

EDA Competition

채움조

20180814 김현석 20182894 현태훈 20192803 주정윤 20202662 조현식 20202639 박세현 20212541 김규리

INDEX

01

배경

어떤 배경으로
EDA분석 방향을 설정하였는가

02

전처리 진행 과정 코드

전처리 과정

03

분석 & 시각화 과정

EDA를 통해
어떤 상관관계를 파악하였는가

04

EDA의 결과로 결론 도출

결론 및 특이점 도출

배경

어떤 데이터를 사용할까?



어떤 열을 사용할까?



유통사별 어떤 특징이 있을까?



유통사 추천하기^{!!}

02 전처리 과정

데이터 불러오기

```
#데이터 불러오기
df_cd = pd.read_csv('customer_data.csv',encoding = 'utf-8')
df_pd = pd.read_csv('product_data.csv',encoding = 'utf-8')
df_od = pd.read_csv('order_data.csv')
df_md = pd.read_csv('market_data.csv')
```

열 이름 변경

```
#열 이름 바꾸기
df_cd.columns = ['고객ID' , '성별' , '연령대' , '거주지_대분류']
df_od.columns = ['고객ID' , '영수증번호' , '온,오프라인구분' , '유통사' , '점포코드' , '상품코드' , '구매일자' , '구매시간' , '구매금액' , '구매수량']
df_pd.columns = ['상품코드' , '상품_소분류' , '상품_대분류' , '상품_중분류']
df_md.columns = ['점포코드' , '유통사' , '거주지_대분류' , '거주지_소분류']
```

중복 열 제거

```
#중복 열 제거
df_md=df_md.drop(['유통사','거주지_대분류'], axis=1)
```

병합

```
#병합
df_merge1 = pd.merge(df_cd, df_od, on='고객ID', how = 'outer')
df_merge2 = pd.merge(df_merge1, df_pd, on='상품코드', how = 'outer')
df = pd.merge(df_merge2, df_md, on='점포코드', how = 'outer')
```

범주 축소

```
#범주 축소화
df = df.drop(['거주지_대분류' , '거주지_소분류' , '점포코드' , '상품코드'], axis=1)
```

02 전처리 과정

결측치 확인 및 처리

```
#결측치 확인  
df.isnull().sum()
```

```
고객ID      0  
성별        0  
연령대      0  
거주지_대분류    0  
영수증번호  2996  
온,오프라인구분  2996  
유통사      2996  
점포코드   431497  
상품코드    2996  
구매일자    2996  
구매시간    2996  
구매금액    2996  
구매수량    2996  
상품_소분류  2996  
상품_대분류  2996  
상품_중분류  2996  
거주지_소분류  431497  
dtype: int64
```

```
#결측치 처리  
df=df.dropna(subset=['구매일자'])
```

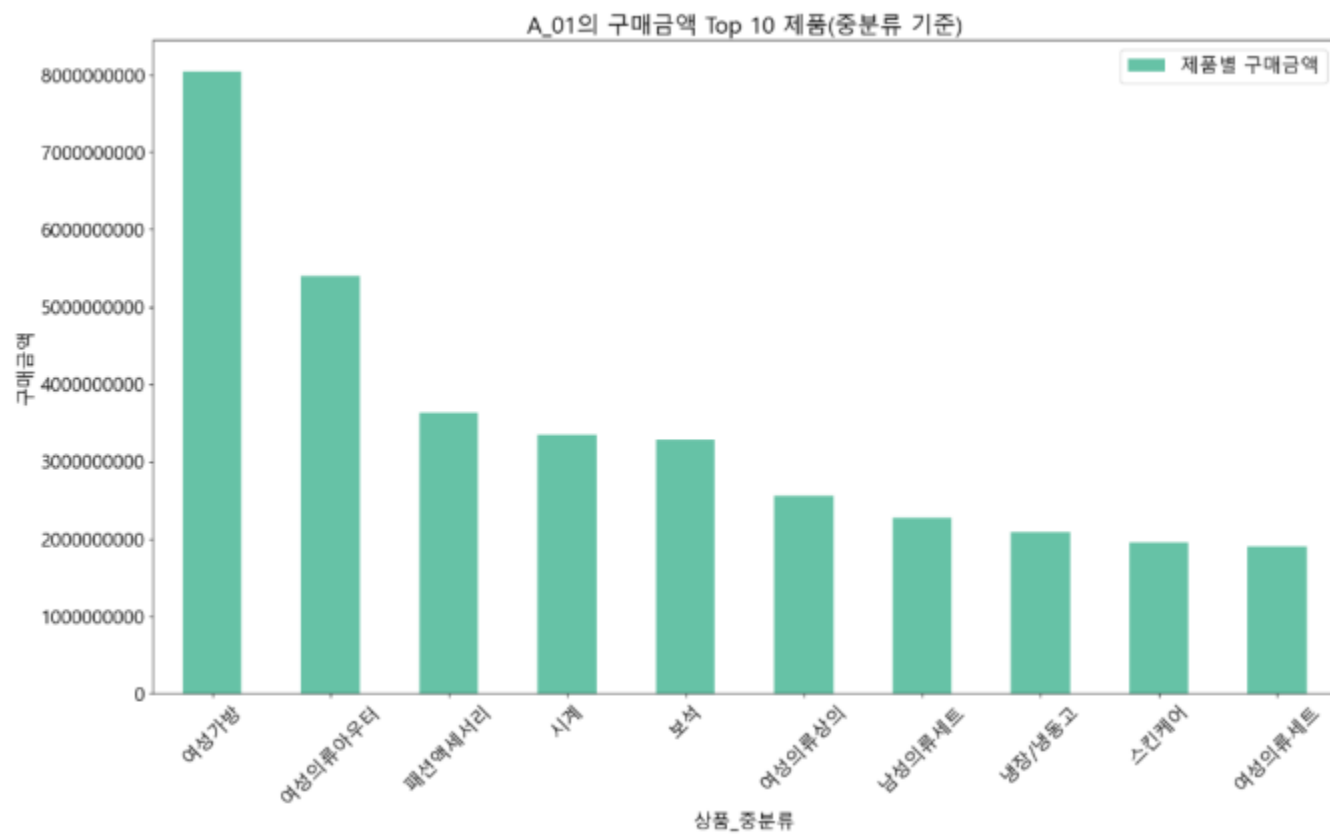
```
고객ID      0  
성별        0  
연령대      0  
거주지_대분류    0  
영수증번호    0  
온,오프라인구분    0  
유통사        0  
점포코드   428501  
상품코드      0  
구매일자      0  
구매시간      0  
구매금액      0  
구매수량      0  
상품_소분류    0  
상품_대분류    0  
상품_중분류    0  
거주지_소분류  428501  
dtype: int64
```

전처리 결과

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 4381743 entries, 0 to 4384738  
Data columns (total 13 columns):  
#   Column      Dtype  
---  -  
0   고객ID      object  
1   성별        object  
2   연령대      object  
3   영수증번호  object  
4   온,오프라인구분  float64  
5   유통사      object  
6   구매일자    float64  
7   구매시간    float64  
8   구매금액    float64  
9   구매수량    float64  
10  상품_소분류  object  
11  상품_대분류  object  
12  상품_중분류  object  
dtypes: float64(5), object(8)  
memory usage: 468.0+ MB
```

03 분석 & 시각화 과정

A01 : 백화점



구매금액 TOP10 제품(중분류)

가방, 시계, 보석등 사치품류 다수



99% 오프라인

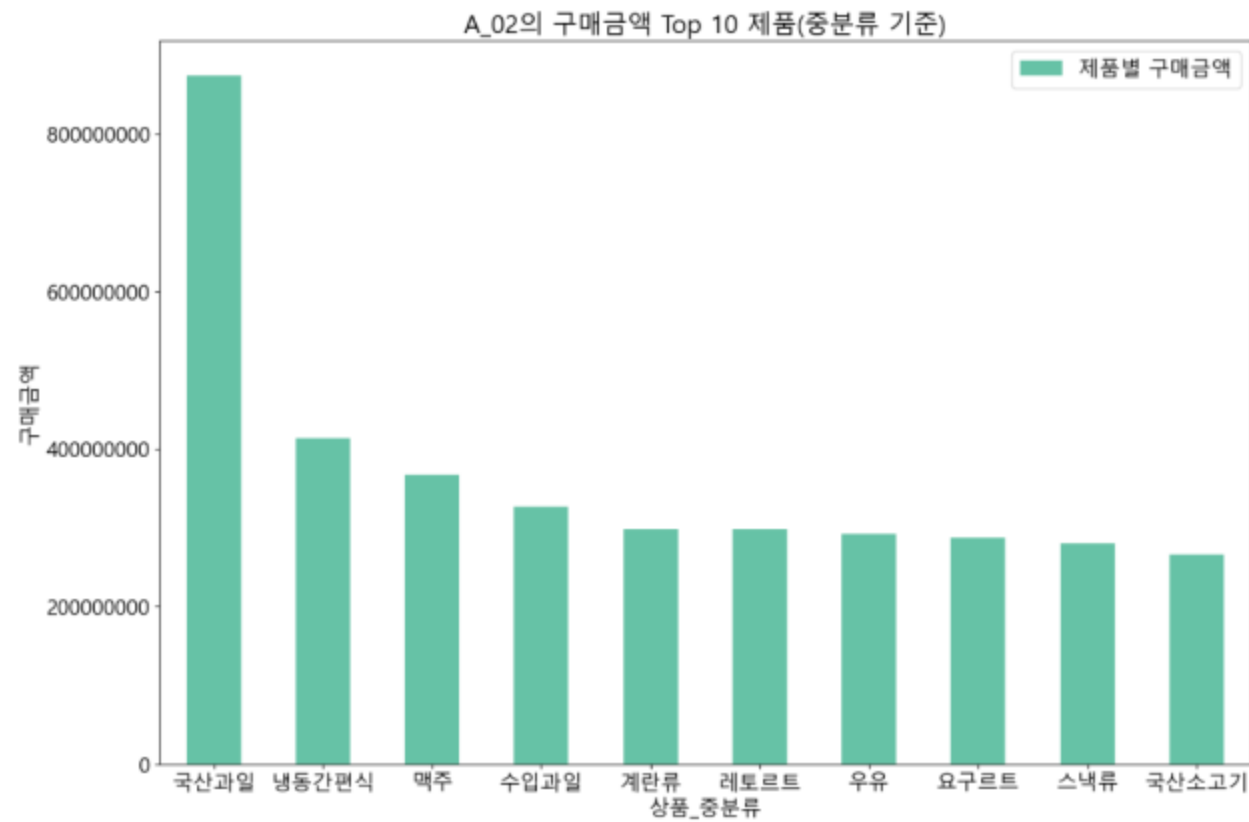


월별 최다 구매 제품

임대매출 다수

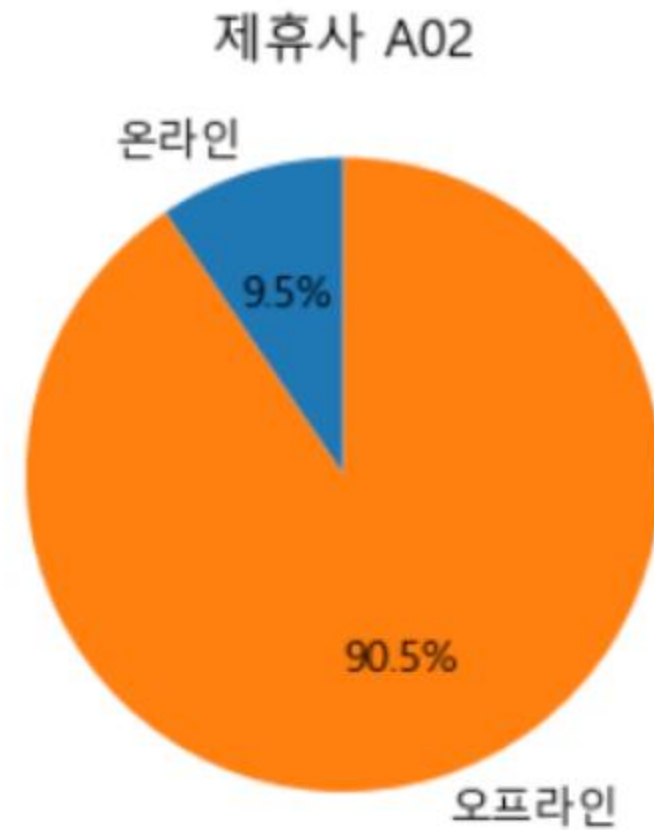
03 분석 & 시각화 과정

A02 : 대형마트

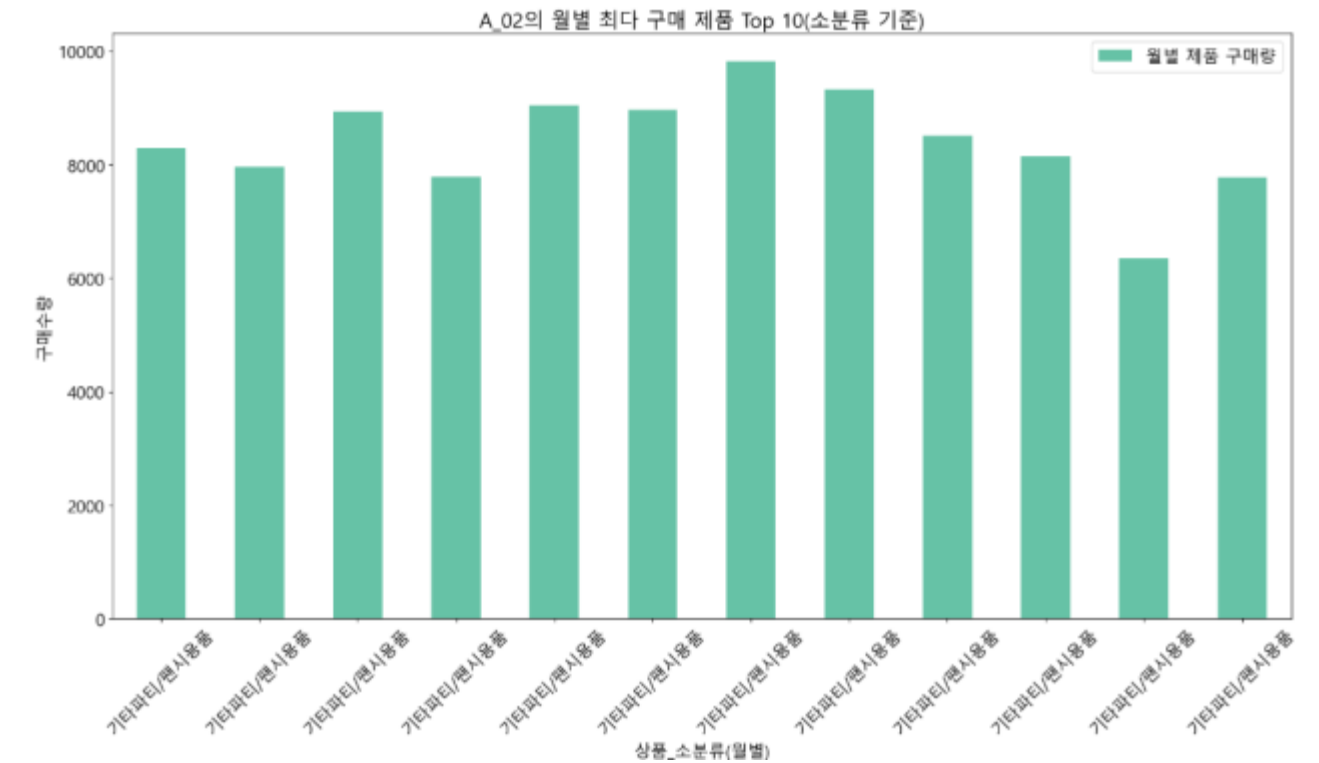


구매금액 TOP10 제품(중분류)

국산과일, 냉동간편식 등 식품류 다수



90% 오프라인

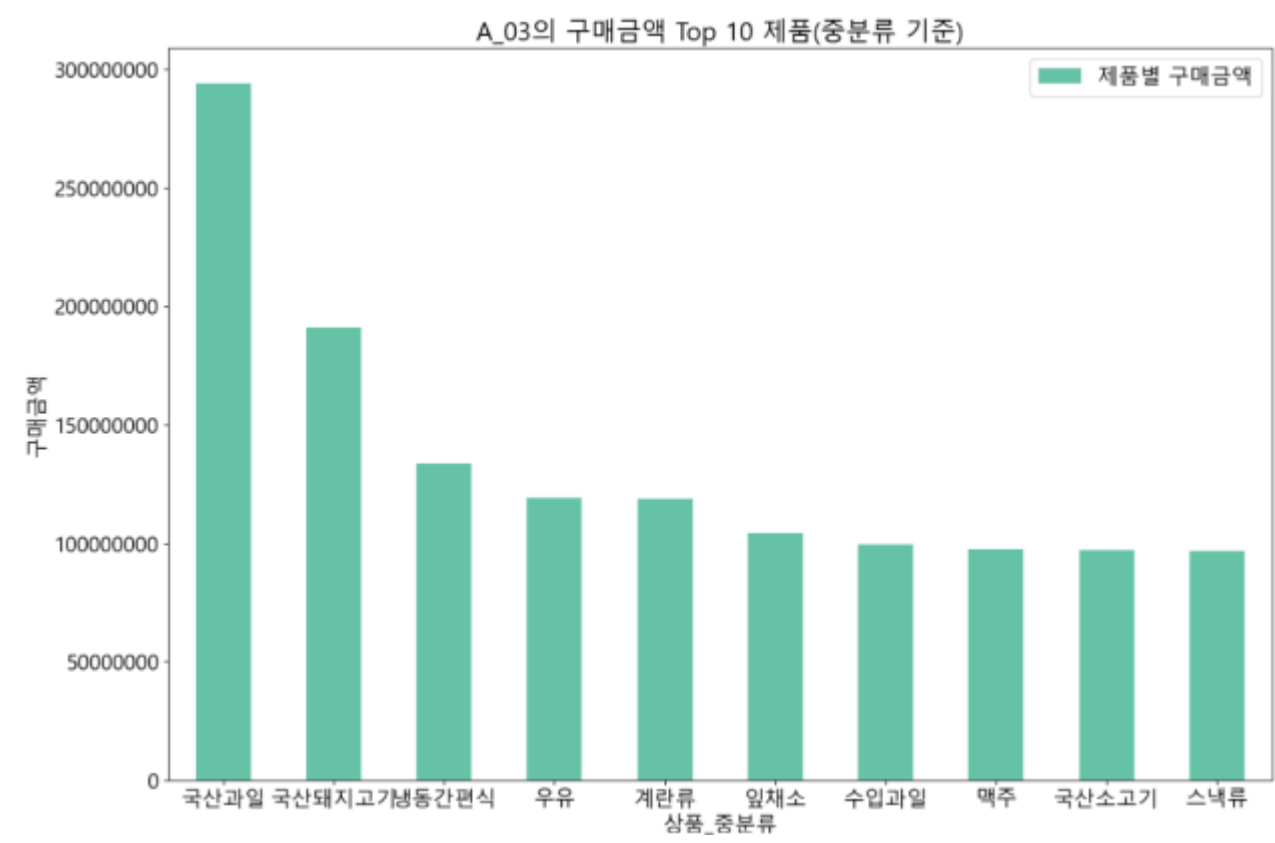


월별 최다 구매 제품

팬시용품 다수

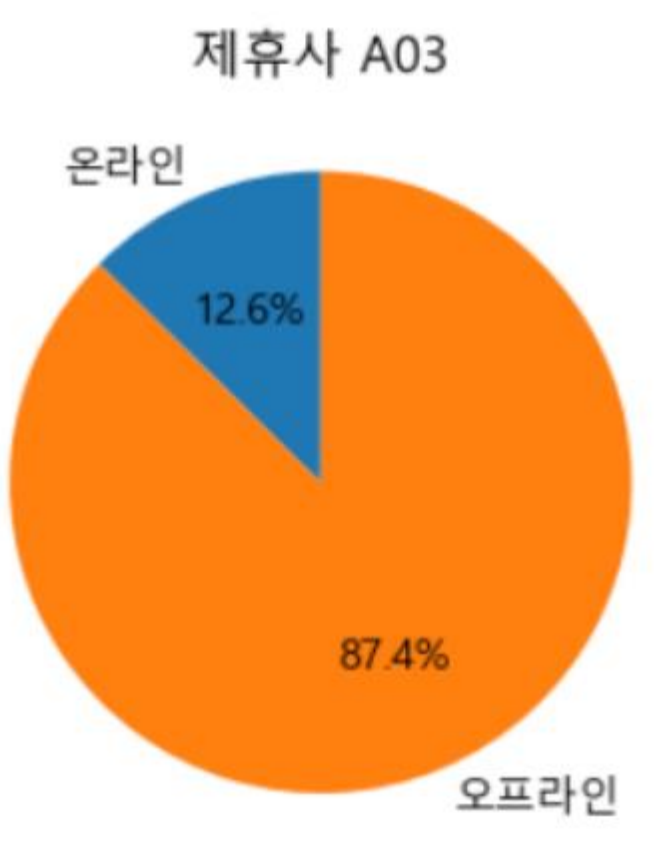
03 분석 & 시각화 과정

A03 : SSM

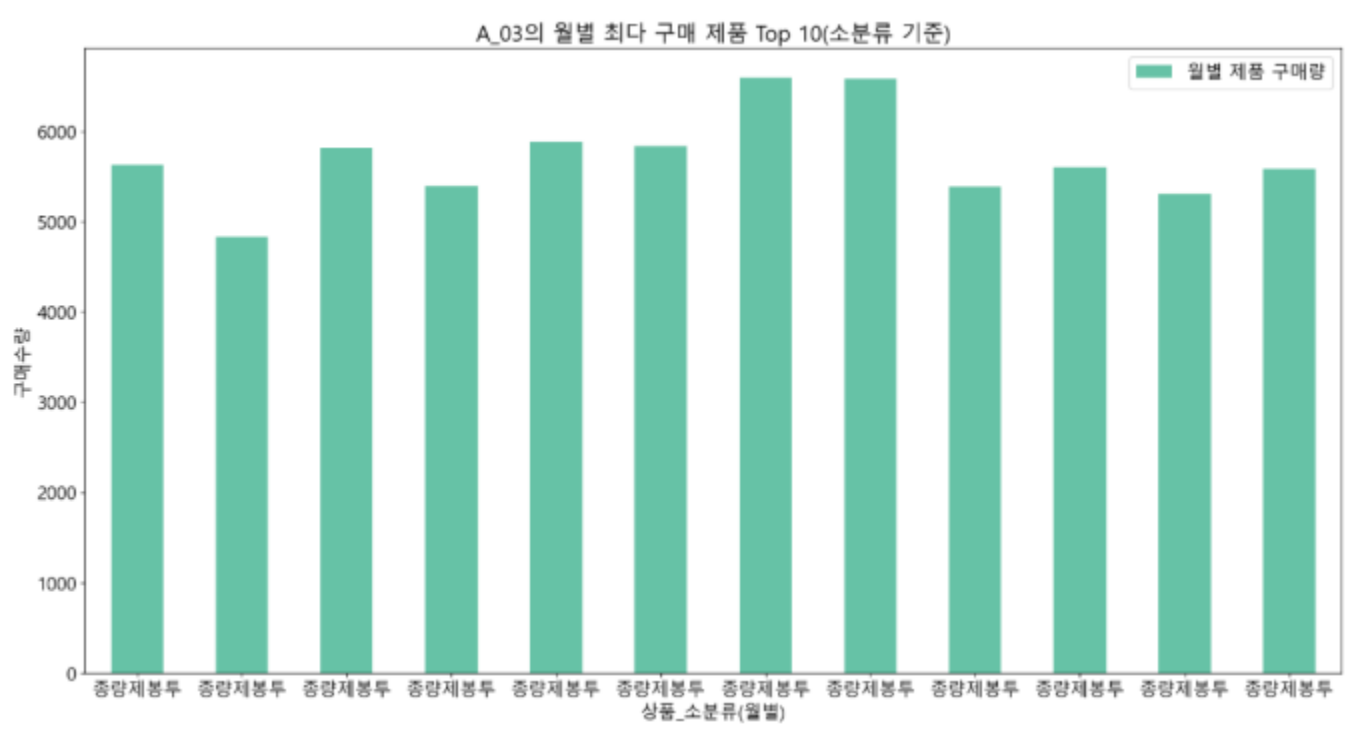


구매금액 TOP10 제품(중분류)

국산과일, 국산돼지고기 등 식품류 다수



87% 오프라인

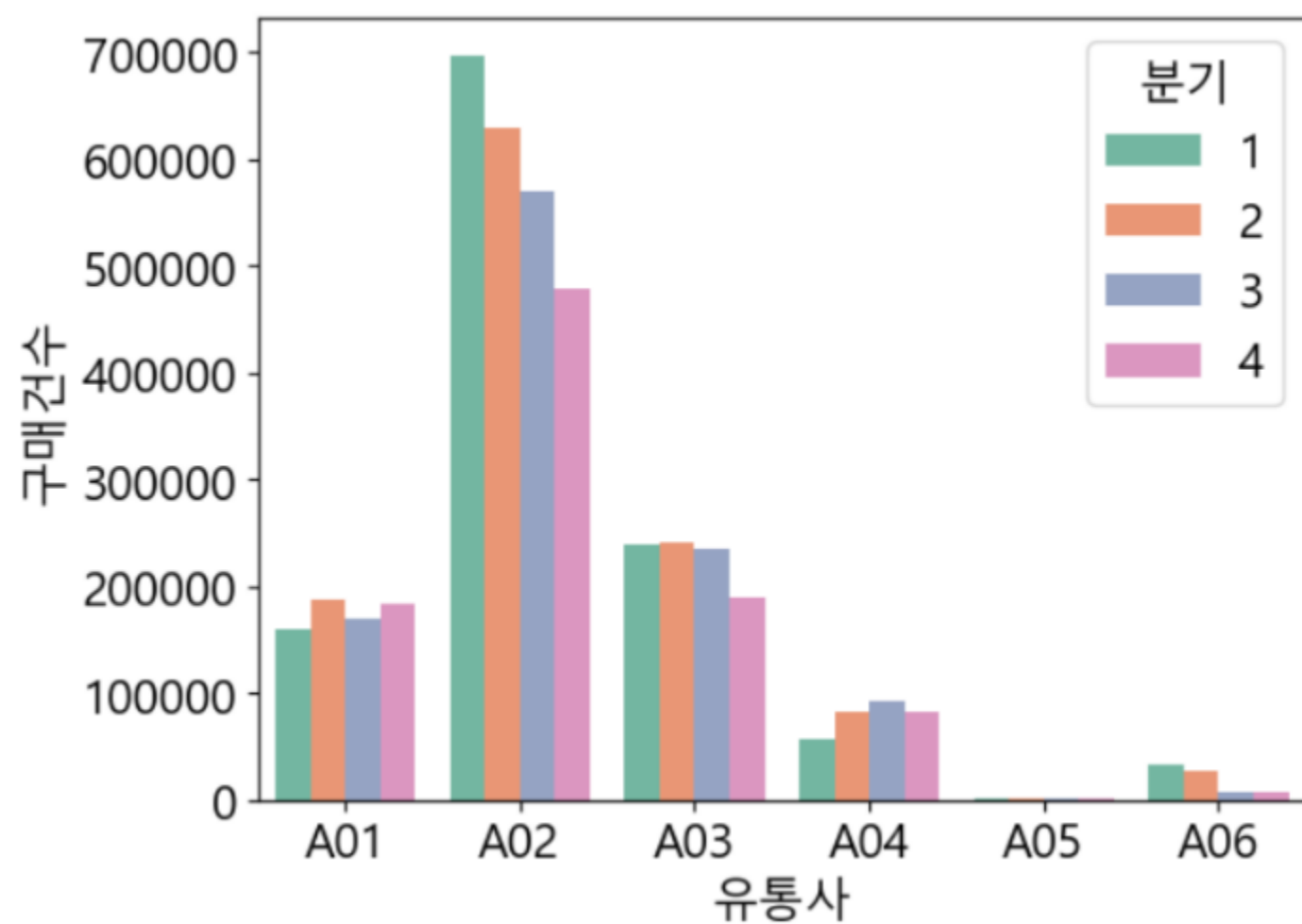


월별 최다 구매 제품

종량제봉투 다수

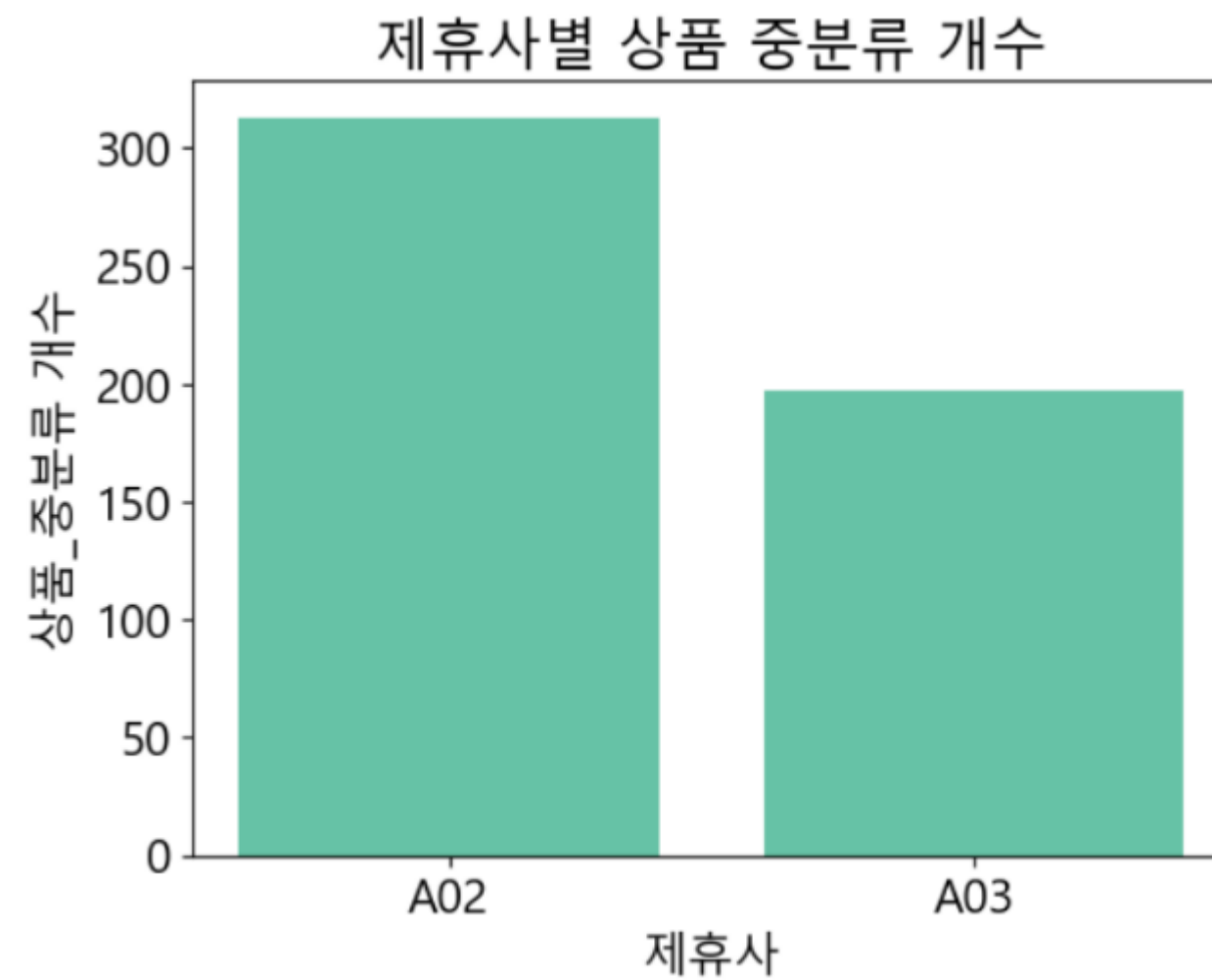
03 분석 & 시각화 과정

A02 vs A03



유통사별 구매건수

A02 > A03

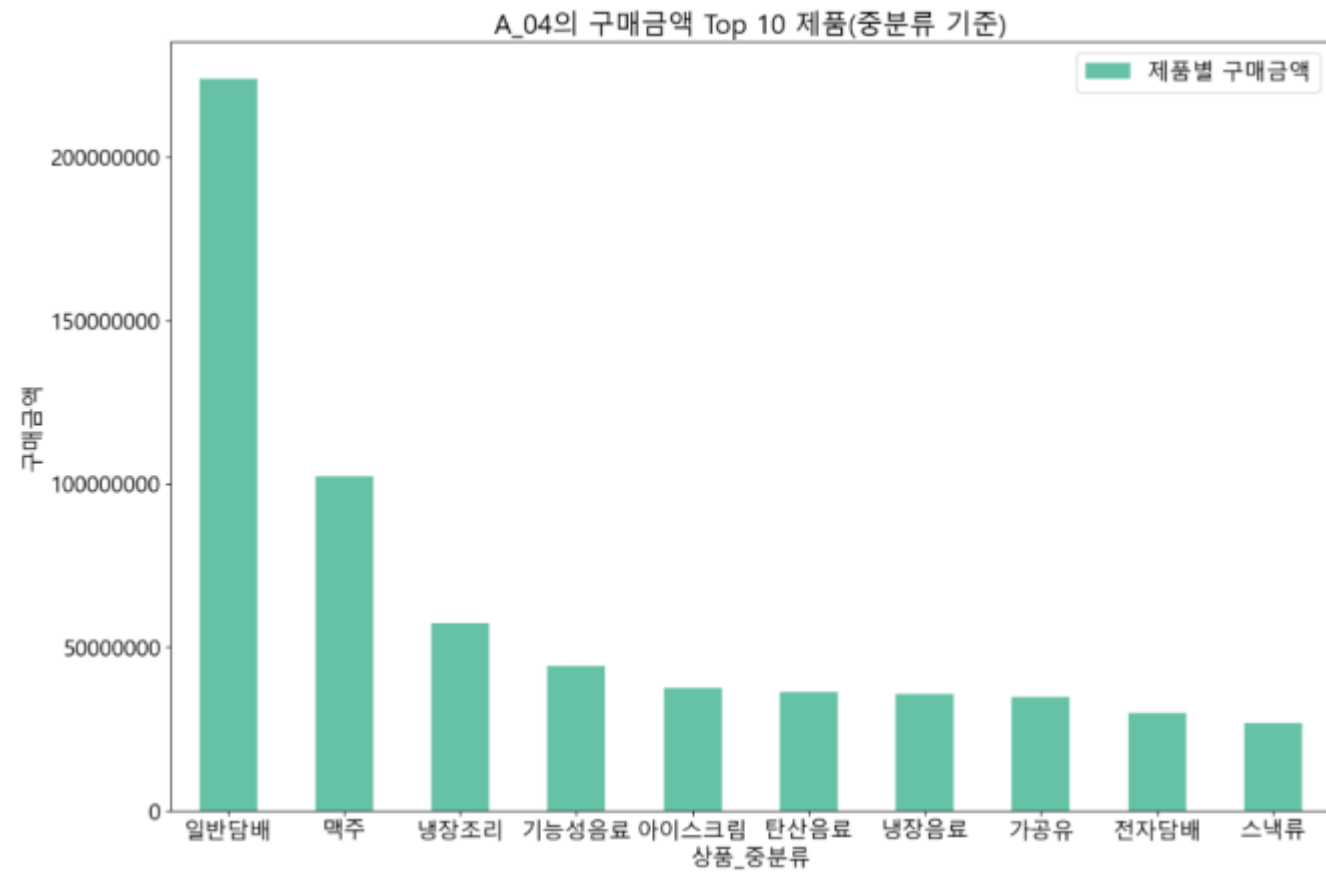


제휴사별 상품 중분류 개수

A02 > A03

03 분석 & 시각화 과정

A04 : 편의점

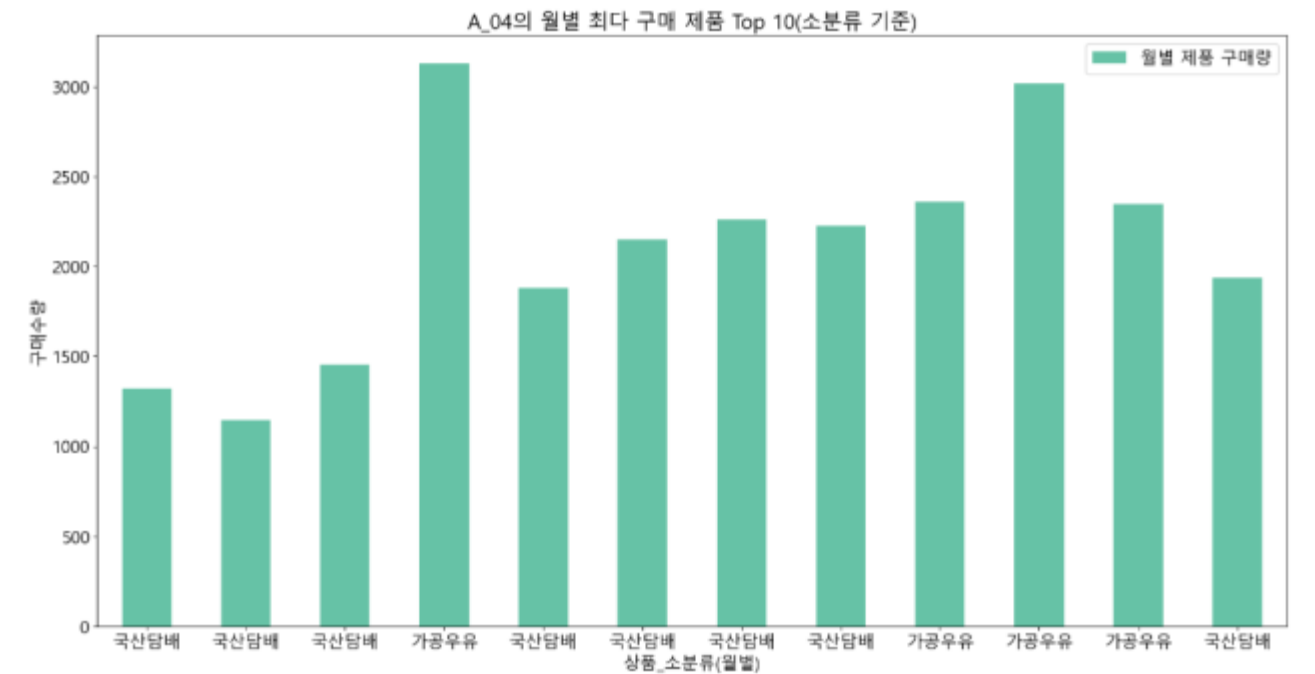


구매금액 TOP10 제품(중분류)

일반담배 구매금액이 전체금액의 21% 차지



100% 오프라인

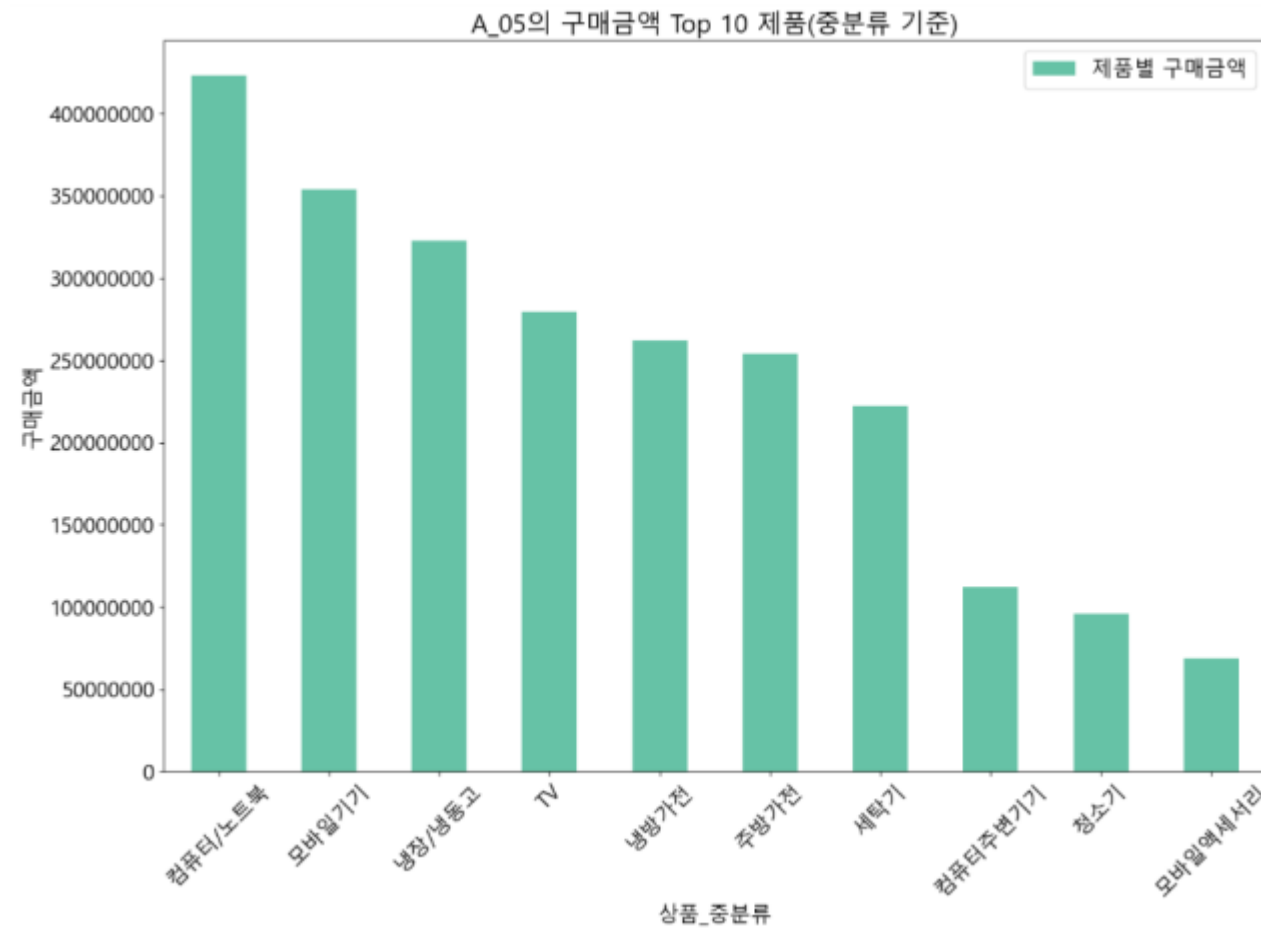


월별 최다 구매 제품

국산담배 다수

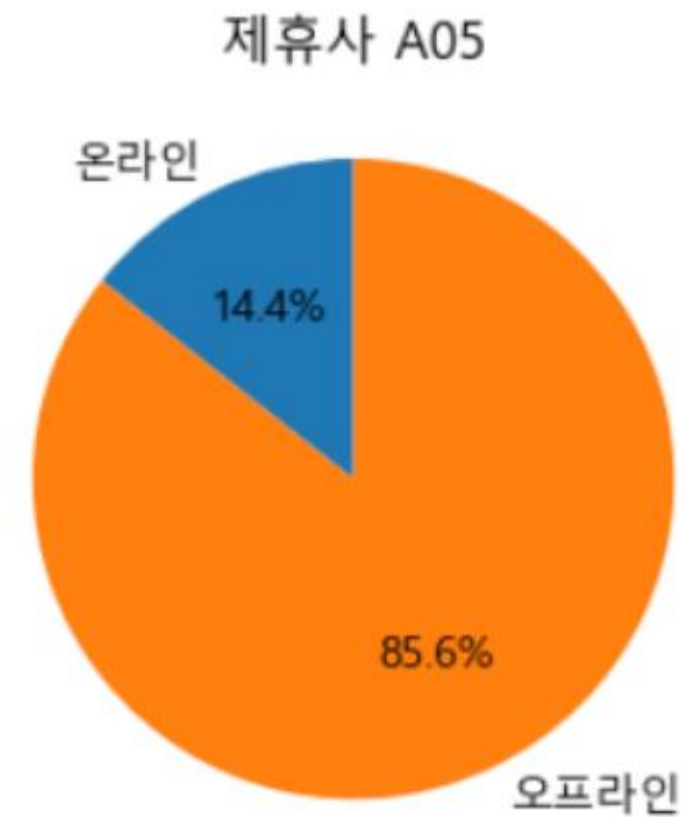
03 분석 & 시각화 과정

A05 : 전자제품 매장

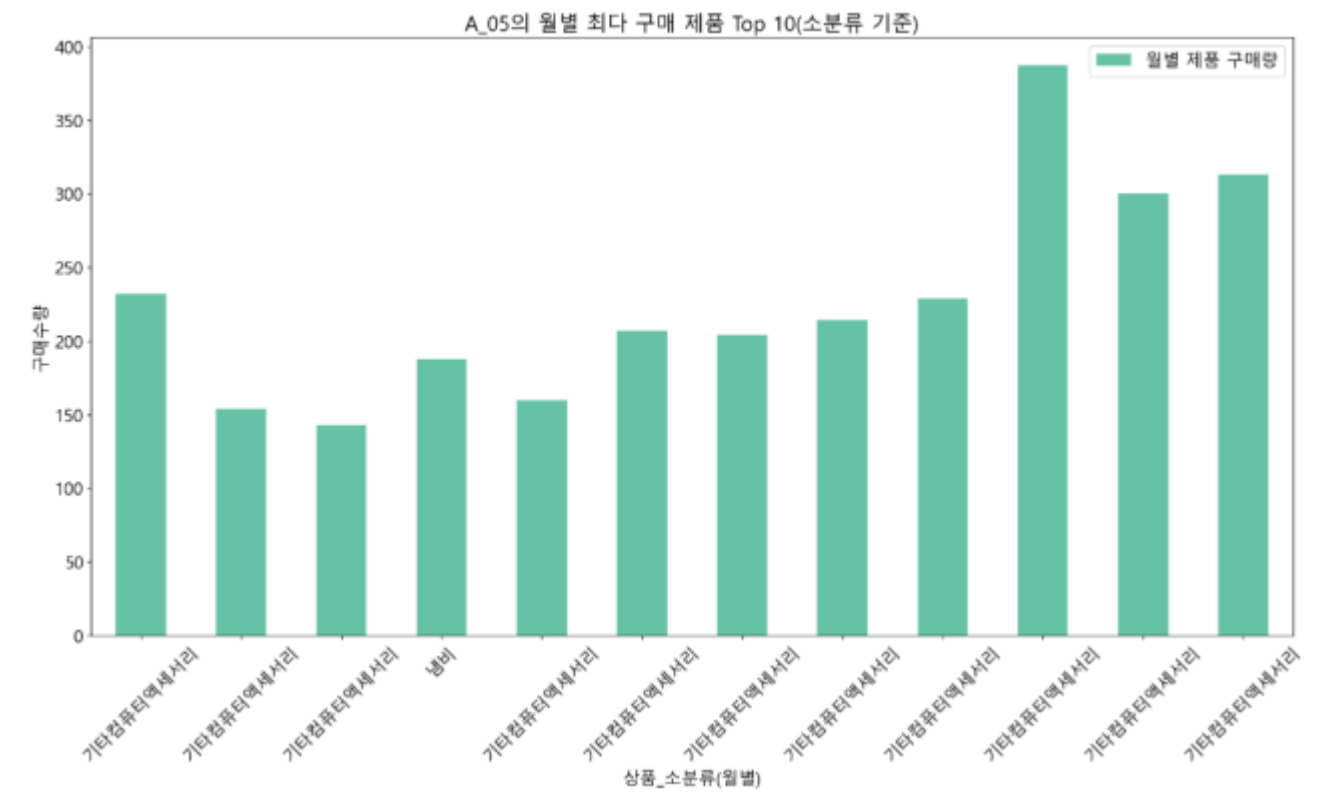


구매금액 TOP10 제품(중분류)

컴퓨터/노트북, 모바일 기기 다수



85% 오프라인

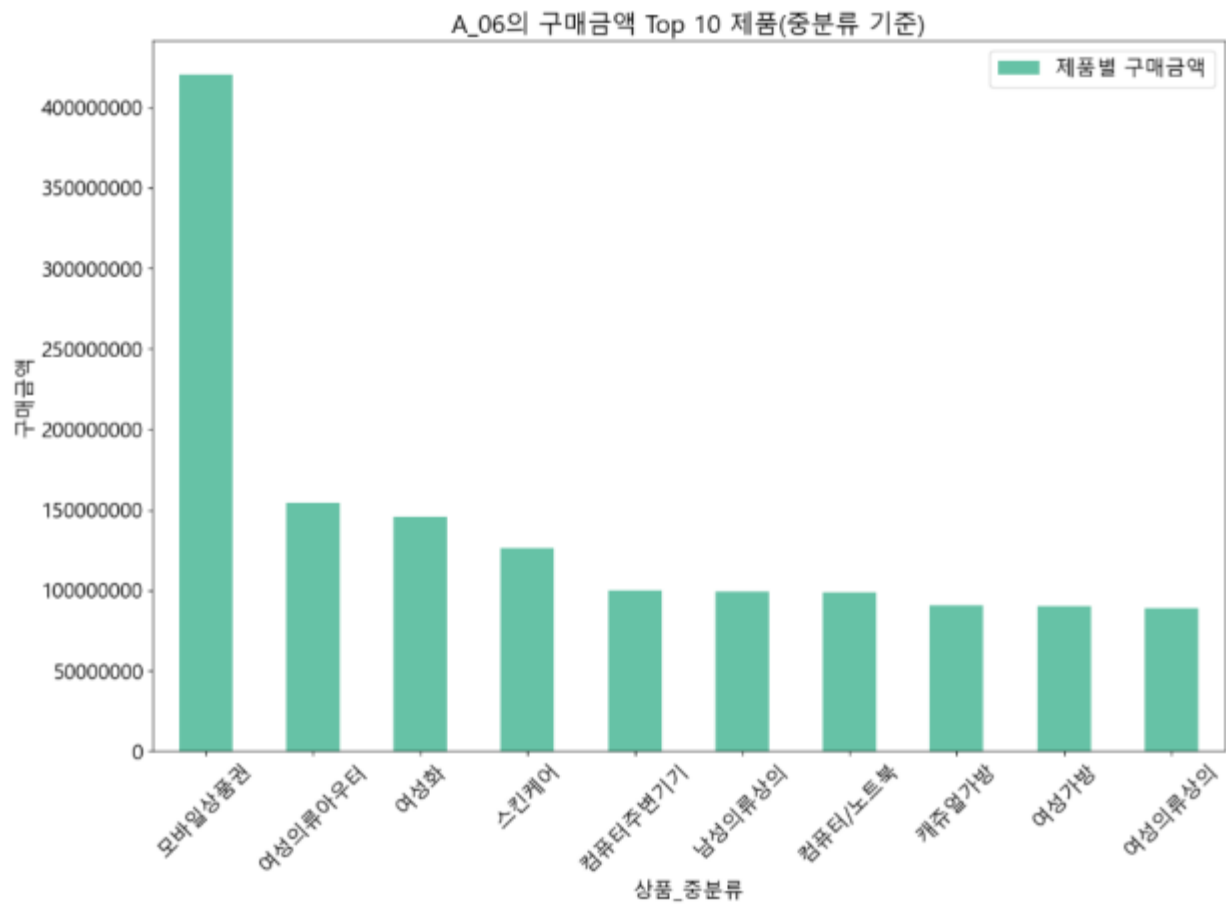


월별 최다 구매 제품

컴퓨터액세서리 다수

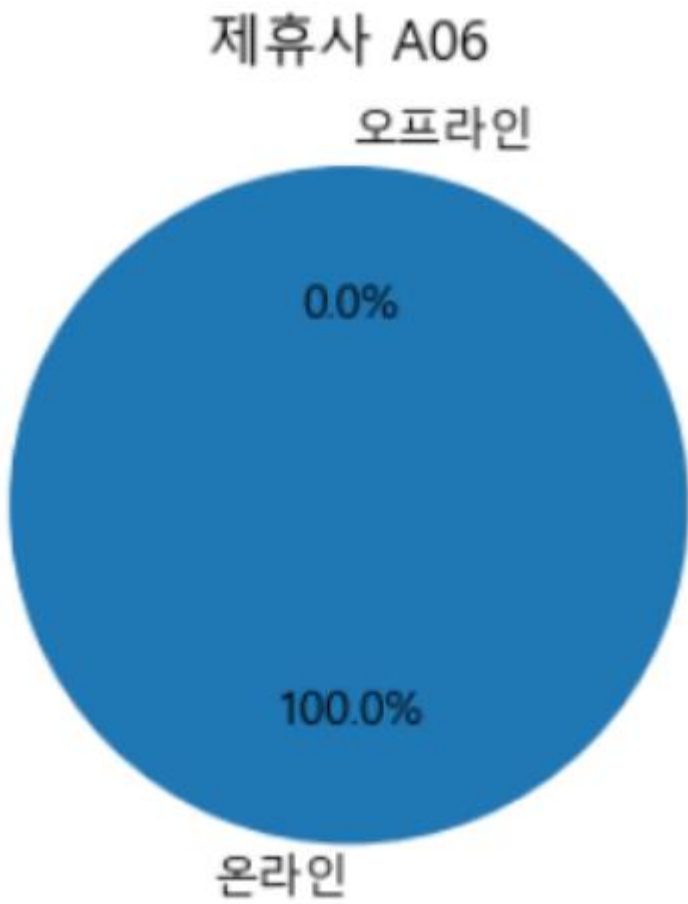
03 분석 & 시각화 과정

A06 : 온라인 매장

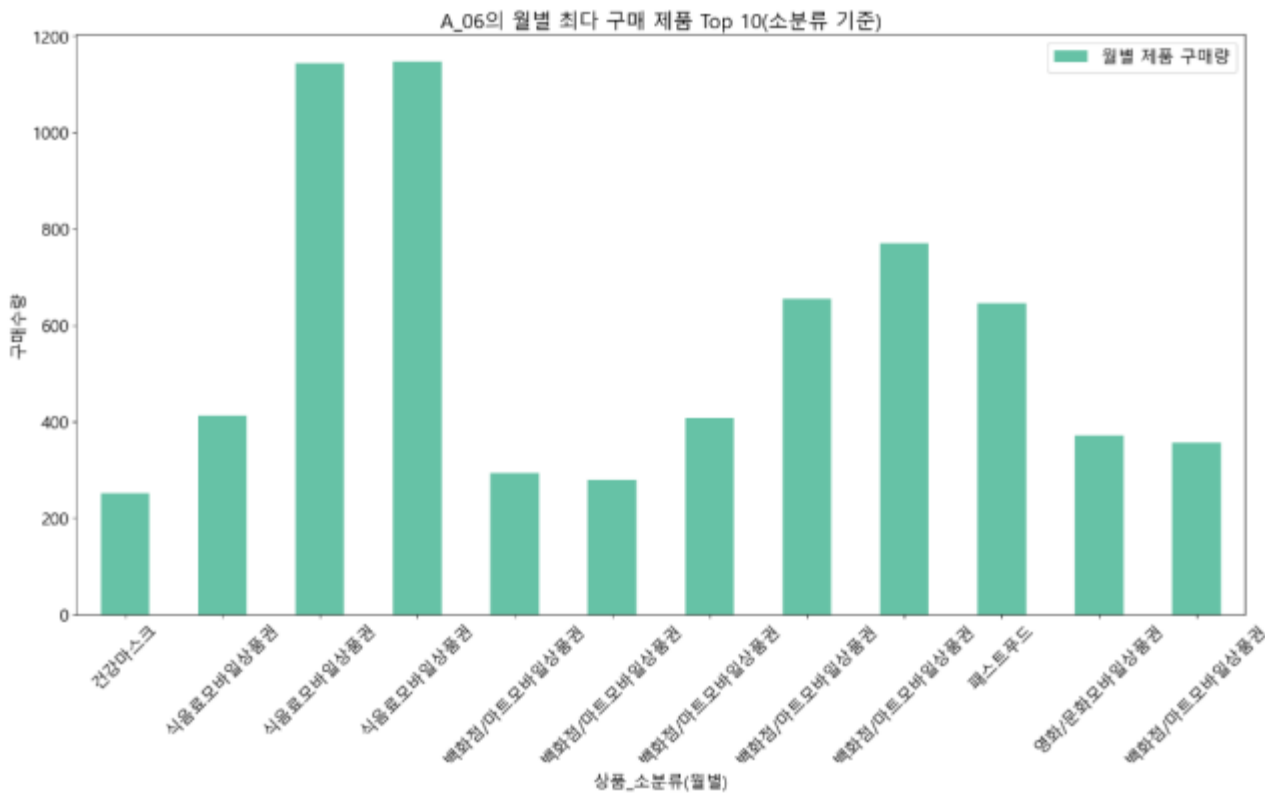


구매금액 TOP10 제품(중분류)

모바일상품권 다수



100% 온라인

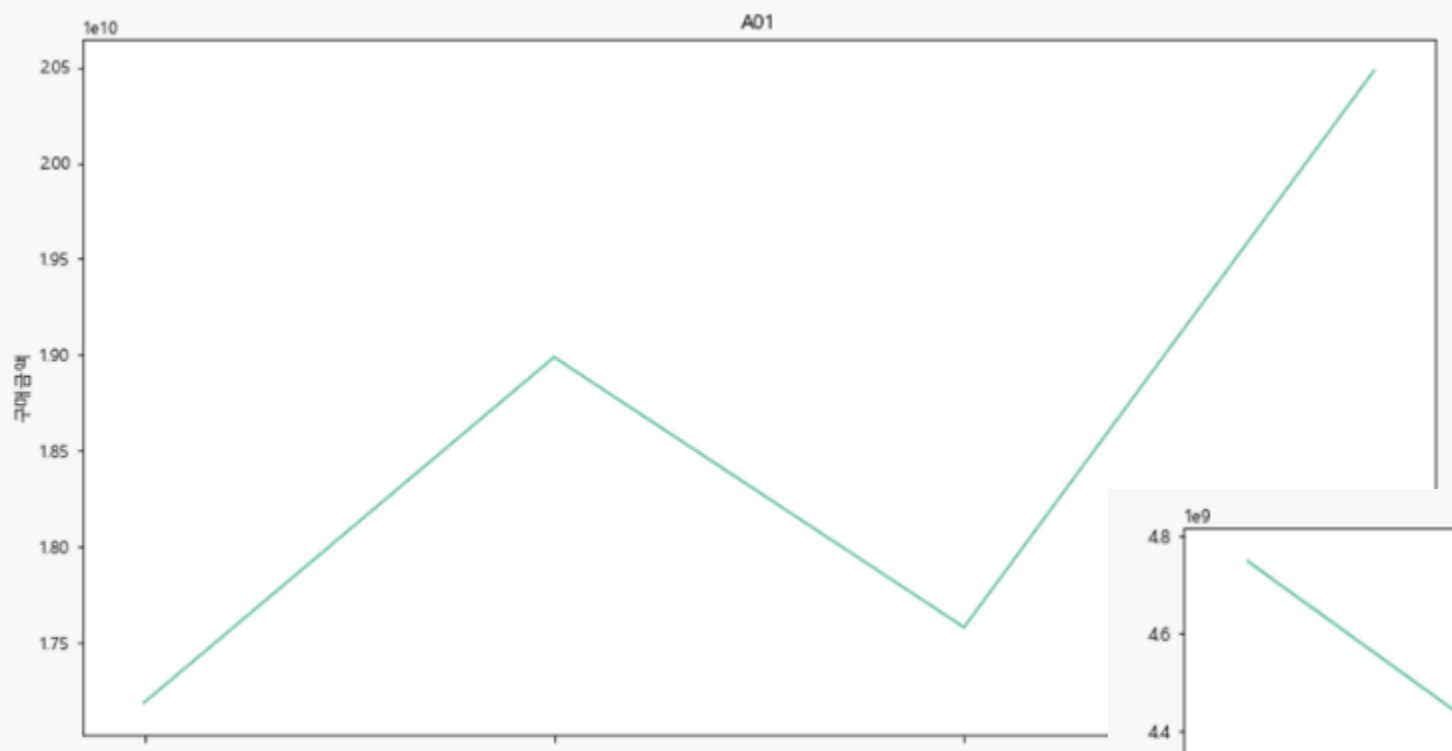


월별 최다 구매 제품

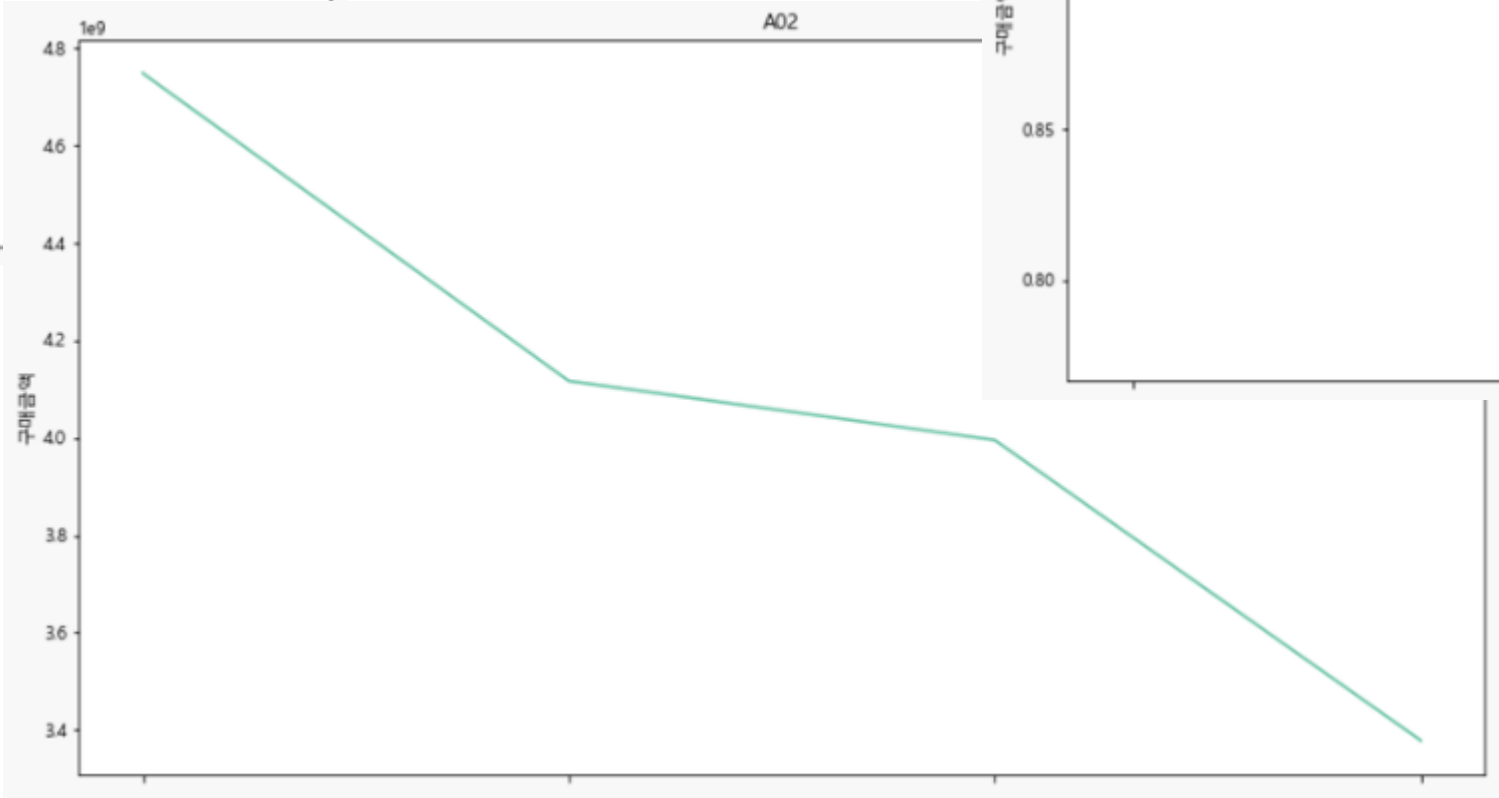
모바일상품권 다수

04 결론 및 특이점 도출

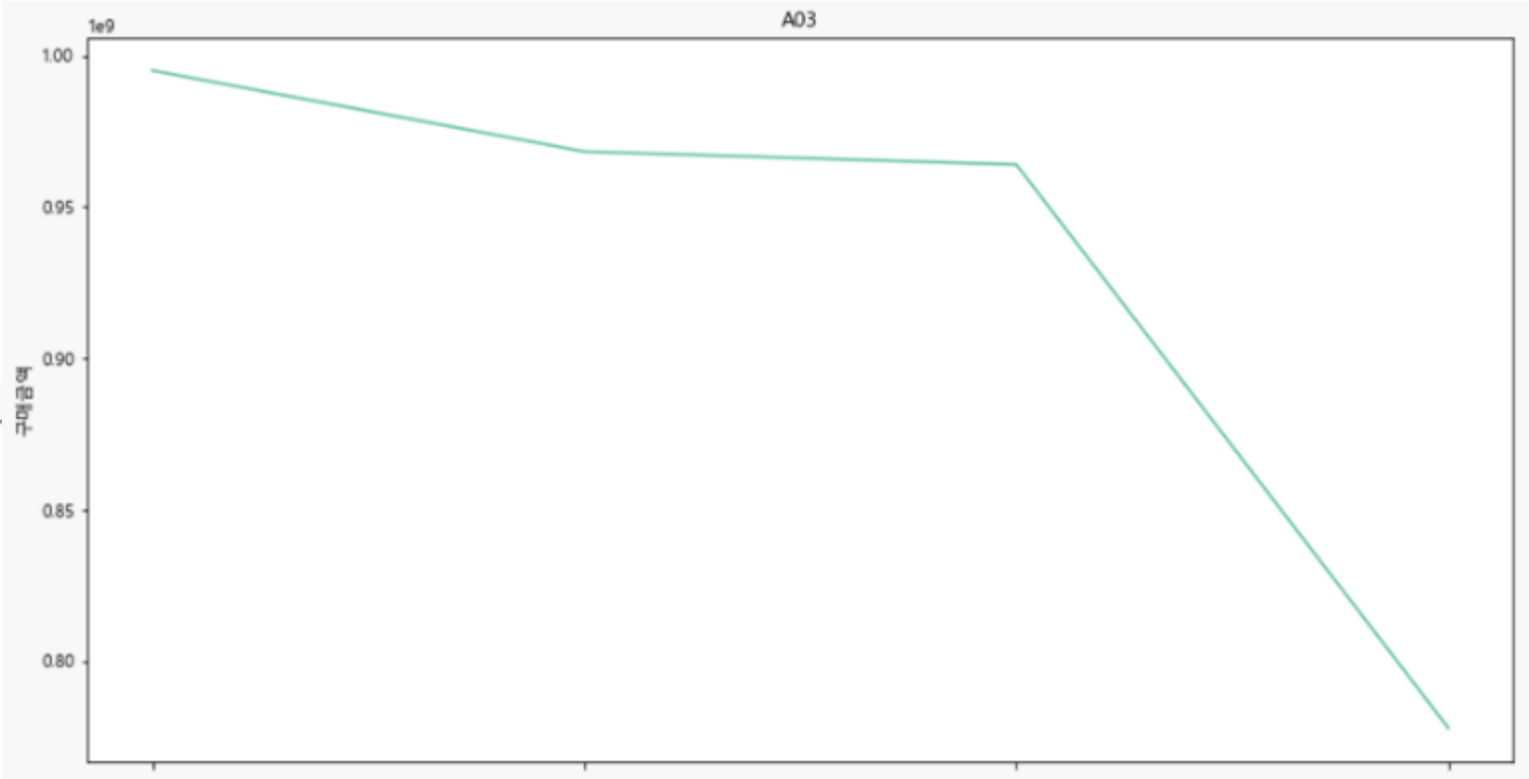
분기별 매출액 - A01, A02, A03



A01



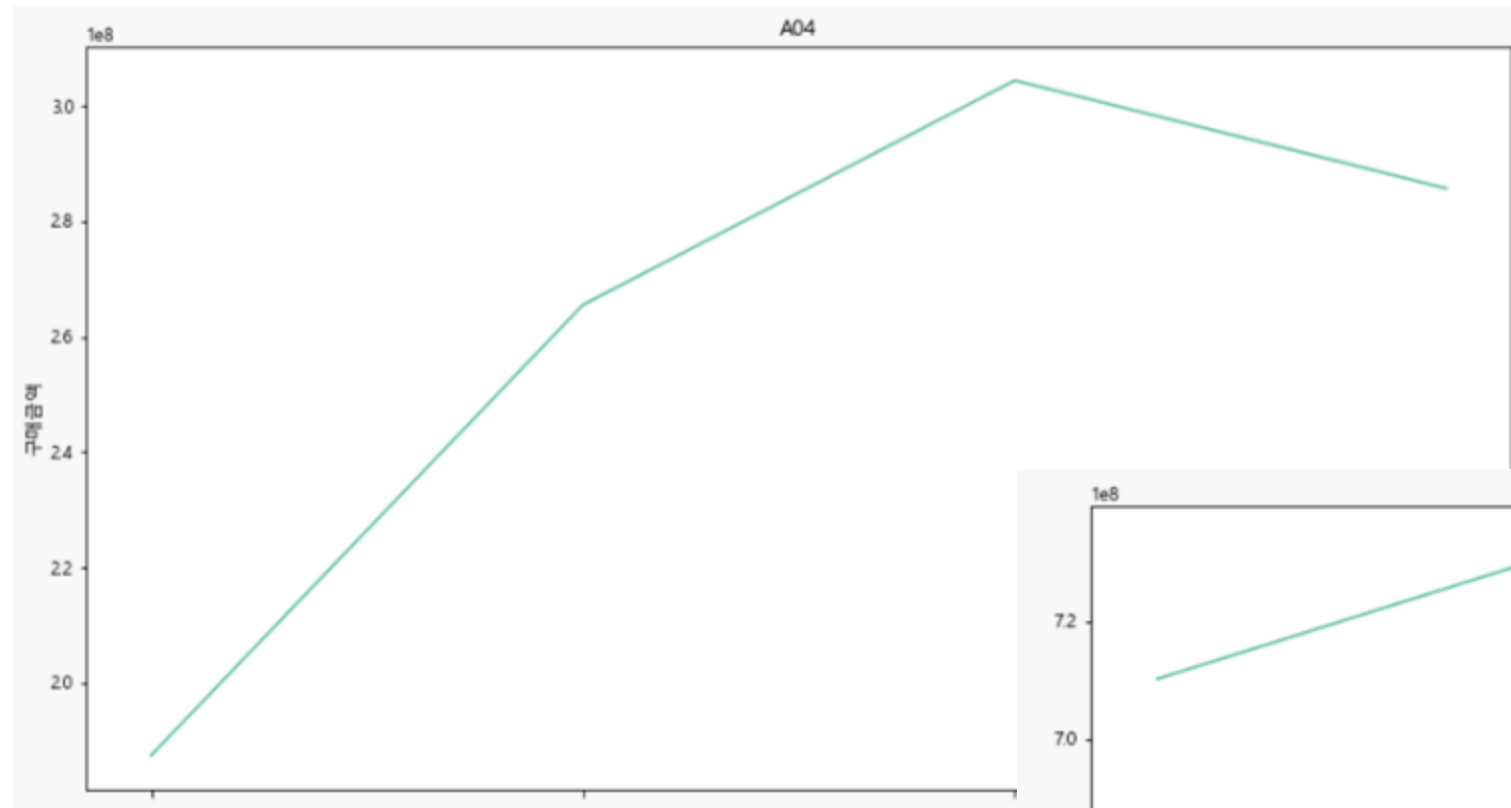
A02



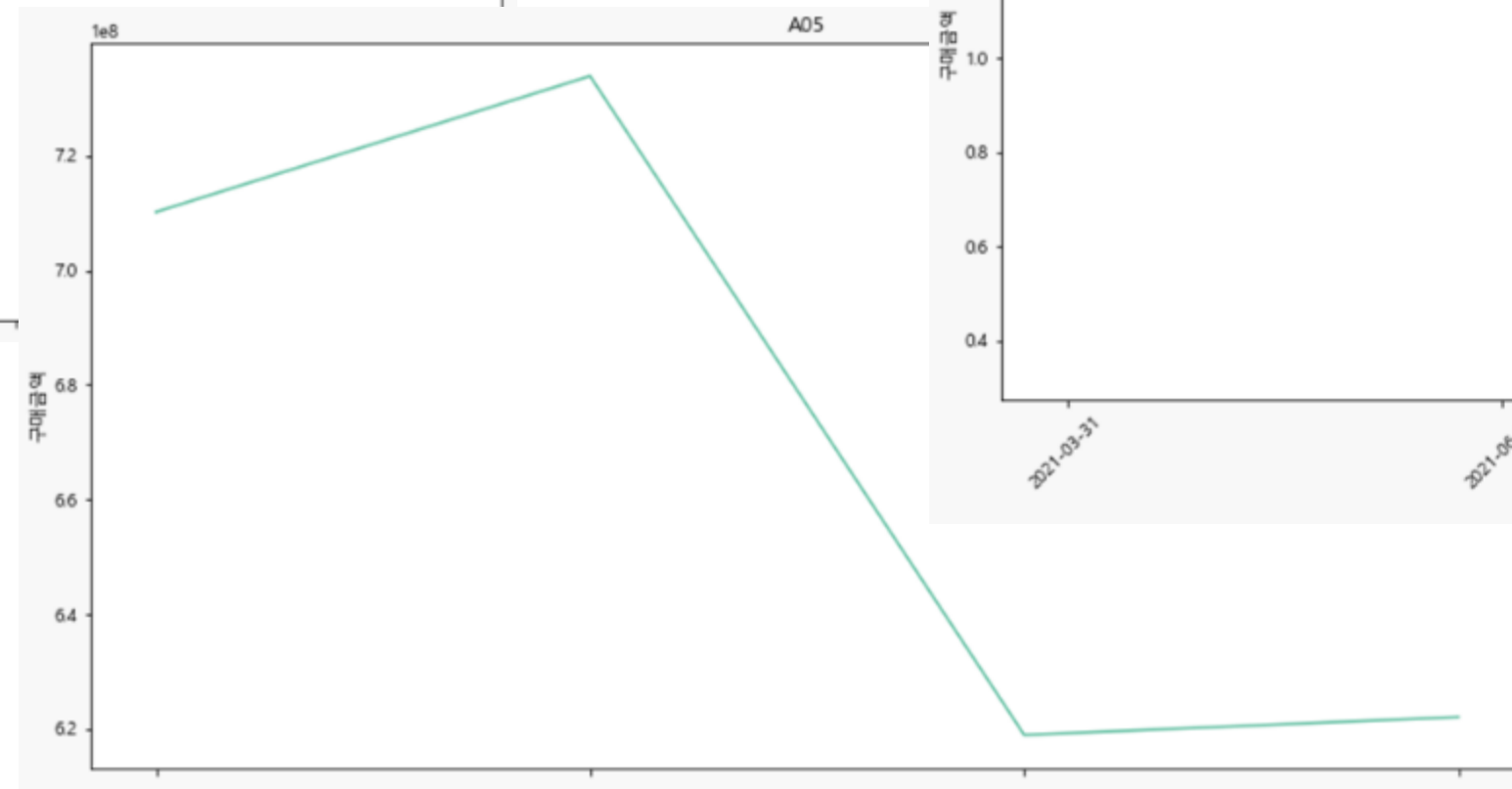
A03

04 결론 및 특이점 도출

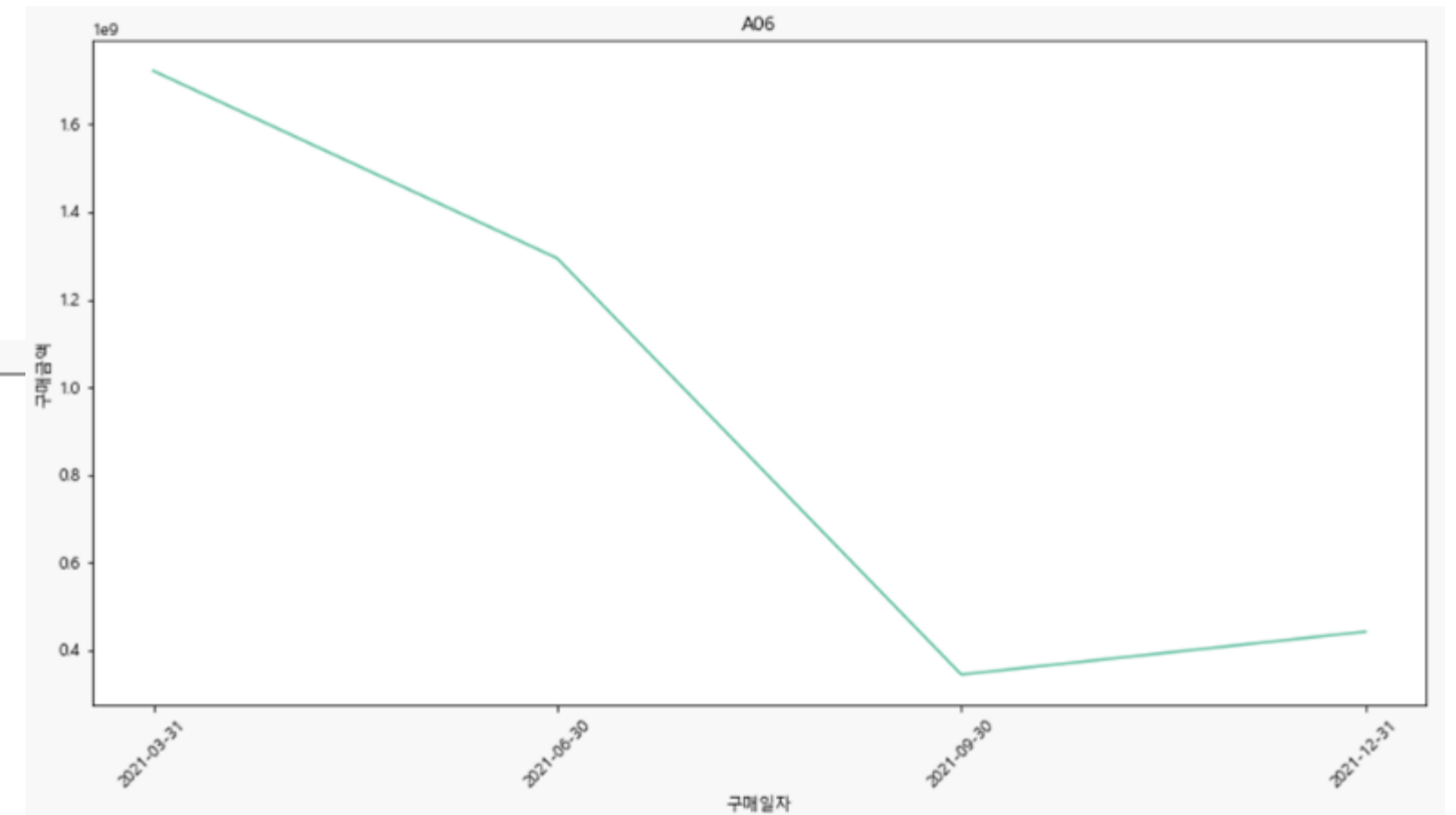
분기별 매출액 - A04, A05, A06



A04



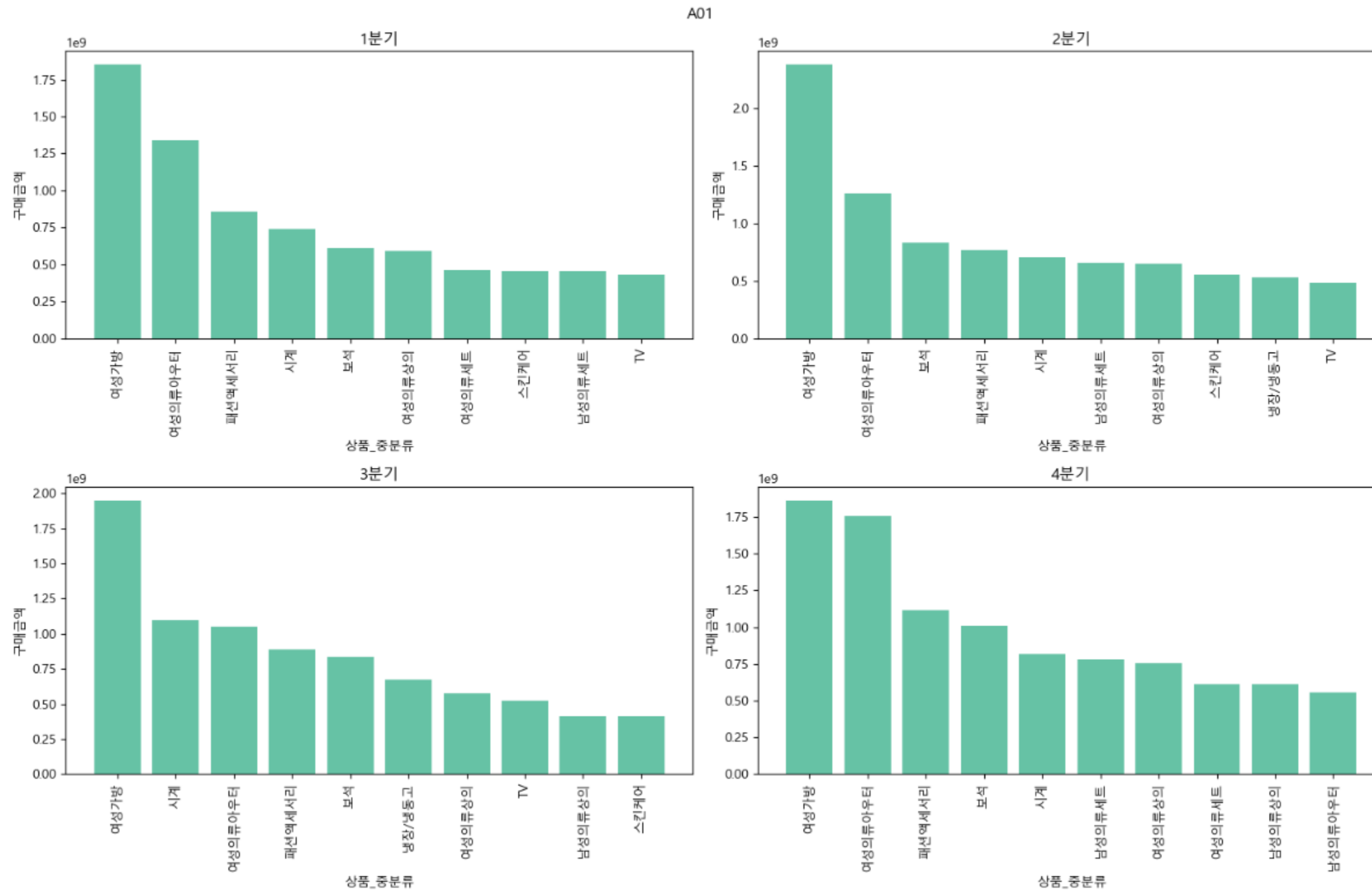
A05



A06

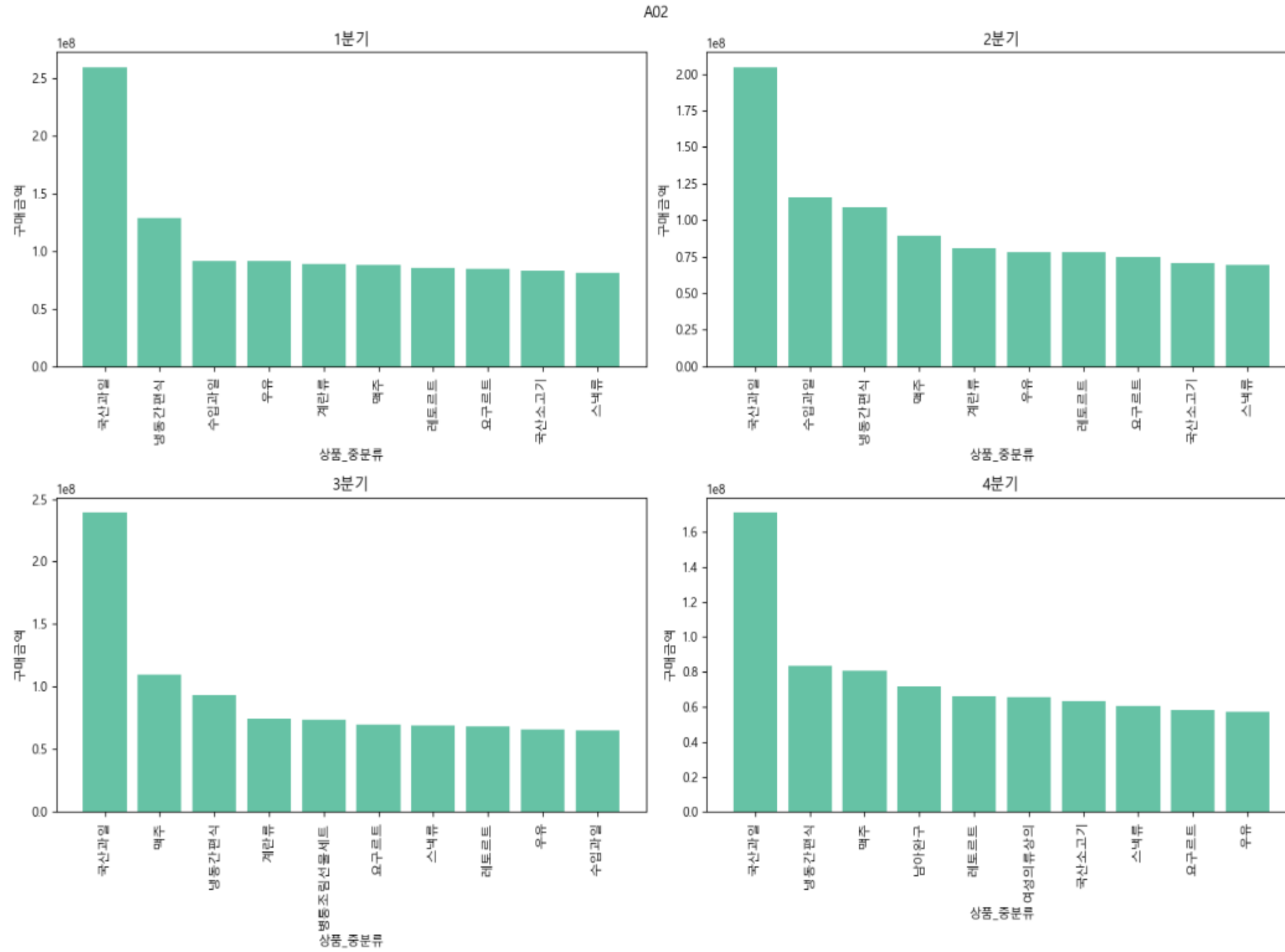
04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A01



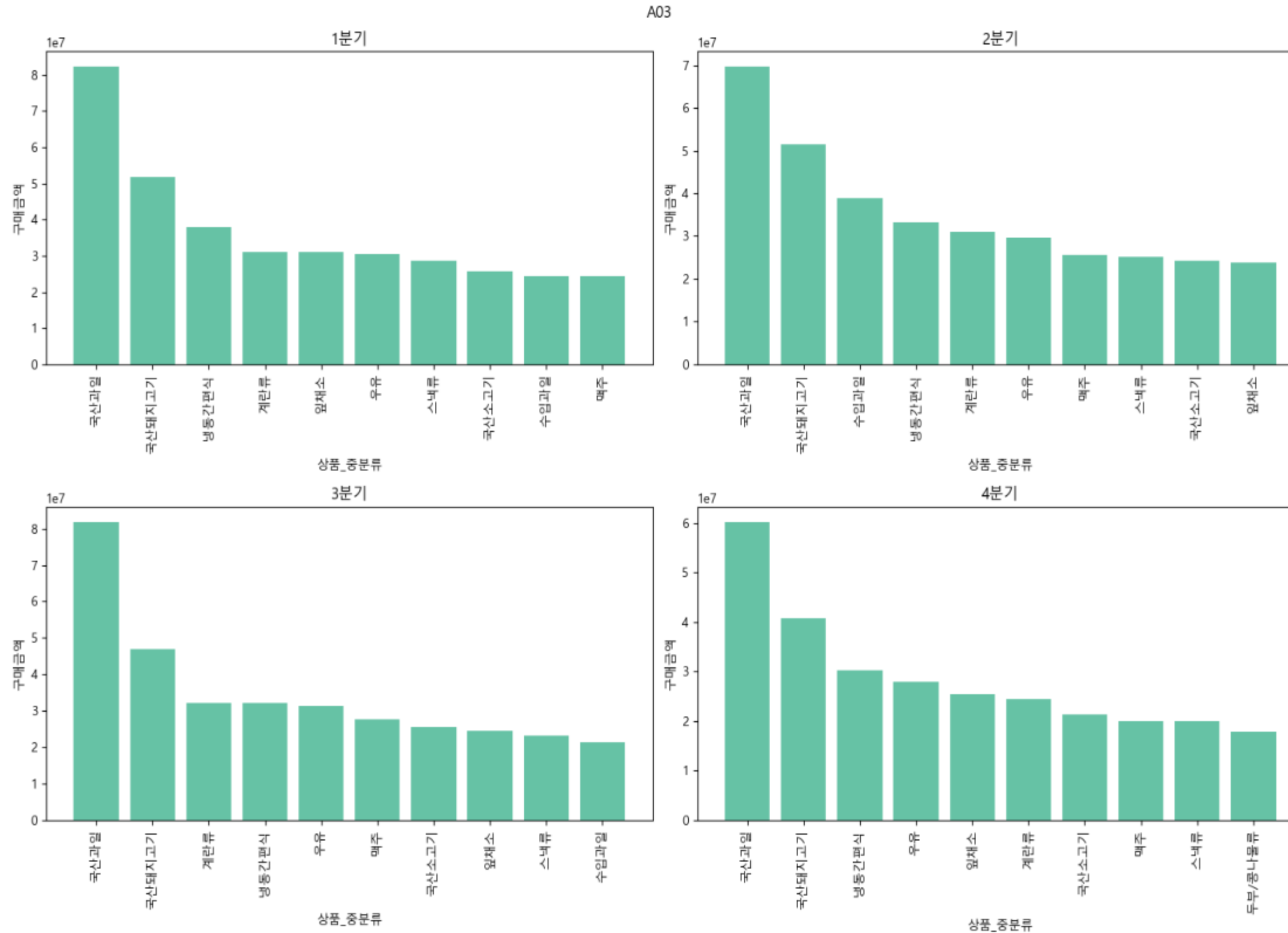
04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A02



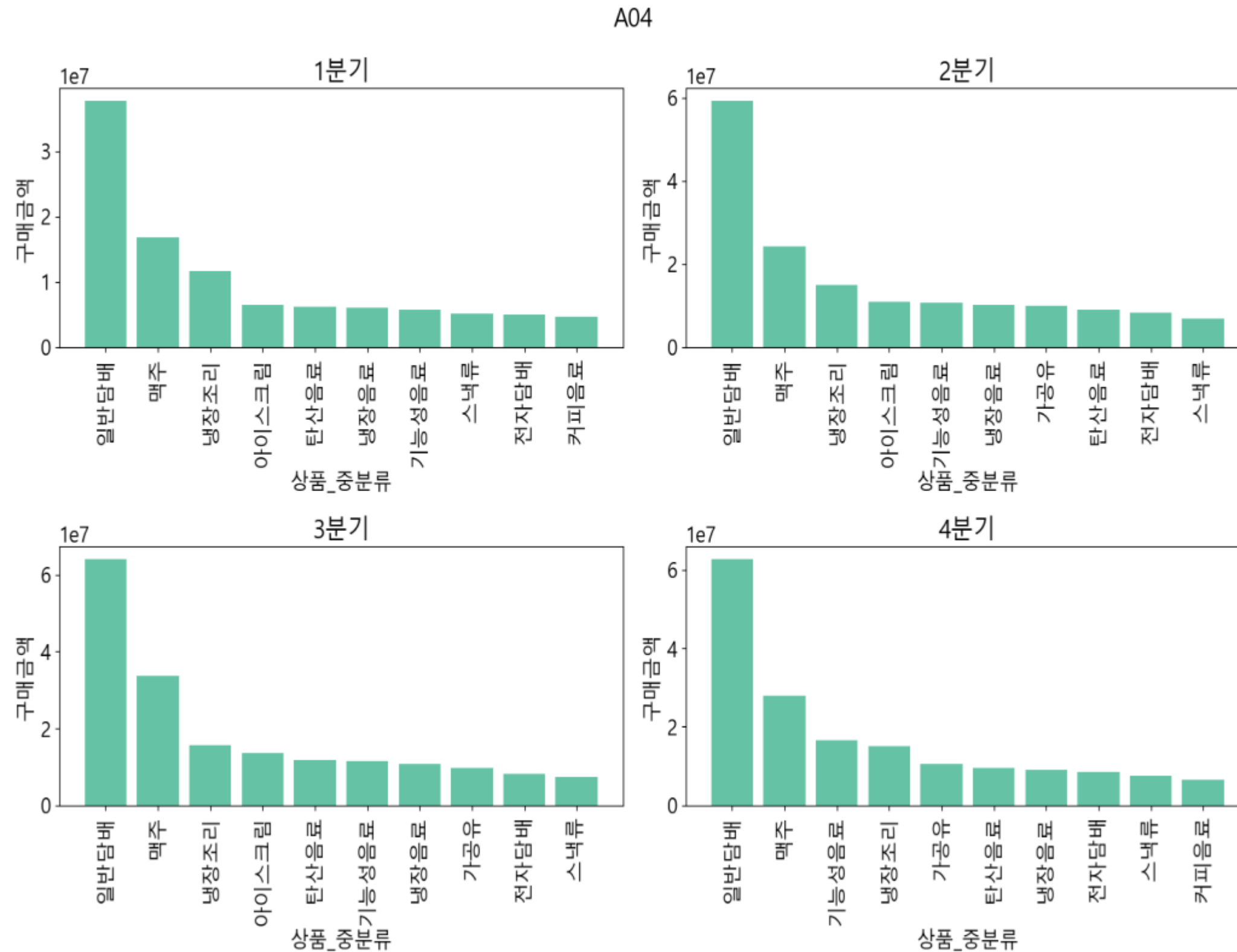
04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A03



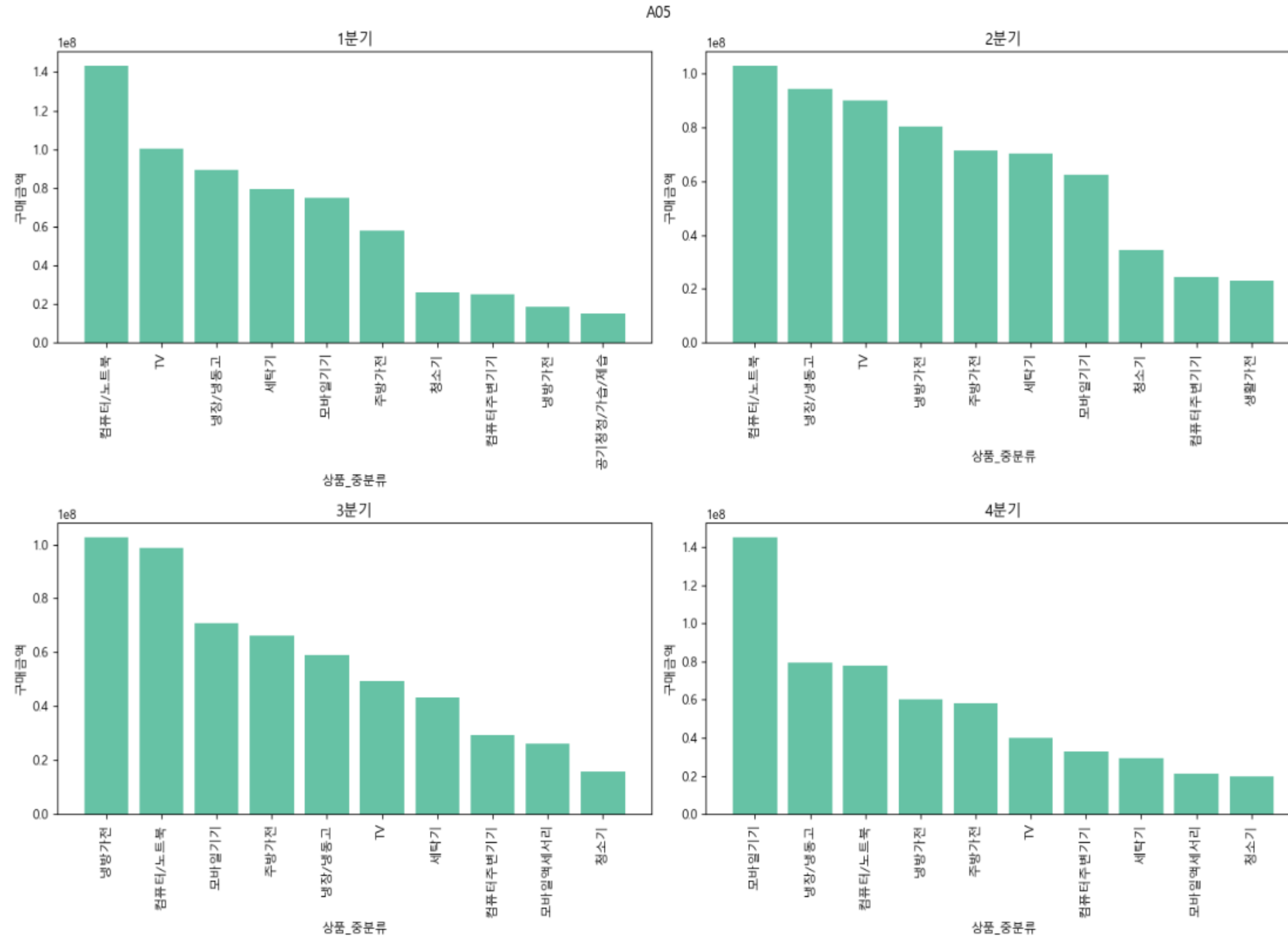
04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A04



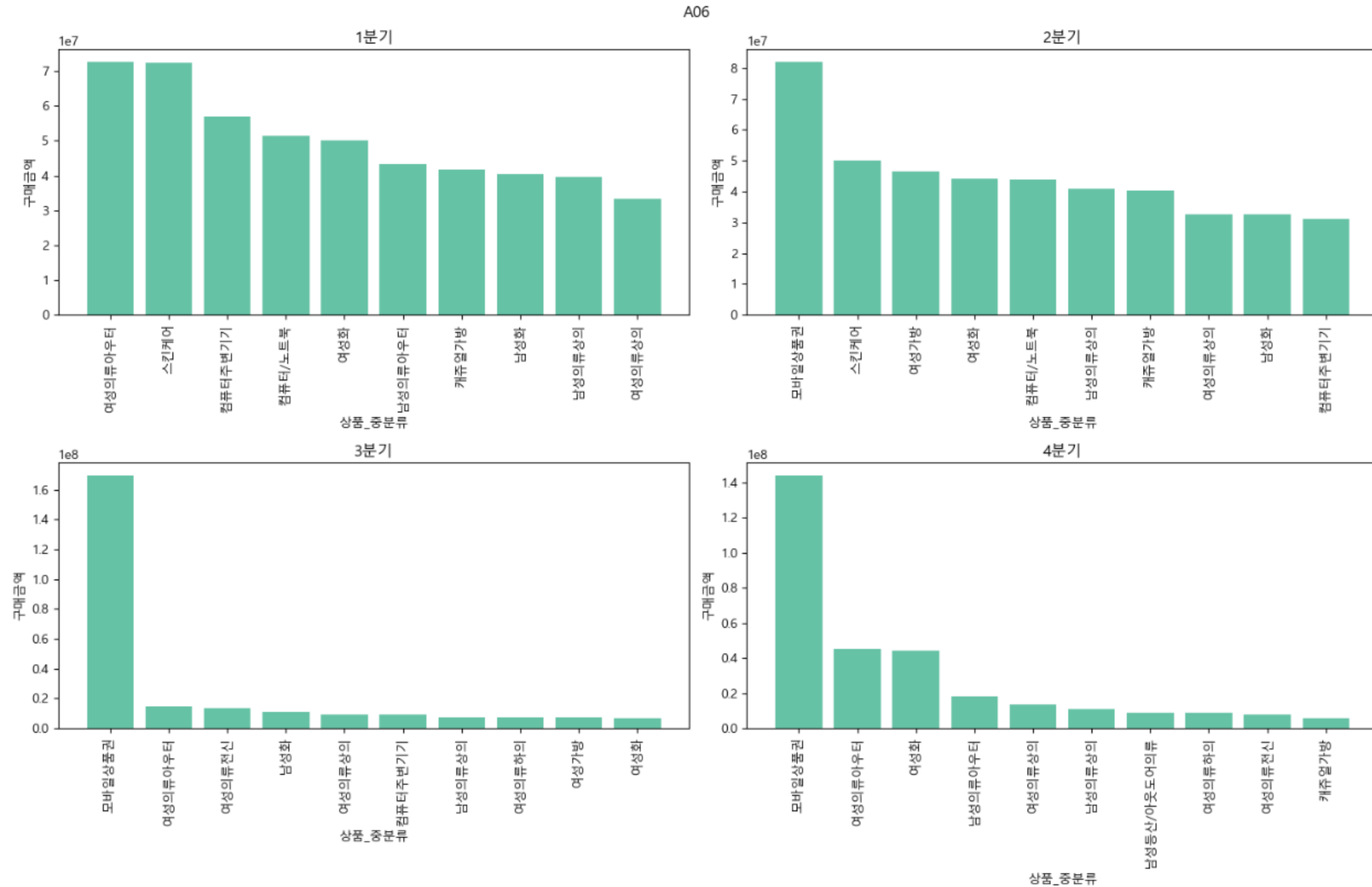
04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A05

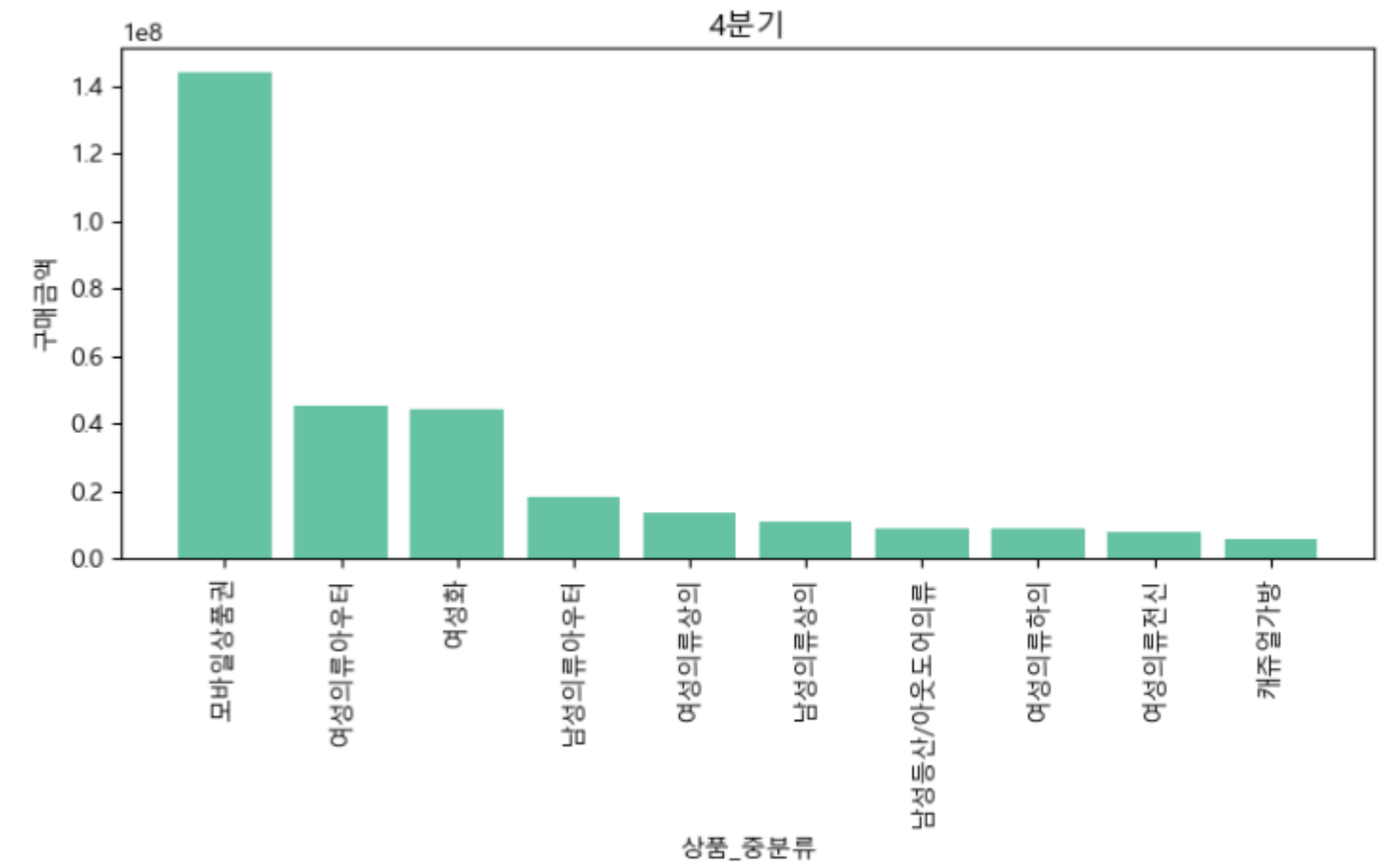
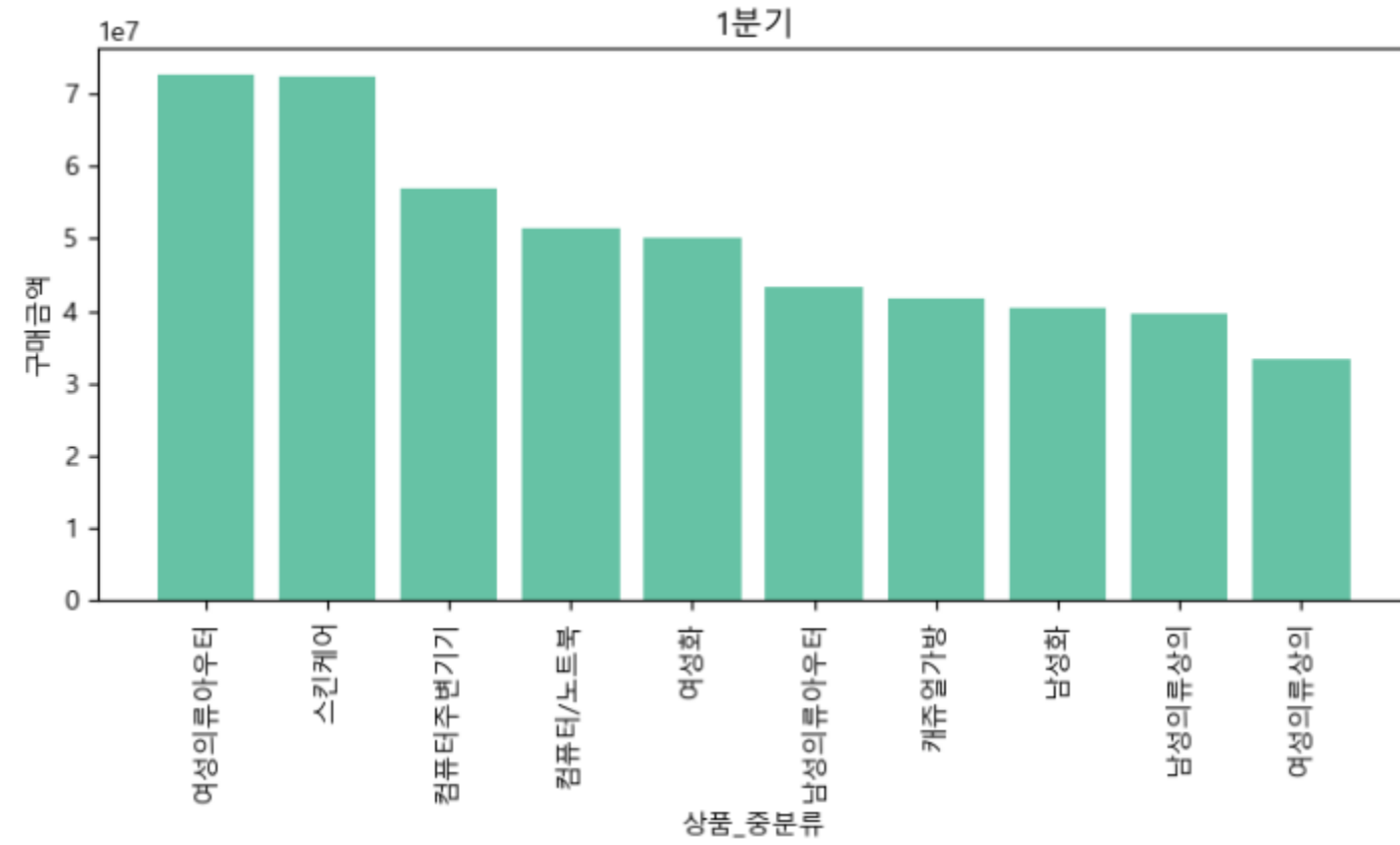


04 결론 및 특이점 도출

분기별 매출액 상위 10개 품목 - A06

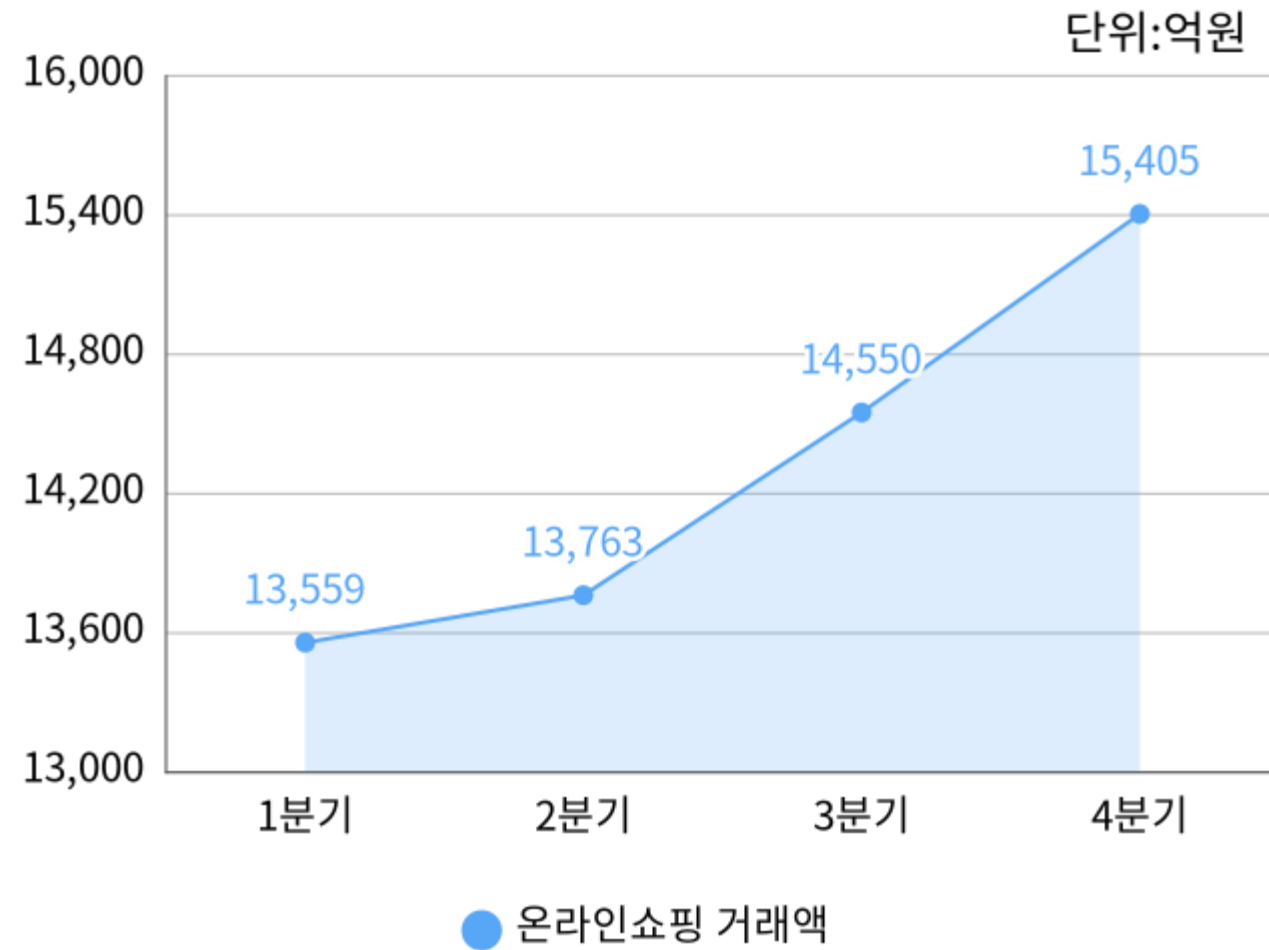


04 결론 및 특이점 도출



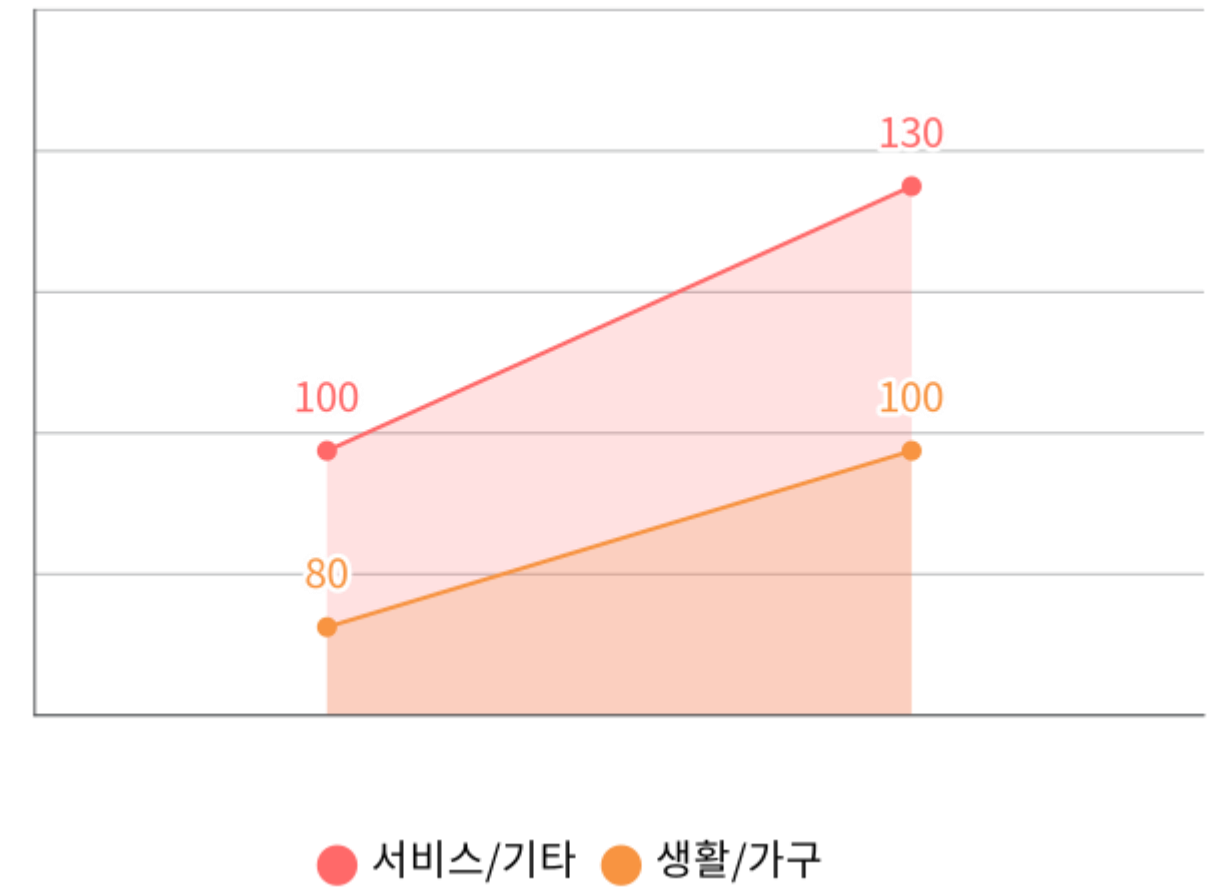
유통사 A6의 총 매출액 중 **모바일상품권**의 비중이 크게 증가

04 결론 및 특이점 도출



온라인쇼핑 거래액 상승

2021 분기별 온라인쇼핑 거래액
(취급상품범위별/상품군별)



실내 생활문화 확산

공연·음식배달·e쿠팤관련 상품 등의 지속적인 판매 호조
> 서비스/기타 매출 30.3% 증가
실내 생활문화 확산에 따른 주방용품·침실가구구매 증가
-> 생활/가구(26.9%)의 매출이 크게 증가

04 결론 도출

추론 결과

- A01 백화점
- A02 대형마트
- A03 SSM
- A04 편의점
- A05 대형마트
- A06 온라인 매장

감사합니다

'채움'