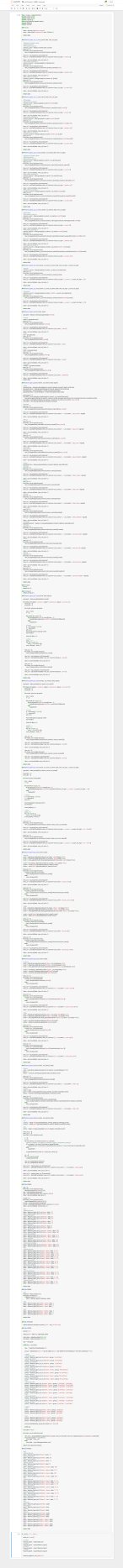
## 代码运行

[点击查看原图](a.png)

图中代码未运行，因为耗时较长，且无输出，有需要建议使用cmd运行，特征文件将产生在代码文件所在目录下



### 讨论分析

1. 本小题仅为提取特征，并不包含任何预测部分。
2. 特征提取对象：
   1. 单对象的特征：针对用户、品牌、类别、商品分提取特征
   2. 联合特征：用户-品牌之间的交互特征（用户-类别、用户-商品...）
3. 特征类型：
   1. 整个时间段简单统计（购买次数、购买天数、消费金额...）
   2. 每月统计（同上）
   3. 聚合特征：按月统计的特征进行聚合...
   4. 最近时间段内的特征（最近一个月、最近一周）
   5. 复杂特征（用户购买的趋势、用户-品牌间的相似度）
   6. 时间特征（用户购买周期、最近一次购买到待预测时间段的距离...）
4. 具体特征实现：[详见PDF](feature_ref.pdf)
   1. 总阶段特征：品类、商品、品牌分别被用户购买的次数、天数，用户购买的品类、商品、品牌种类的数量，品类、商品、品牌被购买的买家数
   2. 每月特征：以上各特征的当月度计算值
   3. 聚合特征：针对月度特征的mean、std、max、median的月度聚类值，全阶段特征针对用户的四种聚类值，全阶段特征针对品类、商品、品牌的四种聚类
   4. 近期特征：针对当前记录发生7天之内的周度特征，包含之前全部的全阶段特征种类
   5. 复杂特征：月度特征的变化率、均值、标准差，品类、商品、品牌被重复购买的比率、被重复购买的人数、被同一人重复购买的天数，用户重复购买的天数、重复购买所占购买次数的比率、占购买天数的比率
   6. 市场特征：某品牌某品类对该品牌、该品类的市场占比、商品数量占比，以及特定购买人群对特定商品、品牌、品类的购买率
5. 特征统计：特征参考共提供6+128+36+36+104+96+24+4+8=500个特征，实际生产464个特征（去除原数据所有列）

C:\Users\ThinkPad\Documents\Tencent Files\799498265\Image\C2C\$Q8H4{0@6U_]]2RAF7{{ESU.png

1. 特征审查：
   1. 代码内特征生产按照type\_page\_level的方式进行索引命名，可重用的部分代码则取最早的命名，对于特定特征的审查可直接在代码内查找，附有一定注释，由于特征过多，此处不做赘述。
   2. 产生的特征文件按照type1-4.csv进行命名，含义为包含type1-k的全部特征。