

个人近一年消费数据分析报告

项目概述:

- 关键词: Python · Pandas · 数据清洗 · 可视化 · 探索性分析
- 项目周期: 2026 年 2 月 · 独立完成 · 总计约 20 小时
- 价值: 通过分析近一年、3128 条(实际有效条数 2855 条)消费记录,识别出 16% 的月度支出优化空间,建立个人财务管理模型,实现合理消费

分析内容与目标:

- 消费情况: 钱主要花在哪里? 何时容易过度, 冲动消费? 哪些支出可优化?
- 分析任务: 支出结构分析、时间模式挖掘、可优化点识别
- 产出标准: 形成可视化报告, 提出具体、可执行的优化建议

分析过程:

1.数据准备: 导出支付宝近一年账单, 用 `pd.read_csv` 加载数据至 `pandas`

2.数据质量报告:

- 原始数据: 3128 条记录, 9 个字段
- 缺失值: 3 列存在缺失(共 28 处, 占比<1%)
- 异常值: 发现 15 条金额异常记录(已处理)
- 清洗后: 2855 条有效记录, 数据完整度 91.3%

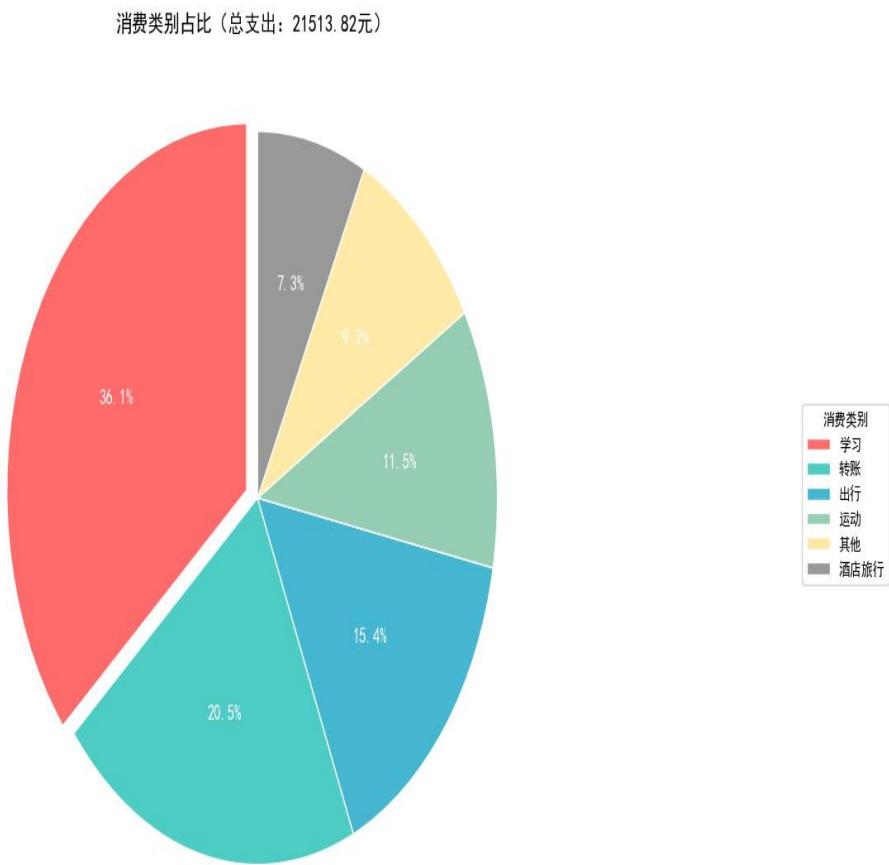
3.数据清洗:

- 删除无用列: `df.drop(columns=['Unnamed: 8'])`

- 筛选支出数据: `df = df[df['收支类型'] == '支出']`
- 清洗金额字段: 使用正则表达式移除非数字字符, 转换为 float 类型
- 过滤无效金额: 移除金额 ≤ 0 的记录
- 转换「记录时间」为日期格式
- 统一「分类」名称, 分为学习, 转账, 出行, 运动, 其他六类, 进行消费结构分析
- 最终有效数据: 2855 行

分析结果:

1. 消费结构

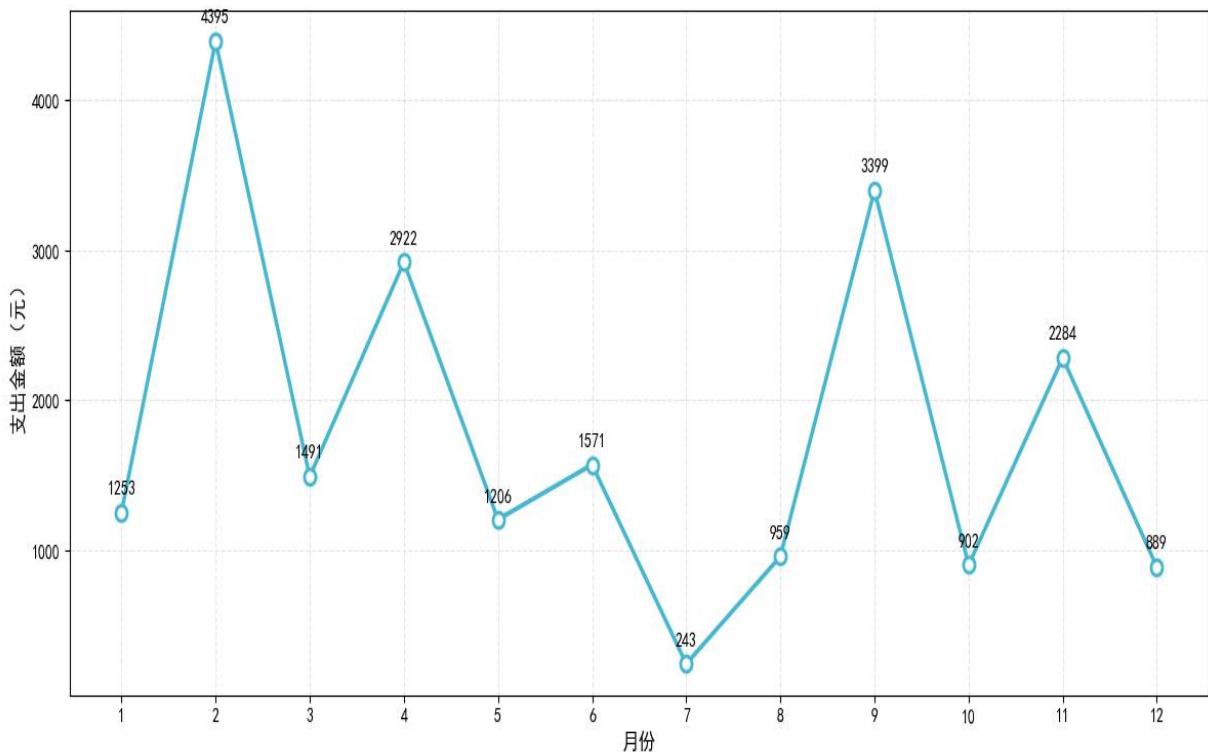


结论与建议: 转账类消费占比达 20.5% (4400 元), 其中 76% 为朋友

聚餐后 AA，建议将降低聚餐频次从每月 6-8 次降至 3-4 次，预计每月可节省 500-600 元

2.月度消费趋势

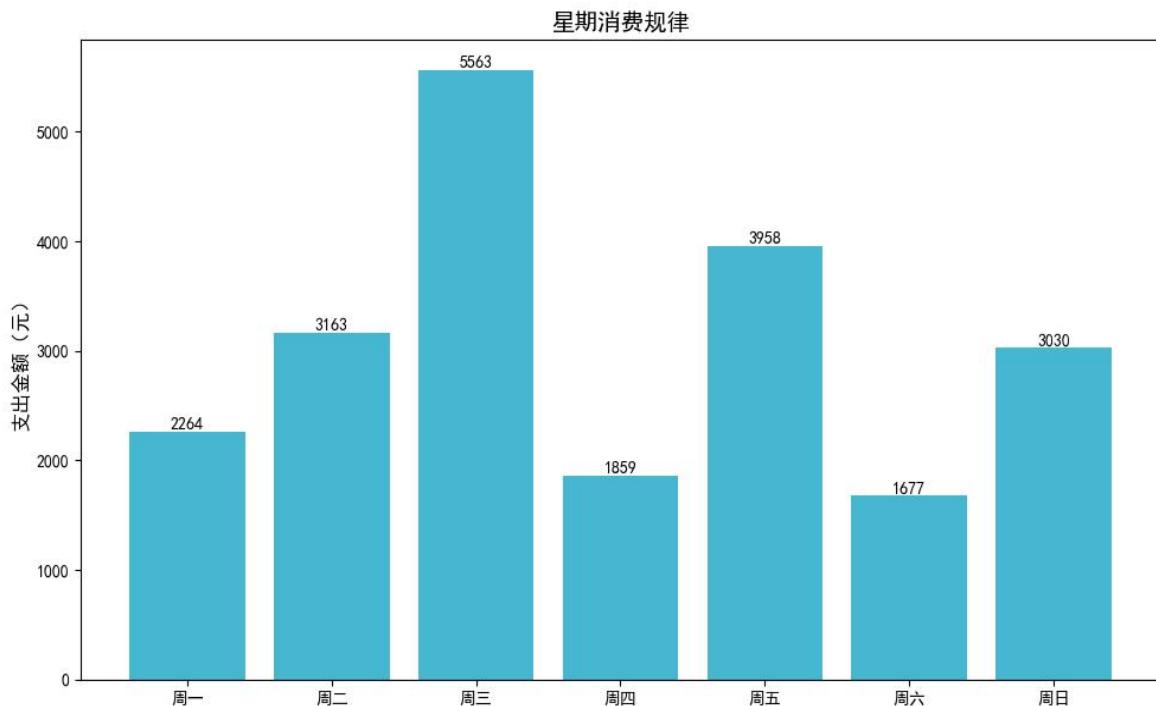
月度支出趋势折线图（1-12月）



结论与建议：2 月与 9 月均为假期与学期过渡期，朋友聚餐多，4 月和 9 月有假期旅游，导致支出偏高

3.消费集中度：金额最大的前 10 笔支出总计 5128 元（占比 23.8%），其中有 7 笔属于旅游类（共计 3640 元，占比 71%），且均为飞机票，以后中短途旅行采用火车出行，可节省 30%-50% 费用

4.星期消费趋势



结论与建议：因为周三，周五课少，聚餐集中于这两天

最终结论与行动

结论

- 可优化空间：每月约 16% 的支出（主要来自聚餐与出行方式）
- 行动建议：设立聚餐预算上限、建立周三，周五消费预警
- 预期效果：每月节省开支 344 元，每年节省 4000 元-4500 元

项目复盘

- 技术收获：掌握了数据导入，Pandas 数据清洗、统一消费分类为学习，转账，出行等 6 类、时间序列分析、数据可视化，总结分析与建议，自己完成了数据分析完整流程
- 难点突破：选择什么图表来进行可视化，如何提高消费自动分类边界的准确率
- 分析局限性：

1. 仅包含支付宝数据，未覆盖现金/微信支付
 2. 消费分类依赖商户名称关键词，可能存在误分类
 3. 未考虑收入端数据，无法计算真实储蓄率
- 可改进点：未来可加入收入数据计算储蓄率，或建立自动化系统