

주말과제 - 외부데이터 적재

1. 서울시 거주자 중 대기 오염 물질 배출 시설이 많은 지역(구 단위)고객(Customer) 대상 건강 제품 마케팅

데이터 전처리

축소된 데이터를 기준으로 테이블 생성

데이터 적재

데이터 전처리 (결측치 정제)

오염물질 배출 설치 사업장이 많은 지역

고객(Customer) 데이터와 연계하여 마케팅 대상자를 파일로 추출

Customer의 ADDRESS1 형식 변환

고객 데이터와 '경기도 화성시'를 연계하여 마케팅 대상자 추출

최종 결과물(1) - JOIN 사용 X

최종 결과물(2) - JOIN 사용 O

시사점

.csv 파일로 추출

어려웠던 점

전체 SQL 코드

1. 서울시 거주자 중 대기 오염 물질 배출 시설이 많은 지역(구 단위)고객(Customer) 대상 건강 제품 마케팅

주말 과제

주제: 서울시 거주자중

대기오염 물질 배출 시설이 많은
지역(구 단위)고객(Customer) 대상
건강제품 마케팅

기간: 04/07~04/09

- ① 공공 개방 데이터
서울시 빅데이터 캠퍼스 <https://bigdata.seoul.go.kr>
공공 데이터 포털 <https://www.data.go.kr>
- ② 데이터 수집 : 대기오염물질배출설치 사업장
→ 대기오염물질배출설치사업장.xlsx
- ③ Excel 에서 데이터 관찰후 데이터 축소(Data Reduction)
* 고객 데이터(Customer)와 [연결 고리를 검토](#)
* 축소 대상 ??
- ④ 축소된 데이터를 기준으로 테이블 생성 [\[각자 공부 해서\]](#)
- ⑤ 데이터 적재(Loading)
- ⑥ [\[Optional\]](#)데이터 전처리(축소/변환/정제/통합) , 데이터 정제 (불일치, 결측치, 이상치, 잡음)
- ⑦ 오염물질 배출 설치 사업장이 가장 많은 구 조사(ex 동작구)
- ⑧ 고객(Customer) 데이터와 [연계\(Join\)](#)하여 마케팅 대상자를 [파일로 추출](#)하여 제출,
[임의의 주제로 분석하여 데이터 시각화 결과 및 SQL 제출 \[시각화는 기존 프로젝트 활용\]](#)

대상	SQL
기존 테이블 복사	<pre>CREATE TABLE DEPT_BK1 AS SELECT * FROM DEPT WHERE 1=0 ; CREATE TABLE DEPT_BK2 AS SELECT DEPTNO,LOC FROM DEPT WHERE DEPTNO IN (10,20); CREATE TABLE DEPT_BK3 AS SELECT DEPTNO,DNAME ' ' LOC AS NAME_LOC FROM DEPT WHERE DEPTNO IN (10,20);</pre>
컬럼 추가	<pre>ALTER TABLE DEPT_BK2 ADD (TELNO VARCHAR2(12)); ALTER TABLE DEPT_BK2 ADD (ADDR VARCHAR2(100) DEFAULT 'SEOUL'); CREATE TABLE DEPT_BK3 AS SELECT DEPTNO, LOC, DEPTNO ' ' LOC AS D_LOC FROM DEPT_BK2;</pre>
컬럼 길이 변경	<pre>ALTER TABLE DEPT_BK2 MODIFY (TELNO VARCHAR2(2)); DESC DEPT_BK2; ALTER TABLE DEPT_BK2 MODIFY (TELNO VARCHAR2(22)); ALTER TABLE DEPT_BK2 MODIFY (ADDR VARCHAR2(30)); ALTER TABLE DEPT_BK2 MODIFY (ADDR VARCHAR2(4)); ALTER TABLE DEPT_BK2 MODIFY (ADDR VARCHAR2(22));</pre>
컬럼 삭제	<pre>ALTER TABLE DEPT_BK2 DROP COLUMN TELNO; ALTER TABLE DEPT_BK2 DROP COLUMN ADDR;</pre>
대상	SQL
제약사항 추가	<pre>ALTER TABLE DEPT ADD CONSTRAINT DEPT_DEPTNO_PK PRIMARY KEY(DEPTNO); ** 문제 해결후 P.K 추가 !!!! ALTER TABLE DEPT DISABLE PRIMARY KEY; ALTER TABLE DEPT DISABLE CONSTRAINT DEPT_DEPTNO_PK ; ALTER TABLE DEPT ENABLE CONSTRAINT DEPT_DEPTNO_PK ;</pre>

데이터 전처리

	1	2	3
1	소재지전체주소	도로명전체주소	상세영업상태
2	서울특별시 종로구 수송동 156번지	서울특별시 종로구 삼봉로 71 (수송동, G타워)	영업
3	서울특별시 종로구 무악동 66-3번지	서울특별시 종로구 통일로16길 4-1 (무악동, 올리비움산후조리원)	영업
4	서울특별시 종로구 종로6가 289-3번지	서울특별시 종로구 청계천로 279 (종로6가)	영업
5	서울특별시 종로구 공평동 5-1번지	서울특별시 종로구 우정국로 26 (공평동)	영업
6		서울특별시 종로구 새문안로 50 (신문로2가, 시티빌딩)	영업
7		서울특별시 종로구 북촌로 15 (재동, 헌법재판소)	영업
8	서울특별시 종로구 계동 140-2번지	서울특별시 종로구 율곡로 75 (계동, 현대건설빌딩)	영업
9	서울특별시 종로구 삼청동 25-23번지	서울특별시 종로구 북촌로 112 (삼청동)	영업
10		서울특별시 종로구 세종대로 163 (세종로, 현대해상빌딩)	영업
11	서울특별시 종로구 신문로1가 238번지	서울특별시 종로구 새문안로3길 12 (신문로1가)	영업
12	서울특별시 종로구 서린동 99번지	서울특별시 종로구 종로 26 (서린동, SK빌딩)	영업
13	서울특별시 종로구 윤니동 98-78번지	서울특별시 종로구 율곡로 84 (윤니동, 가든타워)	영업
14		서울특별시 종로구 율곡로 88 (윤니동, 삼환빌딩)	영업
15	서울특별시 종로구 종로4가 5번지	서울특별시 종로구 종로 199 (종로4가, 한일빌딩)	영업
16	서울특별시 종로구 수송동 146-2번지	서울특별시 종로구 삼봉로 50 (수송동, 종로구청)	영업
17	서울특별시 종로구 수송동 83-1번지	서울특별시 종로구 율곡로2길 19 (수송동, 수송타워빌딩)	영업
18	서울특별시 종로구 내자동 219번지	서울특별시 종로구 새문안로5가길 33 (내자동)	영업
19	서울특별시 종로구 수송동	서울특별시 종로구 종로1길 28 (수송동, 종로소방서)	영업
20	서울특별시 종로구 소격동 165번지	서울특별시 종로구 삼청로 30 (소격동)	영업
21		서울특별시 종로구 새문안로5길 31 (도렴동, 센터포인트 광화문)	영업
22	서울특별시 종로구 평동 149-1번지	서울특별시 종로구 새문안로 27 (평동)	영업
23	서울특별시 종로구 세종로 1번지	서울특별시 종로구 청와대로 1 (세종로)	영업
24	서울특별시 종로구 서린동	서울특별시 종로구 종로 6 (서린동, 광화문우체국)	영업
25	서울특별시 종로구 동숭동	서울특별시 종로구 대학로 86 (동숭동, 한국방송통신대학교)	영업
26	서울특별시 종로구 낙원동	서울특별시 종로구 인사동4길 18 (낙원동, 로담코 인사빌딩)	영업
27	서울특별시 종로구 관훈동 198-42번지	서울특별시 종로구 인사동5길 38 (관훈동)	영업
28	서울특별시 종로구 구기동 105-1		영업
29		서울특별시 종로구 종로5길 86 (수송동, 국세청)	영업



엑셀 파일 내,

소재지전체주소, 도로명전체주소, 상세영업상태명 컬럼 만을 남기고 모든 컬럼을 삭제했다.

		재개업
	서울특별시 중구 퇴계로 65 (회현동1가, 주차장)	폐쇄
17	서울특별시 중구 소월로2길 12, 제일빌딩 (남대문로5가)	폐쇄
18	서울특별시 중구 소월로 3, 롯데손해보험빌딩 (남창동)	폐쇄
	서울특별시 중구 을지로 51 (을지로2가, 내외빌딩)	폐쇄
	서울특별시 중구 퇴계로 334 (광희동2가)	폐쇄
	서울특별시 중구 퇴계로 307 (광희동1가, 광희빌딩)	폐쇄
	서울특별시 중구 청계천로 54 (수표동)	폐쇄
8		폐쇄
	서울특별시 성동구 성수이로7길 28 (성수동2가, 대성ENG)	폐쇄
	서울특별시 성동구 성수이로7길 28 (성수동2가, 대성ENG)	폐쇄
	서울특별시 성동구 고산자로 270 (행당동, 성동구청)	폐쇄
	서울특별시 성동구 마장로 210 (홍익동, 재단법인한국기원)	폐쇄
	서울특별시 성동구 연무장길 97-1 (성수동2가)	폐쇄
	서울특별시 성동구 왕십리로 342 (도선동, 교보생명빌딩)	폐쇄
	서울특별시 성동구 한림말길 43 (옥수동, 옥정초등학교)	폐쇄
	서울특별시 성동구 한림로 555 (행당동)	폐쇄



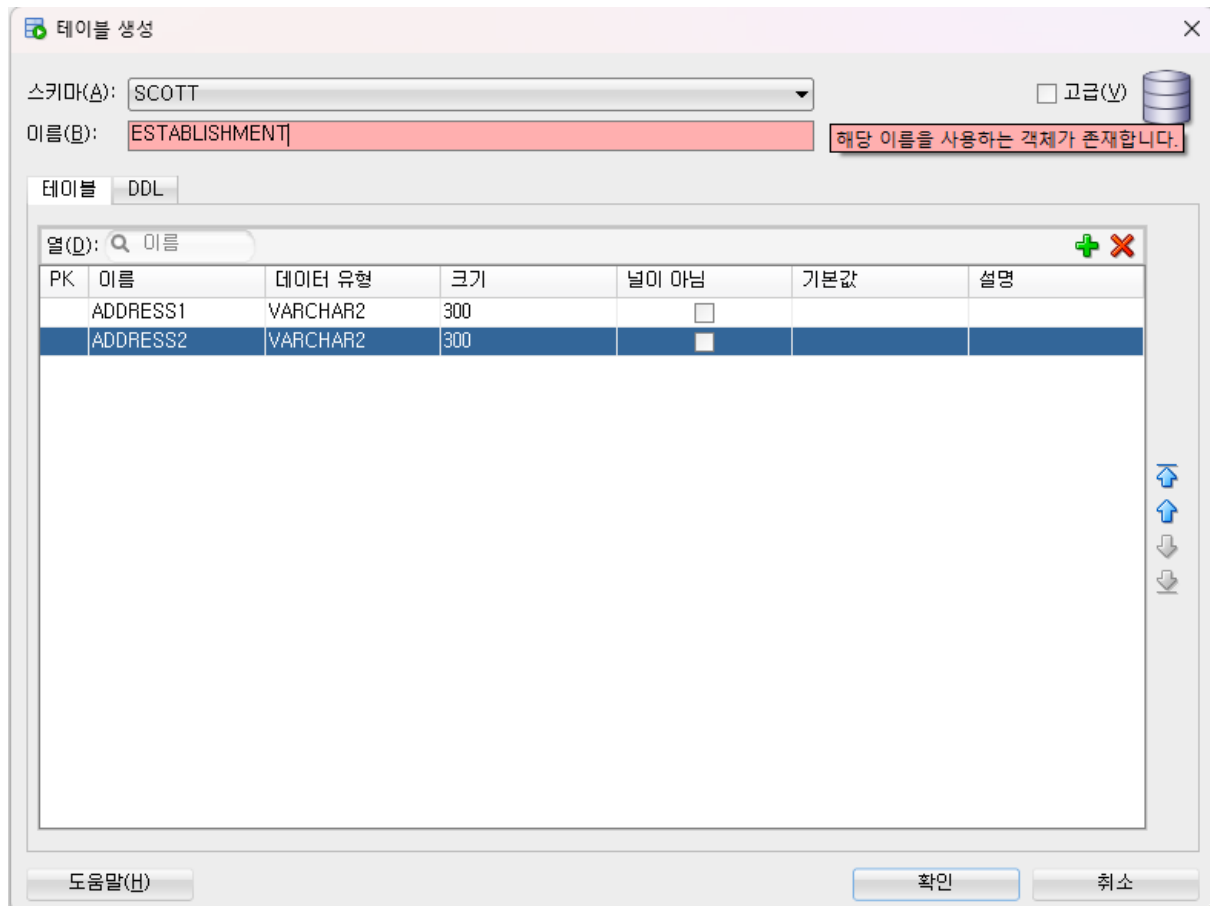
엑셀 - 필터 기능을 통해 정렬 후 폐쇄, 폐업, 휴업에 해당하는 레코드를 모두 삭제했다.

1	소재지전체주소	도로명전체주소
2	서울특별시 종로구 수송동 156번지	서울특별시 종로구 삼봉로 71 (수송동, G타워)
3	서울특별시 종로구 무악동 66-3번지	서울특별시 종로구 통일로16길 4-1 (무악동, 올리비움산후조리원)
4	서울특별시 종로구 종로6가 289-3번지	서울특별시 종로구 청계천로 279 (종로6가)
5	서울특별시 종로구 공평동 5-1번지	서울특별시 종로구 우정국로 26 (공평동)
6		서울특별시 종로구 새문안로 50 (신문로2가, 시티빌딩)
7		서울특별시 종로구 북촌로 15 (재동, 헌법재판소)
8	서울특별시 종로구 계동 140-2번지	서울특별시 종로구 율곡로 75 (계동, 현대건설빌딩)
9	서울특별시 종로구 삼청동 25-23번지	서울특별시 종로구 북촌로 112 (삼청동)
10		서울특별시 종로구 세종대로 163 (세종로, 현대해상빌딩)
11	서울특별시 종로구 신문로1가 238번지	서울특별시 종로구 새문안로3길 12 (신문로1가)
12	서울특별시 종로구 서린동 99번지	서울특별시 종로구 종로 26 (서린동, SK빌딩)
13	서울특별시 종로구 윤니동 98-78번지	서울특별시 종로구 율곡로 84 (윤니동, 가든타워)
14		서울특별시 종로구 율곡로 88 (윤니동, 삼환빌딩)
15	서울특별시 종로구 종로4가 5번지	서울특별시 종로구 종로 199 (종로4가, 한일빌딩)
16	서울특별시 종로구 수송동 146-2번지	서울특별시 종로구 삼봉로 50 (수송동, 종로구청)
17	서울특별시 종로구 수송동 83-1번지	서울특별시 종로구 율곡로2길 19 (수송동, 수송타워빌딩)
18	서울특별시 종로구 내자동 219번지	서울특별시 종로구 새문안로5가길 33 (내자동)
19	서울특별시 종로구 수송동	서울특별시 종로구 종로1길 28 (수송동, 종로소방서)
20	서울특별시 종로구 소격동 165번지	서울특별시 종로구 삼청로 30 (소격동)
21		서울특별시 종로구 새문안로5길 31 (도렴동, 센터포인트 광화문)
22	서울특별시 종로구 평동 149-1번지	서울특별시 종로구 새문안로 27 (평동)
23	서울특별시 종로구 세종로 1번지	서울특별시 종로구 청와대로 1 (세종로)
24	서울특별시 종로구 서린동	서울특별시 종로구 종로 6 (서린동, 광화문우체국)
25	서울특별시 종로구 동숭동	서울특별시 종로구 대학로 86 (동숭동, 한국방송통신대학교)
26	서울특별시 종로구 낙원동	서울특별시 종로구 인사동4길 18 (낙원동, 로담코 인사빌딩)
27	서울특별시 종로구 관훈동 198-42번지	서울특별시 종로구 인사동5길 38 (관훈동)
28	서울특별시 종로구 구기동 105-1	
29		서울특별시 종로구 종로5길 86 (수송동, 국세청)
30	서울특별시 종로구 종로1가	서울특별시 종로구 종로 19 (종로1가, 르메이에르종로타운)
31	서울특별시 종로구 청진동 246번지	서울특별시 종로구 종로3길 17 (청진동, D타워)

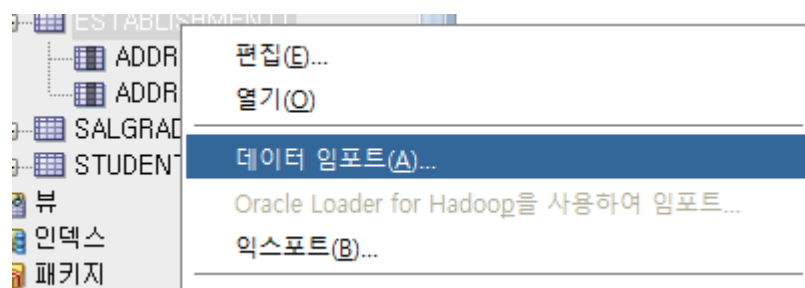


그 후 상세영업상태 컬럼 또한 삭제했다.

축소된 데이터를 기준으로 테이블 생성



데이터 적재



해당 테이블 우클릭 - 데이터 импорт

데이터 임포트 마법사 - 단계 1/5

데이터 미리보기

데이터 미리보기

임포트 방식

열 선택

열 정의

완료

소스: 로컬 파일

파일: C:\Users\Wudej\Desktop\대기오염물질배출설치사업장.xlsx

찾아보기(R)...

파일 형식

☒ 헤더 행 건너뛰기(K): 0

형식(F): excel 95-2003 (*.xls) ☒ 미리보기 행 제한(B): 100

파일 내용

소재지전...	도로명전...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...

도움말(H) < 뒤로(B) 다음(N) > 완료(F) 취소



찾아보기 - 엑셀 파일 선택

데이터 임포트 마법사 - 단계 2/4

임포트 방식

데이터 미리보기

임포트 방식

열 정의

완료

데이터 임포트 방식을 지정하십시오. 스테이징 외부 테이블 방식의 경우 대상 테이블을 임포트할 수 있도록 외부 테이블이 스테이징 테이블로 생성됩니다. 기타 임포트 방식의 경우 데이터가 테이블로 바로 임포트됩니다.

임포트 방식: 삽입

☐ SQL 워크시트로 생성 스크립트 전송

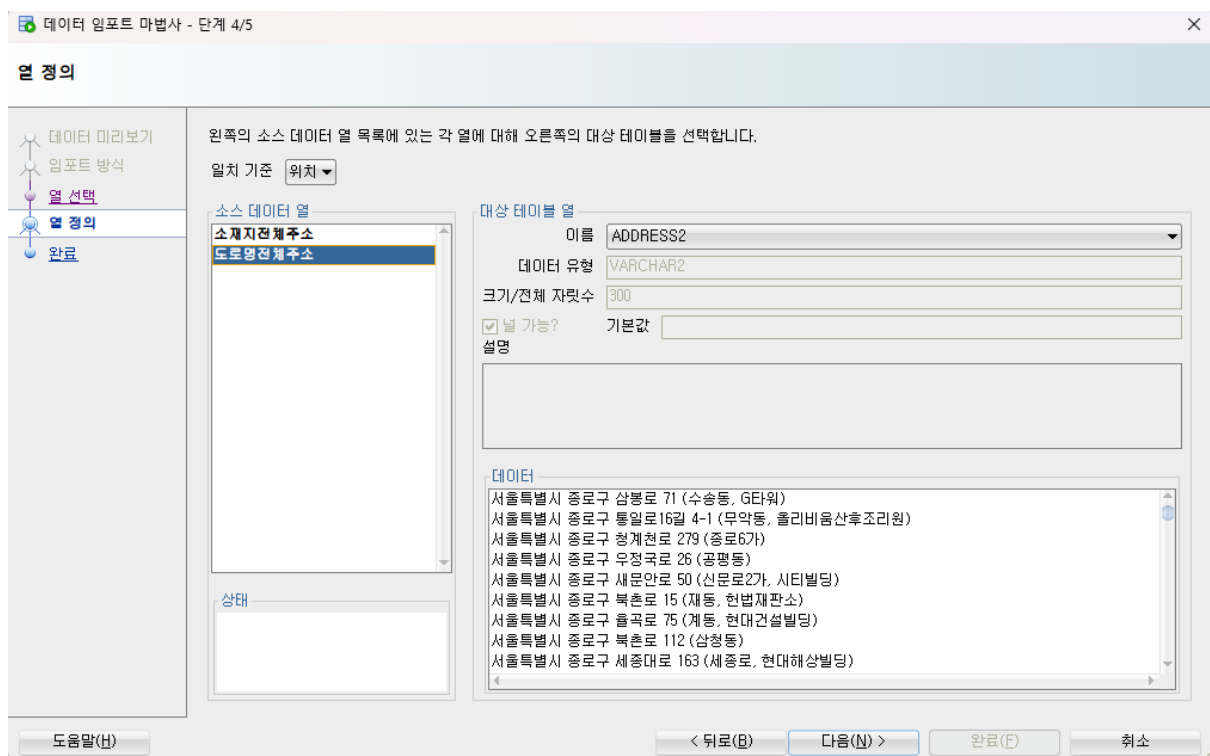
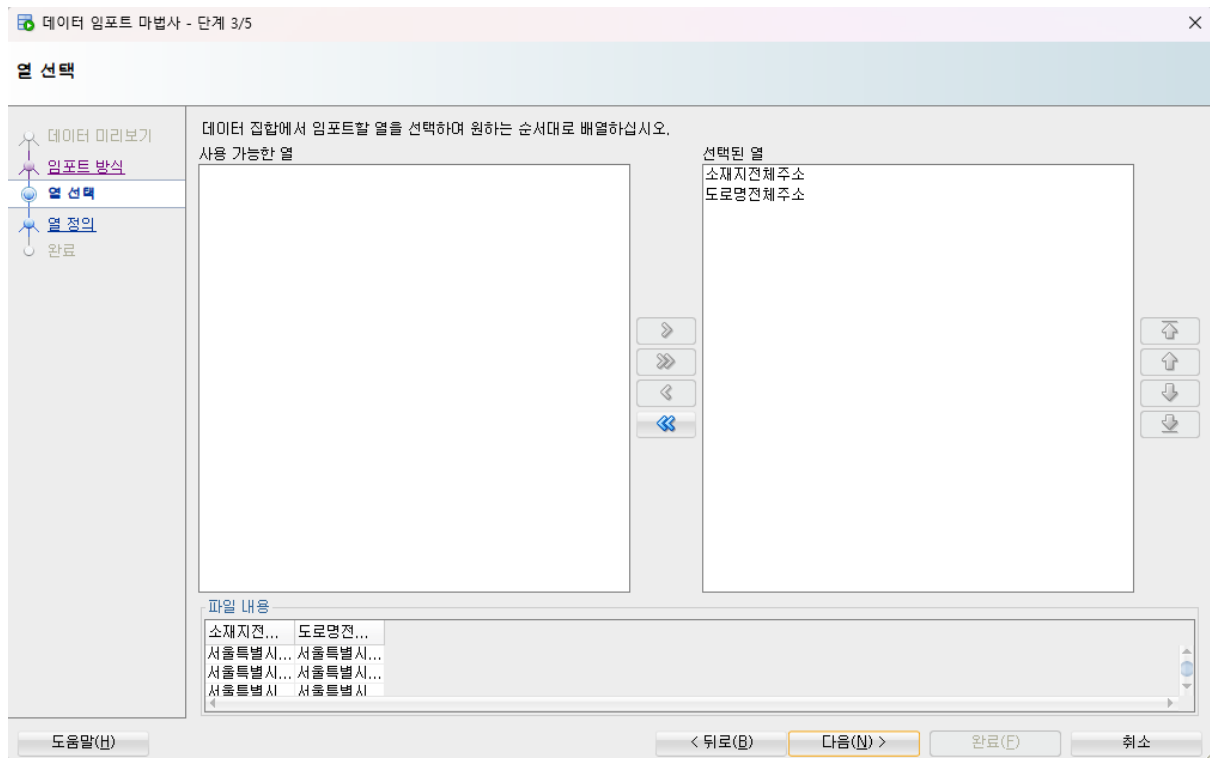
테이블 이름: ESTABLISHMENT1

☐ 임포트 행 제한(I): 100

파일 내용

소재지전...	도로명전...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...
서울특별시...	서울특별시...

도움말(H) < 뒤로(B) 다음(N) > 완료(F) 취소



```
SELECT * FROM ESTABLISHMENT;
```

스크립트 출력 x 질의 결과 x	
SQL 50개의 행이 인출됨(0.022초)	
ADDRESS1	ADDRESS2
1 서울특별시 종로구 관수동 130-2번지	서울특별시 종로구 수표로18가길 10 (관수동)
2 서울특별시 종로구 관수동 140번지	서울특별시 종로구 수표로18가길 6-1, 4층 (관수동)
3 서울특별시 종로구 관수동 113-1번지	(null)
4 서울특별시 종로구 청진동 119-1번지	서울특별시 종로구 종로 33 (청진동)
5 (null)	서울특별시 종로구 새문안로 29 (평동, 강북삼성병원)
6 서울특별시 종로구 평동 164번지	서울특별시 종로구 새문안로 9 (평동, 적십자병원)
7 서울특별시 종로구 종로6가 261-1번지	서울특별시 종로구 종로 266 (종로6가)
8 서울특별시 종로구 서린동 33번지 영풍빌딩	서울특별시 종로구 청계천로 41 (서린동, 영풍빌딩)
9 서울특별시 종로구 명륜3가 53번지	서울특별시 종로구 성균관로 25-2 (명륜3가)
10 서울특별시 종로구 도림동 60번지	서울특별시 종로구 새문안로5길 37 (도림동)
11 서울특별시 종로구 인사동 194-4번지	서울특별시 종로구 인사동5길 25-1008 (인사동)
12 서울특별시 종로구 서린동 136번지 한국무역보험공사빌딩	서울특별시 종로구 종로 14 (서린동, 한국무역보험공사빌딩)
13 서울특별시 종로구 인의동 112-14번지	서울특별시 종로구 창경궁로 109 (인의동, 세운스퀘어)
14 서울특별시 종로구 화동 1번지 정독도서관	서울특별시 종로구 북촌로5길 48 (화동, 정독도서관)
15 서울특별시 종로구 신문로2가 2-1번지	서울특별시 종로구 새문안로 55 (신문로2가)
16 서울특별시 종로구 신문로1가 163번지	서울특별시 종로구 새문안로 92 (신문로1가)



데이터 임포트를 정상적으로 성공했다.

데이터 전처리 (결측치 정제)

ADDRESS1	ADDRESS2
1 서울특별시 종로구 관수동 130-2번지	서울특별시 종로구 수표로18가길 10 (관수동)
2 서울특별시 종로구 관수동 140번지	서울특별시 종로구 수표로18가길 6-1, 4층 (관수동)
3 서울특별시 종로구 관수동 113-1번지	(null)
4 서울특별시 종로구 청진동 119-1번지	서울특별시 종로구 종로 33 (청진동)
5 (null)	서울특별시 종로구 새문안로 29 (평동, 강북삼성병원)
6 서울특별시 종로구 평동 164번지	서울특별시 종로구 새문안로 9 (평동, 적십자병원)



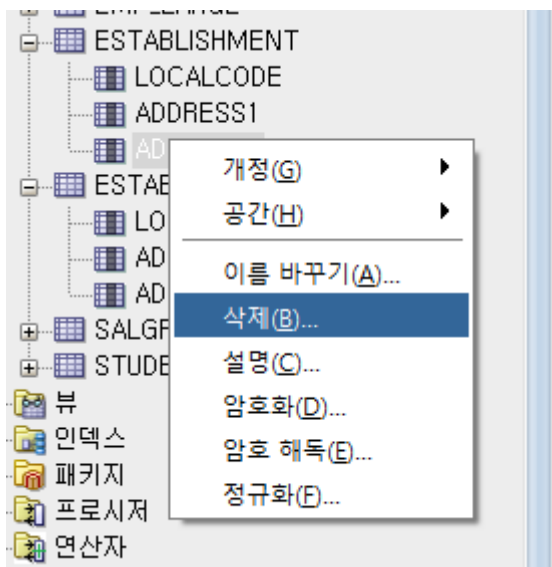
초기 도로명주소와, 행정동 주소간 null (결측치)가 있다.

```
-- ADDRESS1 결측치 - 주소2로 대체하기 --
UPDATE ESTABLISHMENT SET ADDRESS1 = ADDRESS2 WHERE ADDRESS1 is null;
SELECT * FROM ESTABLISHMENT;
```

스크립트 출력 x 질의 결과 x	
SQL 50개의 행이 인출됨(0.006초)	
ADDRESS1	ADDRESS2
1 서울특별시 종로구 관수동 130-2번지	서울특별시 종로구 수표로18가길 10 (관수동)
2 서울특별시 종로구 관수동 140번지	서울특별시 종로구 수표로18가길 6-1, 4층 (관수동)
3 서울특별시 종로구 관수동 113-1번지	(null)
4 서울특별시 종로구 청진동 119-1번지	서울특별시 종로구 종로 33 (청진동)
5 서울특별시 종로구 새문안로 29 (평동, 강북삼성병원)	서울특별시 종로구 새문안로 29 (평동, 강북삼성병원)
6 서울특별시 종로구 평동 164번지	서울특별시 종로구 새문안로 9 (평동, 적십자병원)
7 서울특별시 종로구 종로6가 261-1번지	서울특별시 종로구 종로 266 (종로6가)
8 서울특별시 종로구 서린동 33번지 영풍빌딩	서울특별시 종로구 청계천로 41 (서린동, 영풍빌딩)
9 서울특별시 종로구 명륜3가 53번지	서울특별시 종로구 성균관로 25-2 (명륜3가)
10 서울특별시 종로구 도렴동 60번지	서울특별시 종로구 새문안로5길 37 (도렴동)
11 서울특별시 종로구 인사동 194-4번지	서울특별시 종로구 인사동5길 25-1008 (인사동)
12 서울특별시 종로구 서린동 136번지 한국무역보험공사빌딩	서울특별시 종로구 종로 14 (서린동, 한국무역보험공사빌딩)
13 서울특별시 종로구 인의동 112-14번지	서울특별시 종로구 창경궁로 109 (인의동, 세운스퀘어)
14 서울특별시 종로구 화동 1번지 정독도서관	서울특별시 종로구 북촌로5길 48 (화동, 정독도서관)
15 서울특별시 종로구 신문로2가 2-1번지	서울특별시 종로구 새문안로 55 (신문로2가)
16 서울특별시 종로구 신문로1가 163번지	서울특별시 종로구 새문안로 92 (신문로1가)
17 서울특별시 종로구 사직동 284-1번지 종로문화체육센터	서울특별시 종로구 인왕산로1길 21, 종로문화체육센터 (사직동)



ADDRESS1의 결측치가 있다면, ADDRESS2로 대체하는 방식으로 결측치를 정제했다.



또한 ADDRESS2는 필요 없으므로, 해당 컬럼을 삭제했다.

추가 X

프로그래밍 SQL

소유자 SCOTT

이름 ESTABLISHMENT1

열 이름 ADDRESS

데이터 유형 VARCHAR2

전체 자릿수 300

소수점 이하 자릿수

도움말(H) 적용(A) 취소



주소를 00도(시) 00구 로 분리하기 위한 새로운 ADDRESS 컬럼 을 만들었다.

```
-- 공백 2개 기준 주소 update
UPDATE ESTABLISHMENT SET ADDRESS = substr(address, 0, instr(address, ' ', 1, 2));
commit;

select * from establishment;
```

ADDRESS1	ADDRESS
1 서울특별시 종로구 관수동 130-2번지	서울특별시 종로구
2 서울특별시 종로구 관수동 113-1번지	서울특별시 종로구
3 서울특별시 종로구 청진동 119-1번지	서울특별시 종로구
4 서울특별시 종로구 새문안로 29 (평동, 강북삼성병원)	서울특별시 종로구
5 서울특별시 종로구 평동 164번지	서울특별시 종로구
6 서울특별시 종로구 서린동 33번지 영풍빌딩	서울특별시 종로구
7 서울특별시 종로구 명륜3가 53번지	서울특별시 종로구
8 서울특별시 종로구 도림동 60번지	서울특별시 종로구
9 서울특별시 종로구 서린동 136번지 한국무역보험공사빌딩	서울특별시 종로구
10 서울특별시 종로구 인의동 112-14번지	서울특별시 종로구
11 서울특별시 종로구 화동 1번지 정독도서관	서울특별시 종로구
12 서울특별시 종로구 신문로2가 2-1번지	서울특별시 종로구
13 서울특별시 종로구 사직동 284-1번지 종로문화체육센터	서울특별시 종로구
14 서울특별시 종로구 관훈동 155-2번지	서울특별시 종로구
15 서울특별시 종로구 연건동 28-21번지	서울특별시 종로구
16 서울특별시 종로구 인의동 112-1번지 더케이손해보험빌딩	서울특별시 종로구



위 코드와 같이 ADDRESS 컬럼에 공백 2개를 기준으로 주소를 나누어 원하는 형태인,

00시 00구 형태로 저장했다.

오염물질 배출 설치 사업장이 많은 지역

```
-- 시설 구 단위별 대기오염 시설이 가장 많은 지역 확인 -> 경기도 화성시
SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
FROM ESTABLISHMENT1
GROUP BY ADDRESS
HAVING COUNT(ADDRESS) > 10 -- 잡음 제거
ORDER BY 시설수 DESC;
```

주소	시설수
1 경기도 화성시	2673
2 인천광역시 서구	2419
3 경기도 김포시	2273
4 경상남도 김해시	2185
5 경기도 포천시	2156
6 인천광역시 남동구	2042
7 경기도 용인시	1683
8 경기도 안산시단원구	1664
9 충청북도 음성군	1507
10 경상남도 창원시	1428
11 경기도 평택시	1421
12 경기도 파주시	1412
13 경기도 안산시	1398
14 충청남도 천안시	1313
15 경기도 시흥시	1283
16 경기도 양주시	1264



위와 같이 경기도 화성시가 가장 오염물질 배출 설치 사업장이 많은 것을 볼 수 있었다.

*** 화성시의 행정 구역은 4읍 9면 15동으로, [구]가 없는 [시]이기 때문에 화성시를 오염물질 배출 설치 사업장이 가장 많은 지역으로 채택했다.**

고객(Customer) 데이터와 연계하여 마케팅 대상자를 파일로 추출

Customer의 ADDRESS1 형식 변환

ID	PWD	NAME	ZIPCODE	ADDRESS1	
1	00000098	ctzytjnxuo	정은옥	542-800	전남 구례군 구례읍
2	00000099	ctfbrzatu	김오남	513-832	전남 영광군 불갑면
3	00000100	gqqapxoowk	천승권	703-045	대구 서구 비산5동
4	00000101	bgjlmnrqj	전대백	210-939	강원 강릉시 포남2동
5	00000102	fqvfwzhsq	구건신	670-823	경남 거창군 웅양면
6	00000103	dthmpryhzz	김범결	621-921	경남 김해시 외동
7	00000104	ydaadwyxlt	라석훈	791-711	경북 포항시 북구 흥해읍
8	00000105	oxkazsbjrh	장석승	312-807	충남 금산군 금산읍
9	00000106	qmvwtrdxxq	이성혜	437-090	경기 의왕시 학의동
10	00000107	arvmdnoktr	최나리	355-884	충남 보령시 남포면
11	00000108	teptdaphle	이익안	412-802	경기 고양시 덕양구 판산동
12	00000109	vhojknwvta	김충구	137-732	서울 서초구 서초3동
13	00000110	kloangjxdr	조동유	210-957	강원 강릉시 포남2동
14	00000111	mghiaokdrg	길유인	422-823	경기 부천시 소사구 심곡본동
15	00000112	bvekawhfnl	문설선	601-803	부산 동구 범일2동
16	00000113	pmzvrshxn	김초은	340-931	충남 예산군 고덕면
17	00000114	pwrmhfbldpx	이원국	120-101	서울 서대문구 홍은1동
18	00000115	bkeoajcypp	김연후	503-837	광주 남구 주월2동
19	00000116	gcwtmfmbfm	라채화	502-826	광주 서구 양3동

```

SELECT NAME, MOBILE_NO, EMAIL,
DECODE(SUBSTR(TMP, 0, 2), '서울', '서울특별시',
      '울산', '울산광역시',
      '부산', '부산광역시',
      '대전', '대전광역시',
      '인천', '인천광역시',
      '대구', '대구광역시',
      '광주', '광주광역시',
      '세종', '세종특별자치시',
      '경기', '경기도',
      '강원', '강원도',
      '전남', '전라남도',
      '전북', '전라북도',
      '충남', '충청남도',
      '충북', '충청북도',
      '경남', '경상남도',
      '경북', '경상북도',

```

```

        '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM (
    SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
    FROM CUSTOMER
);

```

	NAME	MOBILE_NO	EMAIL	ADDRESS
1	정은옥	01087685618	wecqdy@kopo.ac.kr	전라남도 구례군
2	김오남	01080545398	thatkcuy@kopo.ac.kr	전라남도 영광군
3	천승권	01063785330	saidwmgv@korea.com	대구광역시 서구
4	전대백	01021523450	oneyydu@dreamwiz.com	강원도 강릉시
5	구건신	01049940680	inswrd@yahoo.com	경상남도 거창군
6	김범결	01992053820	ofbevz@yahoo.com	경상남도 김해시
7	라석훈	01089322713	wordphwv@kopo.ac.kr	경상북도 포항시
8	장석승	01021279238	buttydt@lg.com	충청남도 금산군
9	이성혜	01083615334	oneemgb@naver.com	경기도 의왕시
10	최나리	01176167438	saidvjys@korea.com	충청남도 보령시
11	이익안	01930271389	invtp@naver.com	경기도 고양시
12	김충구	01148933860	atacgx@naver.com	서울특별시 서초구
13	조동유	01028906950	wehqsi@daum.net	강원도 강릉시
14	길유인	01154978503	forqion@kopo.ac.kr	경기도 부천시
15	문설선	01668186529	eachqcbi@daum.net	부산광역시 동구
16	김초은	01059505815	whatlmsg@yahoo.com	충청남도 예산군
17	이원국	01094826225	werexibv@aol.com	서울특별시 서대문구
18	김연후	01154436019	atemej@gmail.com	광주광역시 남구
19	라채화	01028378602	whenpizj@korea.com	광주광역시 서구
20	손필선	01054235165	whichsxlx@cau.ac.kr	충청북도 단양군



Customer의 주소값은 시/도가 fullname이 아닌, 두글자로 잘려있다.

따라서, ESTABLISHMENT 테이블과 Join 하기 위해 `DECODE()` 함수를 활용하여, CUSTOMER의 주소 형식을 변환해주었다.

고객 데이터와 '경기도 화성시'를 연계하여 마케팅 대상자 추출

최종 결과물(1) - JOIN 사용 X

```

-- 최종 결과물(1) --
SELECT NAME, MOBILE_NO, EMAIL,

```

```

DECODE(SUBSTR(TMP, 0, 2),
      '서울', '서울특별시',
      '울산', '울산광역시',
      '부산', '부산광역시',
      '대전', '대전광역시',
      '인천', '인천광역시',
      '대구', '대구광역시',
      '광주', '광주광역시',
      '세종', '세종특별자치시',
      '경기', '경기도',
      '강원', '강원도',
      '전남', '전라남도',
      '전북', '전라북도',
      '충남', '충청남도',
      '충북', '충청북도',
      '경남', '경상남도',
      '경북', '경상북도',
      '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM(
  SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
  FROM CUSTOMER
)
WHERE DECODE(SUBSTR(TMP, 0, 2), '서울', '서울특별시',
      '울산', '울산광역시',
      '부산', '부산광역시',
      '대전', '대전광역시',
      '인천', '인천광역시',
      '대구', '대구광역시',
      '광주', '광주광역시',
      '세종', '세종특별자치시',
      '경기', '경기도',
      '강원', '강원도',
      '전남', '전라남도',
      '전북', '전라북도',
      '충남', '충청남도',
      '충북', '충청북도',
      '경남', '경상남도',
      '경북', '경상북도',
      '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) =
(
  SELECT 주소
  FROM(
    SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
    FROM ESTABLISHMENT
    GROUP BY ADDRESS
    HAVING COUNT(ADDRESS) > 10
    ORDER BY 시설수 DESC
  )
  FETCH FIRST ROW ONLY
);

```


NAME	MOBILE_NO	EMAIL	ADDRESS
1 합영권	01195831097	ofgtib@kopo.ac.kr	경기도 화성시
2 안관아	01074198378	sheicpp@hotmail.com	경기도 화성시
3 박서나	01087393957	areiswu@gmail.com	경기도 화성시
4 김영삼	01084412317	withrwjq@gmail.com	경기도 화성시
5 위명구	01097249891	agyea@sk.com	경기도 화성시
6 이충범	01646304746	forfjhq@gmail.com	경기도 화성시
7 강영세	01991994899	usejywm@naver.com	경기도 화성시
8 염오신	01046385589	washuco@kopo.ac.kr	경기도 화성시
9 성상기	01099554489	iikas@naver.com	경기도 화성시
10 연필표	01188258343	bexfdo@dreamwiz.com	경기도 화성시
11 공건상	01071353243	wasajkj@yahoo.com	경기도 화성시
12 박청찬	01620198735	sheiwbx@naver.com	경기도 화성시
13 김흠해	01036922476	whenowyc@samsung.com	경기도 화성시
14 권군하	01046441082	whatgmle@yahoo.com	경기도 화성시
15 성동라	01038167770	beckqo@korea.com	경기도 화성시
16 도도연	01643550037	butyqzn@sk.com	경기도 화성시
17 모희해	01165575904	allnftg@naver.com	경기도 화성시
18 윤도승	01623044736	asicka@lg.com	경기도 화성시
19 김국지	01060240444	weavkw@naver.com	경기도 화성시
20 채숙림	01036850147	forewmo@dreamwiz.com	경기도 화성시



JOIN 을 사용하지 않고, **WHERE** 조건과 **서브 쿼리** 를 통해 마케팅 대상자를 추출했다.

최종 결과물(2) - JOIN 사용 O

```
-- 최종 결과물 (2) Inner Join 사용 --
SELECT c.*
FROM (
    SELECT ADDRESS, COUNT(ADDRESS) AS 시설수
    FROM ESTABLISHMENT
    GROUP BY ADDRESS
    HAVING COUNT(ADDRESS) > 10
    ORDER BY 시설수 DESC
    FETCH FIRST ROW ONLY
) e
INNER JOIN (
    SELECT NAME, MOBILE_NO, EMAIL,
    DECODE(SUBSTR(TMP, 0, 2),
    '서울', '서울특별시',
    '울산', '울산광역시',
    '부산', '부산광역시',
```

```

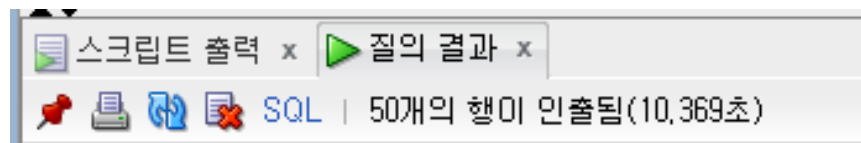
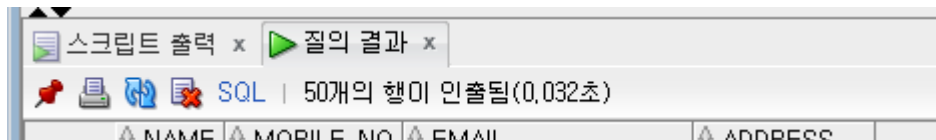
'대전', '대전광역시',
'인천', '인천광역시',
'대구', '대구광역시',
'광주', '광주광역시',
'세종', '세종특별자치시',
'경기', '경기도',
'강원', '강원도',
'전남', '전라남도',
'전북', '전라북도',
'충남', '충청남도',
'충북', '충청북도',
'경남', '경상남도',
'경북', '경상북도',
'제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM (
    SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
    FROM CUSTOMER
)
) c
ON c.ADDRESS = e.ADDRESS;

```



JOIN 을 활용하여 마케팅 대상자를 추출했다.

시사점



JOIN 연산을 사용하지 않고, **WHERE** 와 **서브쿼리** 를 활용한 쿼리가 **JOIN** 을 활용한 쿼리보다 질의 결과 속도가 많이 빠른 것을 볼 수 있었다.

이를 통해 **JOIN** 이 많은 연산을 필요로 함을 체감할 수 있었다.

.csv 파일로 추출

```
set heading on
set feedback off
set echo off
set term off

SPOOL assignment01.csv

SELECT NAME, MOBILE_NO, EMAIL,
DECODE(SUBSTR(TMP, 0, 2),
'서울', '서울특별시',
'울산', '울산광역시',
'부산', '부산광역시',
'대전', '대전광역시',
'인천', '인천광역시',
'대구', '대구광역시',
'광주', '광주광역시',
'세종', '세종특별자치시',
'경기', '경기도',
'강원', '강원도',
'전남', '전라남도',
'전북', '전라북도',
'충남', '충청남도',
'충북', '충청북도',
'경남', '경상남도',
'경북', '경상북도',
'제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM(
SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
FROM CUSTOMER
)
WHERE DECODE(SUBSTR(TMP, 0, 2), '서울', '서울특별시',
'울산', '울산광역시',
'부산', '부산광역시',
'대전', '대전광역시',
'인천', '인천광역시',
'대구', '대구광역시',
'광주', '광주광역시',
'세종', '세종특별자치시',
'경기', '경기도',
'강원', '강원도',
'전남', '전라남도',
'전북', '전라북도',
'충남', '충청남도',
'충북', '충청북도',
'경남', '경상남도',
'경북', '경상북도',
'제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) =
(
SELECT 주소
FROM(
SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
FROM ESTABLISHMENT
GROUP BY ADDRESS
HAVING COUNT(ADDRESS) > 10
```

```

ORDER BY 시설수 DESC
)
FETCH FIRST ROW ONLY
);

SPPOOL OFF

```

```

oracle@Baseline:~
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[oracle@Baseline ~]$ sqlplus scott/tiger

SQL*Plus: Release 12.2.0.1.0 Production on 일 4월 9 22:25:08 2023

Copyright (c) 1982, 2016, Oracle. All rights reserved.

마지막 성공한 로그인 시간: 일 4월 09 2023 21:20:03 +09:00

다음에 접속됨:
Oracle Database 12c Enterprise Edition Release 12.2.0.1.0 - 64bit Production

SCOTT>@assignment01.sql
SCOTT>

```

assignment01.csv

열기(O) 저장(S) ~/

NAME	MOBILE_NO	
EMAIL		
ADDRESS		
함영권	01195831097	ofgtib@kopo.ac.kr
경기도 화성시		
안관아	01074198378	sheicpp@hotmail.com
경기도 화성시		
박서나	01087393957	areiswu@gmail.com
경기도 화성시		
김영삼	01084412317	withrwjq@gmail.com
경기도 화성시		
위명구	01097249891	agyea@sk.com
경기도 화성시		
이충범	01646304746	forfjhq@gmail.com
경기도 화성시		
강영세	01991994899	usejywm@naver.com
경기도 화성시		

CSV 탭 너비: 8 1행 1열



SPPOOL 명령어를 통해 마케팅 대상자 쿼리 결과를 **CSV파일로 저장했다.**

어려웠던 점

	A	B	C	D
61924	100609	6310000	울산광역시 울주군 삼남면 가천리 8	울산광역시 울주군 삼남면 가천지내길 21-3
61925	100610	6410000	경기도 화성시 장안면 금의리 장안침	
61926	100611	6410000	경기도 부천시원미구 역곡동 24-39	
61927	100612	6410000	경기도 부천시원미구 역곡동 24번지	
61928	100613	6410000	경기도 부천시원미구 역곡동 24번지	경기도 부천시 원미구 부일로815번길 94 (역곡동)
61929	100614	6410000	경기도 시흥시 정왕동 2173-1번지	경기도 시흥시 옥구천동로 16, 0호 (정왕동,시흥스마트허브2바808호)
61930	100615	6410000	경기도 시흥시 정왕동 1377-11단지	
61931	100616	6410000	경기도 시흥시 정왕동 1271-14 시회	
61932	100617	6410000	경기도 안산시단원구 원시동 739번	경기도 안산시 단원구 산단로 63 (원시동)
61933	100618	6410000	경기도 의정부시 장암동 산 5	
61934	100619	6410000	경기도 안산시단원구 원시동 725번	경기도 안산시 단원구 산단로 241 (원시동)
61935	100620	6410000	경기도 안산시단원구 원시동 833-1	
61936	100621	6410000	경기도 안산시단원구 신길동 1048-5	
61937	100622	6410000	경기도 안산시단원구 성곡동 647번	경기도 안산시 단원구 만해로229번길 20 (성곡동)
61938	100623	6410000	경기도 안산시단원구 성곡동 637번	경기도 안산시 단원구 성곡로 90 (성곡동)
61939	100624	6410000	경기도 안산시단원구 성곡동 638번	경기도 안산시 단원구 별망로269번길 86 (성곡동)
61940	100625	6410000	경기도 안산시단원구 성곡동 626번	경기도 안산시 단원구 별망로 253 (성곡동)
61941	100626	6410000	경기도 안산시상록구 팔곡이동 68-2	
61942	100627	6410000	경기도 안산시단원구 초지동 621번	경기도 안산시 단원구 삼성로 107 (초지동,반월공단 12B-4L)
61943	100628	6410000	경기도 안산시단원구 초지동 653-4	경기도 안산시 단원구 해안로 309-14 (초지동)
61944	100629	6410000	경기도 안산시단원구 초지동 653번	경기도 안산시 단원구 해안로 309-14 (초지동)
61945	100630	6410000	경기도 안산시단원구 성곡동 599번	경기도 안산시 단원구 강촌로139번길 9 (성곡동)
61946	100631	6410000	경기도 안성시 서운면 신능리 219번	경기도 안성시 서운면 제3공단2길 10 (안성 제3산업단지)
61947	100632	6410000	경기도 안성시 죽산면 두교리 945번	경기도 안성시 죽산면 용대길 38-21
61948	100633	6410000	경기도 화성시 마도면 쌍송리 674번	경기도 화성시 마도면 마도공단로4길 38 (마도산단 7-4)
61949	100634	6410000	경기도 화성시 마도면 쌍송리 677번	경기도 화성시 마도면 마도공단로5길 38 (마도지방산업단지 10B-2L)

- 처음에는 **개방자치단체코드**를 활용하여 **Group by** 하려고 했으나, 6410000 코드에 경기도 내 다른 여러 시가 포함되어 있어 데이터의 신뢰도에 문제가 생겼다.
- 따라서, 주소 값 공통 **Key값**으로 처리하기 위해 **Customer의 주소 포맷을 맞추는 작업**이 가장 어려웠다.

전체 SQL 코드

```
-- 주말 과제 1번 --
DESC CUSTOMER;
DESC ESTABLISHMENT;
select * from customer;
select * from establishment;

-- ADDRESS1 결측치 - 주소2로 대체하기 --
UPDATE ESTABLISHMENT SET ADDRESS1 = ADDRESS2 WHERE ADDRESS1 is null;

-- 공백 2개 기준 주소 분리
UPDATE ESTABLISHMENT
SET ADDRESS1 = SUBSTRING_INDEX(전체주소, ' ', 1);
select substr(address, 0, instr(address, ' ', 1, 2)) from establishment;

-- 공백 2개 기준 주소 update
UPDATE ESTABLISHMENT SET ADDRESS = substr(address, 0, instr(address, ' ', 1, 2));
commit;
```

```

-- 시설 구 단위별 대기오염 시설이 가장 많은 지역 확인 -> 경기도 화성시
SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
FROM ESTABLISHMENT
GROUP BY ADDRESS
HAVING COUNT(ADDRESS) > 10
ORDER BY 시설수 DESC;

-- 최대값 하나만 표시
SELECT 주소, 시설수
FROM (
    SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
    FROM ESTABLISHMENT
    GROUP BY ADDRESS
    HAVING COUNT(ADDRESS) > 10
    ORDER BY 시설수 DESC
)
FETCH FIRST ROW ONLY;

-- CUSTOMER 분리 --
SELECT NAME, MOBILE_NO, EMAIL,
DECODE(SUBSTR(TMP, 0, 2), '서울', '서울특별시',
        '울산', '울산광역시',
        '부산', '부산광역시',
        '대전', '대전광역시',
        '인천', '인천광역시',
        '대구', '대구광역시',
        '광주', '광주광역시',
        '세종', '세종특별자치시',
        '경기', '경기도',
        '강원', '강원도',
        '전남', '전라남도',
        '전북', '전라북도',
        '충남', '충청남도',
        '충북', '충청북도',
        '경남', '경상남도',
        '경북', '경상북도',
        '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM (
    SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
    FROM CUSTOMER
);

-- 최종 결과물(1) --
--
SELECT NAME, MOBILE_NO, EMAIL,
DECODE(SUBSTR(TMP, 0, 2),
        '서울', '서울특별시',
        '울산', '울산광역시',
        '부산', '부산광역시',
        '대전', '대전광역시',
        '인천', '인천광역시',
        '대구', '대구광역시',
        '광주', '광주광역시',
        '세종', '세종특별자치시',
        '경기', '경기도',
        '강원', '강원도',
        '전남', '전라남도',

```

```

        '전북', '전라북도',
        '충남', '충청남도',
        '충북', '충청북도',
        '경남', '경상남도',
        '경북', '경상북도',
        '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM(
    SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
    FROM CUSTOMER
)
WHERE DECODE(SUBSTR(TMP, 0, 2), '서울', '서울특별시',
    '울산', '울산광역시',
    '부산', '부산광역시',
    '대전', '대전광역시',
    '인천', '인천광역시',
    '대구', '대구광역시',
    '광주', '광주광역시',
    '세종', '세종특별자치시',
    '경기', '경기도',
    '강원', '강원도',
    '전남', '전라남도',
    '전북', '전라북도',
    '충남', '충청남도',
    '충북', '충청북도',
    '경남', '경상남도',
    '경북', '경상북도',
    '제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) =
(
    SELECT 주소
    FROM(
        SELECT ADDRESS AS 주소, COUNT(ADDRESS) AS 시설수
        FROM ESTABLISHMENT
        GROUP BY ADDRESS
        HAVING COUNT(ADDRESS) > 10
        ORDER BY 시설수 DESC
    )
    FETCH FIRST ROW ONLY
);

-- 최종 결과물 (2) Inner Join 사용 --
SELECT c.*, e.*
FROM (
    SELECT ADDRESS, COUNT(ADDRESS) AS 시설수
    FROM ESTABLISHMENT
    GROUP BY ADDRESS
    HAVING COUNT(ADDRESS) > 10
    ORDER BY 시설수 DESC
    FETCH FIRST ROW ONLY
) e
INNER JOIN (
    SELECT NAME, MOBILE_NO, EMAIL,
    DECODE(SUBSTR(TMP, 0, 2),
        '서울', '서울특별시',
        '울산', '울산광역시',
        '부산', '부산광역시',
        '대전', '대전광역시',
        '인천', '인천광역시',

```

```

'대구', '대구광역시',
'광주', '광주광역시',
'세종', '세종특별자치시',
'경기', '경기도',
'강원', '강원도',
'전남', '전라남도',
'전북', '전라북도',
'충남', '충청남도',
'충북', '충청북도',
'경남', '경상남도',
'경북', '경상북도',
'제주', '제주특별자치도') || ' ' || SUBSTR(TMP, 4) AS ADDRESS
FROM (
    SELECT SUBSTR(ADDRESS1,1,INSTR(ADDRESS1,' ',1,2)) AS TMP, NAME, MOBILE_NO, EMAIL
    FROM CUSTOMER
)
) c
ON c.ADDRESS = e.ADDRESS;

```