

题目十三：1870-2018 年 Nino4 区海平面温度异常分析

【背景】

厄尔尼诺事件是指赤道中、东太平洋海表大范围持续异常偏暖的现象，其评判标准在国际上还有一定差别。一般将 NINO 3 区海温距平指数连续 6 个月达到 0.5℃以上定义为一次厄尔尼诺事件，美国则将 NINO 3.4 区海温距平的 3 个月滑动平均值达到 0.5℃以上定义为一次厄尔尼诺事件。

为更加充分地反映赤道中、东太平洋的整体状况，目前，中国气象局国家气候中心在业务上主要以 NINO 综合区（NINO 1 + 2 + 3 + 4 区）的海温距平指数作为判定厄尔尼诺事件的依据，指标如下：NINO 综合区海温距平指数持续 6 个月以上≥0.5℃（过程中间可有单个月份未达指标）为一次厄尔尼诺事件；若该区指数持续 5 个月≥0.5℃，且 5 个月的指数之和≥4.0℃，也定义为一次厄尔尼诺事件。

【数据说明】

数据集为 1870-2018 年 NINO4 区(5° N-5° S, 160° E-150° W)海平面温度异常滑动平均值，如下图所示。

Year	January	February	March	April	May	June	July	August	September	October	November	December
1870	-0.48	-1.16	-1.01	-1	-1.08	-1.11	-0.98	-0.86	-0.44	-0.51	-0.62	-0.67
1871	-0.55	-0.41	-0.44	-0.4	-0.7	-0.71	-0.87	-0.55	-0.45	-0.43	-0.52	-0.72
1872	-0.76	-0.51	-0.39	-0.68	-0.61	-0.79	-0.75	-1.13	-1.15	-0.9	-0.74	-0.83
1873	-0.77	-0.86	-1.09	-0.72	-0.54	-0.76	-0.71	-0.62	-0.62	-0.79	-0.7	-0.47
1874	-0.61	-0.68	-0.97	-0.64	-0.61	-0.66	-0.83	-0.88	-1.21	-1.27	-1.22	-0.95
1875	-0.65	-0.41	-0.68	-0.99	-1.1	-1.04	-1.18	-1.13	-1.19	-0.93	-0.66	-0.83
1876	-1.01	-0.95	-0.76	-0.87	-0.71	-0.59	-0.66	-0.47	-0.39	-0.27	-0.22	-0.11
1877	-0.1	0.47	0.25	0.14	0.22	0.47	0.53	0.55	0.6	0.8	0.91	1.16
1878	1.05	1.18	0.71	0.51	0.37	0.65	-0.03	-0.3	-0.5	-0.61	-0.48	-0.45
1879	-0.27	-0.09	-0.09	-0.32	-0.69	-1	-1.03	-0.86	-0.84	-1	-1.15	-0.84

- 字段说明：
- (1) Year 列，表示年份
 - (2) 列 January~December 分别表示对应月份 NINO4 区海平面温度异常滑动平均值。

【任务】

- 1、用 pandas 库读取 “nino4.long.anom.data” 文件，将所有时间抽取为单独的列 Date（形式为 YYYY-MM-01），所有异常平均值抽取为一个单独的列 Nino4，将所有缺失值丢弃处理，并导出到新的 txt 文件 “nino4_dropnan.txt”，第一行为表头，列名分别为 Date 和 Nino4，且表头和数据行中的不同字段信息都是用逗号分割，如下图所示。

```
Date,Nino4
1870-01-01,-0.48
1870-02-01,-1.16
1870-03-01,-1.01
1870-04-01,-1
1870-05-01,-1.08
1870-06-01,-1.11
1870-07-01,-0.98
1870-08-01,-0.86
1870-09-01,-0.44
1870-10-01,-0.51
1870-11-01,-0.62
1870-12-01,-0.67
```

- 2、重新读取新的数据集 “nino4_dropnan.txt”，选择 Nino4 字段，统计最大值 maxValue、最小值 minValue、平均值 meanValue。
- 3、重新读取文件 “nino4_dropnan.txt”，利用第三步统计结果最大值 maxValue、最小值 minValue，利用 category = [minValue, -0.5, 0, 0.5, maxValue] 和 labels = ['LaNinaTemp', 'Cold', 'Warm', 'NinoTemp'] 将 Nino4 进行离散化；并将离散化结果作为一个新的列 Label 添加到原始数据集，并保存为 “nino4_dropnan_result.csv”，从左到右三个列名分别为 Date、Nino4、Label；根据离散化结果画出饼状图，保存为 “nino4_pie.png”，要求分辨率不低于 300dpi。
- 4、重新读取文件 “nino4_dropnan_result.csv”，根据列 Label 判断，假设若连续出现 5 次 'LaNinaTemp' 则判定为出现了一次 LaNina 事件，选择 LaNina 事件出现的开始时间存储到列表 LaNinaList 中；假设若连续出现 5 次 NinoTemp 则认为出现了 Nino 事件，选择 Nino 事件出现的开始时间存储到列表 NinoList 中。最后，将列表 LaNinaList、NinoList 分别保存到文本文件 “LaNinaStartDate.txt”、“NinoStartDate.txt” 中。

参考案例：计算一个列表中连续相同的元素个数，并返回连续出现 4 次及以上的 1 所在的开始位置。

```
import itertools
mylist = [1,1,0,1,1,1,0,0,0,1,1,1,1,0,1,0,11,0]
num_times = [(k, len(list(v))) for k, v in itertools.groupby(mylist)]
print(num_times)
sumIndexList = []#存储符合要求的元素首次出现的位置
for i in range(len(num_times)):
    if num_times[i][0] == 1 and num_times[i][1] >= 4:
```

```
sumIndex = 0
while i >= 1:
    sumIndex += num_times[i-1][1]
    i = i - 1
    sumIndexList.append(sumIndex)
print("符合要求的元素首次出现的位置列表为: ",sumIndexList)结果:
```

输出结果:

[(1, 2), (0, 1), (1, 3), (0, 4), (1, 4), (0, 1), (1, 1), (0, 1), (11, 1), (0, 1)]

符合要求的元素首次出现的位置列表为: [10]

【要求】

- 1、根据以上数据处理任务，设计并编程实现“数据分析与可视化系统”，要求
 - ① 各个任务选择用菜单实现（菜单可用字符串输出模拟，或者 Tkinter 形式实现）。
 - ② 各个任务名称自己定义，须由独立的函数实现，且每个任务执行成功与否须给出必要的文字提示。
 - ③ 数据输入和结果输出的文件名须由人工输入，且输出结果都要以文件形式保存。
 - ④ 为保持程序的健壮性，各个任务执行过程中需要进行必要的判断（如文件是否存在、输入是否合法等）、程序异常控制等。
- 2、根据以上统计结果，书写不少于 300 字的结果分析。