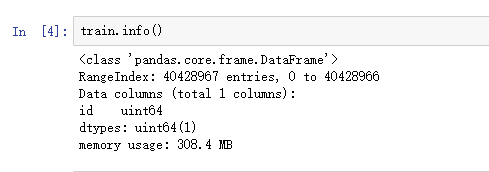
# 各列特征分析

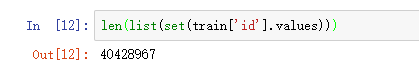
## id

说明: ad identifier （广告 ID）



Int64类型 , 总40428967条记录。

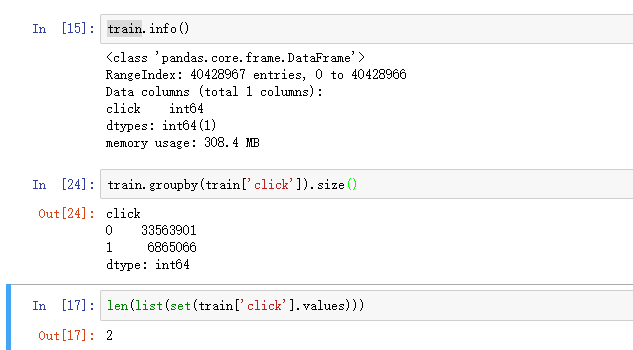
Id中去重之后，记录数不变，id中没有重复记录



## click

说明: 是否被点击，其中 0 为不被点击，1 为不被点击，此列为目

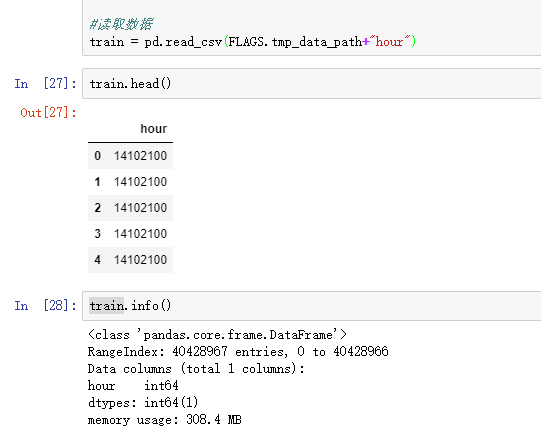
标变量

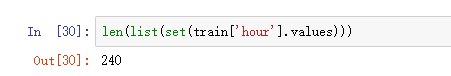


样本中0的样本数量是1样本数量的5倍

## hour

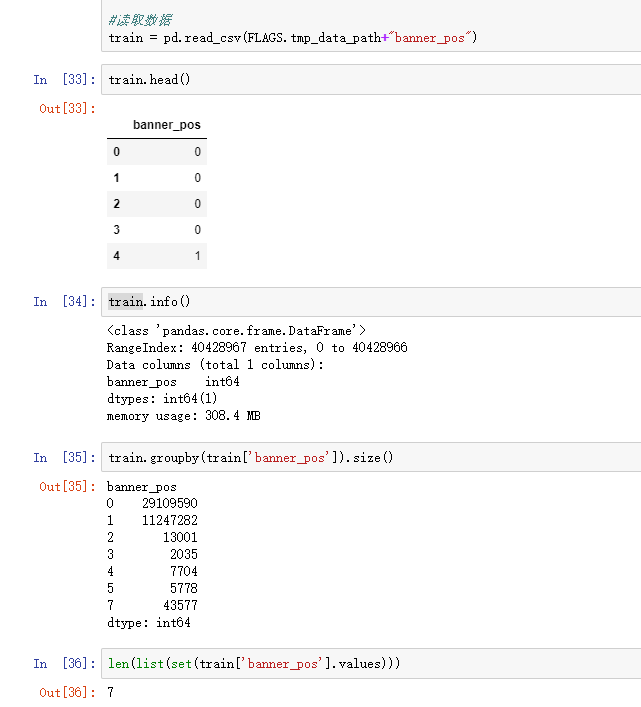
说明: 广告被点击的时间戳





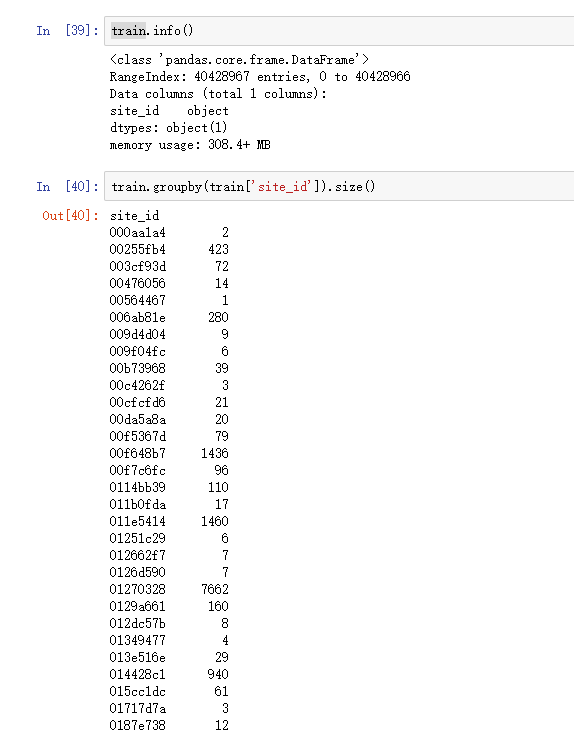
## banner\_pos

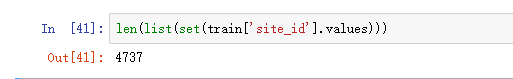
说明：广告所处位置



## site\_id

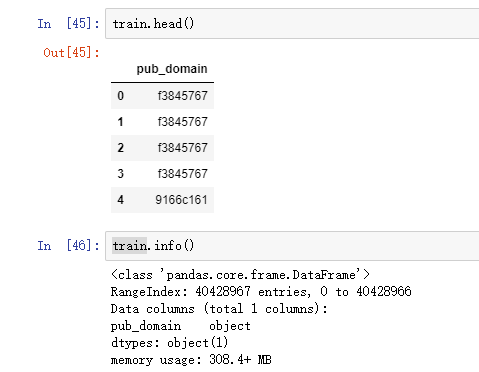
说明：站点 ID

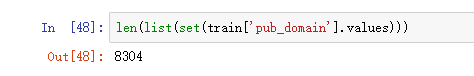




## site\_domain

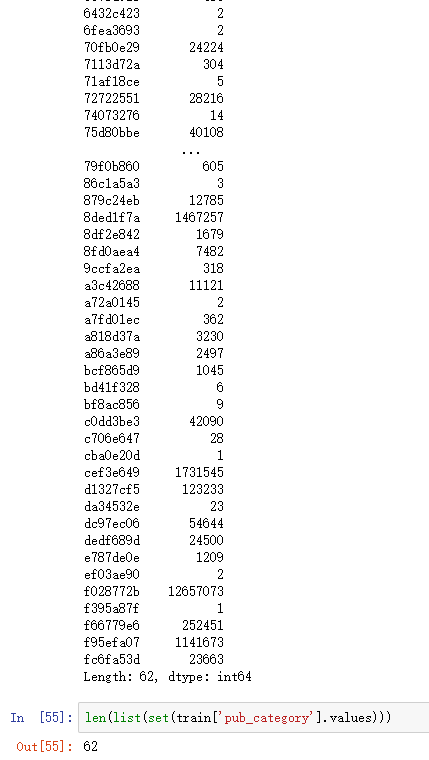
说明: 站点领域





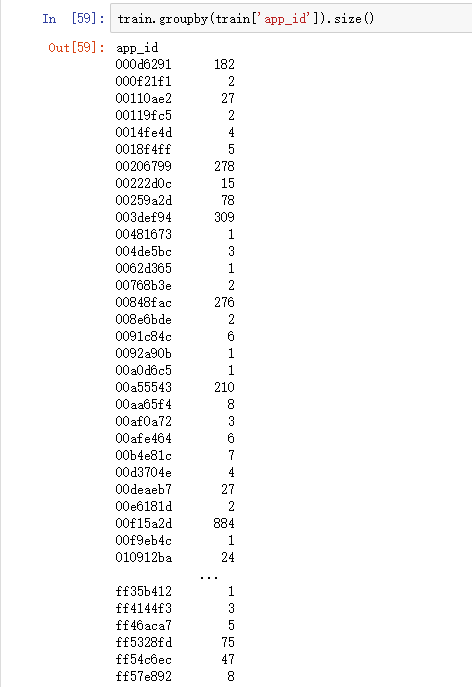
## pub\_category

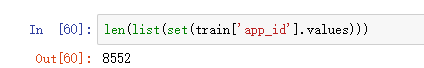
说明：站点类别



## app\_id

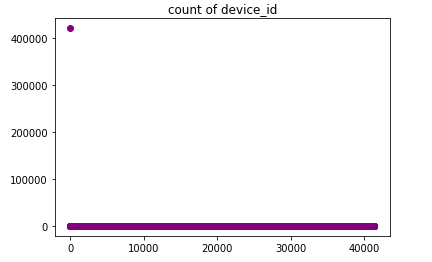
说明：移动端id

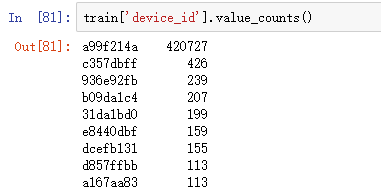




## device\_id

说明:设备id,肯定能用户登入的手机型号或者电脑型号





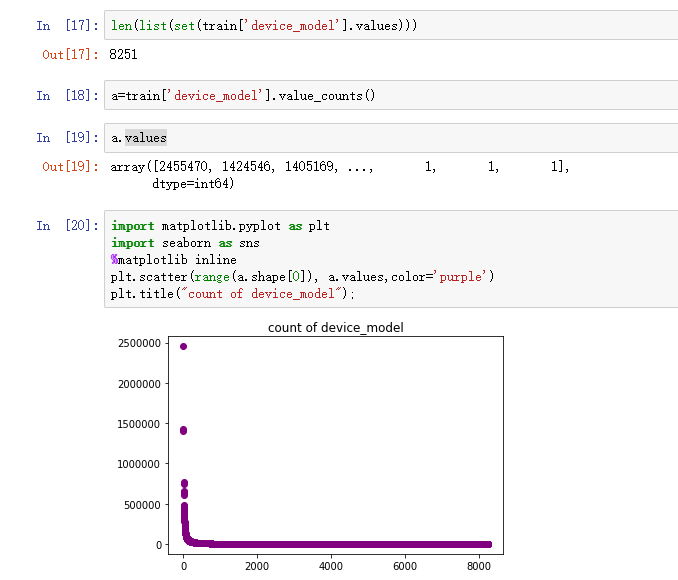
## device\_ip

说明:设备的ip地址



## device\_model

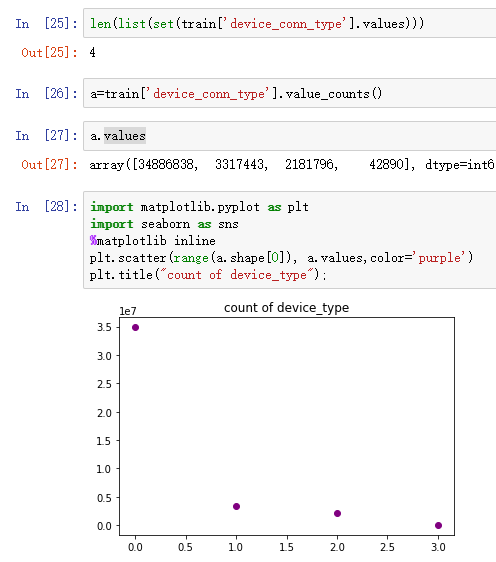
说明: 设备模式



## device\_type

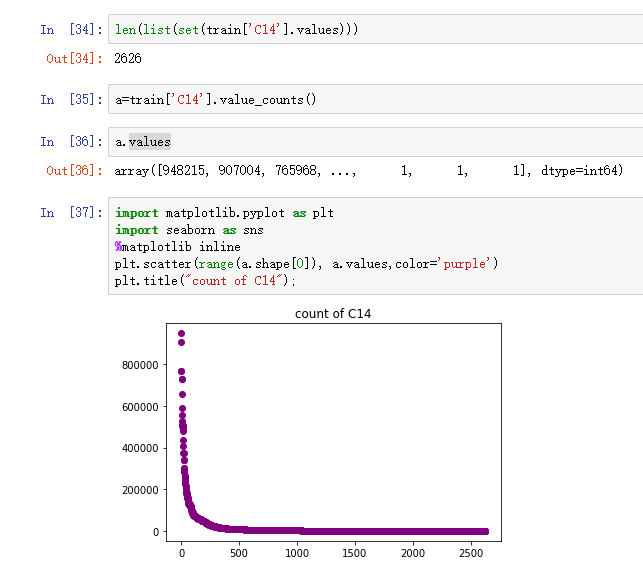
## device\_conn\_type

说明:设备联网方式



## C14

说明:应该是某一字段脱敏后的id或者索引值



## C15

## C16

## C17

## C18

## C19

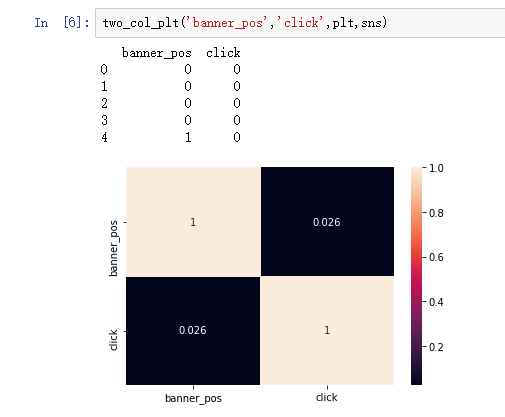
## C20

## C21

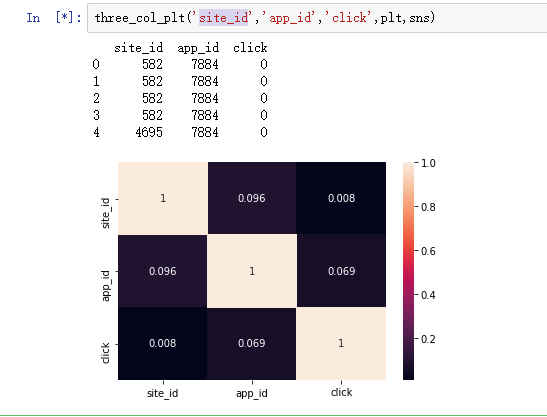
# 特征之间的联合分布

## banner\_pos + click

说明：广告所处位置应该和是否点击存在很强的关联关系



## site\_id or app\_id +click



## device\_model + device\_conn\_type

