1.1. In this question, I read the earthquake data first. Then, I used function groupby and sum() to get the total number of deaths caused by earthquake since 2150 B.C. in each country. I defined 'T_deaths' as a dataframe so that I could do the following work, and then I found that groupby returned series except dataframe, so I used pd.DataFrame to convert T_deaths to dataframe.

To print the top ten countries along with the total number of deaths, I used sort_values, and since we wanted to know the top ten, I used 'ascending = False' to get the descending order. Head() helped me to print the top ten.

1.2. To get the total number of earthquakes with magnitude larger than 6.0 worldwide each year, I filtered the data using ['Mag'] >= 6 first, and then I used groupby to count the times of earthquakes each year. Finally, I plotted it.

As is shown in Figure 1, I found that the total number of earthquakes increases as time goes by. In my opinion, it would be caused by missing records in the past, and at present, we have sufficient labor and equipment to report all of the earthquakes. Therefore, the total number of earthquakes goes up as time series.

1.3. In this question, I defined a function with two independent variables. First, I extracted data of one country using loc. After getting the number of earthquakes in this country, I writed one line to fine the date with the largest earthquake, and then the function would return the number and the date.

After that, I used a for cycle to run the function with all of countries, and added the result to my dataframe. Then I found the dataframe contained a lot of duplicate lines, and I got inspired by reading https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.drop_duplicates.html. By this way, I deleted the duplicates, and reported the results in a descending order to excel file. It would be shown in chart 1.

2. After reading the user guide, I read the file and extracted two columns [['DATE', 'WND']] to the dataframe. With the help in https://blog.csdn.net/Asher117/article/details/84346073, I separated the parameters of wind to some columns using map() and lambda. I got months and years data by the same method, with assuming the date as 1 (we don't need to get the date).

Then I assumed the values of speed rate greater than 900 were invalid data as the user guide showed. At this time, I had a problem that the type of the ['SR'] column was object, which cannot calculate as numerical type. I got help in website: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.to_numeric.html.

The raw data was in chronological order, so I didn't need to sort the values. With groupby, I plotted the time series of speed rate as Figure 2. I found that the wind speed goes up in general as time goes by.

3.1. In this part, I downloaded some data on CASEarth, about rice exportation and importation in China. After reading the csv file, I used dropna() to clean possible data points with missing values or bad quality with reading https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.dropna.html.

3.2. In the data, it would be found that it contains many duplicate years. Thus I used groupby

to plot the time series of annual rice exportation in China as Figure 3.

3.3. In this question, I did 5 simple statistical checks including the ten countries with the largest export volumes (Côte d'Ivoire, Indonesia, Republic of Korea, Japan, Philippines, Cuba, Russian Federation, Democratic People's Republic of Korea, China, Hong Kong SAR, Liberia), the ten countries with the largest import volumes (Thailand, Viet Nam, Pakistan, Cambodia, Myanmar, China, Taiwan Province of, Democratic People's Republic of Korea, Lao People's Democratic Republic, China, mainland, India), the countries count which exist in both list (the result is 1), annual net exportation in China (net exportation equals to exportation minus importation), and plot annual exportation, importation, net exportation. It shows as Figure 4. I have found that most of the top ten countries have a diplomatic friendship with China. However, I can’t conclude a general character of the time series, it seems not too characteristic. The only I could say is that the importation of rice went up at high speed for some reasons.

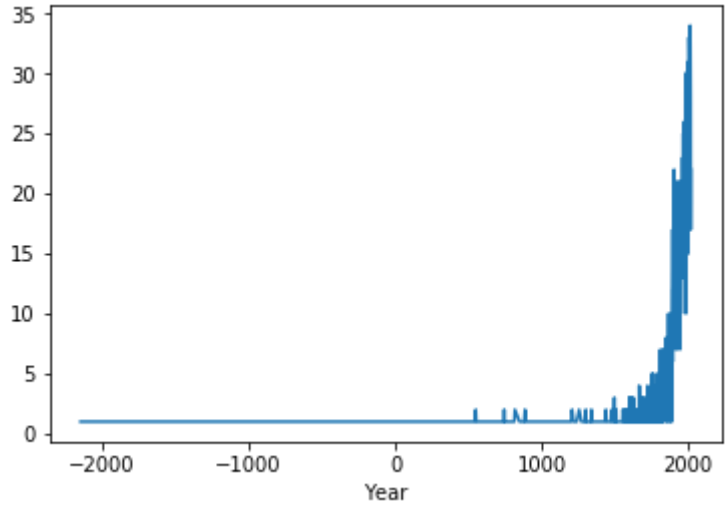

Figure 1. Counts of Earthquakes (magnitude > 6.0) in each year (since 2150 B.C.)

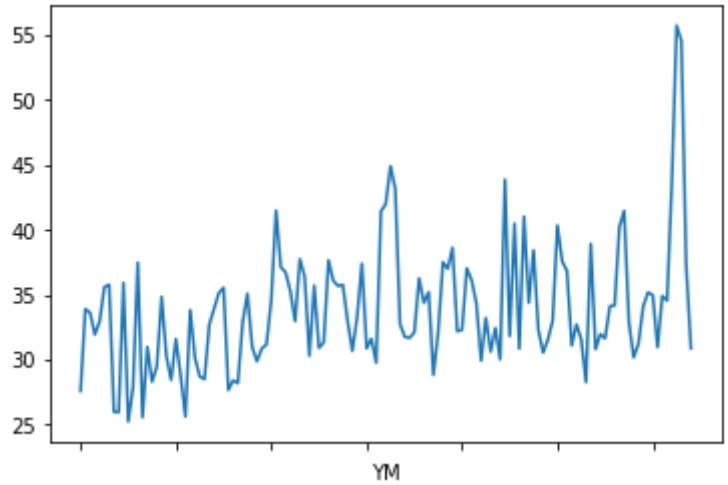

Figure 2. Monthly average wind speed in Shenzhen (2010 - 2020)

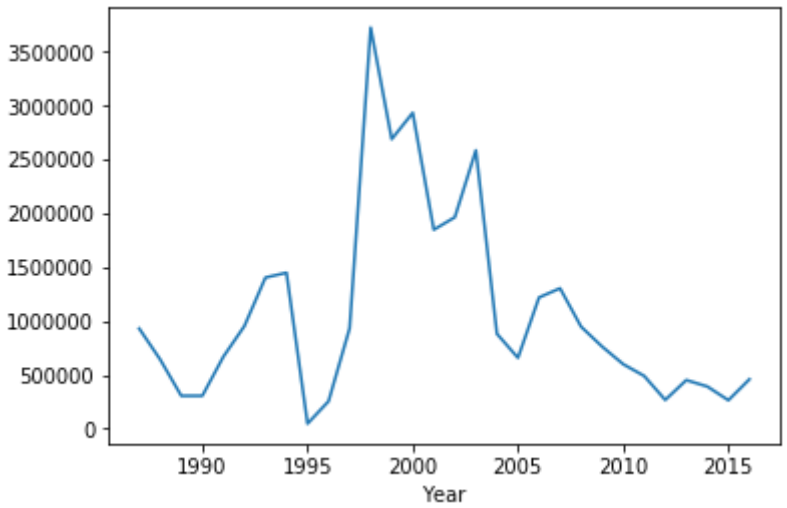

Figure 3. Rice Exportation in each year in China

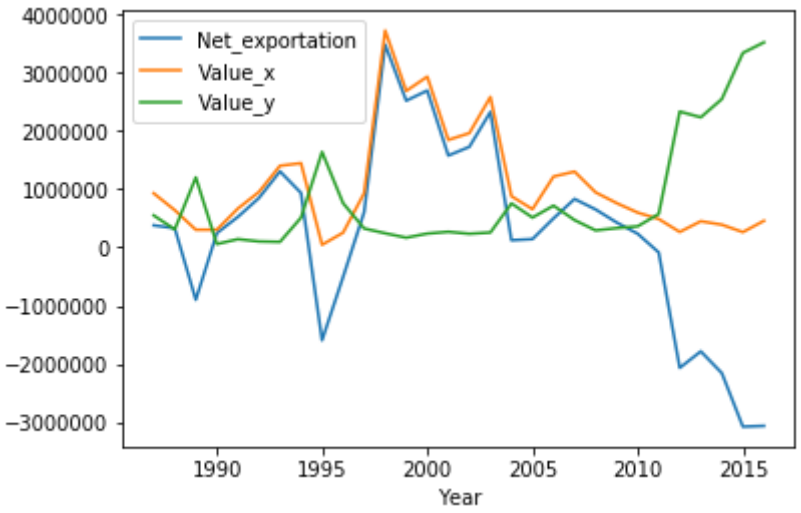

Figure 4. Exportation (Value_x), Importation (Value_y), Net Exportation in each year in China