

1. 数据获取前期准备

- 1.1 了解清楚业务方和研发说的是哪张表、哪份日志
- 1.2 了解这些表和日志的筛选条件是什么，为什么要这么筛选
- 1.3 了解这些表和日志之前有什么坑，是不是哪天数据有缺失
- 1.4 验证现在是否有坑，select *，先跑一个核心数据看下

2. SQL 指数常见问题

- 2.1 被各种各样的其他事情打扰
 - 2.1.1 解决办法----> 要有集中时间管理的意识；
 - 2.1.1.1 早上时间一定要利用好
 - 2.1.1.2 提前了解好会议主题，确定是否必须参加
 - 2.1.1.3 中间所有进来的排队需求，先靠边站
 - 2.1.1.4 在晚上或周末刻意练习“专注”这件事，成为习惯！
- 2.2 遇到一些坑需要思考
 - 2.2.1 通过“踩坑”来观人闻事
 - 2.2.2 对踩过的坑进行文档化
- 2.3 突然找到一个新的点，就一直往下深挖
 - 2.3.1 解决办法：一定要跳出小的圈子，我们最要的事情是先把问题的拆解模块做完！
- 2.4 自己本身不会指数和分析，不知道如何看数据
 - 2.4.1 解决办法：学习如何指数！！！！

2.4.1.1 学习同事如何指数

2.4.1.2 多接需求多指数，熟能生巧

先聚合再计算

如要统计某个维度下的用户数，不要直接count (distinct user)

```
select city,count (id) as user from (select city,imei,count (id) from a group by city,imei) t1 group by city
```

一列多行

AI通过中会讲一个用户打更多电话，而这里就是根据存在一个字段中

所以要根据需要数据，提前将数据准备好并先行验证

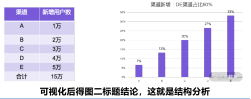
```
select * from t1 lateral view explode (id) table as b
```

取TOP

取每个分类下的top10高指数子分类 (去重--防止作弊)

```
select * from t1 over (partition by id order by id desc) as rank from table t1
```

2.4.1.3 总结如题



3.1 结构分析

3.1.1 所谓的结构分析，在这里就是：渠道的构成比例。

可视化后得图二结论，这就是结构分析

3.2 对比分析

3.2.1 所有数据只有对比才有意义！

渠道名称	渠道占比	渠道占比	渠道占比	渠道占比	
渠道A	0.05	渠道B	0.05	渠道C	0.05
渠道D	0.05	渠道E	0.05	渠道F	0.05
渠道G	0.1	渠道H	0.05	渠道I	0.05
渠道J	0.05	渠道K	0.05	渠道L	0.05
渠道M	0.05	渠道N	0.05	渠道O	0.05
渠道P	0.05	渠道Q	0.05	渠道R	0.05
渠道S	0.05	渠道T	0.05	渠道U	0.05
渠道V	0.05	渠道W	0.05	渠道X	0.05
渠道Y	0.05	渠道Z	0.05	渠道AA	0.05

3.2.2 最常见的对比对象就是：大盘！

微软案例：利用对比分析可以算出微软与抖音的重合用户更加偏爱哪些分类

日期	渠道	流量
5月24日	A	0.38
5月24日	B	0.41
5月24日	C	0.26
5月25日	A	0.27
5月25日	B	0.4
5月25日	C	0.28
5月26日	A	0.38
5月26日	B	0.42
5月26日	C	0.27
5月27日	A	0.39
5月27日	B	0.36
5月27日	C	0.29
5月28日	A	0.4
5月28日	B	0.34
5月28日	C	0.3
5月29日	A	0.37
5月29日	B	0.35
5月29日	C	0.32
5月30日	A	0.28
5月30日	B	0.33
5月30日	C	0.31