

# 《数据预处理与分析》练习03

## 练习题：用户注册规模年度变化折线图分析

### 【实操任务描述】

平台希望分析不同年份用户注册规模的变化趋势，以评估平台的发展情况和用户增长态势。请你结合用户信息表 `users.csv` 与订单表 `orders.csv`，完成用户注册年份统计与趋势可视化分析。

### 【任务要求】

#### 步骤 1：导入库并设置中文显示

使用 `pandas` 完成数据处理，使用 `matplotlib` 绘制折线图，并正确设置中文字体与负号显示。

```
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 `users.csv` (将 `register_date` 解析为日期类型) 和 `orders.csv`，并分别查看前 5 行数据。

```
users = pd.read_csv("users.csv", parse_dates=["register_date"])
orders = pd.read_csv("orders.csv")

users.head(5)
orders.head(5)
```

#### 步骤 3：表联合获取有效用户数据

以 `user_id` 为关联键，对用户表与订单表进行内连接 (`inner join`)，仅保留存在下单记录的用户数据，生成新的 DataFrame。

##### 3.1 表联合

```
df = users.merge(
    orders[["user_id", "order_id"]],
    on="user_id",
    how="inner"
)
```

## 步骤 4：提取用户注册年份

从用户注册时间 `register_date` 中提取注册年份，并生成新的字段 `year`。

```
df["year"] = df["register_date"].dt.year
```

## 步骤 5：按年份统计注册用户人数

以注册年份 `year` 分组，统计每一年中注册并产生订单的用户人数，要求统计去重后的用户数，并按年份升序排列。

```
year_counts = (  
    df.groupby("year")["user_id"]  
        .nunique()  
        .sort_index()  
)
```

## 步骤 6：准备折线图绘制数据

将统计结果中的年份作为横轴数据，注册用户人数作为纵轴数据。

```
x = year_counts.index  
y = year_counts.values
```

## 步骤 7：绘制用户注册趋势折线图

绘制“平台注册用户年度增长趋势折线图”，具体要求如下：

- 使用 `plt.figure(figsize=(10, 6))` 设置画布大小
- 使用 `plt.plot()` 绘制折线图
  - 横轴为注册年份
  - 纵轴为注册用户人数
- 设置折线颜色为蓝色，节点使用圆点标记
- 设置图表标题为“平台注册用户年度增长趋势”
- X轴标签设置为“注册年份”
- Y轴标签设置为“注册用户数量（人）”
- 显示图例并添加网格线
- 使用 `tight_layout()` 优化布局并显示图表

```
plt.figure(figsize=(10, 6))  
  
plt.plot(  
    x, y,
```

```
        color="blue",
        marker="o",
        linestyle="--",
        linewidth=2,
        label="年度新增注册用户"
    )

plt.title("平台注册用户年度增长趋势", fontsize=16)
plt.xlabel("注册年份")
plt.ylabel("注册用户数量 (人) ")

plt.xticks(x)
plt.legend()
plt.grid(True, linestyle="--", alpha=0.5)

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