

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”来观人间事
 - 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 (1) as user from (select city,imei,count (1) from a group by city,imei) t1 group by city
```

直接聚合

如要统计某个用户打更多电话，而该用户是存在一个字段中

```
select user,count (1) as call from (select user,call from a) t1 group by user
```

聚合后再计算

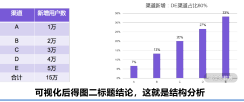
如要统计某个用户的top10消费用户分类 (去重+默认排序)

```
select user,rank (1) over (partition by user order by 1 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。



3.2 对比分析

3.2.1 所有数据只有对比才有意义！

3.2.2 最常见的对比对象就是：大盘！

微视觉案例：利用对比分析可以算出微视与抖音的重合用户更加偏爱哪些分类

| 微视觉/抖音 | 微视觉/抖音 | 微视觉/抖音 | 微视觉/抖音 | 微视觉/抖音 |
|--------|--------|--------|--------|--------|
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |
| 微视觉 | 0.05 | 微视觉 | 0.05 | 微视觉 |

数据分析思维与实战23讲 17 | 数据获取与分析：常见的 SQL 技巧和分析方法

- Who: 指用户基础属性、用户画像。
- Where: 渠道分析，渠道入口，用户从哪里来。
- When: 时间上的特征。
- What: 用户使用了什么功能，哪些行为更加重要。
- Why: 为什么要这么做，用户是主动还是被动做的。
- How: 怎么做的，行为路径是什么。

5W1H分析模板 (如果你记不住前面的分析方法，就用这个模板思考问题！)

3. 常用的分析方法

3.3 时间序列分析

直接看这个“新增用户的次日留存近半年出现下降”案例：



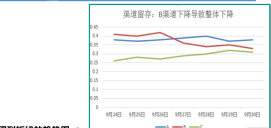
3.3.1 看原始数据

| 日期 | 渠道 | 留存 |
|-------|----|------|
| 9月24日 | A | 0.38 |
| 9月24日 | B | 0.41 |
| 9月24日 | C | 0.26 |
| 9月25日 | A | 0.37 |
| 9月25日 | B | 0.4 |
| 9月25日 | C | 0.28 |
| 9月26日 | A | 0.38 |
| 9月26日 | B | 0.42 |
| 9月26日 | C | 0.27 |
| 9月27日 | A | 0.39 |
| 9月27日 | B | 0.36 |
| 9月27日 | C | 0.29 |
| 9月28日 | A | 0.4 |
| 9月28日 | B | 0.34 |
| 9月28日 | C | 0.3 |
| 9月29日 | A | 0.37 |
| 9月29日 | B | 0.35 |
| 9月29日 | C | 0.32 |
| 9月30日 | A | 0.38 |
| 9月30日 | B | 0.33 |
| 9月30日 | C | 0.31 |

3.3.2 利用透视功能插入透视图

| 日期 | A | B | C |
|-------|------|------|------|
| 9月24日 | 0.38 | 0.41 | 0.26 |
| 9月25日 | 0.37 | 0.4 | 0.28 |
| 9月26日 | 0.38 | 0.42 | 0.27 |
| 9月27日 | 0.39 | 0.36 | 0.29 |
| 9月28日 | 0.4 | 0.34 | 0.3 |
| 9月29日 | 0.37 | 0.35 | 0.32 |
| 9月30日 | 0.38 | 0.33 | 0.31 |

3.3.3 得到这样一个模型的二维数据



3.3.4 利用插入图表功能得到所线的趋势图

结论：由上图可知渠道B渠道出现下滑，需要对B渠道进行优化，这就是时间序列分析的效果。

3.4 相关性分析

请看09课时的总结，谢谢！

3.5 机器学习

请看04课时的总结，谢谢！

4. 最后小建议：

- 4.1 指数和分析类报告急着写报告！！！！
- 4.2 把关键数据和结论同步给业务核心人员看
- 4.3 最后基于前面的理解，去写报告。