

《数据预处理与分析》练习05

练习题：不同商品类别的平均用户评分柱状图

【实操任务描述】

平台希望从用户评分数据中分析不同商品类别的整体口碑水平，以评估各类商品在用户中的受欢迎程度。

请你结合商品评分数据 `product_ratings.csv` 与商品信息表 `products.csv`，完成商品类别平均评分的统计分析，并使用柱状图进行可视化展示。

【任务要求】

步骤 1：导入库并设置中文显示

使用 `pandas` 完成数据处理，使用 `matplotlib` 进行图表绘制，并正确配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：读取数据

分别读取用户评分表 `product_ratings.csv` 与商品信息表 `products.csv`。

```
ratings = pd.read_csv("product_ratings.csv")
products = pd.read_csv("products.csv")
```

步骤 3：表联合补充商品类别信息

以 `product_id` 为关联键，对评分表与商品表执行内连接（inner join），在评分数据中补充商品所属类别 `category`。

```
data = ratings.merge(
    products[["product_id", "category"]],
    on="product_id",
    how="inner"
)
```

步骤 4：按商品类别计算平均评分

以商品类别 `category` 分组，对评分字段 `rating` 计算平均值，并将统计结果整理为 DataFrame，汇总字段命名为 `avg_rating`。

```
category_rating = (  
    data.groupby("category")["rating"]  
        .mean()  
        .reset_index(name="avg_rating")  
)
```

步骤 5：筛选平均评分最高的前 5 个商品类别

按平均评分 `avg_rating` 从高到低排序，筛选评分排名前 5 的商品类别，用于后续可视化展示。

```
top5 = category_rating.sort_values(  
    by="avg_rating",  
    ascending=False  
) .head(5)
```

步骤 6：绘制商品类别评分分析柱状图

绘制“商品类别用户评分 Top5 柱状图”，要求如下：

- 使用 `plt.figure(figsize=(10, 6))` 设置画布大小；
- 使用 `plt.bar()` 绘制柱状图：
 - 横轴为商品类别
 - 纵轴为平均评分
- 设置图表标题为“用户评分较高的商品类别 Top5”；
- X 轴标签设置为“商品类别名称”；
- Y 轴标签设置为“平均用户评分”；
- 在每个柱子上方标注对应的平均评分值（保留两位小数）；
- 使用 `tight_layout()` 优化布局并显示图表。

```
plt.figure(figsize=(10, 6))  
  
plt.bar(top5["category"], top5["avg_rating"], color="skyblue")  
  
plt.title("用户评分较高的商品类别 Top5", fontsize=16)  
plt.xlabel("商品类别名称")  
plt.ylabel("平均用户评分")  
  
# 添加数值标注  
for i, v in enumerate(top5["avg_rating"]):  
    plt.text(i, v, f"{v:.2f}", ha="center", va="bottom")
```

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