

题目十五：1866-2018 年南方涛动指数 SOI 数据分析

【背景】

太平洋与印度洋间存在的一种大尺度的气压升降振荡。当太平洋上气压变高（低）时，印度洋上从非洲到澳大利亚气压变低（高），即两地气压的距平有反向的变化。这种“跷跷板”现象大约 3-7 年重现。南方涛动实际是反映了具有行星尺度的热带大气系统在季节时间尺度上的振荡，这种振荡正是对赤道太平洋冷海温、暖海温时期交替变化所造成的下垫面海温场的热力强迫的响应。

南方涛动指数 SOI 大于 0，即强指数时期，反映热带西太平洋和东印度洋的低气压与东南太平洋的高气压都比较强；东西方向上气压梯度大，南太平洋上有较强的东南信风和赤道东风；SOI 大于 0 与赤道东太平洋冷水事件及 La Nina 具有很强的相关性。

南方涛动指数 SOI 小于 0，即弱指数时期，反映热带西太平洋和东印度洋的低气压与东南太平洋的高气压都比较弱。东西方向上气压梯度弱，南太平洋东南信风减弱，赤道东风也减弱，甚至会出现西风。SOI 小于 0 与赤道东太平洋暖水事件及 El Nino 具有很强的相关性。

【数据说明】

数据集为 1866-2018 年南方涛动指数 SOI 数据，如下图所示。

Year	January	February	March	April	May	June	July	August	September	October	November	December
1866	-0.62	-0.12	-0.62	-0.65	0.04	-0.82	-0.34	0.36	-0.18	0.07	1.1	-0.16
1867	0.09	-0.01	-0.09	0.83	0.5	-0.48	0.44	0.34	0.12	-0.56	-0.65	-0.89
1868	-0.16	-0.34	-1.56	0.3	-1.34	-2.2	-0.4	-1.41	-1.23	-1.24	-1.49	0.52
1869	-1.9	-0.26	-0.59	2.12	1.4	1.53	1.42	0.94	0.12	0.85	0.56	0.42
1870	1.06	0.18	-0.75	0.47	-0.11	-1.1	0.17	0.98	0	-0.58	-0.86	-1.29
1871	-1.13	-0.13	-2	-0.46	-0.47	0.24	0.31	0.38	0.58	-0.14	-0.02	0.97
1872	2.69	1.12	1.69	-0.54	1.35	2.45	2.3	0.99	3.14	1.9	3.09	2.8
1873	2.47	1.12	1.63	2.29	1.19	-3.09	-1.14	-0.59	-2.13	-1.78	-0.36	1.33
1874	0.9	-0.28	-0.52	0.98	1.29	-1.92	-0.14	0.75	2.03	2.19	0.53	1.28
1875	0.65	-0.11	-1	1.4	0.75	2.57	-0.49	-0.21	0.83	0.93	-1.54	-0.54
1876	1.16	0.98	-0.1	0.94	0.75	1.98	-0.51	1.08	0.98	-1	-0.46	-0.45

字段说明：

- (1) Year 列，表示年份
- (2) 列 January~December 分别表示对应月份的 SOI 值。

【任务】

- 1、用 pandas 库读取“soi.long.data.csv”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 SOI，将所有缺失值丢弃处理，并导出到新的 txt 文件“soi_dropnan.txt”，第一行为表头，列名分别为 Date 和 SOI，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,SOI
1866-01-01,-0.62
1866-02-01,-0.12
1866-03-01,-0.62
1866-04-01,-0.65
1866-05-01,-0.04
1866-06-01,-0.82
1866-07-01,-0.34
1866-08-01,-0.36
1866-09-01,-0.18
1866-10-01,-0.07
1866-11-01,-1.1
1866-12-01,-0.16
```

- 2、重新读取新的数据集“soi_dropnan.txt”，选择 SOI 字段，统计最大值、最小值、平均值。
- 3、重新读取文件“soi_dropnan.txt”，利用第三步统计结果最大值 `maxValue`、最小值 `minValue`，利用 `category = [minValue, 0,maxValue]`和 `labels = ['NinoRelate', 'LaNinaRelate']`将 SOI 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“soi_dropnan_result.csv”，从左到右三个列名分别为 Date、SOI、Label；根据离散化结果画出饼状图，保存为“soi_pie.png”，要求分辨率不低于 300dpi。
- 4、重新读取文件“soi_dropnan_result.csv”，利用 `matplotlib` 库，可视化显示 SOI 值，要求包括图例、图标题，x 轴刻度以年显示且间隔为 10，y 轴显示刻度值，曲线颜色为蓝色。

【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。