



链滴

数据分析项目 --SQL 的用户消费分析

作者: [Margaretlxh](#)

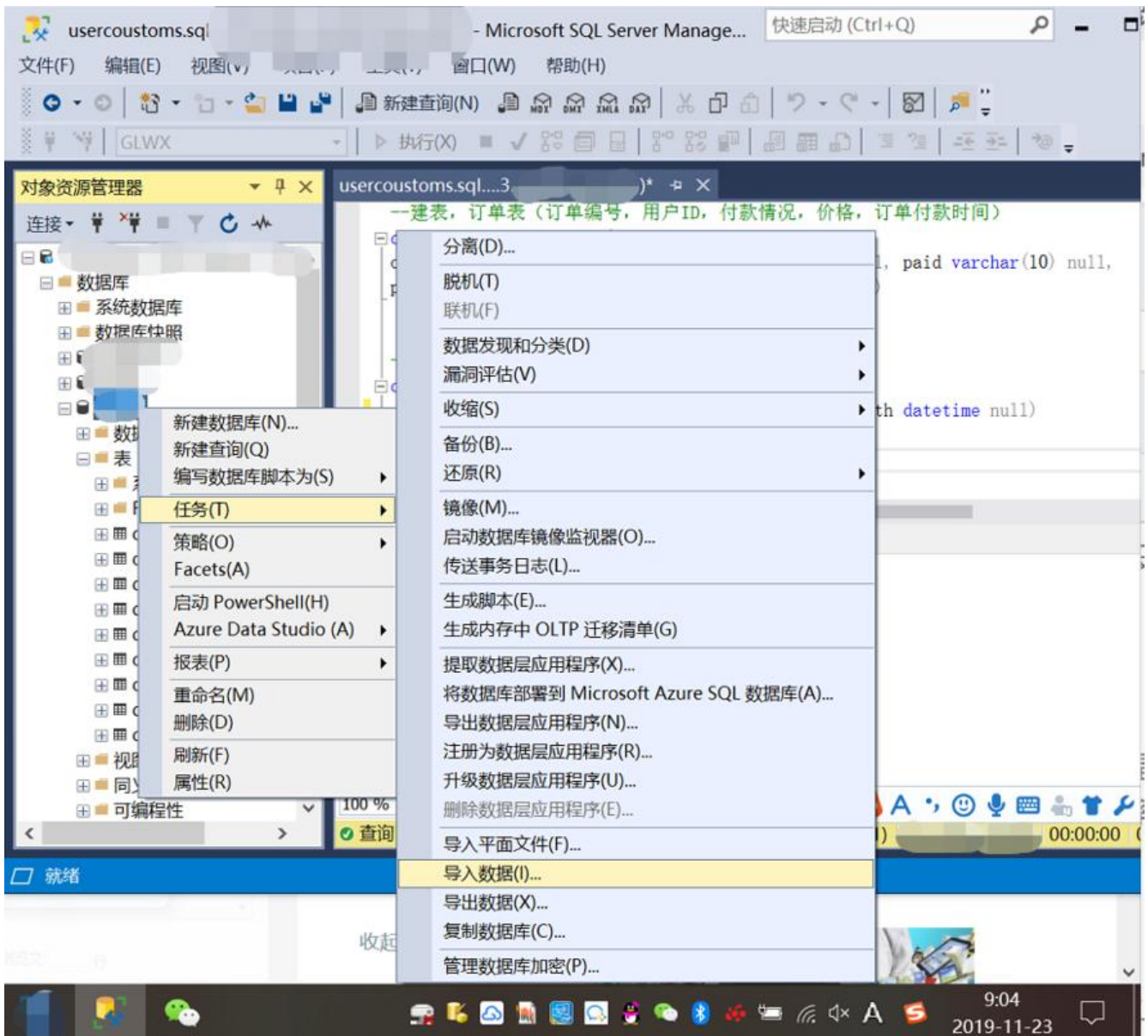
原文链接: <https://ld246.com/article/1575446048208>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

1、导入数据

方法1 直接导入CSV数据，不用建表做关联，如下图



方法2 建表做关联后导入数据

建表, 订单表 (订单编号, 用户ID, 付款情况, 价格, 订单付款时间)

```
create table orderform(  
orderid varchar(20) null, userid varchar(10) null, paid varchar(10) null,  
price varchar(10) null, payment_time datetime null)
```

建表, 客户信息表

```
create table userlist(  
userid varchar(10) null,gender varchar(5) null,birth datetime null)
```

1.1 需要更新下列名，更新列名的语句如下，不过还是表设计里面改起来最快

```
exec sp_rename 'order_info_utf.column1','order_info_utf.orderid'
```

2、用户消费行为分析相关

有关用户消费行为的统计指标，主要的分析内容主要是：

销量和消费金额分析、消费的时间节点、复购率回购率分析、用户分层、用户生命周期和消费周期等具体可以看看

2.1 可以统计不同月份订单数、用户数、下单总金额

```
select convert(varchar(6),payment_time,112) as year_month, count(orderid) as order_num ,
count(distinct userid) as user_num,sum(price) as sum from order_info_utf where paid='已支付'

group by convert(varchar(6),payment_time,112)
order by year_month
```

	year_month	order_num	user_num	sum
1	201603	238474	54799	158866078.3928
2	201604	223324	43967	159632117.3969
3	201605	7	6	4885.7614

2.2 统计用户复购率、回购率（这两个不一样的）

a. 复购率 当月消费一次以上(叫复购)人数的占比情况 (消费次数>1的人数÷消费总人数)

这里遇到了一个难点，就是SQL转换成带%号的比率，折磨人，计算完毕的比率是float型，%是字符，所以拼接是要将float转换后再加%，然后是char和varchar的区别导致的，也可以用convert转换

```
select t.*,concat(cast(round((cast(t.repeat_user as float)/t.user_num)*100,2)as varchar),'%') as repeat_rate from
(select year_month,count(distinct userid)as user_num,
sum(case when b.frequency>1 then 1 else 0 end) as repeat_user from
(select convert(varchar(6),payment_time,112) as year_month ,userid,count(userID) as frequency from order_info_utf
where paid='已支付' group by convert(varchar(6),payment_time,112), userID) b
group by year_month) t
order by year_month
```

	year_m...	user...	repeat...	repeat_rate
1	201604	43967	10283	23.39 %
2	201605	6	1	16.67 %
3	201603	54799	16916	30.87 %

	year_month	user_num	repeat_user	repeat_rate
1	201603	54799	16916	30.87%
2	201604	43967	10283	23.39%
3	201605	6	1	16.67%

b.回购率，回购率一般以月为单位区分，上月购买过的人在本月又买了,先找出本月购买的

```
select a.year_month1 年月,a.购买人数,b.回购人数 from
(select convert(varchar(6),payment_time,112) as year_month1,count( DISTINCT userid) 购买
数
from order_info_utf where paid='已支付' group by convert(varchar(6),payment_time,112))a
left join
(select convert(varchar(6),payment_time,112) year_month2,count (DISTINCT a.userid )回购人数

from order_info_utf a where a.paid='已支付' and a.userid in
(select userid from order_info_utf B where paid='已支付'
and convert(varchar(6),B.payment_time,112)=convert(varchar(6),a.payment_time,112)-1)
group by convert(varchar(6),payment_time,112))b on a.year_month1=b.year_month2
order by a.year_month1
```

结果	消息																
	<table border="1"> <thead> <tr> <th></th> <th>年月</th> <th>购买人数</th> <th>回购人数</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>201603</td> <td>54799</td> <td>NULL</td> </tr> <tr> <td>2</td> <td>201604</td> <td>43967</td> <td>13119</td> </tr> <tr> <td>3</td> <td>201605</td> <td>6</td> <td>4</td> </tr> </tbody> </table>		年月	购买人数	回购人数	1	201603	54799	NULL	2	201604	43967	13119	3	201605	6	4
	年月	购买人数	回购人数														
1	201603	54799	NULL														
2	201604	43967	13119														
3	201605	6	4														

2.3.统计男女的消费频次是否有差异，一般用户分层可以从性别、年龄段来着手

a.从性别分层找差异

```
select u.gender,sum(b.frenquency) as sum_frequency,count(distinct u.userid),
round(convert(float,sum(b.frenquency))/convert(float,count(distinct u.userid)),2)
as avg_frequency from user_info_utf u left join
(select userid,count(userid) as frenquency from order_info_utf where paid='已支付'
group by userid) b on u.userid=b.userid where gender is not null group by gender
```

结果	消息															
	<table border="1"> <thead> <tr> <th></th> <th>gender</th> <th>sum_frequency</th> <th>(无列名)</th> <th>avg_frequency</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>女</td> <td>24989</td> <td>17444</td> <td>1.43</td> </tr> <tr> <td>2</td> <td>男</td> <td>28306</td> <td>19197</td> <td>1.47</td> </tr> </tbody> </table>		gender	sum_frequency	(无列名)	avg_frequency	1	女	24989	17444	1.43	2	男	28306	19197	1.47
	gender	sum_frequency	(无列名)	avg_frequency												
1	女	24989	17444	1.43												
2	男	28306	19197	1.47												

结论：可见在性别方面，平均频次相差不大，那么可以顺便分析下不同性别的消费总额与平均额

```
select u.gender,sum(b.expenditure) as sum_expenditure,count(distinct u.userid), round(conver
(float,sum(b.expenditure))/convert(float,count(distinct u.userid)),2)
as avg_expenditure from user_info_utf u left join
(select userid,sum(price) as expenditure from order_info_utf where paid='已支付' group by us
rid)b
on u.userid=b.userid where u.gender is not null
group by u.gender
```

	gender	sum_expenditure	(无列名)	avg_expenditure
1	男	17698691.2669	19197	921.95
2	女	16634344.6802	17444	953.59

b.年龄分层，从年龄段来分析，先将年龄分层，最小的3岁，最大的1947，显然大于119的基本属于常值，可以将其剔除

```
select c.age_level,count(c.userid) as sum_frequency ,sum(a.expenditure) as sum_expenditure,
round(sum(a.expenditure)/count(c.userid),4) as avg_expenditure from
(select b.userid,age_level=(case when age <=17 then '0-17岁' when age between 18 and 25 t
en '18-25岁'when age between 26 and 35 then '26-35岁'
when age between 36 and 45 then '36-45岁' when age between 46 and 55 then
'46-55岁' else '56岁以上' end)
from (select userid,age=datediff(yy,convert(varchar(10),birth,23),GETDATE()) from user_info_u
f where birth is not null
and datediff(yy,convert(varchar(10),birth,23),GETDATE())<119) b )c
left join
(select userid,sum(price) as expenditure from order_info_utf where paid='success' group by u
serid) a on c.userid=a.userid
group by c.age_level
```

	age_level	sum_frequency	sum_expenditure	avg_expenditure
1	0-17岁	381	166061.309	435.8565
2	18-25岁	2464	1404875.0638	570.1603
3	26-35岁	12629	11371023.4611	900.3899
4	36-45岁	9936	10376578.2983	1044.3416
5	46-55岁	4395	4054134.9354	922.4425
6	56岁以上	2267	2054454.3351	906.2436

可以看出26-35和36-45年龄段的消费热刺和消费额都遥遥领先，但是在人均方面优势并不是十分明显，只有36-45年龄段的人均比较高，总消费额和总消费频次最多的26-35年龄段，人均消费和56岁以上年龄段的人均消费相差无几。

2.4.统计多次消费的用户，第一次和最后一次消费时间的间隔

```
select distinct userid,count(userid)as frequency,
convert(varchar(10),min(payment_time),120) as first_payment_time,
convert(varchar(10),max(payment_time),120) as last_payment_time,
datediff(dd,min(payment_time),max(payment_time)) as time_lag
from order_info_utf where paid='已支付'
group by userid having count(userid)>=2
order by frequency desc
```


	userid	frequency	first_payment_...	last_payment_...	time_lag
1	11211	318674	2016-03-01	2016-05-01	61
2	57282	713	2016-03-01	2016-04-30	60
3	62590	350	2016-03-01	2016-04-30	60
4	53616	337	2016-03-01	2016-04-30	60
5	68226	242	2016-03-01	2016-03-29	28
6	14427	240	2016-03-01	2016-04-30	60
7	14271	167	2016-03-01	2016-04-25	55
8	5775	158	2016-03-01	2016-04-10	40
9	44050	146	2016-03-01	2016-04-29	59
10	25263	144	2016-03-02	2016-04-30	59

2.5.消费黄金法则，二八原则,订单数量最多的20%客户，创造了多少的消费额，占总消费额的比例以及订单额最多的客户，占到80%的订单额的用户数占比总用户数的比率

```
select count(userid),sum(user_price),c.sum_user,c.sum_price,round(sum(b.user_price)/c.sum_price,2) as rate
from (select top 20 percent userid,count(userid) as user_frequency,sum(price) as user_price
from order_info_utf
where paid='success' group by userid order by count(userid) desc) b,
(select count(distinct userid) as sum_user ,sum(price) as sum_price from order_info_utf where
paid='success') c
group by c.sum_price,c.sum_user
```

	(无列名)	(无列名)	sum_user	sum_price	rate
1	17130	267538048.6018	85649	318503081.551103	0.84

以上，符合二八法则，关键用户满足，如果比率小于0.8，可以着重发展先大客户