

问题：预测双十一后的6个月内的返客情况

# count/ratio

day count：每月有某个动作的天数

product diversity：每月在不同类别下的商品点击数

Penetration features.：回头客的数量

# aggregation

Monthly aggregation：每个月对以上feature的均值，方差，最大值等

User aggregation：反映商户的指标，关于用户的聚合。如某件商品各用户买过几次的均值，方差等。对不同的年龄性别的聚合。用户点击量的聚合

Merchant aggregation：反映用户的指标，关于商户的聚合。如一个用户一个月进行过几次购买

# Recent activity

Double 11：双十一当天的动作情况

Latest one-week / one-month：双十一前一周/月的情况

# Complex features

Trend：反映各月趋势平均情况的指标

Repeat buyer：返客率

Market share：一个种类/品牌对商户的贡献率；或一个商户对一个种类的贡献率

User-merchant similarity：对于某个种类/品牌，商户和用户的相似度。商户对某种类占率和用户对次种类购买占总购买的购买率的内积。

PCA features：商户间的相似度矩阵（每件物品的用户购买量来衡量），矩阵过大，用pca降维

# Age/gender related

不同年龄性别对商品种类的喜好程度不同