

택배 물류량과 사회적 요인 간의 관계 분석

신지혜, 김준희



목차



분석 목적 및 배경



데이터 및
테이블 소개



영유아 수와 물류량
간 관계 분석



명절 물류량 분석



오배송
발생 현황 분석



결론 및 한계

분석 목적 및 배경

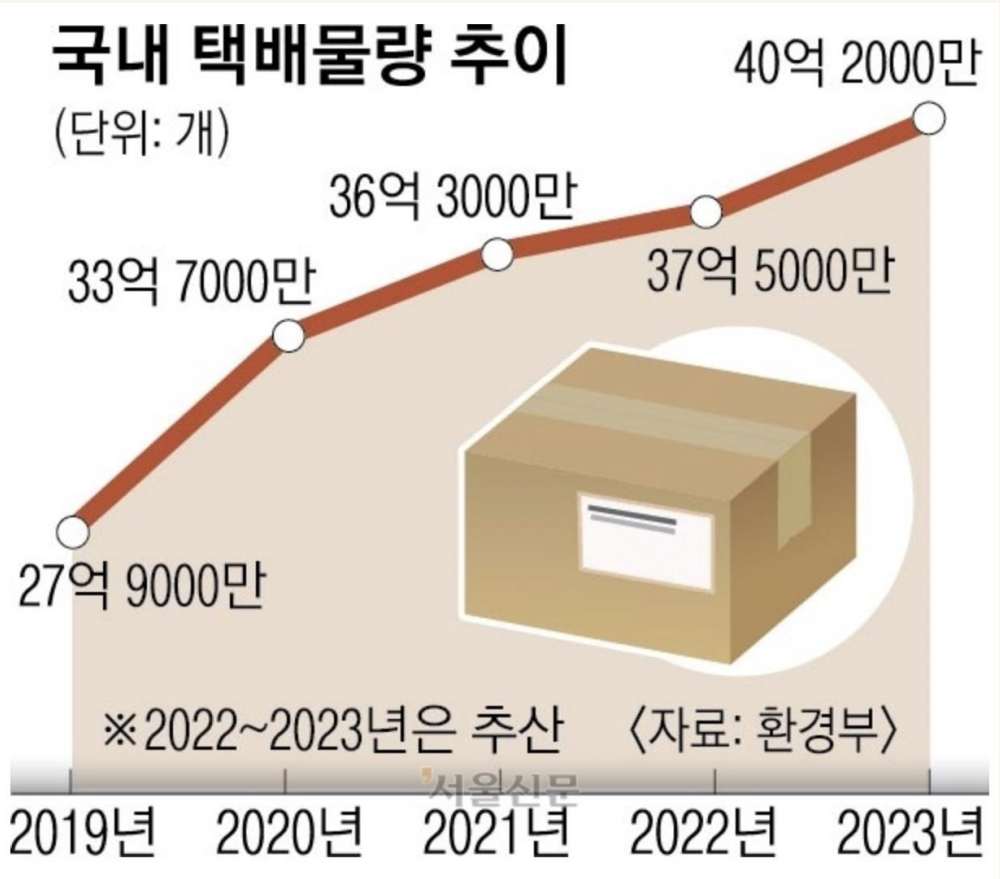
주제: 택배 물류량과 사회적 요인 간의 관계 분석 및
오배송 관련 전략 수립

목적

택배 물류량과 사회적 요인 간의 관계를 분석하고, 오배송 발생 현황에 대한 전략을
모색하는 것을 목표로 함

기대효과

- 다양한 사회적 요인과 택배 물류량 간의 관계를 이해하는 데 기여
- 사회적 요인과의 관계 분석 데이터를 토대로 택배사의 배송 관리 및 생산업
체의 공급망 관리를 최적화하여 관리 효율성을 재고
- 오배송 발생 현황을 분석하여 서비스 개선 방안을 도출



데이터 및 테이블 소개



Main data: 서울시 생활물류 데이터

CJ대한통운의 택배송장데이터를 활용한 2018~2023년 서울시 상품
유형별 택배 물류량 정보 데이터

출처: [서울 열린데이터광장](#)



Sub data: 전국 지역별 영유아 수, 출산율, 혼인 건수

출처: [KOSIS 국가통계포털](#)

데이터 내 택배의 정의

- 가로, 세로 및 높이의 합이 60cm 이하(단, 최장변 100cm 이하)이며, 30kg 이하인 소형화물
- 소형·소량의 운송물을 송하인의 주택, 사무실 또는 기타의 장소에서 수탁하여 수하인의 주택, 사무실 또는 기타의 장소까지 수송하여 인도하는 것을 의미(택배 표준약관 제2조 제1항. 2007)
- 유통사와 음식점의 배달 서비스는 포함하지 않음

데이터 특징

- 지역권별 시민 소비자의 품목별 택배 물류량을 파악하여 온라인구매 트렌드를 파악할 수 있음
- CJ대한통운의 택배시장 점유율은 50%로 온라인 시장을 대변할 수 있는 데이터임

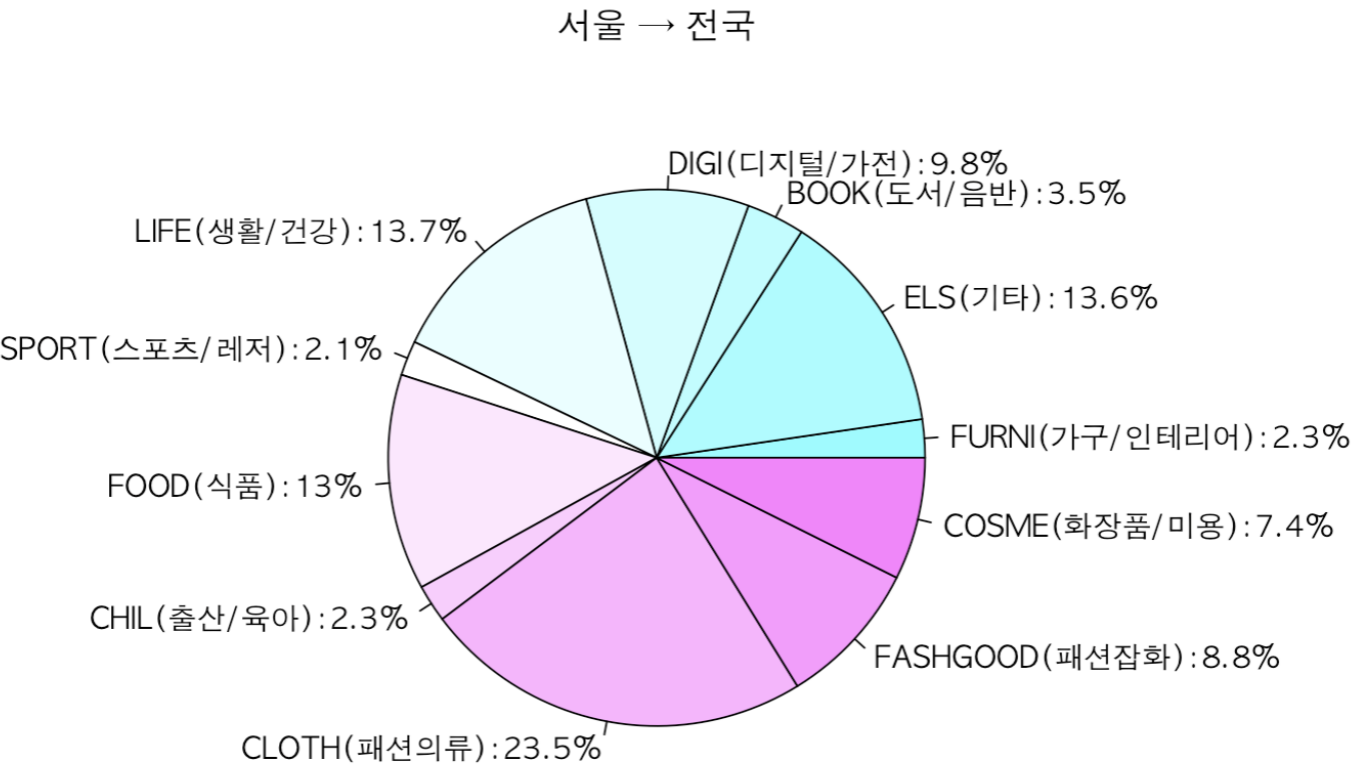
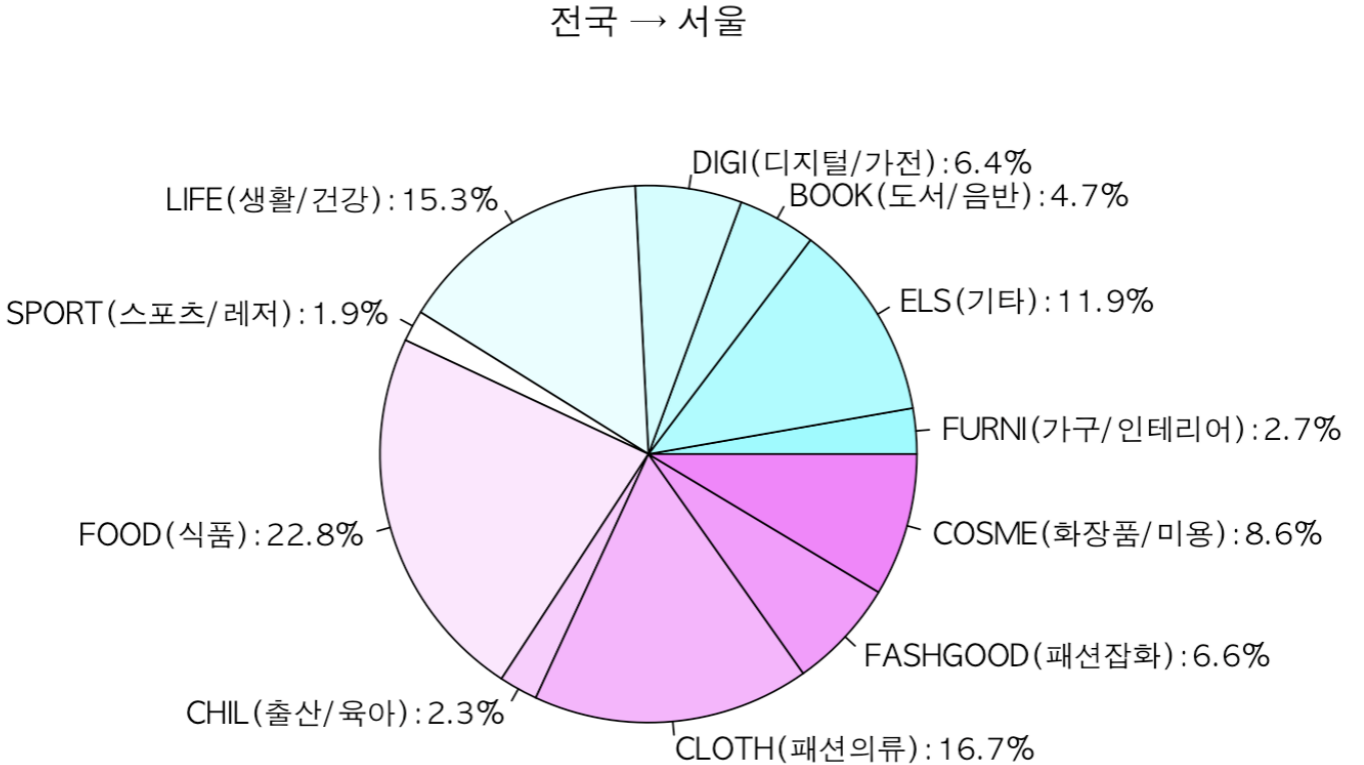
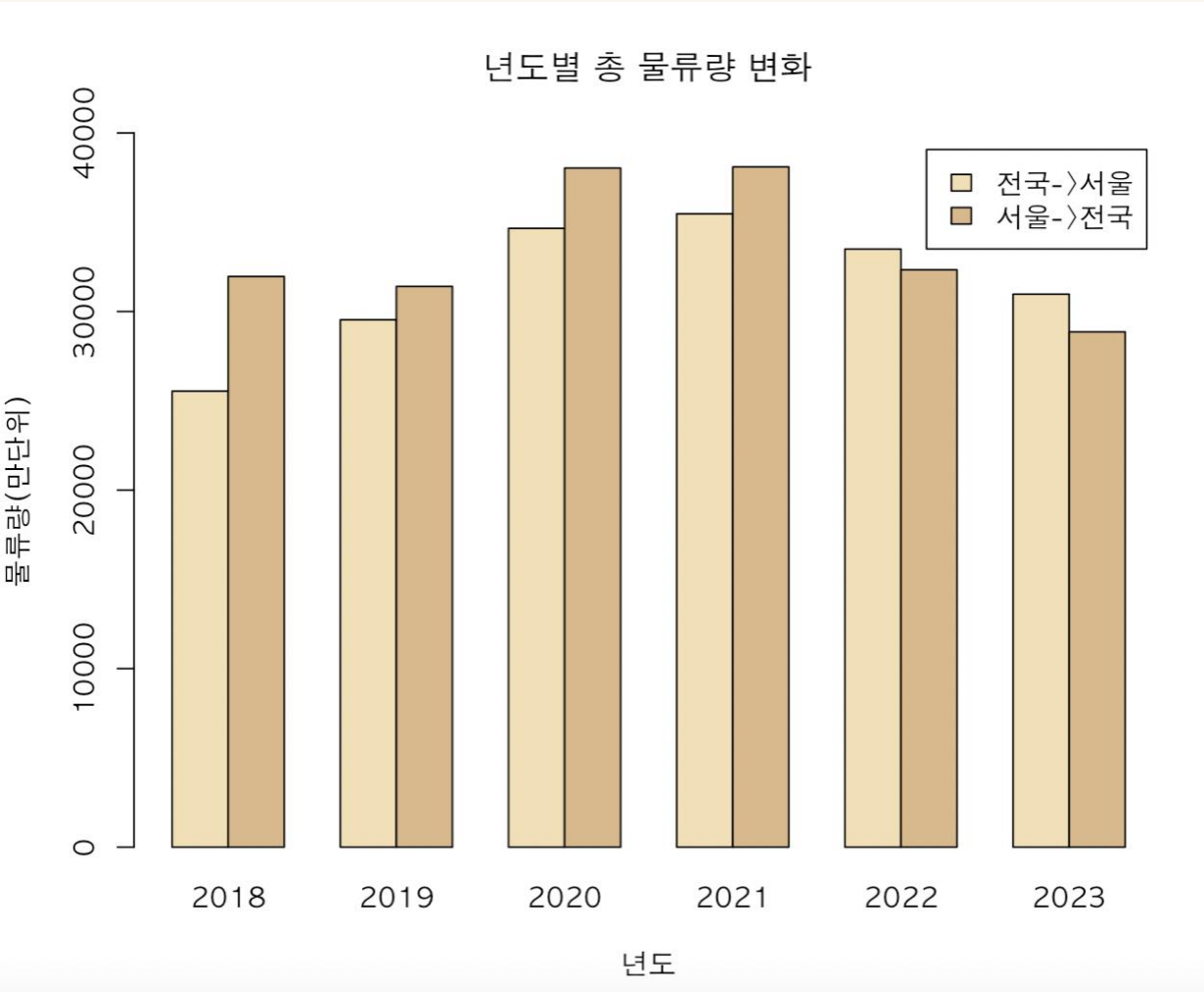
데이터의 분류

- 택배 도착일
- 상품 분류(총 10개 대분류)
- 지역 분류(서울은 자치구 단위, 서울외 지역 광역 시도 단위로 나뉨)

데이터 및 테이블 소개

전국에서 서울로 들어오는 물류량 테이블		서울에서 전국으로 가는 물류량 테이블	
컬럼명	comment	컬럼명	comment
ar_date	배송 년월일	ar_date	배송 년월일
st_lo	출발지 시명	st_lo1	출발지 시명
st_lo_co	출발지 시 코드	st_lo1_co	출발지 시 코드
ar_lo1	도착지 시명	st_lo2	출발지 구명
ar_lo1_co	도착지 시 코드	st_lo2_co	출발지 구 코드
ar_lo2	도착지 구명	ar_lo	도착지 시명
ar_lo2_co	도착지 구 코드	ar_lo_co	도착지 시 코드
furni	대분류 착지 물류량 가구/인테리어	furni	대분류 착지 물류량 가구/인테리어
els	대분류 착지 물류량 기타	els	대분류 착지 물류량 기타
book	대분류 착지 물류량 도서/음반	book	대분류 착지 물류량 도서/음반
digi	대분류 착지 물류량 디지털/가전	digi	대분류 착지 물류량 디지털/가전
life	대분류 착지 물류량 생활/건강	life	대분류 착지 물류량 생활/건강
sport	대분류 착지 물류량 스포츠/레저	sport	대분류 착지 물류량 스포츠/레저
food	대분류 착지 물류량 식품	food	대분류 착지 물류량 식품
chil	대분류 착지 물류량 출산/육아	chil	대분류 착지 물류량 출산/육아
cloth	대분류 착지 물류량 패션의류	cloth	대분류 착지 물류량 패션의류
fashgood	대분류 착지 물류량 패션잡화	fashgood	대분류 착지 물류량 패션잡화
cosme	대분류 착지 물류량 화장품/미용	cosme	대분류 착지 물류량 화장품/미용

데이터 및 테이블 소개



데이터 및 테이블 소개



인덱스

- 조인 속도를 높이기 위해 지역과 날짜 컬럼에 인덱스 추가
- 출발지 분류를 위해 전국->서울 테이블의 출발지 컬럼 추가
- 도착지 분류를 위해 서울->전국 테이블의 도착지 컬럼 인덱스 추가
- 요일별 분류를 위해 요일을 나타내는 함수 기반 인덱스 추가

제약

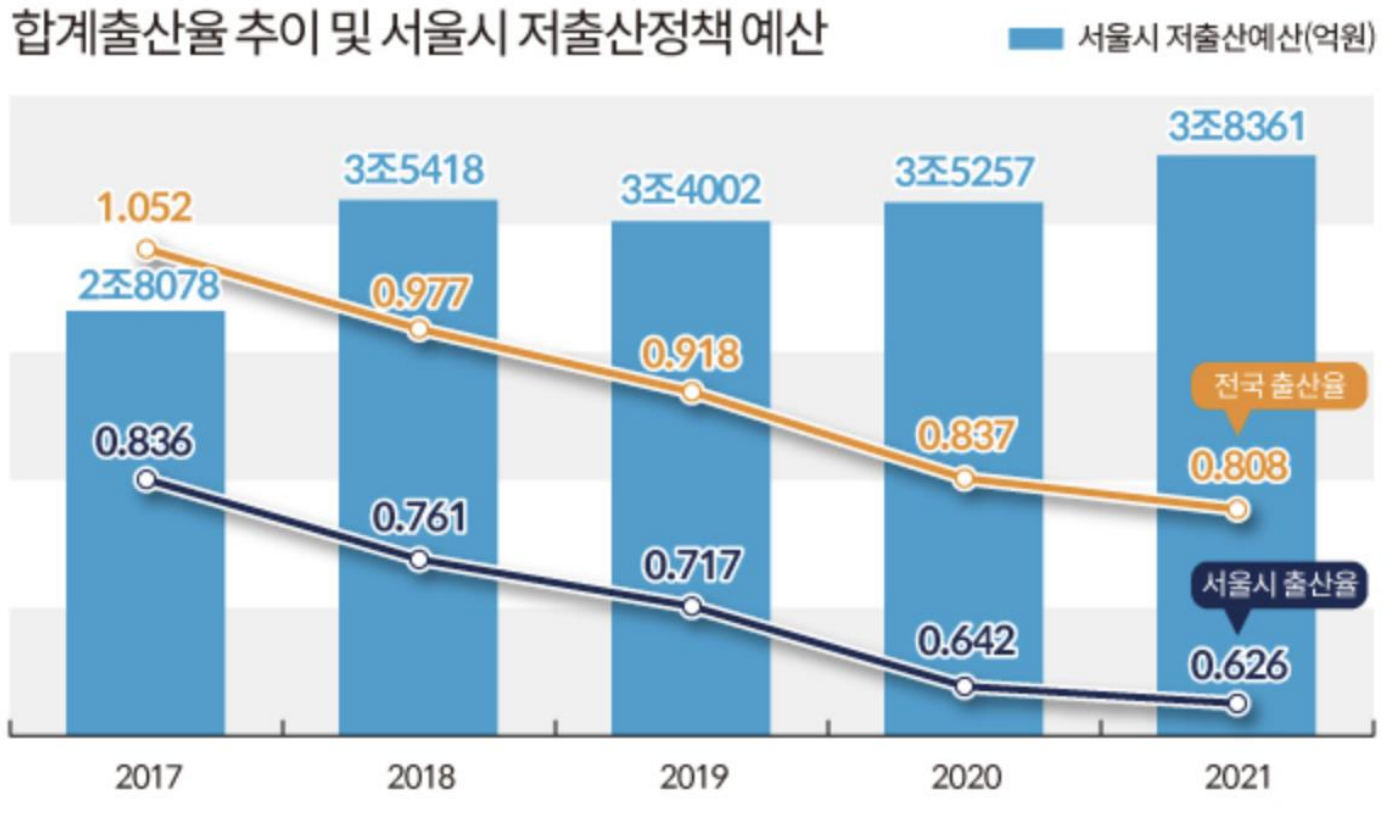
- 회귀분석에서 행번호로 사용할 컬럼에 primary key 제약 추가

비정규화 컬럼

- 사용 빈도를 고려해 모든 대분류를 합한 건 당 배송량 수 sumall 컬럼 추가

영유아 수와 물류량 간 관계 분석

- 질문 소개



출산율이 감소하고 택배 물류량은 증가하는 상황 속에서
출산 및 육아상품의 물류량과 출산 관련 지표 간의 관계 파악

영유아 수와 물류량 간 관계 분석

- 사용 테이블과 전처리

영유아 수 테이블	
컬럼명	comment
region	지역
dt	년도와 월
babyc	영유아 수

```
CREATE TABLE region_chil
(지역별 VARCHAR2(26 BYTE),
M201801 NUMBER(38,0),M201802 NUMBER(38,0),M201803 NUMBER(38,0),M201804 NUMBER(38,0),
M201805 NUMBER(38,0),M201806 NUMBER(38,0),M201807 NUMBER(38,0),M201808 NUMBER(38,0),
M201809 NUMBER(38,0),M201810 NUMBER(38,0),M201811 NUMBER(38,0),M201812 NUMBER(38,0),
M201901 NUMBER(38,0),M201902 NUMBER(38,0),M201903 NUMBER(38,0),M201904 NUMBER(38,0),
M201905 NUMBER(38,0),M201906 NUMBER(38,0),M201907 NUMBER(38,0),M201908 NUMBER(38,0),
M201909 NUMBER(38,0),M201910 NUMBER(38,0),M201911 NUMBER(38,0),M201912 NUMBER(38,0),
M202001 NUMBER(38,0),M202002 NUMBER(38,0),M202003 NUMBER(38,0),M202004 NUMBER(38,0),
M202005 NUMBER(38,0),M202006 NUMBER(38,0),M202007 NUMBER(38,0),M202008 NUMBER(38,0),
M202009 NUMBER(38,0),M202010 NUMBER(38,0),M202011 NUMBER(38,0),M202012 NUMBER(38,0),
M202101 NUMBER(38,0),M202102 NUMBER(38,0),M202103 NUMBER(38,0),M202104 NUMBER(38,0),
M202105 NUMBER(38,0),M202106 NUMBER(38,0),M202107 NUMBER(38,0),M202108 NUMBER(38,0),
M202109 NUMBER(38,0),M202110 NUMBER(38,0),M202111 NUMBER(38,0),M202112 NUMBER(38,0),
M202201 NUMBER(38,0),M202202 NUMBER(38,0),M202203 NUMBER(38,0),M202204 NUMBER(38,0),
M202205 NUMBER(38,0),M202206 NUMBER(38,0),M202207 NUMBER(38,0),M202208 NUMBER(38,0),
M202209 NUMBER(38,0),M202210 NUMBER(38,0),M202211 NUMBER(38,0),M202212 NUMBER(38,0),
M202301 NUMBER(38,0),M202302 NUMBER(38,0),M202303 NUMBER(38,0),M202304 NUMBER(38,0),
M202305 NUMBER(38,0),M202306 NUMBER(38,0),M202307 NUMBER(38,0),M202308 NUMBER(38,0),
M202309 NUMBER(38,0),M202310 NUMBER(38,0),M202311 NUMBER(38,0),M202312 NUMBER(38,0));
```



```
CREATE TABLE region_chil_c
AS
SELECT *
FROM region_chil
UNPIVOT (coun for mon in (M201801,M201802,M201803,M201804,M201805,M201806,
M201807,M201808,M201809,M201810,M201811,M201812,
M201901,M201902,M201903,M201904,M201905,M201906,
M201907,M201908,M201909,M201910,M201911,M201912,
M202001,M202002,M202003,M202004,M202005,M202006,
M202007,M202008,M202009,M202010,M202011,M202012,
M202101,M202102,M202103,M202104,M202105,M202106,
M202107,M202108,M202109,M202110,M202111,M202112,
M202201,M202202,M202203,M202204,M202205,M202206,
M202207,M202208,M202209,M202210,M202211,M202212,
M202301,M202302,M202303,M202304,M202305,M202306,
M202307,M202308,M202309,M202310,M202311,M202312));
```



```
CREATE TABLE region_chil_count
AS
SELECT 지역별 as region,
to_date(substr(mon,2,4)||'/'||substr(mon,6,2)||'/01','RR/MM/DD') as dt,
coun as babyc
FROM region_chil_c;
```

영유아 수와 물류량 간 관계 분석

- 사용 테이블과 전처리

출산/육아 물류량 테이블	
컬럼명	comment
dt	년도와 월
region	지역
chil	대분류 착지 물류량 출산/육아

출산율과 출생아 수 테이블	
컬럼명	comment
region	지역
b_year	년도
b_count	출생아 수
b_rate	출산율

```
CREATE TABLE fromsall
AS
SELECT * FROM froms2018y
UNION ALL
SELECT * FROM froms2019y
UNION ALL
SELECT * FROM froms2020y
UNION ALL
SELECT * FROM froms2021y
UNION ALL
SELECT * FROM froms2022y
UNION ALL
SELECT * FROM froms2023y;
```



```
CREATE TABLE froms_baby
AS
SELECT to_date(to_char(ar_date,'RR/MM')||'/01') as dt,ar_lo as region,
       sum(chil) as chil
FROM fromsall
GROUP BY ar_lo,to_char(ar_date,'RR/MM');
```

```
CREATE TABLE birth_rate
(region varchar2(100),
 b_year number(10),
 b_count number(10),
 b_rate number(10,4));
```

영유아 수와 물류량 간 관계 분석

- 영유아 수와의 상관관계

지역과 월을 기준으로 나눈 상관관계

상관관계
0.961

지역을 기준으로 나눈 상관관계

지역	유아용품	영유아수	상관관계
서울특별시	8,394,969	19,680,264	0.997
부산광역시	2,539,437	7,055,938	0.997
대구광역시	1,943,791	5,361,378	0.997
인천광역시	2,619,698	7,111,568	0.997
광주광역시	1,266,081	3,564,156	0.997
대전광역시	1,312,109	3,537,889	0.997
울산광역시	961,295	2,981,276	0.997
세종특별자치시	467,446	1,454,655	0.997
경기도	13,318,965	35,009,812	0.997
강원도	1,308,450	3,249,531	0.997
충청북도	1,417,714	3,785,792	0.997
충청남도	1,985,803	5,273,604	0.997
전라북도	1,351,124	3,737,128	0.997
전라남도	1,340,822	4,016,972	0.997
경상북도	2,017,312	5,783,992	0.997
경상남도	2,749,832	7,897,098	0.997
제주특별자치도	343,776	1,796,882	0.997

```
SELECT round(corr(f.chil,r.babyc),3)as 상관관계
FROM froms_baby f,region_chil_count r
WHERE f.region = r.region and f.dt = r.dt;
```

```
SELECT f.region 지역,to_char(sum(f.chil),'999,999,999') 유아용품,
to_char(sum(r.babyc),'999,999,999') 영유아수,
round(corr(sum(f.chil),sum(r.babyc)) over ( ),3)as 상관관계
FROM froms_baby f,region_chil_count r
WHERE f.region = r.region and f.dt = r.dt
GROUP BY f.region;
```

영유아 수와 물류량 간 관계 분석

- 영유아 수와의 상관관계

년도를 기준으로 나눈 상관관계

년도	유아용품	영유아수	상관관계
2018	7,129,570	24,237,605	0.123
2019	7,136,119	22,826,044	0.123
2020	8,602,584	21,008,490	0.123
2021	8,952,251	19,094,847	0.123
2022	7,071,492	17,633,488	0.123
2023	6,446,608	16,497,461	0.123

```
SELECT to_char(f.dt,'RRRR') 년도,t0_char(sum(f.chil),'999,999,999') 유아용품,
       t0_char(sum(r.babyc),'999,999,999') 영유아수,
       round(corr(sum(f.chil),sum(r.babyc)) over (),3) as 상관관계
FROM froms_baby f,region_chil_count r
WHERE f.region = r.region and f.dt = r.dt
GROUP BY to_char(f.dt,'RRRR');
```

코로나 시기를 제외한 상관관계

년도	유아용품	영유아수	상관관계
2018	7,129,570	24,237,605	0.725
2019	7,136,119	22,826,044	0.725
2022	7,071,492	17,633,488	0.725
2023	6,446,608	16,497,461	0.725

```
SELECT to_char(f.dt,'RRRR') 년도,t0_char(sum(f.chil),'999,999,999') 유아용품,
       t0_char(sum(r.babyc),'999,999,999') 영유아수,
       round(corr(sum(f.chil),sum(r.babyc)) over (),3) as 상관관계
FROM froms_baby f,region_chil_count r
WHERE f.region = r.region and f.dt = r.dt
      and to_char(f.dt,'RRRR') not in ('2020','2021')
GROUP BY to_char(f.dt,'RRRR');
```

영유아 수와 물류량 간 관계 분석

- 출산율과의 상관관계

년도를 기준으로 나눈 상관관계

년도	유아용품	출산율	상관관계
2018	7,129,570	0.977	0.052
2019	7,136,119	0.918	0.052
2020	8,602,584	0.837	0.052
2021	8,952,251	0.808	0.052
2022	7,071,492	0.778	0.052
2023	6,446,608	0.72	0.052

```
WITH froms_baby_ch AS
  (SELECT to_char(dt,'RRRR') as dt,sum(chil) as chil
   FROM froms_baby
   GROUP BY to_char(dt,'RRRR'))
SELECT f.dt 년도,to_char(f.chil,'999,999,999') 유아용품,b.b_rate 출산율,
       round(corr(f.chil,b.b_rate) over(),3) as 상관관계
FROM froms_baby_ch f,birth_rate b
WHERE b.region = '전국' and f.dt = b.b_year;
```

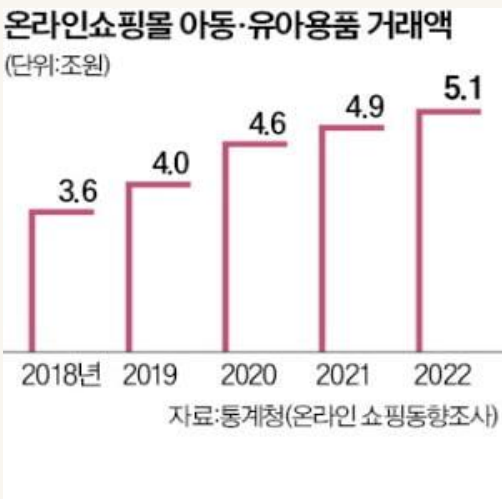
코로나 시기를 제외한 상관관계

년도	유아용품	출산율	상관관계
2018	7,129,570	0.977	0.769
2019	7,136,119	0.918	0.769
2022	7,071,492	0.778	0.769
2023	6,446,608	0.72	0.769

```
WITH froms_baby_ch AS
  (SELECT to_char(dt,'RRRR') as dt,sum(chil) as chil
   FROM froms_baby
   GROUP BY to_char(dt,'RRRR'))
SELECT f.dt 년도,to_char(f.chil,'999,999,999') 유아용품,b.b_rate 출산율,
       round(corr(f.chil,b.b_rate) over(),3) as 상관관계
FROM froms_baby_ch f,birth_rate b
WHERE b.region = '전국' and f.dt = b.b_year
      and f.dt not in ('2020','2021');
```


영유아 수와 물류량 간 관계 분석

- 출산율과의 상관관계



감소하는 영유아 수와 출산율의 영향으로, 유아용품 판매량 감소 추세
따라서 유아용품 시장규모 감소를 예상하게 되지만, 아이러니하게도 시장규모와 아이 1인당 투입 금액은 증가 추세
유아용품의 프리미엄화 진행 중

코로나 시기를 제외한 상관관계

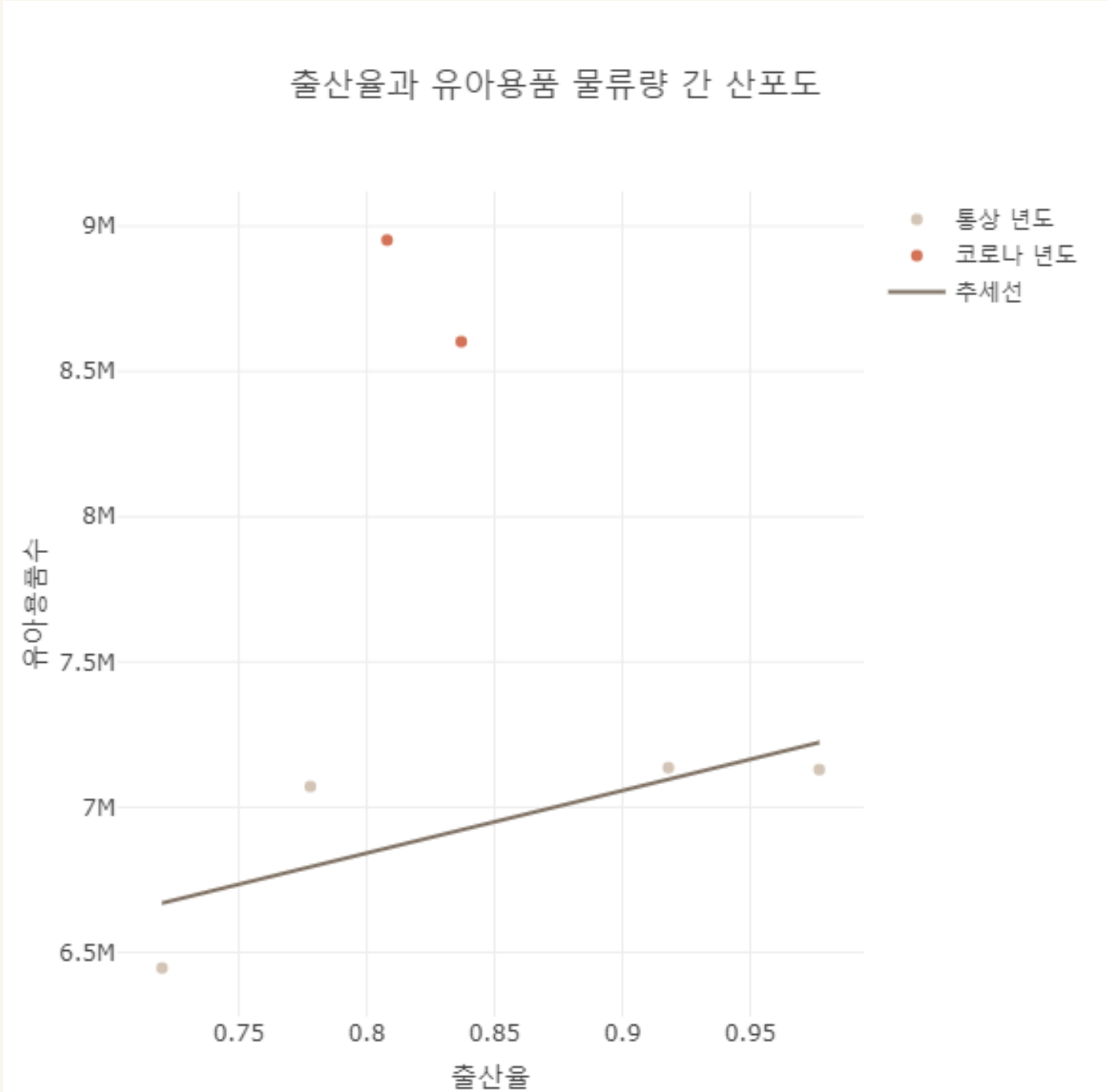
연도	유아용품	출산율	상관관계
2018	7,129,570	0.977	0.769
2019	7,136,119	0.918	0.769
2022	7,071,492	0.778	0.769
2023	6,446,608	0.72	0.769

```
WITH froms_baby_ch AS
  (SELECT to_char(dt,'RRRR') as dt,sum(chil) as chil
   FROM froms_baby
   GROUP BY to_char(dt,'RRRR'))
SELECT f.dt 년도,to_char(f.chil,'999,999,999') 유아용품,b.b_rate 출산율,
       round(corr(f.chil,b.b_rate) over(),3) as 상관관계
FROM froms_baby_ch f,birth_rate b
WHERE b.region = '전국' and f.dt = b.b_year
      and f.dt not in ('2020','2021');
```

영유아 수와 물류량 간 관계 분석

- 출산율과의 상관관계

출산율이 높을수록 유아용품 물류량도 증가하는 패턴을 보임
예외적으로 코로나 시기인 2020년과 2021년에는
통상적인 추세선에 비해 유아용품 물류량이 크게 높은 것으로 나타남
코로나 시기에 오프라인 구매가 어려워
유아용품 구매에 있어 온라인 배송 비율이 크게 증가했기 때문인 것으로 보임



영유아 수와 물류량 간 관계 분석

- 다중회귀분석

혼인 건수 테이블

컬럼명	comment
region	지역
p_year	년월
chil	대분류 착지 물류량 출산/육아

```
CREATE TABLE region_marry
(region varchar2(100),
p_year varchar2(20),
marry_count number(10));
```

다중회귀분석 사용 테이블

컬럼명	comment
num	행번호
chil	대분류 착지 물류량 출산/육아
babyc	영유아 수
marry_count	월별 혼인 건수

```
CREATE TABLE baby_c
AS
SELECT rownum as num,f.chil,r.babyc,c.marry_count
FROM froms_baby_sample f,region_chil_count r,region_marry_c c
WHERE f.dt = r.dt and f.region = r.region
and c.dt = r.dt and c.region = r.region;
```

- 다중회귀분석

R	결정계수
R_SQ	0.931
ADJUSTED_R_SQUARE	0.931

- ## 다중회귀분석 모델

$$Y = 0.2214X_1 + 15.3688X_2 - 762.1224$$

$$Y = \text{출산/육아용품 물류량}$$

X1 = 영유아 수

X2 = 결혼 건수

```
BEGIN
    DBMS_DATA_MINING.CREATE_MODEL(
        MODEL_NAME          => 'MD_REG_MODEL2',
        MINING_FUNCTION      => DBMS_DATA_MINING.REGRESSION,
        DATA_TABLE_NAME    => 'BABY_C',
        CASE_ID_COLUMN_NAME => 'NUM',
        TARGET_COLUMN_NAME  => 'CHIL',
        SETTINGS_TABLE_NAME => 'SETTINGS_REG2' );
END;
```

명절 물류량 분석

- 질문 소개

분석 목적

- 명절 기간에 급증하는 택배 수요를 정량화하여 물류 관리 및 운영의 최적화를 도모하고자 함

조사 항목

- 조사 기념일: 설, 추석
- 조사 기간: 명절 당일 전후 4일
- 설과 추석은 음력으로 쇠기 때문에 해당 년도의 양력 날짜로 변경해 입력함
- 명절 당일을 포함한 전후 4일(총 9일)을 조사 기간으로 설정
- 연휴, 주말, 대체공휴일로 인해 택배가 제때 도착하지 못하는 것을 감안,
조사 기간 중 가장 택배량이 많은 날을 명절 배송량의 기준일로 삼음

명절 물류량 분석

- 분석 결과

```
CREATE VIEW 추석23
AS
WITH 기념일택배 as
(
SELECT ar_date as 배송일,
CASE WHEN ar_date < to_date('23.09.29', 'RR.MM.DD')
THEN '추석' || (ar_date - to_date('23.09.29', 'RR.MM.DD'))
WHEN ar_date > to_date('23.09.29', 'RR.MM.DD')
THEN '추석+' || (ar_date - to_date('23.09.29', 'RR.MM.DD'))
WHEN ar_date = to_date('23.09.29', 'RR.MM.DD')
THEN '추석'
END
as 주변일,
to_char(ar_date, 'day') as 요일,
to_char(sum(sumall), '999,999,999') as 배송량,
dense_rank() over (order by sum(sumall) desc) as 순위,
to_char(round(avg(sum(sumall)) over ()), '999,999,999') as 연평균
FROM froms23
GROUP BY ar_date
),
표준편차값 as
(
SELECT ar_date as 배송일,
round(stddev(sumall), 2) as 표준편차
FROM froms23
GROUP BY ar_date
)
SELECT a.배송일, a.주변일, a.요일, a.배송량, a.연평균, b.표준편차, a.순위
FROM 기념일택배 a, 표준편차값 b
WHERE a.배송일=b.배송일
AND a.배송일 between to_date('23.09.29', 'RR.MM.DD')-4 and to_date('23.09.29', 'RR.MM.DD')+4
ORDER BY a.순위 asc fetch first 1 rows only;
```

```
SELECT * FROM 추석23
UNION ALL
SELECT * FROM 추석22
UNION ALL
SELECT * FROM 추석21
UNION ALL
SELECT * FROM 추석20
UNION ALL
SELECT * FROM 추석19
UNION ALL
SELECT * FROM 추석18;
```

명절 물류량 분석

- 분석 결과

배송일	주변일	요일	배송량	연평균	표준편차	순위
23/09/25	추석-4	월요일	1,480,036	790,735	5263.21	46
22/09/13	추석+3	화요일	2,251,284	885,936	7317.33	3
21/09/23	추석+2	목요일	2,499,630	1,043,863	8395.26	4
20/10/05	추석+4	월요일	3,293,826	1,039,243	11476.71	1
19/09/16	추석+3	월요일	2,327,642	860,497	7981.78	1
18/09/27	추석+3	목요일	2,234,219	875,686	7722.5	1
배송일	주변일	요일	배송량	연평균	표준편차	순위
23/01/25	설+3	수요일	1,974,798	790,735	6871.7	4
22/02/03	설+2	목요일	2,312,313	885,936	8013.92	2
21/02/15	설+3	월요일	3,292,898	1,043,863	12764.43	1
20/01/28	설+3	화요일	2,250,448	1,039,243	7767.96	13
19/02/07	설+2	목요일	2,313,089	860,497	7883.54	2
18/02/19	설+3	월요일	2,182,351	875,686	7437.04	2

- 명절 전후 4일의 조사 기간 중 택배량이 가장 많은 날은 일반적으로 명절 당일의 2, 3일 후인 것으로 확인됨
- 명절 배송량은 2020년과 2021년을 기점으로 감소하는 추세를 보이고 있으며, 표준편차와 일일 배송량 순위 역시 하락하고 있음. 명절 물류량이 연평균 물류량에 대해 갖는 강한 우위가 줄어들고 있음
- 이는 과거보다 명절의 중요성을 낮게 평가하는 경향이 물류량에도 반영되었음을 시사함
- 명절 특수성의 영향력이 감소하면서 생산업체와 유통업체는 명절 전략을 재고할 필요가 있음
- 다만, 코로나 시기에 고향 방문이 어려워지며 명절 선물 물량이 증가한 점과 코로나 시기에 급증했던 연평균 택배량을 고려해야 함. 이러한 요인들을 포함한 해석이기에 이후의 추이를 추가로 살펴봐야 할 것임

오배송 발생 현황 분석

- 질문 소개

분석 목적

- 데이터 전처리 과정에서 결측치는 없었으나, 배송 날짜, 출발지, 도착지가 기록되어 있음에도 불구하고
- 배송된 상품이 없는 오배송 데이터가 발견되었음
- 오배송이 어떠한 상황에서 발생하는지 현황을 분석하여, 오배송을 줄이기 위한 방안을 모색하는 것을 목표로 함

조사 항목

- 전체적인 오배송 발생 현황을 파악하기 위한 연간 오배송량 테이블
- 배송도착일, 출발지, 도착지 컬럼을 활용하여 분석
- 이 중 유의미한 차이가 나타난 요일별 요배송량 및 출발지별 오배송량 테이블을 수록

오배송 발생 현황 분석

- 연간 오배송 현황

서울 -> 전국			전국 -> 서울		
년도	오배송량	오배송 비율	년도	오배송량	오배송 비율
2018	5,195	.00163%	2018	3,036	.00119%
2019	5,107	.00163%	2019	3,246	.0011%
2020	4,354	.00114%	2020	2,449	.00071%
2021	4,090	.00107%	2021	2,291	.00065%
2022	4,578	.00142%	2022	2,341	.0007%
2023	5,158	.00179%	2023	1,891	.00061%
총계	28,482	.00142%	총계	15,254	.0008%

```
WITH froms_0 AS (
    SELECT to_char(ar_date, 'RRRR') as 년도, count(*) as 오배송량
    FROM (
        SELECT ar_date, sumall FROM froms23 WHERE sumall = 0
        UNION ALL
        SELECT ar_date, sumall FROM froms22 WHERE sumall = 0
        UNION ALL
        SELECT ar_date, sumall FROM froms21 WHERE sumall = 0
        UNION ALL
        SELECT ar_date, sumall FROM froms20 WHERE sumall = 0
        UNION ALL
        SELECT ar_date, sumall FROM froms19 WHERE sumall = 0
        UNION ALL
        SELECT ar_date, sumall FROM froms18 WHERE sumall = 0
    )
    GROUP BY to_char(ar_date, 'RRRR')
),
froms_all AS (
    SELECT to_char(ar_date, 'RRRR') AS 년도, sum(sumall) AS 전체배송량
    FROM (
        SELECT ar_date, sumall FROM froms23
        UNION ALL
        SELECT ar_date, sumall FROM froms22
        UNION ALL
        SELECT ar_date, sumall FROM froms21
        UNION ALL
        SELECT ar_date, sumall FROM froms20
        UNION ALL
        SELECT ar_date, sumall FROM froms19
        UNION ALL
        SELECT ar_date, sumall FROM froms18
    )
    GROUP BY to_char(ar_date, 'RRRR')
)
SELECT
    NVL(f0.년도, '총계') AS 년도,
    TO_CHAR(SUM(f0.오배송량), '999,999') AS 오배송량,
    ROUND(SUM(f0.오배송량) / SUM(fa.전체배송량) * 100, 5) || '%' AS "오배송 비율"
FROM froms_0 f0, froms_all fa
WHERE f0.년도 = fa.년도
GROUP BY ROLLUP(f0.년도)
ORDER BY CASE WHEN f0.년도 IS NULL THEN 1 ELSE 0 END, f0.년도;
```

연간 오배송 건수: 각 테이블 당 약 4700건, 2500건

총 오배송 비율: 약 0.001 수준

절대적인 비율은 낮지만, 꾸준히 발생하고 있으므로 발생 원인을 분석하고 개선하는 것이 필요함

오배송 발생 현황 분석

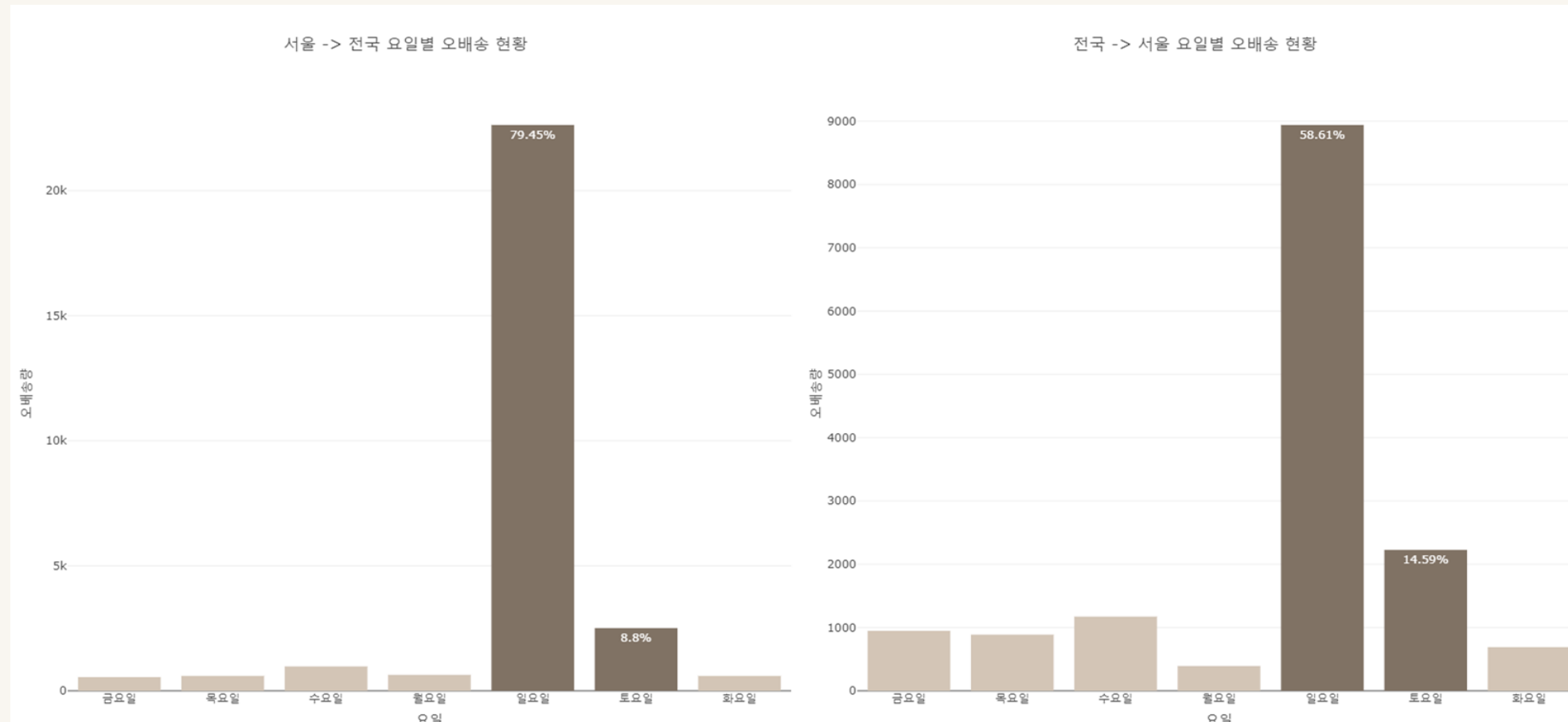
- 요일별 오배송 현황

서울 -> 전국				전국 -> 서울			
요일	오배송량	요일별 비율	오배송 비율	요일	오배송량	요일별 비율	오배송 비율
일요일	22,628	79.45%	.1353%	일요일	8,941	58.61%	.01946%
토요일	2,506	8.8%	.00485%	토요일	2,226	14.59%	.00409%
수요일	973	3.42%	.00028%	수요일	1,172	7.68%	.00036%
월요일	636	2.23%	.00012%	금요일	947	6.21%	.00033%
목요일	596	2.09%	.00017%	목요일	887	5.81%	.00028%
화요일	595	2.09%	.00015%	화요일	689	4.52%	.00019%
금요일	548	1.92%	.00018%	월요일	392	2.57%	.00008%

```
WITH froms_0 AS (  
  SELECT to_char(ar_date, 'DAY') as 요일, count(*) as 오배송량  
    FROM (  
      SELECT ar_date      FROM froms23      WHERE sumall = 0  
    UNION ALL  
      SELECT ar_date      FROM froms22      WHERE sumall = 0  
    UNION ALL  
      SELECT ar_date      FROM froms21      WHERE sumall = 0  
    UNION ALL  
      SELECT ar_date      FROM froms20      WHERE sumall = 0  
    UNION ALL  
      SELECT ar_date      FROM froms19      WHERE sumall = 0  
    UNION ALL  
      SELECT ar_date      FROM froms18      WHERE sumall = 0      )  
  GROUP BY to_char(ar_date, 'DAY')      ),  
froms_all AS (  
  SELECT to_char(ar_date, 'DAY') as 요일, sum(sumall) as 전체배송량  
    FROM (  
      SELECT ar_date, sumall FROM froms23  
    UNION ALL  
      SELECT ar_date, sumall FROM froms22  
    UNION ALL  
      SELECT ar_date, sumall FROM froms21  
    UNION ALL  
      SELECT ar_date, sumall FROM froms20  
    UNION ALL  
      SELECT ar_date, sumall FROM froms19  
    UNION ALL  
      SELECT ar_date, sumall FROM froms18      )  
  GROUP BY to_char(ar_date, 'DAY')      )  
SELECT  
  f0.요일,  
  to_char(sum(f0.오배송량), '999,999') as 오배송량,  
  round(ratio_to_report(sum(f0.오배송량)) over () * 100, 2) || '%' as "요일별 비율",  
  round(sum(f0.오배송량)/sum(fa.전체배송량) * 100,5) || '%' as "오배송 비율"  
FROM froms_0 f0, froms_all fa  
WHERE f0.요일=fa.요일  
GROUP BY f0.요일  
ORDER BY 오배송량 desc;
```

오배송 발생 현황 분석

- 요일별 오배송 현황



전체 오배송의 약 70%가 일요일에 나타나며,

토요일이 약 11%로 뒤를 이음.

주말에 오배송이 빈발하는 패턴을 확인할 수 있음

이는 주말에 운영되는 물류 체계나 인력 배치에 문제가 있을 가능성을 시사함

오배송 발생 현황 분석

- 출발지별 오배송 현황

출발지	오배송량	지역별 비율	오배송 비율
울산광역시	4,022	26.37%	.05131%
광주광역시	2,802	18.37%	.01382%
세종특별자치시	1,492	9.78%	.01625%
제주특별자치도	1,102	7.22%	.0069%
부산광역시	970	6.36%	.00135%
강원도	619	4.06%	.00333%
충청남도	597	3.91%	.00151%
대구광역시	535	3.51%	.00087%
경상북도	521	3.42%	.00134%
전라북도	498	3.26%	.0017%
전라남도	492	3.23%	.00227%
경상남도	477	3.13%	.00135%
충청북도	441	2.89%	.00086%
대전광역시	240	1.57%	.00046%
서울특별시	221	1.45%	.00005%
인천광역시	123	.81%	.0001%
경기도	102	.67%	.00001%

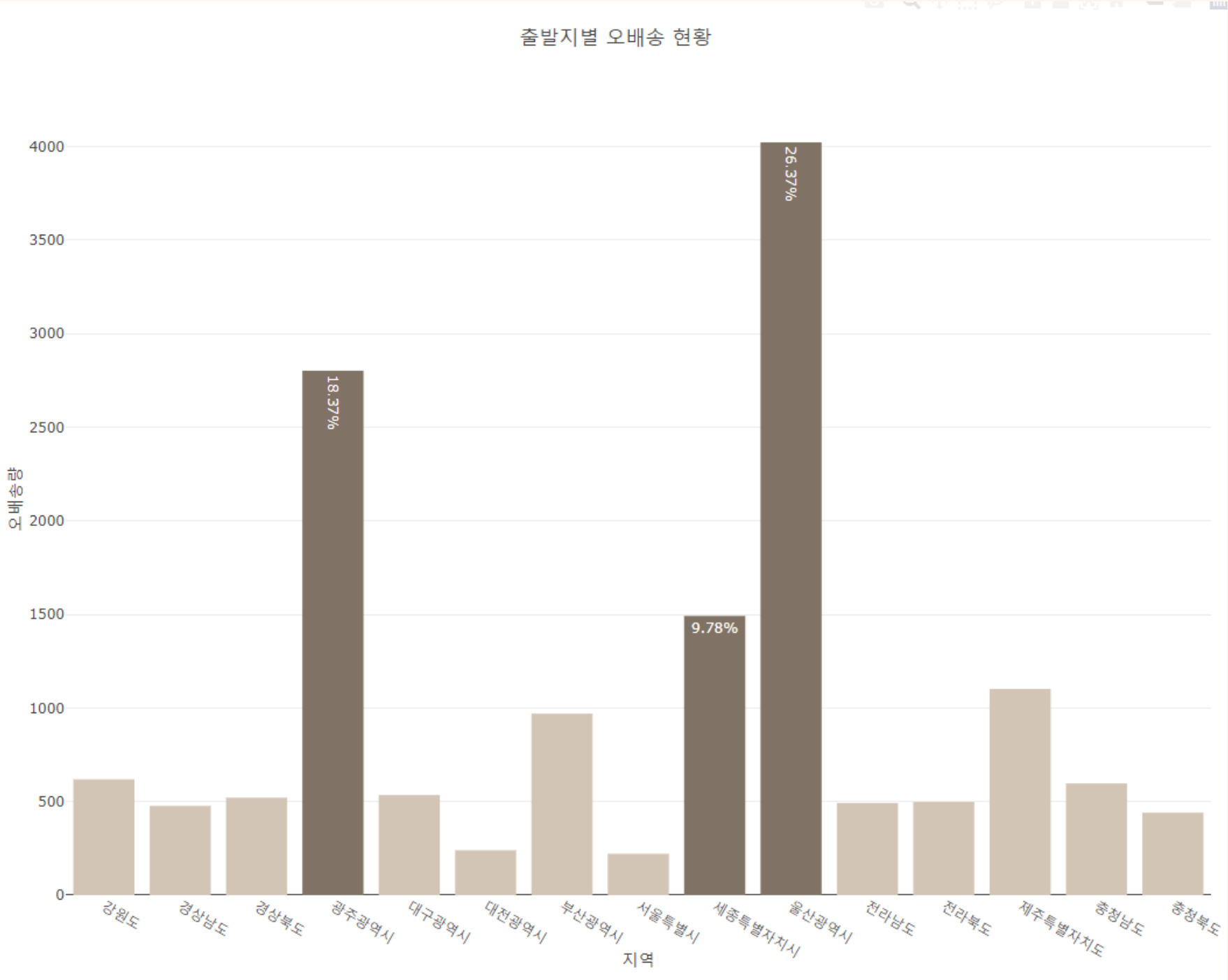
```
WITH tos_0 AS (
SELECT st_lo as 출발지, count(*) as 오배송량
FROM (
SELECT st_lo FROM tos23 WHERE sumall = 0
UNION ALL
SELECT st_lo FROM tos22 WHERE sumall = 0
UNION ALL
SELECT st_lo FROM tos21 WHERE sumall = 0
UNION ALL
SELECT st_lo FROM tos20 WHERE sumall = 0
UNION ALL
SELECT st_lo FROM tos19 WHERE sumall = 0
UNION ALL
SELECT st_lo FROM tos18 WHERE sumall = 0 )
GROUP BY st_lo
),
tos_all AS (
SELECT st_lo as 출발지, sum(sumall) as 전체배송량
FROM (
SELECT st_lo, sumall FROM tos23
UNION ALL
SELECT st_lo, sumall FROM tos22
UNION ALL
SELECT st_lo, sumall FROM tos21
UNION ALL
SELECT st_lo, sumall FROM tos20
UNION ALL
SELECT st_lo, sumall FROM tos19
UNION ALL
SELECT st_lo, sumall FROM tos18 )
GROUP BY st_lo
)
SELECT
f0.출발지,
to_char(sum(f0.오배송량), '999,999') AS 오배송량,
round(ratio_to_report(sum(f0.오배송량)) over () * 100, 2) || '%' as "지역별 비율",
round(sum(f0.오배송량)/sum(fa.전체배송량) * 100,5) || '%' as "오배송 비율"
FROM tos_0 f0, tos_all fa
WHERE f0.출발지=fa.출발지
GROUP BY f0.출발지
ORDER BY 오배송량 desc;
```

오배송 발생 현황 분석

- 출발지별 오배송 현황

전체 오배송의 약 26%가
울산광역시에서 출발하였으며,
광주광역시가 18%,
세종특별자치시가 10%로 뒤를 이음

특정 출발지에서 유독 많은 오배송이 발생하는 것은
해당 지역의 물류센터 운영 방식 또는
지역적 물류 인프라에 문제가 있을 가능성이 있음
해당 지역의 물류 운영 방식을 면밀히 분석하고
개선책을 마련하는 것이 요구됨



결론 및 한계

결론

- 출산/육아용품 물류량과 영유아수, 출산율은 높은 상관관계를 가짐. 출산율 감소에 따라 출산/육아용품 판매량 역시 감소 중
다중회귀분석 모델을 통해 영유아수, 결혼 건수를 통해 출산/육아용품 물류량을 예측할 수 있음
- 대부분의 명절 물류량은 명절 2,3일 후에 도착하며, 명절 물류량이 연평균 물류량에 대해 갖는 우위가 감소하고 있음.
이는 한국인들이 과거보다 명절의 중요성을 낮게 평가하는 경향이 물류량에도 반영되었음을 시사함
- 꾸준한 비율을 유지하는 오배송 데이터를 추적 관찰할 필요성이 있으며,
특히 주말 및 특정 지역에서 오배송이 집중되므로 이와 관련한 개선방안이 요구됨

한계

- 내부 입력 절차의 불명확성: 데이터의 내부 입력 절차에 대해 공개된 바가 적어 오배송 데이터가 발생하는 정확한 이유가 불명확함
- 코로나 시기로 인한 데이터 일관성 부족: 6년 간의 데이터에 코로나 시기의 영향이 커서 데이터의 일관적인 추이를 분석하는 데 어려움이 있음

Thank You