

8-3. SQL Developer를 위한 데이터 분석

홍형경

chariehong@gmail.com

2020.11

1. Titanic 데이터 생성

- titanic_data.csv 파일을 이용해 테이블 생성 및 데이터 임포트
- titanic 테이블

TITANIC	설명
PASSENGERID	승객 아이디
SERVIVED	생존여부, 0: 사망, 1:생존
PCLASS	등급, 1, 2, 3
NAME	이름
SEX	성별, male : 남성, female : 여성
AGE	나이
SIBSP	같이 탑승한 형제 및 배우자 수
PARCH	같이 탑승한 부모 및 자녀 수
TICKET	티켓 번호
FARE	요금
CABIN	객실 번호
EMBARKED	탑승 항구, C: 세르부르(프랑스), Q: 퀸즈타운(아일랜드), S:사우샘프턴(영국)

2. Titanic 데이터 가공

- 좀 더 파악하기 쉽게 다음 쿼리를 실행해 TATANIC2 테이블 생성

```
CREATE TABLE titanic2 AS
```

```
SELECT passengerid
```

```
,CASE WHEN survived = 0 THEN '사망' ELSE '생존' end survived
```

```
,TO_CHAR(pclass) || '등급' pclass ,name
```

```
,DECODE(sex, 'male','남성', 'female','여성', '없음') gender
```

```
,age, sibsp ,parch ,ticket ,fare ,cabin
```

```
,CASE embarked WHEN 'C' THEN '프랑스-세르부르'
```

```
    WHEN 'Q' THEN '아일랜드-퀸즈타운'
```

```
    WHEN 'S' THEN '영국-사우샘프턴'
```

```
    ELSE ''
```

```
END embarked
```

```
FROM titanic
```

```
ORDER BY 1;
```

PASSENGERID	SURVIVED	PCLASS	NAME	GENDER	AGE	SIBSP	PARCH	TICKET	FARE	CABIN	EMBARKED
1	사망	3등급	Braund, Mr. Owen ...	남성	22	1	0A/5	21171	7.25 (null)		영국-사우샘프턴
2	생존	1등급	Cummings, Mrs. Joh...	여성	38	1	0PC	17599	71.2833	C85	프랑스-세르부르
3	생존	3등급	Heikkinen, Miss. ...	여성	26	0	0STON/02.	3101282	7.925 (null)		영국-사우샘프턴
4	생존	1등급	Futrelle, Mrs. Ja...	여성	35	1	0113803		53.1C	123	영국-사우샘프턴
5	사망	3등급	Allen, Mr. Willia...	남성	35	0	0373450		8.05 (null)		영국-사우샘프턴
6	사망	3등급	Moran, Mr. James	남성	(null)	0	0330877		8.4583 (null)		아일랜드-퀸즈타운
7	사망	1등급	McCarthy, Mr. Tim...	남성	54	0	017463		51.8625	E46	영국-사우샘프턴
8	사망	3등급	Palsson, Master. ...	남성	2	3	1349909		21.075 (null)		영국-사우샘프턴
9	생존	3등급	Johnson, Mrs. Osc...	여성	27	0	2347742		11.1333 (null)		영국-사우샘프턴
10	생존	2등급	Nasser, Mrs. Nich...	여성	14	1	0237736		30.0708 (null)		프랑스-세르부르
11	생존	3등급	Sandstrom, Miss. ...	여성	4	1	1PP	9549	16.7G	6	영국-사우샘프턴
12	생존	1등급	Bonnell, Miss. El...	여성	58	0	0113783		26.55C	103	영국-사우샘프턴
13	사망	3등급	Saunderscock, Mr. ...	남성	20	0	0A/5.	2151	8.05 (null)		영국-사우샘프턴
14	사망	3등급	Andersson, Mr. An...	남성	39	1	5347082		31.275 (null)		영국-사우샘프턴
15	사망	3등급	Vestrom, Miss. Hu...	여성	14	0	0350406		7.8542 (null)		영국-사우샘프턴

2. Titanic 데이터 가공

· TITANIC vs TITANIC2 테이블

컬럼	설명	TITANIC	TITANIC2
PASSENGERID	승객 아이디		
SERVIVED	생존여부	0, 1	생존, 사망
PCLASS	등급	1, 2, 3	1등급, 2등급, 3등급
NAME	이름		
SEX	성별	male, female	남성, 여성
AGE	나이		
SIBSP	같이 탑승한 형제 및 배우자 수		
PARCH	같이 탑승한 부모 및 자녀 수		
TICKET	티켓 번호		
FARE	요금		
CABIN	객실 번호		
EMBARKED	탑승 항구	C, Q, S	세르부르(프랑스), 퀸즈타운(아일랜드), 사우샘프턴(영국)

3. Titanic 데이터 분석

(1-1) 성별 생존/사망자 수

SELECT **gender**, **survived**, **COUNT(*)** cnt
FROM titanic2
GROUP BY **gender**, **survived**
ORDER BY 1, 2;

	♣ GENDER	♣ SURVIVED	♣ CNT
1	남성	사망	468
2	남성	생존	109
3	여성	사망	81
4	여성	생존	233

3. Titanic 데이터 분석

(1-2) 성별 생존/사망자 수와 비율

SELECT **gender**, **survived**, **cnt**,
ROUND(cnt / SUM(cnt) OVER (PARTITION BY gender ORDER BY gender),2) 비율
FROM (SELECT **gender**, **survived**, count(*) cnt
FROM titanic2
GROUP BY gender, survived
) t ;

	GENDER	SURVIVED	CNT	비율
1	남성	사망	468	0.81
2	남성	생존	109	0.19
3	여성	사망	81	0.26
4	여성	생존	233	0.74

3. Titanic 데이터 분석

(2-1) 등급별 생존/사망자 수

SELECT **pclass**, **survived**, **COUNT(*)** cnt
FROM titanic2
GROUP BY **pclass**, **survived**
ORDER BY 1, 2;

	PCLASS	SURVIVED	COUNT(*)
1	1등급	사망	80
2	1등급	생존	136
3	2등급	사망	97
4	2등급	생존	87
5	3등급	사망	372
6	3등급	생존	119

3. Titanic 데이터 분석

(2-2) 등급별 생존/사망자 수와 비율

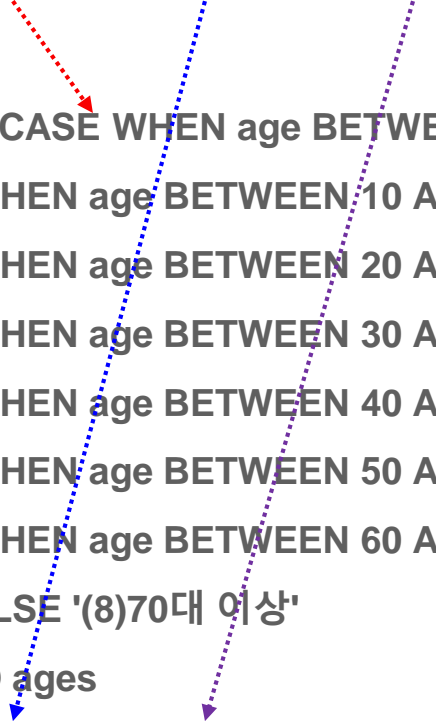
SELECT **pclass**, **survived**, **cnt**,
ROUND(cnt / SUM(cnt) OVER (PARTITION BY pclass ORDER BY pclass),2) 비율
FROM (SELECT **pclass**, **survived**, count(*) cnt
FROM titanic2
GROUP BY pclass, survived
) t ;

	PCLASS	SURVIVED	CNT	비율
1	1등급	사망	80	0.37
2	1등급	생존	136	0.63
3	2등급	사망	97	0.53
4	2등급	생존	87	0.47
5	3등급	사망	372	0.76
6	3등급	생존	119	0.24

3. Titanic 데이터 분석

(3) 연령대별 생존/사망자 수

```
SELECT CASE WHEN age BETWEEN 1 AND 9 THEN '(1)10대 이하'
        WHEN age BETWEEN 10 AND 19 THEN '(2)10대'
        WHEN age BETWEEN 20 AND 29 THEN '(3)20대'
        WHEN age BETWEEN 30 AND 39 THEN '(4)30대'
        WHEN age BETWEEN 40 AND 49 THEN '(5)40대'
        WHEN age BETWEEN 50 AND 59 THEN '(6)50대'
        WHEN age BETWEEN 60 AND 69 THEN '(7)60대'
        ELSE '(8)70대 이상'
        END ages
        ,survived, COUNT(*)
FROM titanic2
```



GROUP BY

```
CASE WHEN age BETWEEN 1 AND 9 THEN '(1)10대 이하'
        WHEN age BETWEEN 10 AND 19 THEN '(2)10대'
        WHEN age BETWEEN 20 AND 29 THEN '(3)20대'
        WHEN age BETWEEN 30 AND 39 THEN '(4)30대'
        WHEN age BETWEEN 40 AND 49 THEN '(5)40대'
        WHEN age BETWEEN 50 AND 59 THEN '(6)50대'
        WHEN age BETWEEN 60 AND 69 THEN '(7)60대'
        ELSE '(8)70대 이상'
        END
        ,survived
ORDER BY 1,2;
```

3. Titanic 데이터 분석

70대 이상 인원이 너무 많지 않나?

(3) 연령대별 생존/사망자 수

```
SELECT CASE WHEN age BETWEEN 1 AND 9 THEN '(1) 10대 이하'
           WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
           WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
           WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
           WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
           WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
           WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
           ELSE '(8) 70대 이상'
        END ages
      ,survived, COUNT(*)
  FROM titanic2
GROUP BY CASE WHEN age BETWEEN 1 AND 9 THEN '(1) 10대 이하'
           WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
           WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
           WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
           WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
           WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
           WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
           ELSE '(8) 70대 이상'
        END
      ,survived
ORDER BY 1,2;
```

	AGES		SURVIVED	COUNT(*)
1	(1) 10대 이하	사망		24
2	(1) 10대 이하	생존		31
3	(2) 10대	사망		61
4	(2) 10대	생존		41
5	(3) 20대	사망		143
6	(3) 20대	생존		77
7	(4) 30대	사망		94
8	(4) 30대	생존		73
9	(5) 40대	사망		55
10	(5) 40대	생존		34
11	(6) 50대	사망		28
12	(6) 50대	생존		20
13	(7) 60대	사망		13
14	(7) 60대	생존		6
15	(8) 70대 이상	사망		131
16	(8) 70대 이상	생존		60

3. Titanic 데이터 분석

(3) 연령대별 생존/사망자 수

```
SELECT age
FROM titanic2
ORDER BY 1 DESC;
```

	AGE
1	(null)
2	(null)
3	(null)
4	(null)
5	(null)
6	(null)
7	(null)
8	(null)
9	(null)
10	(null)
11	(null)
12	(null)
13	(null)
14	(null)
15	(null)
16	(null)
17	(null)
18	(null)

* 나이(age) 컬럼에서 null을 제거해야 한다!!

```
SELECT age
FROM titanic2
ORDER BY 1;
```

	AGE
1	0.42
2	0.67
3	0.75
4	0.75
5	0.83
6	0.83
7	0.92
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	2
16	2
17	2
18	2
19	2

* 나이가 1부터 시작하는게 아니었구나!!!

3. Titanic 데이터 분석

(3) 연령대별 생존/사망자 수 -- 쿼리 보정(1)

```
SELECT CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
          WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
          WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
          WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
          WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
          WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
          WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
          ELSE '(8) 70대 이상'
        END ages
        ,survived, COUNT(*)
FROM titanic2
WHERE age IS NOT NULL -- NULL 제거
GROUP BY CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
          WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
          WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
          WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
          WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
          WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
          WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
          ELSE '(8) 70대 이상'
        END
        ,survived
ORDER BY 1,2;
```

* 나이가 1부터 시작하는게 아니었구나!!!

	AGES	SURV...	COUNT(*)
1	(1) 10대 이하	사망	24
2	(1) 10대 이하	생존	38
3	(2) 10대	사망	61
4	(2) 10대	생존	41
5	(3) 20대	사망	143
6	(3) 20대	생존	77
7	(4) 30대	사망	94
8	(4) 30대	생존	73
9	(5) 40대	사망	55
10	(5) 40대	생존	34
11	(6) 50대	사망	28
12	(6) 50대	생존	20
13	(7) 60대	사망	13
14	(7) 60대	생존	6
15	(8) 70대 이상	사망	6
16	(8) 70대 이상	생존	1

* 나이(age) 컬럼에서 null을 제거해야 한다!!

3. Titanic 데이터 분석

(3) 연령대별 생존/사망자 수 -- 쿼리 보정(2)

- 나이가 NULL인 건을 제거하는 것이 맞을까?
 - NULL인 건이 누락되어 정확한 분석을 위해 부적절
 - 나이가 NULL인 건은 **알수없음** 으로 표시

3. Titanic 데이터 분석

(3) 연령대별 생존/사망자 수 -- 쿼리 보정(2)

```

SELECT CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
            WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
            WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
            WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
            WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
            WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
            WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
            WHEN age >= 70 THEN '(8) 70대 이상'
            ELSE '(9) 알수없음'
        END ages
        ,survived, COUNT(*)
FROM titanic2
GROUP BY CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
            WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
            WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
            WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
            WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
            WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
            WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
            WHEN age >= 70 THEN '(8) 70대 이상'
            ELSE '(9) 알수없음'
        END
        ,survived
ORDER BY 1,2;
    
```

	AGES	SURVIVED	COUNT(*)
1	(1) 10대 이하	사망	24
2	(1) 10대 이하	생존	38
3	(2) 10대	사망	61
4	(2) 10대	생존	41
5	(3) 20대	사망	143
6	(3) 20대	생존	77
7	(4) 30대	사망	94
8	(4) 30대	생존	73
9	(5) 40대	사망	55
10	(5) 40대	생존	34
11	(6) 50대	사망	28
12	(6) 50대	생존	20
13	(7) 60대	사망	13
14	(7) 60대	생존	6
15	(8) 70대 이상	사망	6
16	(8) 70대 이상	생존	1
17	(9) 알수없음	사망	125
18	(9) 알수없음	생존	52

3. Titanic 데이터 분석

(4) 연령대, 성별 생존/사망자 수

```
SELECT CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
           WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
           WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
           WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
           WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
           WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
           WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
           WHEN age >= 70 THEN '(8) 70대 이상'
           ELSE '(9) 알수없음'
        END ages
,gender, survived, COUNT(*)
FROM titanic2
GROUP BY gender,
        CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
        WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
        WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
        WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
        WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
        WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
        WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
        WHEN age >= 70 THEN '(8) 70대 이상'
        ELSE '(9) 알수없음'
        END
,gender, survived
ORDER BY 1,3,2;
```

	AGES	GENDER	SURVIVED	COUNT(*)
1	(1) 10대 이하	남성	사망	13
2	(1) 10대 이하	여성	사망	11
3	(1) 10대 이하	남성	생존	19
4	(1) 10대 이하	여성	생존	19
5	(2) 10대	남성	사망	50
6	(2) 10대	여성	사망	11
7	(2) 10대	남성	생존	7
8	(2) 10대	여성	생존	34
9	(3) 20대	남성	사망	123
10	(3) 20대	여성	사망	20
11	(3) 20대	남성	생존	25
12	(3) 20대	여성	생존	52
13	(4) 30대	남성	사망	84
14	(4) 30대	여성	사망	10
15	(4) 30대	남성	생존	23
16	(4) 30대	여성	생존	50
17	(5) 40대	남성	사망	45
18	(5) 40대	여성	사망	10
19	(5) 40대	남성	생존	12
20	(5) 40대	여성	생존	22
21	(6) 50대	남성	사망	26
22	(6) 50대	여성	사망	2
23	(6) 50대	남성	생존	4
24	(6) 50대	여성	생존	16
25	(7) 60대	남성	사망	13
26	(7) 60대	여성	생존	2
27	(7) 60대	남성	생존	4
28	(8) 70대 이상	남성	사망	6
29	(8) 70대 이상	여성	생존	1
30	(9) 알수없음	남성	사망	108
31	(9) 알수없음	여성	사망	17
32	(9) 알수없음	남성	생존	16
33	(9) 알수없음	여성	생존	36

3. Titanic 데이터 분석

(5) 형제,배우자 수별 부모자식수별 생존/사망자 수

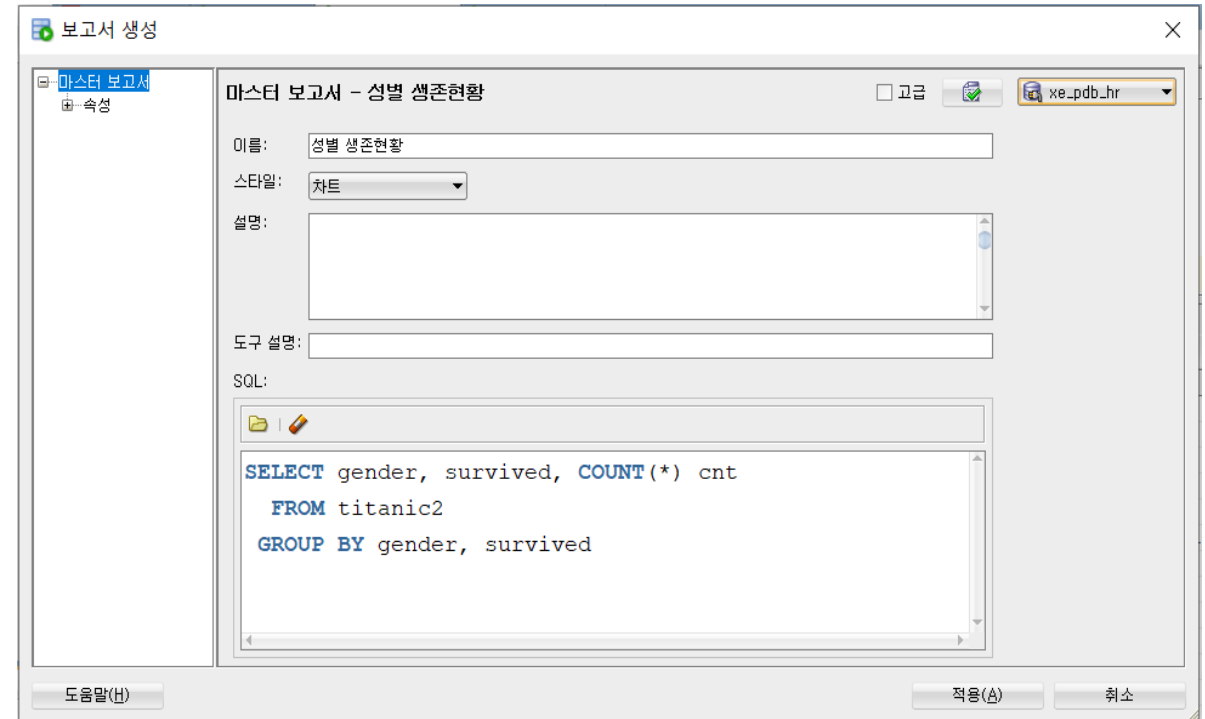
```
SELECT sibsp, parch, survived, count(*)
FROM titanic2
GROUP BY sibsp, parch, survived
ORDER BY 1, 2, 3;
```

	SIBSP	PARCH	SURVIVED	COUNT(*)
1	0	0	사망	374
2	0	0	생존	163
3	0	1	사망	13
4	0	1	생존	25
5	0	2	사망	8
6	0	2	생존	21
7	0	3	생존	1
8	0	4	사망	1
9	0	5	사망	2
10	1	0	사망	59
11	1	0	생존	64
12	1	1	사망	23
13	1	1	생존	34
14	1	2	사망	7
15	1	2	생존	12
16	1	3	사망	2
17	1	3	생존	1
18	1	4	사망	3
19	1	5	사망	2
20	1	5	생존	1
21	1	6	사망	1
22	2	0	사망	12
23	2	0	생존	4
24	2	1	사망	1
25	2	1	생존	6
26	2	2	사망	2
27	2	2	생존	2
28	2	3	생존	1
29	3	0	생존	2
30	3	1	사망	7
31	3	2	사망	5
32	3	2	생존	2
33	4	1	사망	9
34	4	2	사망	6
35	4	2	생존	3
36	5	2	사망	5
37	8	2	사망	7

4. Titanic 데이터 분석 시각화

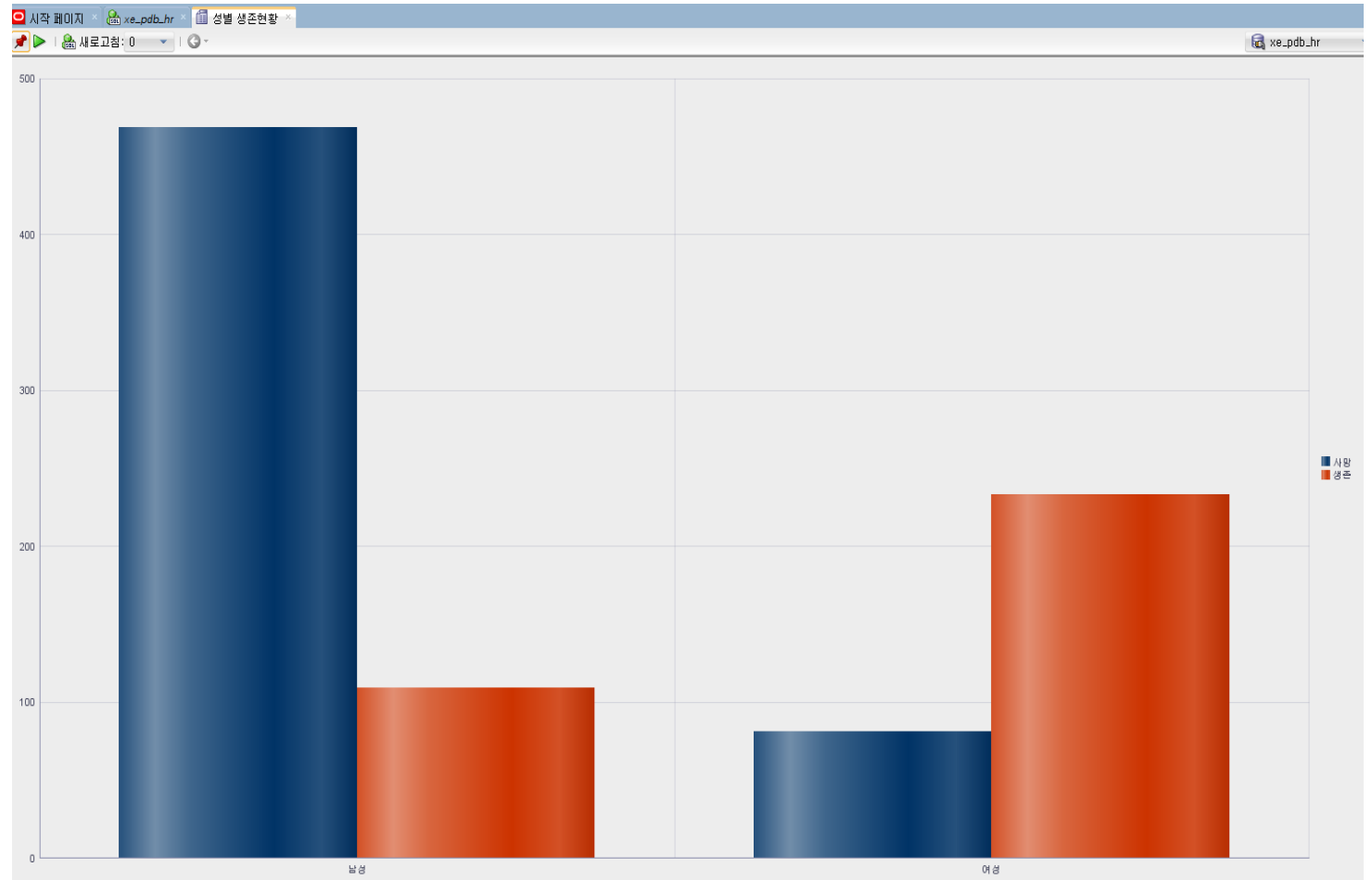
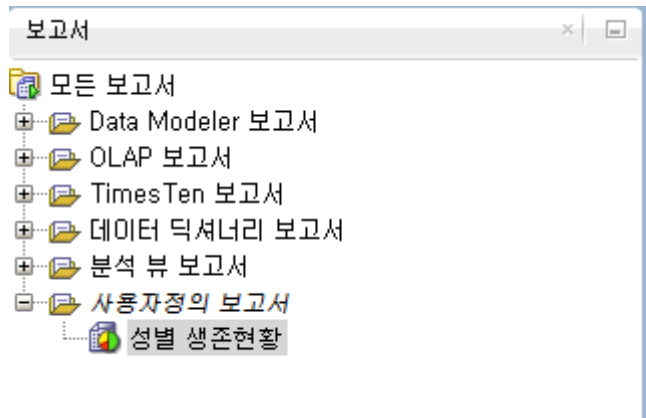
- SQL Developer 실행 → 보기 메뉴 → 보고서 선택
- 보고서 탭 → 사용자 정의 보고서 선택 → 마우스 오른쪽 버튼 → 새 보고서 선택
- - 이름: 성별 생존현황
 - 스타일: 차트
 - SQL :

```
SELECT gender, survived, COUNT(*) cnt
FROM titanic2
GROUP BY gender, survived
```
 - 우측 상단에서 접속 정보 선택 후, 적용 버튼 클릭



4. Titanic 데이터 분석 시각화

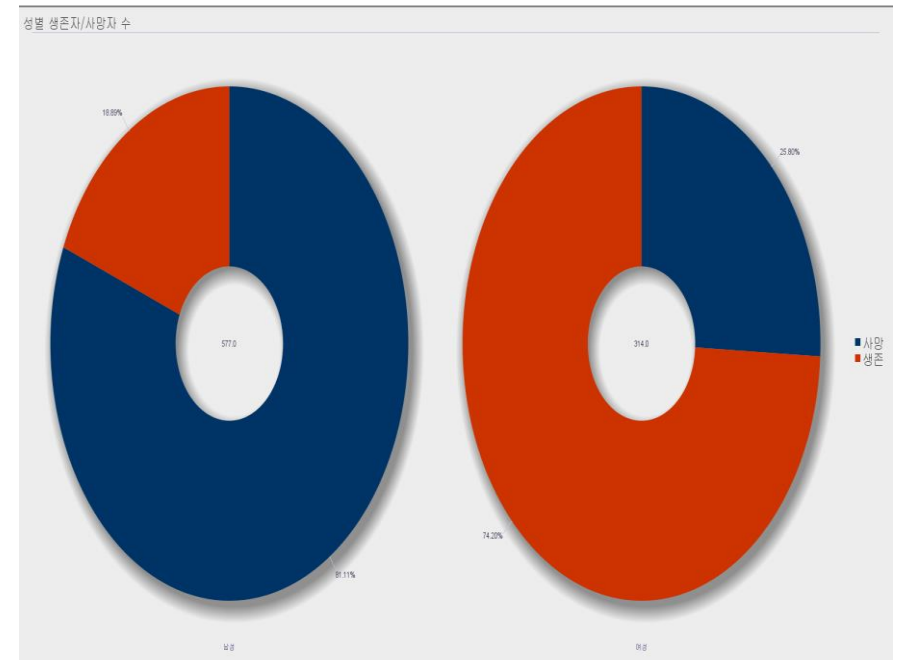
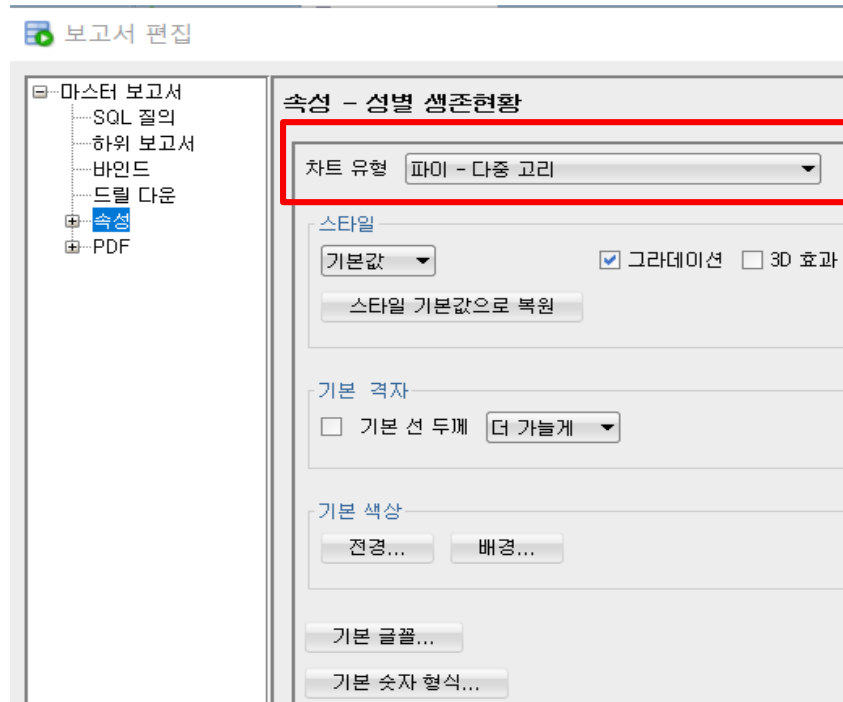
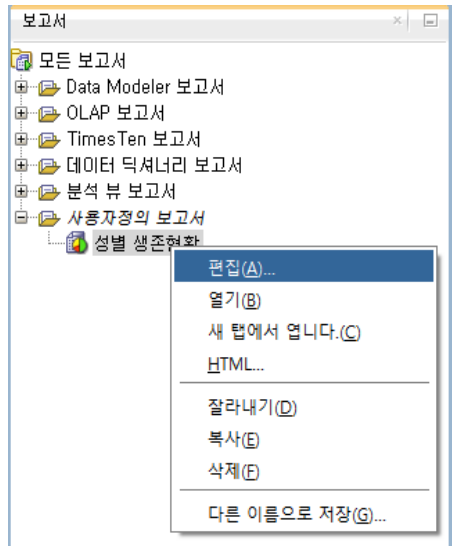
(1) 성별 생존/사망자 수 그래프



4. Titanic 데이터 분석 시각화

(1) 성별 생존/사망자 수 그래프

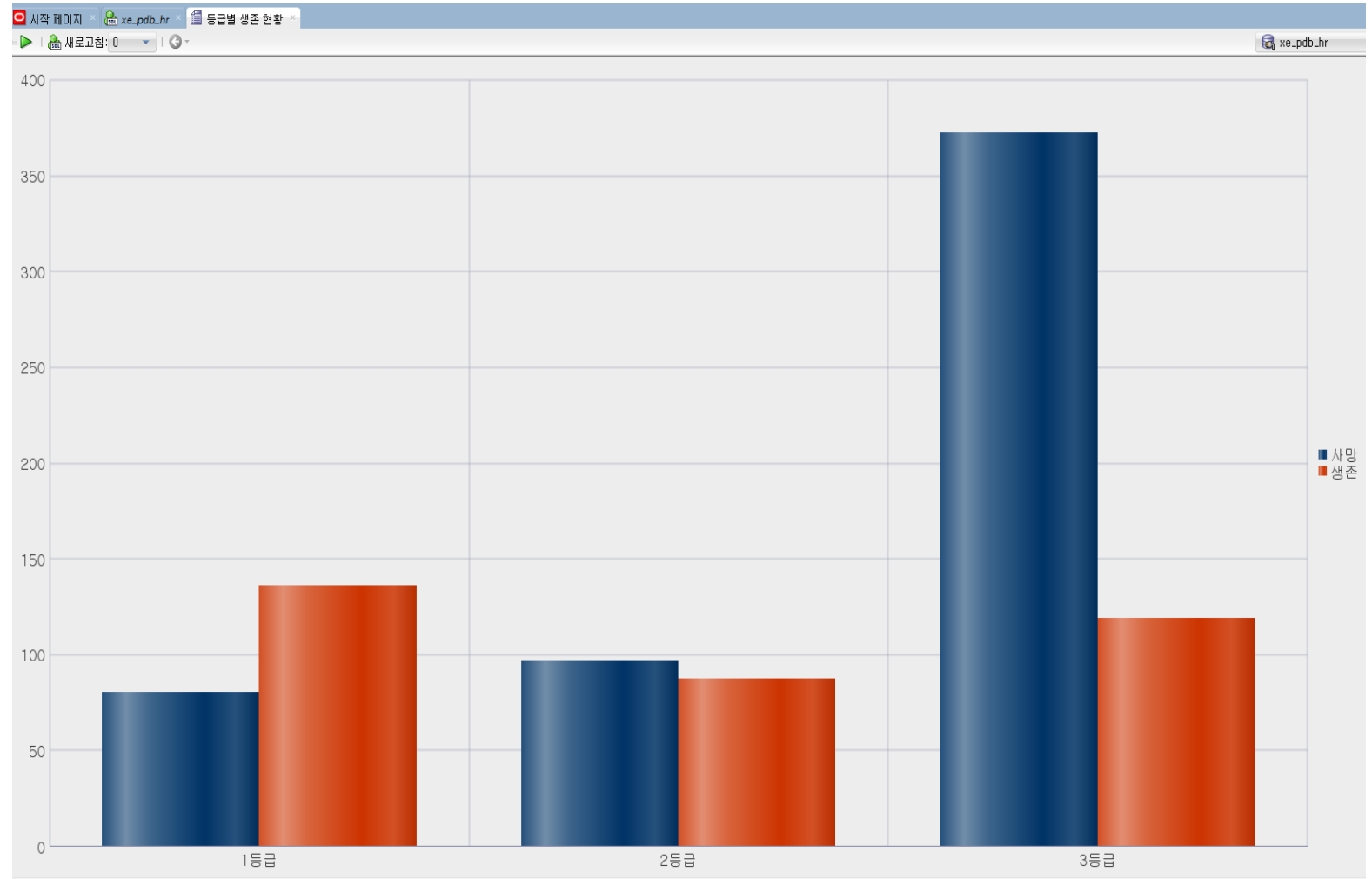
- 편집 → 보고서 편집 창에서 속성 선택 → 다양한 차트 유형 선택 가능



4. Titanic 데이터 분석 시각화

(2) 등급별 생존/사망자 수 그래프

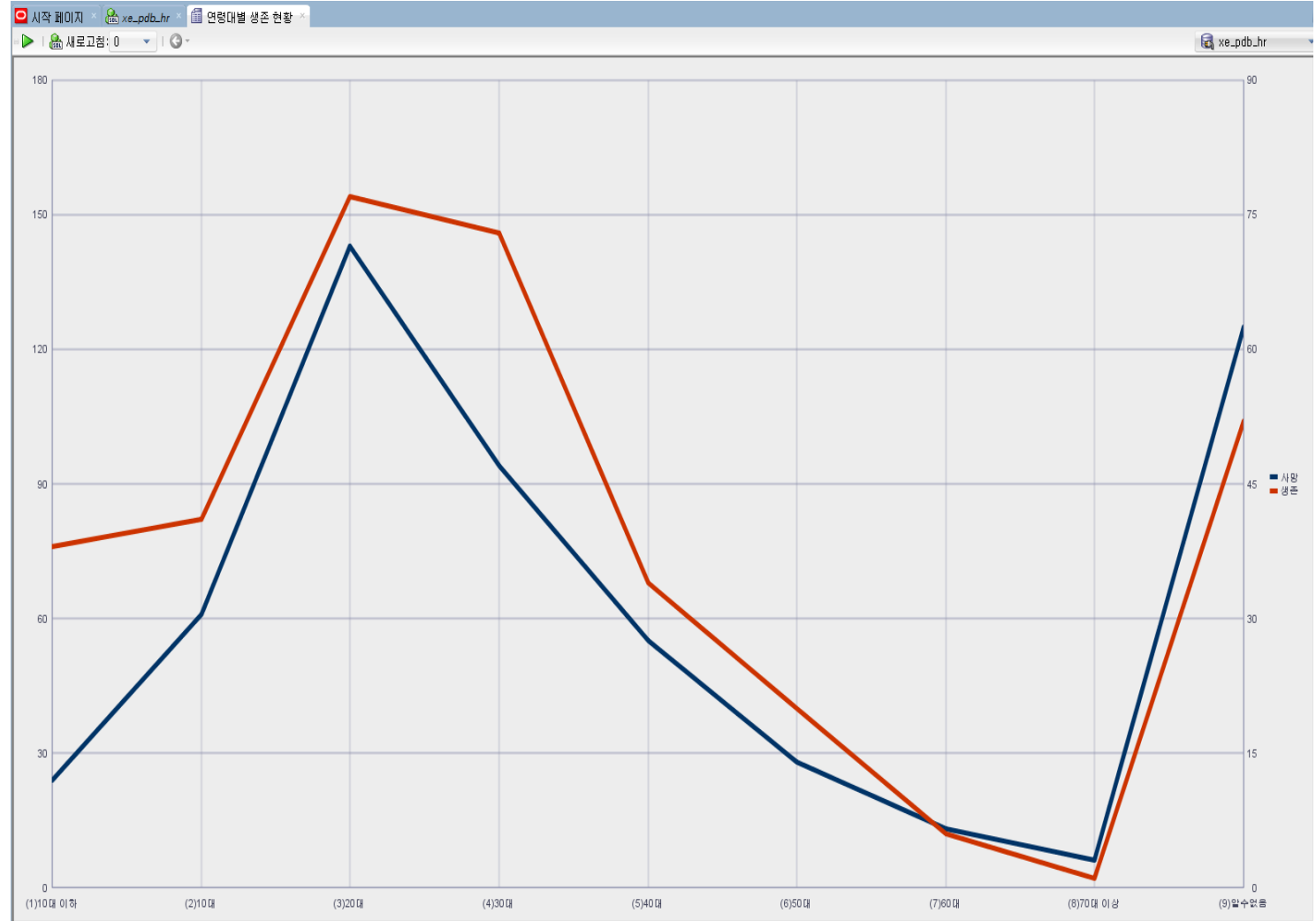
```
SELECT pclass, survived, count(*)  
FROM titanic2  
GROUP BY pclass, survived  
ORDER BY pclass, survived
```



4. Titanic 데이터 분석 시각화

(3) 연령대별 생존/사망자 수 그래프

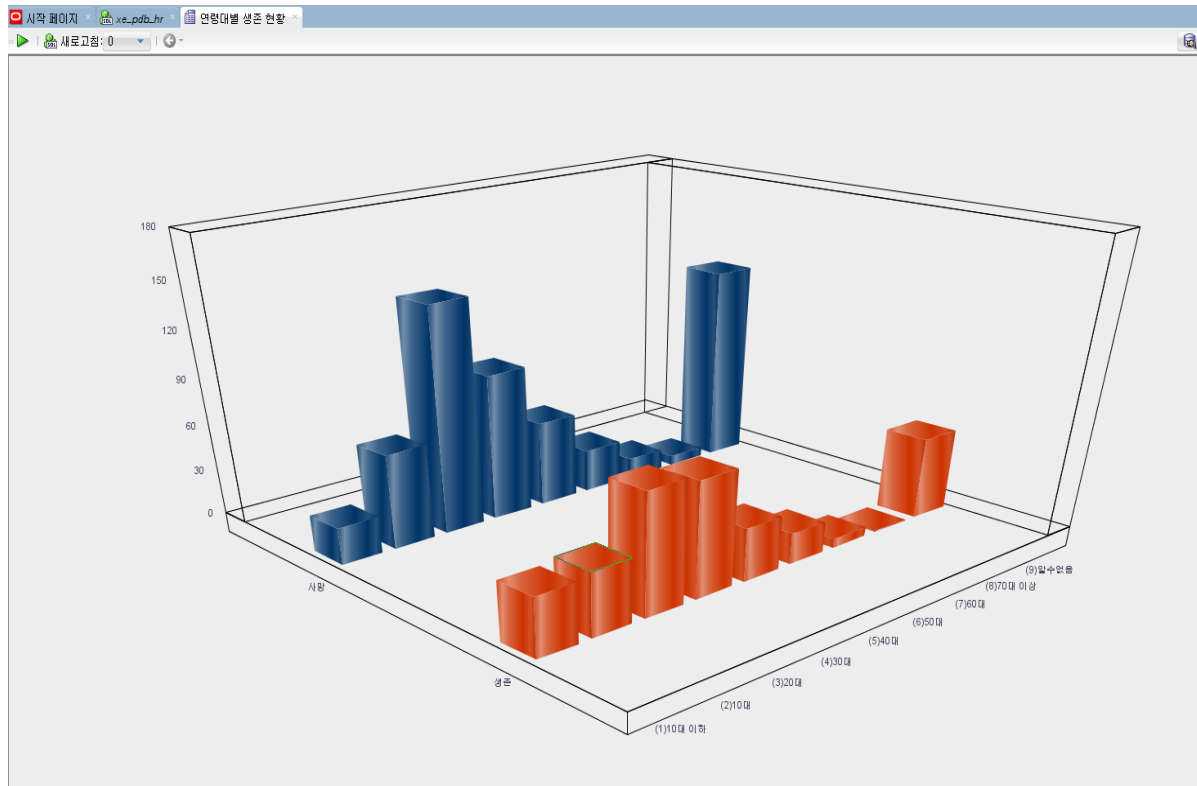
```
SELECT CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
          WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
          WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
          WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
          WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
          WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
          WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
          WHEN age >= 70 THEN '(8) 70대 이상'
          ELSE '(9) 알수없음'
        END ages
, survived, COUNT(*)
FROM titanic2
GROUP BY CASE WHEN age BETWEEN 0 AND 9 THEN '(1) 10대 이하'
          WHEN age BETWEEN 10 AND 19 THEN '(2) 10대'
          WHEN age BETWEEN 20 AND 29 THEN '(3) 20대'
          WHEN age BETWEEN 30 AND 39 THEN '(4) 30대'
          WHEN age BETWEEN 40 AND 49 THEN '(5) 40대'
          WHEN age BETWEEN 50 AND 59 THEN '(6) 50대'
          WHEN age BETWEEN 60 AND 69 THEN '(7) 60대'
          WHEN age >= 70 THEN '(8) 70대 이상'
          ELSE '(9) 알수없음'
        END
, survived
ORDER BY 1, 2;
```



4. Titanic 데이터 분석 시각화

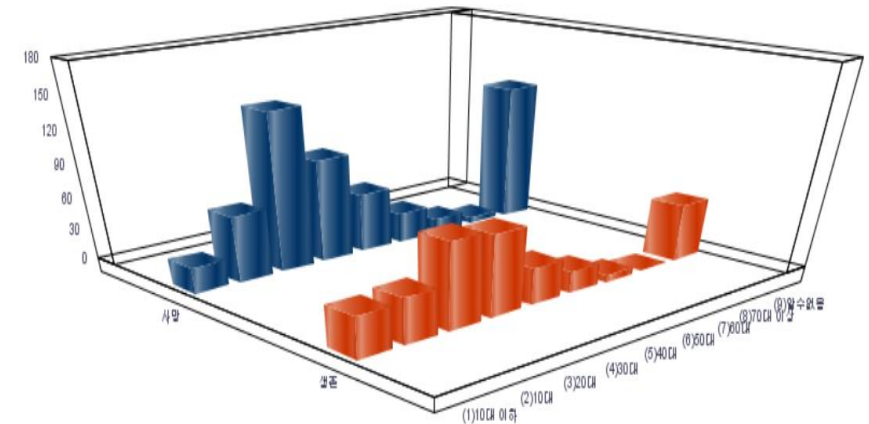
(3) 연령대별 생존/사망자 수 그래프

3D 막대형 그래프



HTML 변환

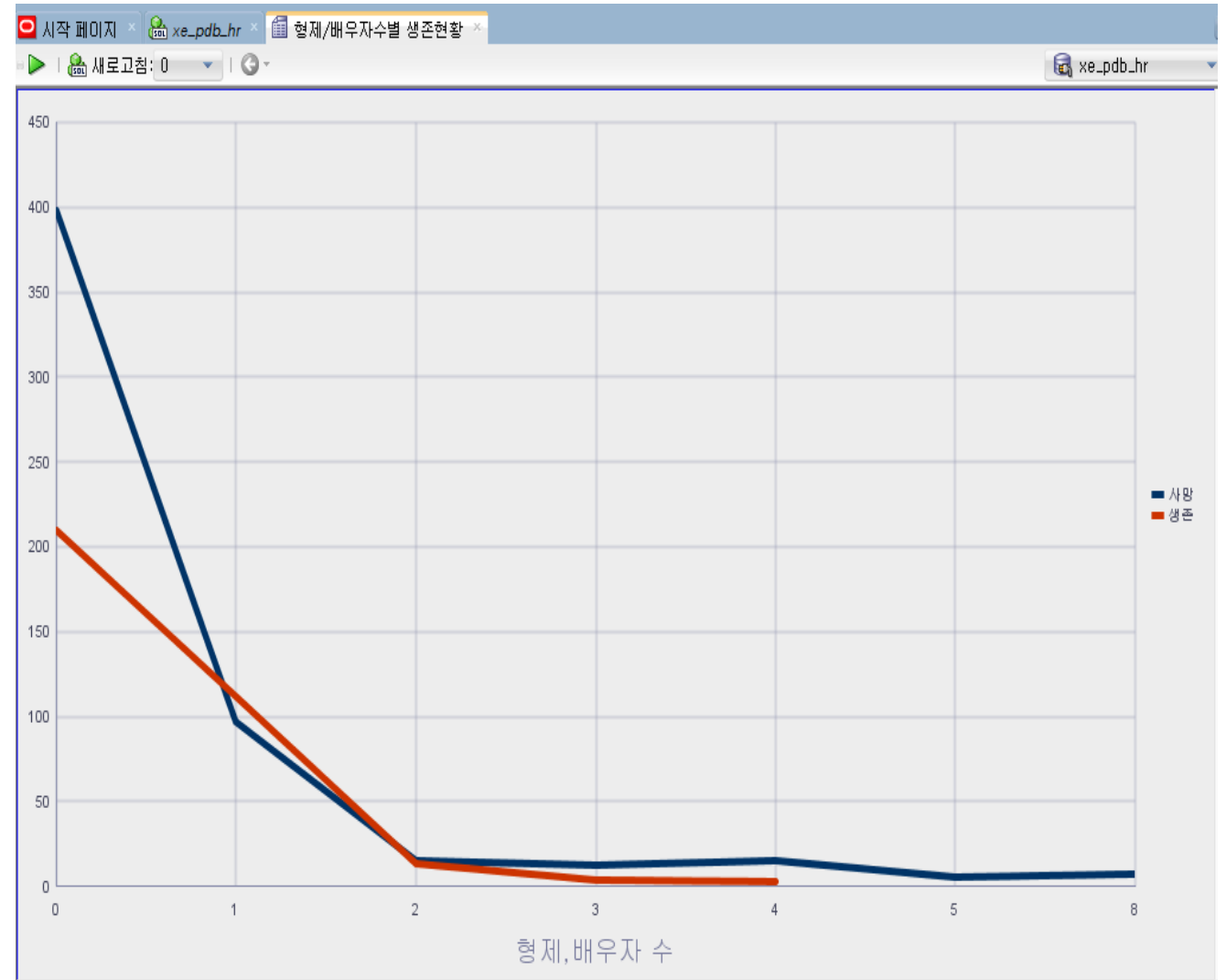
연령대별 생존 현황



4. Titanic 데이터 분석 시각화

(4) 형제/배우자 수 별 생존/사망자 수 그래프

```
SELECT sibsp, survived, count(*)  
FROM titanic2  
GROUP BY sibsp, survived  
ORDER BY 1, 2;
```



4. Titanic 데이터 분석 시각화

(5) 부모/자식 수 별 생존/사망자 수 그래프

```
SELECT parch, survived, count(*)  
FROM titanic2  
GROUP BY parch, survived  
ORDER BY 1, 2;
```

