

一、背景

某电商平台为了合理的投入人力物力创造更大的销售利润，现对已有的销售数据进行用户分析，提出合理的促销计划。

围绕**产品**和**用户**两大方面展开为电商平台制定策略提供分析及建议。

二、需求

- 用户分析：从性别、年龄、职业、城市、居住年限，婚姻状况等维度找到高质量用户，并查看高质量用户人群的占比，为其提供高价值消费品(定位高价值消费品以销售金额评估)。针对其他的用户，主要引导用户进行购买，多推荐一些热销的商品(定位热销产品)
- 产品分析：从销量、销售额都高的产品并以二八法则找到高贡献的一级产品类目

三、数据介绍

假定每条记录为一单

字段名称	字段描述
User_ID	顾客ID
Product_ID	商品ID
Gender	顾客性别
Age	顾客年龄
Occupation	顾客从事职业ID
City_Category	城市类别
Stay_In_Current_City_Years	在现城市呆的年数
Marital_Status	婚姻状况
Product_Category_1	商品类别1
Product_Category_2	商品类别2
Product_Category_3	商品类别3
Purchase	消费金额

```
--建表
create table model2_datas
(User_ID int,
Product_ID string,
```

```

Gender string,
Age string,
Occupation int,
City_Category string,
Stay_In_Current_City_Years string,
Marital_Status int,
Product_Category_1 int,
Product_Category_2 int,
Product_Category_3 int,
Purchase double
)
row format delimited fields terminated by ',' --指定分隔符csv 为逗号分割
tblproperties(
"skip.header.line.count"="1" --跳过文件行首1行
);

--装载数据
load data local inpath '/home/hadoop/datas/model2_datas.csv' overwrite into
table model2_datas

```

四、需求实现

1. 查询订单整体的消费情况（包括：总销售额、人均消费、平均每单消费）（10分）
2. 用户分析（找到高质量人群）（30分）
 - 统计各性别消费情况(字段包含性别、人数、人数占比、人均消费、消费金额、消费占比) 并以消费占比降序
 - 统计各年龄段消费情况(字段包含年龄段、人数、人数占比、消费金额、人均消费、消费占比)并以消费占比降序
 - 统计各职业消费情况(字段包含职业、人数、人数占比、消费金额、人均消费、消费占比) 并以消费占比降序
 - 统计各婚姻状况消费情况(字段包含婚姻状况、人数、人数占比、消费金额、人均消费、消费占比) 并以消费占比降序
 - 依据以上查询结果找到高质量人群（如：性别为...，年龄段为...，职业为...，婚姻状况为...）
3. 产品分析（30分）
 - 查询出订单量TOP10的产品(包含字段 排名编号、商品ID、销量、销量占比)按销量降序显示
 - 查询出销售额TOP10的产品(包含字段 排名编号、商品ID、销售额、销售额占比)
 - 统计各一级产品类目的订单量、销售额、订单量占比、销售额占比、累计销售额占比（以销售额占比降序），并根据查询结果找到累计销售额达到20%的几个一级产品类目
4. 细化分析（30分）
 - 查询出各性别销售额TOP10产品（字段包含商品ID、订单量、销售额、销售额占比、类别1、类别2、类别3）
 - 查询高质量用户群年龄段的订单量TOP10产品（字段包含商品ID、订单量、销售额、订单量占比、

销售额占比、类别1、类别2、类别3)