PS2.ipynb

All code is annotated in the .ipynb file

1. Significant earthquakes since 2150 B.C.

1.1

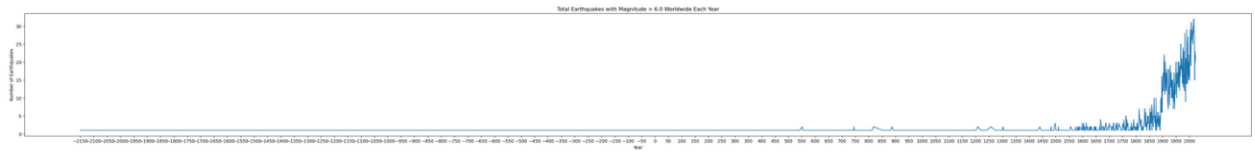
Read the data first, separated by '\t', the total number of deaths is then calculated according to country groups. Finally, sort and display the top 10 data

```
The top10 countries along with the total number of deaths:
Country
CHINA         2075045.0
TURKEY        1188881.0
IRAN          1011449.0
ITALY          498478.0
SYRIA          439224.0
HAITI          323478.0
AZERBAIJAN     317219.0
JAPAN          279085.0
ARMENIA        191890.0
PAKISTAN       145083.0
Name: Deaths, dtype: float64
```

1.2

Firstly, the data of earthquake level greater than 6 is extracted, and then the number of earthquakes is obtained by grouping according to the year and counting the number of earthquake magnitude data.

Trend: As can be seen from the graph, the total number of earthquakes shows an increasing trend with each year. I think the main reason is the progress of modern earthquake monitoring technology. Secondly, I suspect that the crust movement has become more intense in recent years.



1.3

First assign the null value in the table to 9999, otherwise an error will be reported. Then I designed a function, CountEq_LargestEq, in which I counted the number of earthquakes recorded in each country to get the total number of earthquakes. They then took data on the year, month and day of the largest earthquakes in each country and put them together to get the final date. Extract all country names and discard the first line, because the first line is nan(learn unique function from CSDN). Generate a DataFrame

to store all the information required by the problem. Finally, replace the original '9999' with nan to get the final table.

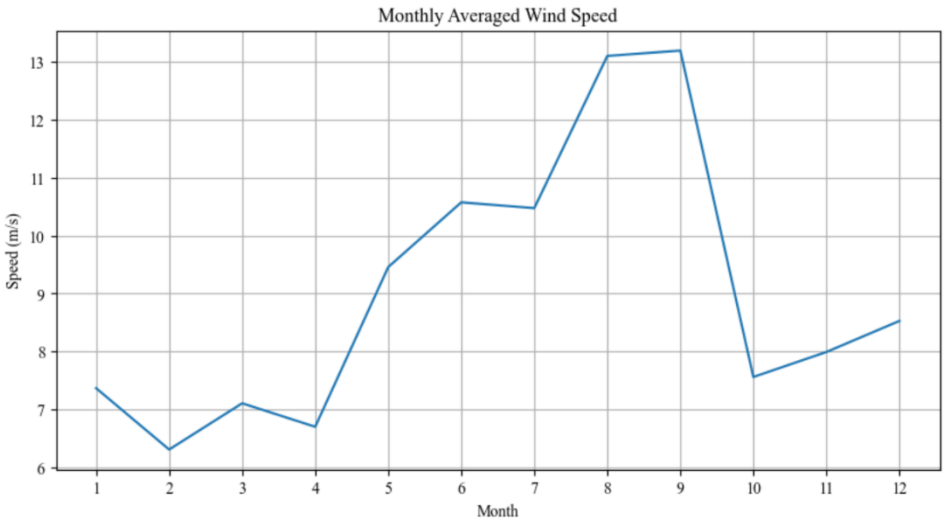
Out[3]:

	Country	total number of eqs	Date of the largest eqs
14	CHINA	620	-47-4-17
33	JAPAN	414	1545-2-7
68	INDONESIA	411	1629-8-1
7	IRAN	384	662-4-26
9	TURKEY	335	-282-nan-nan
...	...	...	...
93	NORWAY	1	1819-8-31
126	CENTRAL AFRICAN REPUBLIC	1	1921-9-16
124	PALAU	1	1914-10-23
118	KIRIBATI	1	1905-6-30
155	COMOROS	1	2018-5-15

156 rows × 3 columns

2. Wind speed in Shenzhen during the past 10 years

According to the user guide, we can know the wind speed information in the fourth separated data bit of each row in the WND column, that is, the data like '0200'. To process all the data quickly and avoid using for loops, I put the WND data into array form. Converts all of the elements in the array to *str* form, separating all of them with the separator comma. Then the wind speed data can be read quickly by list derivation(learn list derivation from CSDN). The to_datetime function is used to group DataFrame by month and obtain the average wind speed to get the final result.



Trend: The general trend is: the wind speed gradually increases from February to September, with September being the largest, then the wind speed rapidly decreases in October, and finally the wind speed fluctuates within a certain range from November to February.

3. Explore a data set

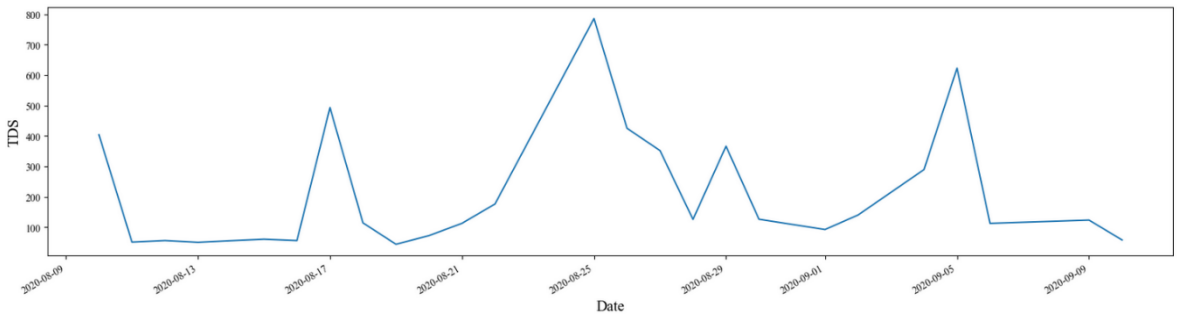
3.1

File name is: 2020 夏季雅江全流域.xlsx

Clear data points without water temperature and type

3.2

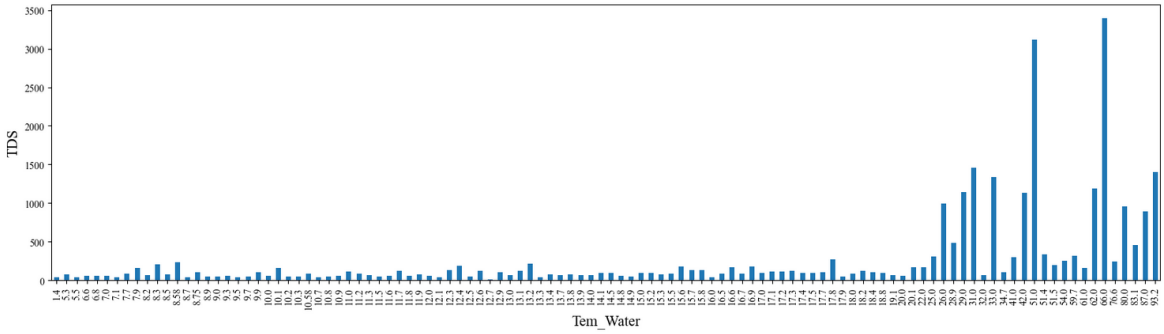
Draw line charts for dates and TDS using the groupby function



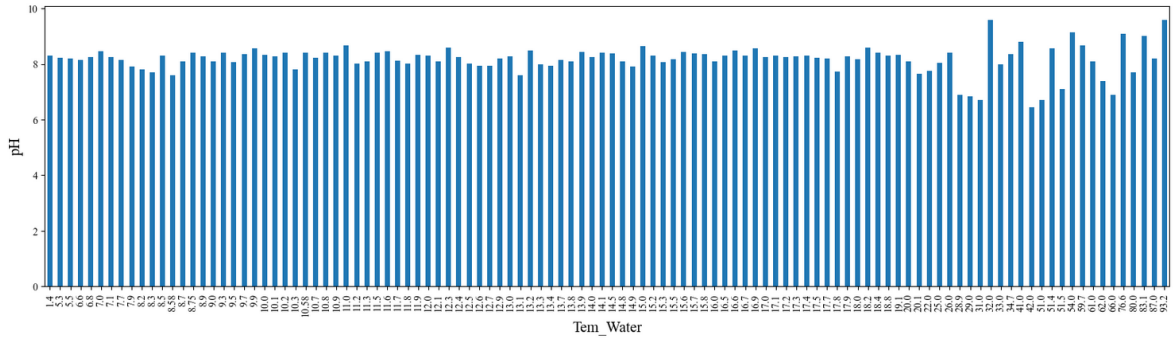
3.3

Because not all data grouped by TDS can find a relationship, here are five sets of statistics that can find some relationship.

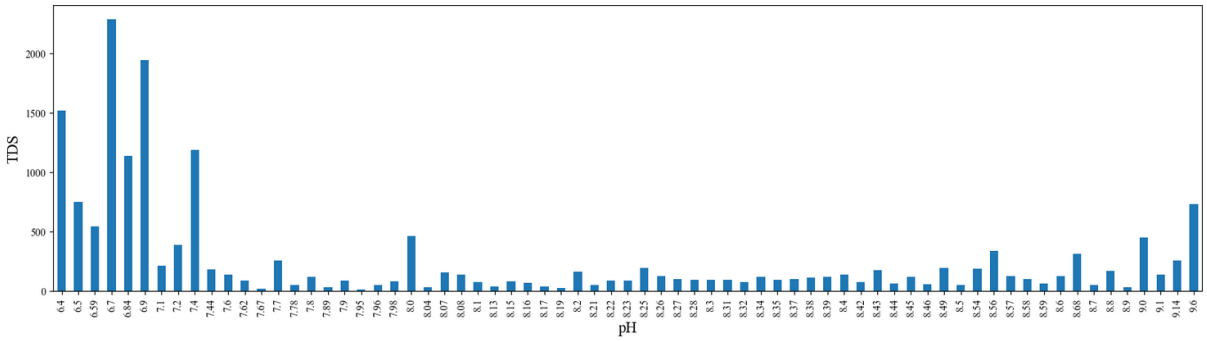
3.3.1 Tem_water and TDS: The general trend is that TDS increases with water temperature



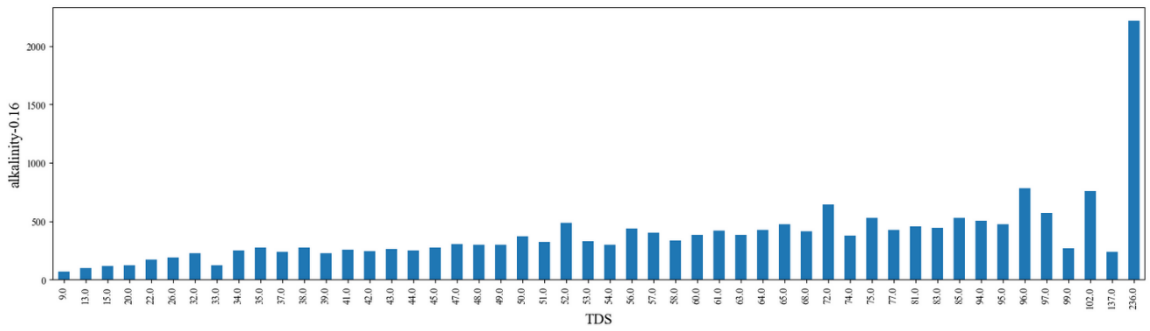
3.3.2 Tem_water and pH: No significant relationship was found



3.3.3 pH and TDS: TDS decreases as pH increases



3.3.4 TDS and alkalinity-0.16 : The general trend of alkalinity-0.16 is increasing with TDS



3.3.5 PSU and alkalinity-0.16: The general trend of alkalinity-0.16 is increasing with PSU