

阶段二模块二作业：

1.0

-- 建立用户活跃日期表

```
CREATE TABLE temp_user_act (user_id INT, dates date);
```

1.1

-- 用用户活跃日期表做自连接

```
select * from
```

```
temp_user_act a
```

```
left join
```

```
temp_user_act b
```

```
on a.user_id=b.user_id;
```

如下图所示：

信息	结果 1	剖析	状态
user_id	dates	user_id(1)	dates(1)
98047837	2019-12-0	98047837	2019-12-05
98047837	2019-12-0	98047837	2019-11-18
98047837	2019-12-0	98047837	2019-12-18
98047837	2019-12-0	98047837	2019-11-30
98047837	2019-12-0	98047837	2019-12-10
98047837	2019-12-0	98047837	2019-11-25
98047837	2019-12-0	98047837	2019-11-28
98047837	2019-12-0	98047837	2019-12-11
98047837	2019-12-0	98047837	2019-12-04
98047837	2019-12-0	98047837	2019-11-24
98047837	2019-12-0	98047837	2019-11-22

1.2

-- 筛选出右表日期大于或等于左表日期的内容

```
select * from
```

```
temp_user_act a
```

```
left join
```

```
temp_user_act b
```

```
on a.user_id=b.user_id
```

```
where datediff(b.dates,a.dates)>=0;
```

如图所示：

信息	结果 1	剖析	状态
user_id	dates	user_id(1)	dates(1)
98047837	2019-12-02	98047837	2019-12-06
98047837	2019-12-01	98047837	2019-12-06
98047837	2019-11-29	98047837	2019-12-06
98047837	2019-12-05	98047837	2019-12-06
98047837	2019-11-18	98047837	2019-12-06
98047837	2019-11-30	98047837	2019-12-06
98047837	2019-11-25	98047837	2019-12-06
98047837	2019-11-28	98047837	2019-12-06

1.3

/* 计算以左表日期为基准日的当日用户数，第二日回访用户数，第三日回访用户数，第四日回访用户数，第八日回访用户数；*/

```
select a.dates,
count(distinct b.user_id) 当日用户数,
count(distinct if(datediff(b.dates, a.dates)=1,b.user_id,null)) 第二日回访用户数,
count(distinct if(datediff(b.dates, a.dates)=2,b.user_id,null)) 第三日回访用户数,
count(distinct if(datediff(b.dates, a.dates)=3,b.user_id,null)) 第四日回访用户数,
count(distinct if(datediff(b.dates, a.dates)=7,b.user_id,null)) 第八日回访用户数
from
temp_user_act a
LEFT JOIN temp_user_act b ON a.user_id = b.user_id
WHERE
datediff( b.dates, a.dates ) >= 0
group by a.dates;
```

如图所示：

信息	结果 1	剖析	状态		
dates	当日用户数	第二日回访用户数	第三日回访用户数	第四日回访用户数	第八日回访用户数
2019-12-06	250	163	161	161	154
2019-12-07	241	167	165	168	150
2019-12-08	247	171	173	181	153
2019-12-09	251	173	177	202	149
2019-12-10	254	190	207	166	163
2019-12-11	272	224	174	166	150
2019-12-12	324	203	193	195	0
2019-12-13	245	155	162	146	0
2019-12-14	235	161	152	151	0
2019-12-15	243	168	158	140	0
2019-12-16	245	166	143	0	0
2019-12-17	238	154	0	0	0
2019-12-18	224	0	0	0	0

1.4

利用上述数据计算出每日的当日用户数以及次日留存率，二日留存率，三日留存率，7日留存率(率需要使用百分比表示结果)

```
with p as(
select a.dates dates,
count(distinct b.user_id) user_count,
```

```

count(distinct if(datediff(b.dates, a.dates)=1,b.user_id,null)) remain1,
count(distinct if(datediff(b.dates, a.dates)=2,b.user_id,null)) remain2,
count(distinct if(datediff(b.dates, a.dates)=3,b.user_id,null)) remain3,
count(distinct if(datediff(b.dates, a.dates)=7,b.user_id,null)) remain7
from
temp_user_act a
LEFT JOIN temp_user_act b ON a.user_id = b.user_id
WHERE
datediff( b.dates, a.dates ) >= 0
group by a.dates
)
select dates,
concat(round((remain1 / user_count)100,2),"%") day1,
concat(round((remain2 / user_count)100,2),"%") day2,
concat(round((remain3 / user_count)100,2),"%") day3,
concat(round((remain7 / user_count)100,2),"%") day7
from p;

```

如图所示：

信息	结果 1	剖析	状态	
dates	day1	day2	day3	day7
2019-11-18	65.04%	64.60%	60.18%	60.18%
2019-11-19	69.37%	63.51%	61.71%	57.21%
2019-11-20	63.20%	59.31%	61.90%	57.14%
2019-11-21	62.07%	62.93%	65.09%	62.50%
2019-11-22	69.47%	64.16%	65.49%	63.27%
2019-11-23	66.80%	66.39%	58.51%	63.49%
2019-11-24	69.07%	63.98%	66.53%	65.25%

1.5

-- 求出每日的次留与次留的周环比

```

with p1 as
(with p as(
select a.dates dates,
count(distinct b.user_id) user_count,
count(distinct if(datediff(b.dates, a.dates)=1,b.user_id,null)) remain1
from
temp_user_act a
LEFT JOIN temp_user_act b ON a.user_id = b.user_id
WHERE
datediff( b.dates, a.dates ) >= 0
group by a.dates
)
select dates,
concat(round((remain1 / user_count)*100,2),"%") day1
from p
)
select dates,

```

```

day1,
lag(day1,7) over (order by dates) day1_1
from p1;

```

如图所示：

信息	结果 1	剖析	状态
dates	day1	day1_1	
2019-11-23	66.80%	(Null)	
2019-11-24	69.07%	(Null)	
2019-11-25	64.61%	65.04%	
2019-11-26	67.87%	69.37%	
2019-11-27	70.69%	63.20%	
2019-11-28	67.92%	62.07%	
2019-11-29	69.55%	69.47%	

2.1

--最大断更天数

```

with b as
  (with a as
    (select author_id,
      dates,
      lag(dates) over (partition by author_id order by dates) as dates2,
      @r := 0
    from temp_author_act)
    select author_id,
      dates,
      dates2,
      if(datediff(dates, dates2) = 1, @r := @r + 1, @r := 0) as lag
    from a)

select author_id,
  max(lag)
from b
group by author_id
;

```