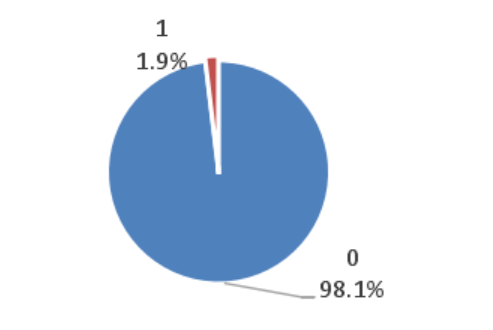
**描述性统计及可视化**

1. train\_m.csv
2. **饼图**，使用**变量label**绘制，取值为0、1，其中0为未转化，1为转化，**效果大概如图：**



1. adFeature.csv
2. **变量creativeSize（素材大小），**绘制条形图。

一共有四个取值（即四个类目）4、6、9、11。

横坐标（Count）即四个类目，纵坐标（Size**\_**Category）为计数即数据集中每个类目的数量。

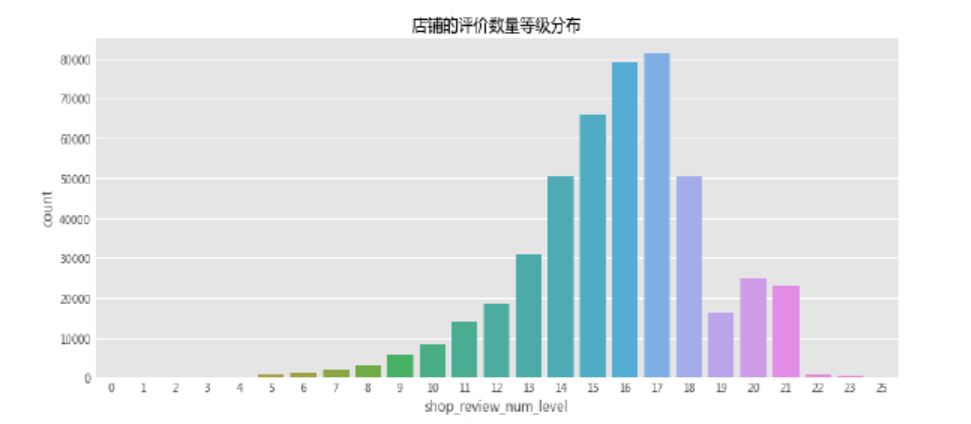
1. **变量productType（商品类型），**绘制条形图

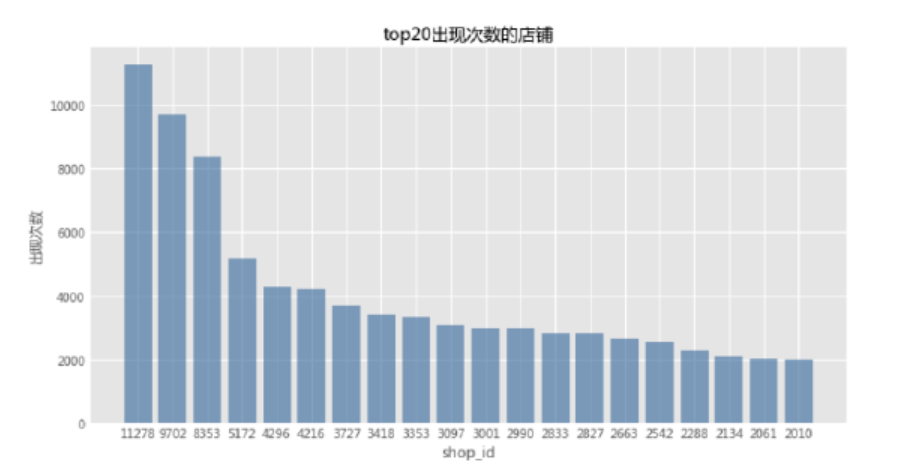
首先将取值分成6段，1=（0-20]、2=（20-40]、3=（40-60]、4=（60-80]、5=（80-100]、6=（100-120]。横坐标（Count）即六个分段，纵坐标（ProductType\_Level）为计数，即数据集中每个分段样本的数量。

1. **变量adCategoryId（广告类目），**绘制条形图

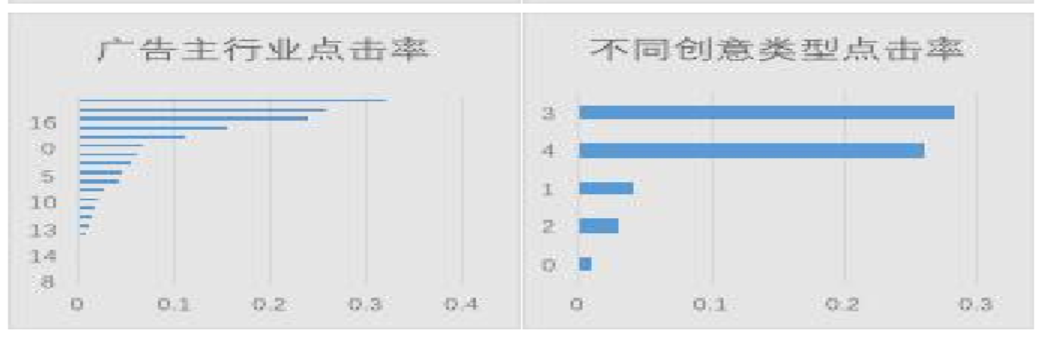
首先将取值分成12段，1=（0-25]、2=（25-50]、3=（50-75]、4=（75-100]、5=（100-125]、6=（125-150]，7=（150-175]、8=（175-200]、9=（200-225]、10=（225-250]、11=（250-275]、12=（275-300]。横坐标（Count）即12个分段，纵坐标（AdCategoryId\_Level）为计数，即数据集中每个分段样本的数量。

**效果大概如图：（不用有标题）**





1. userFeature\_old.csv
2. **变量age（年龄），**绘制条形图。横坐标是计数，纵坐标为分类。效果：



**2、变量education（学历），**绘制条形图。横坐标是计数，纵坐标为分类。效果同上

**3、变量carrier（移动运营商），**绘制条形图。横坐标是计数，纵坐标为分类。效果同上

**4、变量ct（上网链接类型），**绘制条形图。横坐标是计数，纵坐标为分类。效果同上

**5、变量gender（性别），**绘制饼图。

**6、变量consumptionAbility（消费能力），**绘制饼图

**7、变量house（有房），**绘制饼图

**8、变量os（操作系统），**绘制饼图



1. train\_m.csv
2. 不同广告素材大小对转化率的影响，需要使用**变量creativeSize（素材大小）和变量label（是否转化）。**creativeSize（素材大小），一共有四个取值（即四个类目）4、6、9、11，分别统计四个类目中label=1的样本数量。例如：

**“creativeSize=4”的转化率 =**

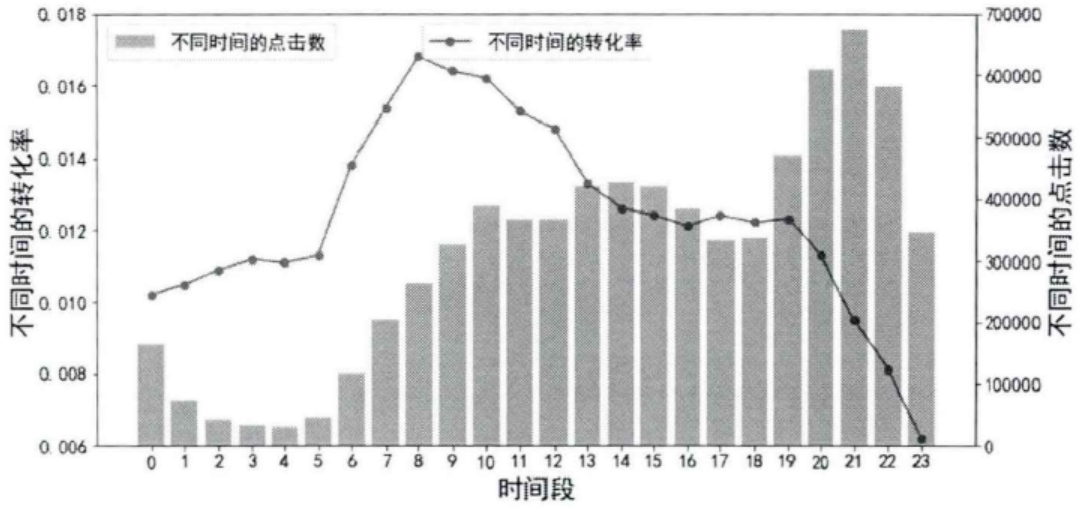
**数据集中”creativeSize=4”的样本中”label=1”的样本数量 / ”creativeSize=4”的样本数量**

**“creativeSize=6”的转化率 =**

**数据集中”creativeSize=6”的样本中”label=1”的样本数量 / ”creativeSize=6”的样本数量**

**以此类推。**

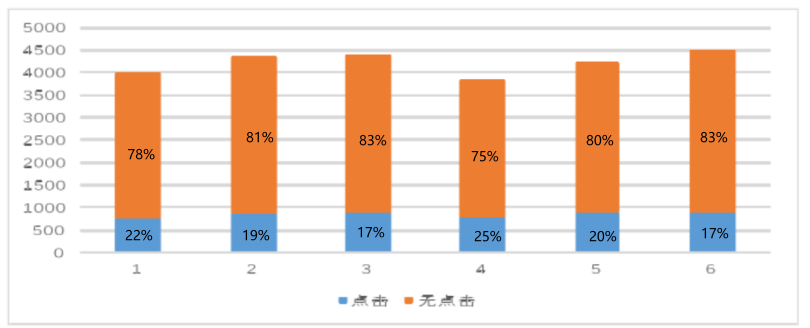
**效果如图：**



**绘制 折线图+条形图，横坐标为分组，**

**左边纵坐标为当下分组的转化率，右边纵坐标为当下分组 ”label=1”的count计数。**

1. 不同商品类型对转化率的影响，需要使用**变量productType（商品类型）和变量label（是否转化），转化率计算方法与绘图同上。**
2. 不同广告类目对转化率的影响，需要使用**变量adCategoryId（广告类目）和变量label（是否转化）。转化率计算方法与绘图同上。**不同操作系统对转化率的影响，需要使用**变量os（操作系统）和变量label（是否转化），转化率计算方法与绘图同上。**
3. 不同移动运营商对转化率的影响，需要使用**变量carrier（移动运营商）和变量label（是否转化），转化率计算方法与绘图同上。**
4. 不同上网链接类型对转化率的影响，需要使用**变量ct（上网链接类型）和变量label（是否转化），转化率计算方法与绘图同上。**
5. 不同移动运营商对转化率的影响，需要使用**变量os（移动运营商）和变量label（是否转化），转化率计算方法与绘图同上。**
6. 不同年龄对转化率的影响，需要使用**变量age（年龄）和变量label（是否转化），绘制堆积图，转化率计算同上。横坐标为分类，纵坐标该分类总样本量。蓝色为转化，橙色为未转化，图形上标明比例。效果如图：**



1. 不同学历对转化率的影响，需要使用**变量education（学历）和变量label（是否转化），绘制堆积图，转化率计算同上。**
2. 不同性别对转化率的影响，需要使用**变量gender（性别）和变量label（是否转化），绘制堆积图，转化率计算同上。**
3. 不同消费能力对转化率的影响，需要使用**变量consumptionAbility（消费能力）和变量label（是否转化），绘制堆积图，转化率计算同上。**
4. 是否有房对转化率的影响，需要使用**变量house（有房）和变量label（是否转化），绘制堆积图，转化率计算同上。**

**五、统计表**

使用train\_m.csv，将四中的11个图的信息进行统计

对于每个变量：（共11个表格）



**六、相关性分析（可删除缺失样本）**

1、对train\_m.csv中，**creativeSize（素材大小）productType（商品类型）adCategoryId（广告类目）三个变量进行相关性分析**

**2、**、对train\_m.csv中，**age（年龄）education（学历）carrier（移动运营商）ct（上网链接类型）gender（性别）consumptionAbility（消费能力）house（有房）os（操作系统）八个变量进行相关性分析**

两个表格，效果如图：

