

## 课前测试

在开始学习之前，检测一下自己已经掌握的知识吧！亲爱的学员，你需要在之后的课程中找到本周课前测试的所有答案~

### 1.相比 Excel，SQL 的优势在于（多选）

- A. 依托于更强大的计算资源
- B. 更简洁高效的数据筛选方法
- C. 无需写代码，而是点选操作
- D. 让分析思维和方法能脱离于数据单独呈现出来

### 2.在数据收集上，以下哪个工作不是通过数据库进行的

- A. 抽取
- B. 拼接
- C. 聚合
- D. 可视化

### 3.SQL 语言的特点有哪些？

- A. 高度的“易学易用”
- B. 在企业界有很广泛的应用
- C. 成熟的标准语句写法
- D. SQL 语言对入门用户友善的工作方式

## 课前准备：MySQL 安装方法与常见问题



小灶 能力派

在学习 MySQL 之前，先了解 MySQL 安装方法以及 SQL 语言的书写规则吧。

**SQL 安装教程包.rar**

3.1 MB

**SQL 语言书写规则.pdf**

152.3 KB

## 课前准备：SQL 课程案例的数据



小灶 能力派

在本周的课程中我们将通过电商数据的若干个小任务，逐步了解 SQL 语言中的一些基本语法，这也是我们在第一周课程所提到的，应用 SQL 对数据进行“抽取，拼接，聚合”的过程。

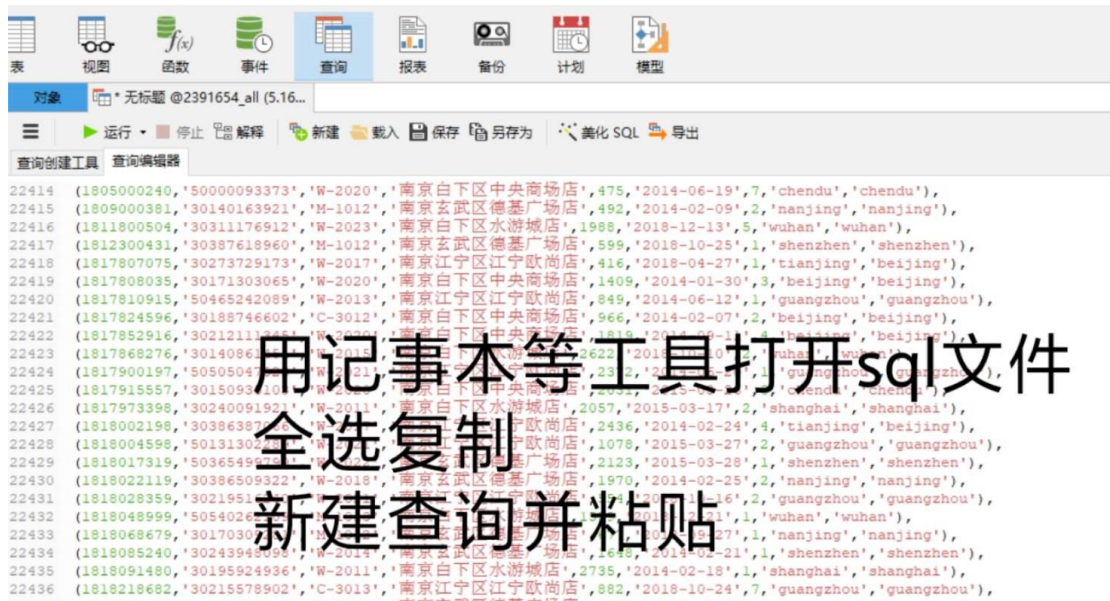
最后，我们将把数据扩展一步加入新的微信推送活动数据表，增加新的业务问题以及对应的 SQL 解决方法，以此完成我们对 SQL 语言的学习。

为了更好的学习和应用课程内容，下载下方的课程数据吧！本周课程案例中的表单都在其中。👉

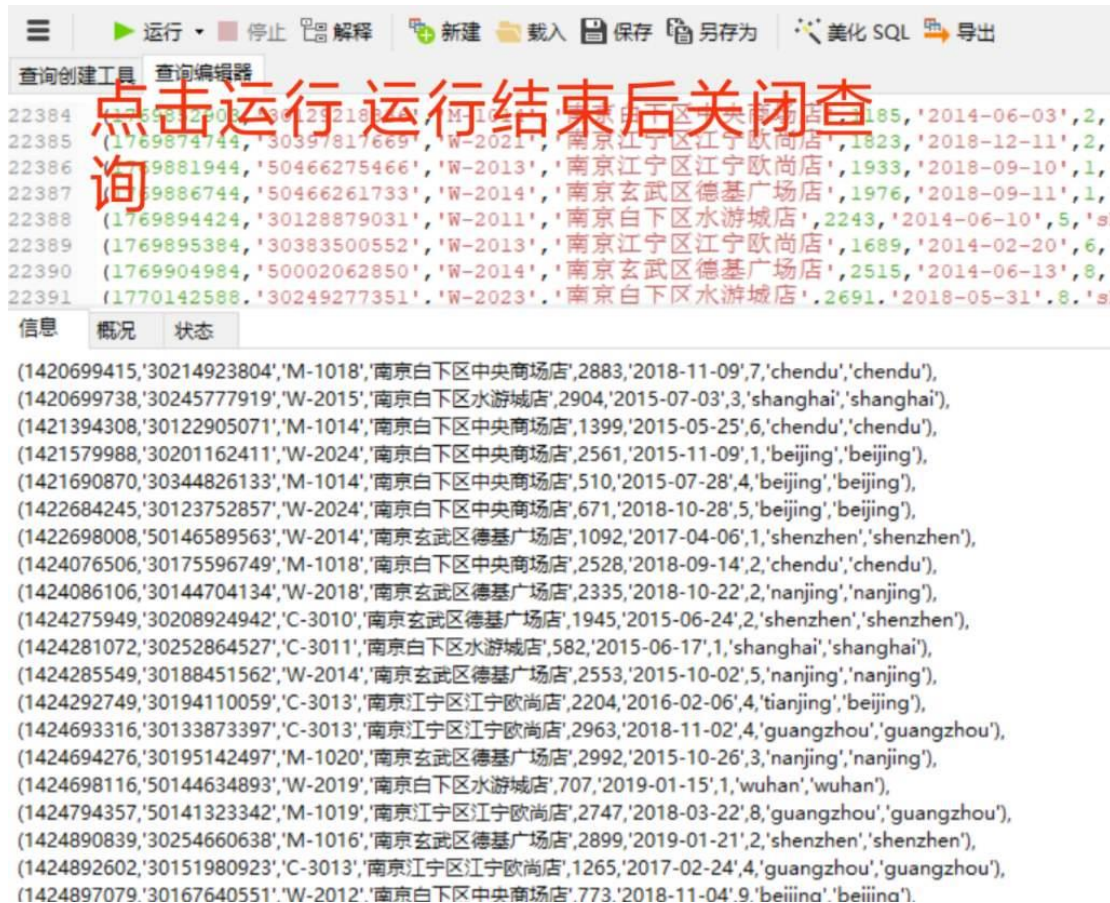
**课程数据.zip**

318.2 KB

在 SQL 数据导入教程



第一步 ↑



第二步 ↑



第三步↑

## 第 10 课：SQL 语言入门



小灶 能力派

SQL 语言，是所有数据分析师最重要也最常使用的编程语言。

SQL 语言强大且易于理解，是数据分析中需要掌握的基础工具。

### SQL

- SQL——Structured Query Language
- 是目前对数据库进行查询和编辑的主流编程语言

(视频 2.1 SQL 语言入门 0507\_batch.mp4)

## 关键词

### Excel 和 SQL 的异同

—

- Excel 处理数据存储在一个工作簿中，工作簿又以若干个工作表组成。在 SQL 环境中，数据是存储在数据库中，而数据库又是由若干个数据表（table）组成的。
- SQL 的计算资源远超过 Excel，在具体的分析场景中，运行速度和代码可读性也非常有优势。

### SQL 语言的 3 个特点

—

- 高度的“易学易用”
- 成熟的标准语句写法
- 对入门用户友善的工作方式

## 第 11 课：用 SQL 查询数据



小灶 能力派

在本小节，我们将通过数据集和若干个由浅入深的小任务，来分别解释 SQL 语言的几个基本语法：

使用最初的案例数据，运行下列 SQL，看下效果吧👉

```
SELECT COUNT(city),COUNT(distinct city),COUNT(*) FROM members;
```

将上面这条代码放入 SQL，看一下效果吧！

### 关键语句

#### DESCRIBE 语句

—

- “描述”指定表或视图中的所有列，可以帮助我们了解数据表中包含了哪些列
- 语法：Describe table\_name;

#### SELECT 语句

—

查询数据

语法： SELECT "字段名" FROM "表格名";

- SELECT \* FROM"表格名"; ——\*代表数据表中包含的所有列变量
- SELECT \* FROM "表格名" limit 5; ——“Limit”限制前 5 条
- SELECT Distinct "字段名" \* FROM "表格名"; ——“Distinct”去重

#### COUNT 函数

-

查询满足条件的记录总数

语法: SELECT COUNT("字段名") FROM "表格名";

- SELECT COUNT(\*) FROM "表格名"; ——计算数据总数
- SELECT COUNT("字段名") FROM "表格名"; ——计算变量 (去除了缺失值)
- SELECT COUNT(Distinct ("字段名")) FROM "表格名"; ——计算去重后的变量

### MIN/MAX/SUM/AVG 函数

-

- 最小值 min, 最大值 max, 平均值 avg 以及求和 sum
- 语法: SELECT min("字段名"),max("字段名"),avg("字段名"),sum("字段名"),count("字段名") FROM "表格名";

(视频 [2.2 用 SQL 认识和理解数据 0508\\_batch.mp4](#))

课程案例中的代码下载 

第 11 课 用 SQL 查询数据.pdf

80.8 KB

## 第 12 课：用 SQL 筛选数据



小灶 能力派

随着数据的剧增，我们实际上比较少在全表，或者某个字段的全表数据进行操作。

在分析中必须要进行筛选以聚焦特定的分析对象，以减少计算的数据集大小，提升 SQL 语言的运行速度，也方便验证对数据的理解。

在这里我们可以使用 WHERE 语句， WHERE 语句是 SQL 语句中的筛选语句。

```
SELECT * FROM Members where name like '张%' and phone  
like '09%'
```

将上面这条代码输入 SQL，猜猜会出现什么？

### 关键语句

#### WHERE 语句

-

- WHERE 语句是 SQL 语句中的筛选语句。它的主要功能是对数据按照特定的条件去进行筛选，一般位置在 FROM 的后面。
- 语法：SELECT 列名称 FROM 表名称 WHERE 列运算符值

#### 操作符

-

- 操作符是精确匹配
- 除了等号之外，我们还可以使用以下符号来满足我们不同的了解数据的需求。这类操作符的特点是：在操作符的右边，都是只有一个固定值，不管是数字还是字符。
  - 不等于<>或者!= ,
  - 大于> ,



- 小于<
- 大于等于>=
- 小于等于<=

## IS NULL

-

- 用 IS NULL 和 IS NOT NULL 这两个来筛选 NULL 值

## IN/BETWEEN/LIKE

-

- IN/BETWEEN 是范围匹配
  - IN
    - 只要满足其中一个就属于满足了 WHERE 的条件
    - 语法：SELECT "字段名"FROM "表格名"WHERE "字段名" IN ('值一', '值二', ...);
  - BETWEEN
    - 任何在两个取值范围之间的记录都被认为是满足了 WHERE 条件
    - 语法：SELECT "字段名"FROM "表格名"WHERE "字段名" BETWEEN '值一' AND '值二';
- LIKE 是模糊匹配，通常搭档百分号符%
  - 语法：SELECT "字段名"FROM "表格名"WHERE "字段名" LIKE '值%';

## AND/OR

-

- AND 操作符的逻辑是筛选出满足所有条件的数据
- OR 操作符的逻辑是筛选出满足任何一个条件的数据

## AS

—

- 对变量进行命名以及重命名
  - 语法：SELECT "字段名"AS"新字段名"FROM "表格名"

(视频 2.3 用 SQL 筛选数据 0507\_batch.mp4)

课程案例中的代码下载 

第 12 课 用 SQL 筛选数据.pdf

160.6 KB

## 第 13 课：用 SQL 拼接数据



小灶 能力派

我们之前使用的数据中包含销售，产品和会员三个方面的信息。

从业务角度上来说，把三个表进行连接可以帮助我们回答很多重要的商业问题，例如：

1. 由于销售表本身只记录了产品的编号，为了获取具体的产品类别及其它信息，这时候就需要将销售表和产品表进行匹配连接，以找到产品大类，小类的销售情况
2. 同样的销售表本身只记录了购买会员的编号，因此需要和会员表进行连接匹配，进而找出购买产品和未购买产品的的顾客年龄，性别分布是怎样的

为了实现表与表之间的匹配连接，我们需要使用到 SQL 语句的连接语句 JOIN 语句。

```
Select S.Product_ID
      ,S.Amount
      ,P.Base_category
      ,P.category
from Sales s
inner join Products p on S.Product_ID=P.Product_ID
;
```

运行以上代码，初步探索第一个商业问题。

## SQL 语句的连接语句 JOIN 语句

### INNER JOIN...ON 语句（最常用）

-

- INNER JOIN 语句两边是需要进行链接的两张表，ON 后面的部分则是告诉

SQL 是根据两张表里的哪个字段去进行表与表之间的连接。

- 语法：
  - SELECT column\_name(s)
  - FROM table1
  - INNER JOIN table2
  - ON table1.column\_name=table2.column\_name;

## LEFT/RIGHT JOIN 语句

—

- LEFT JOIN: 包含 LEFT JOIN 语句左边表的全部记录以及语句右边表能匹配上的部分记录
- RIGHT JOIN: 包含 RIGHT JOIN 语句右边表的全部记录以及语句左边表能匹配上的部分记录

## FULL OUTER JOIN 语句

—

- inner join 加上 left join 再加上 right join 三种 join 方式的结合

(视频 2.4 用 SQL 回答业务问题 0510\_batch.mp4)

课程案例中的代码下载 

第 13 课 用 SQL 拼接数据.pdf

96.9 KB

## 第 14 课：用 SQL 聚合数据



小灶 能力派

在上一节课，我们提出了两个商业问题：

1. 由于销售表本身只记录了产品的编号，为了获取具体的产品类别及其它信息，这时候就需要将销售表和产品表进行匹配连接，以找到产品大类，小类的销售情况。
2. 同样的销售表本身只记录了购买会员的编号，因此需要和会员表进行连接匹配，进而找出购买产品和未购买产品的的顾客年龄，性别分布是怎样的

我们将在本节课中，完成这两个商业问题所需要的数据整合工作。

```
Select p.Base_category
,P.Category
,sum(s.amount) as total
from Sales s
inner join Products p on S.Product_ID=P.Product_ID
group by base_category,category
order by total desc
;
```

运行上述代码，你找到问题一的答案了吗？

```
Select m.Member_ID
,m.Age
,m.Gender
,case when s.Member_ID is not null
Then 'Y'
Else 'N'
End as Flag
from Members m
left join Sales s on m.Member_ID=s.Member_ID;
```

运行上述代码，为解答问题二做好准备！

## 数据聚合

### GROUP BY 语句

-

- GROUP BY 语句可以实现按照特定字段包含的分类进行汇总计算的效果（比如求最小值，最大值，平均值求和等等）

## ORDER BY 语句

—

- 在 select from 语句后面加上 ORDER BY 语句可以实现将显示结果按照特定字段进行排序的效果。默认为升序，最后加 desc 为降序。

## HAVING 语句

—

- 将 HAVING 语句添加在 GROUP BY 语句的最后，可以起到筛选出我们想要的结果的效果。
- HAVING 语句中使用的筛选的变量一般是做过聚合运算处理之后计算出来的新变量。
- HAVING 语句的前面一定要加一条 GROUP BY 的汇总语句。

## CASE WHEN 语句

—

- CASE WHEN 语句主要是根据是否满足语句中的判断条件会落入不同的取值，对数据进行重新分类整理和命名。
- CASE WHEN 语句在结尾部分必须有个 END，来提示系统整个循环语句已经结束了。

(视频 2.5 用 SQL 聚合数据\_batch.mp4)

课程案例中的代码下载 

14 用 SQL 聚合数据.pdf

248.6 KB

## 第 15 课: SQL 进阶案例



小灶 能力派

在前一节，我们仍然是单个 SQL 语句的来看问题，那实际上我们的数据分析工作中应该是若干个 SQL 语句，一个个串起来来完成。

在本节中，我们将引入新的业务场景，迎接新的挑战。

思考题：用 SQL 评估微信推送活动的效果。

```
select a.camp_id
      ,b.total_amount
      ,a.cost
      ,b.total_amount/a.cost as ROI
from push
inner join (select p.camp_id
               ,sum(q.amount) as total_amount
             from push_detail o
             inner join push p on o.camp_id=p.camp_id
             inner join sales q on o.member_id=q.member_id
             where o.read_date>=p.drop_dt
             and o.read_date<=date(p.drop_dt)+1
             and q.transaction_date>=o.read_date
             and q.transaction_date<=date(o.read_date)+1
             group by p.camp_id
           ) b on a.camp_id=b.camp_id
order by ROI desc
;
```

运行上述代码，你得到答案了吗？

## 关键 SQL 语句

### CREATE TABLE 语句

—

- 创建新表
- 语法：CREATE TABLE New\_table\_name AS + SELECT 语句

### DROP TABLE 语句

- 删除表
- 语法: DROP TABLE table\_name;

(视频 2.6 SQL 进阶案例 0510\_batch.mp4)

## 用 SQL 评估微信推送活动的效果

### 【解题思路】

问题: 哪些微信推送活动给我们带来了好的销售业绩, 有比较好的效果?

- 评价标准:  $ROI = \text{收入} / \text{活动成本}$ , 找到 ROI 高的微信推送活动
- 定义: 因为微信推送带来的用户购买
- 三个关键的时间点:
  - 微信活动推送的时间
  - 用户阅读推送内容的时间
  - 以及用户发生购买行为的时间
- 总结: 对于每一条用户阅读时间是在推送时间的 24 小时以内, 并且交易发生时间是在阅读时间的 24 小时以内, 我们就认为这个交易是由微信推送活动所带来的。

### 【SQL 语句】

- 分步思考:
  - DESCRIBE 语句: 了解两张表格信息“微信推送活动表(push)”和“推送明细表(push\_detail)”
  - GROUP BY 语句 & ORDER BY 语句: 将推送活动按照其推送效果从好到坏进行排列



- WHERE 语句：保留微信推送活动中，那些阅读日期是在推送日期一天以内的记录
- CREATE TABLE 语句：新建一个“阅读过微信文章的会员表（push\_reader）”
- JOIN 语句：把满足条件的用户（push\_reader）和发生过购买行为的用户进行匹配链接
- AS 语句：返回到微信活动表，从中提取出推送活动的成本信息，计算出每个推送活动的投资回报率 ROI，并进行从高到低排序
- 语句整合：
  - 子查询 subquery 语句：在 FROM 的语句中，用括号包含住另一个 SELECT 语句，可得到一样的结果（即课程最初代码）。

## 第 16 课：从写 SQL 到写好 SQL



小灶 能力派

本节课中，将会为你讲述写好 SQL 语言的小技巧：

1. 书写代码时提供尽可能多的注释，注释符//
2. 只对所需要的数据进行连接 join，最好在 join 语句中就进行 where 筛选
3. 谨慎选择连接数据表 join 的方式：inner join 是几种 join 方式中最有效率匹配速度最快的
4. 编写代码时请注意及时换行

(视频 2.7 写 SQL 到写好 SQL0510\_batch.mp4)

SQL 语言虽然是从 SELECT 开始，但是实际上执行的顺序是：

1. 首先去找 FROM 和 JOIN 中所提到的数据表，确定我们所需要的数据表是否存在
2. 根据 Where 语句中筛选条件对记录进行筛选，只保留满足条件的记录
3. 如果有的话，执行聚合语句 Group by 和筛选语句 having 的部分
4. 最后执行 Select 部分的相关语句

## 课后测试

经过一周的学习，快来看自己掌握了多少知识吧！

1.还不清楚数据明细的时候，用什么语句可以了解表的数据结构

- A. SELECT FROM TABLE;
- B. DESC TABLE;
- C. SELECT FROM TABLE LIMIT 5;
- D. SELECT \* FROM TABLE;

2.想查询顾客表(customer)中性别是缺失的方法是

- A. DESC customer;
- B. SELECT \* FROM customer where gender=null;
- C. select \* from customer where gender is null;
- D. select \* from customer where gender='null'

3.如果只想取 A 表中有而不存在于 B 表的数据，正确的方式是？（多选）

- A. A INNER JOIN B
- B. A LEFT JOIN B
- C. A RIGHT JOIN B
- D. B RIGHT JOIN A

4.如果要使用 **group by** 和 **where** 来筛选组合，使用它们的正确顺序是

- A. from//group by//where
- B. FROM//WEHRE//GROUP BY
- C. GROUP BY//WHERE//FROM
- D. WHERE//FROM//GROUP BY

## 第二周作业：用 SQL 回答索尼的业务问题

索尼公司是源自日本的跨国综合企业，以研制电子产品为主要事业，经营领域横跨消费电子产品、专业性电子产品、电子游戏、金融、娱乐等，拥有全世界的品牌知名度。目前在国内，索尼的产品主要覆盖在照相机，摄像机，电视，和智能手机四个主要领域。

My Sony 俱乐部是索尼官方专为索尼会员设立的会员中心，分注册会员、铜牌会员、银牌会员、金牌会员四个等级，会员可根据会员等级享受不同的会员权益。同时，索尼还为尊敬的索尼会员提供了丰富的线上线下活动以及影像课堂等福利。

My Sony 会员，共分为注册会员、铜牌会员、银牌会员、金牌会员 4 个等级。会员的等级由"成长值"数值决定，成长值越高，会员等级越高。加入索尼会员可以享受会员尊享价，官方延保以及累积成长等多个激励。用户可以通过索尼中国在线商城，索尼产品，索尼 app，微信等渠道注册。

请将提供的数据导入到 MySQL 中，并试图回答下列问题

- (1) 所提供数据中，销售最好的产品分别是哪些？
- (2) 会员和非会员在购买金额和产品类别方面，有什么主要区别？
- (3) 不同注册渠道的会员，在购买金额和产品类别方面有什么主要区别？

【重要】记得下载下方的数据哦！👉👉👉

**sony.zip**

1 MB