

摩拜单车客群分层

数据维度

原数据字典

user_id 用户编码
start_time 开始时间
end_time 结束时间
timeduration 骑行时长
bikeid 自行车编码
tripduration 骑行距离
from_station_id 开始站编码
from_station_name 开始站名字
to_station_id 结束站编码
to_station_name 结束站名字
usertype 用户种类
gender 性别
birthyear 出生年份
age 年龄

数据维度分类

用户数据

- user_id 用户编码
- usertype 用户种类
- gender 性别
- birthyear 出生年份
- age 年龄

骑行数据

- start_time 开始时间
- end_time 结束时间
- timeduration 骑行时长
- bikeid 自行车编码
- tripduration 骑行距离
- from_station_id 开始站编码
- from_station_name 开始站名字
- to_station_id 结束站编码
- to_station_name 结束站名字

数据分析流程

数据概况分析

info()

- 查看空值
- 查看无意义变量
- 查看类别型变量 发现age是类别型变量

describe()

- 查看异常值

head()

- 查看数据前5行

单变量分析

删除对业务没有实际作用的变量

- Unnamed: 0
- user_id
- from_station_id
- to_station_id
-

删除空值

- gender
- birthyear
-

删除异常值

- birthyear

处理时间数据

- start_time
- end_time

将类别型变量转化为数字型变量

- age与tripduration调整数据格式
- 其他类别型变量转化为哑变量

聚类分析模型

模型建立

- 选择变量并进行数据标准化
- 建立聚类分析模型，创建群数
- 查看分群效果
 - groupby()
 - 散点图

模型评估

- 计算轮廓系数

模型优化

- 调整模型变量
- 调整群数

模型的业务解读

导出csv
解读模型