《Python编程》大作业

姓名： 学号： 班级：

一、题目要求

提供的csv文件是某商店的销售纪录，其前5行记录如下：

user\_id,item\_id,behavior\_type,user\_geohash,item\_category,time

10001082,285259775,1,97lk14c,4076,2014-12-08 18

10001082,4368907,1,,5503,2014-12-12 12

10001082,4368907,1,,5503,2014-12-12 12

10001082,53616768,1,,9762,2014-12-02 15

可以看出，每行记录都包含5个字段，数据集中的字段及其含义如下：

user\_id（用户id）

item\_id(商品id)

behaviour\_type（包括浏览、收藏、加购物车、购买，对应取值分别是1、2、3、4）

user\_geohash(用户地理位置哈希值，数据处理时暂不考虑)

item\_category（商品分类）

time（该记录产生时间）

要求进行Python编程实现以下功能（可以导入任意的库来辅助你完成）：

1. 找出“购买效率达人”：根据behaviour\_type的数据，找出购买行为占比最高的前10个用户id
2. 找出用户10001082的“同道中人”：即找出与该用户浏览、收藏、加购物车、购买的商品分类近似的前10个用户id
3. 给出三种数据画像：这里不给出具体的要求，你可以从自己的分析角度对数据给出各种效果显示，如用户关注的商品分类饼图；某用户的购物活动时间折线图；商品分类关注度词云图；用户活跃度（前10）柱状图等。

二、解答

1、

2、

3、