
백화점 BI(Business Intelligence) 분석 보고서

(백화점 비정제 데이터 분석)

2021.02.08

남윤호, 박수진, 박창서, 송은이, 선은빈, 우지원

목차

| | |
|--|----|
| 0. 프로젝트 개요 | 1 |
| 1. 데이터 처리 | 3 |
| 1-1. 데이터 기본탐색 | 3 |
| 1-1-1) 원천 데이터셋 로딩과 기본 탐색 | 3 |
| 1-1-2) 원천 데이터셋 조인 | 4 |
| 1-2. 데이터 전처리 | 4 |
| 1-2-1) '구매액', '할인액' 칼럼을 가공 '실구매액(sales)' 칼럼 생성 | 5 |
| 1-2-2) 데이터에 있는 공백을 제거 | 5 |
| 1-2-3) 동일 데이터 이름 통일 | 5 |
| 1-2-4) 더 정확한 파트 분류 | 6 |
| 2. BI 분석 내용 | 7 |
| 2-1. 매출 기준 시각화 | 7 |
| 2-1-1) 전체 매출액 내역 | 7 |
| 2-1-2) 파트, PC, 코너, 브랜드 별 매출액 TOP 10 | 8 |
| 2-1-2) 파트별 Top10 브랜드 | 8 |
| 2-2. 지점 기준 시각화 | 10 |
| 2-2-1) 각 지점별 순매출액 비교 | 10 |

| | |
|--------------------------------|----|
| 2-2-2) 각 지점의 파트별 매출액 비교 | 11 |
| 2-2-3) 각 지점별 PC 매출액 비교 | 12 |
| 2-3. 날짜, 날씨 기준 시각화 | 14 |
| 2-3-1) 날짜 데이터 처리 | 14 |
| 2-3-2) 월별 매출액 분석 | 15 |
| 2-3-3) 시간대별 매출액 분석..... | 16 |
| 2-3-4) 요일별 구매 파트 비율 분석 | 16 |
| 2-3-5) 영업 시간 외 매출 내역 분석 | 17 |
| 2-3-6) 날씨 기준 시각화 | 18 |
| 2-4. 성별 기준 시각화..... | 20 |
| 2-4-1) 성별 백화점 구매 고객 수 비교 | 20 |
| 2-4-2) 성별 매출액 비교 | 21 |
| 2-5-3) 성별 파트별 구매액 비교 | 21 |
| 2-5. 고객등급 기준 시각화 | 22 |
| 2-5-1) 등급기준표..... | 22 |
| 2-5-2) 등급별 비중(고객수, 매출액) | 23 |
| 2-5-3) 등급별 관심품목 | 24 |

| | |
|-----------------------------------|----|
| 2-6. 수입품/ 국내품 기준 시각화..... | 25 |
| 2-6-1) 매출의 수입품과 국내품 비중 | 26 |
| 2-6-2) 지점별 수입품과 국내품 매출 비교..... | 26 |
| 2-6-3) 파트별 수입품과 국내품 매출 비교..... | 27 |
| 2-7. 할인율 기준 시각화..... | 27 |
| 2-7-1) 지점 평균 할인율 | 28 |
| 2-7-2) 파트, PC, 브랜드 별 평균 할인율 | 28 |
| 2-7-3) 월별 평균 할인율 시각화..... | 31 |
| 2-8. 환불 기준 시각화..... | 31 |
| 2-8-1) 지점별 환불액 | 32 |
| 2-8-2) 지점별 상위 환불 비율 브랜드 | 32 |
| 2-8-3) 지점별 상위 환불 횟수 브랜드 | 33 |
| 2-8-4) 상위 환불 파트, PC..... | 34 |

3. BI 분석 결과 요약 및 분석 37

1. 분석 목적

■ 백화점 데이터를 시각화하여 차후 경영 현황 파악

- 데이터를 통해 백화점의 현황을 객관적으로 파악가능
- 데이터 시각화를 이용하여 백화점의 현황 파악 및 경영전략에 보다 쉽게 접근가능

■ 데이터 시각화를 통하여 비즈니스 문제 및 이슈 해결에 도움

- 분석 결과를 바탕으로 해결 가능한 비즈니스 문제 및 이슈 예시 데이터를 시각화를 통하여 경영의 다양한 분야에 활용 및 참고 가능
- 상품 및 서비스 기획 시 참고 가능
- 전략적인 마케팅 계획 수립 가능
- Part 별 사업 예산 책정 시 참고 가능
- 비효율적인 매장 및 파트 파악과 개선점 도출가능
- 일/월/시즌 별 할인 프로모션 시 참고가능

2. 자료 내용

■ 서울 소재 백화점 회원들의 제품 구입 내역 및 성별

3. 자료 분석 범위 및 대상

■ 분석 기간 : 2000 년 05 월 ~ 2001 년 4 월

■ 분석 가맹점 : 본점, 무역점, 천호점, 신촌점

■ 분석 대상 : 분석 가맹점 회원들의 모든 제품 구입 내역 (총 1,036,653 건)

■ 분석 내용 :

- 백화점의 BI 분석을 위한 각 세분화된 데이터 시각화
- 경영사항에 기본이 되는 매출 기준 시각화에서 시작
- 백화점 운영에 영향을 미칠 수 있는 외적 요인인 날씨별 시각화, 기후별 시각화
진행
- 고객과 입점 브랜드, 품목 등의 현황을 잘 파악 할 수 있는 성별, 고객 등급,
수입품/국내품, 할인율, 환불 시각화 진행
- 매출액 기준 시각화
- 날씨, 날씨 기준 시각화
- 성별 기준 시각화
- 고객등급 기준 시각화
- 수입품/국내품 기준 시각화
- 할인율 기준 시각화
- 환불 기준 시각화

1

데이터 처리

1-1. 데이터 기본 탐색

1-1-1) 원천 데이터셋 로딩과 기본 탐색

■ 기본 데이터 X_train, Y_train 의 데이터 설명

| 제공 데이터 설명 (X_train.csv) | | |
|------------------------------|-------------|-------------------------|
| 일련번호 | 필드명 | 필드 내용 |
| 1 | custid | 고객ID |
| 2 | date_time | 판매일시 (yyyy-mm-dd hh:mm) |
| 3 | store | 지점명 |
| 4 | product | 상품코드 |
| 5 | brand | 브랜드명 |
| 6 | corner | 코너명 |
| 7 | pc | PC명 |
| 8 | part | 파트명 |
| 9 | imported | 수입상품여부(1: 수입상품) |
| 10 | amount | 구매액 |
| 11 | discount | 할인액 |
| 12 | installment | 할부기간 |

| (Y_train.csv) | | |
|---------------|--------|-----------------|
| 일련번호 | 필드명 | 필드 내용 |
| 1 | custid | 고객ID |
| 2 | gender | 성별 (0:여자, 1:남자) |

- X_train 은 거래 관련 데이터, Y_train 은 성별 데이터

■ X_train 로딩 및 기본 탐색

```
x_train = pd.read_csv('X_train.csv', encoding='cp949')
x_train.head()
```

| | custid | date_time | store | product | brand | corner | pc | part | imported | amount | discount | installment |
|---|--------|------------------|-------|---------------|-------|---------|------|------|----------|--------|----------|-------------|
| 0 | 0 | 2000-06-25 12:12 | 무역점 | 2116050008000 | 에스티로더 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 90000 | 9000 | 3 |
| 1 | 0 | 2000-06-25 12:42 | 무역점 | 4125440008000 | 시슬리 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 39000 | 3900 | 1 |
| 2 | 0 | 2000-08-26 18:10 | 본점 | 2116052008000 | 크리닉 | 수입종합화장품 | 화장품 | 잡화파트 | 1 | 175000 | 17500 | 3 |
| 3 | 0 | 2000-08-26 18:30 | 본점 | 4106430119900 | 듀폰 | 수입의류 | 명품토털 | 잡화파트 | 1 | 455000 | 45500 | 3 |
| 4 | 0 | 2000-09-03 18:02 | 무역점 | 2139141008000 | 랑콤 | 수입종합화장품 | 화장품 | 명품잡화 | 0 | 100000 | 10000 | 3 |

■ Y_train 로딩 및 기본 탐색

```
y_train = pd.read_csv('Y_train.csv', encoding='cp949')
y_train.head()
```

| | custid | gender |
|---|--------|--------|
| 0 | 0 | 1 |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 3 | 0 |
| 4 | 4 | 0 |

1-1-2) 원천 데이터셋 조인

■ X_train 과 Y_train 데이터 조인

■ 공통된 필드명인 custid 를 기준으로 merge 하여 새로운 데이터 셋 'train' 생성

```
train = pd.merge(x_train, y_train, on='custid', how='left')
train.head()
```

| | custid | date_time | store | product | brand | corner | pc | part | imported | amount | discount | installment | gender |
|---|--------|------------------|-------|---------------|-------|---------|------|------|----------|--------|----------|-------------|--------|
| 0 | 0 | 2000-06-25 12:12 | 무역점 | 2116050008000 | 에스티로더 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 90000 | 9000 | 3 | 1 |
| 1 | 0 | 2000-06-25 12:42 | 무역점 | 4125440008000 | 시슬리 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 39000 | 3900 | 1 | 1 |
| 2 | 0 | 2000-08-26 18:10 | 본점 | 2116052008000 | 크리닉 | 수입종합화장품 | 화장품 | 잡화파트 | 1 | 175000 | 17500 | 3 | 1 |
| 3 | 0 | 2000-08-26 18:30 | 본점 | 4106430119900 | 듀폰 | 수입의류 | 명품토탈 | 잡화파트 | 1 | 455000 | 45500 | 3 | 1 |
| 4 | 0 | 2000-09-03 18:02 | 무역점 | 2139141008000 | 랑콤 | 수입종합화장품 | 화장품 | 명품잡화 | 0 | 100000 | 10000 | 3 | 1 |

1-2. 데이터 전처리

1-2-1) '구매액', '할인액' 칼럼을 가공 '실구매액(sales)' 칼럼 생성

```
train["sales"] = train["amount"] - train["discount"]
train.head()
```

| | custid | date_time | store | product | brand | corner | pc | part | imported | amount | discount | installment | gender | sales |
|---|--------|------------------|-------|---------------|-------|---------|------|------|----------|--------|----------|-------------|--------|--------|
| 0 | 0 | 2000-06-25 12:12 | 무역점 | 2116050008000 | 에스티로더 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 90000 | 9000 | 3 | 1 | 81000 |
| 1 | 0 | 2000-06-25 12:42 | 무역점 | 4125440008000 | 시슬리 | 수입종합화장품 | 화장품 | 명품잡화 | 1 | 39000 | 3900 | 1 | 1 | 35100 |
| 2 | 0 | 2000-08-26 18:10 | 본점 | 2116052008000 | 크리닉 | 수입종합화장품 | 화장품 | 잡화파트 | 1 | 175000 | 17500 | 3 | 1 | 157500 |
| 3 | 0 | 2000-08-26 18:30 | 본점 | 4106430119900 | 듀폰 | 수입의류 | 명품토탈 | 잡화파트 | 1 | 455000 | 45500 | 3 | 1 | 409500 |
| 4 | 0 | 2000-09-03 18:02 | 무역점 | 2139141008000 | 랑콤 | 수입종합화장품 | 화장품 | 명품잡화 | 0 | 100000 | 10000 | 3 | 1 | 90000 |

1-2-2) 데이터에 있는 공백을 제거

- ex) 'brand' 칼럼의 '밀라노 신촌'을 '밀라노신촌'으로 변경
- 아래 그림을 참고하면 공백 제거 전에는 분명 데이터가 있음에도 불구하고 추출 되지 않아 공백 제거 후 포함된 데이터가 추출됨

```
train[train["brand"]=="밀라노 신촌"]
```

| custid | date_time | store | product | brand | corner | pc | part | imported | amount | discount | installment | gender | sales |
|--------|-----------|-------|---------|-------|--------|----|------|----------|--------|----------|-------------|--------|-------|
|--------|-----------|-------|---------|-------|--------|----|------|----------|--------|----------|-------------|--------|-------|

```
train["brand"] = train["brand"].apply(lambda x: x.replace(' ', ''))
train["pc"] = train["pc"].apply(lambda x: x.replace(' ', ''))
train["corner"] = train["corner"].apply(lambda x: x.replace(' ', ''))
```

```
train[train["brand"]=="밀라노신촌"].head(2)
```

| | custid | date_time | store | product | brand | corner | pc | part | imported | amount | discount | installment | gender | sales |
|------|--------|------------------|-------|---------------|-------|---------|--------|------|----------|--------|----------|-------------|--------|--------|
| 3868 | 103 | 2001-01-05 11:30 | 신촌점 | 4800429713020 | 밀라노신촌 | 뉴베이직캐주얼 | 어덜트캐주얼 | 영캐릭터 | 0 | 46400 | 0 | 1 | 0 | 46400 |
| 3871 | 103 | 2001-01-06 17:10 | 신촌점 | 4800429713020 | 밀라노신촌 | 뉴베이직캐주얼 | 어덜트캐주얼 | 영캐릭터 | 0 | -46400 | 0 | 1 | 0 | -46400 |

1-2-3) 동일 데이터 이름 통일

- '로얄부틱'과 '로얄부티크' 와 같이 같은 파트를 의미하지만 다른 이름으로 저장된 소들을 제거
- 그 결과 파트 명의 개수가 '31'개에서 '12'개로 줄어듬

```
print("파트 명 개수 :", len(train["part"].unique()))
print("\n파트 명 :", train["part"].unique())
```

파트 명 개수 : 31

파트 명 : ['명품잡화' '잡화파트' '남성의류' '골프/유니캐주얼' '로얄부띠끄' '캐주얼,구두,아동' '여성캐주얼' '여성의류파트' '가정용품' '가정용품파트' '남성정장스프츠' '아동' '잡화' '영어덜트캐주얼' '영라이브' '공산품' '스포츠캐주얼' '여성정장' '아동문화' '공산품파트' '생식품파트' '영플라자' '생식품' '패션잡화' '아동,스포츠' '로얄부틱' '여성캐주얼' '스포츠캐주얼' '영캐릭터' '상품개발영업1과' '인터넷백화점']

전처리

```
train["part"] = train["part"].apply(lambda x: "생식품" if x == "생식품파트"
                                     else "가정용품" if x=="가정용품파트"
                                     else "여성의류" if x in ["여성정장", "여성의류파트", "여성캐주얼", "여성캐주얼"]
                                     else "명품" if x in ["로얄부틱", "로얄부띠끄", "명품잡화"]
                                     else "공산품" if x=="공산품파트"
                                     else "잡화" if x in ["잡화파트", "패션잡화"]
                                     else "스포츠캐주얼" if x in ["골프/유니캐주얼", "스포츠캐주얼"]
                                     else "아동" if x in ["아동문화", "아동,스포츠"]
                                     else "남성의류" if x in ["남성정장스프츠", "인터넷백화점"]
                                     else "영플라자" if x in ["영어덜트캐주얼", "영라이브", "영캐릭터"]
                                     else "기타" if x in ["상품개발영업1과"]
                                     else x)
```

```
print("파트 명 개수 :", len(train["part"].unique()))
print("\n파트 명 :", train["part"].unique())
```

파트 명 개수 : 12

파트 명 : ['명품' '잡화' '남성의류' '스포츠캐주얼' '캐주얼,구두,아동' '여성의류' '가정용품' '아동' '영플라자' '공산품' '생식품' '기타']

1-2-4) 정확한 파트 분류

■ 포괄적인 파트를 세분화하여 더 정확한 파트 분류를 함

■ Ex) "케주얼,구두 아동" 파트의 "여성캐주얼" PC 를 "여성의류" 파트 아래의 PC 로 옮김

■ 그 결과 파트명의 개수가 12 개에서 11 개로 변경

```
print("케주얼,구두,아동 파트의 pc :",train[train['part']=='케주얼,구두,아동']['pc'].unique())
```

케주얼,구두,아동 파트의 pc : ['아동복' '내의란제리' '영트렌디' '트레디셔널' '진캐주얼' '문화용품' '미확인pc' '남여구두' '캐릭터정장' '니트,단품,모피' '여성캐주얼' '셔츠' '화장품']

```
import warnings
warnings.filterwarnings(action='ignore')

train.part[(train['pc'] == '내의란제리') & (train['part']=='케주얼,구두,아동')] = "여성의류"
train.part[(train['pc'] == "문화용품") & (train['part']=='케주얼,구두,아동')] = "공산품"
train.part[(train['pc'] == "아동복") & (train['part']=='케주얼,구두,아동')] = "아동"
train.part[(train['pc'] == "영트렌디") & (train['part']=='케주얼,구두,아동')] = "아동"
train.part[(train['pc'] == "트레디셔널") & (train['part']=='케주얼,구두,아동')] = "스포츠캐주얼"
train.part[(train['pc'] == "진캐주얼") & (train['part']=='케주얼,구두,아동')] = "스포츠캐주얼"
train.part[(train['pc'] == "캐릭터정장") & (train['part']=='케주얼,구두,아동')] = "스포츠캐주얼"
train.part[(train['pc'] == "미확인pc") & (train['part']=='케주얼,구두,아동')] = "잡화"
train.part[(train['pc'] == "남여구두") & (train['part']=='케주얼,구두,아동')] = "잡화"
train.part[(train['pc'] == "여성캐주얼") & (train['part']=='케주얼,구두,아동')] = "여성의류"
train.part[(train['pc'] == "셔츠") & (train['part']=='케주얼,구두,아동')] = "스포츠캐주얼"
train.part[(train['pc'] == "화장품") & (train['part']=='케주얼,구두,아동')] = "잡화"
train.part[(train['pc'] == "니트,단품,모피") & (train['part']=='케주얼,구두,아동')] = "잡화"
```

```
print("파트 명 개수 :", len(train["part"].unique()))
print("파트 명 :", train["part"].unique())
```

파트 명 개수 : 11

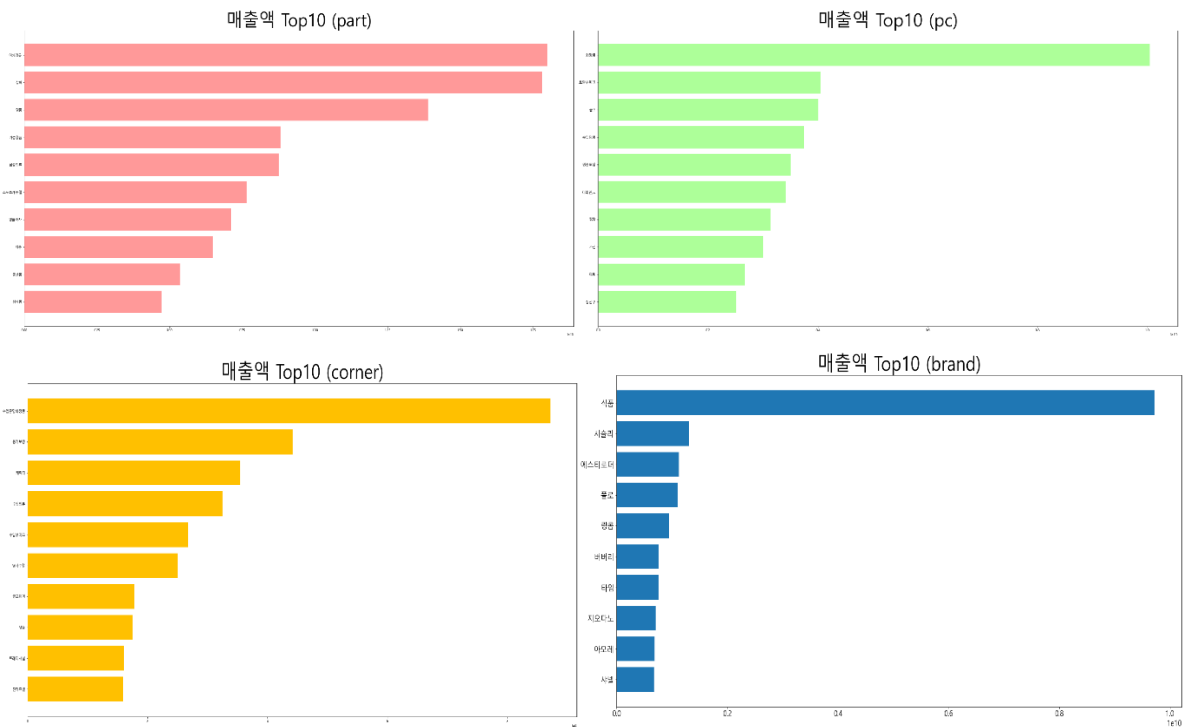
파트 명 : ['명품' '잡화' '남성의류' '스포츠캐주얼' '아동' '여성의류' '가정용품' '영플라자' '공산품' '생식품' '기타']

2-1. 매출액 기준 시각화

2-1-1) 전체 매출액 내역

■ 총매출 : 98,605, 451,012 원

2-1-2) 파트, PC, 코너, 브랜드별 매출액 TOP 10



■ 파트별 매출액 비중에선 여성의류와 잡화가 큰 비중을 차지. 생식품과 공산품의 대부분이 식품이기 때문에 식품의 비중도 매우 큼

■ PC별 매출액 비중에선 화장품이 다른 PC에 비해 월등히 높음

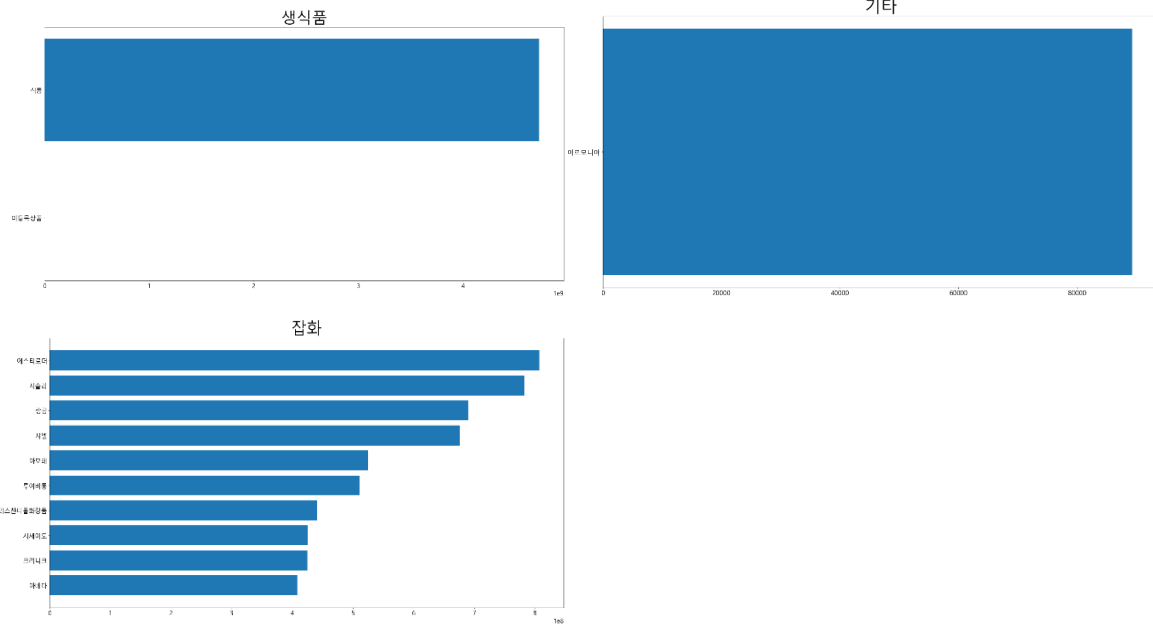
■ 코너별 매출액 또한 수입종합화장품 비율이 높아 PC와 비슷한 결과를 보임

■ 브랜드별 매출액에선 식품이 압도적으로 높은 비중을 보이는데 이것은 식품의 세부 브랜드를 분할하지 않고 '식품' 단일 브랜드로 통합했기 때문. 그 뒤로는 시슬리, 에스

티로더 등 화장품 브랜드의 매출이 높았음

2-1-3) Part 별 Top10 브랜드





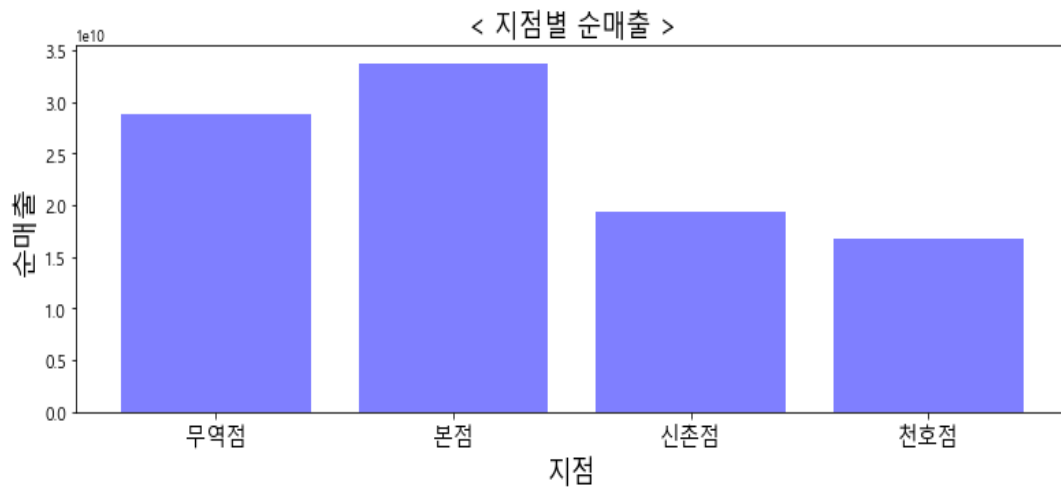
- Part별 상위 Top10 브랜드 그래프를 결과, 전 Part에서 1위 브랜드의 매출 성과가 도드라지는 특징이 나타남
- 특히 공산품의 경우 라면, 과자 등 식품이 대부분을 차지

2-2. 지점별 순매출 시각화

■ 백화점의 각 지점별 순매출액의 내역을 시각화하여 지점별 운영상태 파악

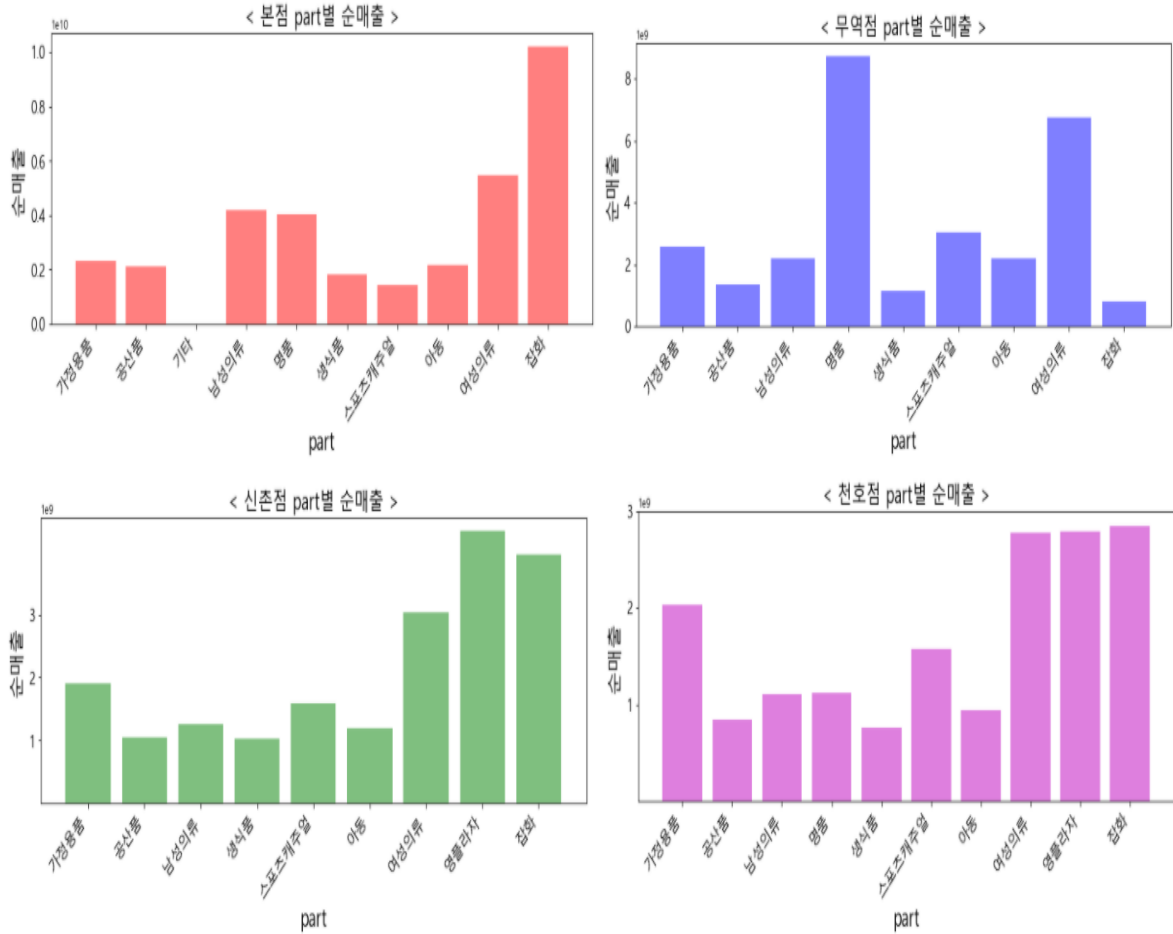
- 지점별 순매출액을 시각화하여 비교
- 각 지점의 Part 별, PC 별 매출 내역을 시각화하여 파악

2-2-1) 각 지점별 순매출액 비교



■ 각 지점별 순매출액은 본점(33,748,842,949 원), 무역점(28,734,742,934 원)
신촌점(19,335,308,652 원), 천호점(16,786,556,477 원) 순으로 순매출액이 높음

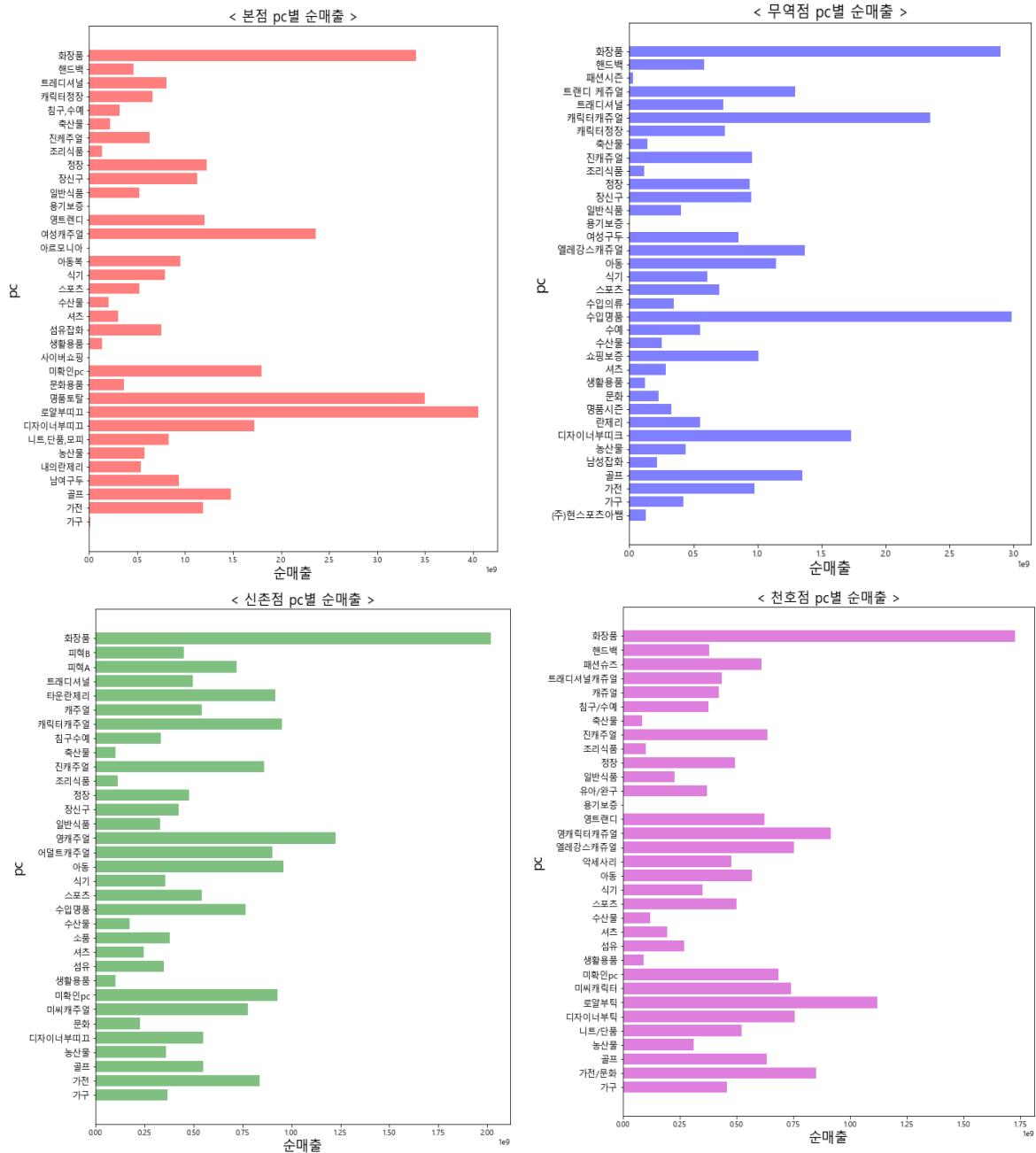
2-2-2) 각 지점의 Part 별 매출액 비교



- 본점은 잡화 Part 가 가장 높은 순매출액을 보였고, 상품개발영업 1 과, 인터넷백화점이 10 만원이하의 적은 순매출액을 가짐
- 무역점은 명품 파트가 압도적인 순매출액을 보였고, 그 외 Part 는 순매출액 비중이 상대적으로 고르게 나타나고, 명품 Part 의 압도적인 순매출액은 무역점의 특성이 반영된 것으로 보임
- 신촌점은 잡화 Part 가 가장 높은 순매출액을 발생시키고 있으며, 영플라자, 가정용품이 그 다음 순매출액을 발생시킴

- 천호점은 잡화 part 가 가장 높은 순매출액을 발생시키고 있으며, 가정용품 순으로 나타남 상대적으로 순매출액이 비슷한 신촌점보다 가정용품 Part 의 매출액 비중이 높은 것을 파악할 수 있음

2-2-3) 각 지점의 PC 매출액 비교



- 본점은 로얄부띠끄, 명품토탈, 화장품 PC 가 압도적으로 높은 매출을 발생
- 무역점은 수입명품과 화장품이 가장 높은 매출을 발생시키며 매장의 특성에 기인한 것으로 보임
- 신촌점은 화장품 PC 가 가장 높은 매출을 발생시키며 다른 지점에 없는 '피혁 A', '피혁 B' 항목이 존재하고 다른 부분의 매출은 상대적으로 미약한 편
- 천호점은 화장품 PC 가 압도적으로 높은 매출액을 발생시키며 신촌점과 마찬가지로 다른 부분의 매출은 상대적으로 미약한 편

2-3. 날짜, 날씨 기준 시각화

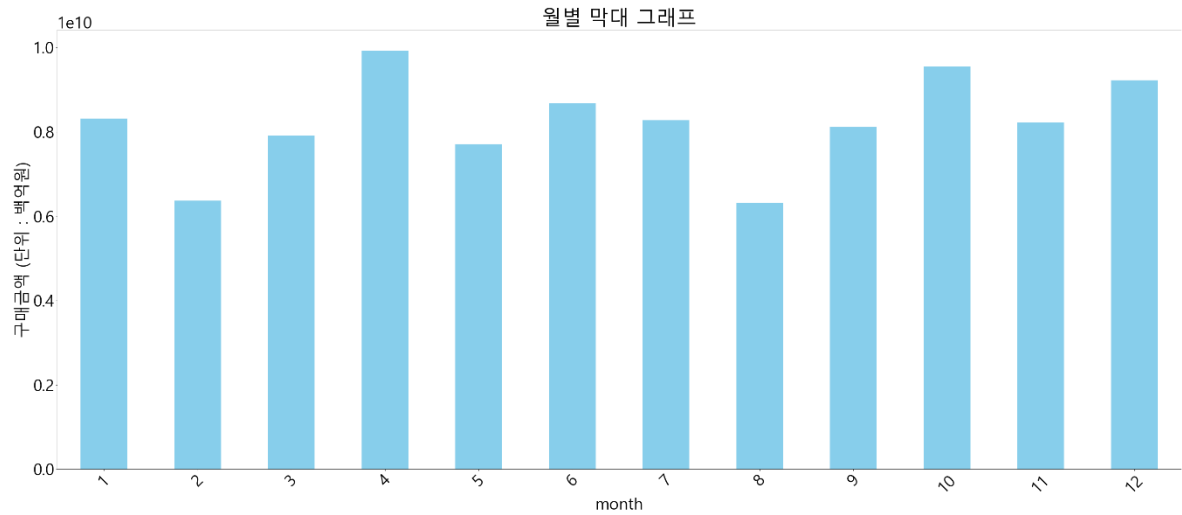
2-3-1) 날짜 데이터 처리

- 날짜 데이터가 'str'로 저장되어있기 때문에 매번 'datetime' 형식으로 변환 필요
- 시간(time), 시(hour), 날짜(date), 월(month), 시간대(timeline) 칼럼을 추가로 생성
- 아래와 같이 7 개의 칼럼을 추가 생성

| day | dayweek | timeline | hour | month | time | date |
|-----|---------|----------|------|-------|----------|------------|
| 6 | 일 | 오후 | 12 | 6 | 12:12:00 | 2000-06-25 |
| 6 | 일 | 오후 | 12 | 6 | 12:42:00 | 2000-06-25 |
| 5 | 토 | 6시이후 | 18 | 8 | 18:10:00 | 2000-08-26 |
| 5 | 토 | 6시이후 | 18 | 8 | 18:30:00 | 2000-08-26 |
| 6 | 일 | 6시이후 | 18 | 9 | 18:02:00 | 2000-09-03 |

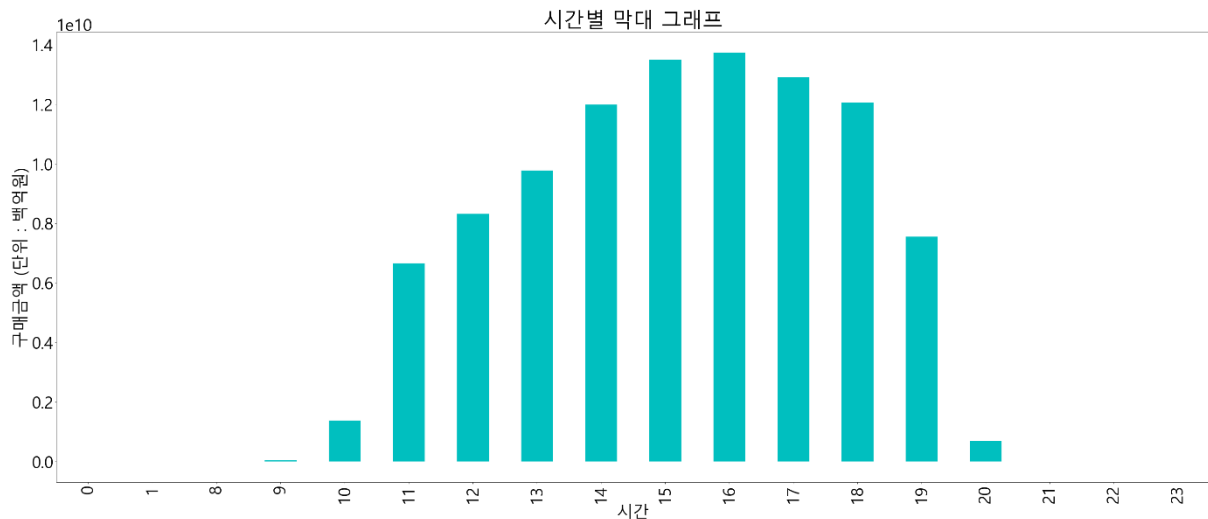
2-3-2) 월별 매출액 분석

- 일별 매출액은 일별로 편차가 매우 커 그래프 분석에 용이하지 않아 월별로 그룹화 후 분석

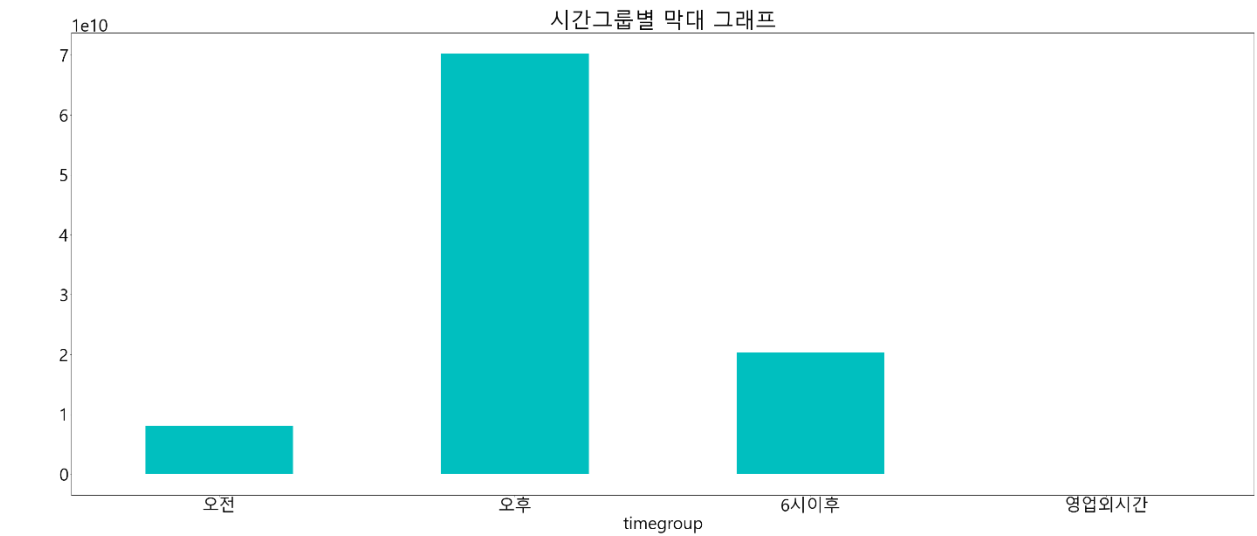


- 매출액은 2 월 6,362,640,770 원, 8 월 6,316,979,826 원으로 가장 낮고, 4 월 9,927,584,799 원, 10 월 9,551,179,558 원, 12 월 9,221,803,489 원 순으로 높음
- 매출액이 가장 높은 4 월과 가장 낮은 2 월의 금액 차이가 약 36 억원

2-3-3) 시간대별 매출액 분석

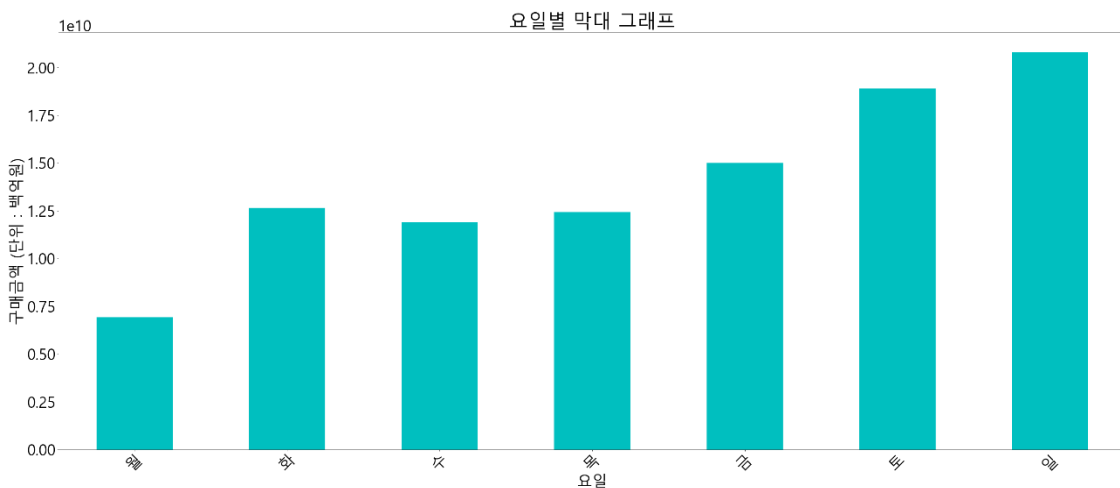


- 개장 후 매출액은 점점 높아지며 16 시에 정점을 찍은 후 점차적으로 감소함



- '오전'은 9 시-12 시 (3 시간), '오후'는 12 시-18 시 (4 시간), '6 시이후'는 18 시- 22 시 (4 시간), 그 외 시간은 '영업 외 시간'
- 같은 4 시간임에도 '오후' 시간대가 '6 시 이후' 시간대보다 3 배 가량의 매출액 발생
- 대부분의 매출이 '오후' 시간대에 발생

2-3-4) 요일별 구매 파트 비율 분석



- 주말(토, 일) 매출액이 가장 높으며, 월요일이 가장 매출액이 낮고, 그 외의 평일의 매출액은 거의 균일함

2-3-5) 영업 시간 외 매출 내역 분석

- 시간대별 매출액 분석 중 새벽시간대 등 영업 외 시간에 거래된 내역 확인

관련된 거래는 총 15 건으로 전체 약 170 만건에 데이터에 비해 매우 적은 비중

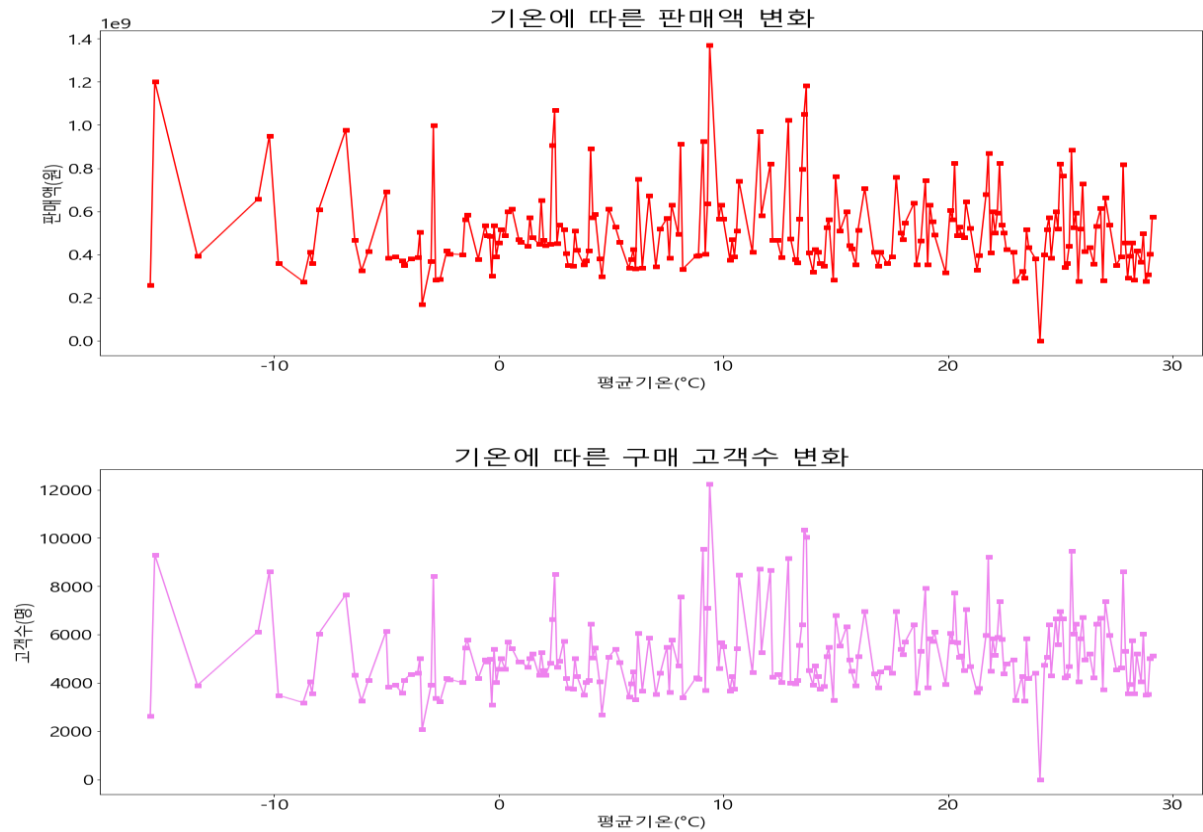
- 15 건의 반 이상이 환불 거래인 것으로 보아 매장 마감 정산 등의 오류를 수정한 것으로 추정

| train[train['timeline']=='영업외시간'] | | | | | | | | | | | | | | | | | | | |
|-----------------------------------|---------------------|-------|---------------|---------|--------|--------------|--------|----------|----------|-----|--------|----------|-----|---------|----------|------|---|--|--|
| custid | date_time | store | product | brand | corner | pc | part | imported | amount | ... | gender | sales | day | dayweek | timeline | hour | r | | |
| 193906 | 2001-04-07 22:11:00 | 본점 | 4404830025000 | 레노마플 | 애플어 | 애플 | 남성외복 | 0 | -236000 | ... | 1 | -212400 | 5 | 토 | 영업외시간 | 22 | | | |
| 232702 | 2000-07-23 22:00:00 | 본점 | 6300032019930 | 빈폴 | 트래디셔널 | 트래디셔널 | 스포츠캐주얼 | 0 | 52500 | ... | 1 | 52500 | 6 | 일 | 영업외시간 | 22 | | | |
| 302562 | 2000-05-05 01:18:00 | 본점 | 4116900003200 | 샤보 | 캐주얼 | 캐주얼 | 남성캐주얼 | 0 | -158000 | ... | 0 | -150100 | 4 | 금 | 영업외시간 | 1 | | | |
| 371984 | 2000-05-03 01:15:00 | 무역점 | 4502161930270 | 삼성 | 가전 | 가전 | 가전 | 0 | -3290000 | ... | 0 | -3290000 | 2 | 수 | 영업외시간 | 1 | | | |
| 608568 | 2000-08-31 23:46:00 | 무역점 | 6421880029901 | 에일바 | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 36000 | ... | 0 | 36000 | 3 | 목 | 영업외시간 | 23 | | | |
| 608569 | 2000-08-31 23:47:00 | 무역점 | 6240990026001 | 잠지스포르츠 | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 29000 | ... | 0 | 29000 | 3 | 목 | 영업외시간 | 23 | | | |
| 614114 | 2000-08-03 00:33:00 | 본점 | 4105203210000 | 맥스앤스 | 수입부 | 로탈부 | 명품 | 1 | -77000 | ... | 0 | -77000 | 5 | 토 | 영업외시간 | 0 | | | |
| 614115 | 2000-06-03 00:33:00 | 본점 | 4300500019900 | 올덴 | 트래디셔널 | 트래디셔널 | 스포츠캐주얼 | 0 | -310000 | ... | 0 | -294500 | 5 | 토 | 영업외시간 | 0 | | | |
| 815516 | 2000-06-02 00:23:00 | 본점 | 4125270008400 | 슈리미아 | 화장 | 화장 | 화장 | 0 | -25000 | ... | 0 | -23750 | 4 | 금 | 영업외시간 | 0 | | | |
| 853474 | 2000-05-05 01:18:00 | 본점 | 4512410943020 | 태극봉예 | 음식 | 음식 | 음식 | 0 | -54400 | ... | 0 | -54400 | 4 | 금 | 영업외시간 | 1 | | | |
| 906935 | 2000-06-17 22:04:00 | 무역점 | 6240990026000 | 잠지스포르츠 | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 43000 | ... | 0 | 40850 | 5 | 토 | 영업외시간 | 22 | | | |
| 1003980 | 2001-01-14 08:40:00 | 본점 | 6300031019930 | 컬렉션/카디날 | 내셔널 | 정장 | 남성외복 | 0 | 795000 | ... | 1 | 795000 | 6 | 일 | 영업외시간 | 8 | | | |
| 1024402 | 2000-07-08 22:00:00 | 무역점 | 6227692026030 | MLB | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 23100 | ... | 0 | 23100 | 5 | 토 | 영업외시간 | 22 | | | |
| 1024403 | 2000-07-08 22:00:00 | 무역점 | 6421871028420 | 리본슈즈 | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 71200 | ... | 0 | 71200 | 5 | 토 | 영업외시간 | 22 | | | |
| 1024404 | 2000-07-08 22:04:00 | 무역점 | 6421861028401 | 아디다스슈즈 | 미확인 | (주)원스포르츠아일랜드 | 아름 | 0 | 119000 | ... | 0 | 113050 | 5 | 토 | 영업외시간 | 22 | | | |

2-3-6) 날씨 기준 시각화

■ 기상청 데이터를 이용한 백화점 매출액, 구매 고객 수 분석

■ 기온에 따른 일 평균 매출액, 구매 고객 수 분석

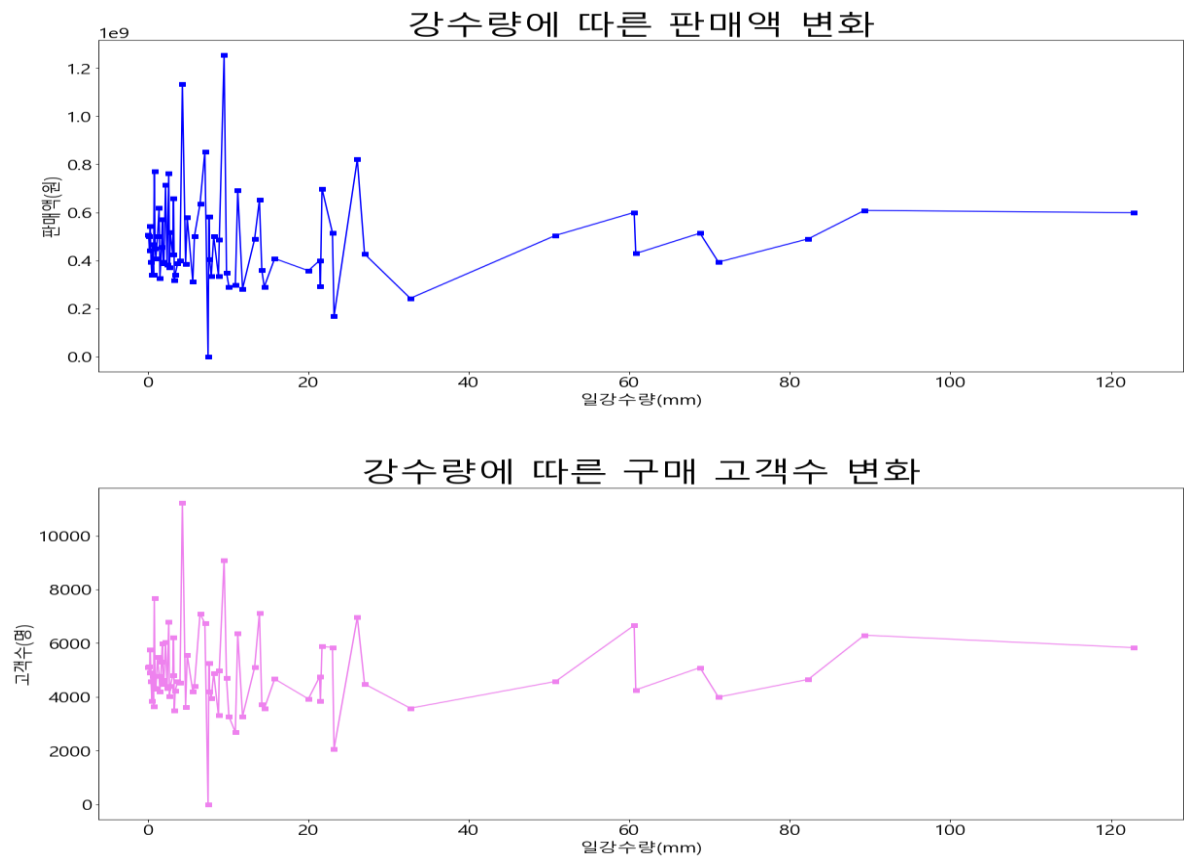


■ 기온에 관계없이 매출액과 고객수가 나타남

■ 기온에 따른 매출액과 구매 고객수를 같이 비교하였을 때 거의 동일한 그래프의 형태를 볼 수 있으며, 구매 고객수가 많을수록 매출액이 높은 것으로 나타남

■ 구매 이력이 있는 고객의 데이터만 존재하기 때문에 백화점 방문 고객수는 정확히 알 수 없음

■ 강수량에 따른 일 평균 매출액, 구매 고객 수 분석

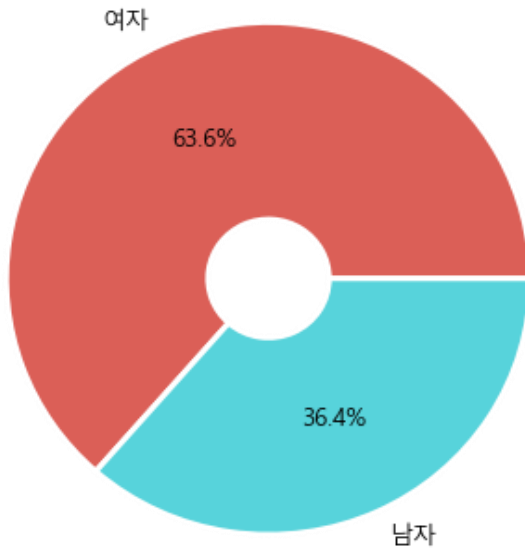


■ 강수량에 관계없이 매출액과 고객수가 나타남

- 강수량에 따른 매출액과 구매 고객수를 같이 비교하였을 때 거의 동일한 그래프의 형태를 볼 수 있으며, 구매 고객수가 많을수록 매출액이 높음

2-4. 성별 기준 시각화

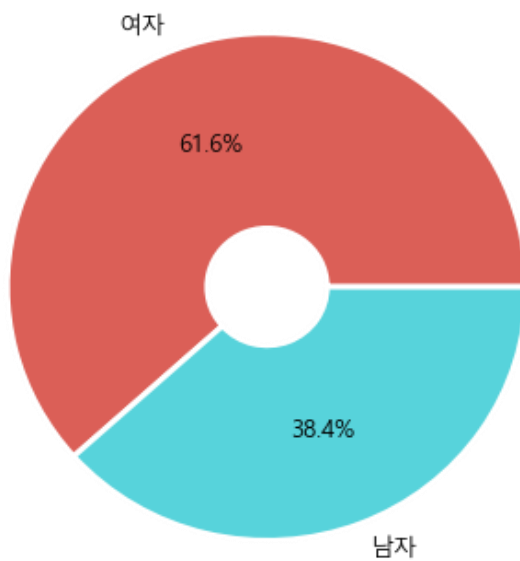
2-4-1) 성별 백화점 구매 고객 수 비교



■ 여성 고객 : 658,863 명, 남성 고객 : 377,790 명, 총 고객 : 1,036,653 명

■ 백화점 총 고객 수의 63%가 여성으로 약 3분의 2 정도로 여성 고객 비중이 높음

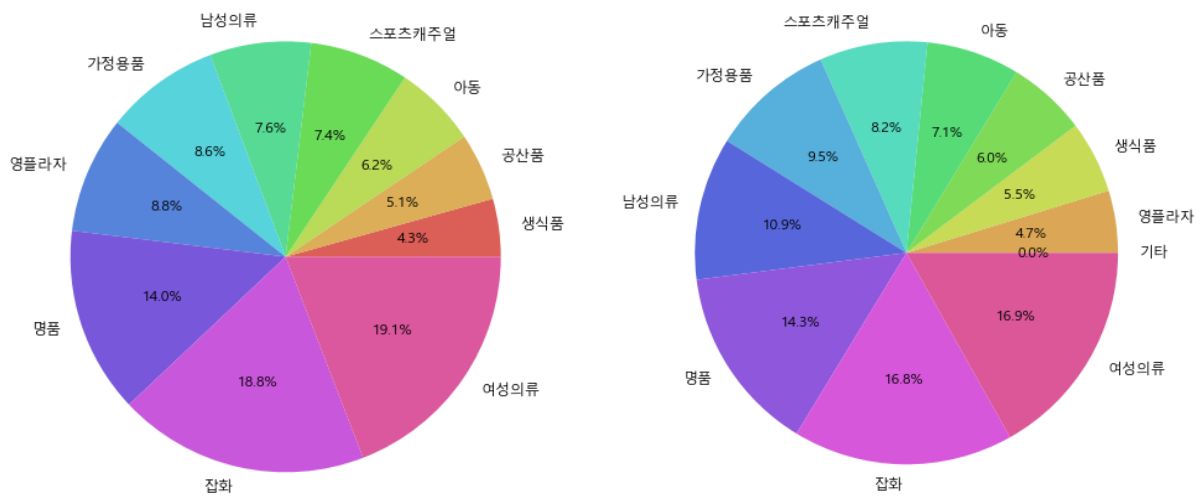
2-4-2) 성별 매출액 비교



■ 여성고객 매출 : 60,785,840,928 원, 남성고객 매출 : 37,819,610,084 원, 전체 판매액 : 98,605,451,012

■ 백화점 총 판매액의 61.6%가 여성

2-4-3) 성별 Part 별 구매액 비교



[좌 : 여성 고객 / 우 : 남성 고객]

- 여성 고객의 경우 여성의류 19.1%로 가장 구매 비중이 높고, 남성은 여성의류 16.9%, 잡화 16.8%로 비등하게 구매 비중이 높은 것으로 나타남
- 영플라자의 경우 여성의 구매비중에서 8.8%, 남성의 구매비중에서 4.7%를 차지하여 약 4%p 차이를 보임
- 분석 전에는 남녀별 구매 비중의 차이가 클 것으로 예상했지만, 분석 결과 성별 구매 비중의 차이는 매우 미세함
- 이는 부부의 경우 카드를 성별 구분 없이 혼합해서 사용하는 경우가 대다수이고, 선물 목적으로 여성이 남성 의류를 구매하는 경우와 남성이 여성 의류를 구매하는 경우가 많기 때문으로 추

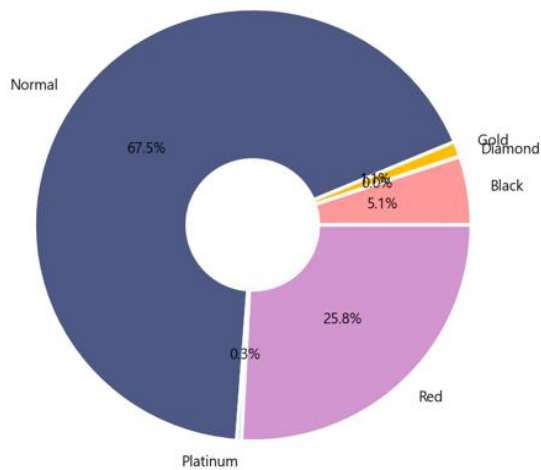
2-5. 고객등급 기준 시각화

2-5-1) 등급 기준표 : 2020 년 등급표에 소비자물가지수를 반영한 추정치 사용

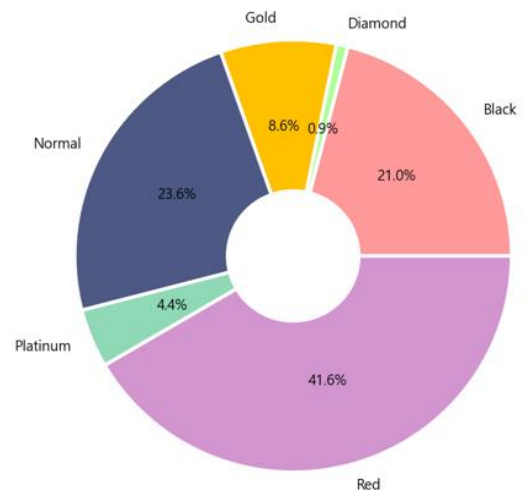
| 등급 | 연간구매액 |
|----------|------------|
| Diamond | 6600 만원 이상 |
| Platinum | 3300 만원 이상 |
| Gold | 2000 만원 이상 |
| Black | 1000 만원 이상 |
| Red | 300 만원 이상 |
| Normal | 300 만원 이하 |

2-5-2) 등급별 비중 (고객수, 매출액)

고객 구매등급별 비중(고객수 기준)



고객 구매등급별 비중(매출액 기준)

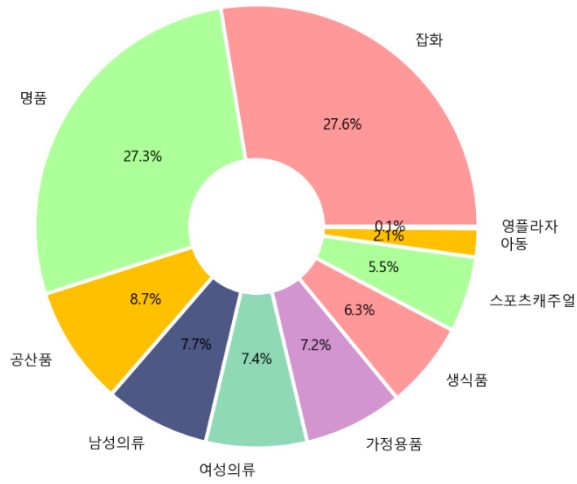


| class | 고객수 | 매출액 (단위:억) | 인당평균매출액(단위:만원) |
|----------|-------|---------------|----------------|
| Diamond | 10 | 8.5 | 8500 |
| Platinum | 103 | 43.4 | 4214 |
| Gold | 338 | 84.5 | 2500 |
| Black | 1539 | 206.7 | 1343 |
| Red | 7748 | 409.8 | 529 |
| Normal | 20262 | 233.2 | 115 |
| SUM | 30000 | 986.1 | 329 |

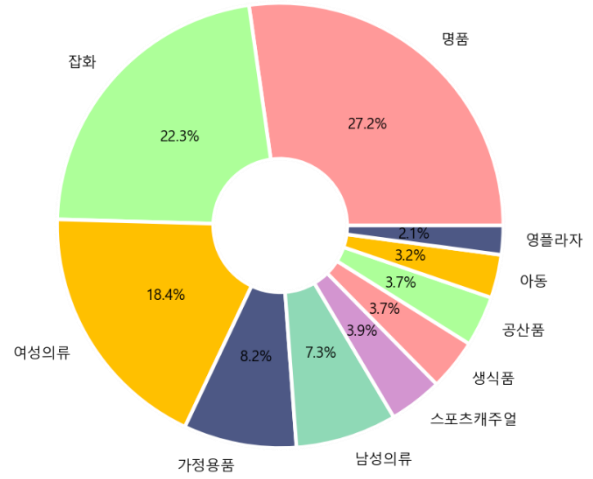
- 등급별 고객수는 상위 등급에서 하위 등급으로 갈수록 많아지며, Normal 및 Red 등급이 전체의 93%로 고객 대부분을 차지함
- Gold, Platinum, Diamond 등급이 고객 수에서 1.5%이지만 백화점 전체 매출액의 14%를 차지하기 때문에 별도의 고객 맞춤형 고급 서비스 전략도 꾸준히 필요할 것으로 예상
- 등급별 매출액은 Red, Normal, Black 순으로 많고, 이들 3개 등급이 전체 매출액의 86%를 차지함에 따라 해당 등급의 고객을 타겟으로 한 마케팅 전략이 주효할 것으로 예상

2-5-3) 등급별 관심품목

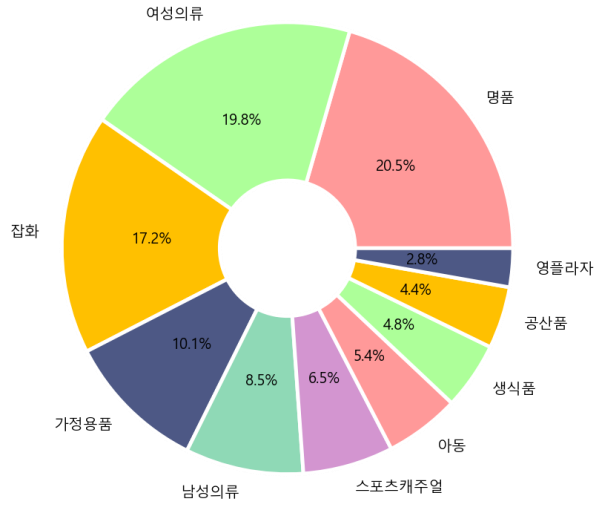
< Diamond >



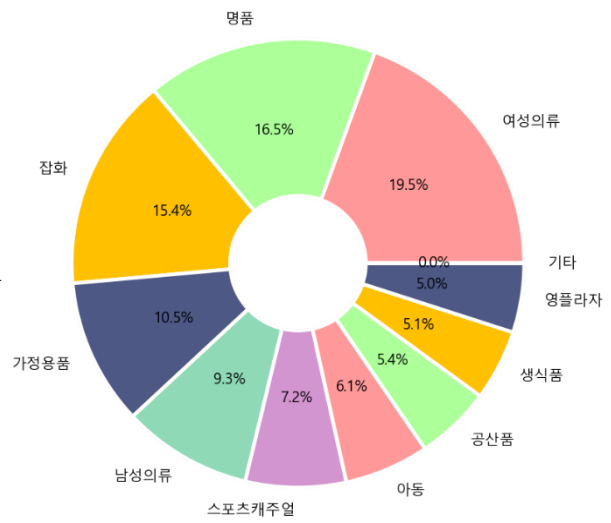
< Platinum >

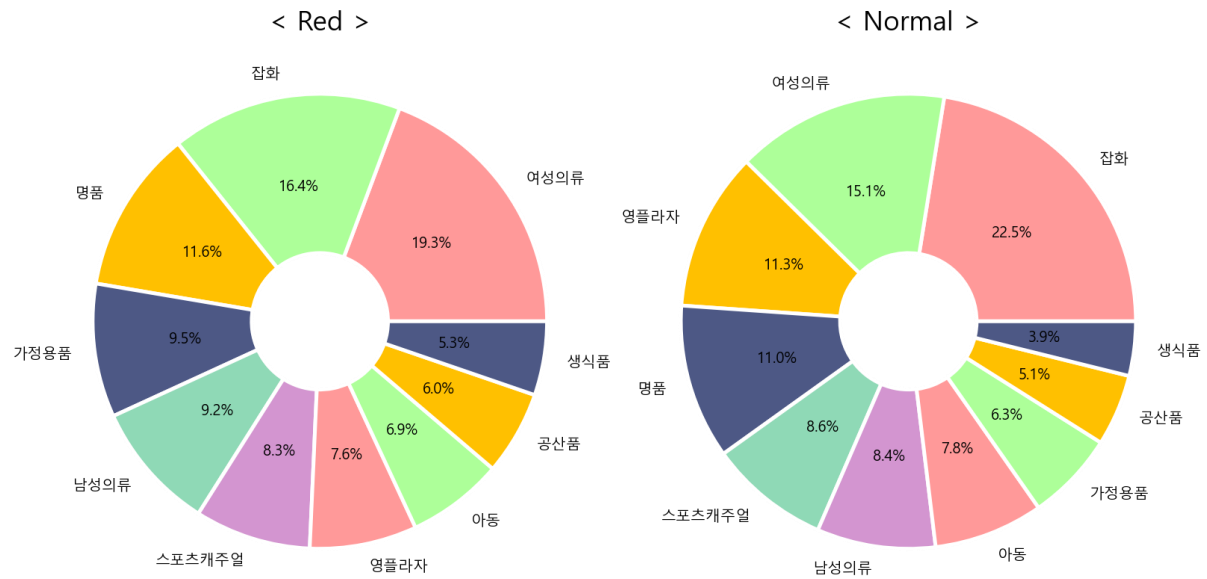


< Gold >



< Black >



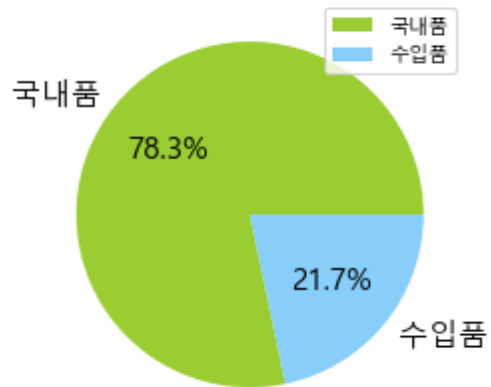


- 전 등급에서 명품, 잡화, 여성의류 3개 파트가 소비의 50~60% 가량을 차지함
- 이외의 파트의 경우 등급에 따른 소비 비중 변화가 거의 나타나지 않음
- 상위 등급에서 하위 등급으로 갈수록 소비에서 명품의 비중이 감소하며, 명품 대신 여성의류 및 잡화로 대체함을 추정 가능함

2-6. 수입품/국내품 시각화

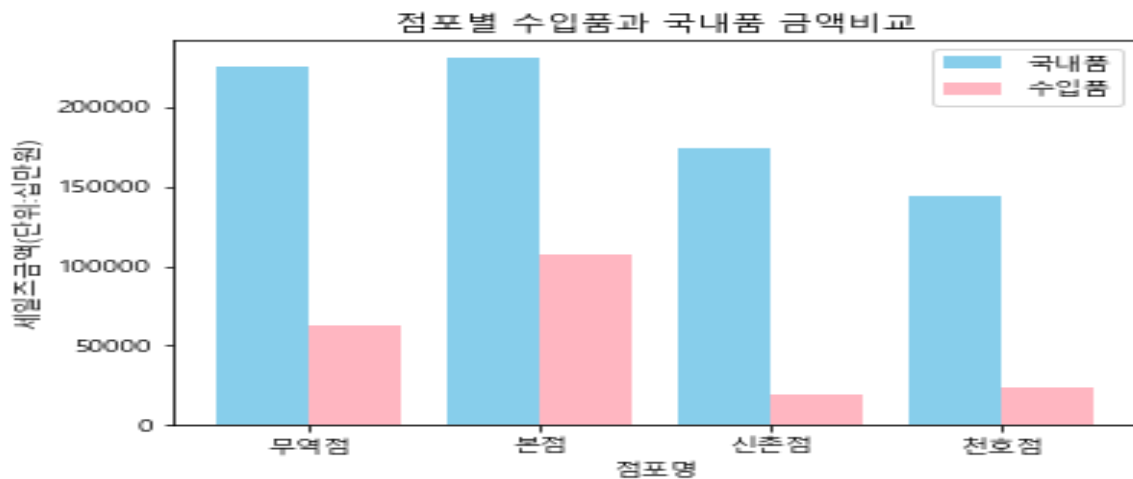
2-6-1) 매출별 수입품과 국내품 비중

매출별 수입품/국내품 비중



■ 매출에서 국내품이 차지하는 비중이 78.3 퍼센트로 수입품에 비해 4 배정도 높음

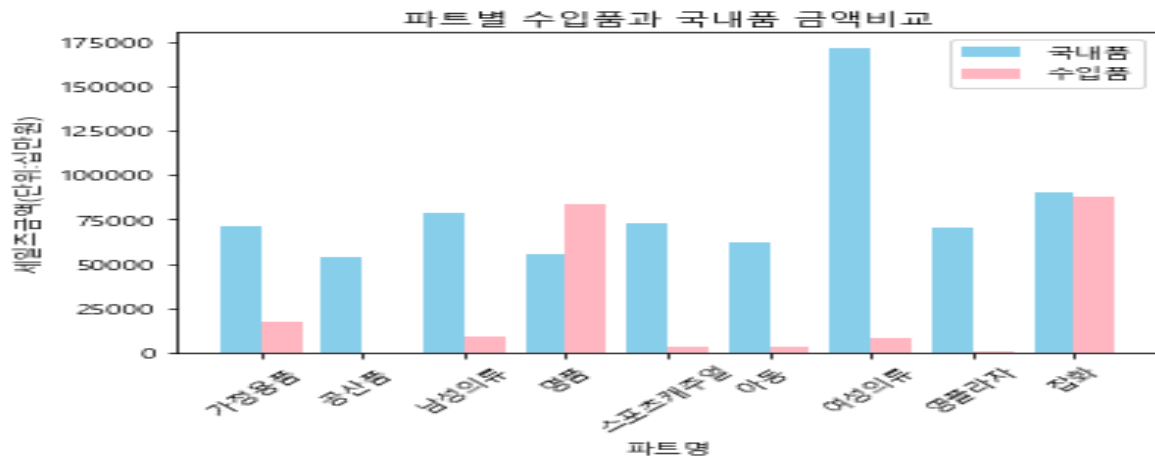
2-6-2) 지점별 수입품과 국내품 매출 비교



■ 모든 점포에서 수입품보다는 국내품의 매출이 높게 나타남

■ 본점에서는 수입품 매출이 타 지점보다는 크게 나타나지만 국내품 매출의 50%에 그침

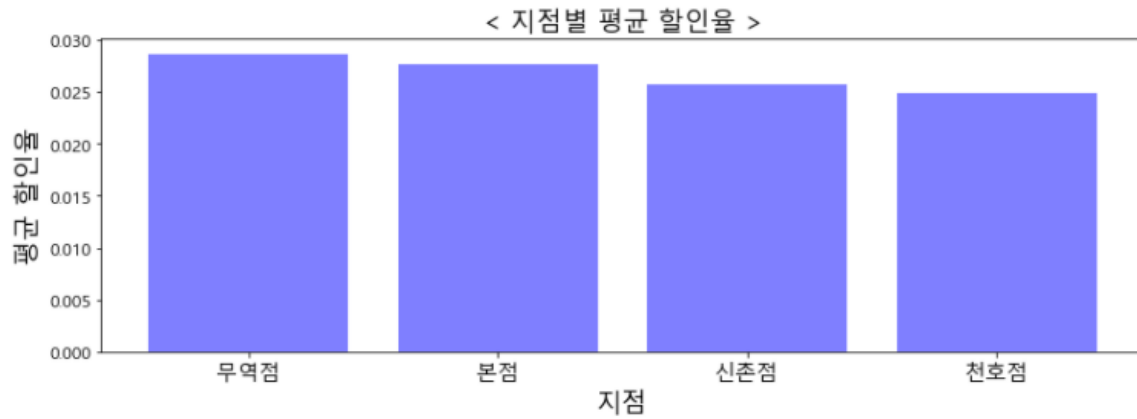
2-6-3) 파트별 수입품과 국내품 매출 비교



- 잡화와 명품의 수입품 매출이 가장 높음
- 명품과 잡화를 제외한 모든 파트에서는 수입품보다 국내품의 매출이 월등히 높음
- 생식품 파트의 수입품 매출은 나타나지 않아 배제하였고, 공산품의 경우에도 수입품 매출이 139,550 원으로 막대 그래프상 수치가 거의 나타나지 않음

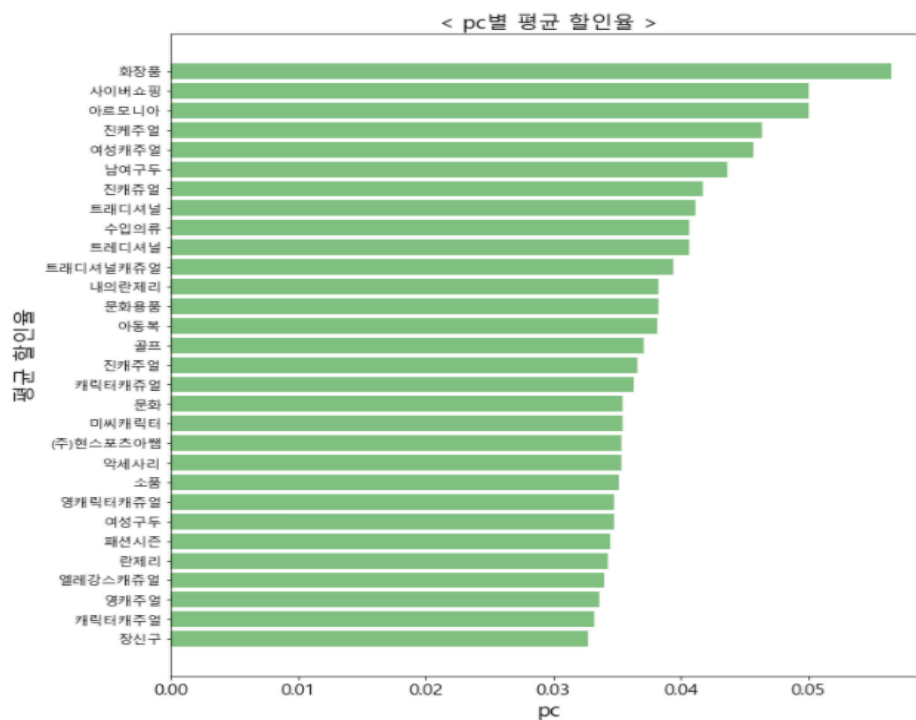
2-7 할인율 기준 시각화

2-7-1) 지점 평균 할인율

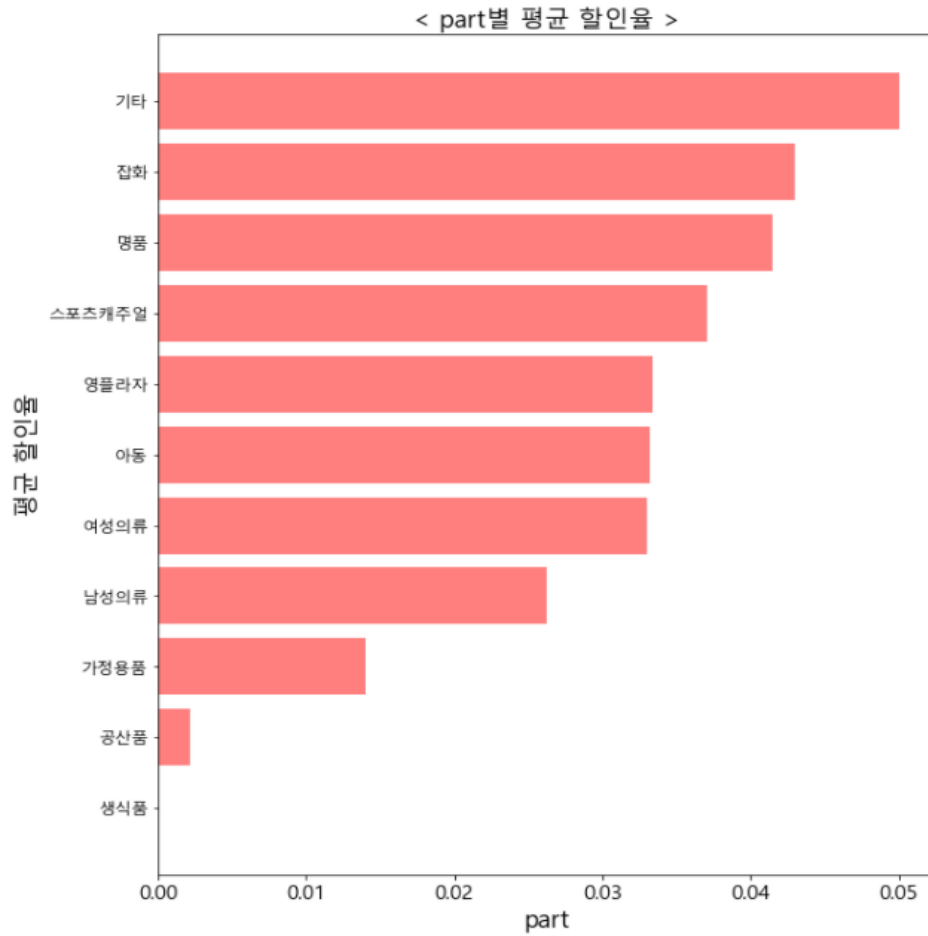


■ 모든 지점 2.5%~3%사이의 고른 평균할인율을 가짐

2-7-2) part, pc, brand 별 평균 할인율

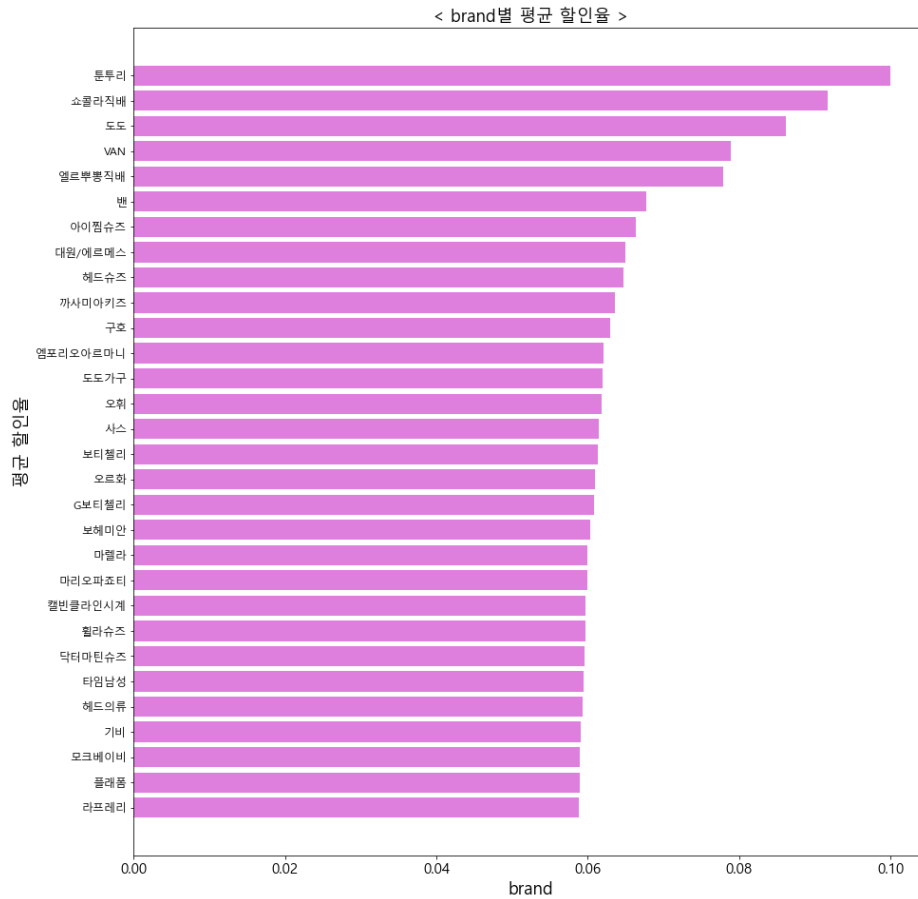


■ 화장품, 사이버 쇼핑, 항목이 약 5%로 가장 큰 할인율을 보임



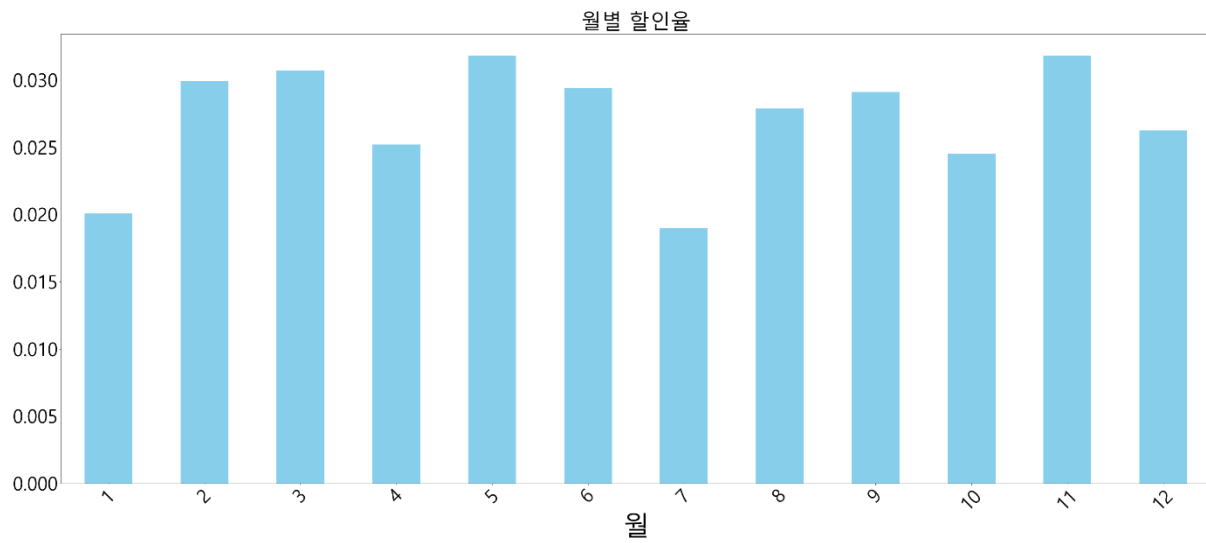
■ 잡화, 명품 항목이 순으로 큰 할인율을 보임

■ 공산품은 1%미만의 평균할인을 하였고, 생식품은 할인을 거의 하지 않음



■ 툰투리가 10%로 평균할인율이 가장 큰 브랜드였고 2,3 번째는 쇼콜라직배, 도도가 9%, 8%의 할인율을 높은 평균할인율을 가짐

2-7-3) 월별 할인율 평균 시각화

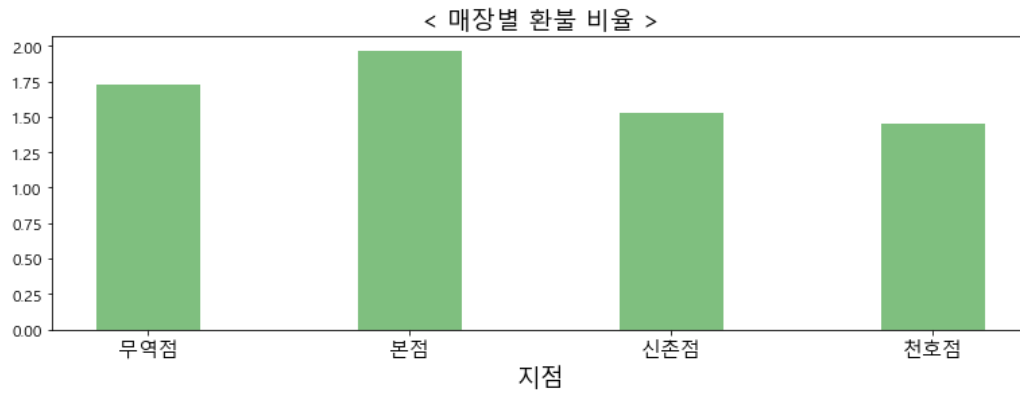


■ 모든 월에서 2%~3%사이의 고른 평균할인율을 가짐

■ 비교적 1 월, 7 월이 낮은 할인율을 가지며 5 월과 11 월이 높은 할인율을 가짐

2-8. 환불 기준 시각화

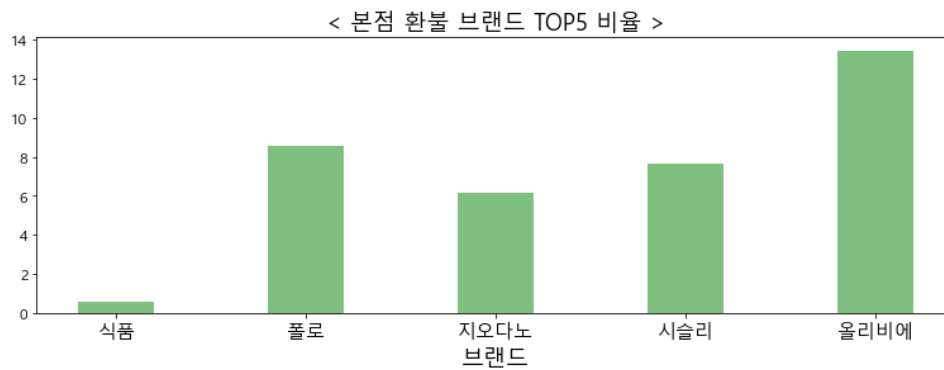
2-8-1) 지점별 환불액

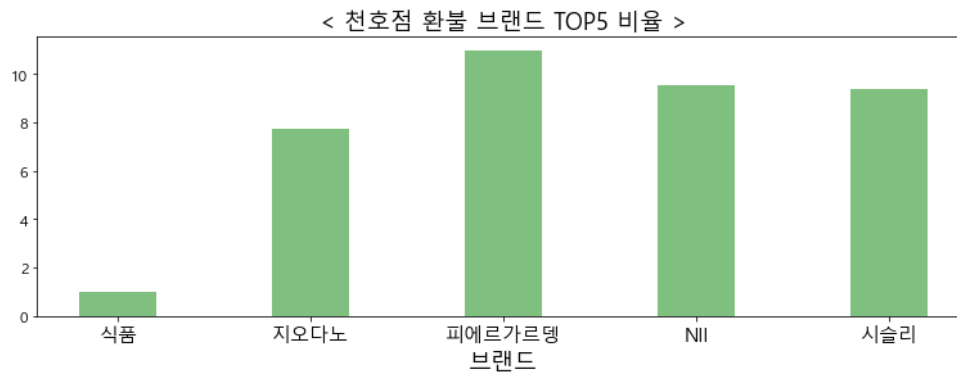
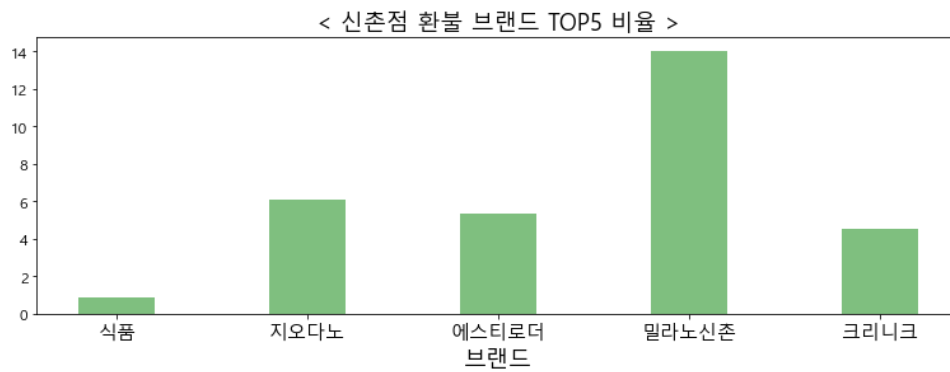
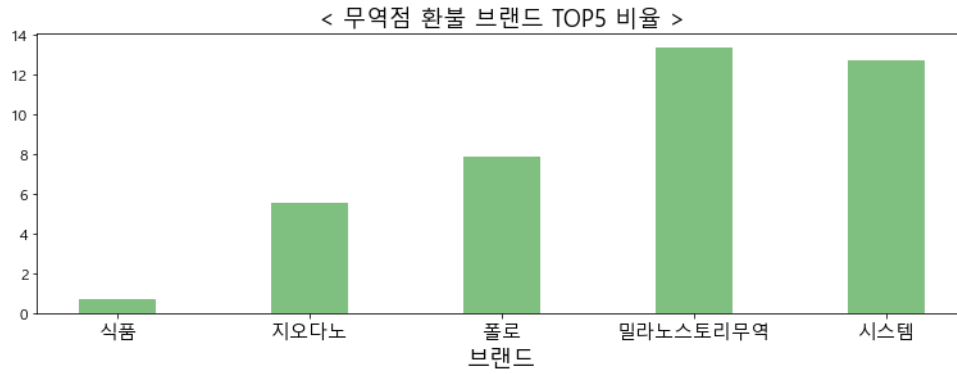


■ 본점, 무역점, 신촌점, 천호점 순으로 매장별 환불 비율이 높음

■ 환불비율은 '환불횟수 / 전체 구매횟수'

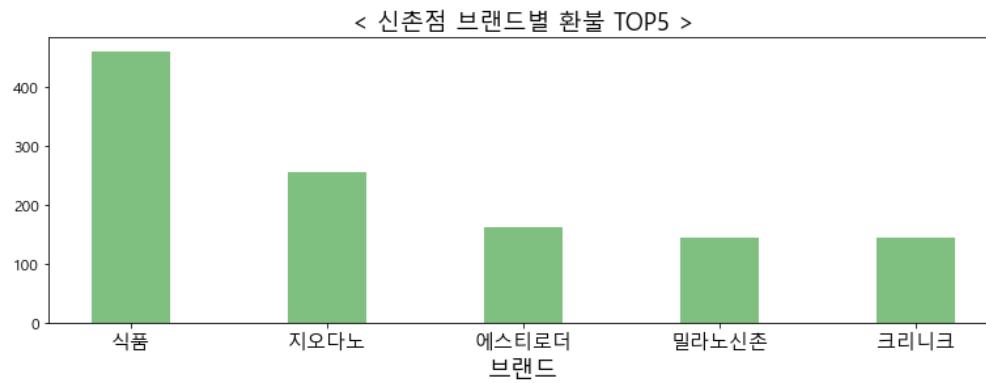
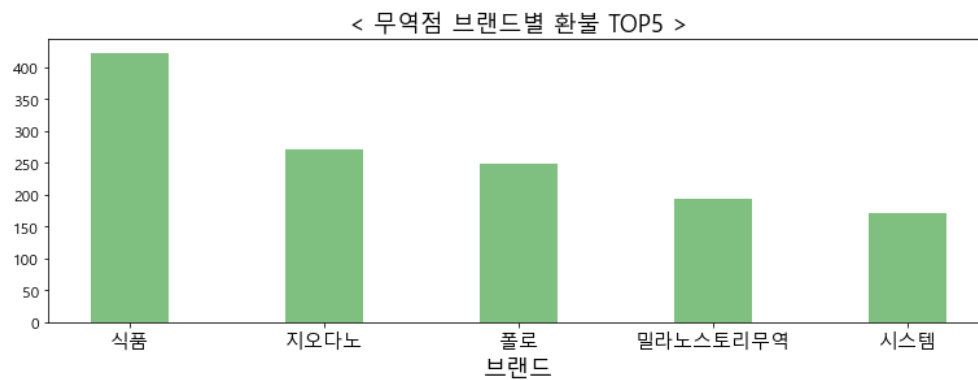
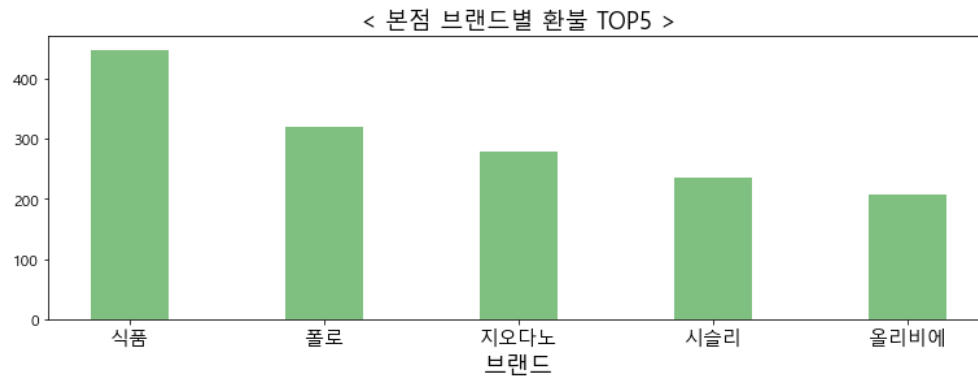
2-8-2) 지점별 상위 환불 비율 브랜드

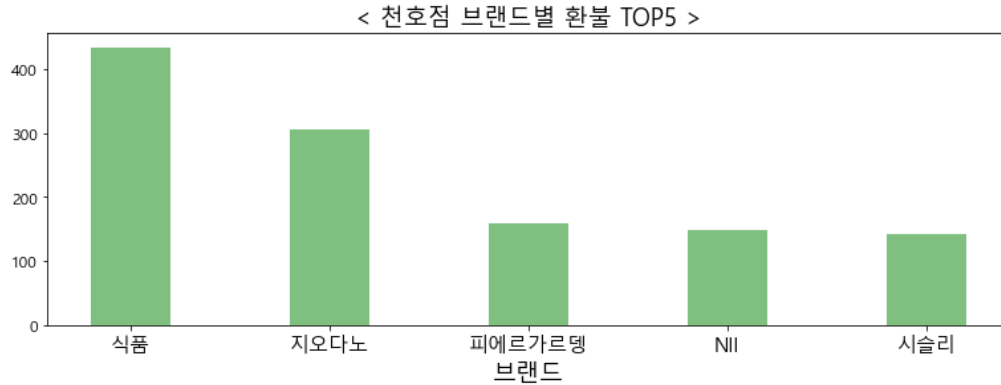




■ 지점별 환불 횟수가 많은 브랜드와 환불 비율이 높은 브랜드의 차이가 있으므로 조금 더 자세한 분석 필요

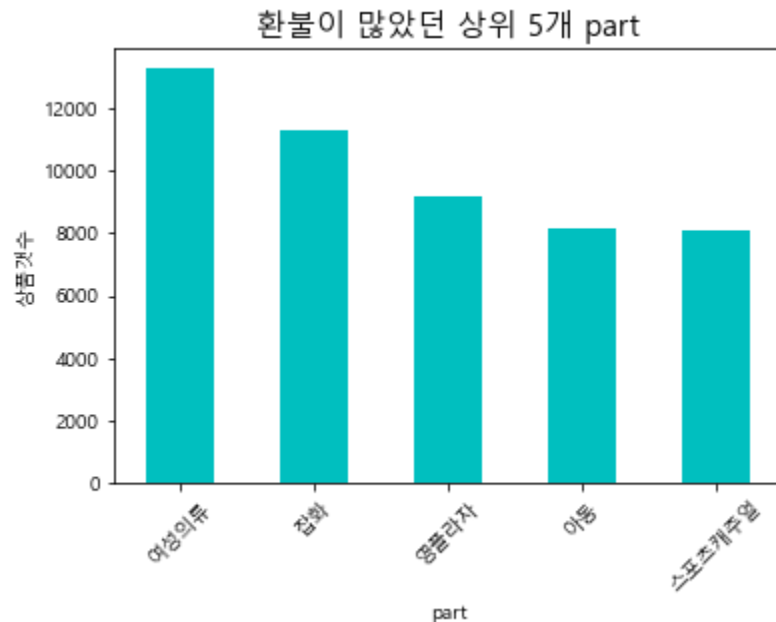
2-8-3) 지점별 상위 환불 횟수 브랜드



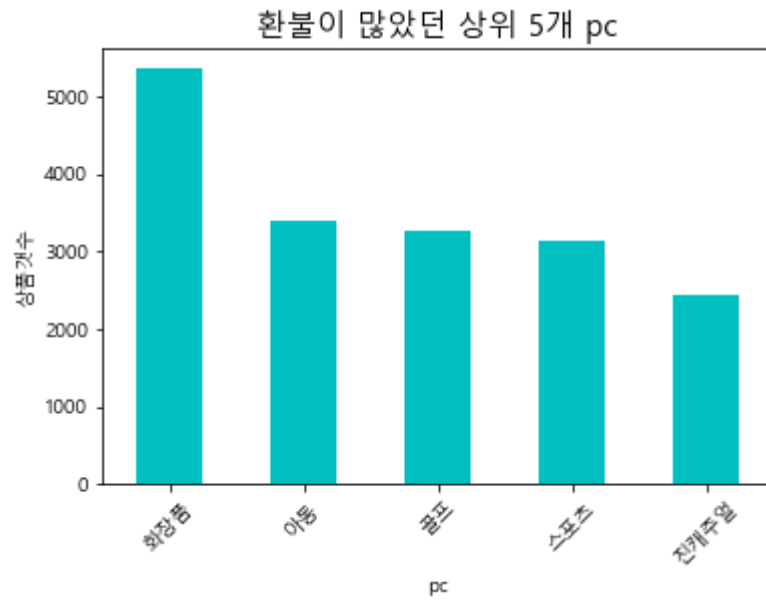


- 지점별로 환불 브랜드의 차이가 있으나 식품과 지오다노의 비중이 높은 것을 알 수 있음
- 식품의 경우 공산품 파트와 생식품 파트를 포함하면 백화점 매출의 10%를 차지하는 부분이며, 고객들에게 접근성이 가장 높은 파트이기 때문에 환불 횟수 또한 많음

2-8-4) 상위 환불 pc, part



- Part 별 환불 TOP5 의 경우 여성의류, 스포츠캐주얼 등 의류 부분에 많이 치중되어 있음



- PC 별 경우 Part 별과는 다르게 다양한 품목별로 환불이 되는 것을 알 수 있음

1. 데이터 처리

- 실제 데이터이기 때문에 통계분석 이전에 세부 데이터를 확인하여 전처리 하는 것이 중요함. 전처리를 하지 않는다면 오타, 띄어쓰기, 통일되어있지 않은 명칭 등으로 통계 분석 결과를 오염시킬 수 있음.

2. 매출액 기준 분석

- Part별 상위 10개 브랜드의 그래프를 도출해 본 결과, 모든 Part에서 1위 브랜드의 매출이 도드라짐. 해당 브랜드로부터의 지속적인 매출과 더불어 해당 브랜드를 이용하여 다른 상품이나 파트의 매출창출이 가능하도록 연결고리를 생각해 볼 필요가 있음.

3. 지점기준 분석

- 본점과, 무역점은 유의미하게 특정 파트의 매출 비중이 높음. 해당 지점이 그 파트에 주력하고 있음을 알 수 있음. 그 파트의 쇼핑 용이성 등을 지점 홍보에 활용 가능.

4. 날짜 및 날씨 기준 분석

- 주말에 고객 수와 매출이 쏠려 있기 때문에 주말에 매장 내 혼잡을 줄이고 편안한 쇼핑을 할 수 있는 환경 조성에 노력이 필요함.
- 오전시간대와 월요일 매출이 낮기 때문에 이를 높이기 위한 할인 행사와 기획이 필요해보임.
- 날씨와 강수량은 매출과 구매 고객수에 유의미한 상관관계가 나타나지 않음.

5. 성별 기준 분석

- 여성고객이 63% 이상의 비중으로 남성고객보다 많음. 성별로 파트별 구매비율에 있어서는 큰 차이가 있지 않지만, 여성 고객이 많다는 점을 향후 백화점 운영 기획에 유의하는 것이 필요함.

6. 고객등급 기준 분석

- 매출액의 86%를 차지하는 Red, Normal, Black 등급을 타겟으로 한 마케팅 전략이 주력이 되어야 함. 한편 상위 1.5%의 고객이 약 15%의 매출을 차지하고 있기 때문에 고객 맞춤형 프리미엄 서비스를 지속적으로 제공해야 타 백화점으로서의 중요고객 유출을 방지할 수 있음.

7. 수입품/국내품 기준 분석

- 국내품 매출액이 수입품 매출액 보다 약 4 배정도 높은 매출을 보이고 있음. 하지만 국내품목이 차지하는 비중이 수입품목보다 훨씬 많기 때문에 고객들이 국내품을 더 선호한다고 보기에는 어려움. 수입품의 판매를 높이려면 해외 인기브랜드의 추가적인 입점 노력이 필요함.

8. 할인율 기준 분석

- 할인율 데이터와 매출액 데이터, 브랜드별 할인율 데이터와 매출액 데이터를 통하여 할인 행사 기간 및 브랜드를 유동적으로 선정할 수 있음.

9. 환불 기준 분석

- 환불 횟수는 식품의 비중이 월등히 높는데 그 이유는 여러 가지 음식을 식품이라는 하나의 카테고리로 분류하였기 때문으로 파악됨. 식품 코너의 효율적인 운영을 위해 식품의 환불 원인 등을 분석하여 줄여나가는 노력이 필요함.