

Chapter 05. 유저 세그먼트별 분석

유저 세그먼트별 분석 문제 모음

🎯 13. 전체 유저의 Demographic을 알고 싶어요. 성 · 연령별로 유저 숫자를 알려주세요. 어느 세그먼트가 가장 숫자가 많나요? 참고로 기타 성별은 하나로, 연령은 5세 단위로 적당히 묶어주시고 유저 수가 높은 순서대로 보여주세요

- Case when 사용
- Null 값 오류 대처
- Group by, Order by 사용

```
1 • select case when length(gender)<1 then '기타'
2         else gender
3         end as gender
4     , case when age is null then '무응답'
5         when age <= 15 then '0_15세 이하'
6         when age <= 20 then '1_15-20세'
7         when age <= 25 then '2_21-25세'
8         when age <= 30 then '3_26-30세'
9         when age <= 35 then '4_31-35세'
10        when age <= 40 then '5_36-40세'
11        when age <= 45 then '6_41-45세'
12        when age >= 46 then '7_46세 이상'
13        end as age
14        , count(*) as cnt
15
16 from fastcampus.tbl_customer
17 group by 1, 2
18 order by 3 desc
```

🎯 14. Q12 결과의 성 · 연령을 "남성(25-29세)" 와 같이 통합해주시고,
각 성 · 연령이 전체 고객에서 얼마나 차지하는지 분포(%)를 알려주세요,
역시 분포가 높은 순서대로 알려주세요

- Concat 함수 사용
- Scala Subquery 사용

```
1 select concat(case when gender = 'M' then '남성'
2               when gender = 'F' then '여성'
3               when gender = 'Others' then '기타'
4               when length(gender)<1 then '기타'
5               end
6               , '('
7               , case when age is null then '무응답'
8                     when age <= 15 then '15세 이하'
9                     when age <= 20 then '15~20세'
10                    when age <= 25 then '21~25세'
11                    when age <= 30 then '26~30세'
12                    when age <= 35 then '31~35세'
13                    when age <= 40 then '36~40세'
14                    when age <= 45 then '41~45세'
15                    when age >= 46 then '46세 이상'
16                    end
17               , ')') as gen_age
18   , count(*) as cnt
19   , round(count(*)/(select count(*)
20                  from fastcampus.tbl_customer)*100,2) as per
21 from fastcampus.tbl_customer
22 group by 1
23 order by 2 desc
```

🎯 15. 2020년 7월, 성별에 따라 구매 건수와, 총 Revenue를 구해주세요.

이전 처럼 남녀 이외의 성별은 하나로 묶어주세요

- Join과 Group by 함께 사용

```
1  ● ○ select case when length(B.gender) < 1 then 'Others'
2      else B.gender
3      end as gender
4      , count(*) as cnt
5      , sum(price) as revenue
6
7  from fastcampus.tbl_purchase A
8  left join fastcampus.tbl_customer B
9  on A.customer_id = B.customer_id
10 where A.purchased_at >= '2020-07-01'
11      and A.purchased_at < '2020-08-01'
12
13 group by 1
```

🎯 16. 2020년 7월, 성별 · 연령대에 따라 구매 건수와, 총 Revenue를 구해주세요

- Join과 Group by 함께 사용

```
1  ● ○ select case when length(B.gender) < 1 then 'Others'
2      else B.gender
3      end as gender
4      ○ , case when B.age is null then '무응답'
5          when B.age <= 15 then '0_15세 이하'
6          when B.age <= 20 then '1_15-20세'
7          when B.age <= 25 then '2_21-25세'
8          when B.age <= 30 then '3_26-30세'
9          when B.age <= 35 then '4_31-35세'
10         when B.age <= 40 then '5_36-40세'
11         when B.age <= 45 then '6_41-45세'
12         when B.age >= 46 then '7_46세 이상'
13     end age
14     , count(*) as cnt
15     , sum(price) as revenue
16 from fastcampus.tbl_purchase A
17 left join fastcampus.tbl_customer B
18 on A.customer_id = B.customer_id
19 where A.purchased_at >= '2020-07-01'
20      and A.purchased_at < '2020-08-01'
21
22 group by 1, 2
23 order by 4 desc
--
```