CSIC7204 DATA SCIENCE AND DATA VISUALIZATION

Assignment 1-Report

MC464633

Luo Weiming

2024-09-07

Index

1. Introduction

The purpose of this project is to analyze the weather changes in Macau over the last 20 years using Scatter Plot, Line Chart, Box Plot, Heatmap, Bar Chart, Violin chart, and to train a model for future weather prediction using the corresponding data.

2. Dataset

This is an annual weather data provided by the government of Macao, China (*SMG - Geophysical and Meteorological Office of the Macao SAR Government*, n.d.; *What Does Humidity Percentage Mean and What Is High Humidity? | Weather Radio Review*, n.d.) from 1999 to 2019, It contains 7,537 pieces of data. "Date", "Mean maximum (°C)", "Mean(°C) ", "Mean minimum (°C)", "Mean relative humidity (%)", "Insolation duration(hour)", 'Total rainfall (mm)' 7 dimensions, the original data is a total of xlsx format table, in order to facilitate the data processing, the data was converted into a csv file. In the rain fall data, there is "VST" indicating rainfall below 0.2, and "VST" has been replaced with 0.1 for ease of calculation.

3. Data analysis

3.1 Temperature

The data shows three related data: average maximum temperature, average temperature, and average minimum temperature. The temperatures were first averaged for each month to get figure 1

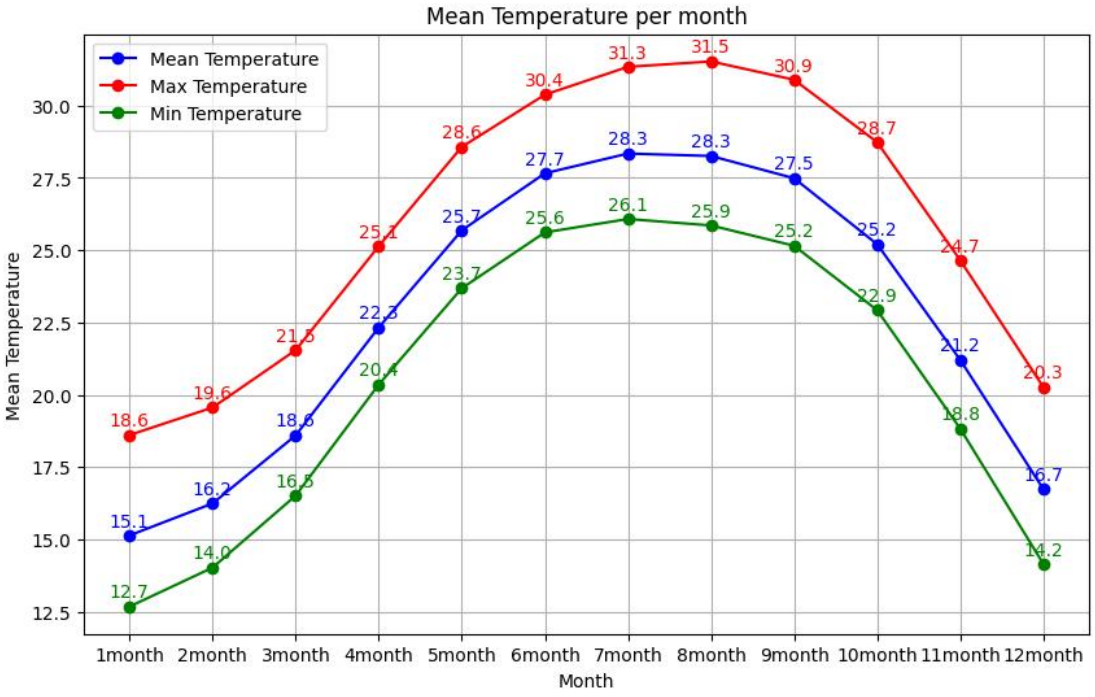

Figure 1 MeanTemperture_Month

We can see from this graph that the overall temperature in Macau is seasonally related, with the lowest in winter and the highest in summer. In terms of changes, Macau has four distinct seasons, and the temperature changes are relatively smooth everyone can adapt well to the temperature changes.

In order to show the temperature difference of each month, a dynamic graph was made using the package dash broad of python to analyze it, and we can freely choose the year to view the corresponding data. The years 2003(figure 2) and 2013(figure 3) have been chosen to do exhibition.

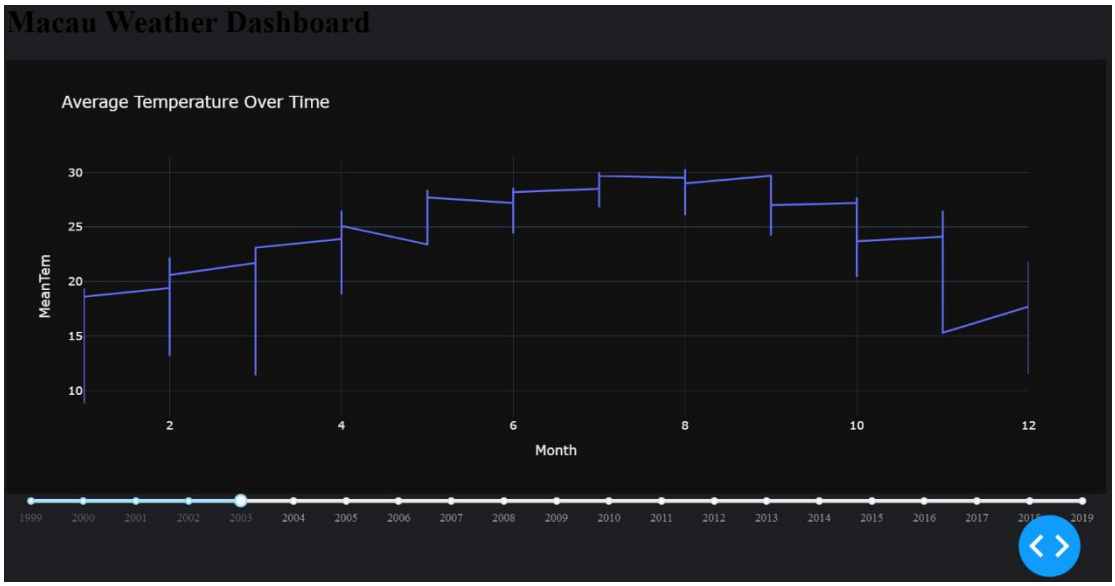

Figure 2 Temperature DashBroad-2003

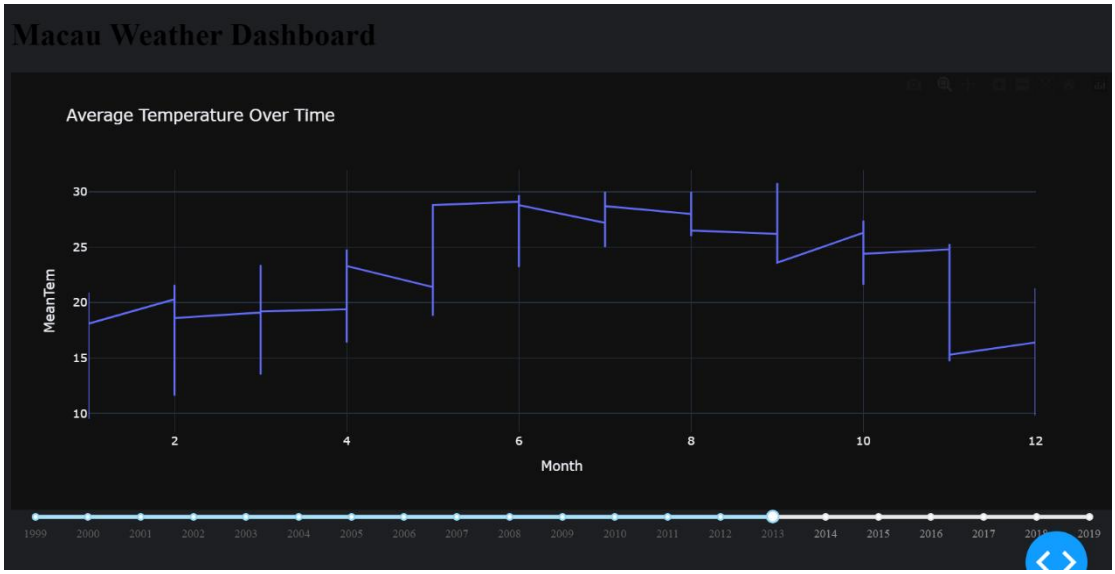

Figure 3 Temperature DashBroad-2013

As you can see from the two dashboards, the temperature difference in Macau is relatively small in the summer (June-September), with a minimum of only 4 °C, and relatively large in the winter (11-2), with a maximum difference of 11.5 °C.

After analyzing it again using the violin chart (figure 4), it is found that the occurrence of cold temperatures in Macau is more concentrated in January-April, while the temperatures in the rest of the months are relatively uniform

.

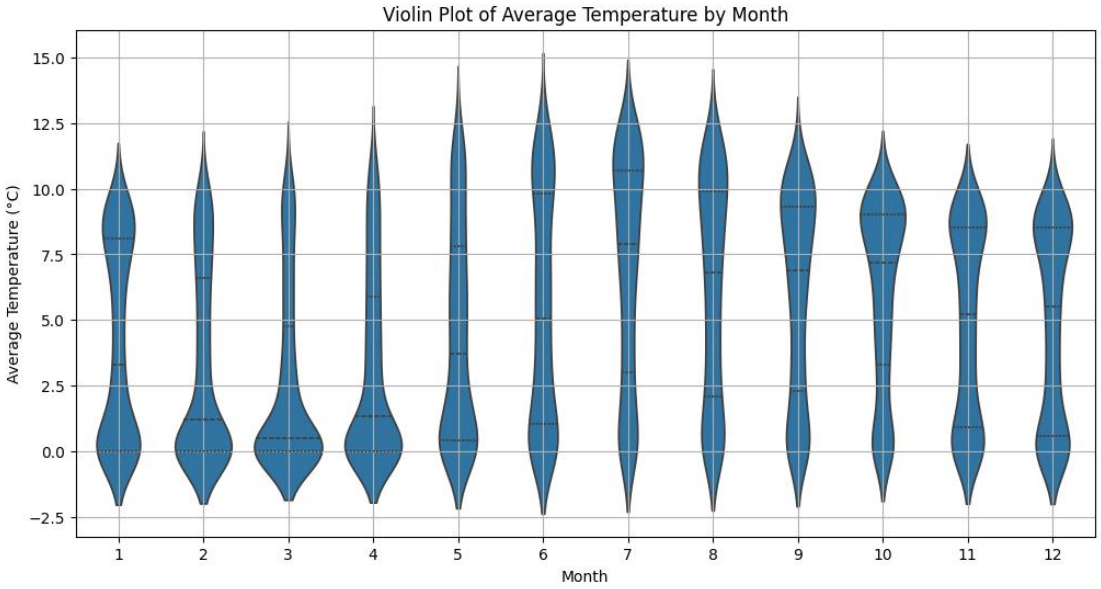

Figure 4 violin chart Temperature

In order to analyze the general trend in temperature over the two decades, a line graph was made using the highest average maximum temperature and the lowest average minimum temperature over the two decades, and a linear regression was used to fit the trend in temperature.

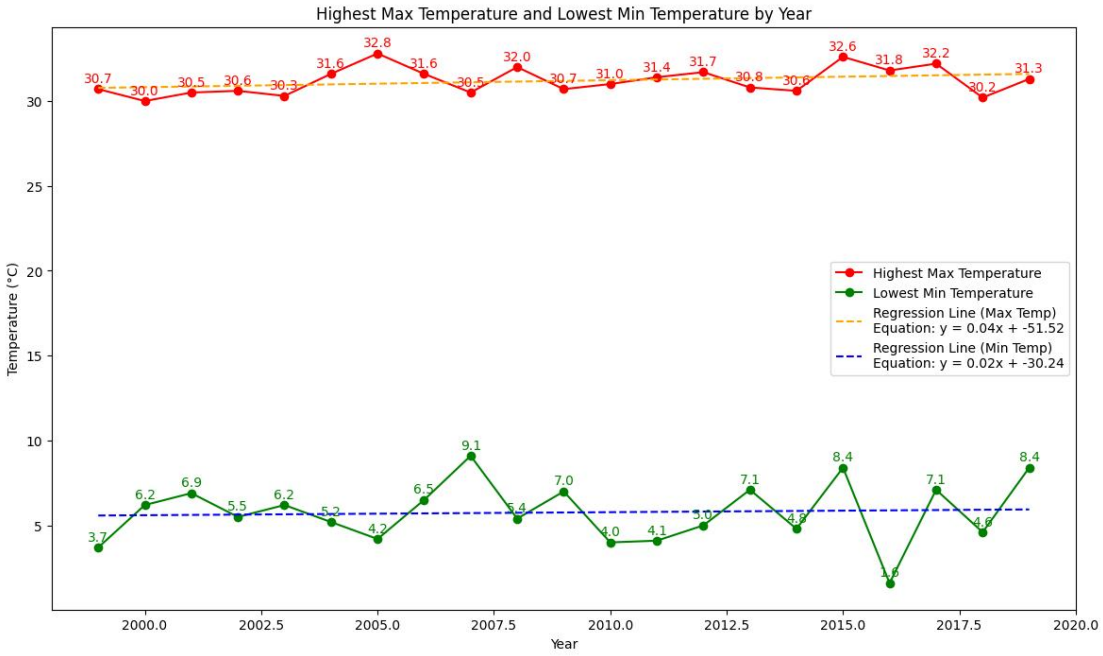

Figure 5 Temperature liner regression

The orange line and blue line is the liner regression expression. It can be noticed that these two figures are almost increasing year by year, and the slope of the regression line is also greater than 0. I speculate that it may be due to global warming that the Macao region is also affected.

3.2 Relative humidity

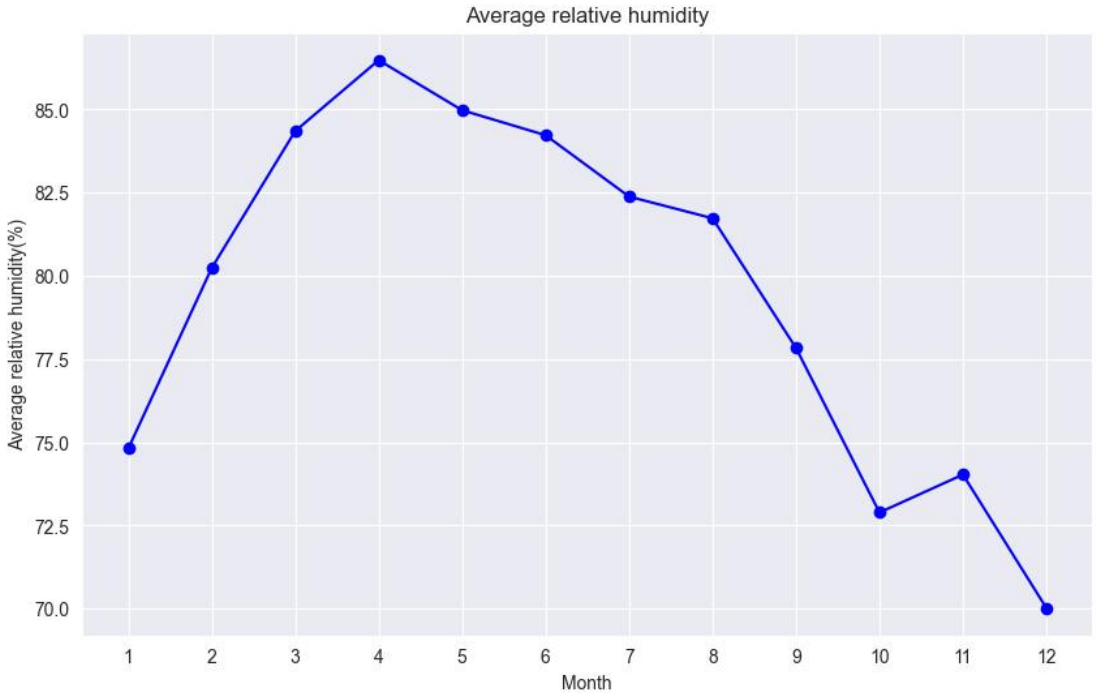

Figure 6 relative humidity monthly

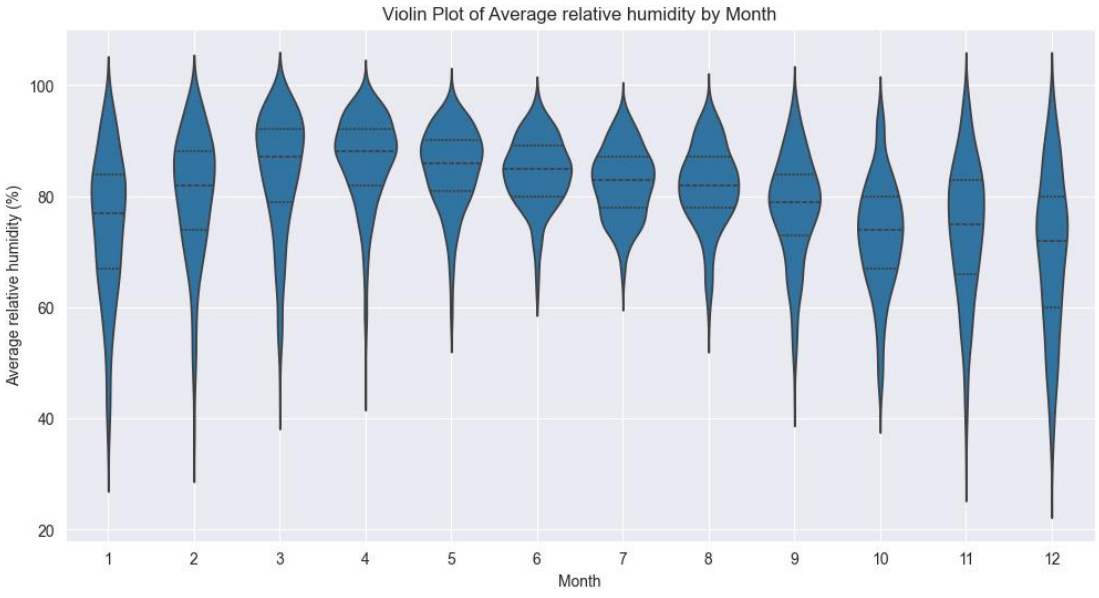

Figure 7 relative humidity Violin

As can be seen from Figure 6, the air humidity in Macao is at a very high level all year round, above 60% throughout the year, and even above 80% from April to August. Figure 7 shows that the distribution of humidity is mostly concentrated, with Macau having

exceptionally high humidity almost most of the time. Studies (*What Does Humidity Percentage Mean and What Is High Humidity? | Weather Radio Review*, n.d.) have shown that relative humidity levels above 60% are usually considered high. Such humidity levels can have negative impacts on comfort and health, such as increasing mold growth, causing breathing problems, and triggering allergic reaction.

3.3 Insolation duration

A Combination graph (figure 8) was created to show the degree of sunlight per day, with colors ranging from cool to warm to indicate that Insolation Duration goes from low to high each day. In the graph we can see that Insolation Duration varies with the seasons. Summer has the most hours of sunlight and winter has the least hours of sunlight.

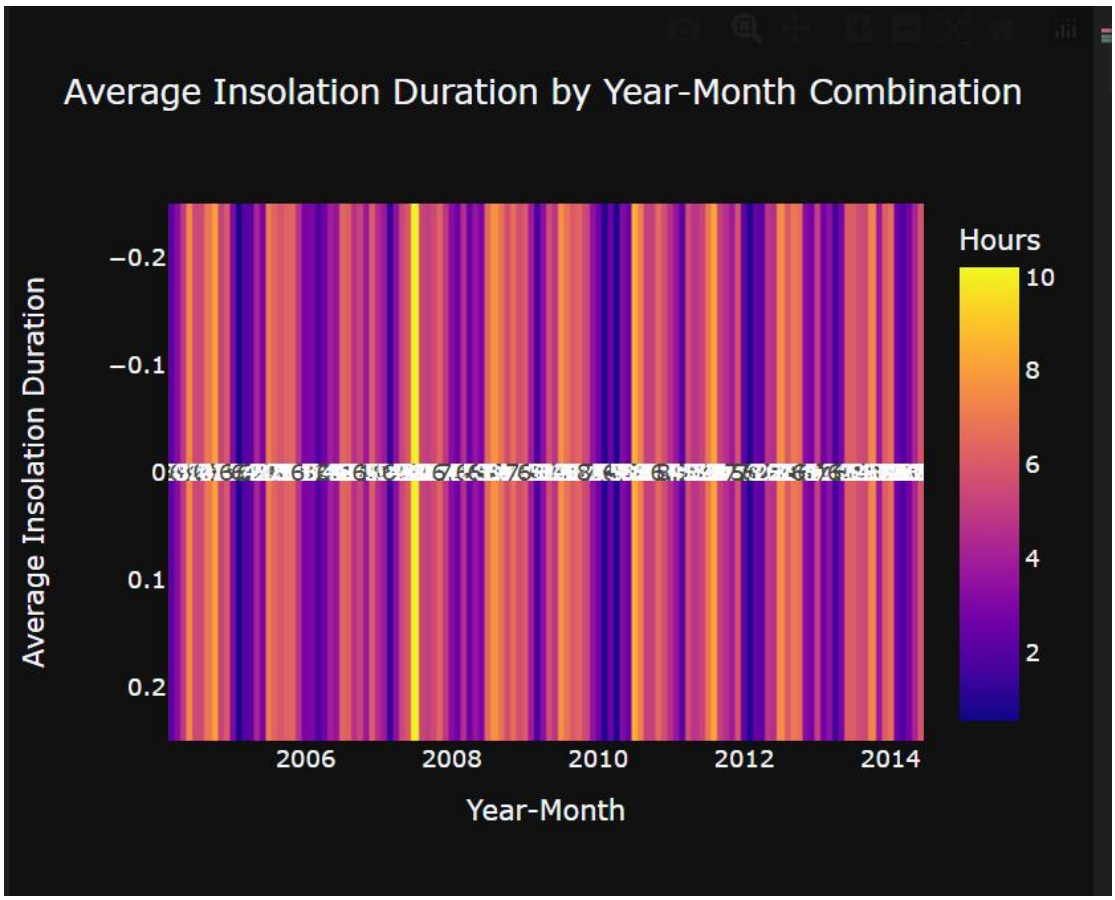

Figure 8 Insolation Duration Combination

3.4 Total rainfall

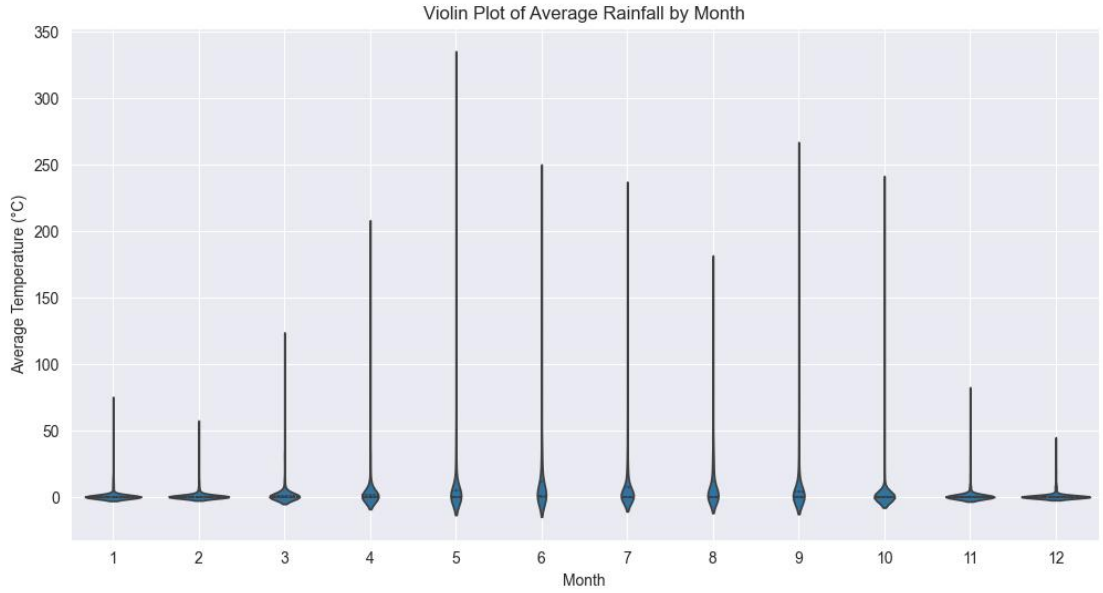

Figure 9 violin rainfall

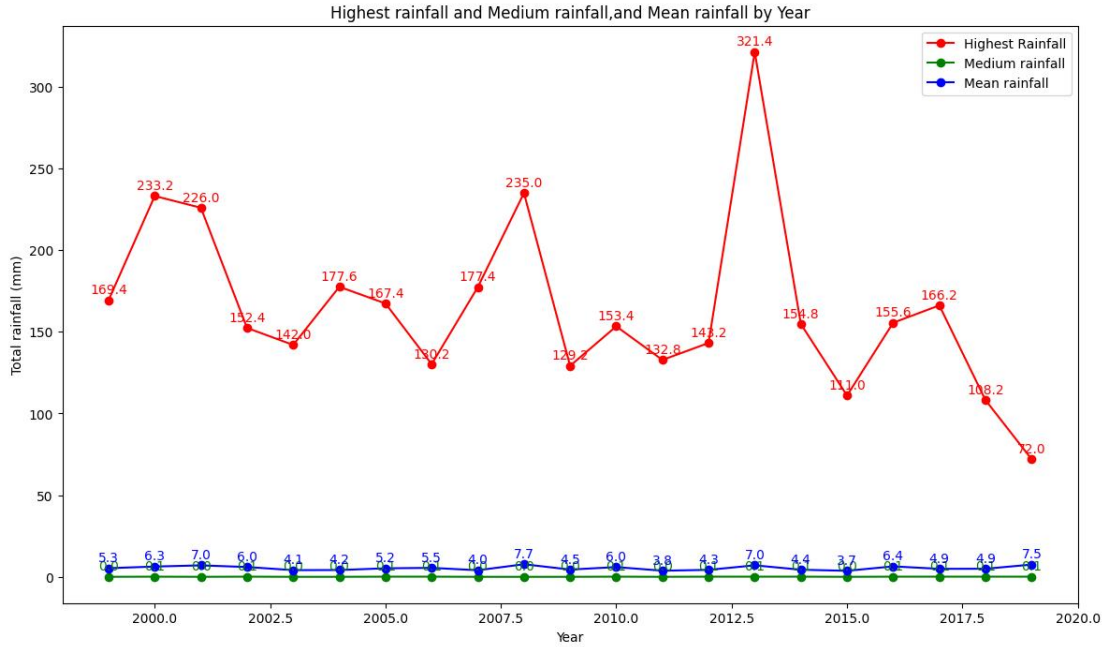

Figure 10 rainfall line graph

Calculated the maximum, minimum and median rainfall, and found that most of the rainfall in Macau is not significant, but occasionally the rainfall is particularly heavy. According code calculation, there are 3779 days with non-zero rainfall, more than 50%, which shows that Macau is a city with high rainfall frequency.(*Climate Change Indicators: Heavy Precipitation | US EPA*, n.d.)

4. Prediction

To predict the temperature, an LSTM model was trained to make predictions. Using rolling forecasts, one month at a time, the temperature for August 2019 is predicted and the results are as follows:

```
[[32.36081   29.107018  26.70735   86.360954   5.7498436  8.289654 ]
 [32.560654  29.176037  26.676277  85.25519    6.259141   8.648929 ]
 [32.63699   29.236591  26.68668   84.837326   6.1957655  8.45953  ]
 [32.67219   29.287025  26.72134   84.914764   5.859714   9.503887 ]
 [32.70191   29.323605  26.73905   85.011215   5.9093246  9.580781 ]
 [32.664436  29.320366  26.752436  85.05941    5.893023   9.606389 ]
 [32.705704  29.360453  26.783813  84.92958    5.931772   9.692518 ]
 [32.757835  29.419298  26.854221  85.011635   5.997264   9.787157 ]
 [32.801296  29.443758  26.882452  85.10125    6.0263343  9.66254  ]
 [32.850235  29.483541  26.923792  85.10345    6.08393    9.5912695]
 [32.90795   29.490011  26.932707  84.89774    6.159218   9.288245 ]
 [32.89516   29.435104  26.873287  84.51697    6.1503825  9.001454 ]
 [32.813183  29.372997  26.83249   84.70718    5.9200125  9.3733425]
```

Figure 11 part of prediction

Let us view the real data

7507	2019/7/20	31.3	28.2	26.2	89	2.9	0.2
7508	2019/7/21	31.4	28.6	26.5	87	3.8	VST
7509	2019/7/22	30.3	27.9	25.2	89	0.9	11.6
7510	2019/7/23	31.2	28	24.6	87	8.5	6.2
7511	2019/7/24	31.4	28.8	25.7	86	10.6	6.2
7512	2019/7/25	31.9	29	27.2	85	9.8	1.2
7513	2019/7/26	31.7	29.5	27.7	82	11.6	0
7514	2019/7/27	32.3	29.5	27.8	82	9.3	0
7515	2019/7/28	31.6	28.2	24.2	87	2.2	22
7516	2019/7/29	33	28.7	27	84	8.1	VST
7517	2019/7/30	30.8	28.2	24.8	86	7.1	7.4
7518	2019/7/31	27.7	25.8	24.4	94	0.3	52.2
7519	2019/8/1	26.9	25.8	24.4	97	0	58
7520	2019/8/2	28.6	26	23.4	95	1	72
7521	2019/8/3	30	26.4	24.7	91	0.5	1.4
7522	2019/8/4	30.3	26.9	24.8	90	1.1	2
7523	2019/8/5	33.8	28.9	25.2	80	12.3	0
7524	2019/8/6	32.8	29.1	27.1	82	2.6	0
7525	2019/8/7	32.7	28.9	26	81	10.4	0
7526	2019/8/8	34.3	29.6	24.4	83	10.7	16.8
7527	2019/8/9	35.7	31.2	25.2	76	10.9	VST
7528	2019/8/10	34.3	29.9	26.8	87	6.9	3.4
7529	2019/8/11	33.2	29.8	28.5	86	9.2	0
7530	2019/8/12	33.6	29.3	25.2	87	6.6	15.8
7531	2019/8/13	34.6	30	28.3	83	8.8	0
7532	2019/8/14	34.5	29.7	25.9	81	8.4	7

Figure 12 ture data of last month

Temperature's, humidity, and sunshine predictions are all very similar to the real values, but the rainfall predictions are not so good.

The reason may have two. Firstly, from figure10 shows that, there so many data is zero, what's more the highest data is over 300 but most of data (mean and medium) even less than 1, it's hard to normalize data. Secondly, the correlation matric (figure 13) shows that, the correlation between rainfall and all other data is low.

The solution of it is add more dimension data like add Add wind speed and direction data to characterize the temperature from other weather dataset support by Macau government (*Macao SAR Government Data Open Platform*, n.d.).

Climate Change Indicators: Heavy Precipitation | US EPA. (n.d.). Retrieved September 7, 2024, from https://www.epa.gov/climate-indicators/climate-change-indicators-heavy-precipitation

SMG - Geophysical and Meteorological Office of the Macao SAR Government. (n.d.). Retrieved September 7, 2024, from https://www.smg.gov.mo/en/subpage/345/embed-path/p/query-weather-e_panel/800

What Does Humidity Percentage Mean and What Is High Humidity? | Weather Radio Review. (n.d.). Retrieved September 7, 2024, from https://weatherradioreview.com/what-does-humidity-percentage-mean-and-what-is-high-humidity/

Macao SAR Government Data Open Platform. (n.d.). Retrieved September 7, 2024, from https://data.gov.mo/Detail?id=712bf9fa-2ec1-4762-aa55-2cdaf23dd563