公司每日的内呼外呼电话数 50万+

每天与订单相关的外呼电话 4万+

"2016-05-31"<=d and d <= "2018-03-27"

|  |  |
| --- | --- |
| 电话被外呼的次数 | 电话数 |
| 1 | 2228143 |
| 2 | 1129139 |
| 3 | 399112 |
| 4 | 210592 |
| 5 | 112702 |
| 6 | 68719 |
| 7 | 44394 |
| 8 | 30522 |
| 9 | 21924 |
| 10 | 16578 |
| 11 | 12876 |
| 12 | 10310 |
| 13 | 8426 |
| 14 | 7104 |
| 15 | 6056 |
| 16 | 5087 |
| 17 | 4522 |
| 18 | 4007 |
| 19 | 3586 |
| 20 | 3327 |

注:有2228143个电话在指定时间段内被打的次数为1

--根据每个电话被拨打的次数进行归类

SELECT phone\_call\_nums, COUNT(\*) as phone\_nums

FROM

(

SELECT dnis\_no, phone\_call\_nums FROM tmp\_htlbidb.tmp\_allcall GROUP BY dnis\_no, phone\_call\_nums

)a

GROUP BY phone\_call\_nums

dw\_htlmaindb.currentfacthtlordersnap 中可以对hotel与orderid关联在一起

关于验证模型的方法：

整个数据集被分成两部分（按时间平均分配），第一部分用来构建统计模型，第二部分用来

验证统计模型的效果，模型会吐出当前时间点当前电话的接通率，如果该电话不存在跳过。

1. 将第一部分数据统计存储在一张临时表中
2. 使用第二部分数据（test数据）去join存储的临时表，并且记录test数据中该通电话是否有效，记录统计模型给出的概率（历史该通电话在该时间点打出去的接通率）
3. 如果test中的数据（电话号码）不在统计模型中则不用考虑该条记录
4. 最后根据提取的数据做出roc曲线，寻找最优阈值

select

maincid, orderid, dnis\_no, time\_in,

hour(time\_in) as time\_quantum,

(case when callduration>0 then 1 else 0 end) as isconn

from OLAP\_CallCenterDB.O\_Call\_013

where "2017-06-01"<=d and d <= "2018-04-09" and UPPER(calltype)="O" and orderid<>-1

统计第二部分数据

use tmp\_htlbidb;

drop table if exists tmp\_allcall\_valid;

create table tmp\_allcall\_valid as

select

a.dnis\_no, b.time\_in,

a.time\_quantum,

a.con\_rate,

b.isconn

from

(

select \* from tmp\_htlbidb.tmp\_allcall\_part1

)a

join

(

select

dnis\_no, time\_in,

hour(time\_in) as time\_quantum,

(case when callduration>0 then 1 else 0 end) as isconn

from OLAP\_CallCenterDB.O\_Call\_013

where "2017-06-01"<=d and d <= "2018-04-09" and UPPER(calltype)="O" and orderid<>-1

)b

on a.dnis\_no == b.dnis\_no and a.time\_quantum == b.time\_quantum

创建一张时间表，一张电话表，用于做相互的笛卡尔积

--use tmp\_htlbidb;

drop table if exists tmp\_htlbidb.tmp\_allcall\_24;

create table tmp\_htlbidb.tmp\_allcall\_24 as

select distinct time\_quantum, (0) as same from tmp\_htlbidb.tmp\_allcall

--use tmp\_htlbidb;

drop table if exists tmp\_htlbidb.tmp\_allcall\_dnis\_no;

create table tmp\_htlbidb.tmp\_allcall\_dnis\_no as

select distinct dnis\_no, (0) as same from tmp\_htlbidb.tmp\_allcall

提高外呼酒店电话接通率的项目做模型的话可以参考提取的特征：

1. 当前电话的订单对应的酒店在当前时间（outcall分配时间）前后5分钟内酒店受理订单数
2. 当前电话的订单对应的酒店在当前时间（outcall分配时间）前后10分钟内酒店受理订单数
3. 当前电话的订单对应的酒店在当前时间（outcall分配时间）前后20分钟内酒店受理订单数
4. 最近两小时酒店订单数
5. 当前酒店当天的订单数
6. 当前电话当前时间电话的接通率
7. 当前时间对应的日期是不是节假日（海外，国内）
8. 当前时间是不是周中（海外，国内）
9. 当前时间是不是工作时间（海外，国内）
10. 当前酒店的砖级
11. 当前酒店的星级
12. 当前酒店的评分
13. 是否是客栈
14. 商务酒店、度假酒店
15. 是否集团对接
16. 是否是保留房
17. 预订渠道