

《数据预处理与分析》练习04

练习题：平台年度下单次数趋势面积图分析

【实操任务描述】

电商平台希望分析不同年份的整体下单规模变化趋势，以评估平台交易活跃度和业务发展情况。

请你基于用户行为日志数据 `browsing_logs.csv`，完成年度下单次数的统计，并绘制趋势面积图进行可视化展示。

【任务要求】

步骤 1：导入库并设置中文显示

使用 `pandas` 进行数据处理，使用 `matplotlib` 进行图表绘制，并完成中文字体与负号显示配置。

```
import pandas as pd
import matplotlib.pyplot as plt

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
```

步骤 2：读取并预览数据

读取 `browsing_logs.csv` 文件，并将行为时间字段解析为日期类型；

```
logs = pd.read_csv("browsing_logs.csv", parse_dates=["browse_time"])
```

步骤 3：筛选有效“下单”行为记录

从行为日志中筛选出 `action_type` 为“下单”的记录，作为后续统计的基础数据。

```
order_logs = logs[logs["action_type"] == "下单"]
```

步骤 4：提取并生成下单年份字段

从下单时间 `browse_time` 中提取年份信息，并生成新的字段 `year`，用于年度统计分析。

```
# 关闭 Pandas 的链式赋值警告
pd.options.mode.chained_assignment = None
# 提取并生成下单年份字段
order_logs["year"] = order_logs["browse_time"].dt.year
```

步骤 5: 按年份统计下单次数

以年份 `year` 为分组依据, 统计每一年的下单次数, 并将统计结果整理为 `DataFrame`, 统计字段命名为 `order_count`。

```
year_order_count = (  
    order_logs.groupby("year")  
        .size()  
        .reset_index(name="order_count")  
)
```

步骤 6: 排序并检查统计结果

按年份升序对统计结果进行排序。

```
year_order_count = year_order_count.sort_values("year")  
print(year_order_count)
```

步骤 7: 绘制年度下单趋势面积图

绘制“平台年度下单次数变化趋势面积图”, 具体要求如下:

- 使用 `plt.figure(figsize=(10, 6))` 设置画布大小;
- 使用 `fill_between()` 绘制面积图, 体现下单规模变化;
- 使用 `plot()` 叠加折线, 突出趋势走向;
- 设置图表标题为“平台年度下单次数变化趋势 (面积图)”;
- X 轴标签为“年份”, Y 轴标签为“下单次数 (笔)”;
- 设置 X 轴刻度为具体年份;
- 显示图例与网格线;
- 使用 `tight_layout()` 优化布局并显示图表。

```
x = year_order_count["year"]  
y = year_order_count["order_count"]  
  
plt.figure(figsize=(10, 6))  
  
# 面积图  
plt.fill_between(  
    x, y,  
    alpha=0.6,  
    label="年度下单次数"  
)  
  
# 折线叠加  
plt.plot(x, y, marker="o", linewidth=2)
```

```
plt.title("平台年度下单次数变化趋势（面积图）", fontsize=14)
plt.xlabel("年份")
plt.ylabel("下单次数（笔）")

plt.xticks(x)
plt.ticklabel_format(style="plain", axis="y")

plt.legend()
plt.grid(axis="y", linestyle="--", alpha=0.5)

plt.tight_layout()
plt.show()
```