

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据：某年淘宝双12部分购物数据；
 - 1.2 分析目的：
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标：PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析：每天访问量
 - 2.6.3 UV分析：每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
 - 3 用户价值分析
 - 3.1 R理解：
 - 3.2 F理解：
 - 3.3 M理解：
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则：
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R，最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF，根据RF进行分类

1 案例:某年淘宝双12部分购物数据

1.1 数据：某年淘宝双12部分购物数据；

来源：天池数据平台

1.2 分析目的：

- 熟悉电商常用分析指标
- 用户行为分析
- 用户价值模型

2 熟悉数据

2.1 导入数据

```
In [2]: 1 pdata[:10]

Out[2]:      user_id  item_id  behavior_type  user_geohash  item_category      time
0  98047837  232431562             1           NaN         4245  2014-12-06 02
1  97726136  383583590             1           NaN         5894  2014-12-09 20
2  98607707   64749712             1           NaN         2883  2014-12-18 11
3  98662432  320593836             1    96nn52n         6562  2014-12-06 10
4  98145908  290208520             1           NaN        13926  2014-12-16 21
5  93784494  337869048             1           NaN         3979  2014-12-03 20
6  94832743  105749725             1           NaN         9559  2014-12-13 20
7  95290487   76866650             1           NaN        10875  2014-11-27 16
8  96610296  161166643             1           NaN         3064  2014-12-11 23
9 100684618   21751142             3           NaN         2158  2014-12-05 23

In [ ]: 1

In [ ]: 1

In [1]: 1 import pandas as pd
        2 fpath = r'G:\data\taobao1920\data.csv'
        3 pdata = pd.read_csv(fpath)
        4 pdata.columns

Out[1]: Index(['user_id', 'item_id', 'behavior_type', 'user_geohash', 'item_category',
              'time'],
              dtype='object')
```

2.2 数据字段说明

字段	说明
user_id	用户ID
item_id	物品ID
behavior_type	用户行为，1:点击 2:收藏 3:加购物车 4:支付
user_geohash	地理位置
item_category	物品种类
time	时间

2.3 数据量大小

```
In [2]: 1 len(pdata)

Out[2]: 12256906
```

2.4 缺失数据检查

- 每个字段缺失数量
- 按照缺失数量排序
- 使用什么知识点？

```
In [3]: 1 #判断缺失数据计算
        2 missAll = pdata.isnull().sum()
        3 missAll

Out[3]: user_id      0
        item_id      0
        behavior_type  0
        user_geohash  8334824
        item_category  0
        time         0
        dtype: int64

In [4]: 1 #统计确实字段
        2 missField = missAll[missAll>0]
        3 missField

Out[4]: user_geohash      8334824
        dtype: int64
```

从结果可以看到：数据中缺少user_geohash字段

2.5 基本分析

2.6 指标：PV与UV

PV(访问量)：Page View,浏览量/点击量(包括重复浏览)。
UV(独立访客)：Unique Visitor, 独立用户或者设备访问量。

常见指标：

- 日访问量

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
 - ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

- 日活
- 月活
- 日/月成交量
- 日/月成交额

指标核心点: 时间

2.6.1 时间处理

问题:

- 时间是否可以直接使用?
- 如何处理当前时间:
- 我们需要什么样的时间

当前时间列为:年-月-日 小时, 将其转化成Timestamp

- 时间戳
- 日期(天为单位)
- 小时

```
In [4]: 1 #将字符串时间转成Timestamp
        2 pdata['ts'] = pd.to_datetime(pdata['time'])

In [5]: 1 pdata['hour'] = pdata.ts.map(lambda item: item.hour)

In [6]: 1 pdata['day'] =pdata.ts.map(lambda item: item.date())
```

In [7]:

1	pdata
---	-------

Out[7]:

Contents

▼ 1 案例:某年淘宝双12部分购物数据

1.1 数据: 某年淘宝双12部分购物数据;

1.2 分析目的:

▼ 2 熟悉数据

2.1 导入数据

2.2 数据字段说明

2.3 数据量大小

2.4 缺失数据检查

2.5 基本分析

▼ 2.6 指标: PV与UV

2.6.1 时间处理

2.6.2 PV分析: 每天访问量

2.6.3 UV分析: 每天独立用户数

2.6.4 PV与UV的数据可视化

2.6.5 用户购买次数分布

▼ 2.6.6 用户复购

2.6.6.1 实现方式1

2.6.6.2 实现方式2

▼ 2.6.7 用户复购时间间隔

2.6.7.1 diff方法

2.6.7.2 实现

▼ 2.7 AARRR模型

2.7.1 理解AARRR

2.7.2 双12用户购物模型

2.7.3 不同层级数量

2.7.4 计算转化率

▼ 3 用户价值分析

3.1 R理解:

3.2 F理解:

3.3 M理解:

3.4 RFM评分模型

3.5 不同指标对应测策略

3.6 用户分类规则:

3.7 RFM评分规则

3.8 灵活使用

▼ 3.9 淘宝双12购物用户RFM模型

3.9.1 整理数据

3.9.2 计算R, 最近购物时间

3.9.3 数据验证

3.9.4 计算R值

3.9.5 计算r

3.9.6 计算F

3.9.7 数据集合并(R与F)

3.9.8 设计RF, 根据RF进行分类

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day
0	98047837	232431562	1	NaN	4245	2014-12-06 02	2014-12-06 02:00:00	2	2014-12-06
1	97726136	383583590	1	NaN	5894	2014-12-09 20	2014-12-09 20:00:00	20	2014-12-09
2	98607707	64749712	1	NaN	2883	2014-12-18 11	2014-12-18 11:00:00	11	2014-12-18
3	98662432	320593836	1	96nn52n	6562	2014-12-06 10	2014-12-06 10:00:00	10	2014-12-06
4	98145908	290208520	1	NaN	13926	2014-12-16 21	2014-12-16 21:00:00	21	2014-12-16
5	93784494	337869048	1	NaN	3979	2014-12-03 20	2014-12-03 20:00:00	20	2014-12-03
6	94832743	105749725	1	NaN	9559	2014-12-13 20	2014-12-13 20:00:00	20	2014-12-13
7	95290487	76866650	1	NaN	10875	2014-11-27 16	2014-11-27 16:00:00	16	2014-11-27
8	96610296	161166643	1	NaN	3064	2014-12-11 23	2014-12-11 23:00:00	23	2014-12-11
9	100684618	21751142	3	NaN	2158	2014-12-05 23	2014-12-05 23:00:00	23	2014-12-05
10	100509623	266020206	3	tfvomgk	4923	2014-12-08 17	2014-12-08 17:00:00	17	2014-12-08
11	101781721	139144131	1	9rgt162	3424	2014-12-13 21	2014-12-13 21:00:00	21	2014-12-13
12	103891828	255365467	1	96sjmho	552	2014-12-01 13	2014-12-01 13:00:00	13	2014-12-01
13	101260672	212072908	1	95q0is4	10984	2014-12-12 11	2014-12-12 11:00:00	11	2014-12-12
14	104811265	322736792	1	NaN	6513	2014-12-11 19	2014-12-11 19:00:00	19	2014-12-11
15	103802946	382807809	1	NaN	13103	2014-11-20 23	2014-11-20 23:00:00	23	2014-11-20
16	104221274	262661866	1	NaN	4370	2014-12-14 12	2014-12-14 12:00:00	12	2014-12-14
17	110418606	144902506	1	NaN	11639	2014-12-09 16	2014-12-09 16:00:00	16	2014-12-09
18	100684618	185630324	1	NaN	5204	2014-11-26 11	2014-11-26 11:00:00	11	2014-11-26
19	100509623	76021805	1	NaN	1308	2014-11-21 12	2014-11-21 12:00:00	12	2014-11-21
20	101781721	19349307	1	9rgt16g	1863	2014-12-12 12	2014-12-12 12:00:00	12	2014-12-12
21	103891828	149380817	1	NaN	7876	2014-12-08 21	2014-12-08 21:00:00	21	2014-12-08
22	101260672	395905225	1	NaN	14018	2014-11-28 09	2014-11-28 09:00:00	9	2014-11-28
23	104811265	328893812	1	NaN	5894	2014-12-05 11	2014-12-05 11:00:00	11	2014-12-05
24	103802946	142178953	1	NaN	12536	2014-11-23 21	2014-11-23 21:00:00	21	2014-11-23
25	104221274	114830658	1	NaN	9516	2014-11-29 09	2014-11-29 09:00:00	9	2014-11-29
26	110418606	335988905	1	NaN	10703	2014-11-25 19	2014-11-25 19:00:00	19	2014-11-25
27	100684618	66146539	1	NaN	3783	2014-12-05 23	2014-12-05 23:00:00	23	2014-12-05
28	100509623	236195070	1	NaN	1863	2014-12-15 14	2014-12-15 14:00:00	14	2014-12-15
29	101781721	46475830	1	9rgt17q	13230	2014-12-16 21	2014-12-16 21:00:00	21	2014-12-16
...
12256876	91530370	277962627	1	NaN	4640	2014-12-16 22	2014-12-16 22:00:00	22	2014-12-16
12256877	91530370	136942019	1	NaN	9820	2014-11-26 13	2014-11-26 13:00:00	13	2014-11-26
12256878	91530370	170765640	1	NaN	3381	2014-12-12 00	2014-12-12 00:00:00	0	2014-12-12
12256879	91530370	309381689	1	NaN	11552	2014-12-16 22	2014-12-16 22:00:00	22	2014-12-16
12256880	91530370	130056262	1	NaN	11552	2014-12-16 22	2014-12-16 22:00:00	22	2014-12-16
12256881	91530370	82501758	1	NaN	552	2014-12-04 23	2014-12-04 23:00:00	23	2014-12-04
12256882	91530370	393936911	3	NaN	8792	2014-11-26 13	2014-11-26 13:00:00	13	2014-11-26
12256883	91530370	393936911	1	NaN	8792	2014-11-26 13	2014-11-26 13:00:00	13	2014-11-26
12256884	91530370	186332216	1	NaN	552	2014-11-26 13	2014-11-26 13:00:00	13	2014-11-26
12256885	91530370	384717078	2	NaN	7876	2014-12-02 23	2014-12-02 23:00:00	23	2014-12-02
12256886	91530370	293543750	3	NaN	552	2014-12-04 23	2014-12-04 23:00:00	23	2014-12-04
12256887	91530370	183076534	1	NaN	9720	2014-12-02 18	2014-12-02 18:00:00	18	2014-12-02
12256888	91530370	183076534	1	NaN	9720	2014-11-28 23	2014-11-28 23:00:00	23	2014-11-28
12256889	91530370	241618446	1	NaN	3381	2014-11-26 12	2014-11-26 12:00:00	12	2014-11-26
12256890	91530370	241618446	1	NaN	3381	2014-11-26 13	2014-11-26 13:00:00	13	2014-11-26
12256891	91530370	183076534	1	NaN	9720	2014-12-02 18	2014-12-02 18:00:00	18	2014-12-02
12256892	91530370	183076534	1	NaN	9720	2014-12-02 18	2014-12-02 18:00:00	18	2014-12-02
12256893	91530370	183076534	1	NaN	9720	2014-11-30 13	2014-11-30 13:00:00	13	2014-11-30
12256894	91530370	131769266	1	NaN	12090	2014-12-03 13	2014-12-03 13:00:00	13	2014-12-03
12256895	91530370	320518586	1	NaN	12090	2014-12-03 13	2014-12-03 13:00:00	13	2014-12-03
12256896	91530370	94732789	1	NaN	12090	2014-12-12 23	2014-12-12 23:00:00	23	2014-12-12

Contents

- ▼ 1 案例: 某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
 - ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day
12256897	91530370	330719178	1	NaN	12090	2014-12-15 01	2014-12-15 01:00:00	1	2014-12-15
12256898	91530370	82755420	1	NaN	12553	2014-12-03 00	2014-12-03 00:00:00	0	2014-12-03
12256899	91530370	58196276	1	NaN	12553	2014-11-26 10	2014-11-26 10:00:00	10	2014-11-26
12256900	91530370	101985395	3	NaN	12090	2014-12-12 21	2014-12-12 21:00:00	21	2014-12-12
12256901	93812622	378365755	1	95q6d6a	11	2014-12-13 21	2014-12-13 21:00:00	21	2014-12-13
12256902	93812622	177724753	1	NaN	12311	2014-12-14 21	2014-12-14 21:00:00	21	2014-12-14
12256903	93812622	234391443	1	NaN	8765	2014-12-11 16	2014-12-11 16:00:00	16	2014-12-11
12256904	93812622	26452000	1	95q6dqc	7951	2014-12-08 22	2014-12-08 22:00:00	22	2014-12-08
12256905	108404535	362699797	1	NaN	9847	2014-12-03 19	2014-12-03 19:00:00	19	2014-12-03

12256906 rows × 9 columns

2.6.2 PV分析: 每天访问量

```
In [8]: 1 daily_pv = pdata.groupby('day')['user_id'].count()
        2 daily_pv
```

```
Out[8]: day
2014-11-18    366701
2014-11-19    358823
2014-11-20    353429
2014-11-21    333104
2014-11-22    361355
2014-11-23    382702
2014-11-24    378342
2014-11-25    370239
2014-11-26    360896
2014-11-27    371384
2014-11-28    340638
2014-11-29    364697
2014-11-30    401620
2014-12-01    394611
2014-12-02    405216
2014-12-03    411606
2014-12-04    399952
2014-12-05    361878
2014-12-06    389610
2014-12-07    399751
2014-12-08    386667
2014-12-09    398025
2014-12-10    421910
2014-12-11    488508
2014-12-12    691712
2014-12-13    407160
2014-12-14    402541
2014-12-15    398356
2014-12-16    395085
2014-12-17    384791
2014-12-18    375597
Name: user_id, dtype: int64
```

2.6.3 UV分析: 每天独立用户数

```
In [9]: 1 tmp = pdata.groupby('day')['user_id']
        2 #每个分组去重, 并统计数量
        3 daily_uv = tmp.apply(lambda x:x.drop_duplicates().count())
```

```
In [10]: 1 daily_uv
```

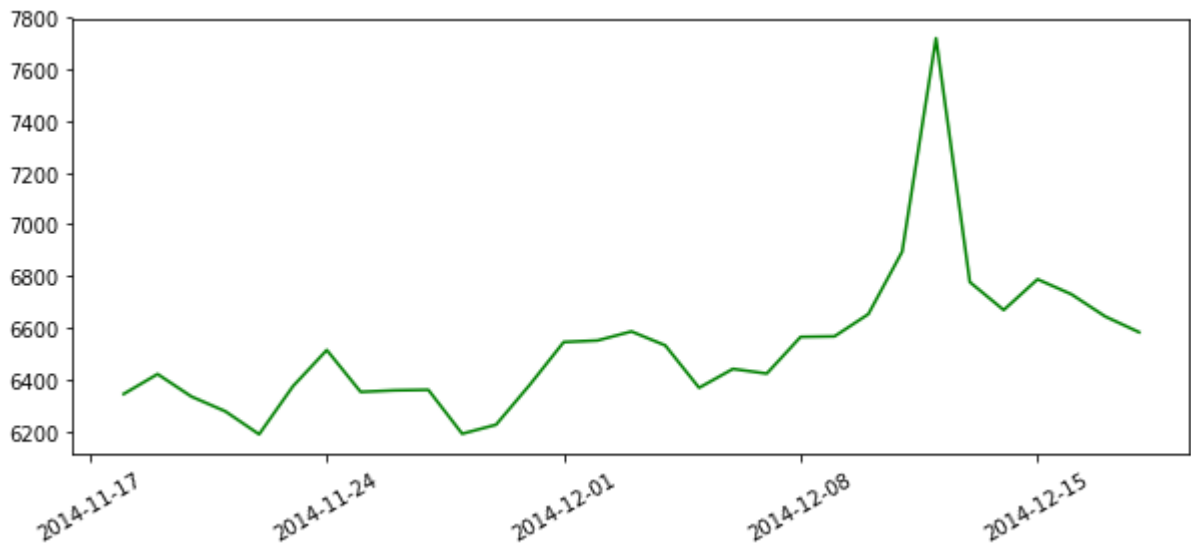
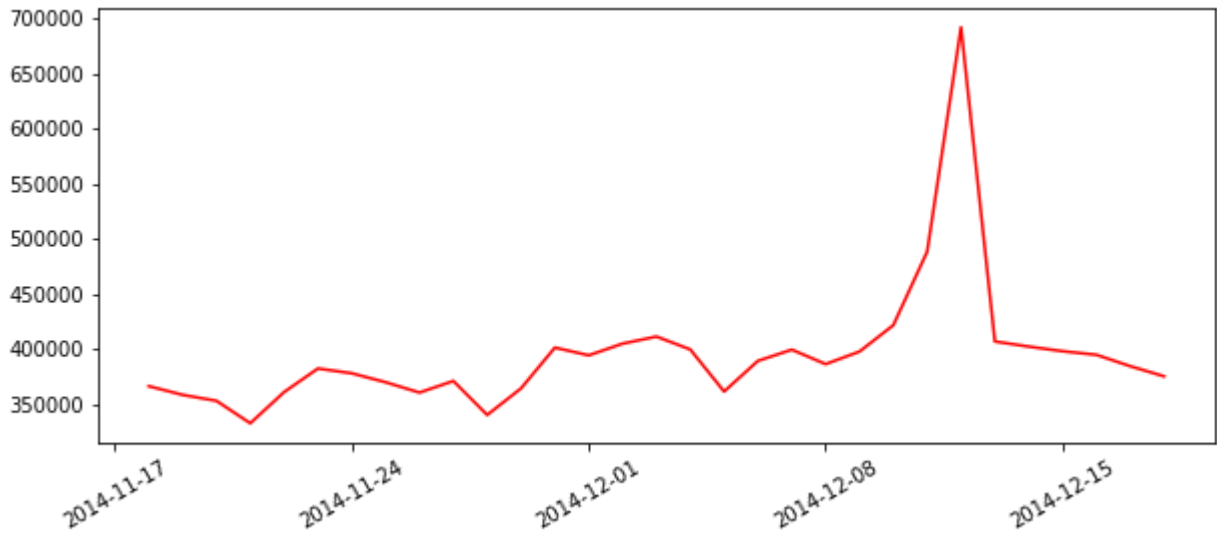
```
Out[10]: day
2014-11-18    6343
2014-11-19    6420
2014-11-20    6333
2014-11-21    6276
2014-11-22    6187
2014-11-23    6373
2014-11-24    6513
2014-11-25    6351
2014-11-26    6357
2014-11-27    6359
2014-11-28    6189
2014-11-29    6224
2014-11-30    6379
2014-12-01    6544
2014-12-02    6550
2014-12-03    6585
2014-12-04    6531
2014-12-05    6367
2014-12-06    6440
2014-12-07    6422
2014-12-08    6564
2014-12-09    6566
2014-12-10    6652
2014-12-11    6894
2014-12-12    7720
2014-12-13    6776
2014-12-14    6668
2014-12-15    6787
2014-12-16    6729
2014-12-17    6643
2014-12-18    6582
Name: user_id, dtype: int64
```

2.6.4 PV与UV的数据可视化

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [11]: 1 import matplotlib.pyplot as plt
2 %matplotlib inline
3 plt.figure(figsize=(10,4))
4 plt.plot(daily_pv.index.values,daily_pv.values, 'r')
5 plt.xticks(rotation=30)
6 plt.figure(figsize=(10,4))
7 plt.plot(daily_uv.index.values,daily_uv.values, 'g')
8 plt.xticks(rotation=30)
9 plt.show()
```



2.6.5 用户购买次数分布

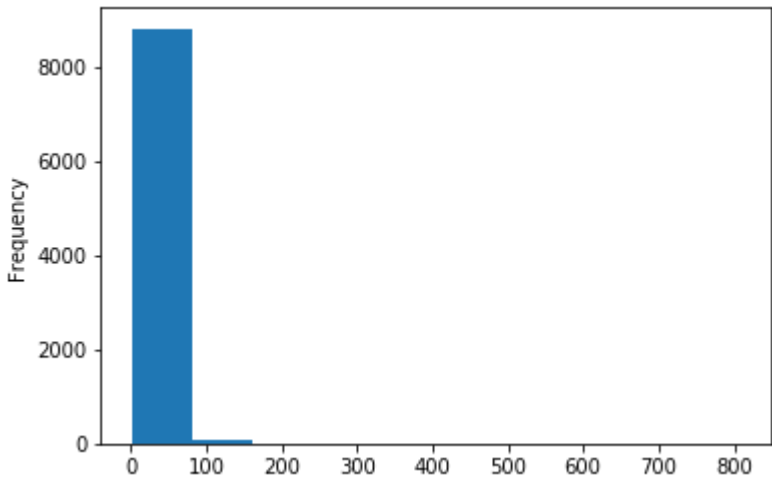
- 这一期间每个用户支付次数
- 统计每个用户支付次数
- 使用直方图展示

步骤:

- 过滤数据: 行为为支付
- 分组: 根据user_id分组.groupby
- 统计: 使用count方法分组
- 可视化: 使用hist方法直方图显示

```
In [12]: 1 user_bynums=pdata[pdata.behavior_type==4].groupby('user_id')['behavior_type'].count()
2 user_bynums.plot.hist()
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x19b3bd18e48>



通过图表发现, 用户支付数量基本在10次左右

2.6.6 用户复购

复购: 用户两天以上有购买行为

复购率: 复购行为用户数/有购买行为的用户总数

思考: 如何统计一个用户, 在不同day发生购物行为?

2.6.6.1 实现方式1

- 使用user_id与day分组, 并统计数量, 目的: 实现user_id与day的去重
- 重新根据user_id分组, 并统计数量

- 根据用户与时间分组,并查看数据

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [13]: 1 redata =pdata[pdata.behavior_type==4].groupby(['user_id', 'day']).count()  
2 redata
```

Out[13]:

		item_id	behavior_type	user_geohash	item_category	time	ts	hour
	user_id	day						
		2014-12-01	1	1	0	1	1	1
		2014-12-07	2	2	2	2	2	2
	4913	2014-12-11	1	1	0	1	1	1
		2014-12-13	1	1	0	1	1	1
		2014-12-16	1	1	0	1	1	1
	6118	2014-12-17	1	1	0	1	1	1
		2014-11-22	1	1	0	1	1	1
		2014-11-26	1	1	0	1	1	1
		2014-11-27	1	1	0	1	1	1
	7528	2014-11-30	1	1	0	1	1	1
		2014-12-03	1	1	0	1	1	1
		2014-12-13	1	1	0	1	1	1
		2014-11-25	1	1	1	1	1	1
		2014-12-02	1	1	1	1	1	1
		2014-12-03	1	1	1	1	1	1
		2014-12-04	3	3	3	3	3	3
	7591	2014-12-06	2	2	2	2	2	2
		2014-12-10	4	4	4	4	4	4
		2014-12-11	3	3	3	3	3	3
		2014-12-12	4	4	0	4	4	4
		2014-12-13	2	2	2	2	2	2
		2014-11-27	4	4	4	4	4	4
	12645	2014-11-29	1	1	1	1	1	1
		2014-12-03	1	1	1	1	1	1
		2014-12-14	2	2	2	2	2	2
	54056	2014-11-18	1	1	0	1	1	1
		2014-12-07	1	1	0	1	1	1
	63348	2014-12-11	1	1	0	1	1	1
	79824	2014-11-18	2	2	0	2	2	2
		2014-11-27	1	1	0	1	1	1

	142368840	2014-12-10	1	1	0	1	1	1
		2014-12-16	1	1	0	1	1	1
	142376113	2014-12-08	1	1	0	1	1	1
		2014-11-27	3	3	3	3	3	3
		2014-11-30	1	1	0	1	1	1
		2014-12-02	2	2	0	2	2	2
	142412247	2014-12-04	2	2	1	2	2	2
		2014-12-10	1	1	1	1	1	1
		2014-12-11	1	1	0	1	1	1
		2014-12-15	2	2	2	2	2	2
		2014-11-29	1	1	0	1	1	1
		2014-12-07	1	1	0	1	1	1
	142430177	2014-12-12	1	1	1	1	1	1
		2014-12-14	1	1	0	1	1	1
		2014-12-18	1	1	1	1	1	1
		2014-11-24	14	14	0	14	14	14
		2014-11-27	2	2	0	2	2	2
		2014-12-01	2	2	2	2	2	2
	142450275	2014-12-04	3	3	3	3	3	3
		2014-12-05	2	2	0	2	2	2
		2014-12-08	4	4	0	4	4	4
		2014-12-12	12	12	9	12	12	12
		2014-12-13	1	1	1	1	1	1
		2014-11-21	5	5	0	5	5	5
		2014-11-22	1	1	0	1	1	1
		2014-11-24	1	1	0	1	1	1
	142455899	2014-11-26	2	2	0	2	2	2
		2014-11-30	1	1	0	1	1	1
		2014-12-03	1	1	0	1	1	1
		2014-12-04	2	2	0	2	2	2

49201 rows × 7 columns

- 重置索引

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [14]: 1 redata = redata.reset_index()
        2 redata
```

Out [14]:

	user_id	day	item_id	behavior_type	user_geohash	item_category	time	ts	hour
	0	4913	2014-12-01	1	1	0	1	1	1
	1	4913	2014-12-07	2	2	2	2	2	2
	2	4913	2014-12-11	1	1	0	1	1	1
	3	4913	2014-12-13	1	1	0	1	1	1
	4	4913	2014-12-16	1	1	0	1	1	1
	5	6118	2014-12-17	1	1	0	1	1	1
	6	7528	2014-11-22	1	1	0	1	1	1
	7	7528	2014-11-26	1	1	0	1	1	1
	8	7528	2014-11-27	1	1	0	1	1	1
	9	7528	2014-11-30	1	1	0	1	1	1
	10	7528	2014-12-03	1	1	0	1	1	1
	11	7528	2014-12-13	1	1	0	1	1	1
	12	7591	2014-11-25	1	1	1	1	1	1
	13	7591	2014-12-02	1	1	1	1	1	1
	14	7591	2014-12-03	1	1	1	1	1	1
	15	7591	2014-12-04	3	3	3	3	3	3
	16	7591	2014-12-06	2	2	2	2	2	2
	17	7591	2014-12-10	4	4	4	4	4	4
	18	7591	2014-12-11	3	3	3	3	3	3
	19	7591	2014-12-12	4	4	0	4	4	4
	20	7591	2014-12-13	2	2	2	2	2	2
	21	12645	2014-11-27	4	4	4	4	4	4
	22	12645	2014-11-29	1	1	1	1	1	1
	23	12645	2014-12-03	1	1	1	1	1	1
	24	12645	2014-12-14	2	2	2	2	2	2
	25	54056	2014-11-18	1	1	0	1	1	1
	26	54056	2014-12-07	1	1	0	1	1	1
	27	63348	2014-12-11	1	1	0	1	1	1
	28	79824	2014-11-18	2	2	0	2	2	2
	29	79824	2014-11-27	1	1	0	1	1	1

	49171	142368840	2014-12-10	1	1	0	1	1	1
	49172	142368840	2014-12-16	1	1	0	1	1	1
	49173	142376113	2014-12-08	1	1	0	1	1	1
	49174	142412247	2014-11-27	3	3	3	3	3	3
	49175	142412247	2014-11-30	1	1	0	1	1	1
	49176	142412247	2014-12-02	2	2	0	2	2	2
	49177	142412247	2014-12-04	2	2	1	2	2	2
	49178	142412247	2014-12-10	1	1	1	1	1	1
	49179	142412247	2014-12-11	1	1	0	1	1	1
	49180	142412247	2014-12-15	2	2	2	2	2	2
	49181	142430177	2014-11-29	1	1	0	1	1	1
	49182	142430177	2014-12-07	1	1	0	1	1	1
	49183	142430177	2014-12-12	1	1	1	1	1	1
	49184	142430177	2014-12-14	1	1	0	1	1	1
	49185	142430177	2014-12-18	1	1	1	1	1	1
	49186	142450275	2014-11-24	14	14	0	14	14	14
	49187	142450275	2014-11-27	2	2	0	2	2	2
	49188	142450275	2014-12-01	2	2	2	2	2	2
	49189	142450275	2014-12-04	3	3	3	3	3	3
	49190	142450275	2014-12-05	2	2	0	2	2	2
	49191	142450275	2014-12-08	4	4	0	4	4	4
	49192	142450275	2014-12-12	12	12	9	12	12	12
	49193	142450275	2014-12-13	1	1	1	1	1	1
	49194	142455899	2014-11-21	5	5	0	5	5	5
	49195	142455899	2014-11-22	1	1	0	1	1	1
	49196	142455899	2014-11-24	1	1	0	1	1	1
	49197	142455899	2014-11-26	2	2	0	2	2	2
	49198	142455899	2014-11-30	1	1	0	1	1	1
	49199	142455899	2014-12-03	1	1	0	1	1	1
	49200	142455899	2014-12-04	2	2	0	2	2	2

49201 rows × 9 columns

- 根据user_id分组, 并统计day的数量

Contents

- 1 案例: 某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [16]:      1  ret = redata.groupby('user_id')['day'].count()
           2  ret

Out[16]: user_id
4913      5
6118      1
7528      6
7591      9
12645     4
54056     2
63348     1
79824     7
88930    10
100539    12
104155     3
109103    11
113251     4
113960    12
120873     3
134658     2
151617    14
156608     4
157097     9
189833     8
190327     5
191366     2
213655    10
217996     1
227293     3
230711     7
239485     5
247543     9
250843     9
263670     1
..
142060721  2
142061900  2
142064691  6
142073213  1
142081324  9
142095821  8
142120051 21
142126535  5
142128942  9
142128951  9
142138619  4
142144275  6
142151675  9
142168798  8
142181816  5
142204924  2
142216376  9
142227202  9
142244794  5
142265405  7
142306250  5
142306361  6
142337230  3
142358910  2
142368840  6
142376113  1
142412247  7
142430177  5
142450275  8
142455899  7
Name: day, Length: 8886, dtype: int64
```

- 获取数量大于1的数据

```
In [17]:      1  ret[ret>1].count()/ret.count()

Out[17]: 0.87170830519918974
```

2.6.6.2 实现方式2

- 根据user_id分组
- 对day进行去重, 并统计数量
- 计算

```
In [26]:      1  re_data=pdata[pdata.behavior_type==4].groupby('user_id')
```

```
In [27]:      1  re_data

Out[27]: <pandas.core.groupby.DataFrameGroupBy object at 0x0000019B3C129400>
```

```
In [ ]:      1
```

- 使用apply方法对分组数据进行处理,
- 每个apply方法处理的user_id对应的series对象
- 使用unique获取唯一值
- 使用len方法获取长度

```
In [28]:      1  re_data = re_data['day'].apply(lambda x:len(x.unique()))
```


Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [21]:      1  re_data

Out[21]:      user_id
4913          5
6118          1
7528          6
7591          9
12645         4
54056         2
63348         1
79824         7
88930        10
100539       12
104155        3
109103       11
113251        4
113960       12
120873        3
134658        2
151617       14
156608        4
157097        9
189833        8
190327        5
191366        2
213655       10
217996        1
227293        3
230711        7
239485        5
247543        9
250843        9
263670        1
...
142060721     2
142061900     2
142064691     6
142073213     1
142081324     9
142095821     8
142120051    21
142126535     5
142128942     9
142128951     9
142138619     4
142144275     6
142151675     9
142168798     8
142181816     5
142204924     2
142216376     9
142227202     9
142244794     5
142265405     7
142306250     5
142306361     6
142337230     3
142358910     2
142368840     6
142376113     1
142412247     7
142430177     5
142450275     8
142455899     7
Name: day, Length: 8886, dtype: int64
```

- 计算复购率

```
In [30]:      1  re_data[re_data>1].count()/re_data.count()

Out[30]:      0.87170830519918974
```

用户复购率为0.8717左右

2.6.7 用户复购时间间隔

复购间隔:

- 按时间排序
- 用户第1次购物时间: day_1
- 用户第2次购物时间: day_2 , 间隔: $day_2 - day_1$
- 用户第3次购物时间: day_3 , 间隔: $day_3 - day_2$
- 用户第N次购物时间: day_n , 间隔: $day_n - day_{n-1}$
- 要求: 时间间隔大于0

问题: 如何根据day, 获取时间间隔

2.6.7.1 diff方法

后一个减去前一个

```
In [54]:      1  p = pd.Series([1,2,7,4,10])
           2  p.diff(1)

Out[54]:      0  NaN
           1  1.0
           2  5.0
           3  -3.0
           4  6.0
dtype: float64

In [53]:      1  #删除缺失省值
           2  p.diff(2).dropna()

Out[53]:      2  6.0
           3  2.0
           4  3.0
dtype: float64
```

2.6.7.2 实现

- 获取支付数据
- 根据用户进行分组
- 计算时间间隔

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [55]: 1 #分组
        2 userbuy=pdata[pdata.behavior_type==4].groupby('user_id')
```

```
In [56]: 1 #对天进行处理
        2 #lambda的每个x为分组后的series对象
        3 # series对象排序, 并使用diff方法计算, 并去除缺省值
        4 day_interval = userbuy.day.apply(lambda x:x.sort_values().diff(1).dropna())
```

```
In [ ]: 1
```

- 问题: day_interval中的value是什么值?

```
In [58]: 1 t = day_interval[0]
        2 t
```

Out[58]: Timedelta('6 days 00:00:00')

```
In [59]: 1 #获取天数
        2 t.days
```

Out[59]: 6

- 获取时间间隔

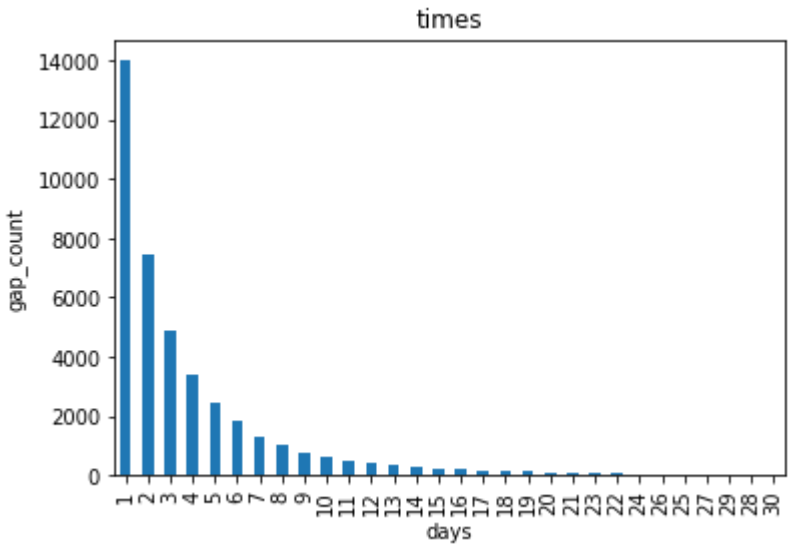
```
In [60]: 1 day_interval=day_interval.map(lambda x:x.days)
```

```
In [ ]: 1
```

```
In [62]: 1 day_interval = day_interval[day_interval>0]
```

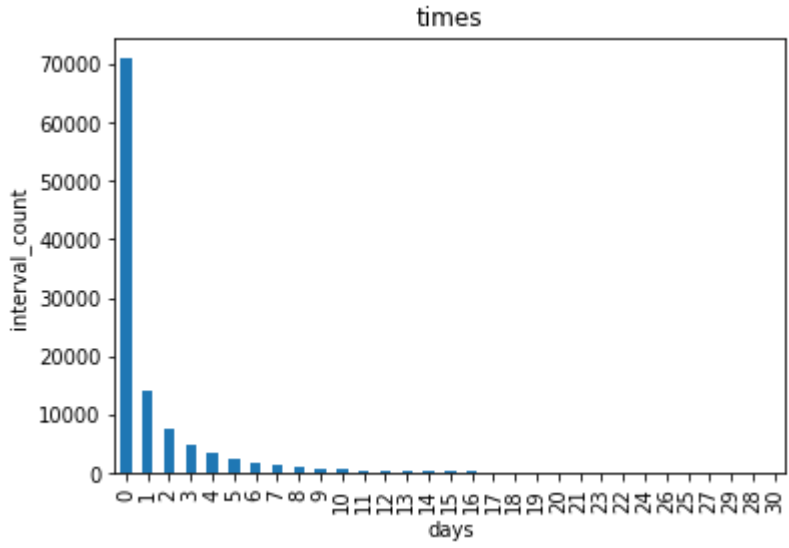
```
In [ ]: 1
```

```
In [66]: 1 #统计每个时间数量, 并绘制柱状图
        2 day_interval.value_counts().plot(kind='bar')
        3 plt.title('times')
        4 plt.xlabel('days')
        5 _ = plt.ylabel('gap_count')
```



- 最终实现

```
In [68]: 1 #用户购物行为及分组
        2 #时间间隔大于1天
        3 day_interval = pdata[pdata.behavior_type == 4].groupby('user_id').day.apply(lambda x:x.sort_values().c
        4 #获取天数
        5 day_interval = day_interval.map(lambda x:x.days)
        6 #统计数量, 并绘制柱状图
        7 day_interval.value_counts().plot(kind='bar')
        8 plt.title('times')
        9 plt.xlabel('days')
        10 _ = plt.ylabel('interval_count')
```



2.7 AARRR模型

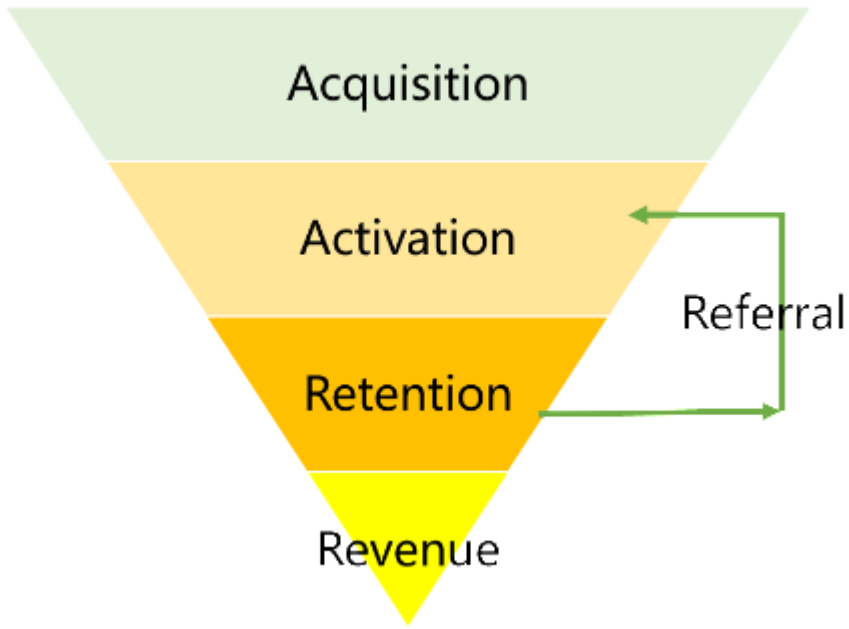
AARRR代表意思:

AARRR	说明
Acquisition	用户获取
Activation	用户激活
Retention	提高留存
Revenue	增加收入
Referral	传播推荐

AARRR模型反应出不同阶段中用户参与行为的深度和类型, 是一个典型的漏斗结构;

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据：某年淘宝双12部分购物数据；
 - 1.2 分析目的：
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标：PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析：每天访问量
 - 2.6.3 UV分析：每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解：
 - 3.2 F理解：
 - 3.3 M理解：
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则：
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R，最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF，根据RF进行分类



AARRR模型描述了用户/客户/访客需经历的五个环节，以便企业获取价值。
关键点：通过模型提高留存和转化率

2.7.1 理解AARRR

- 获取用户：获取用户方式，获取用户渠道，获取用户成本，用户定位等，例如：应用下载量，安装量，注册量
- 活跃度：日活，月活，使用时长，启动次数等
- 留存率：次日留存，周留存率，不同邻域用户，留存周期不同，例如：微博，1周末登录，可以视为流失用户；
- 收入：APA(活跃付费用户数),ARPU(平均每用户收入), ARPPU(平均每付费用户收入)；提高活跃度、提高留存率是增加收入基础。
- 传播推荐：用户自发传播，例如拼多多砍价，核心：产品过硬

2.7.2 双12用户购物模型

- 层级1：用户访问量
- 层级2：收藏
- 层级3：购物车
- 层级4：支付

2.7.3 不同层级数量

```
In [85]: 1 event_type_count = pdata.groupby('behavior_type').size()
2 #重置索引并修改列名
3 event_type_count = event_type_count.reset_index().rename(columns={0:'total'})
4 #重置索引
5 event_type_count = event_type_count.set_index('behavior_type')
6 event_type_count
```

Out[85]:

	total
behavior_type	
1	11550581
2	242556
3	343564
4	120205

2.7.4 计算转化率

- 将层级1视为基准

```
In [86]: 1 event_type_count
```

Out[86]:

	total
behavior_type	
1	11550581
2	242556
3	343564
4	120205

```
In [87]: 1 event_type_count['pre'] = (event_type_count.total/event_type_count.total[1])*100
2 event_type_count
```

Out[87]:

	total	pre
behavior_type		
1	11550581	100.000000
2	242556	2.099946
3	343564	2.974430
4	120205	1.040684

3 用户价值分析

RFM模型是衡量当前用户价值和客户潜在价值的重要工具和手段

- R:最近一次消费（Recency）
- F:消费频率（Frequency）
- M:消费金额（Monetary）

3.1 R理解：

- 用户最近一次消费
- R值越小，说明用户的价值越高，例如：最近一次消费1个月，最近一次消费10个月
- R值非常大，说明该用户可能为流失用户；

3.2 F理解：

- 消费频率，指定时间内购买次数；
- 问题：一些商品购买频次较低，所以会将时间忽略，以购买次数为主；
- 不同品类产品，对F的理解不一样；例如：电子类产品频率较低；化妆品，衣服等消费品购买频次较高

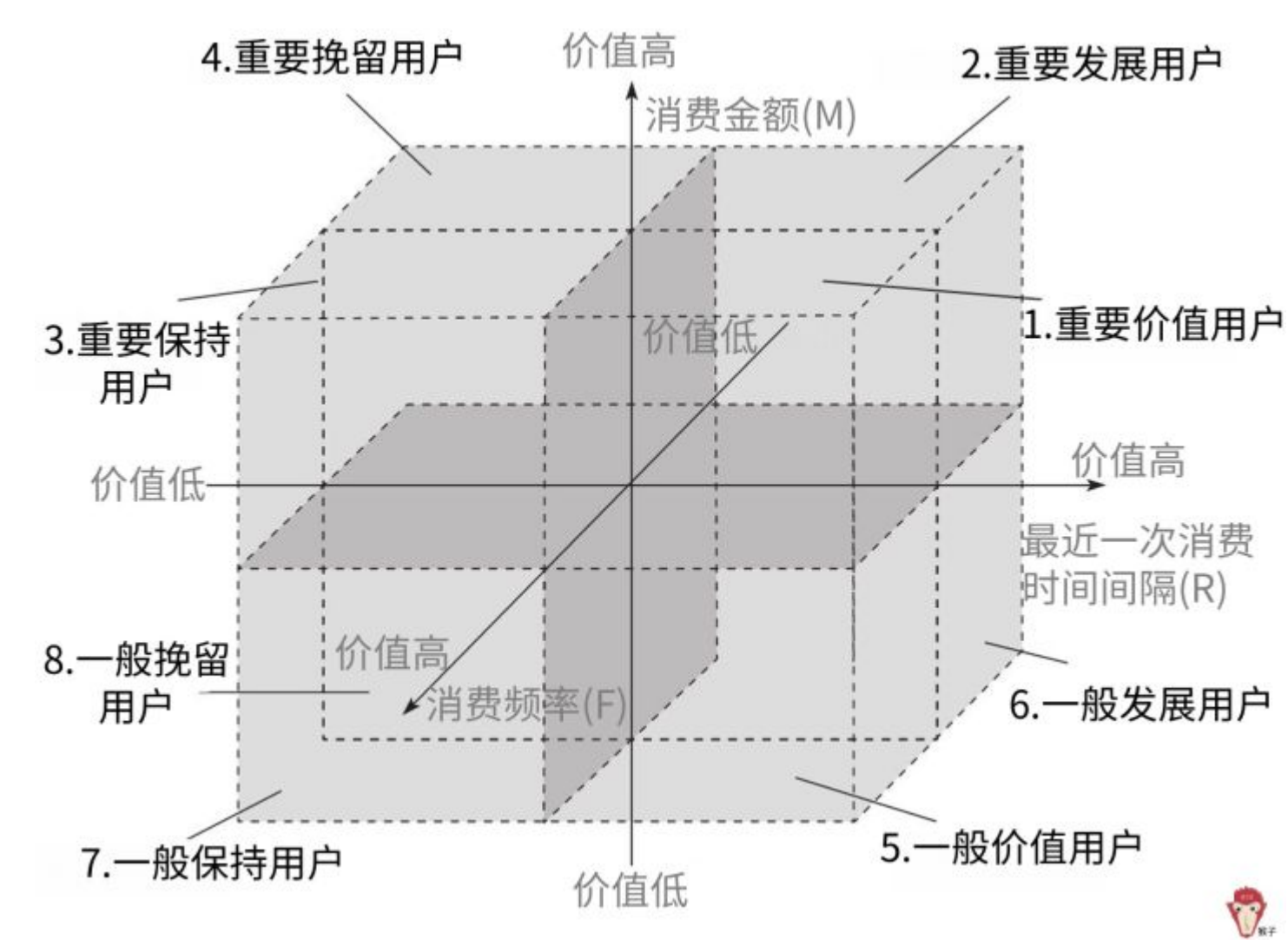
3.3 M理解：

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据：某年淘宝双12部分购物数据；
 - 1.2 分析目的：
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标：PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析：每天访问量
 - 2.6.3 UV分析：每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- ▼ 3 用户价值分析
 - 3.1 R理解：
 - 3.2 F理解：
 - 3.3 M理解：
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则：
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R，最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF，根据RF进行分类

- 一定时间范围内，用户消费金额，消费金额越大，价值越高

3.4 RFM评分模型



3.5 不同指标对应测策略

指标	客户分组	策略	指标分段
R	活跃客户	频繁推送营销信息	一个月未购买
R	沉默用户	推送优惠力度	2个月未购买
R	休眠用户	大型活动促销	3个月未购买
R	流失用户	超大型优惠促销	4个月未购买
F	新客户	促销信息	购买1次
F	老客户	品牌信息	购买2次
F	成熟客户	新品与活动信息	购买3次
F	忠实客户	传递会员活动与权益	购买4次
M	低贡献用户	促销商品与折扣	1/2客单价
M	中贡献用户	促销商品与折扣	1/2客单价~客单价
M	中高贡献用户	形象商品与品牌活动	客单价~2倍客单价
M	高献用户	形象商品与品牌活动	2倍客单价以上

客单价：平均客单价格

3.6 用户分类规则：

用户分类	R	F	M
重要价值用户	高	高	高
重要发展用户	高	低	高
重要保持用户	低	高	高
重要挽留用户	高	高	高
一般价值用户	低	低	高
一般发展用户	高	高	低
一般保持用户	高	低	低
一般挽留用户	低	低	低

3.7 RFM评分规则

- 对于R：时间不同，得分不同，例如：最近1个月10分，最近2个月7分，最近3个月4分...
- 对于F：根据不同次数，设定不同得分
- 对于M：根据不同客单价设置得分
- 分别计算每个维度数值的均值或者中位数，
- 根据每个数据值与每个维度均值或者中位数对比，确认处于哪个象限；

3.8 灵活使用

- 显示场景中，购买频次或者客单价差距不大，可以忽略
- 对于M：商品销售价格一样(例如：虚拟商品等)，可以将一时间累积消费值作为M
- 实际数据中，如果缺少F或者M的值，可以尝试去忽略这个字段

3.9 淘宝双12购物用户RFM模型

3.9.1 整理数据

- 数据中没有商品价格，所以可以忽略该字段
- 根据数据整理出每个用户消费R
- 根据数据整理出每个用户消费F

3.9.2 计算R，最近购物时间

以结束时间为2014-12-20截止，问题：获取用户最后购物日期？

- 按照时间排序
- 获取最后一个值
- 如何验证数据是正确的

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据：某年淘宝双12部分购物数据；
 - 1.2 分析目的：
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标：PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析：每天访问量
 - 2.6.3 UV分析：每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
 - ▼ 3 用户价值分析
 - 3.1 R理解：
 - 3.2 F理解：
 - 3.3 M理解：
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则：
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R，最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF，根据RF进行分类

In [8]:

```
1 #排序
2 tmp = pdata.sort_values('day')
3 #过滤数据，只保留购物的数据
4 tmp = tmp[tmp['behavior_type']==4]
5 #保留最后一个值，生成bool索引,删除最后一个，对应值为Flase
6 tindex = tmp.duplicated(subset=['user_id'], keep='last')
7 #boolean索引变化，将False设置为True
8 sindex = tindex == False
9 #获取过滤后数据
10 tdata = tmp[sindex]
```

In [10]:

1	tdata
---	-------

Out[10]:

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day	
	6946622	49598175	176268691	4	97rjjab	11270	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18
	7917659	141878326	379814689	4	NaN	8095	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18
	7549224	61454609	134625812	4	99ui5e6	12067	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18
	10056408	86985041	105821300	4	NaN	6669	2014-11-18 10	2014-11-18 10:00:00	10	2014-11-18
	1961858	112485959	228534873	4	NaN	292	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18
	11346671	71238931	342295739	4	NaN	9720	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18
	11505802	134020041	253326728	4	NaN	8952	2014-11-18 16	2014-11-18 16:00:00	16	2014-11-18
	4561621	36164872	61293574	4	95bbigf	8099	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18
	659375	141792132	230991857	4	NaN	11914	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18
	9392943	104639948	69768782	4	94olrb4	6576	2014-11-18 20	2014-11-18 20:00:00	20	2014-11-18
	10162132	113445602	389638875	4	NaN	1709	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18
	10212711	95937118	61587564	4	NaN	3472	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18
	1612218	125817021	160186502	4	NaN	2284	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18
	723952	21226209	314041390	4	NaN	3960	2014-11-18 00	2014-11-18 00:00:00	0	2014-11-18
	2557048	122694692	23948496	4	NaN	11115	2014-11-18 23	2014-11-18 23:00:00	23	2014-11-18
	4584854	44048864	200504840	4	NaN	7322	2014-11-18 22	2014-11-18 22:00:00	22	2014-11-18
	6656377	45149630	221677852	4	NaN	9550	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18
	11106545	21754104	12785701	4	NaN	11914	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18
	6451632	28160359	152061953	4	NaN	3660	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18
	3579185	10797721	381491297	4	NaN	3628	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18
	3589838	6596569	279280503	4	NaN	11311	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18
	4946491	56960895	260291631	4	9rf95au	8254	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18
	3170441	85947802	114836496	4	ffwqqbr	4468	2014-11-18 07	2014-11-18 07:00:00	7	2014-11-18
	8514716	122876959	241021997	4	NaN	12067	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18
	4994431	95302160	280054537	4	mtj03d9	5468	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18
	8420014	17785981	156043735	4	NaN	6648	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18
	2083727	130457432	348533843	4	NaN	5599	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18
	5885538	98621566	39051316	4	956o9ug	4582	2014-11-18 12	2014-11-18 12:00:00	12	2014-11-18
	2956807	119112558	222924385	4	NaN	3424	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18
	3414727	139544053	180741456	4	94h6efu	7850	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18

	4927060	139104982	71992019	4	mtdasrg	10703	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18
	1001705	82838416	266422041	4	NaN	2251	2014-12-18 12	2014-12-18 12:00:00	12	2014-12-18
	4925906	130132447	277293784	4	NaN	6936	2014-12-18 07	2014-12-18 07:00:00	7	2014-12-18
	8613649	88120786	402483437	4	NaN	231	2014-12-18 09	2014-12-18 09:00:00	9	2014-12-18
	11127661	31446692	23709809	4	NaN	2754	2014-12-18 16	2014-12-18 16:00:00	16	2014-12-18
	999440	135803085	340015023	4	NaN	7549	2014-12-18 07	2014-12-18 07:00:00	7	2014-12-18
	6704431	151617	343739381	4	9rrsiqf	8749	2014-12-18 16	2014-12-18 16:00:00	16	2014-12-18
	3238355	42966250	360502621	4	NaN	5334	2014-12-18 01	2014-12-18 01:00:00	1	2014-12-18
	6704324	54492776	323332967	4	NaN	2762	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18
	11130091	5686843	225727745	4	NaN	5454	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18
	8605563	45796530	247114107	4	ffqhkbk	12553	2014-12-18 15	2014-12-18 15:00:00	15	2014-12-18
	6696453	93687523	239333787	4	NaN	5214	2014-12-18 13	2014-12-18 13:00:00	13	2014-12-18
	4933061	1805740	306159400	4	94ebusu	231	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18
	6696192	137129651	96160664	4	97t6pq5	10431	2014-12-18 22	2014-12-18 22:00:00	22	2014-12-18
	8627624	83851924	222570726	4	95c6gr2	8986	2014-12-18 20	2014-12-18 20:00:00	20	2014-12-18
	8627447	127552225	176709504	4	NaN	4676	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18
	8627952	3694961	296462974	4	NaN	9130	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18
	317321	38968501	253930616	4	NaN	9517	2014-12-18 06	2014-12-18 06:00:00	6	2014-12-18
	6694445	867381	154354524	4	NaN	10894	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18
	8624702	109048285	136615762	4	95fwge1	7712	2014-12-18 20	2014-12-18 20:00:00	20	2014-12-18
	3220691	128180595	397452112	4	NaN	4333	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day
3221482	131254719	192958143	4	95ll4ir	10141	2014-12-18 14	2014-12-18 14:00:00	14	2014-12-18
8617075	128733919	87278584	4	99cte0p	9053	2014-12-18 09	2014-12-18 09:00:00	9	2014-12-18
3230182	124836406	355540084	4	NaN	7229	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18
1004797	104777513	293440996	4	NaN	2425	2014-12-18 12	2014-12-18 12:00:00	12	2014-12-18
319576	131514258	281172663	4	NaN	1514	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18
11888591	123848007	106982400	4	956k2js	11750	2014-12-18 14	2014-12-18 14:00:00	14	2014-12-18
3227167	49871598	47827433	4	96kajt4	2631	2014-12-18 11	2014-12-18 11:00:00	11	2014-12-18
6696914	136470861	347836659	4	9rf42od	10472	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18
8618913	81958592	332303278	4	9r5fri3	13288	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18

8886 rows × 9 columns

3.9.3 数据验证

- 数量验证

```
In [11]: 1 #数量
         2 len(tmp.groupby('user_id'))
```

Out[11]: 8886

- 指定用户验证, 检查索引值

```
In [12]: 1 tmp[tmp['user_id'] == 81958592]
```

Out[12]:

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day
5479627	81958592	16646074	4	NaN	1843	2014-11-18 10	2014-11-18 10:00:00	10	2014-11-18
3098208	81958592	56354270	4	NaN	2993	2014-11-20 17	2014-11-20 17:00:00	17	2014-11-20
3951826	81958592	332303278	4	NaN	13288	2014-11-21 09	2014-11-21 09:00:00	9	2014-11-21
8024873	81958592	147112095	4	NaN	1956	2014-11-25 06	2014-11-25 06:00:00	6	2014-11-25
10528858	81958592	704791	4	NaN	10079	2014-11-26 22	2014-11-26 22:00:00	22	2014-11-26
2097335	81958592	342323669	4	9r5frdt	10703	2014-11-28 19	2014-11-28 19:00:00	19	2014-11-28
571995	81958592	365660750	4	9r5f70o	10412	2014-11-30 10	2014-11-30 10:00:00	10	2014-11-30
6431696	81958592	308243804	4	NaN	6513	2014-12-02 09	2014-12-02 09:00:00	9	2014-12-02
3951706	81958592	27762745	4	NaN	6253	2014-12-07 12	2014-12-07 12:00:00	12	2014-12-07
3951954	81958592	126523128	4	9r5f7tu	9372	2014-12-07 21	2014-12-07 21:00:00	21	2014-12-07
3951770	81958592	239342258	4	9r5f7ut	6648	2014-12-07 12	2014-12-07 12:00:00	12	2014-12-07
3098384	81958592	379358581	4	NaN	6677	2014-12-07 11	2014-12-07 11:00:00	11	2014-12-07
10528795	81958592	370332314	4	NaN	9516	2014-12-12 08	2014-12-12 08:00:00	8	2014-12-12
3951802	81958592	332303278	4	9r5frog	13288	2014-12-12 20	2014-12-12 20:00:00	20	2014-12-12
5479755	81958592	317573283	4	NaN	9856	2014-12-12 08	2014-12-12 08:00:00	8	2014-12-12
10529075	81958592	232532440	4	9r5f7t9	3289	2014-12-17 20	2014-12-17 20:00:00	20	2014-12-17
8618913	81958592	332303278	4	9r5frl3	13288	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18

3.9.4 计算R值

Contents

- ▼ 1 案例: 某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [13]: 1 pd.to_datetime(tdata[' day'])

Out[13]: 6946622      2014-11-18
7917659      2014-11-18
7549224      2014-11-18
10056408     2014-11-18
1961858      2014-11-18
11346671     2014-11-18
11505802     2014-11-18
4561621      2014-11-18
659375       2014-11-18
9392943      2014-11-18
10162132     2014-11-18
10212711     2014-11-18
1612218      2014-11-18
723952       2014-11-18
2557048      2014-11-18
4584854      2014-11-18
6656377      2014-11-18
11106545     2014-11-18
6451632      2014-11-18
3579185      2014-11-18
3589838      2014-11-18
4946491      2014-11-18
3170441      2014-11-18
8514716      2014-11-18
4994431      2014-11-18
8420014      2014-11-18
2083727      2014-11-18
5885538      2014-11-18
2956807      2014-11-18
3414727      2014-11-18
...
4927060      2014-12-18
1001705      2014-12-18
4925906      2014-12-18
8613649      2014-12-18
11127661     2014-12-18
999440       2014-12-18
6704431      2014-12-18
3238355      2014-12-18
6704324      2014-12-18
11130091     2014-12-18
8605563      2014-12-18
6696453      2014-12-18
4933061      2014-12-18
6696192      2014-12-18
8627624      2014-12-18
8627447      2014-12-18
8627952      2014-12-18
317321       2014-12-18
6694445      2014-12-18
8624702      2014-12-18
3220691      2014-12-18
3221482      2014-12-18
8617075      2014-12-18
3230182      2014-12-18
1004797      2014-12-18
319576       2014-12-18
11888591     2014-12-18
3227167      2014-12-18
6696914      2014-12-18
8618913      2014-12-18
Name: day, Length: 8886, dtype: datetime64[ns]
```


Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [14]: 1 tdl = pd.to_datetime('2014-12-20') - pd.to_datetime(tdata['day'])
2 tdata['tdays'] = tdl.dt.days
3 tdata
```

d:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy> (<http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>)

Out[14]:		user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day	tdays
	6946622	49598175	176268691	4	97rjjab	11270	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18	32
	7917659	141878326	379814689	4	NaN	8095	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18	32
	7549224	61454609	134625812	4	99ui5e6	12067	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18	32
	10056408	86985041	105821300	4	NaN	6669	2014-11-18 10	2014-11-18 10:00:00	10	2014-11-18	32
	1961858	112485959	228534873	4	NaN	292	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18	32
	11346671	71238931	342295739	4	NaN	9720	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18	32
	11505802	134020041	253326728	4	NaN	8952	2014-11-18 16	2014-11-18 16:00:00	16	2014-11-18	32
	4561621	36164872	61293574	4	95bbigf	8099	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18	32
	659375	141792132	230991857	4	NaN	11914	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18	32
	9392943	104639948	69768782	4	94olrb4	6576	2014-11-18 20	2014-11-18 20:00:00	20	2014-11-18	32
	10162132	113445602	389638875	4	NaN	1709	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18	32
	10212711	95937118	61587564	4	NaN	3472	2014-11-18 14	2014-11-18 14:00:00	14	2014-11-18	32
	1612218	125817021	160186502	4	NaN	2284	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18	32
	723952	21226209	314041390	4	NaN	3960	2014-11-18 00	2014-11-18 00:00:00	0	2014-11-18	32
	2557048	122694692	23948496	4	NaN	11115	2014-11-18 23	2014-11-18 23:00:00	23	2014-11-18	32
	4584854	44048864	200504840	4	NaN	7322	2014-11-18 22	2014-11-18 22:00:00	22	2014-11-18	32
	6656377	45149630	221677852	4	NaN	9550	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18	32
	11106545	21754104	12785701	4	NaN	11914	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18	32
	6451632	28160359	152061953	4	NaN	3660	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18	32
	3579185	10797721	381491297	4	NaN	3628	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18	32
	3589838	6596569	279280503	4	NaN	11311	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18	32
	4946491	56960895	260291631	4	9rf95au	8254	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18	32
	3170441	85947802	114836496	4	ffwqqbr	4468	2014-11-18 07	2014-11-18 07:00:00	7	2014-11-18	32
	8514716	122876959	241021997	4	NaN	12067	2014-11-18 19	2014-11-18 19:00:00	19	2014-11-18	32
	4994431	95302160	280054537	4	mtj03d9	5468	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18	32
	8420014	17785981	156043735	4	NaN	6648	2014-11-18 11	2014-11-18 11:00:00	11	2014-11-18	32
	2083727	130457432	348533843	4	NaN	5599	2014-11-18 18	2014-11-18 18:00:00	18	2014-11-18	32
	5885538	98621566	39051316	4	956o9ug	4582	2014-11-18 12	2014-11-18 12:00:00	12	2014-11-18	32
	2956807	119112558	222924385	4	NaN	3424	2014-11-18 13	2014-11-18 13:00:00	13	2014-11-18	32
	3414727	139544053	180741456	4	94h6efu	7850	2014-11-18 21	2014-11-18 21:00:00	21	2014-11-18	32

	4927060	139104982	71992019	4	mtdasrg	10703	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18	2
	1001705	82838416	266422041	4	NaN	2251	2014-12-18 12	2014-12-18 12:00:00	12	2014-12-18	2
	4925906	130132447	277293784	4	NaN	6936	2014-12-18 07	2014-12-18 07:00:00	7	2014-12-18	2
	8613649	88120786	402483437	4	NaN	231	2014-12-18 09	2014-12-18 09:00:00	9	2014-12-18	2
	11127661	31446692	23709809	4	NaN	2754	2014-12-18 16	2014-12-18 16:00:00	16	2014-12-18	2

Contents

- 1 案例: 某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

	user_id	item_id	behavior_type	user_geohash	item_category	time	ts	hour	day	tdays
999440	135803085	340015023	4	NaN	7549	2014-12-18 07	2014-12-18 07:00:00	7	2014-12-18	2
6704431	151617	343739381	4	9rrsiqf	8749	2014-12-18 16	2014-12-18 16:00:00	16	2014-12-18	2
3238355	42966250	360502621	4	NaN	5334	2014-12-18 01	2014-12-18 01:00:00	1	2014-12-18	2
6704324	54492776	323332967	4	NaN	2762	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18	2
11130091	5686843	225727745	4	NaN	5454	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18	2
8605563	45796530	247114107	4	ffqhkbk	12553	2014-12-18 15	2014-12-18 15:00:00	15	2014-12-18	2
6696453	93687523	239333787	4	NaN	5214	2014-12-18 13	2014-12-18 13:00:00	13	2014-12-18	2
4933061	1805740	306159400	4	94ebusu	231	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18	2
6696192	137129651	96160664	4	97t6pq5	10431	2014-12-18 22	2014-12-18 22:00:00	22	2014-12-18	2
8627624	83851924	222570726	4	95c6gr2	8986	2014-12-18 20	2014-12-18 20:00:00	20	2014-12-18	2
8627447	127552225	176709504	4	NaN	4676	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18	2
8627952	3694961	296462974	4	NaN	9130	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18	2
317321	38968501	253930616	4	NaN	9517	2014-12-18 06	2014-12-18 06:00:00	6	2014-12-18	2
6694445	867381	154354524	4	NaN	10894	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18	2
8624702	109048285	136615762	4	95fwge1	7712	2014-12-18 20	2014-12-18 20:00:00	20	2014-12-18	2
3220691	128180595	397452112	4	NaN	4333	2014-12-18 21	2014-12-18 21:00:00	21	2014-12-18	2
3221482	131254719	192958143	4	95ll4ir	10141	2014-12-18 14	2014-12-18 14:00:00	14	2014-12-18	2
8617075	128733919	87278584	4	99cte0p	9053	2014-12-18 09	2014-12-18 09:00:00	9	2014-12-18	2
3230182	124836406	355540084	4	NaN	7229	2014-12-18 23	2014-12-18 23:00:00	23	2014-12-18	2
1004797	104777513	293440996	4	NaN	2425	2014-12-18 12	2014-12-18 12:00:00	12	2014-12-18	2
319576	131514258	281172663	4	NaN	1514	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18	2
11888591	123848007	106982400	4	956k2js	11750	2014-12-18 14	2014-12-18 14:00:00	14	2014-12-18	2
3227167	49871598	47827433	4	96kajt4	2631	2014-12-18 11	2014-12-18 11:00:00	11	2014-12-18	2
6696914	136470861	347836659	4	9rf42od	10472	2014-12-18 10	2014-12-18 10:00:00	10	2014-12-18	2
8618913	81958592	332303278	4	9r5frl3	13288	2014-12-18 17	2014-12-18 17:00:00	17	2014-12-18	2

8886 rows × 10 columns

3.9.5 计算r

- 计算最近一次购买数据的平均时间
- 大于平均时间, 设置为0
- 小于或者等于平均时间, 设置为1

In [15]:

```
1 #重置索引
2 rdata = tdata.set_index('user_id')
3 #设置R值
4 rdata['R'] = (rdata.tdays <= rdata.tdays.mean()).astype('i')
5 #过滤R列
6 rdata = rdata[['R']]
```

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

In [16]:

1	rdata
---	-------

Out[16]:

R		
	user_id	
	49598175	0
	141878326	0
	61454609	0
	86985041	0
	112485959	0
	71238931	0
	134020041	0
	36164872	0
	141792132	0
	104639948	0
	113445602	0
	95937118	0
	125817021	0
	21226209	0
	122694692	0
	44048864	0
	45149630	0
	21754104	0
	28160359	0
	10797721	0
	6596569	0
	56960895	0
	85947802	0
	122876959	0
	95302160	0
	17785981	0
	130457432	0
	98621566	0
	119112558	0
	139544053	0

	139104982	1
	82838416	1
	130132447	1
	88120786	1
	31446692	1
	135803085	1
	151617	1
	42966250	1
	54492776	1
	5686843	1
	45796530	1
	93687523	1
	1805740	1
	137129651	1
	83851924	1
	127552225	1
	3694961	1
	38968501	1
	867381	1
	109048285	1
	128180595	1
	131254719	1
	128733919	1
	124836406	1
	104777513	1
	131514258	1
	123848007	1
	49871598	1
	136470861	1
	81958592	1

8886 rows × 1 columns

3.9.6 计算F

用户在数据集中购物频率

Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [32]: 1 #根据user_id分组, 并且根据购物时间进行数量统计
2 fdata = tmp.groupby(['user_id']).time.count()
3 #重置索引
4 fdata = fdata.reset_index()
5 fdata = fdata.set_index('user_id')
6 fdata.columns = ['F']
7 fdata.F = pd.qcut(fdata.F, 2,['0','1'])
8 fdata
```

Out[32]:

	F
user_id	
4913	0
6118	0
7528	0
7591	1
12645	0
54056	0
63348	0
79824	1
88930	1
100539	1
104155	0
109103	1
113251	0
113960	1
120873	0
134658	0
151617	1
156608	1
157097	1
189833	1
190327	0
191366	0
213655	1
217996	0
227293	0
230711	1
239485	0
247543	1
250843	1
263670	0
...	...
142060721	0
142061900	0
142064691	1
142073213	0
142081324	1
142095821	1
142120051	1
142126535	1
142128942	1
142128951	1
142138619	0
142144275	0
142151675	1
142168798	1
142181816	0
142204924	0
142216376	1
142227202	1
142244794	1
142265405	1
142306250	1
142306361	0
142337230	0
142358910	0
142368840	1
142376113	0
142412247	1
142430177	0
142450275	1
142455899	1

8886 rows × 1 columns

3.9.7 数据集合并(R与F)

```
In [33]: 1 rfdata = pd.merge(rdata, fdata, left_index=True, right_index=True)
```

In []:

```
1
```

```
In [35]: 1 rfdata['RF'] = rfdata.R.str.cat(rfdata.F)
```


Contents

- 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
- 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

In [36]: 1 rfdata

Out[36]:

	R	F	RF
user_id			
49598175	0	0	00
141878326	0	0	00
61454609	0	0	00
86985041	0	0	00
112485959	0	0	00
71238931	0	0	00
134020041	0	0	00
36164872	0	0	00
141792132	0	0	00
104639948	0	0	00
113445602	0	0	00
95937118	0	0	00
125817021	0	0	00
21226209	0	0	00
122694692	0	0	00
44048864	0	0	00
45149630	0	1	01
21754104	0	0	00
28160359	0	0	00
10797721	0	0	00
6596569	0	0	00
56960895	0	0	00
85947802	0	0	00
122876959	0	0	00
95302160	0	0	00
17785981	0	0	00
130457432	0	0	00
98621566	0	0	00
119112558	0	0	00
139544053	0	0	00
...
139104982	1	1	11
82838416	1	1	11
130132447	1	1	11
88120786	1	1	11
31446692	1	1	11
135803085	1	1	11
151617	1	1	11
42966250	1	1	11
54492776	1	0	10
5686843	1	1	11
45796530	1	0	10
93687523	1	0	10
1805740	1	1	11
137129651	1	1	11
83851924	1	1	11
127552225	1	0	10
3694961	1	1	11
38968501	1	1	11
867381	1	1	11
109048285	1	1	11
128180595	1	1	11
131254719	1	1	11
128733919	1	1	11
124836406	1	0	10
104777513	1	1	11
131514258	1	1	11
123848007	1	1	11
49871598	1	1	11
136470861	1	1	11
81958592	1	1	11

8886 rows × 3 columns

3.9.8 设计RF, 根据RF进行分类

用户分类	R	F
重要价值用户	高	高
重要发展用户	高	低
重要保持用户	低	高
一般价值用户	低	低

Contents

- ▼ 1 案例:某年淘宝双12部分购物数据
 - 1.1 数据: 某年淘宝双12部分购物数据;
 - 1.2 分析目的:
- ▼ 2 熟悉数据
 - 2.1 导入数据
 - 2.2 数据字段说明
 - 2.3 数据量大小
 - 2.4 缺失数据检查
 - 2.5 基本分析
 - ▼ 2.6 指标: PV与UV
 - 2.6.1 时间处理
 - 2.6.2 PV分析: 每天访问量
 - 2.6.3 UV分析: 每天独立用户数
 - 2.6.4 PV与UV的数据可视化
 - 2.6.5 用户购买次数分布
 - ▼ 2.6.6 用户复购
 - 2.6.6.1 实现方式1
 - 2.6.6.2 实现方式2
 - ▼ 2.6.7 用户复购时间间隔
 - 2.6.7.1 diff方法
 - 2.6.7.2 实现
 - ▼ 2.7 AARRR模型
 - 2.7.1 理解AARRR
 - 2.7.2 双12用户购物模型
 - 2.7.3 不同层级数量
 - 2.7.4 计算转化率
 - ▼ 3 用户价值分析
 - 3.1 R理解:
 - 3.2 F理解:
 - 3.3 M理解:
 - 3.4 RFM评分模型
 - 3.5 不同指标对应测策略
 - 3.6 用户分类规则:
 - 3.7 RFM评分规则
 - 3.8 灵活使用
 - ▼ 3.9 淘宝双12购物用户RFM模型
 - 3.9.1 整理数据
 - 3.9.2 计算R, 最近购物时间
 - 3.9.3 数据验证
 - 3.9.4 计算R值
 - 3.9.5 计算r
 - 3.9.6 计算F
 - 3.9.7 数据集合并(R与F)
 - 3.9.8 设计RF, 根据RF进行分类

```
In [37]: 1 def func(value):
2         v = '一般用户'
3         if value == '00':
4             v = '一般用户'
5         elif value == '01':
6             v = '重要保持'
7         elif value == '10':
8             v = '重要发展'
9         elif value == '11':
10            v = '重要价值'
11         return v
```

```
In [38]: 1 rfdata['level'] = rfdata.RF.map(func)
2         rfdata
```

Out[38]:

	R	F	RF	level
user_id				
49598175	0	0	00	一般用户
141878326	0	0	00	一般用户
61454609	0	0	00	一般用户
86985041	0	0	00	一般用户
112485959	0	0	00	一般用户
71238931	0	0	00	一般用户
134020041	0	0	00	一般用户
36164872	0	0	00	一般用户
141792132	0	0	00	一般用户
104639948	0	0	00	一般用户

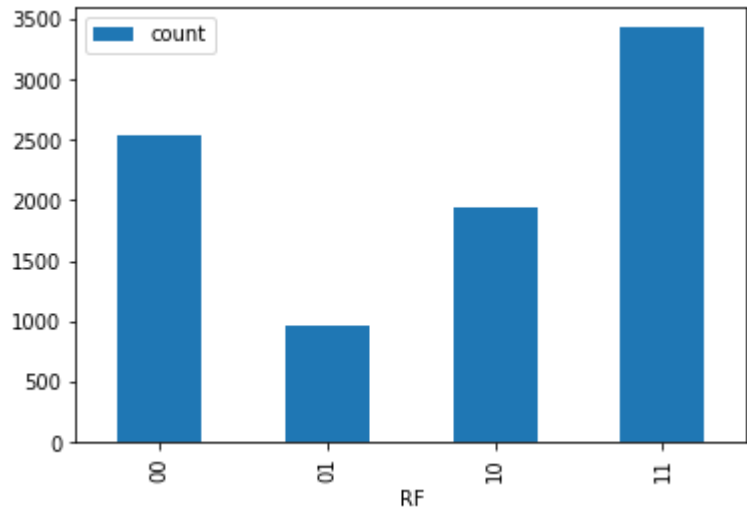
```
In [39]: 1 #分组统计
2 rf = rfdata.groupby('RF').F.count()
3 #创建DataFrame对象
4 rf = rf.reset_index()
5 #修改列名
6 rf = rf.rename(columns = {'F':'count'})
```

```
In [40]: 1 rf
```

Out[40]:

	RF	count
0	00	2540
1	01	966
2	10	1948
3	11	3432

```
In [289]: 1 _ = rf.plot.bar(x = 'RF', y='count')
```



```
In [ ]: 1
```