER WORK

Trend of Climate

The overall trend in the climate of Asia shows that the temperature remains generally in the warmer part and temperature is increasing at a constant rate. Observing above figure 3 to figure 6 indicates that.

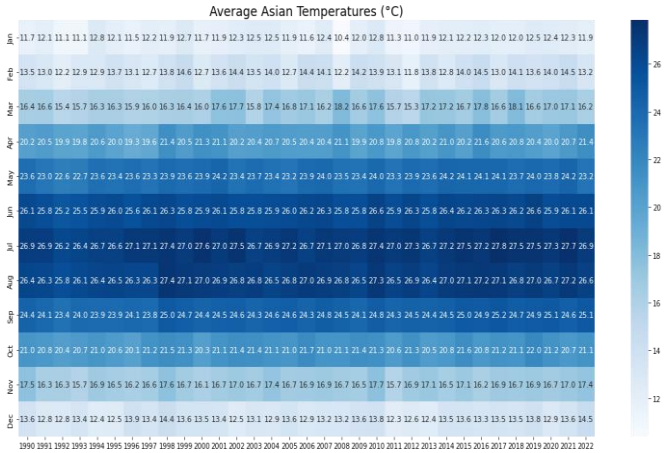


Figure 14

Looking at the heatmap, hotter and cooler months of the year from Asia can be assumed. May to September generally remains the hot, with July being the hottest and over time the temperature in the month of July is seem to be increasing gradually.

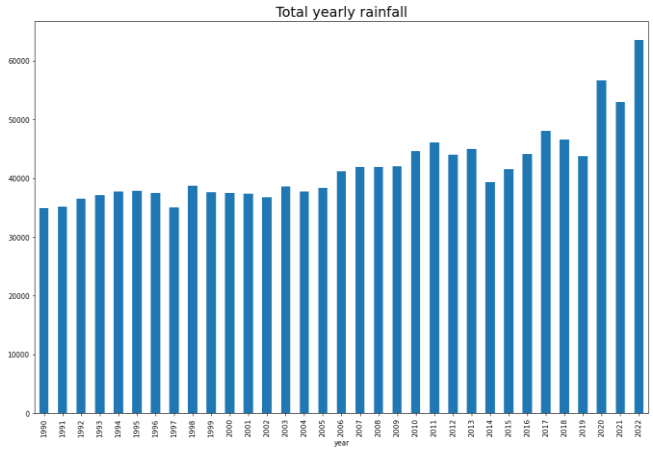
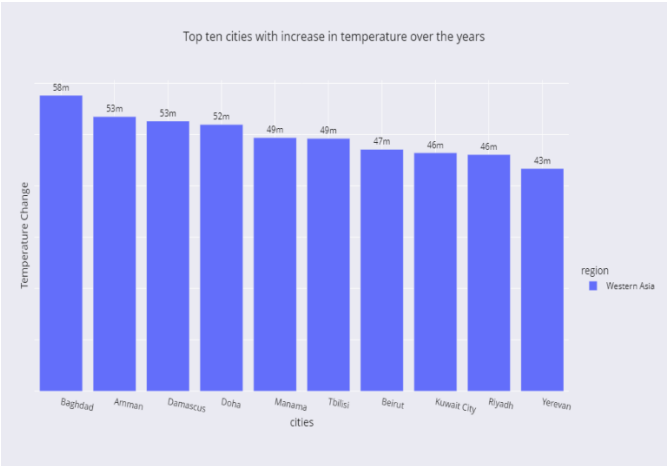


Figure 15

The rainfall graph presented above also gives us reference about the upward rise of rainfalls in Asia. From the analysis conducted while differentiating covid lockdown rainfall and average rainfall it was evident that rainfall is in a positive direction. The above graph is also an evident and a strong reference to that analysis.

Highest climate change experienced region

From the analysis above, with reference to figure 8, it was evident apart from South Eastern Asia, all other Asian regions were highly affected by rise of temperature every year.



In terms of temperature increase, the most affected city is Baghdad, followed by Amman, Damascus, and others. From the analysis it can be observed, the temperature is increasing in Baghdad at a rate of 0.057 degree-Celsius, followed by 0.053 and 0.053 degree-Celsius for Amman and Damascus respectively. Surprisingly, all of the cities in top ten lists for

an increase in temperature belong to the Western Asia region, which is very alarming for that region.

Pausation of human activity in climate change

From the analysis carried out on figure 9, it was totally clear that covid lockdown and human activity pausation at that time had effect on climate.
During that period, normal human activity was hampered and there were reduction in human activity which resulted in making the environment less hampered. As a result of which it had some effect on climate.
Also, after implementing similar machine learning models in detecting future rainfalls, there were some affect on the accuracy and score when the lockdown period data was not considered. The model gave less accuracy and score in predicting the future outcomes. Though the number was very less but still not considering the lockdown period had affect on the overall future result.
Considering this, it can be stated and related to our 3rd and 4th analytical question that there were minor affect on climate due to human activity pausation and if similar condition arises or if human activities are reduced in the future there are chances of that in affecting the overall future climate.

Further Work

The conducted analysis did not consider abnormal weather conditions. Storms, unusual rainfalls, flood etc are the abnormal weather conditions or outliers which can be taken in to consideration while doing further research.

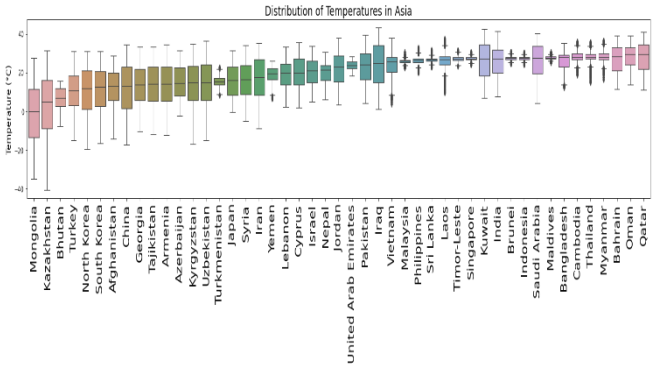


Figure 16

The above graph gives indication of the outliers in case of considering abnormal temperatures which is normal to occur in the world.

Besides, conducted research worked with data from overall Asia, from this similar approach and method can be considered to climate analysis on a global scale. The overall global condition and individual standings of the world countries can be achieved and a overall metrics can be created.

Conclusion

The overall trend of climate of Asia is alarming and is increasing as the day progresses. This problem should be taken into consideration. Besides, from the analysis affect of pausing human activity on present and future climate can be understood. If human activity is paused on a regular basis for a long span of time it can be possible to manipulate the overall climate of the world.

References

[1] "Why is climate important?," *NCAS*. https://ncas.ac.uk/learn/why-is-climate-important

[2] "Asia's Climate Emergency," *International Monetary Fund*, Aug. 31, 2021. Accessed: Aug. 15, 2022. [Online]. Available: https://www.imf.org/en/Publications/fandd/issues/2021/09/asia-climate-emergency-role-of-fiscal-policy-IMF-dabla

[3] Z. Hausfather, "Analysis: Why scientists think 100% of global warming is due to humans," *Carbon Brief*, Dec. 13, 2017. https://www.carbonbrief.org/analysis-why-scientists-think-100-of-global-warming-is-due-to-humans

[4] "NASA POWER | Prediction Of Worldwide Energy Resources," *Nasa.gov*, 2012. https://power.larc.nasa.gov/

Word Count

Introduction- 261 words
Analytical Questions- 219 words
Data- 261 words
Analysis- 883 words
Findings, Reflection and Further Work- 599 words