Modern Data Analytics Report

Analyzing the Impact of Global High Temperatures on Population and Mortality Rates

Liren Xie R0864366

KU LEUVEN

Introduction

Heatwaves are becoming more frequent and intense around the world, and their impact on human health is a growing concern. The purpose of this study is to analyze the relationship between high temperatures and mortality rates, and to identify the key factors that contribute to this relationship. By doing so, we hope to provide insights into the potential consequences of heatwaves on human health, and to inform efforts to mitigate their impact.

Methodology

To conduct this study, we used a variety of data analysis and prediction models, including RNN and XGBoost. Our approach was scientifically rigorous, using established statistical methods and machine learning techniques to analyze the data and draw conclusions.

RNN:

One of the key challenges in analyzing the impact of high temperatures on mortality rates is the time-series nature of the data. Mortality rates and temperature data are both collected over time, and the relationship between the two variables may change over time as well. To account for this, we used a recurrent neural network (RNN) to analyze the data.RNNs are a type of neural network that are designed to handle sequential data,

such as time-series data. They are particularly well-suited for analyzing data that has a temporal structure, as they can capture the dependencies between data points over time. In our analysis, we used an RNN to model the relationship between high temperatures and mortality rates over time, and to make predictions about future trends.

XGBoost:

In addition to RNNs, we also used XGBoost to analyze the data. XGBoost is a powerful machine learning algorithm that is widely used in data analysis and prediction tasks. It is particularly well-suited for handling large datasets with many features, as it can handle both numerical and categorical data.In our analysis, we used XGBoost to identify the key factors that contribute to the relationship between high temperatures and mortality rates. We used XGBoost to analyze the data and identify the most important features, such as temperature thresholds and duration, that contribute to the harmful effects of heatwaves.

Overall, our use of RNN and XGBoost allowed us to analyze the data in a rigorous and comprehensive manner, and to identify the key factors that contribute to the impact of high temperatures on mortality rates. By using these advanced data analysis and prediction models, we were able to extract valuable insights from the data and make predictions about future

trends.

Data Description

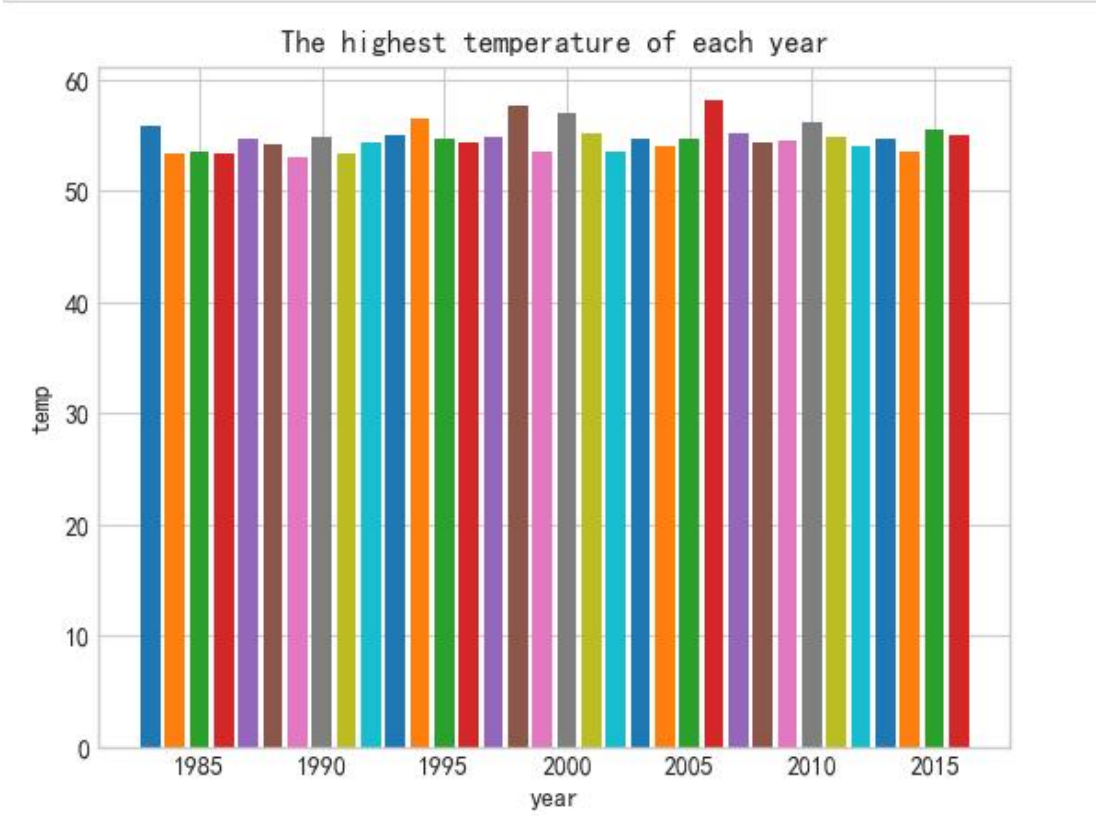
In this study, we used population and mortality data from the Socioeconomic Data and Applications Center (sedac) for 21 regions globally from 1983 to 2016 and the relevant heat and temperature in different countries and cities data as reference.

Results Analysis

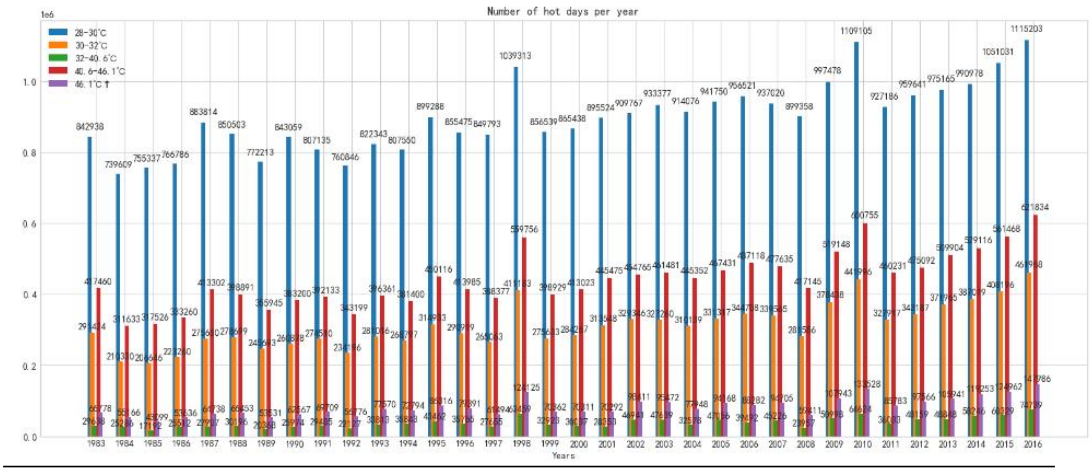
In this chapter, we present the results of our analysis. We describe the relationship between high temperatures and mortality rates, and identify the key factors that contribute to this relationship. We also analyze the impact of temperature thresholds and duration on mortality rates, and make predictions about future trends based on our analysis.

First, we conducted the most basic analysis of the data. The graph below shows the maximum temperature recorded each year, and it can be seen that the highest temperature each year globally is around 55

degrees

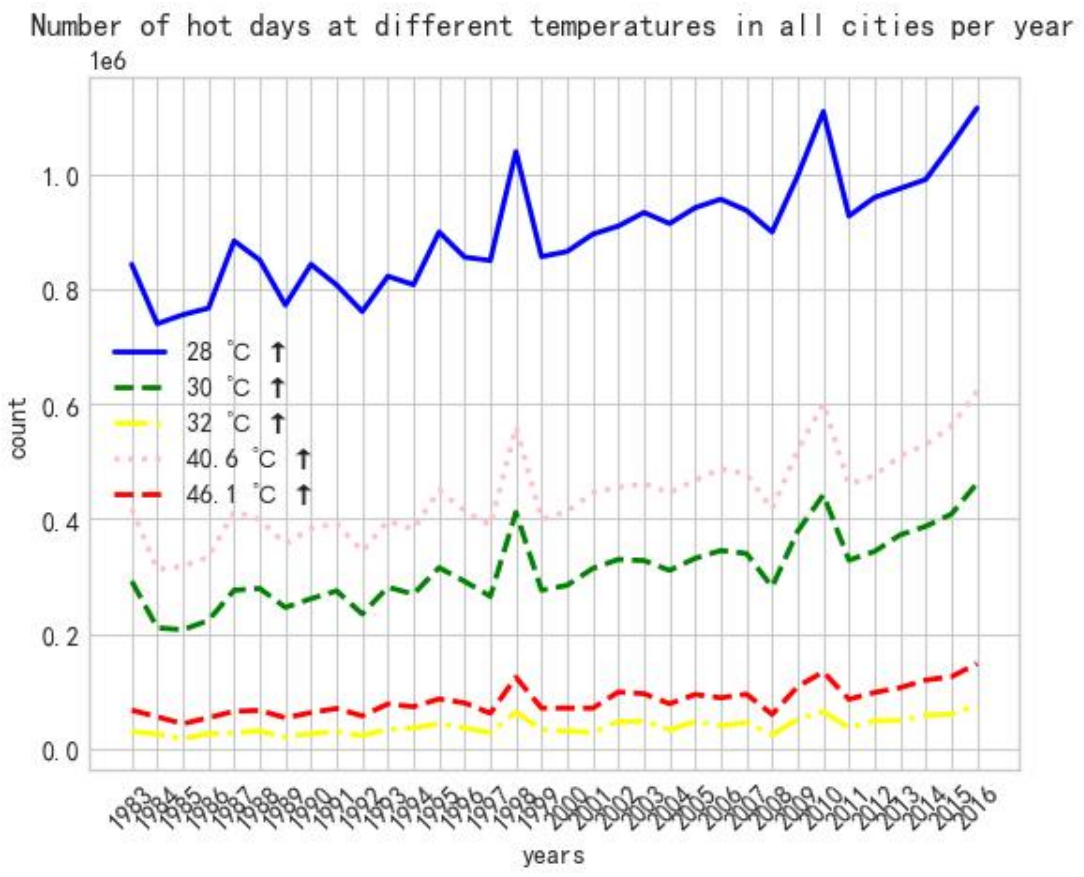


The figure below shows the statistics of the number of high temperature days per year. It can be seen from the histogram that the number of high temperature days is increasing every year, which is also related to global warming.



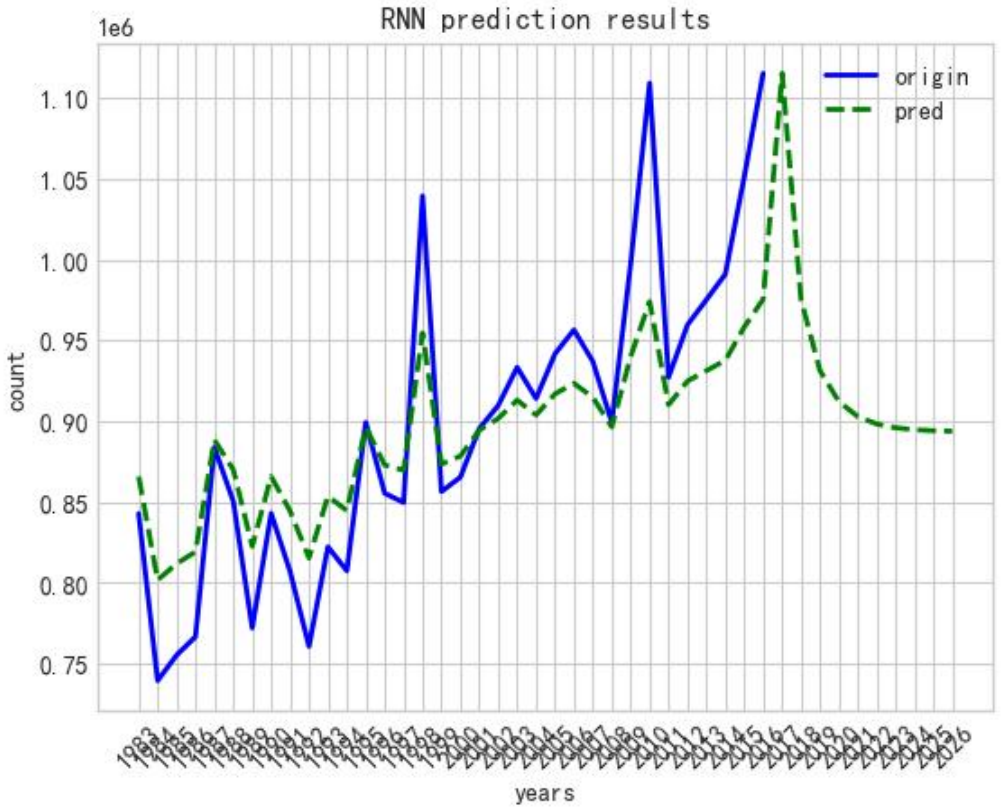
We got the sum of the high temperature days in all cities at different

temperatures each year, and use a line chart to show the sum of the high temperature weather that the world receives each year at different temperatures. We find that the total number of high temperature days in the world is fluctuating over time. Rising, indicating that the global temperature is in the process of rising. For this reason, we need to predict the change trend of the sum of global high temperature weather in each temperature range in the next few years, so as to facilitate our response to high temperature weather in the future.



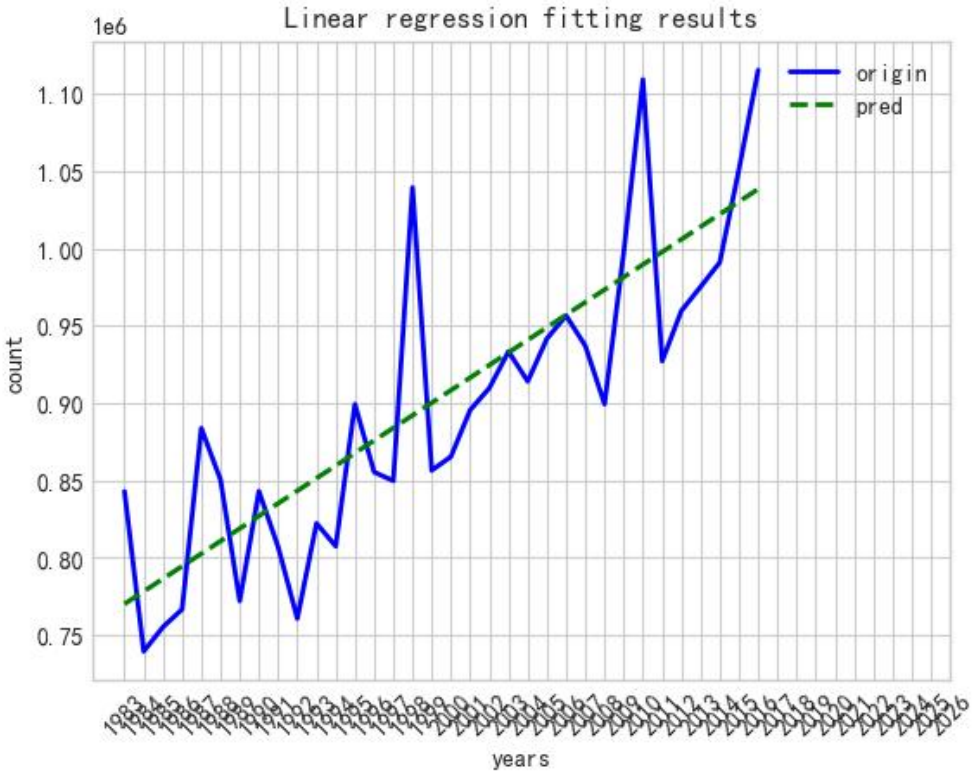
Since this is a very typical time series problem, first we use RNN to train the data. The first step is to normalize the data, because there is a large variance in the data of the same dimension, and the training effect of such

data is often Not ideal, so we normalize the data using MinMaxScaler. Then we divide the data into a training set and a test set according to 8:2, and after preparing the data, transfer the data to the RNN network for training. The prediction of the finally trained model for 2017-2026 is shown in the figure

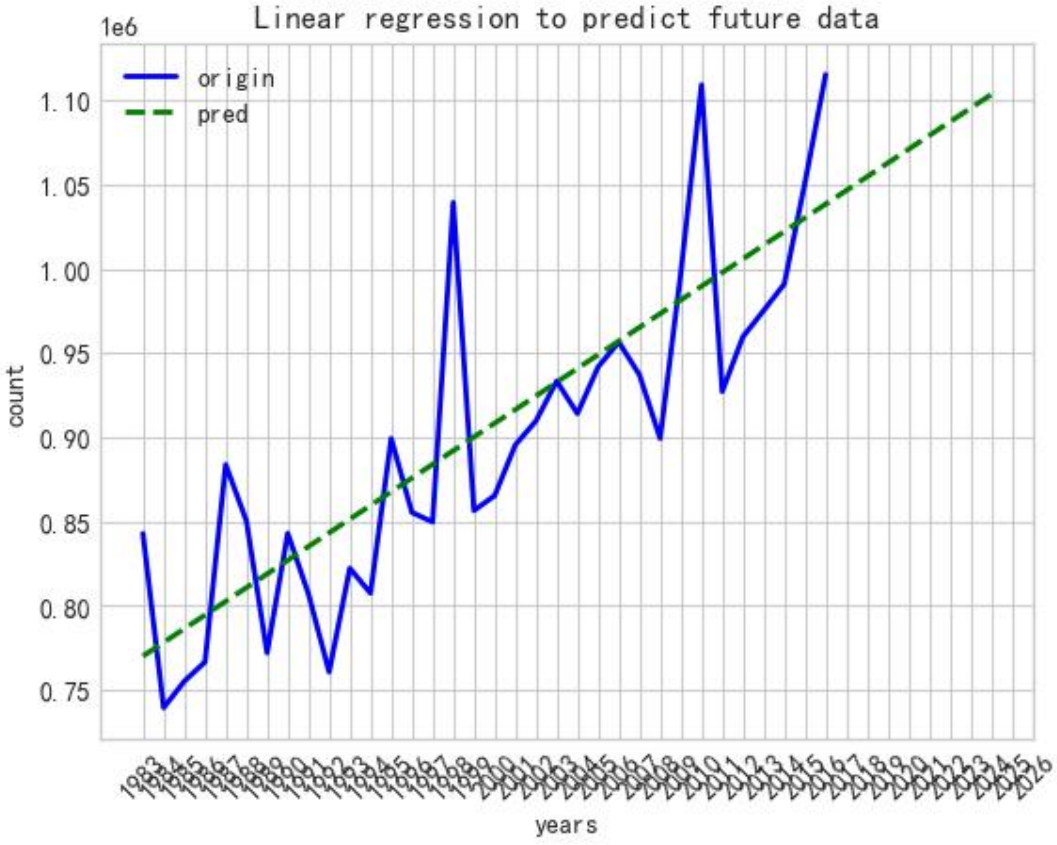


It can be seen that the RNN does not seem to be very accurate in predicting the future temperature. The reason for the analysis is that our data volume is too small, only the data from 1983-2016, so the RNN model has not learned enough data change characteristics, which also As a result, RNN has a lot of errors in each prediction, and the last error is accumulated in the next time, causing the final RNN prediction results to gradually deviate from our expected values.

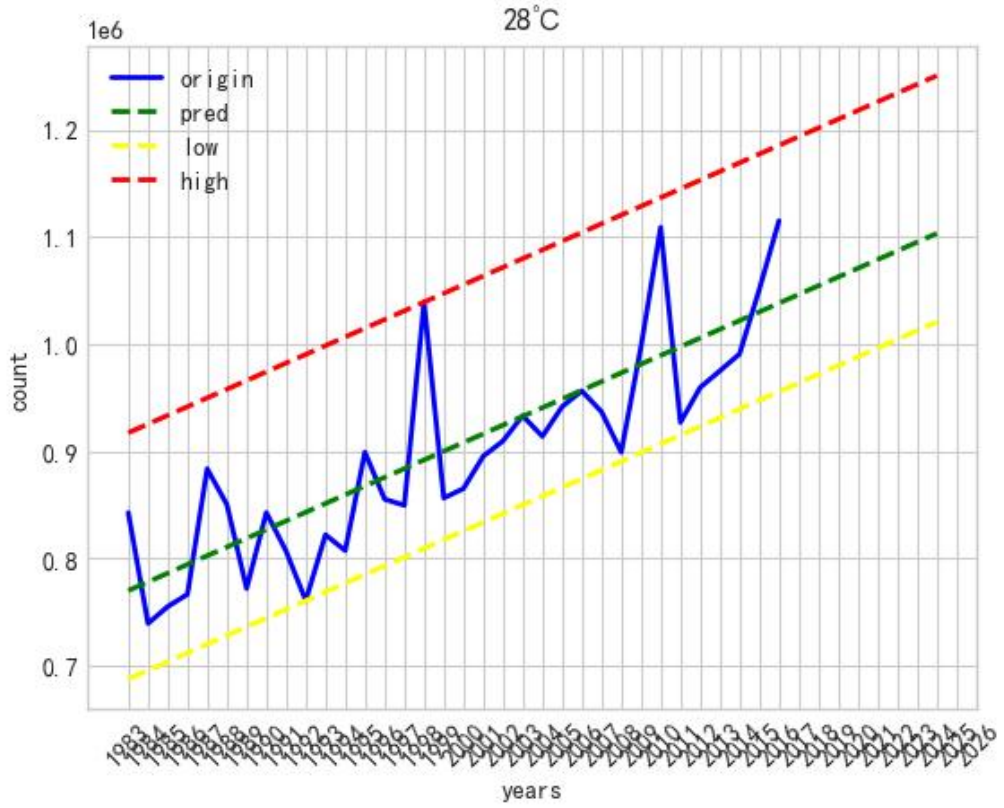
Because the RNN network cannot accurately predict future data, we decided to use simple linear regression for data prediction. The fitting results for the 1983-2016 data are as follows

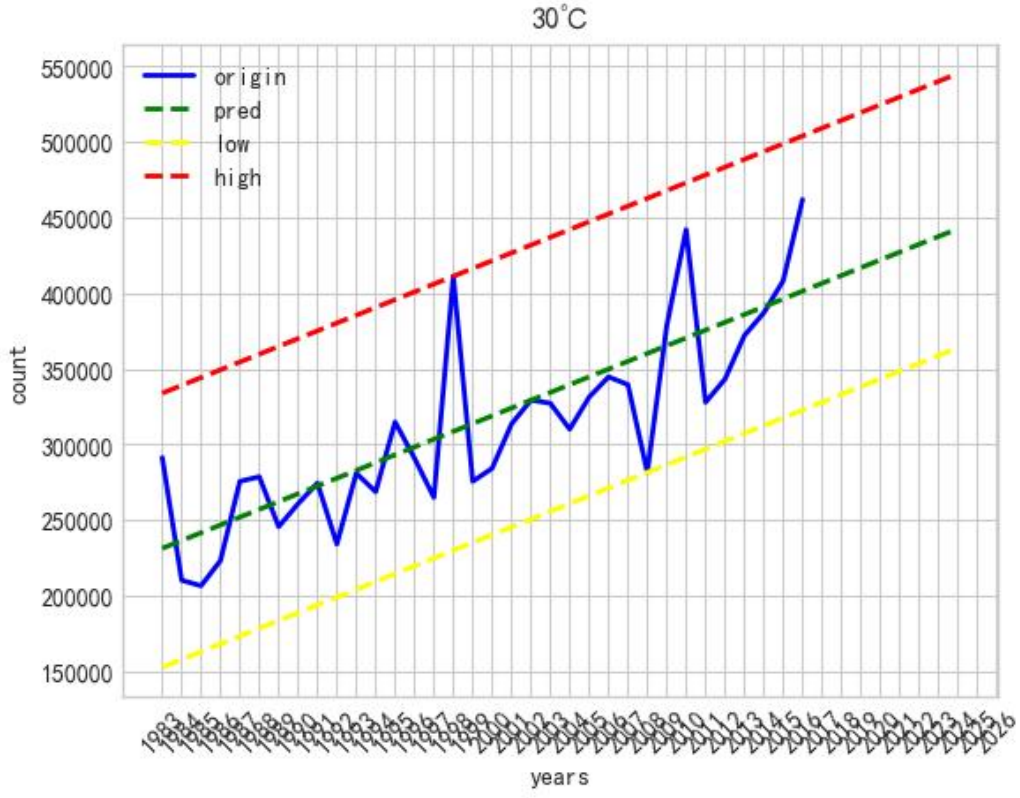


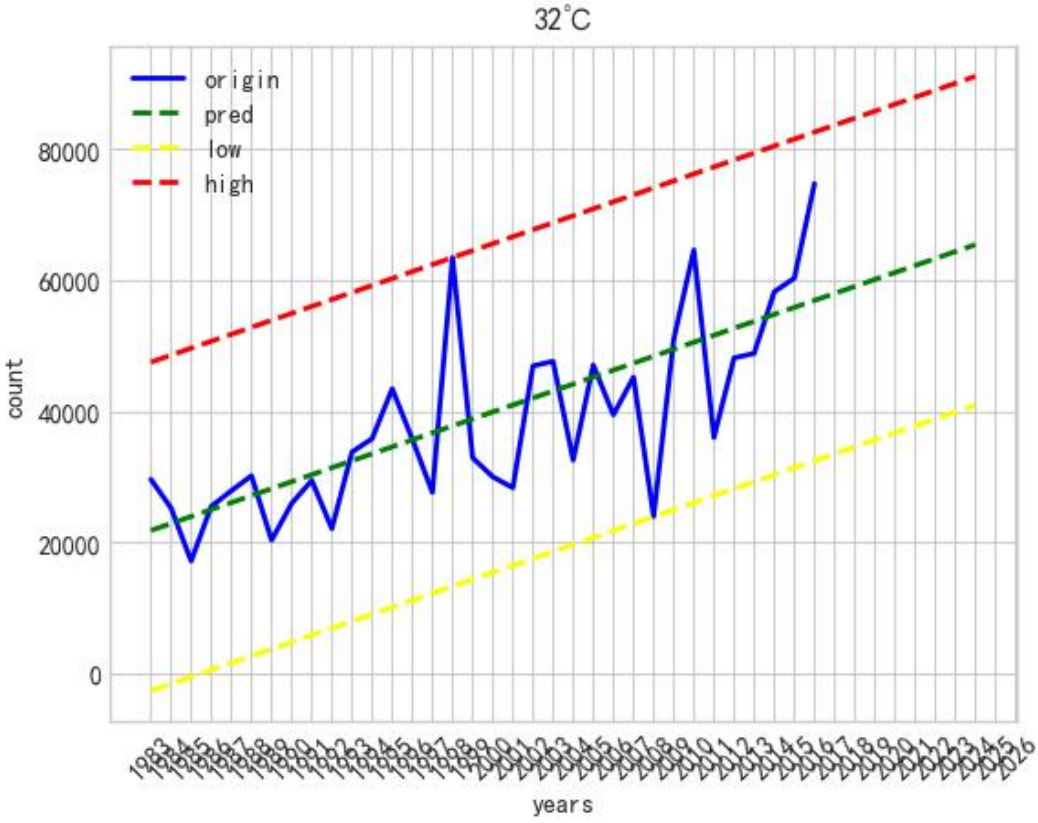
Since there is only one column of data features, we can only do simple linear regression. The prediction results of this model for the future are as follows

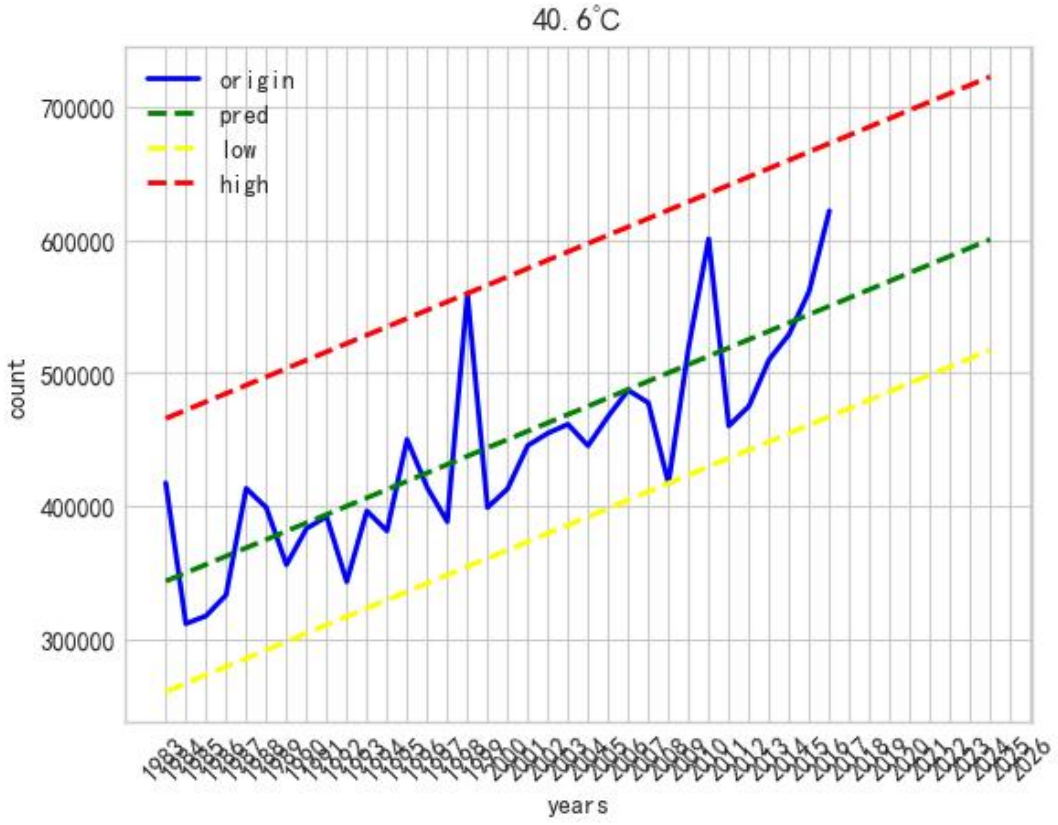


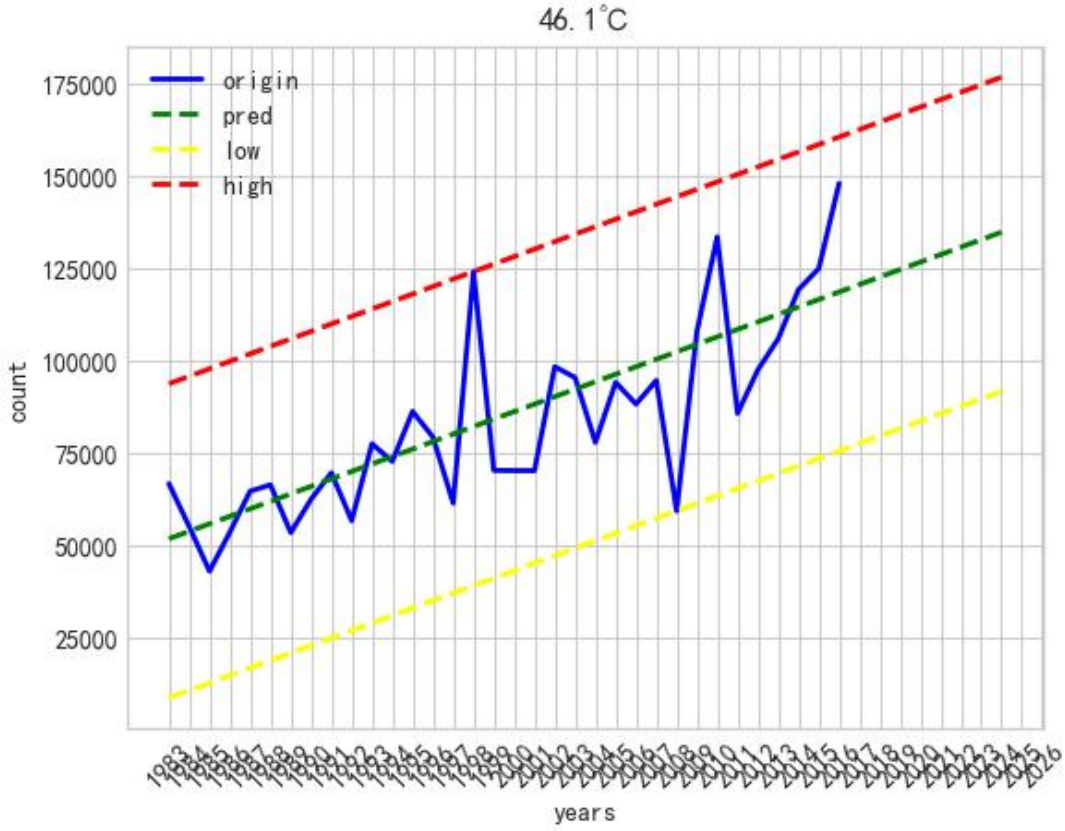
For it is a simple linear regression, the obtained prediction results can only be regarded as the average value of future data, so we calculate the maximum and minimum values that the future data may reach according to the slope of the slope predicted by the model, as shown in the figure below


28°C
1e6
origin
pred
low
high
1.2
1.1
1.0
0.9
0.8
0.7
count
years

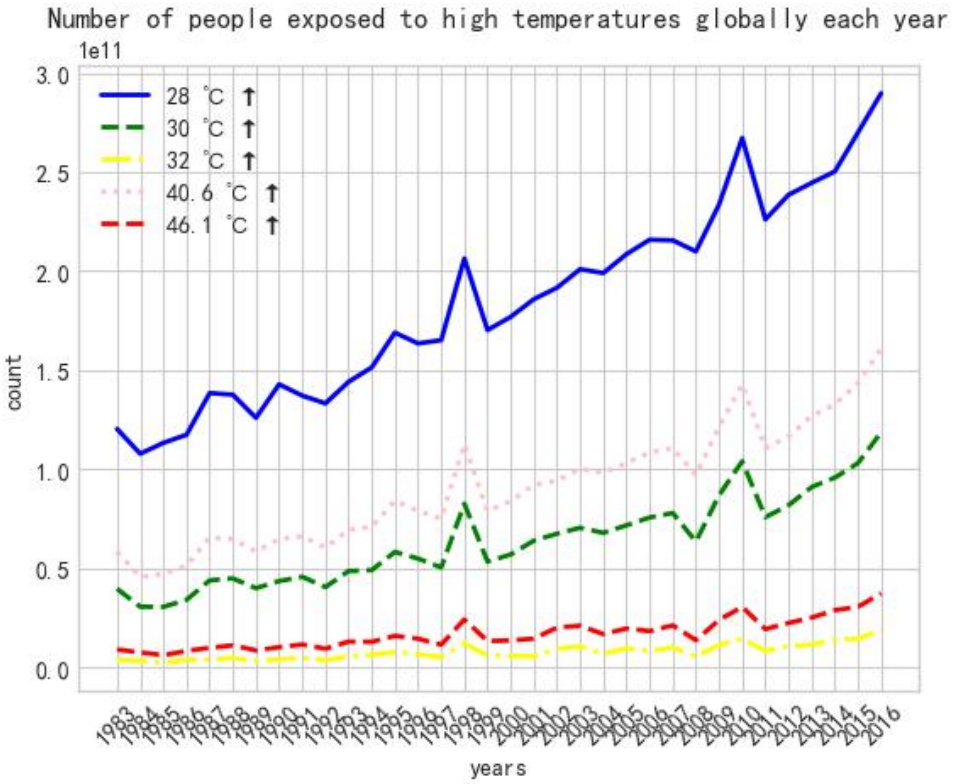

30°C
550000
500000
450000
400000
350000
300000
250000
200000
150000
origin
pred
low
high
count
years


32°C
origin
pred
low
high
80000
60000
40000
20000
0
count
years

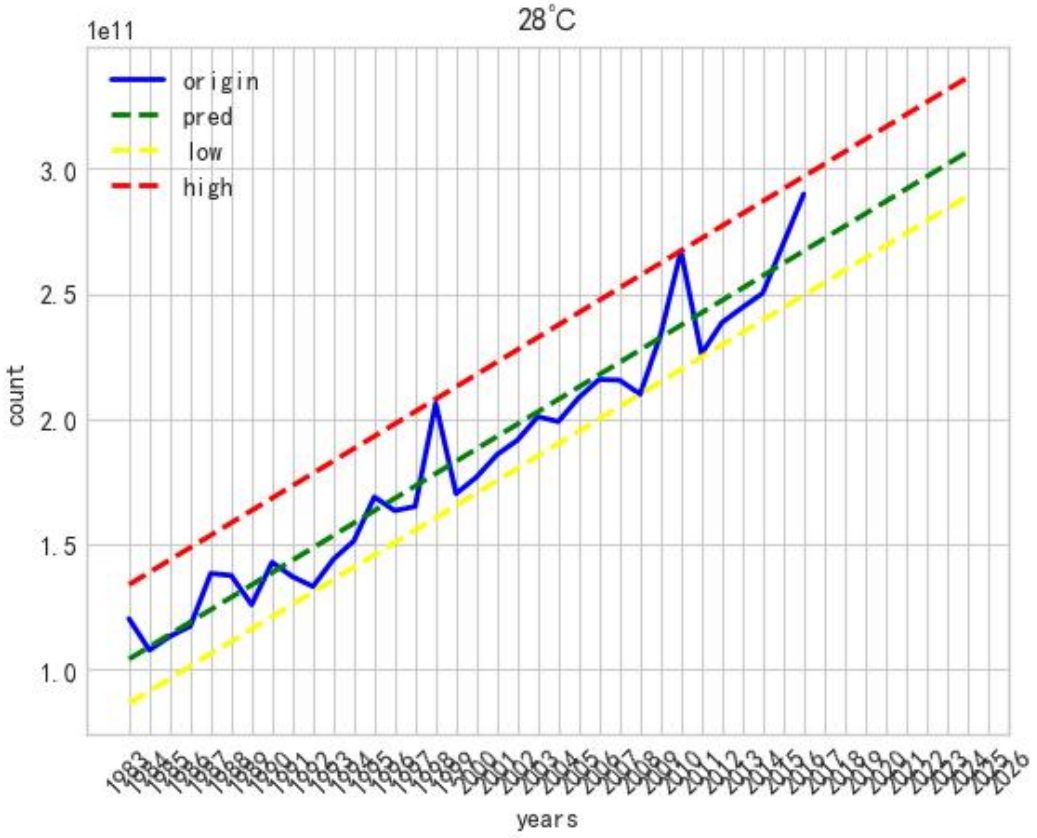

40.6°C
origin
pred
low
high
700000
600000
500000
400000
300000
count
years

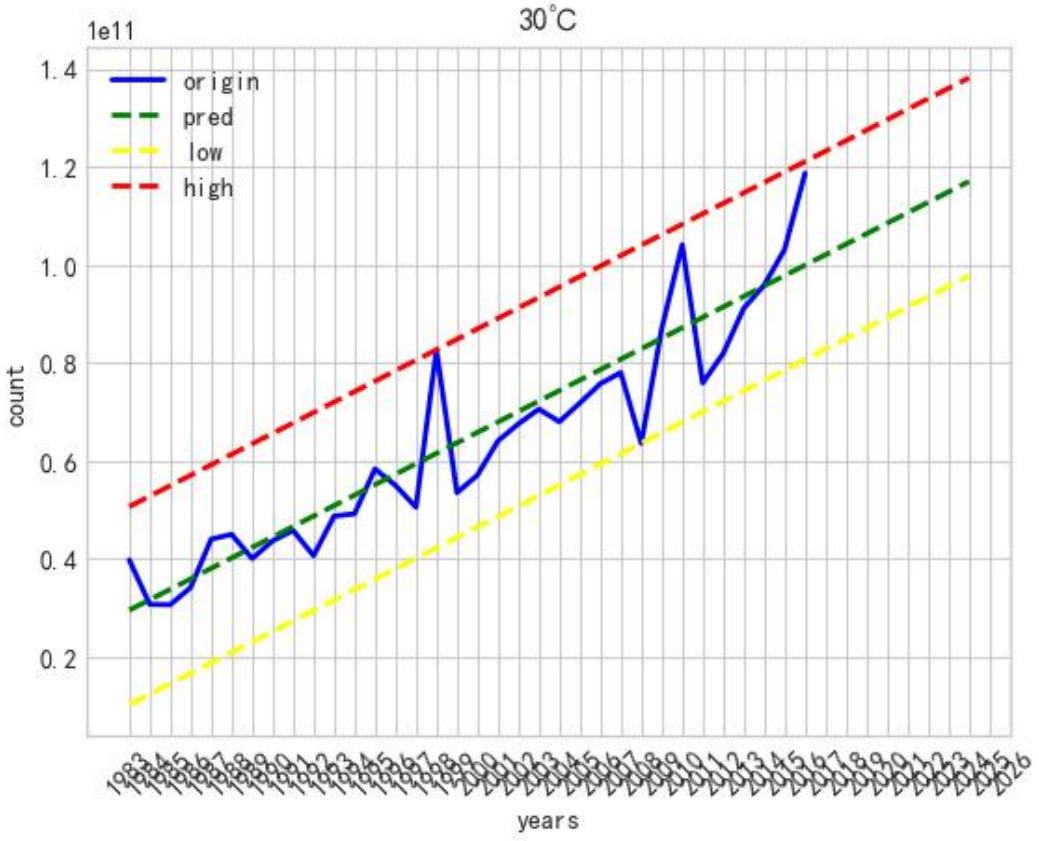


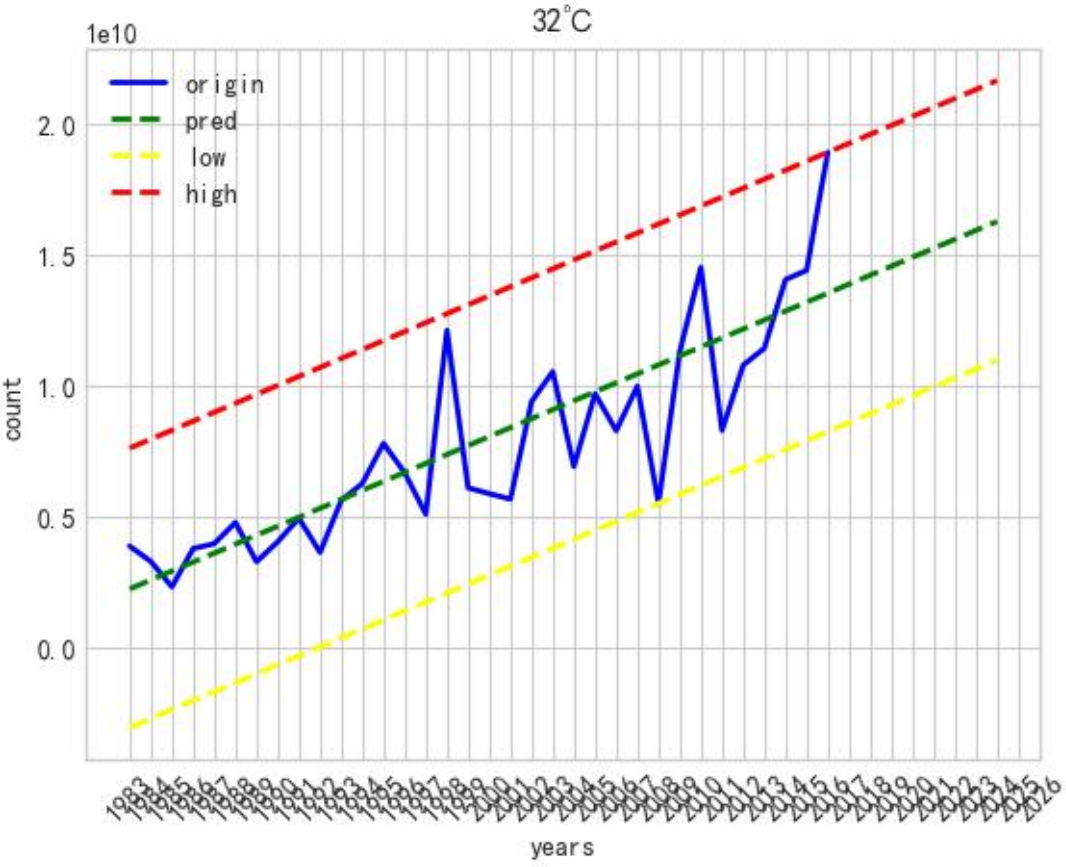
The second part is to calculate how many people around the world are exposed to high temperatures every year. We will represent the number of people exposed to high temperatures above 28 degrees Celsius every year through a line graph

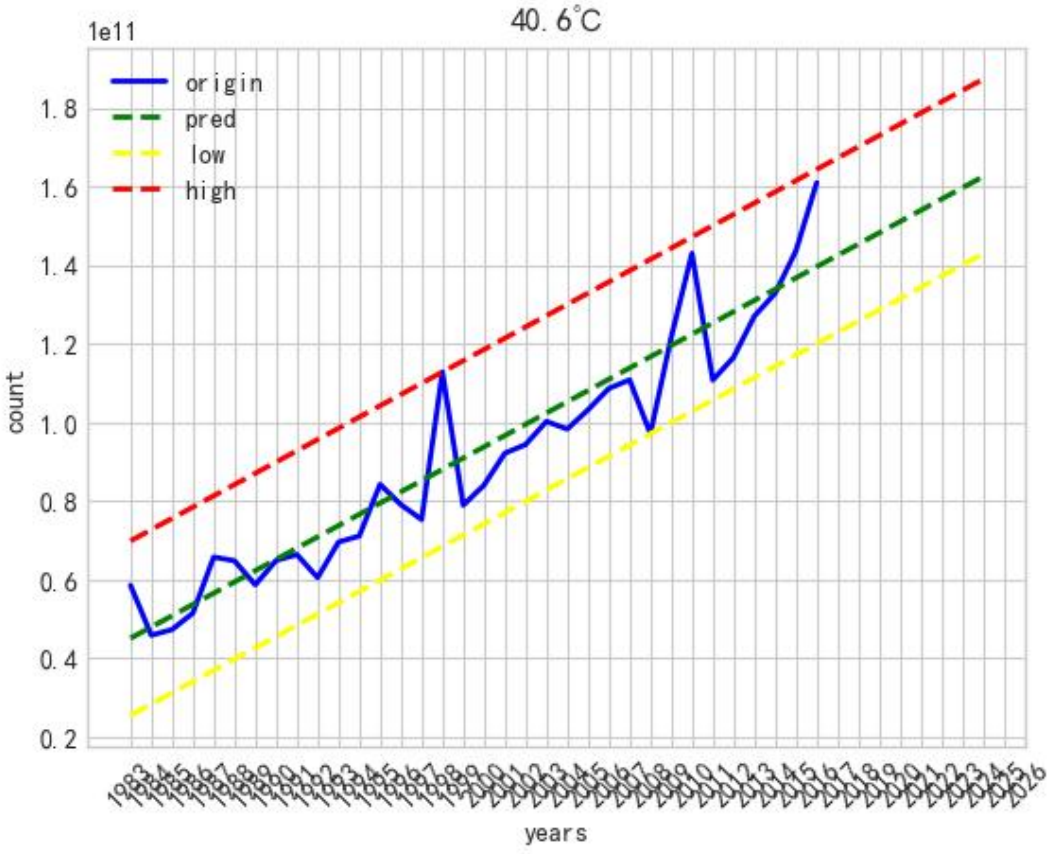


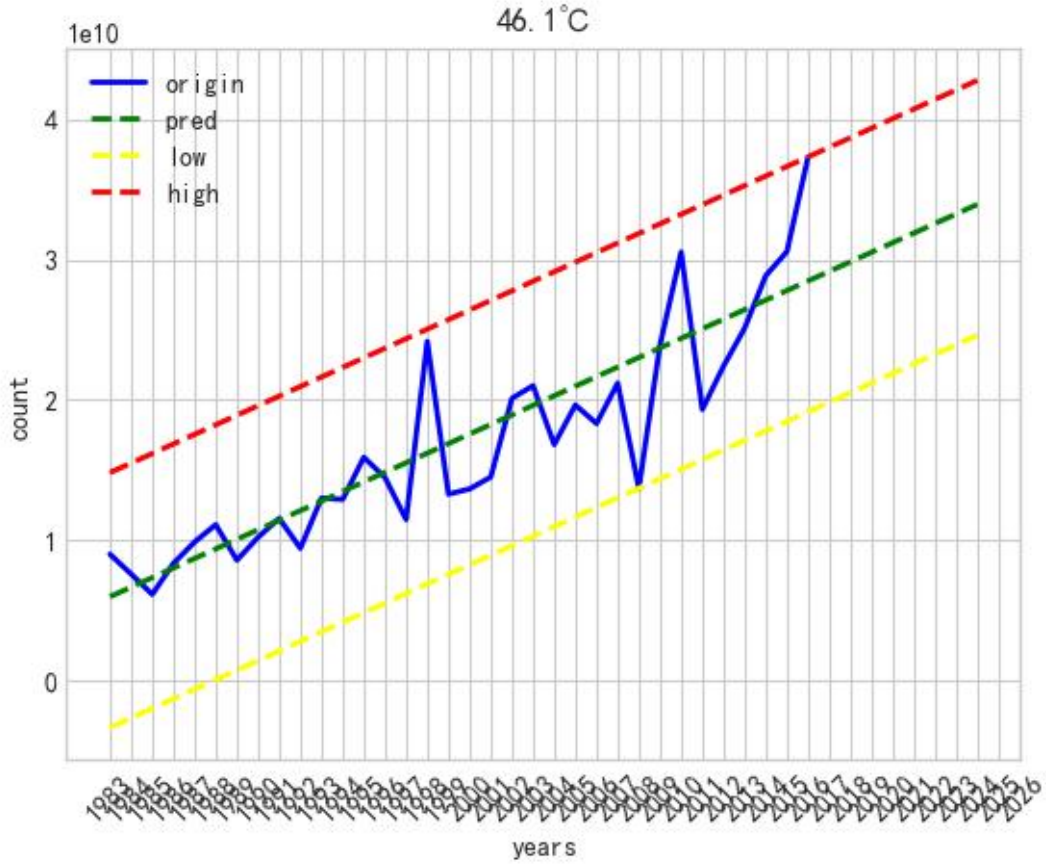
It can be seen that as time goes by, the number of people exposed to high temperature is increasing every year, so we also use the same method as in the first part to predict the number of people exposed to high temperature in the next few years


28°C
1e11
origin
pred
low
high
3.0
2.5
2.0
1.5
1.0
count
years


30°C
1e11
origin
pred
low
high
1.4
1.2
1.0
0.8
0.6
0.4
0.2
count
years


32°C
1e10
origin
pred
low
high
2.0
1.5
1.0
0.5
0.0
count
years


40.6°C
1e11
origin
pred
low
high
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
count
years



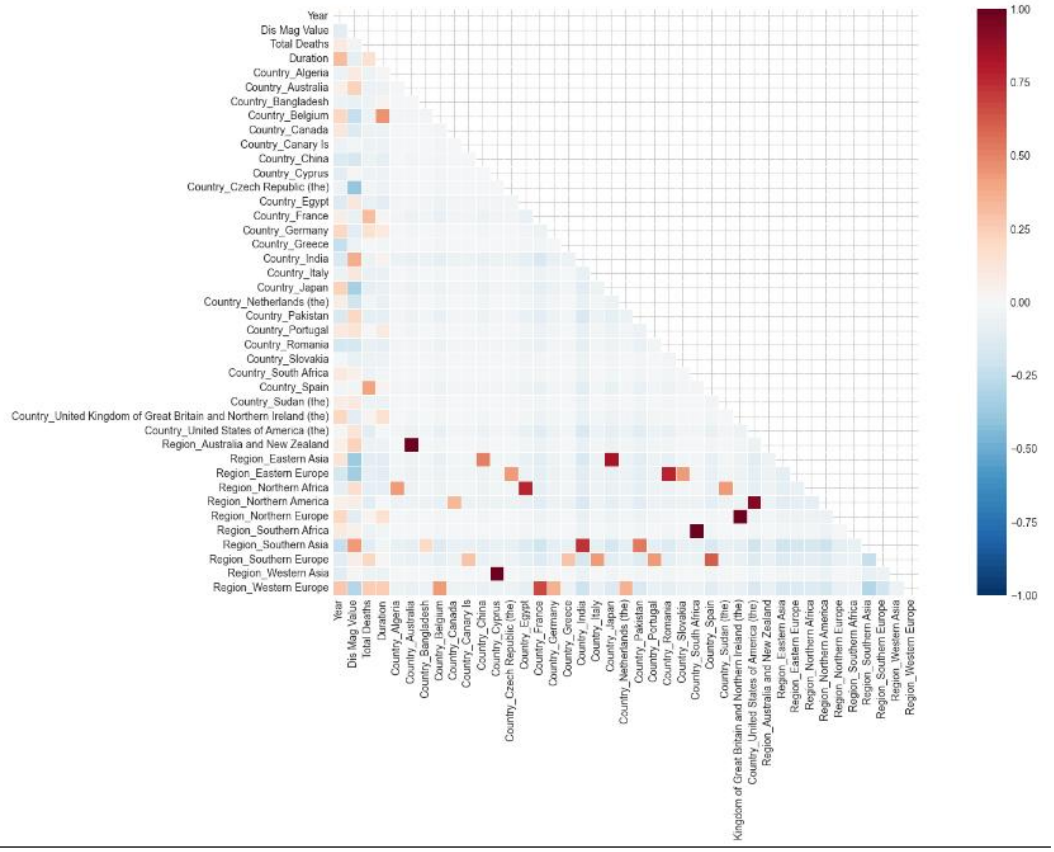
In addition to data analysis on a global scale, the high temperature conditions in various regions of the world are also worthy of our attention. In this part, we will analyze the high temperature conditions in 21 regions around the world. We will show the statistical results in a line chart as follows

The number of people threatened by high temperatures in 21 regions globally gradually increased between 1983 and 2016, especially in Southern Asia, where the increase was rapid. One reason for this is that Southern Asia is located in a low-latitude area, where the impact of global warming on high temperatures is more severe. Another reason is that Southern Asia is mostly composed of developing countries, which are characterized by rapid population growth, such as India. Therefore, the number of people threatened by high temperatures in Southern Asia is growing most rapidly.

After analyzing the impact of high temperatures on the global exposed population, we will analyze the impact of high temperatures on population mortality. In this section, we will use XGBoost, with feature X set as Year, Country, Region, Dis Mag Value, and Duration, and feature Y set as Duration. Since Country and Region are both string-type text, we first encoded these two features using one-hot encoding. One feature of

XGBoost is that it can analyze features, and after model training, the importance of each feature is shown in the table below.

	feature name	importance
2	Duration	0.086807
16	Country_India	0.078827
39	Region_Western Europe	0.073747
33	Region_Northern America	0.069495
32	Region_Northern Africa	0.062151
18	Country_Japan	0.061846
22	Country_Romania	0.054879
17	Country_Italy	0.054049
37	Region_Southern Europe	0.052193
28	Country_United States of America (the)	0.048239
5	Country_Bangladesh	0.048089
24	Country_South Africa	0.045647
3	Country_Algeria	0.045472
1	Dis Mag Value	0.044096
9	Country_China	0.043347
8	Country_Canary Is	0.024387
20	Country_Pakistan	0.020526
36	Region_Southern Asia	0.018436
0	Year	0.018642
25	Country_Spain	0.013068
26	Country_Sudan (the)	0.011166
6	Country_Belgium	0.009318
13	Country_France	0.007967
12	Country_Egypt	0.004374
31	Region_Eastern Europe	0.003216
27	Country_United Kingdom of Great Britain and No...	0.001799
21	Country_Portugal	0.000380
14	Country_Germany	0.000038
38	Region_Western Asia	0.000000
4	Country_Australia	0.000000
35	Region_Southern Africa	0.000000
34	Region_Northern Europe	0.000000
15	Country_Greece	0.000000
7	Country_Canada	0.000000
30	Region_Eastern Asia	0.000000
29	Region_Australia and New Zealand	0.000000
11	Country_Czech Republic (the)	0.000000
23	Country_Slovakia	0.000000
19	Country_Netherlands (the)	0.000000
10	Country_Cyprus	0.000000



It can be seen from the importance of characteristics that the most important factor affecting population death due to high temperature is the duration of high temperature, followed by temperature value.

Conclusion

In conclusion, our study provides valuable insights into the impact of global high temperatures on population and mortality rates. By using rigorous scientific methods and advanced data analysis techniques, we were able to identify the key factors that contribute to the harmful effects of heatwaves. Our findings have important implications for public health policy and planning, and can help to inform efforts to mitigate the impact of heatwaves on human health. By identifying the key factors that

contribute to the harmful effects of heatwaves, we can work to develop more effective strategies for protecting vulnerable populations and reducing the impact of climate change on human health.