1. Significant earthquakes since 2150 B.C.

导入了 pandas、numpy 和 matplotlib 库，设置画图的字体为 Times New Romans，使用 pd.read_csv 导入了 earthquakes-2024-11-01_10-28-41_+0800.tsv，文件。

1.1

根据 Country 列使用 groupby 函数进行分组，再对每个 group 进行求和，接着根据 Deaths 列进行降序排序，最后单独取出 Deaths 列的数据。输出前 20 的国家及死亡总人数。

1.2

根据判断形式提取出 Ms>3.0 的数据，即震级大于 3.0 的地震数据，再根据 Year 列进行分组，并使用 count()函数求出每年的、震级大于 3.0 的地震数量。

使用 plt.plot()函数绘图，可以看到，从公元前 2000 开始，到 500 年前，震级大于 3.0 的地震数量较少，而近 500 年，震级大于 3.0 的地震数量迅速增长，可能有两个原因：一是由于以前的人数较少、文献记载缺失、缺乏计量的技术手段等等，导致地震记载的数量较少，而现在在科技发展的背景，地震数量可以很好地被记录；二是地球板块运动可能逐渐变得越来越活跃，板块相互碰撞和挤压导致地震频繁发生。

1.3

定义了一个 CountEq_LargestEq(country_name)函数，输入参数为国家的名字，接着在函数内定义一个 get_largest(x)函数，传入参数为 dataframe 类型，接着对 x 按照 Ms 进行降序排序，并取出排序后的 dataframe 的第一行，即取出 x 中震级最大的一行数据。

(1) 根据 Country 分组，并使用 count 函数()对地震次数计数，再读取 Id 列和 country_name 行，完成第一个要求；

(2) 根据 Country 分组，应用刚定义的 get_largest(x)函数，取出每个国家 dataframe 数据块的震级最大数据，接着定义一个字符串，使用格式化的方法输出，内容包括日期，位置名字和经纬度，完成第二个要求。

返回[<国家发生地震总数>, <发生地震的日期和地点>]列表，于是得到题目所要求的函数，接着我们使用列表生成式对每个国家应用 CountEq_LargestEq 函数，并将发生地震总数按照降序排序输出。

2. Air temperature in Shenzhen during the past 25 years

读取 Baoan_Weather_1998_2022.csv 数据，接着读取原始数据的日期和温度数据，使用 str()函数取出了数据的月数据（当时还不知道第三题的指定时间列的方式），对 TMP 列，按照指导手册中描述的放缩因子和数据格式，修改并添加

Tmp 列。

接着剔除掉温度大于 61.8℃和小于-93.2℃的数据。提取出 Month 和 Tmp 两列的数据，对 Month 进行分组，并把 Month 列转化为时间格式，然后以 Month 为横坐标，以 Tmp 为纵坐标绘制折线图。

由图可知，逐月气温随时间呈现周期性的波动，所以过去 25 年间，深圳的月平均气温没有发生明显变化。

3. Global collection of hurricanes

示例代码的每个参数解释如脚本所示。

3.1

先将 WMO_WIND 列转化为数值类型，接着使用类似与 1.3 的 get_largest(x)函数得出同 SID 的飓风中 WMO_WIND 最大的值，再对每个飓风根据 WMO_WIND 进行降序排序，输出前 10 行即为所求。

3.2

使用 3.1 的结果绘制前 20 的飓风风速条形图。

3.3

根据 BASIN 分组后，使用 count()函数得出每个区域的数据点数，接着绘制条形图。

3.4

提取出 LAT 列和 LON 列，分别为纬度和经度，接着绘制点位置的六边形图。

3.5

根据 NAME 分组，选取出名为 MANGKHUT 的 dataframe，接着提取出其经纬度，绘制出其轨迹的散点图。

3.6

由后面的题目可知，只需要提取出前 7 列即可。使用不等式提取出 1970 以来的数据，接着对 BASIN 进行分组，然后提取出 WP 和 EP 列的 dataframe，接着将两个 dataframe 拼接，即为所求。

3.7

对 3.6 中的数据，添加一个新列 DAY，其值为 ISO_TIME 的年月日部分，接着对 DAY 分组，并使用 count()函数进行计数。以 DAY 为横坐标，SID 为纵坐标（值等于一天的数据量数）绘制折线图。

3.8

对 3.7 中分组后的数据，添加一个新列 D，其值为 DAY 列的“日”部分，代表一年中的第几天，接着对 D 分组，并使用 mean()函数求出多年日平均值。

接着绘制多年日平均值与 D 的图像，即为所求。

3.9

使用 3.8 分组后的数据，接着对 D 分组，并使用 mean()函数求出多年日平均值，使用 merge 将平均值对应到原始数据中，将 SID_x 和 SID_y 两列作差，求出异常值，最后绘制异常值与 Day 的图像，即为所求。

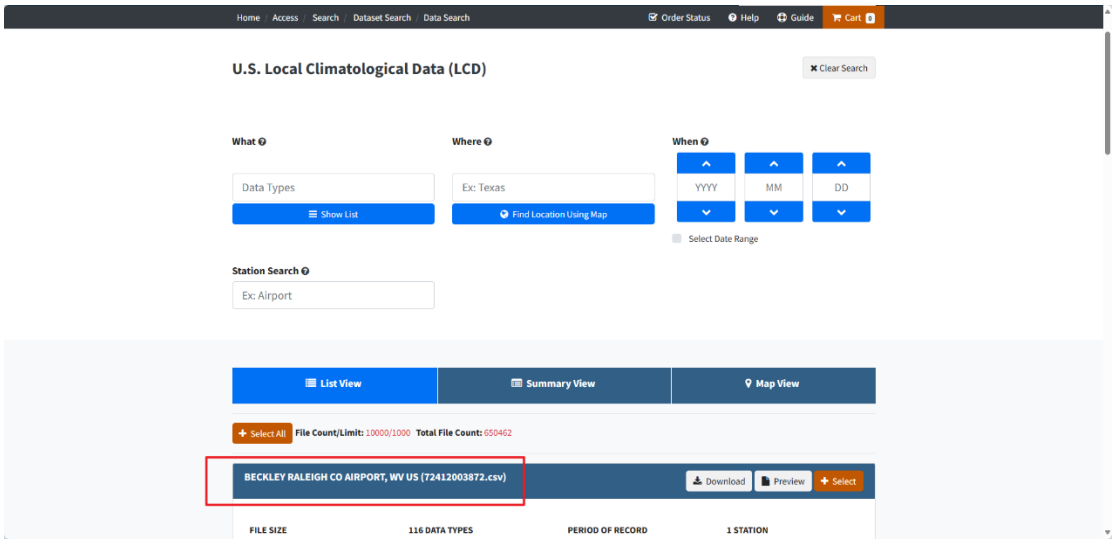
3.10

仿照前面小题，绘制年尺度的异常图。由图可知，1992 年和 1994 四年的飓风活动异常突出。

4. Explore a data set

4.1

我使用的数据是 NCEI 的美国 Beckley Raleligh co 机场气候数据，文件名为 72412003872.csv。



读取后使用 dropna()去除有异常值的行，使用 drop_duplicates()去除重复行。

4.2

选取 HourlyDewPointTemperature 列，绘制其时间序列图，如图所示。

4.3

取出 HourlyDewPointTemperature 和 HourlyDryBulbTemperature 变量，使用 describe()函数，即可得到两个变量的 8 个统计量。