#宠物行业客户流失分析与预测报告

### 1. 代码解析与业务逻辑

#### 1.1 模拟数据生成 (`generate\_simulated\_data`)

- \*\*目标\*\*：创建符合zooplus业务特性的数据集

- \*\*核心字段\*\*：

- `user\_id`：用户唯一标识

- `signup\_date`：注册日期（随机分布在2023年）

- `is\_plus\_member`：是否为付费会员（40%概率是）

- `pet\_type`：宠物类型（狗60%，猫40%）

- `avg\_order\_value`：平均订单价值（付费会员均值€35，免费会员€25）

- `orders\_last\_90d`：近90天订单数（泊松分布，付费会员λ=3.5，免费λ=1.8）

- `last\_order\_days\_ago`：距最后订单天数（随机10-100天）

- `avg\_days\_between\_orders`：平均下单间隔（付费会员25天，免费45天）

- `browsing\_freq`：浏览频率（付费会员均值2.5次/周，免费0.8次）

- `renewal\_date`：续费日期（注册日期+365天）

- `churn`：是否流失（根据业务规则生成）

- \*\*流失规则\*\*：

- \*\*付费会员\*\*：同时满足`最后订单>60天前`且`近90天订单<2`时，70%概率流失；否则20%概率流失

- \*\*免费会员\*\*：`最后订单>90天前`则流失

- \*\*业务意义\*\*：模拟了用户活跃度下降导致流失的场景，特别关注付费会员续费前的沉默现象。

#### 1.2 沉睡客户分析 (`analyze\_dormant\_users`)

- \*\*沉睡定义\*\*：付费会员中`最后订单>60天前`的用户

- \*\*分析步骤\*\*：

1. 计算沉睡用户占比（输出16.9%）

2. 描述性统计（近90天订单数、浏览频率等）

3. 生成5张关键图表：

- \*\*近90天订单量分布dormant\_order\_distribution.png\*\*：直方图+KDE，红线标记干预阈值（1.5单）

- \*\*客单价对比箱线图dormant\_spend\_boxplot.png\*\*：比较沉睡用户与活跃用户的客单价差异

- \*\*行为特征散点图\*dormant\_behavior\_scatter.png\*：以`距最后订单天数`和`浏览频率`为坐标，用宠物类型和客单价标注

- \*\*宠物类型与沉睡比例堆叠图dormant\_pet\_type\_stacked.png\*\*：展示不同宠物类型用户的沉睡比例

- \*\*沉睡状态与流失率热力图dormant\_churn\_heatmap.png\*\*：验证沉睡用户与流失的正相关性

#### 1.3 流失预测模型 (`build\_churn\_model`)

- \*\*特征工程\*\*：

- 删除无关字段（用户ID、日期等）

- 对分类变量`pet\_type`进行独热编码（生成`pet\_type\_cat`字段）

- \*\*模型选择\*\*：随机森林分类器（Pipeline包含标准化和分类器）

- \*\*模型优化\*\*：网格搜索调参（n\_estimators, max\_depth, min\_samples\_split）

- \*\*评估指标\*\*：

- AUC（0.9174）

- 分类报告（precision/recall/f1-score）

- ROC曲线

曲线：模型性能(越靠近左上角越好)

虚线：随机猜测基线

AUC=0.917：优秀区分能力(>0.9)

- 特征重要性排序feature\_importance.png

排名：last\_order\_days\_ago > orders\_last\_90d > browsing\_freq

解读：距最后订单天数是流失最强预测因子

### 2. 关键指标解读

#### 2.1 沉睡客户分析指标

- \*\*沉睡客户占比16.9%\*\*：付费会员中有近17%处于沉默状态（最后订单>60天），需优先干预。

- \*\*订单分布\*\*：沉睡用户近90天平均订单3.4单，但25%用户≤2单（应重点干预≤2单用户）。

- \*\*客单价对比\*\*：沉睡用户客单价略低于活跃用户（箱线图中位数偏低）。

- \*\*宠物类型洞察\*\*：狗主人沉睡比例略高于猫主人（堆叠图显示狗用户沉睡占比更高）。

- \*\*沉睡与流失关系\*\*：热力图显示沉睡用户的流失率高达80.9%，证实沉睡是流失强信号。

#### 2.2 模型评估指标

- \*\*AUC=0.9174\*\*：模型区分流失与非流失用户的能力极强（>0.9属于优秀）。

- \*\*特征重要性\*\*：

1. `last\_order\_days\_ago`（距最后订单天数）：最重要特征，符合业务认知

2. `orders\_last\_90d`（近90天订单数）

3. `browsing\_freq`（浏览频率）

4. `avg\_days\_between\_orders`（平均下单间隔）

- \*\*分类报告\*\*：聚焦流失用户（class=1）的指标：

- Precision：预测为流失的用户中实际流失的比例

- Recall：实际流失用户中被模型正确预测的比例

- F1-score：precision和recall的调和平均

---

### 3. 可视化图表解析

1. \*\*近90天订单量分布\*\*(dormant\_order\_distribution.png)

\*\*洞察\*\*：分布右偏，约30%用户订单≤2单（红线左侧），建议对低订单用户启动干预。

2. \*\*客单价对比箱线图\*\*(dormant\_spend\_boxplot.png)

\*\*洞察\*\*：沉睡用户客单价中位数低于活跃用户，但存在高价值沉睡用户（上须较长）。

3. \*\*行为特征散点图\*\*(dormant\_behavior\_scatter.png)

\*\*洞察\*\*：

- 点越大表示客单价越高，颜色区分宠物类型

- 右上角用户（长时间未购+高浏览）可能是比价用户，需价格锚定策略

- 左下角用户（刚沉默+低浏览）可能已流失，需紧急召回

4. \*\*宠物类型与沉睡比例\*\*(dormant\_pet\_type\_stacked.png)

\*\*洞察\*\*：狗主人沉睡比例(18.5%)高于猫主人(14.2%)，可能与狗粮购买频次更高有关。

5. \*\*沉睡状态与流失率热力图\*\*(dormant\_churn\_heatmap.png)

\*\*洞察\*\*：沉睡用户流失率80.9% vs 非沉睡用户9.7%，证实沉睡是核心预警指标。

6. \*\*ROC曲线\*\*(roc\_curve.png)

\*\*洞察\*\*：曲线靠近左上角，AUC=0.917表明模型预测能力优秀。

7. \*\*特征重要性\*\*(feature\_importance.png)

\*\*洞察\*\*：最后订单天数、近90天订单数是最强预测因子，验证了业务假设。

---

### 4. 分析报告

1. \*\*背景与目标\*\*

- 分析宠物消费付费会员沉睡风险，预测流失用户

- 目标：降低付费会员流失率，提升订阅续费率

2. \*\*核心发现\*\*

- 付费会员沉睡率16.9%，其流失概率高达80.9%

- 关键预警信号：最后订单>60天 + 近90天订单<2

- 高价值用户特征：高客单价 + 高浏览频率的狗主人

3. \*\*行动建议\*\*

- \*\*分层干预策略\*\*：

| 用户类型 | 干预措施 | 预期效果 |

|-------------------------|----------------------------------|-------------------------|

| 高客单价沉睡用户 | 赠送宠物健康指南 + 专属客服通道 | 挽回LTV€120+用户 |

| 低订单频率用户 | 推送“订阅锁价”猫粮/狗粮套餐 | 提升复购频次1.5次/季度 |

| 高浏览低转化用户 | APP弹窗提供限时闪购 | 转化率提升15% |

- \*\*产品优化\*\*：针对狗主人设计专属订阅包（含狗粮、零食、玩具）

4. \*\*模型应用方案\*\*

- 上线流失预警系统：对预测流失概率>70%的用户自动触发干预

- 监控指标：干预ROI = (留存用户LTV - 干预成本) / 总干预成本