

小明哒SQL教程

——数据分析快速入门

小明先森

Bilibili名:小明先森 记得关注和三连呀! (¬_¬)

课程介绍

- 这个课程主要是帮助SQL零基础的数据分析师/数据产品经理快速入门SQL啦!
- 本课程主要涵盖内容包括：关系型数据库基本知识、SQL查询基础、Hive SQL简介、复杂数据分析逻辑的SQL实现。因针对数据分析，所以主要介绍“**数据查询**”部分，并不会涉及过多的数据库开发知识。
- SQL的教学环境为MySQL，但是也会简单介绍其他SQL环境下一些SQL写法的不同。
- 视频纯娱乐之作，PPT比较丑，内容也主要是个人经验提炼，如果有错误欢迎评论区指出~
- 喜欢本视频的话记得**关注和三连**呐，视频后续会持续更新！也欢迎在弹幕和评论区里提问和讨论~
- 课程所用的PPT、教学示例数据、代码、作业习题可以到以下网址下载：
- https://github.com/XiaoMingOwO/Bilibili_SQL_guide

数据分析师为什么要学习SQL呢？

- 查询数据/按指定形式提取分析数据。
- 从数据角度理解业务逻辑，能对数据出现的问题进行归因。
- 知道如何与数据开发有效沟通，提出数据需求。
- 从本质上理解BI工具。
- 掌握一种数据量过大时，仍能便捷处理数据的方法。

.....

下面我们来举一些栗子吧！



查询/提取数据

- 小红是一家公司的数据运营，她负责了一个用户回馈的活动。如果一个用户在30天内，每一天相对于**前一天在该平台花费的金额都提升了1%**，并且这个用户属于**5星级用户**，就给用户发一朵大红花。她应该怎么从公司的数据库里找到这样的用户呢？
- 小红从公司数据平台上下载下来了30天的支付记录，然后非常费劲的使用Excel里面的筛选功能和函数处理找出了符合支付条件的用户，然后又从公司的用户信息表中下载下了所有五星级用户，使用vlookup比对之后选出了发放小红花的用户。整个过程小红花了**几个小时**。
- 小红找到用户后开开心心和老板汇报，老板看了看说：“我觉得发大红花的用户这样筛出来看起来太少了，很多用户会不开心的。而且40岁以上的用户都是成熟的大人了，不会喜欢大红花的。你要不把**5星级用户**的条件换成**40岁以下**吧！”然后，重做一遍。**小红的一天就没了。。。**

业务理解/问题归因

- 数据出现明显下降了，是因为取数的问题还是真的就下降了？下降了的原因我应该怎么核查呢？
- 几处数据算同一个指标对不上是统计口径的差异，还是数据质量的问题？
- 开发做完了我需要的数据表之后，我应该怎么样去验证是否符合业务逻辑的要求呢？

上面这些问题可能都需要去写SQL去验证。用代码去描述业务逻辑表达更加精准~、(´▽`)/

需求沟通

一个完全不懂SQL的数据分析师/数据产品与开发的交流是这样的

数据：帮我加个算这个指标的功能吧。XXXX。。。。

开发：现在底层不支持算。

数据：为什么算不出来？不是这些数都有了吗？

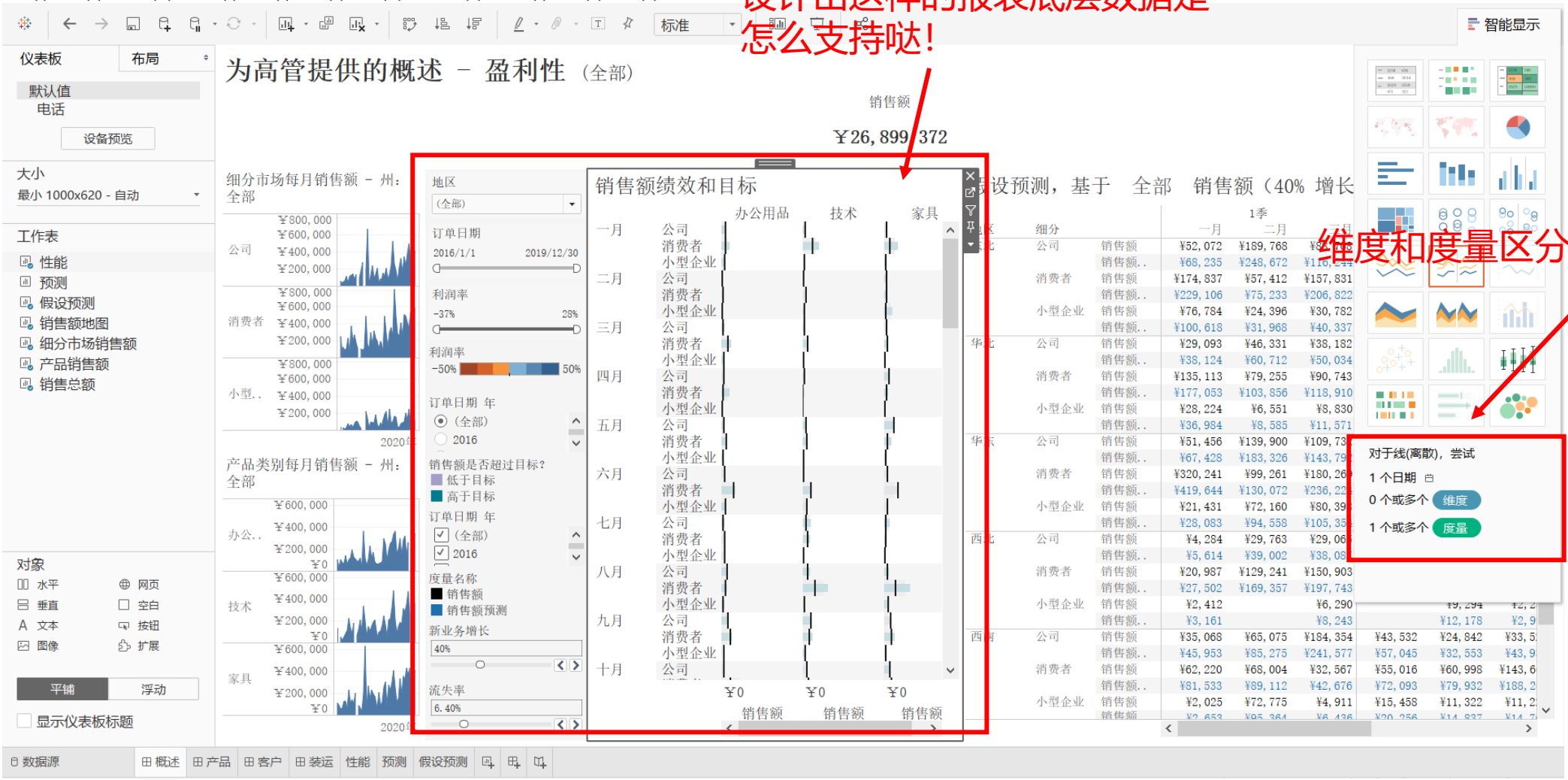
开发：因为这个计算逻辑是XXXX，你要的是XXXX，现在没有XXXX的表，所以没法在系统里直接做算这个。你先提个在底层做这个XXXX表的需求，做完后下次再给你做加XXXX指标的功能。

数据：我听不懂你在说什么！我不管，我就要这个功能嘛！



产品经理说要让APP的
颜色随手机壳的颜色变化

● 以Tableau为例



大规模数据处理

- Excel 2007以上版本，最多显示1048576行数据，并且很多复杂逻辑处理起来会很慢。



- 如果使用Python和R，学习成本相对SQL还是要高一些哒！
- 并且如果对于100GB以上的数据，在Spark之类的大数据平台中，还是得写SQL来处理数据。

课程体系

- 第一部分：关系型数据库基本知识
- 第二部分：SQL语法基础
- 第三部分：分组查询与聚合函数
- 第四部分：数据分析的常用 SQL 函数
- 第五部分：多表连接查询
- 第六部分：数据仓库简介与Hive SQL
- 第七部分：SQL数据分析实战

SQL学习资料推荐

- W3scool. SQL教程. <https://www.w3school.com.cn/sql/index.asp>
- 《SQL必知必会》
- 《MySQL必知必会》
- 王珊、萨师煊. 《数据库系统概论》