

데이터 분석을 위한 중급 SQL - CASE

목적

데이터 분석을 위한 중급 SQL 강의 중 CASE 관련 강의를 수강 후 드는 의문점과 구글링으로 해결 과정을 기록하기 위해 해당 글을 마크다운으로 작성했습니다.

의문점

1. CASE 조건문에 대해서 categoryID = 1을 왜 '음료'라고 쓰고, categoryID = 2를 왜 '조미료' 또는 '소스'라고 썼을까?
2. 표준 SQL 구문의 실행 순서는 FROM - WHERE - GROUP BY - HAVING - SELECT - ORDER BY 순으로 실행된다고 알고 있었다. 하지만 SELECT에서 지정한 ALIAS(별칭)을 어떻게 GROUP BY에서 사용 가능할까?

첫번째 의문점에 대한 해결과정 및 결론

CASE 조건문에 대해서 categoryID = 1을 왜 '음료'라고 쓰고, categoryID = 2를 왜 '조미료' 또는 '소스'라고 썼을까?

SQL Statement:

```
SELECT CASE
  WHEN categoryid = 1 AND SupplierID = 1 THEN '음료'
  WHEN categoryid = 2 THEN '조미료'
  END AS 'categoryName', *
FROM Products;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 77

categoryName	ProductID	ProductName	SupplierID	CategoryID	Unit	Price
음료	1	Chais	1	1	10 boxes x 20 bags	18
음료	2	Chang	1	1	24 - 12 oz bottles	19
조미료	3	Aniseed Syrup	1	2	12 - 550 ml bottles	10
조미료	4	Chef Anton's Cajun Seasoning	2	2	48 - 6 oz jars	22
조미료	5	Chef Anton's Gumbo Mix	2	2	36 boxes	21.35
조미료	6	Grandma's Boysenberry Spread	3	2	12 - 8 oz jars	25

필자는 강의의 주요 내용 외에도 다른 사람들이 신경쓰지 않는 부분도 신경을 쓴다. 강의 내용 중 해당 쿼리를 작성하는 부분이 있었을 것이다. categoryid = 1을 왜 '음료'라고 지정했으며, 왜 categoryid = 2를 '조미료'라고 이름 지었을까? 의문이었다.

다행히 그 의문은 ProductName들을 구글링 검색을 통해 쉽게 해결하였다. 우선 가장 위에 있는 Chais를 검색해보았다.



Chais de Bourlemont

Merlot

[France](#) • [Atlantique](#) • [Chais de Bourlemont](#) • [Red wine](#) • [Merlot](#)

3.7  36 ratings

 Add to Wishlist

검색해 보니 맥주의 한 종류인 듯 하다.



×

궁금해서 Chang도 쳐봤는데 이 역시 맥주라고 나와 있다. 태국 타이 비버리지 사가 제조하는 맥주 브랜드이며, 태국어로 코끼리를 의미한다고 한다. 이외에도 categoryid가 1인 product를 다 검색해 본 결과 맥주 뿐만 아니라 와인, 탄산 음료등도 검색 된 것을 확인했다. 그래서 선미님이 categoryid가 1인 것을 '음료'라고 명시했구나 라는 것을 깨달았다.

그렇다면 categoryid가 4인 것에 대한 ProductName들을 구글링으로 확인해 본 결과 해당 제품의 공통적인 카테고리는 치즈라는 것을 알게 되었으며, 다음과 같이 지정 가능하게 만들어 보았다.

SQL Statement:

```
SELECT CASE
    WHEN categoryid = 1 THEN '음료'
    WHEN categoryid = 2 THEN '조미료'
    WHEN categoryid = 4 THEN '치즈'
    ELSE '기타'
END AS 'categoryName', *
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

이와 같이 product의 제품군을 확인 후 categoryid를 CASE 조건문으로 재정의하고, 확장하는 것이 가능할 것이다. (사실 Categories 테이블에 다 정의되어 있고 연결되어 있다. 데이터 베이스 내의 테이블 간의 연결은 실전반에서 할 것 같다.)

Result:

Number of Records: 8

CategoryID	CategoryName	Description
1	Beverages	Soft drinks, coffees, teas, beers, and ales
2	Condiments	Sweet and savory sauces, relishes, spreads, and seasonings
3	Confections	Desserts, candies, and sweet breads
4	Dairy Products	Cheeses
5	Grains/Cereals	Breads, crackers, pasta, and cereal
6	Meat/Poultry	Prepared meats
7	Produce	Dried fruit and bean curd
8	Seafood	Seaweed and fish

두번째 의문점에 대한 해결 과정 및 결론

표준 SQL 구문의 실행 순서는 FROM - WHERE - GROUP BY - HAVING - SELECT - ORDER BY 순으로 실행된다고 알고 있었다. 하지만 SELECT에서 지정한 ALIAS(별칭)을 어떻게 GROUP BY에서 사용 가능할까?

```
SELECT CASE
    WHEN categoryid = 1 THEN '음료'
    WHEN categoryid = 2 THEN '소스'
    ELSE '기타' END AS new_category, AVG(Price)
FROM Products
GROUP BY new_category
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 3

category	AVG(Price)
기타	28.11716981132075
소스	23.0625
음료	37.979166666666664

카테고리 id 별로 가격의 평균을 알고 싶을 때는 원래대로라면 다음과 같은 검색 쿼리로 작성을 한다.

```
SELECT categoryid
       , AVG(Price)
FROM Products
GROUP BY categoryid;
```

해당 쿼리에 대한 결과

Result:

Number of Records: 8

CategoryID	AVG(Price)
1	37.979166666666664
2	23.0625
3	25.16
4	28.73
5	20.25
6	54.006666666666667
7	32.37
8	20.6825

하지만 해당 쿼리는 카테고리id 별로 가격에 대한 평균을 나타내는 구문에 대한 데이터 셋이 나오므로 가시성이 그렇게 뛰어나지 않는 결과가 나온다.

그래서 다음과 같은 CASE 관련 쿼리문을 써서 데이터 셋에 대한 가시성을 높이는 데이터셋을 만든다.

```
SELECT CASE
    WHEN categoryid = 1 THEN '음료'
    WHEN categoryid = 2 THEN '소스'
    ELSE '이외'
END AS new_category, AVG(Price)
FROM Products
GROUP BY new_category;
```

해당 쿼리에 대한 결과

Result:

Number of Records: 3

new_category	AVG(Price)
소스	23.0625
음료	37.979166666666664
이외	28.11716981132075

그런데 여기서 의문점이 생길 것이다.

SELECT 문이 GROUP BY 문 이후에 실행되는 것으로 알고 있는데 어떻게 GROUP BY에서 SELECT CASE 문의 ALIAS를 사용할 수 있을까? 이 의문은 인프런에 있는 커뮤니티 게시판에 의해 쉽게 해결할 수 있었다.

표준 SQL 구문 실행 순서 참고 블로그 글: <https://myjamong.tistory.com/17>

MYSQL 8.0 표준 문서 중 별칭에 대한 글: <https://dev.mysql.com/doc/refman/8.0/en/problems-with-alias.html>

해당 의문에 관한 스택 오버플로우 글: <https://stackoverflow.com/questions/3841295/sql-using-alias-in-group-by>

결론은 표준 관계형 데이터베이스인 Oracle이나 MS SQL Server에서는 GROUP BY가 SELECT 절 보다 먼저 실행되기 때문에 SELECT 절에서 정의한 별칭은 GROUP BY 절에서 사용할 수 없는게 맞다. 하지만 예외 사항이 있다. **MYSQL**이나 **Postgres SQL**은 이를 가능하게 하는 기능을 만들어 **SELECT**에서 정의한 별칭을 **GROUP BY**에서도 사용할 수 있다.

