

CHAI Data Analytics

Team 도원결의

양대용, 이민정, 이진규, 정주훈

CONTENTS

01. 데이터 소개

02. 기본 분석

03. 심화 분석

04. 결과 및 결론



2019년 6월 출시 ⚡

은행 계좌 기반의 간편 결제 서비스

2019년 8월 - 2020년 3월 ⚡

해당 기간의 차이 결제 데이터

약 562만건 · 1150억 · 54만 ⚡

총 거래건수 · 총 결제금액 · 총 유저수

- 데이터 컬럼명

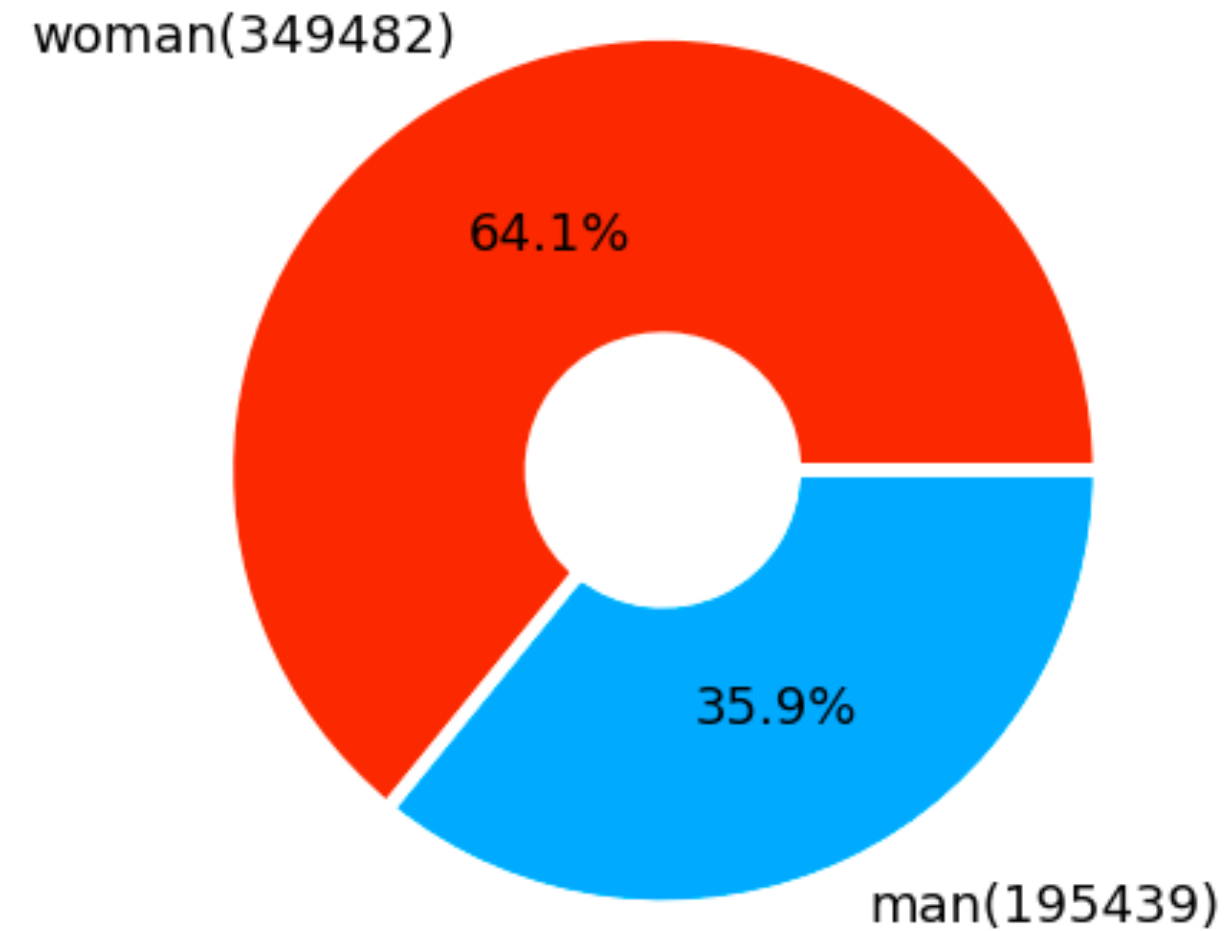
customer_id	gender	is_foreigner	push_permission	birthday	sign_up_date	created_at
유저 ID	성별	외국인 여부	Push 알림 설정 여부	생일	가입일	결제 시각

pre_discount	post_discount	cashback_amount	discount_amount	total_promotion	merchant_id
할인 전 금액	할인 후 금액 (캐시백 X)	캐시백 금액	할인 금액	캐시백 & 할인 금액	산업 코드

기본 분석⚡

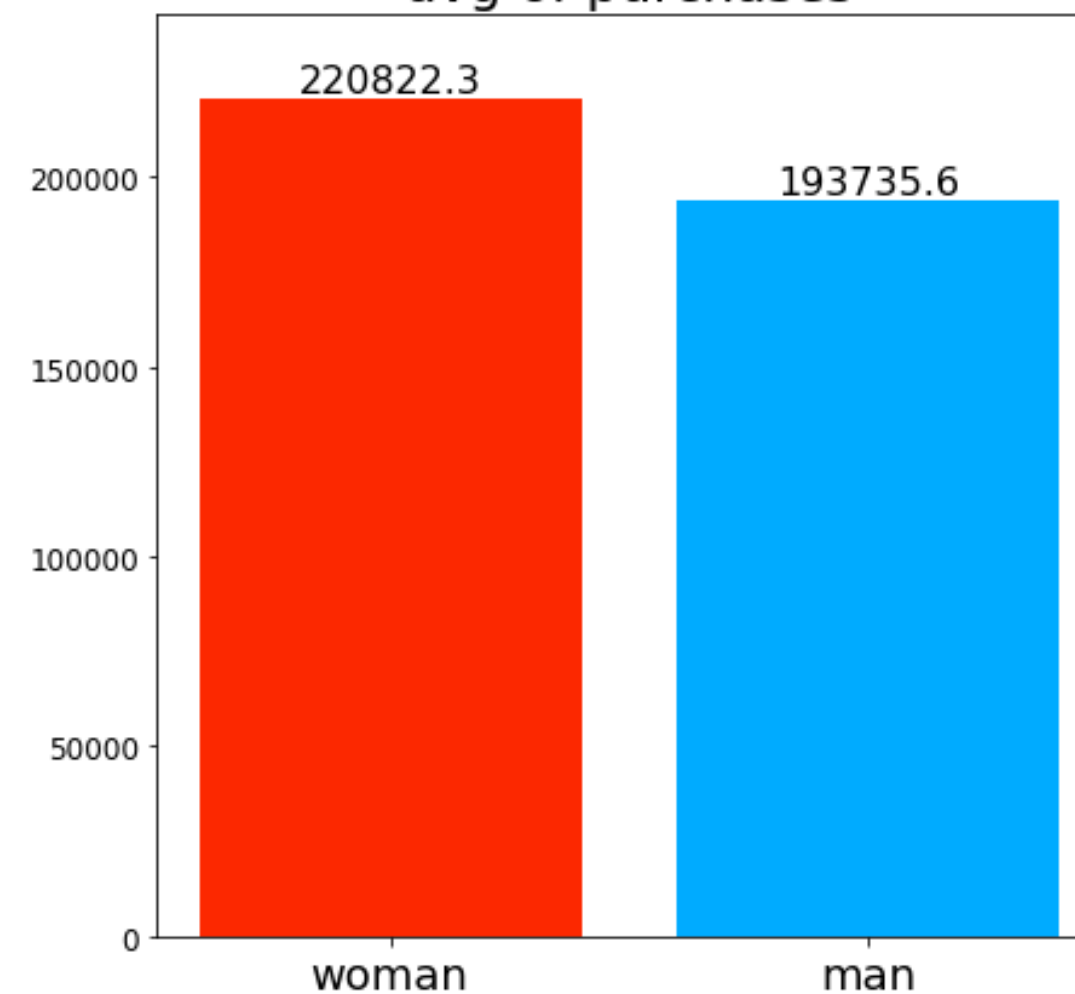
1. 유저별 분포

distributions of gender (user)



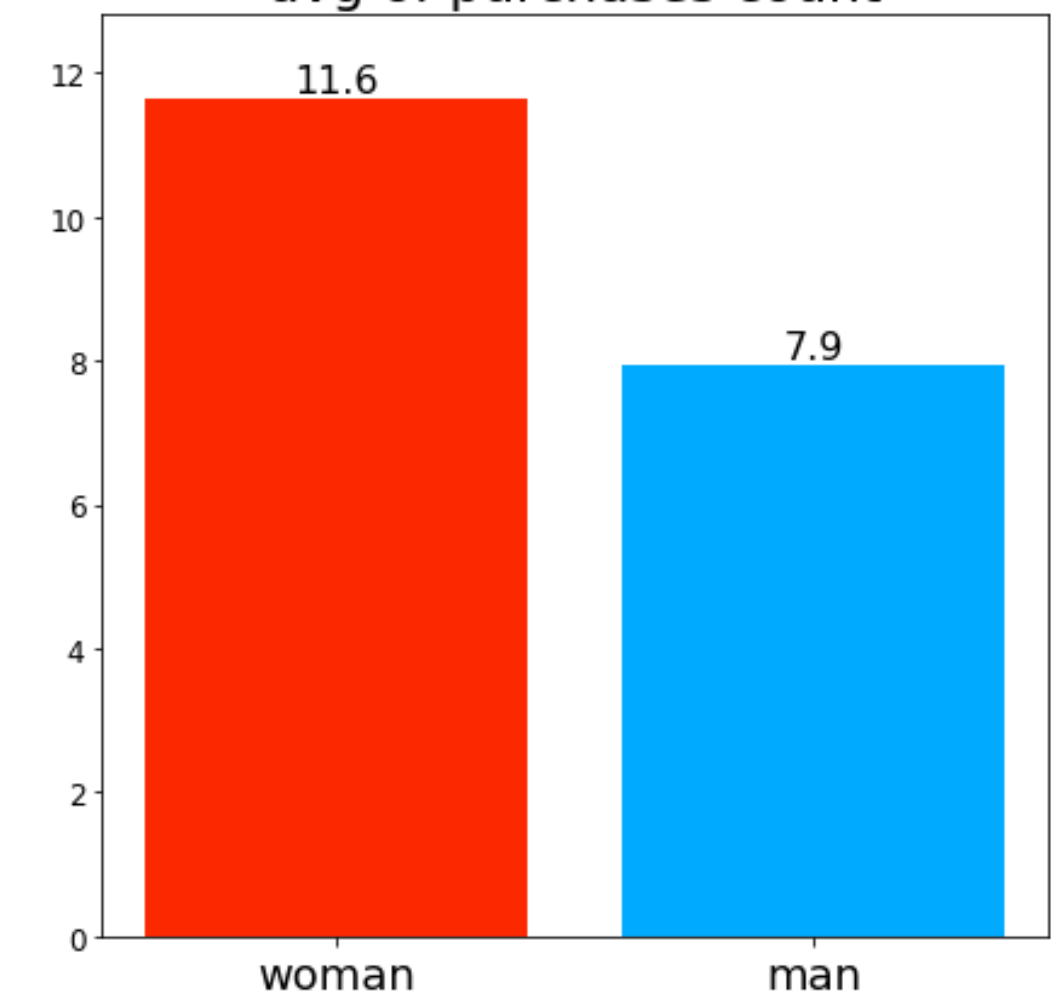
<유저별 성별 분포>

avg of purchases



<유저별 총 구매액 평균>

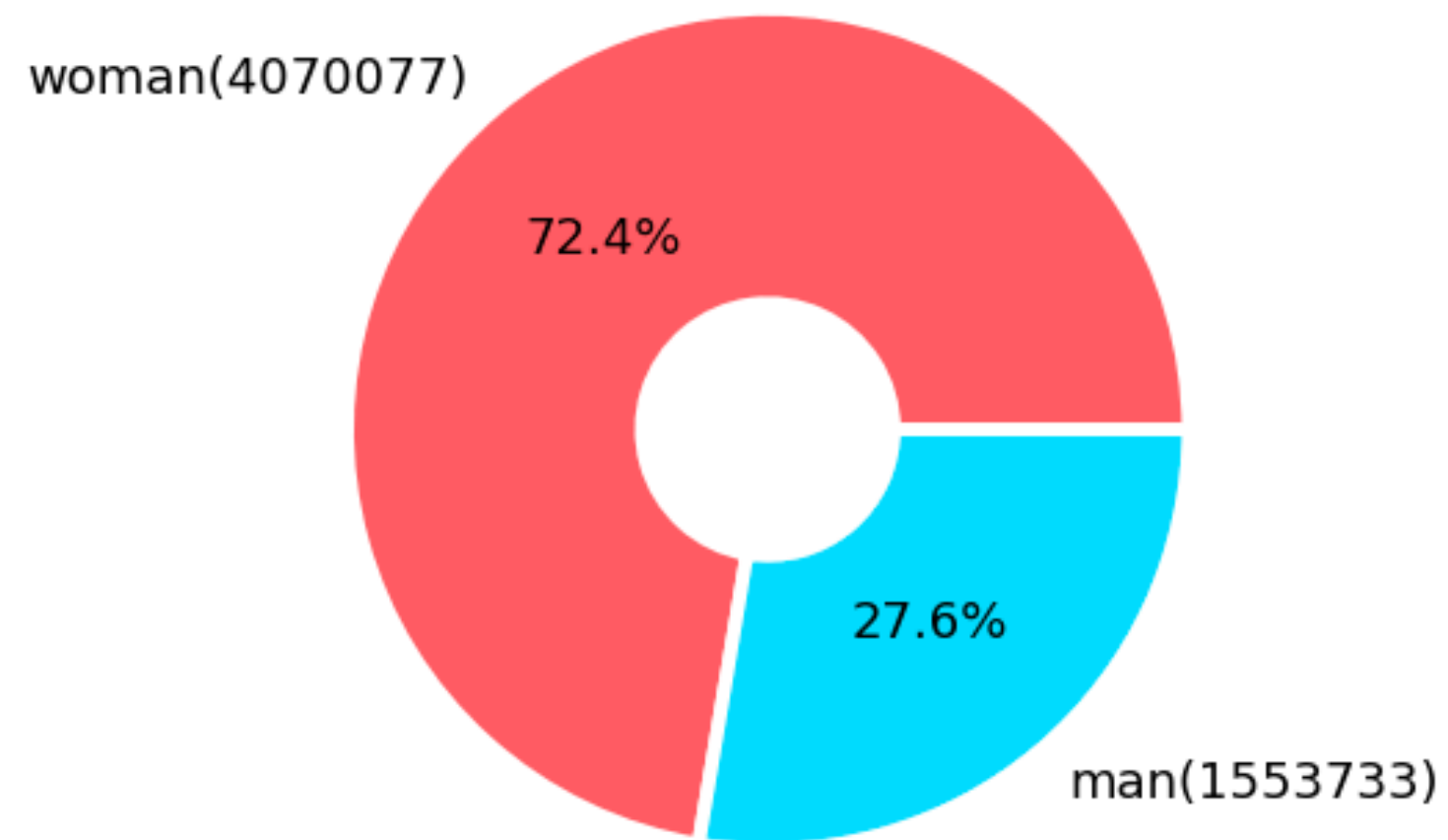
avg of purchases count



<유저별 구매횟수 평균>

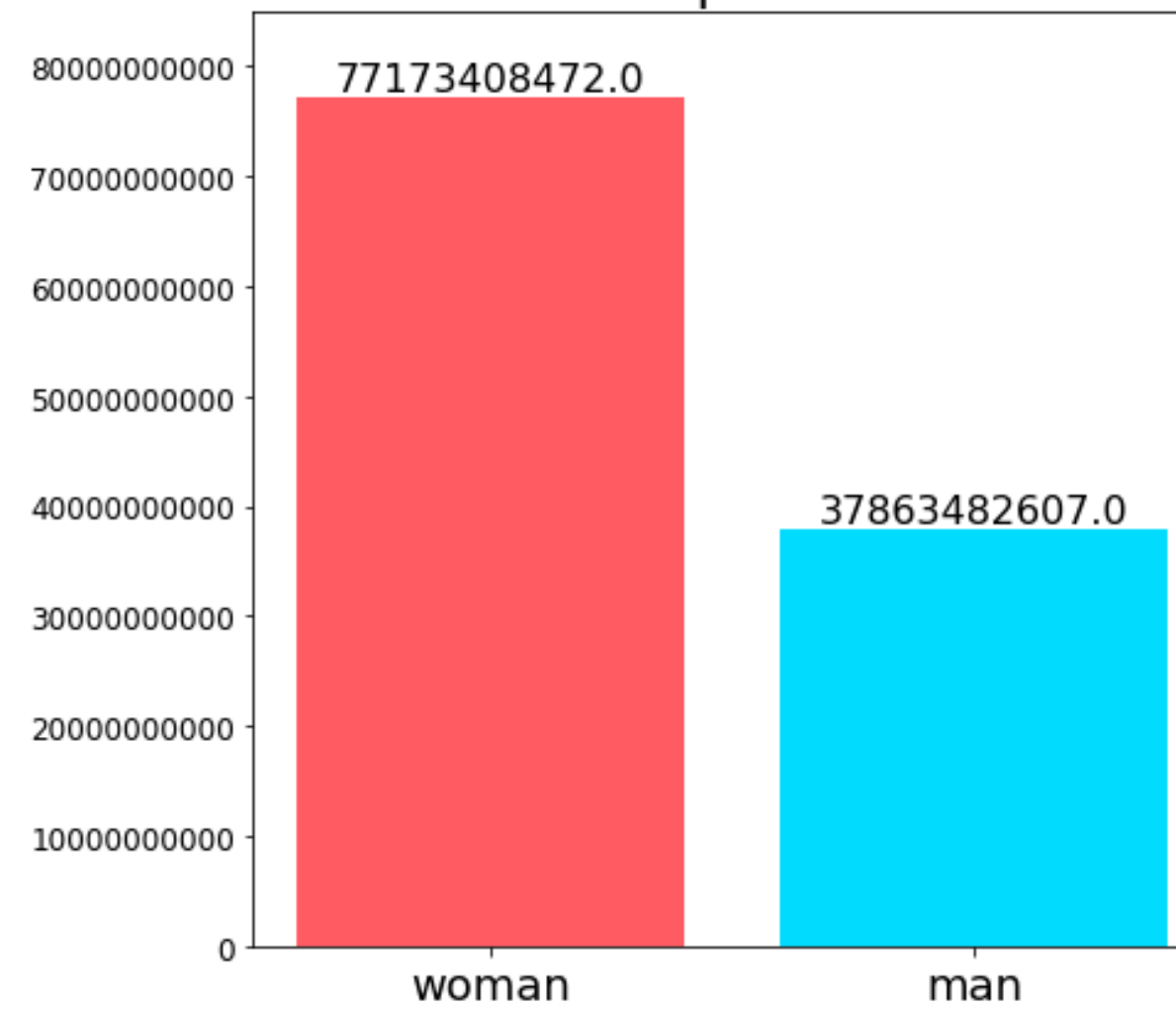
2. 거래별 분포

distributions of gender (case)



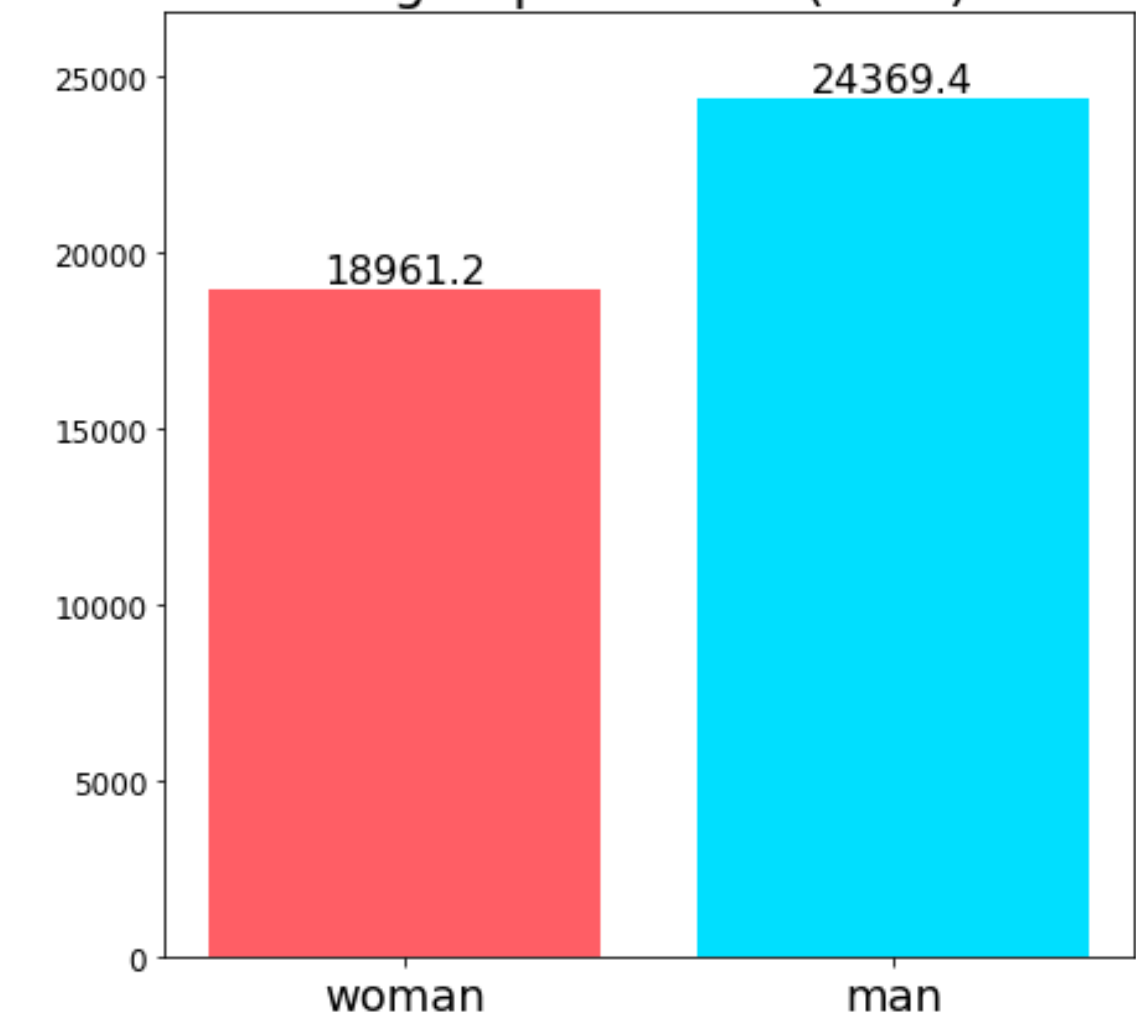
<거래별 성별 분포>

sum of purchases



<거래별 총 구매액>

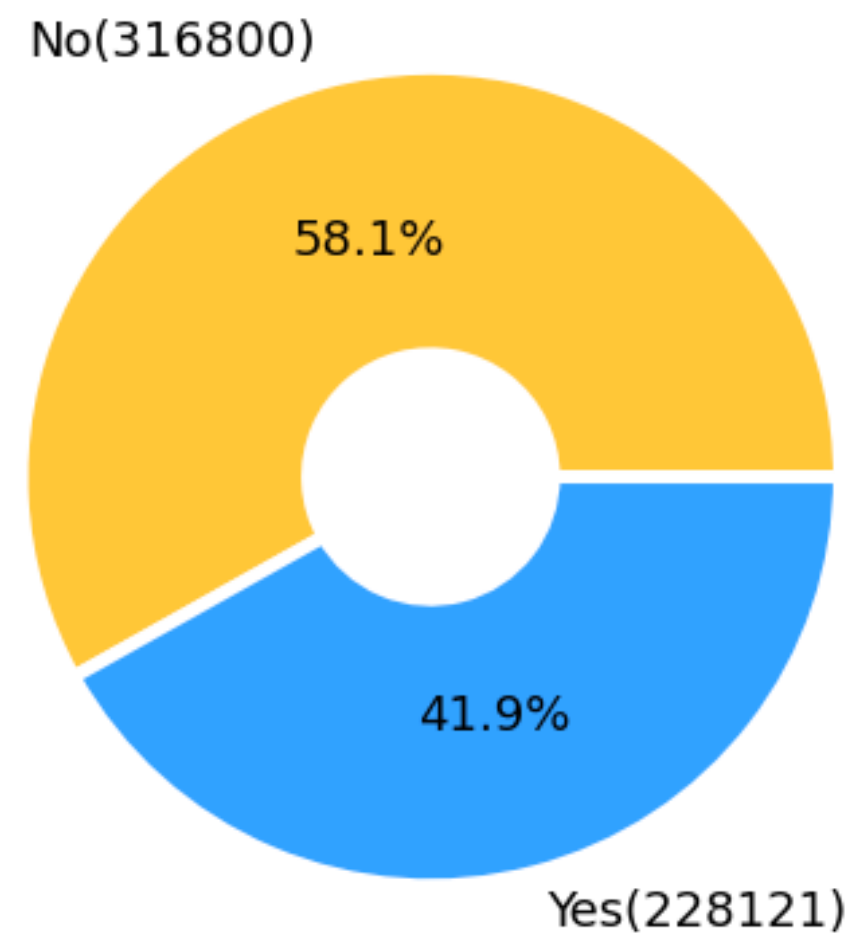
avg of purchases (case)



<거래별 구매액 평균>

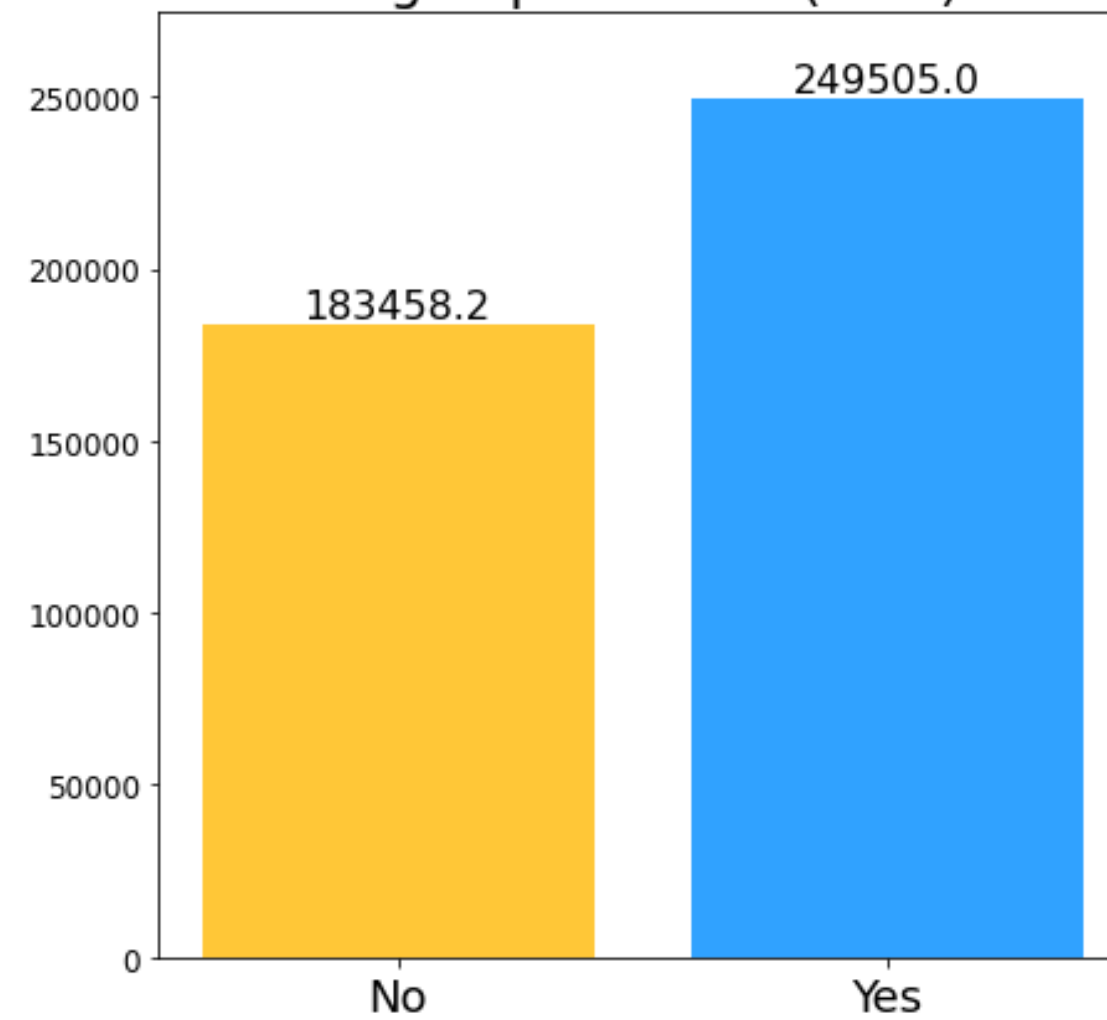
1. 유저별 분포

distributions of push_permission (user)



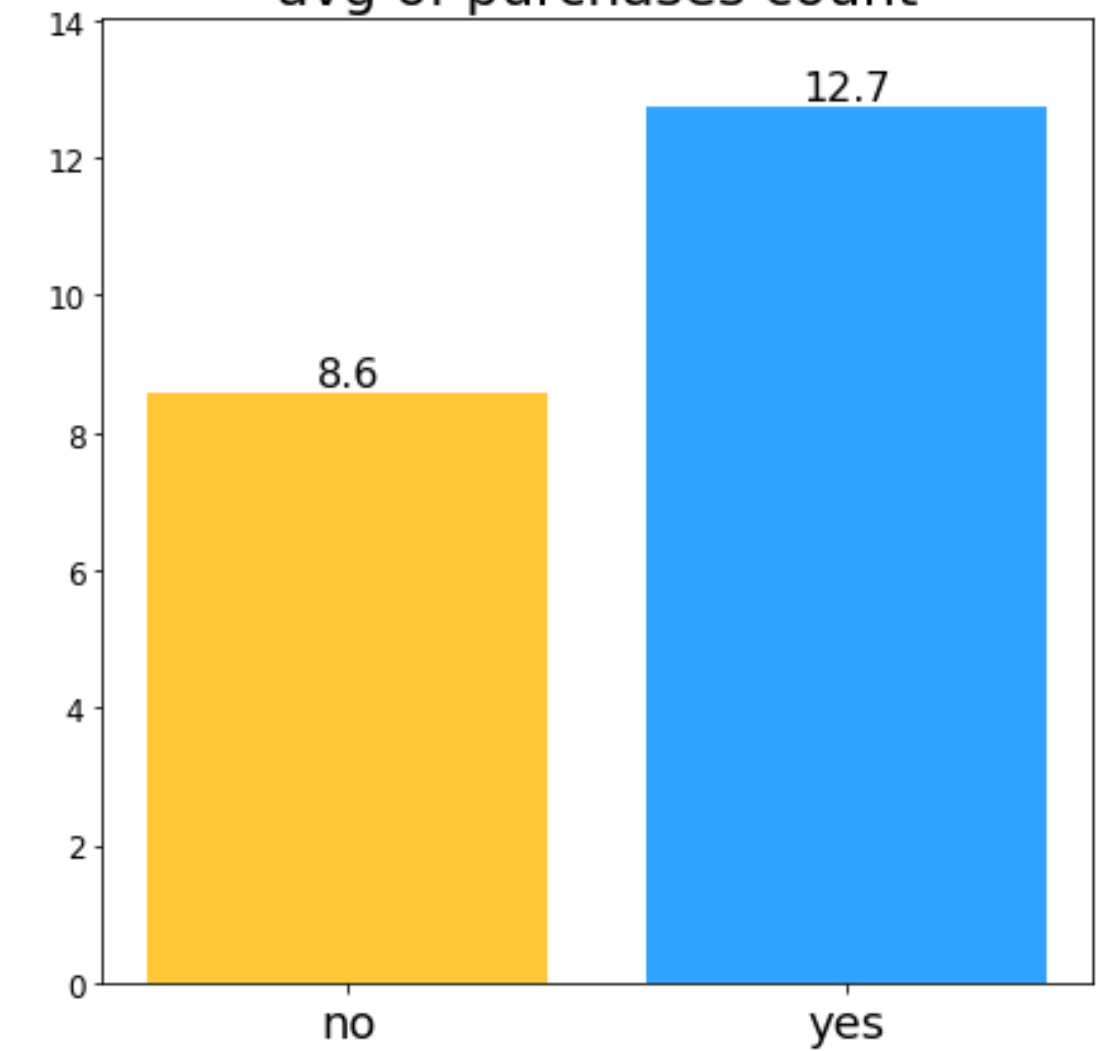
<유저별 Push 설정 분포>

avg of purchases (user)



<Push 유저별 총 구매액 평균>

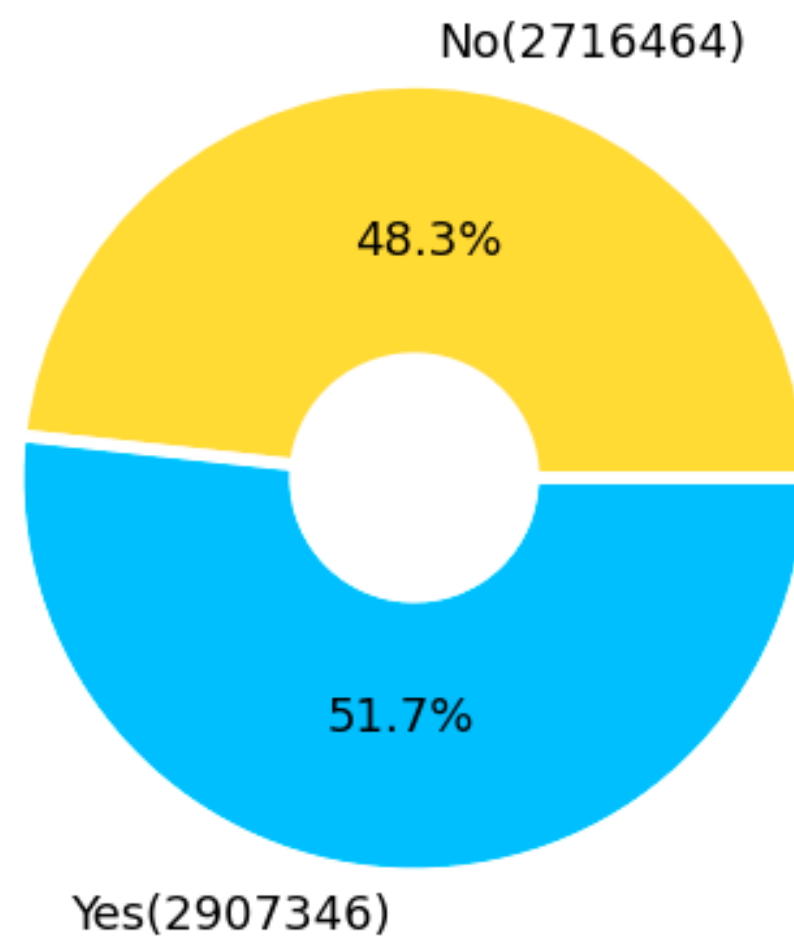
avg of purchases count



<Push 유저별 구매횟수 평균>

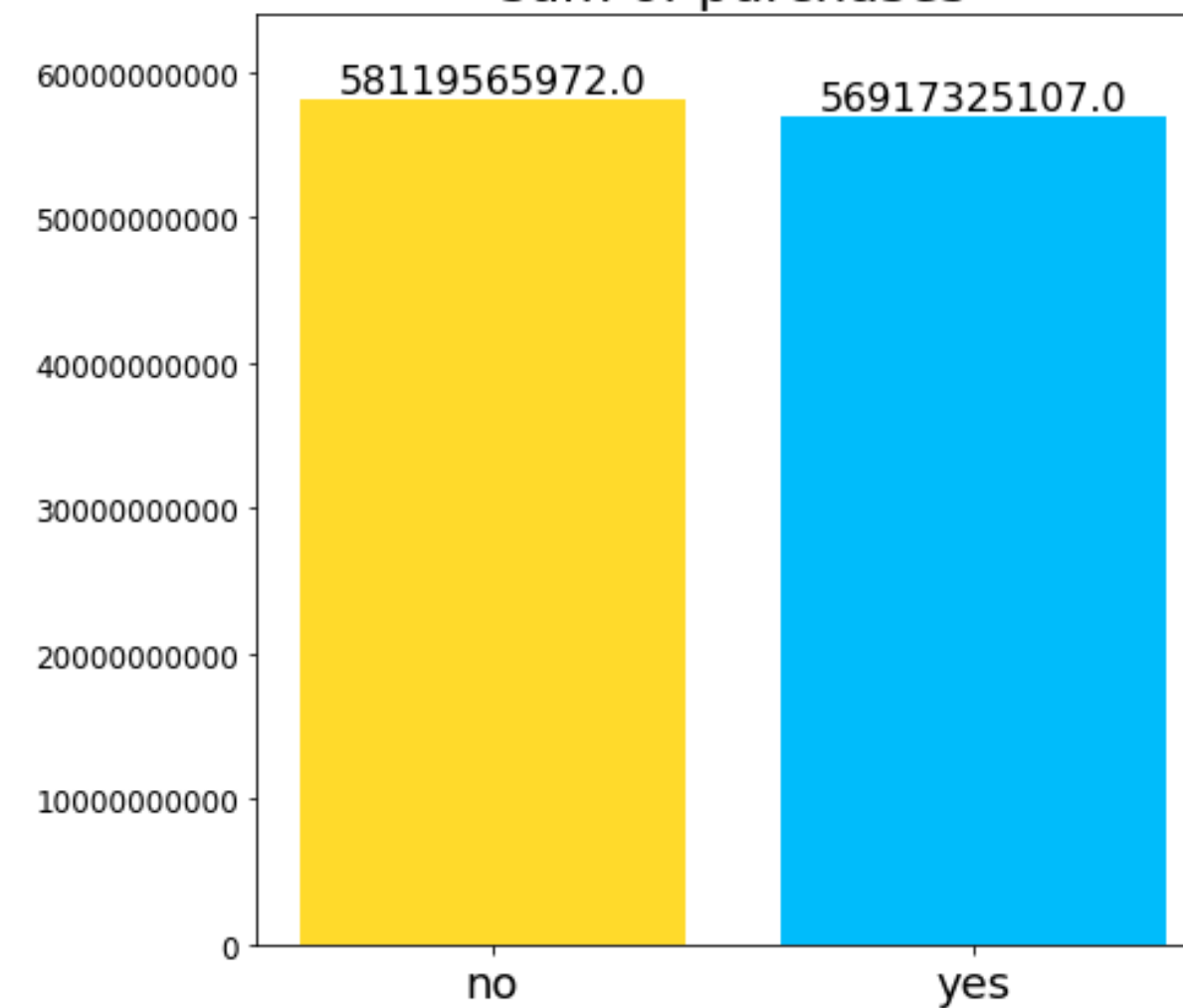
2. 거래별 분포

distributions of push_permission (case)



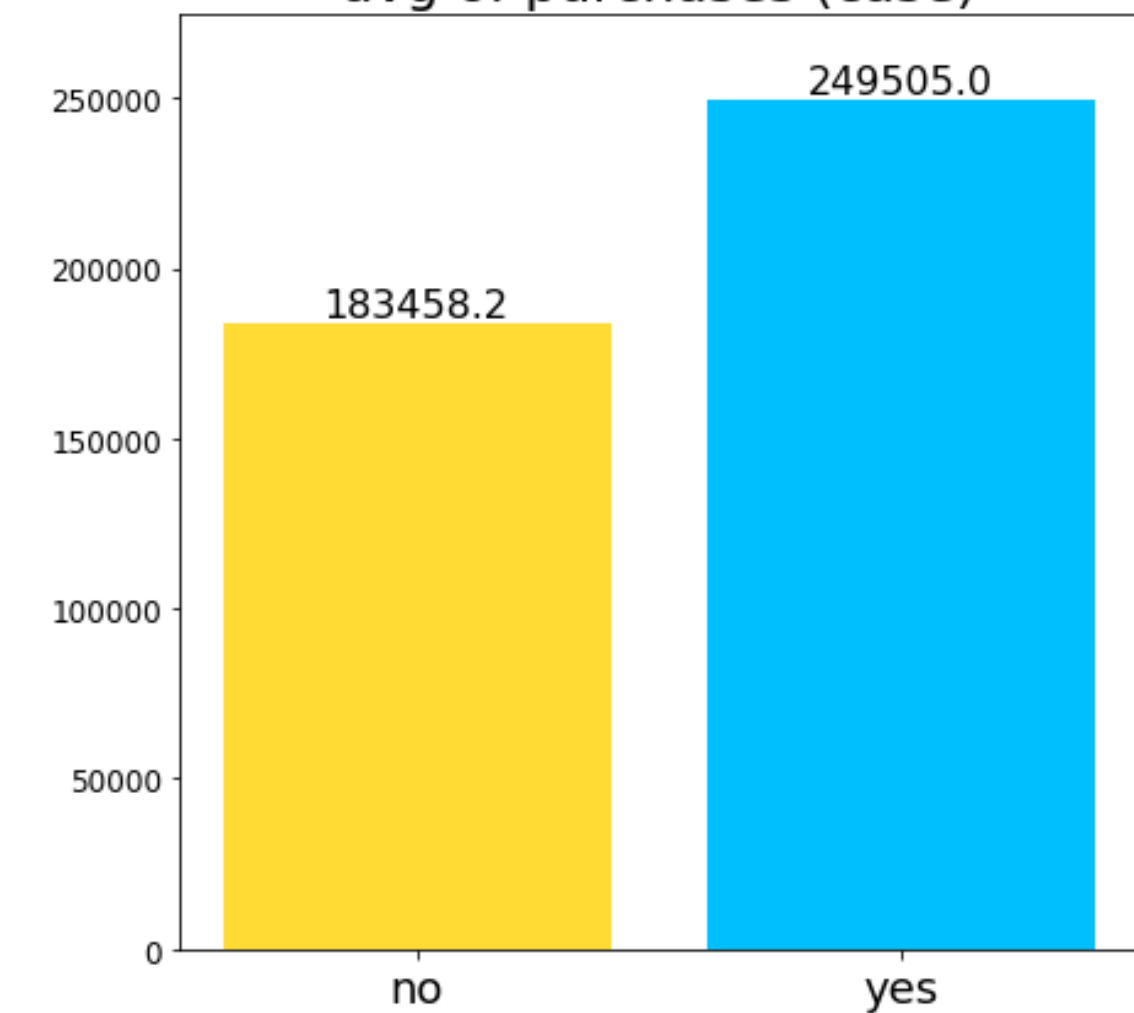
<거래별 Push 설정 분포>

sum of purchases



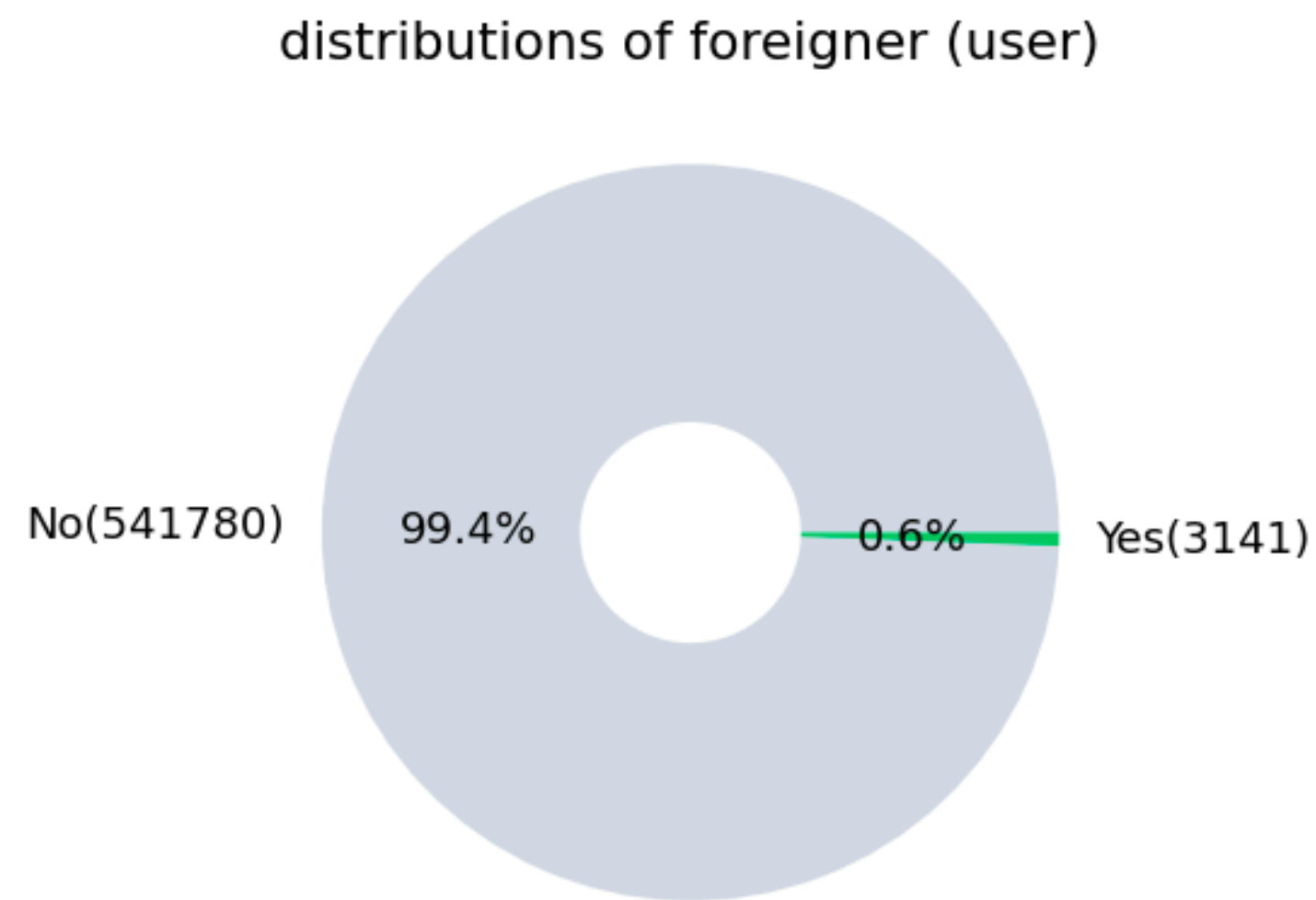
<Push 거래별 총 구매액>

avg of purchases (case)

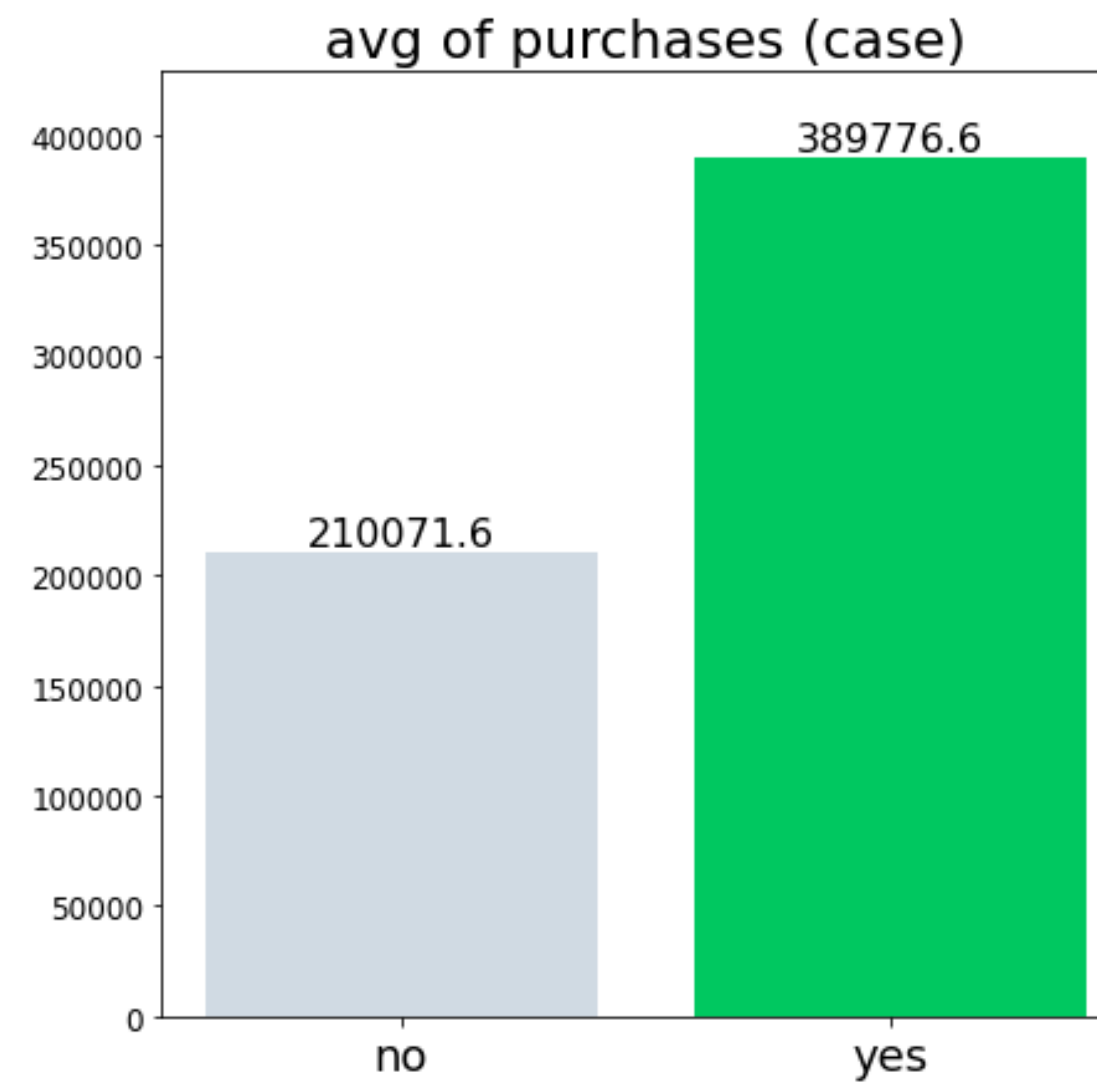


<Push 총 구매 금액 평균>

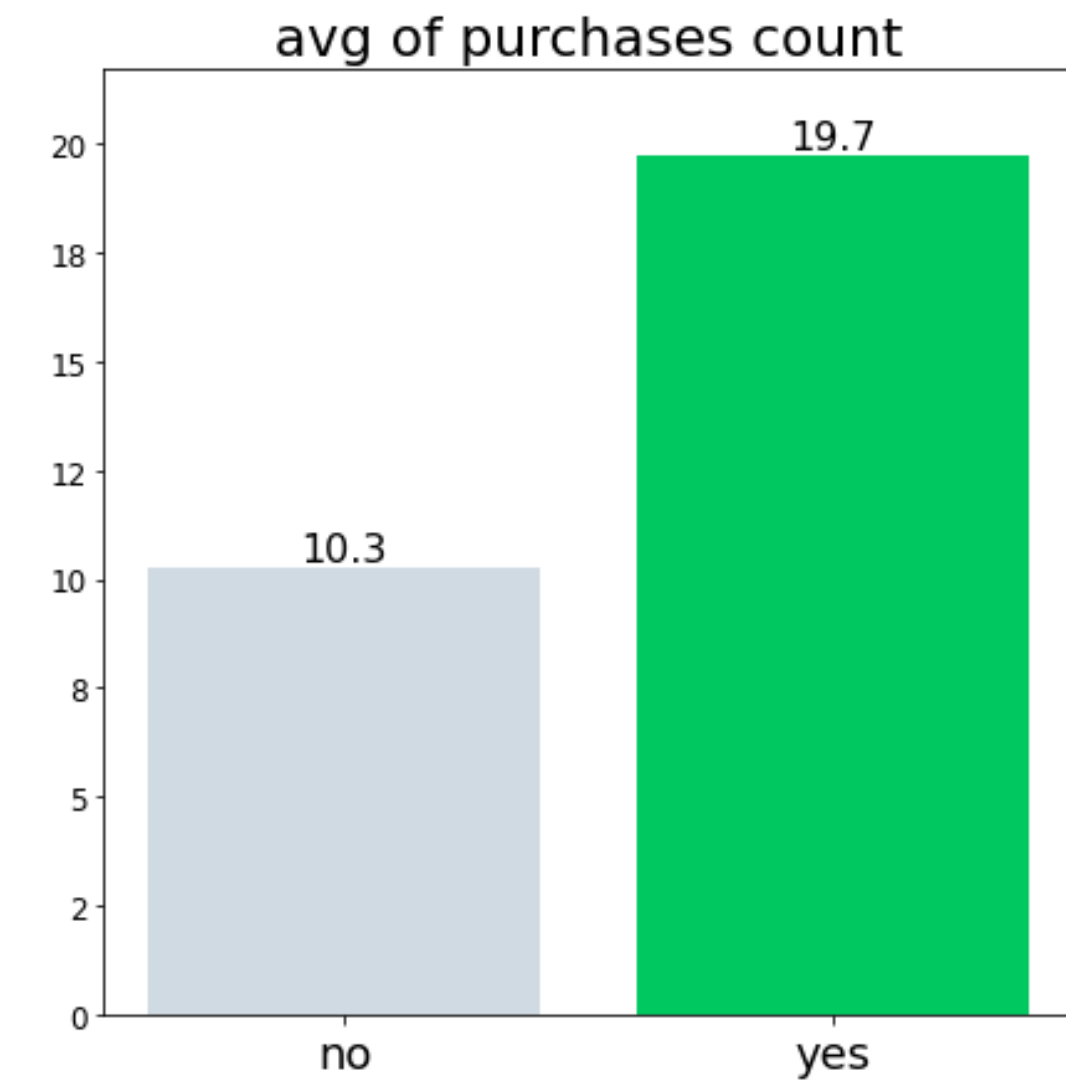
1. 유저별 분포



<유저별 외국인 분포>

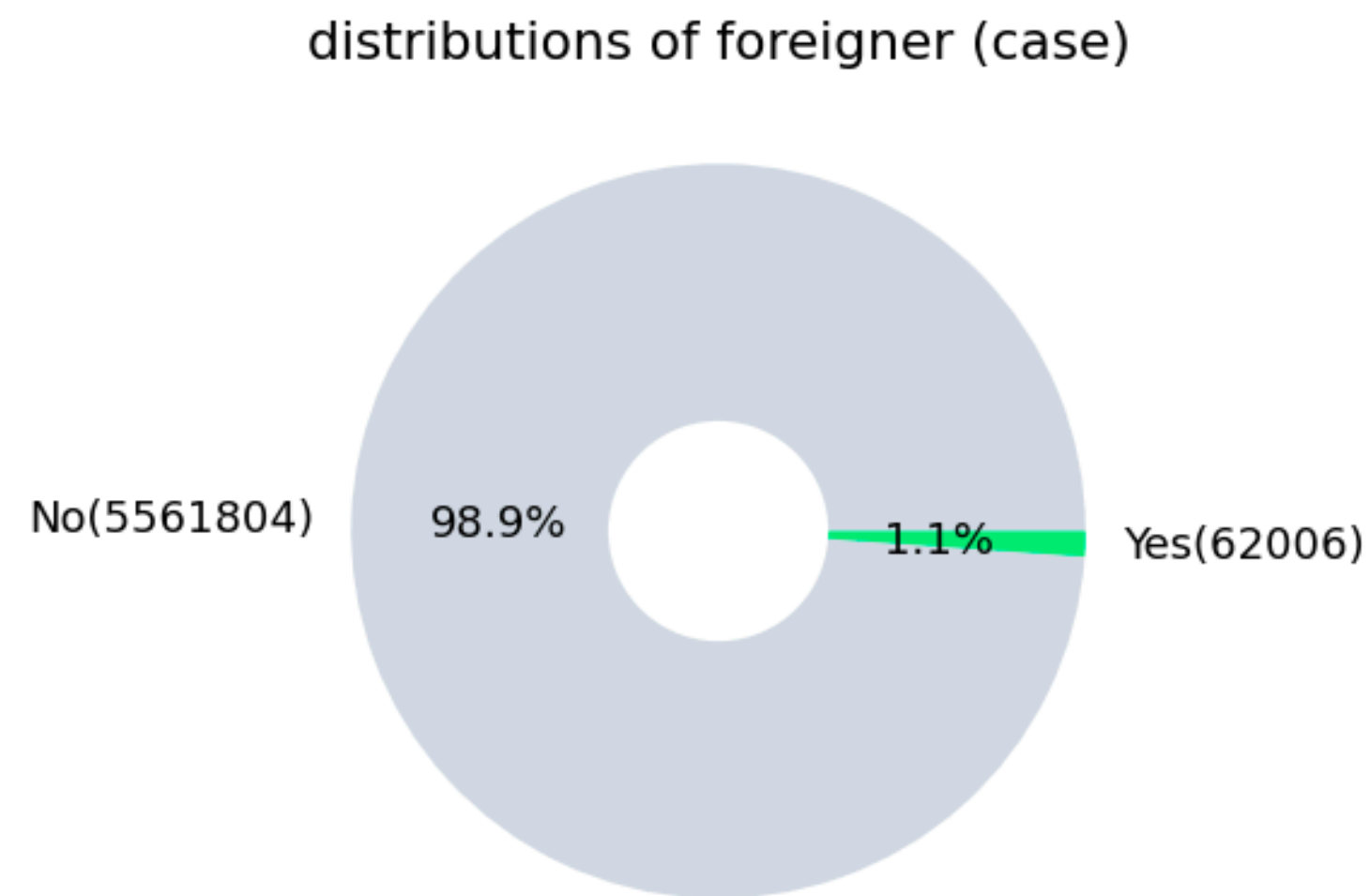


<유저별 총 구매액 평균>

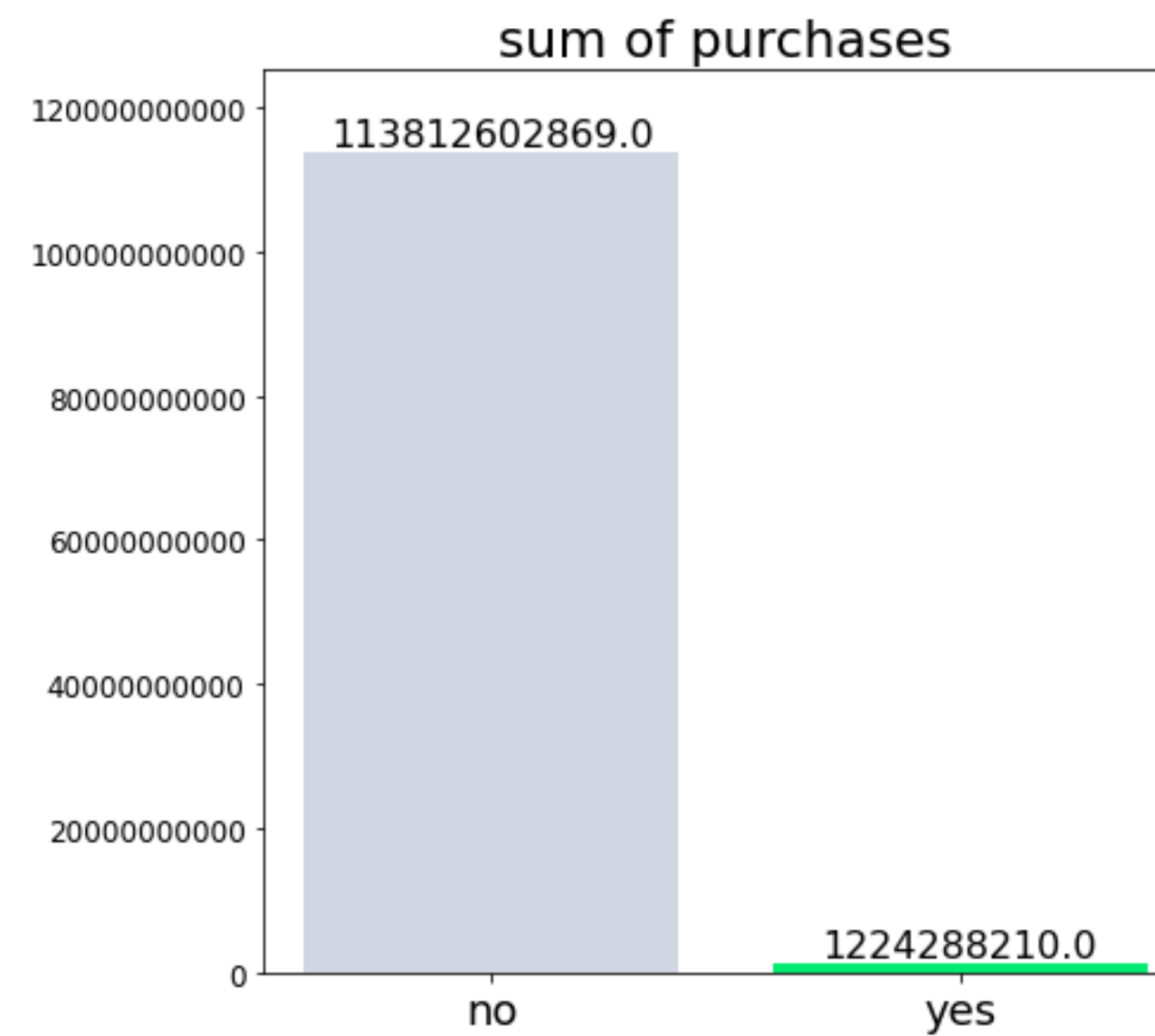


<유저별 구매 횟수 평균>

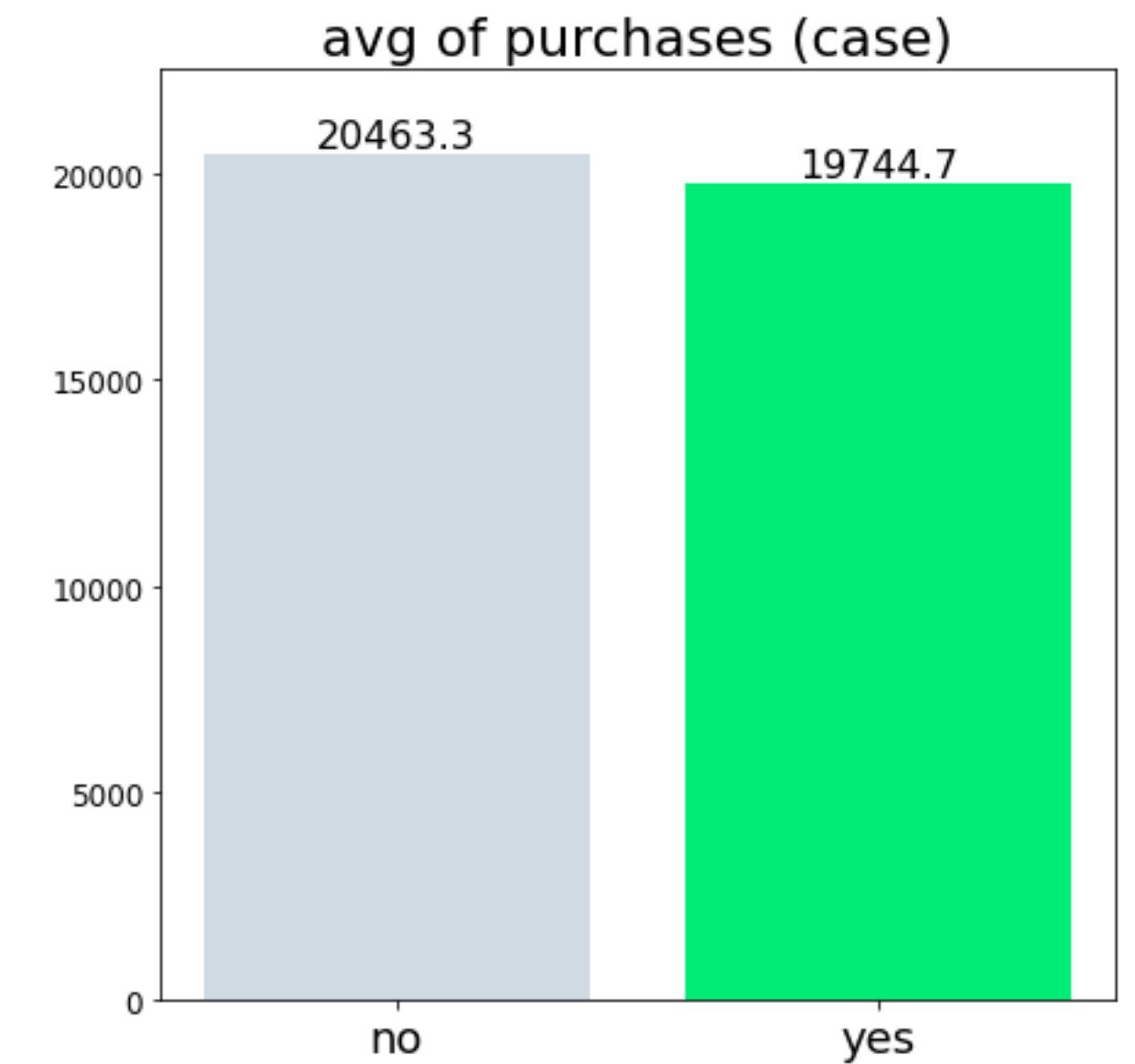
2. 거래별 분포



<거래별 외국인 분포>

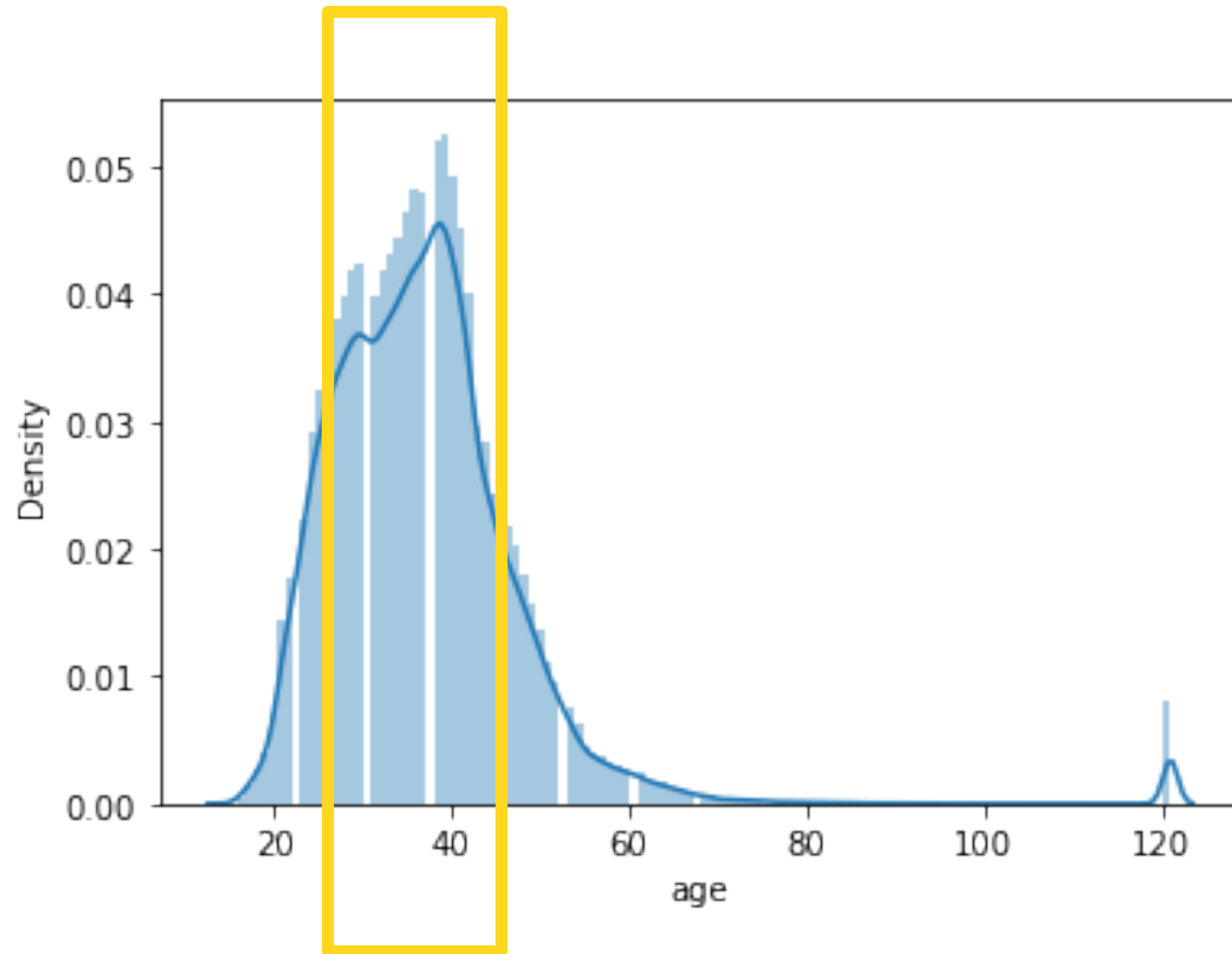


<거래별 총 구매액>



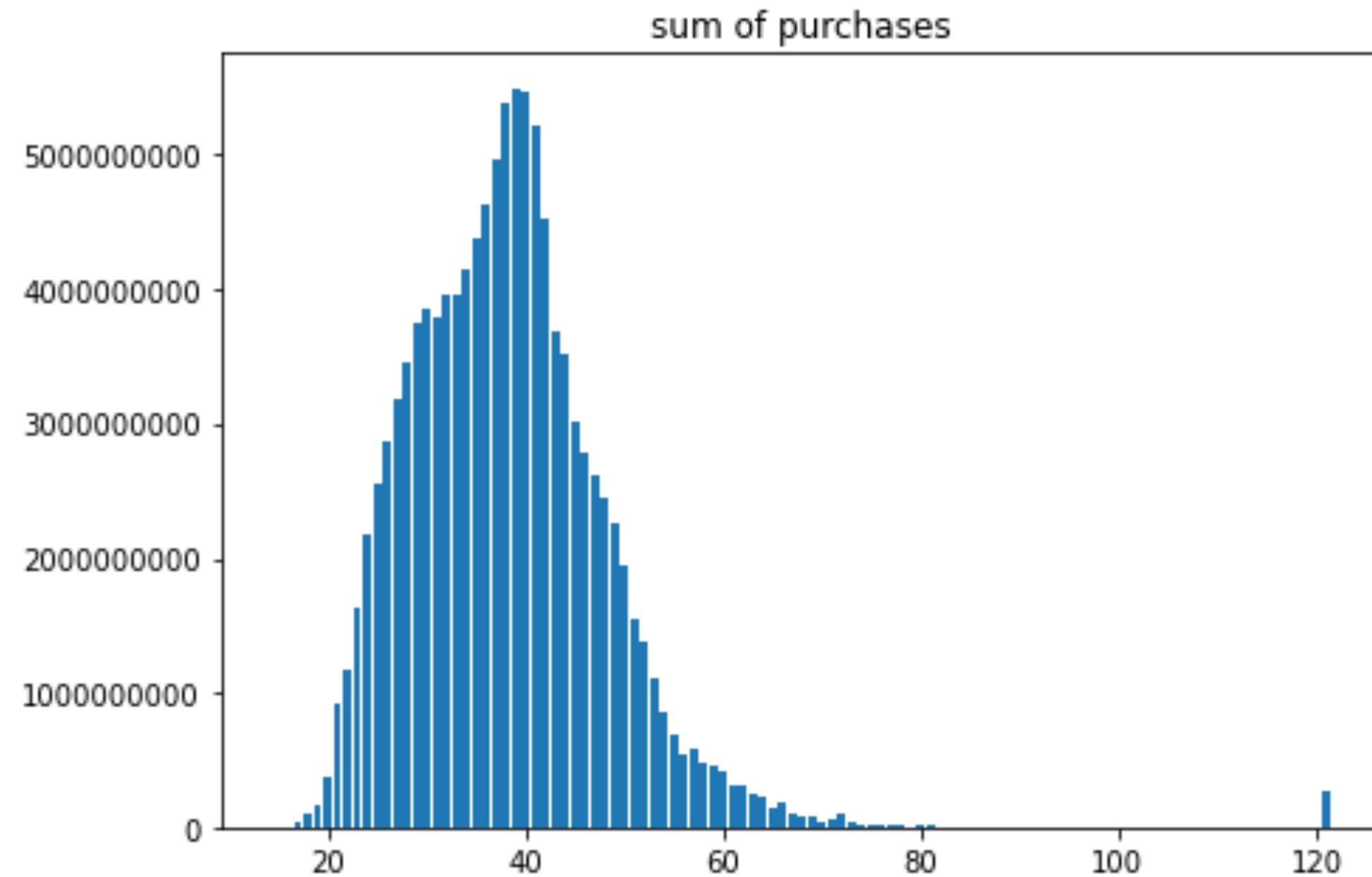
<거래별 구매 금액 평균>

유저 분포 : 연령대 기준

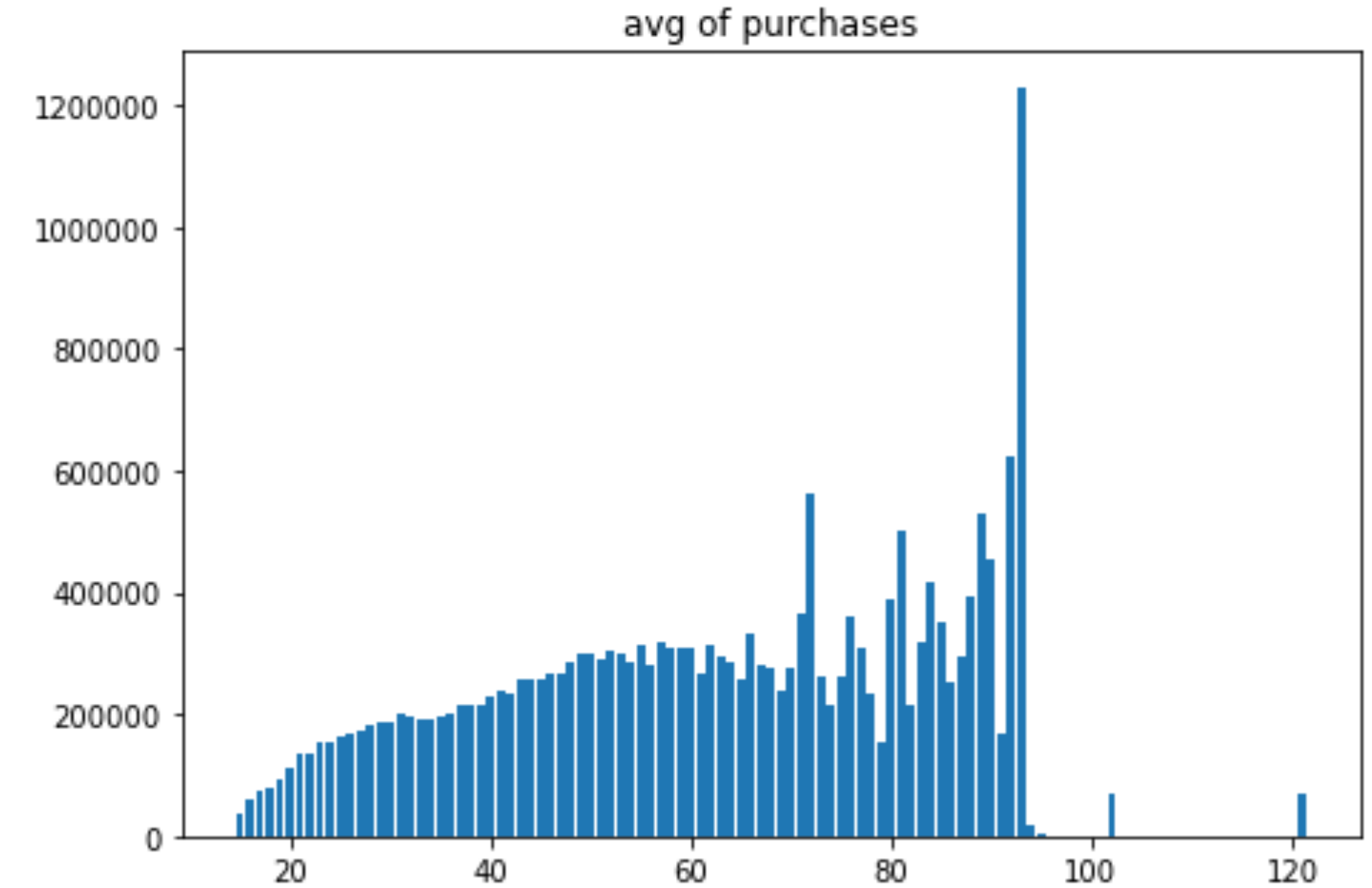


- 평균 이용자 연령은 약 25-48세 추정
- 사용자가 가장 많은 나이는 39세
- 가장 적은 나이는 93세, 102세
- 1900년대생(121세) 데이터는 3902명
 - 생일값만 Null인 것을 확인

유저 분포 : 연령대 기준



- 총 결제 금액도 39세가 가장 많다.

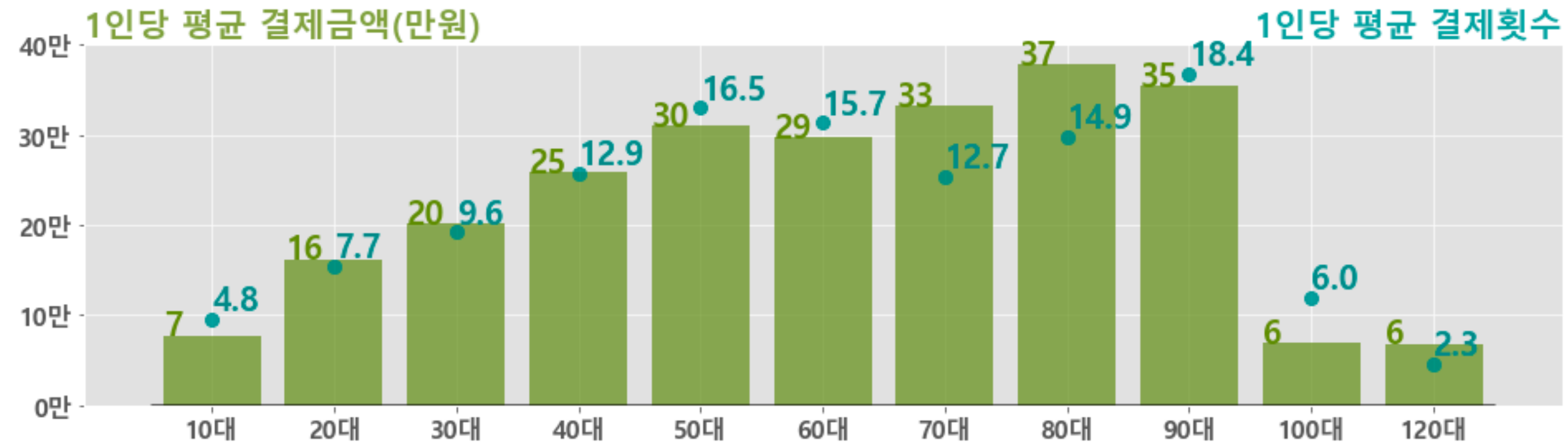


- 평균 결제 금액은 93세가 가장 높다. (고객 1명)

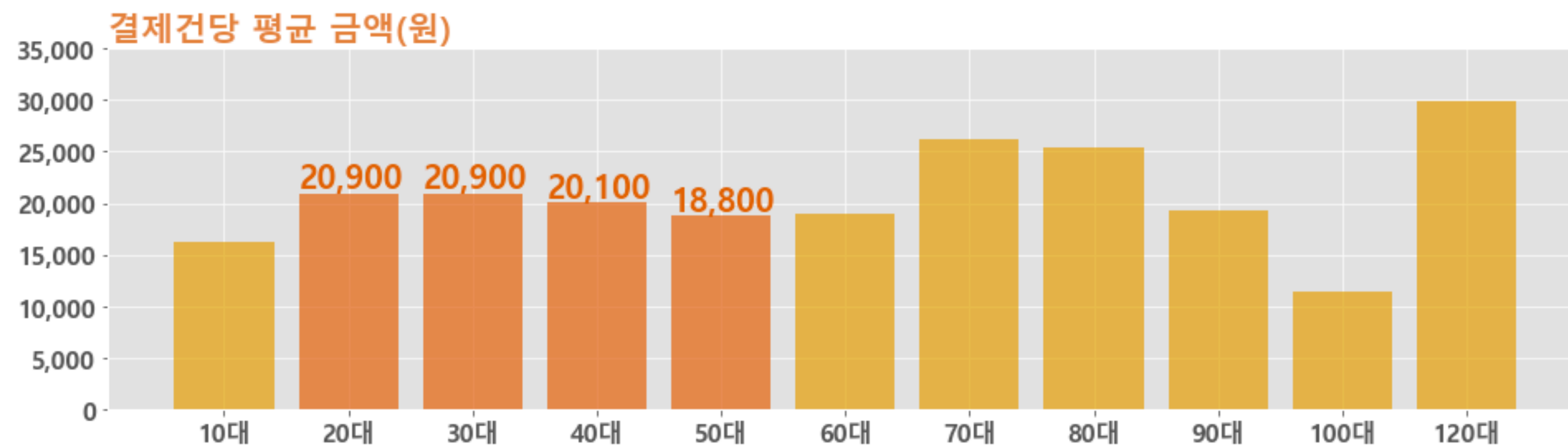


평균 결제 금액은 **나이가 많을수록** 함께 많아진다.

유저 분포 : 연령대 기준

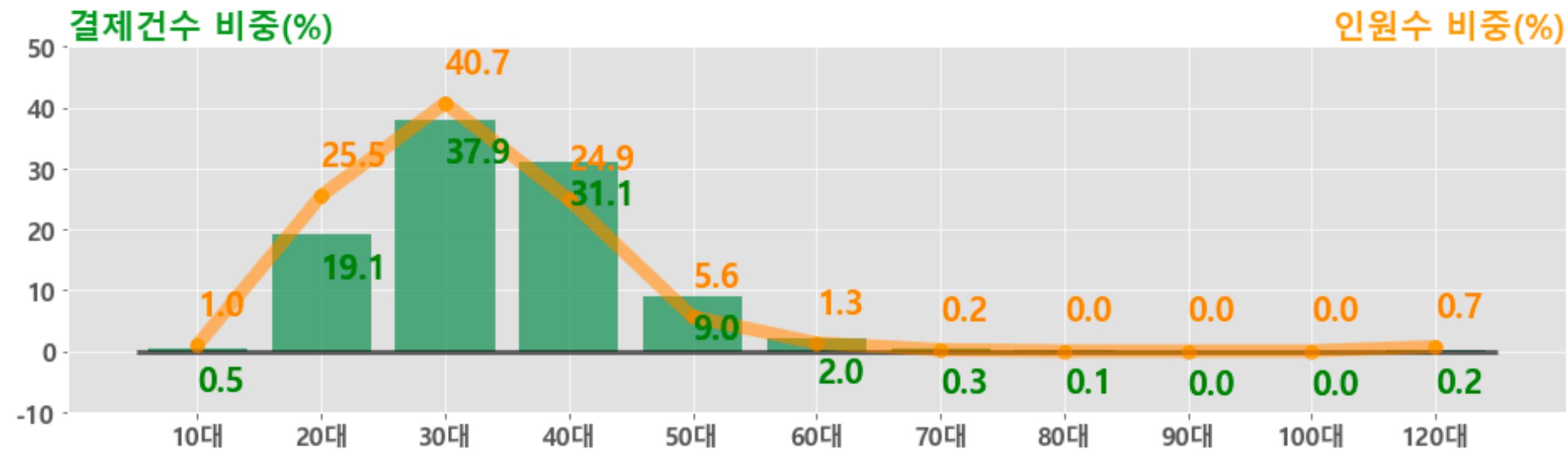


<1인당 평균 결제금액 및 결제횟수>

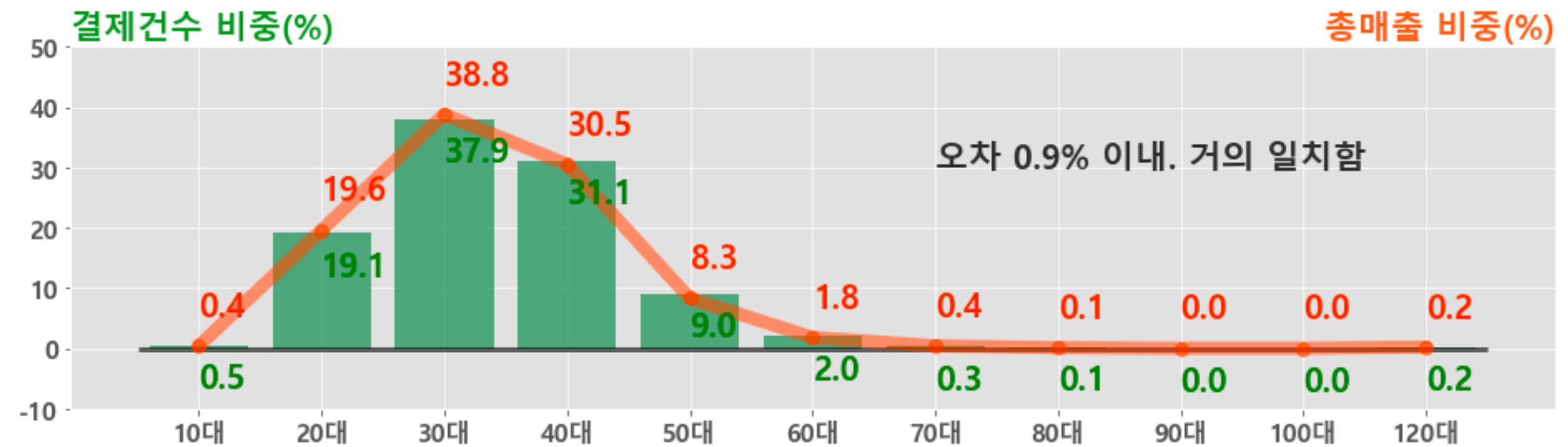


<결제 건당 평균 금액>

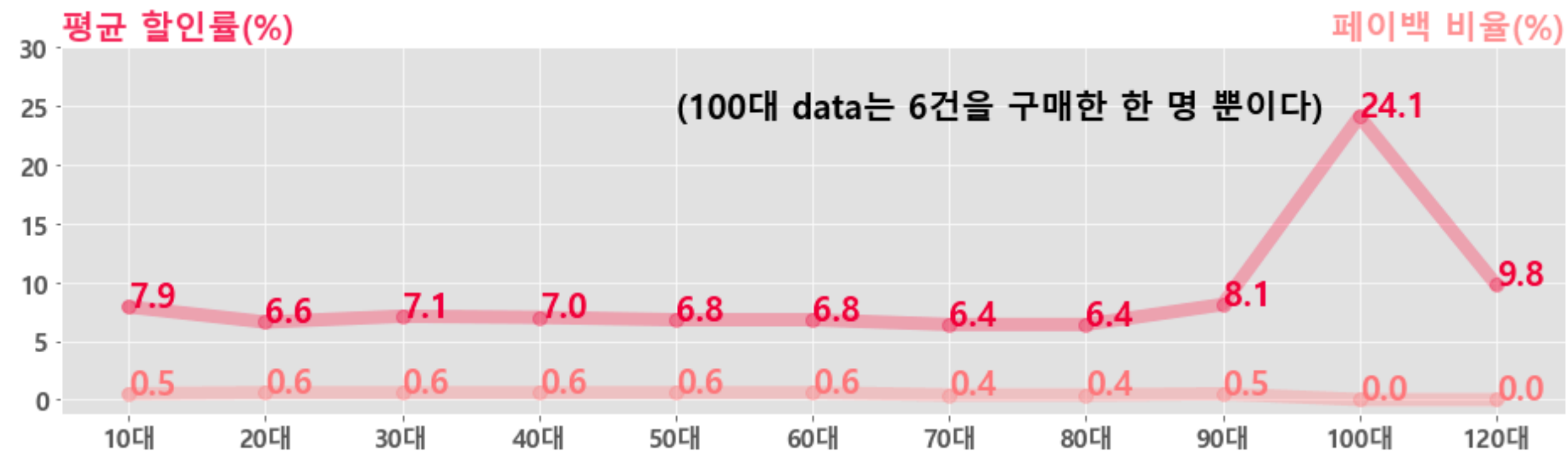
기본 분석 column별 탐색 : 회원 번호(customer_id)



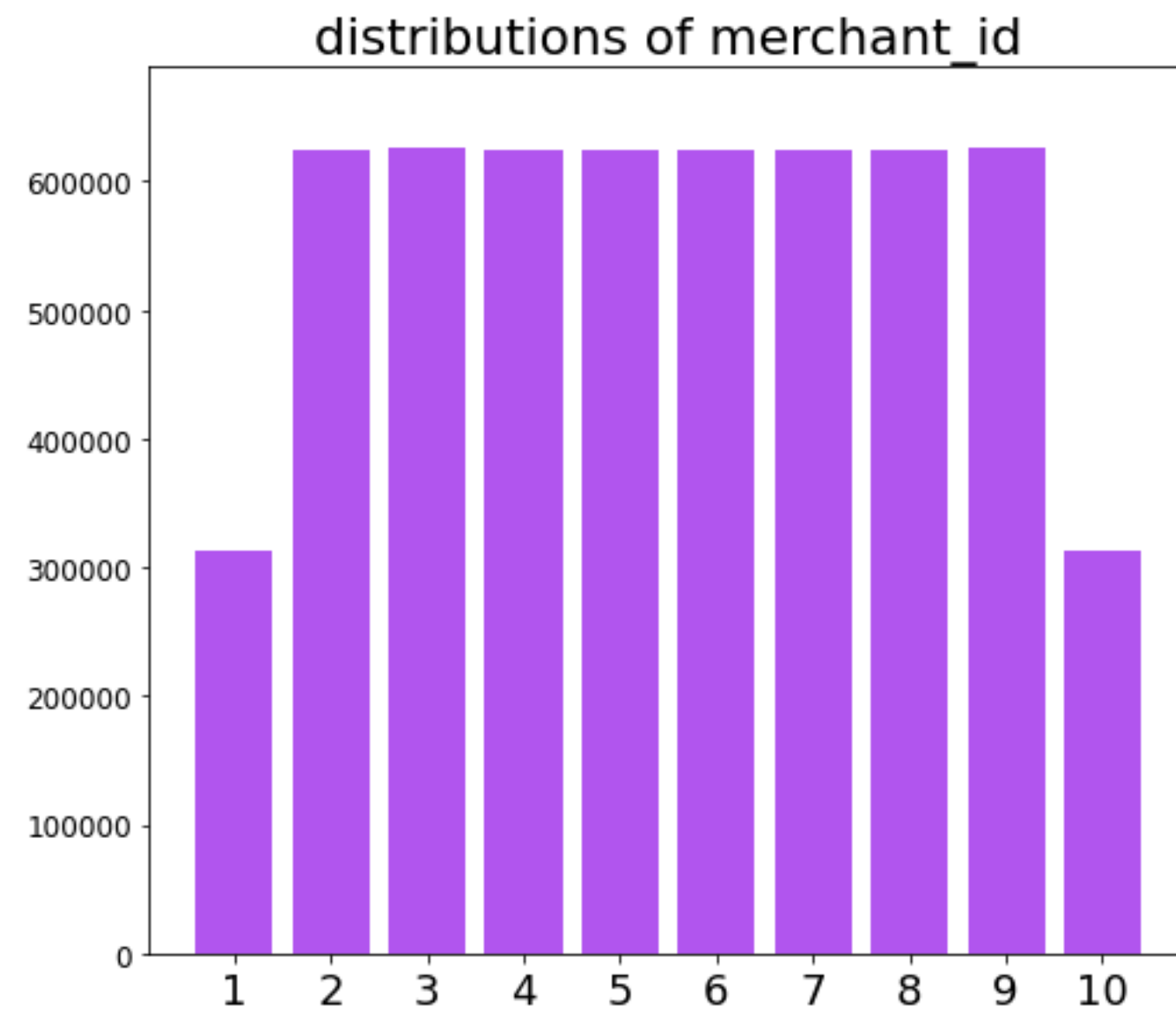
<연령별 결제 건수 및 인원수 비중>



<연령별 결제 건수 및 총매출 비중>

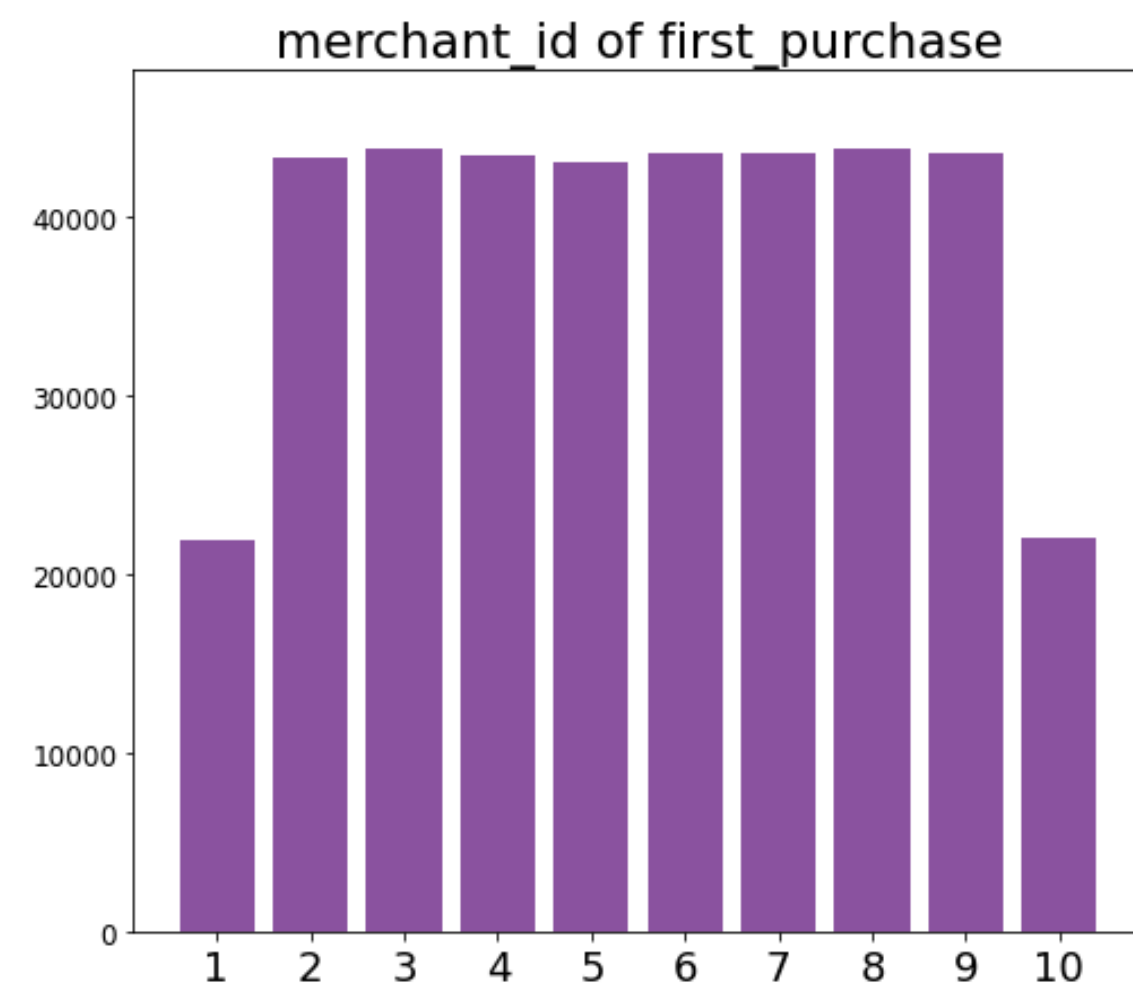


<연령별 평균 할인율 및 페이백 비율>

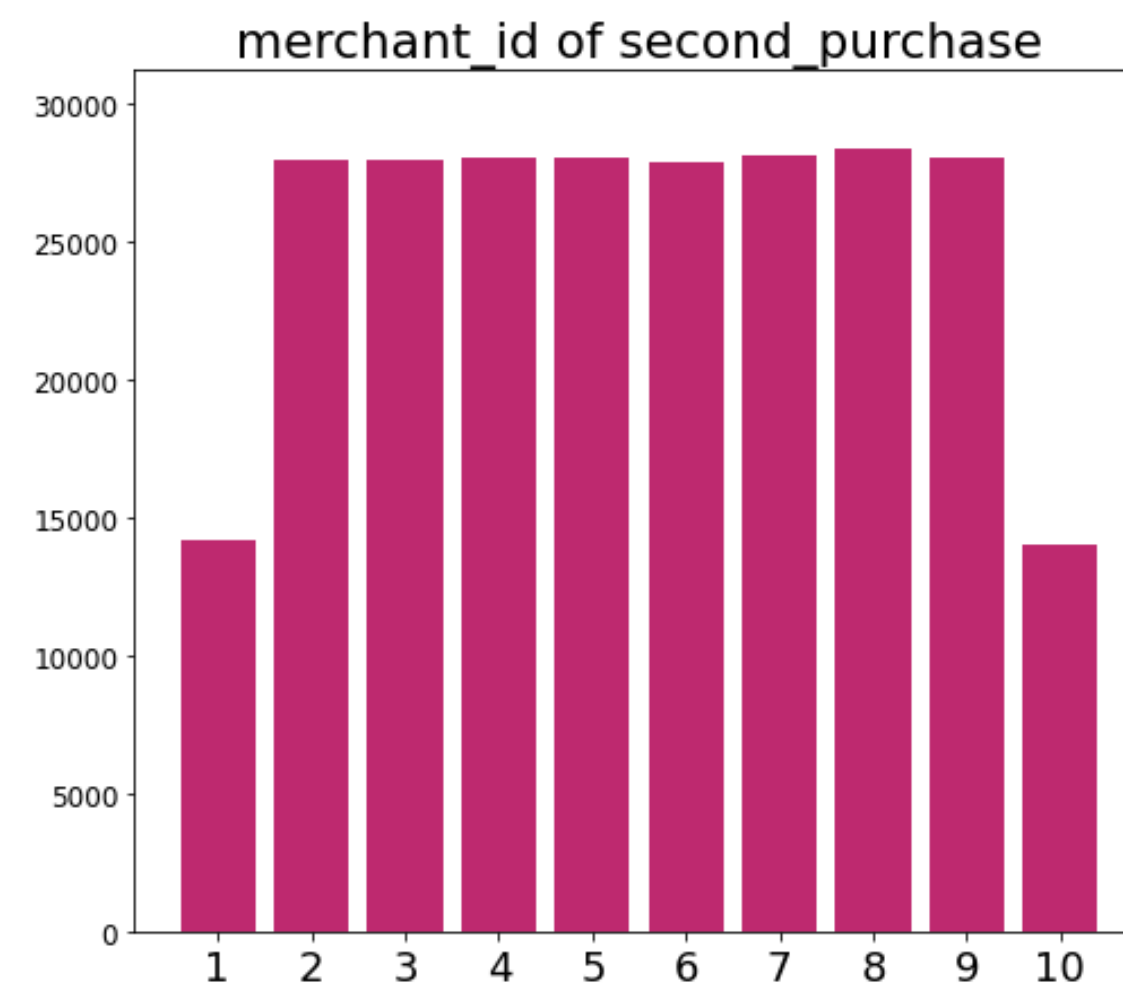


- merchant_id의 거래별 분포가 고르게 나타난다.
- 1번, 10번의 수를 합치면 다른 코드와 비슷한 비율이 된다.
- 이런 분포가 나타날 수 있는 상황
 1. 실제로 모든 사람들이 고르게 구매할 경우
 2. 가공된 데이터일 경우

merchant_id 분포 : 신규 유저 거래 기준



<첫번째 거래의 분포>

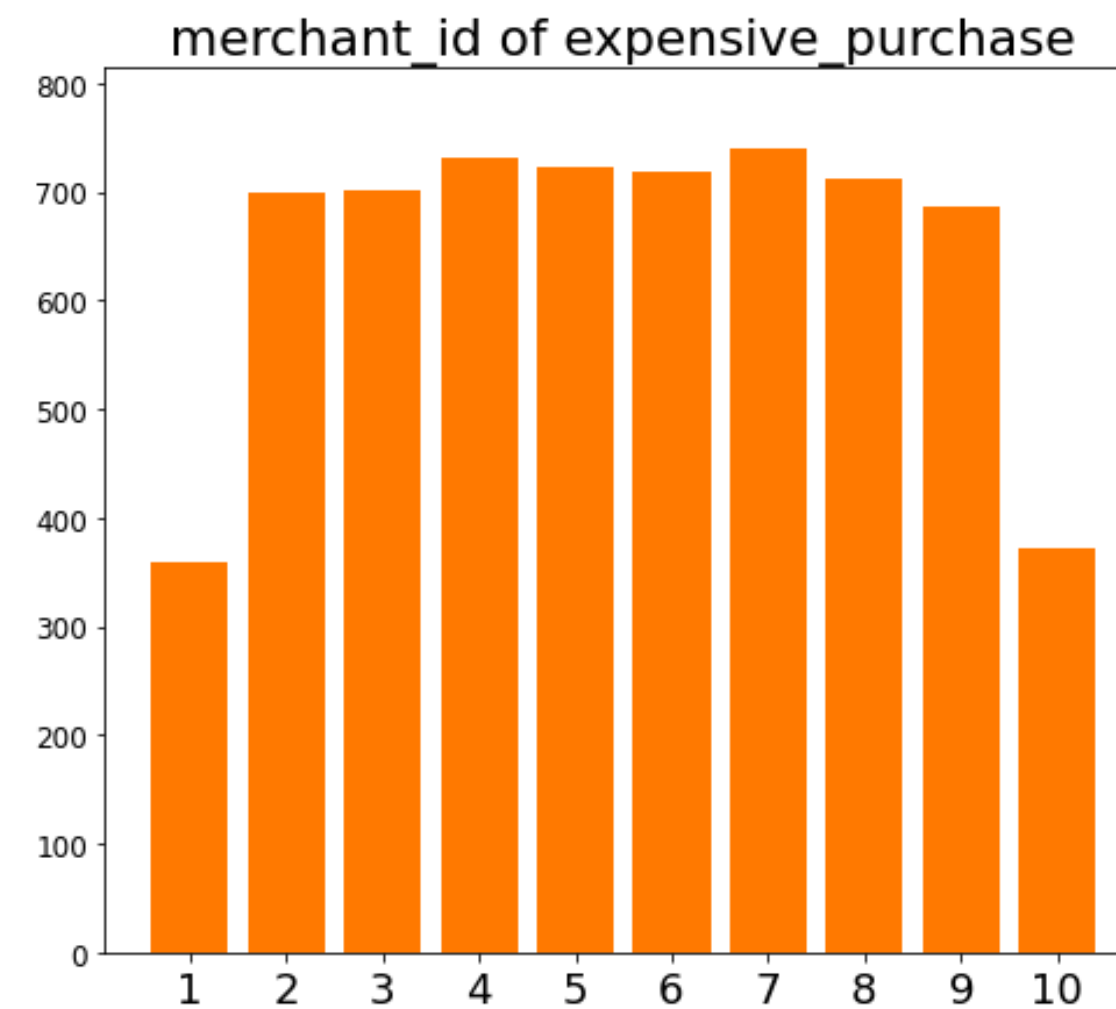


<두번째 거래의 분포>

- 두 거래 모두 매우 유사한 분포를 띈다
- 8월 이후 신규 유저의 수는 약 39만명
- 이들의 첫번째, 두번째 거래의 분포가
완전 비슷하게 나타나는 것은 정상으로 보기 어렵다

merchant_id 분포 : 50만원 이상(고가) 거래 유저 기준

<고가 거래 유저의 merchant_id 분포>



- 고가 거래도 이전과 마찬가지로 분포가 고르게 나타난다



결론, merchant_id는 **가공된 데이터**로 보임

예상하는 이유는? **보안** 유지

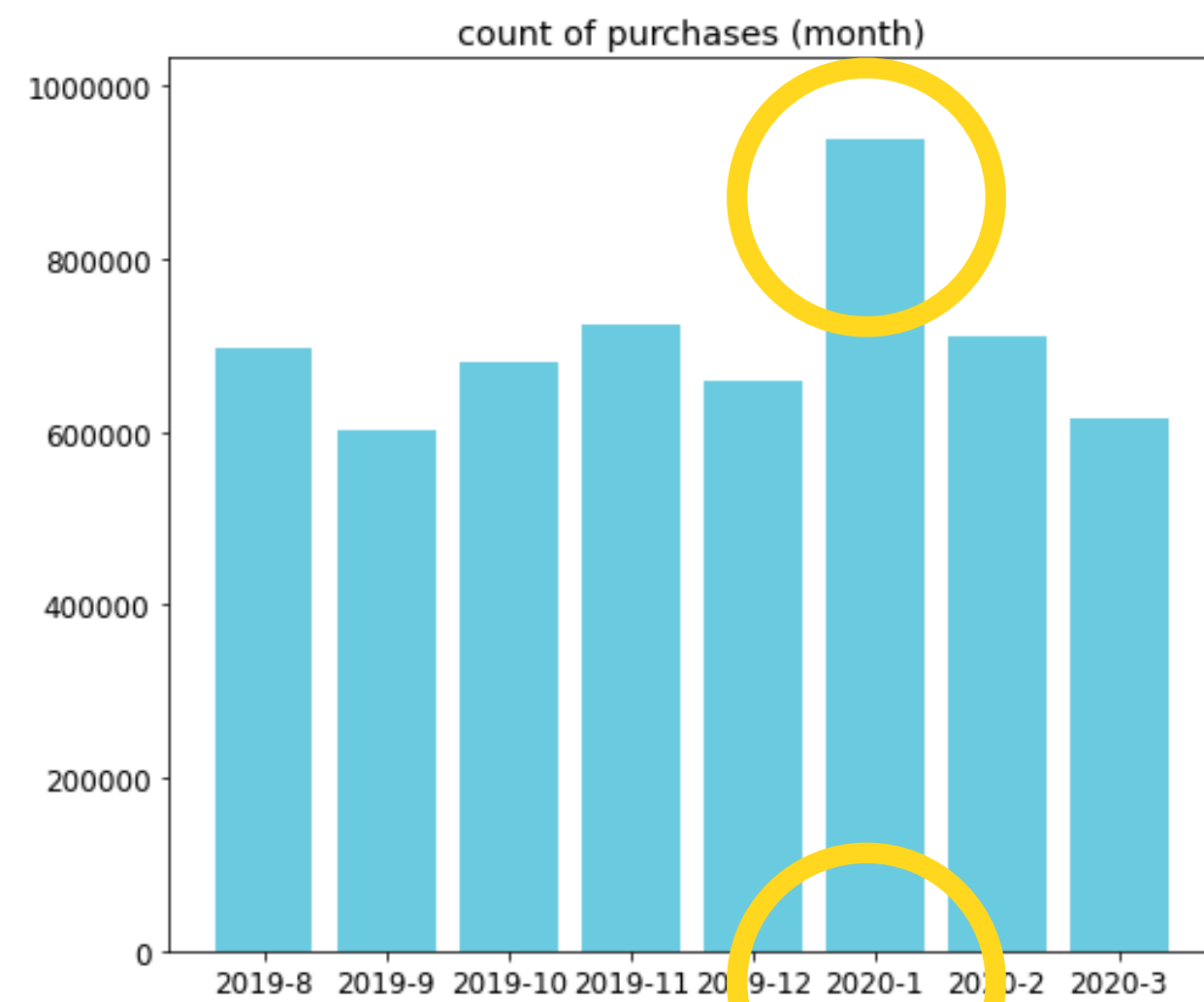


산업 코드는 협력사 및 차이의 매출 및 실적을 나타내는 매우 중요한 지표

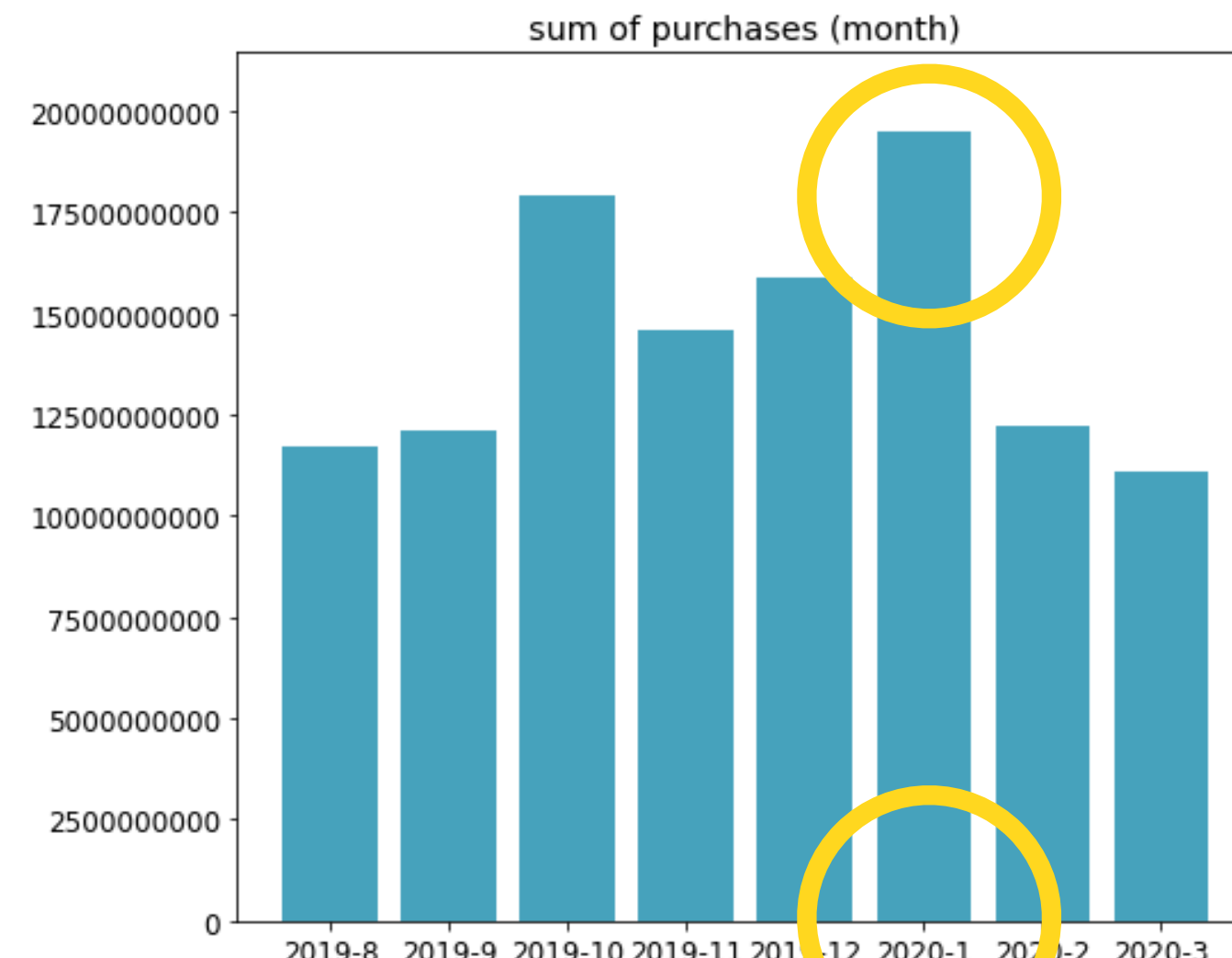
보안 상의 이유로 마스킹 처리를 하는 과정 중 고의적으로 난수(Random) 처리를 한 것으로 예상

다만 해당 데이터를 마스킹 및 난수 처리를 함으로 인해 이를 활용한 분석 등은 어려워보인다

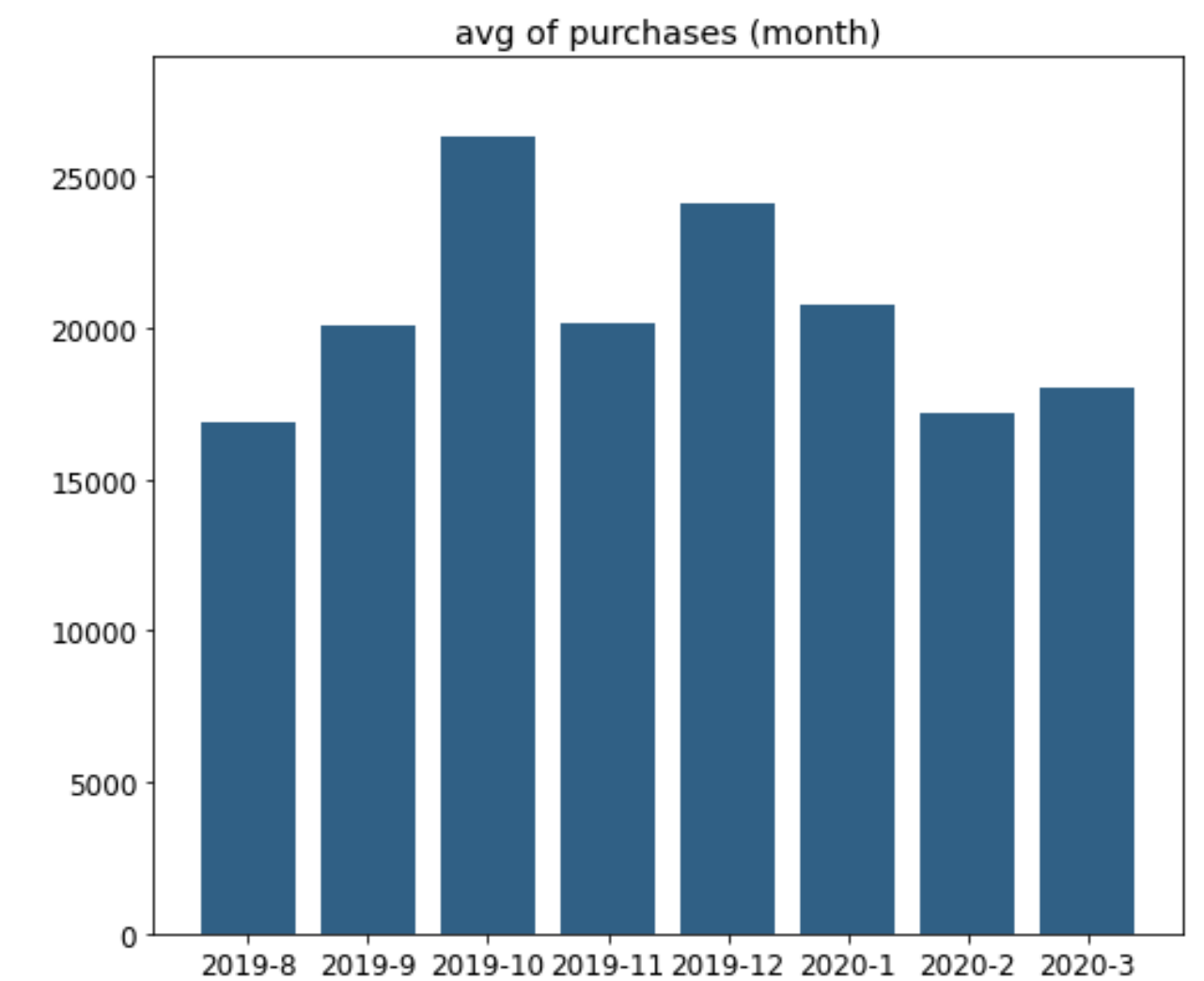
기본 분석 column별 탐색 : 날짜 (월 단위)



<월별 구매 횟수>



<월별 총 매출>

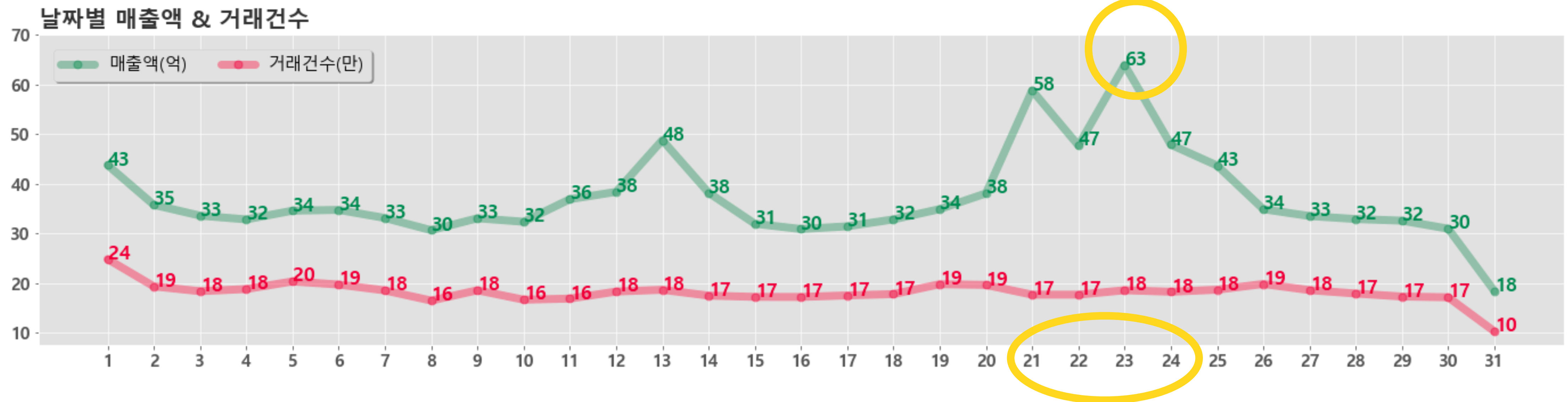


<월별 평균 구매 금액>



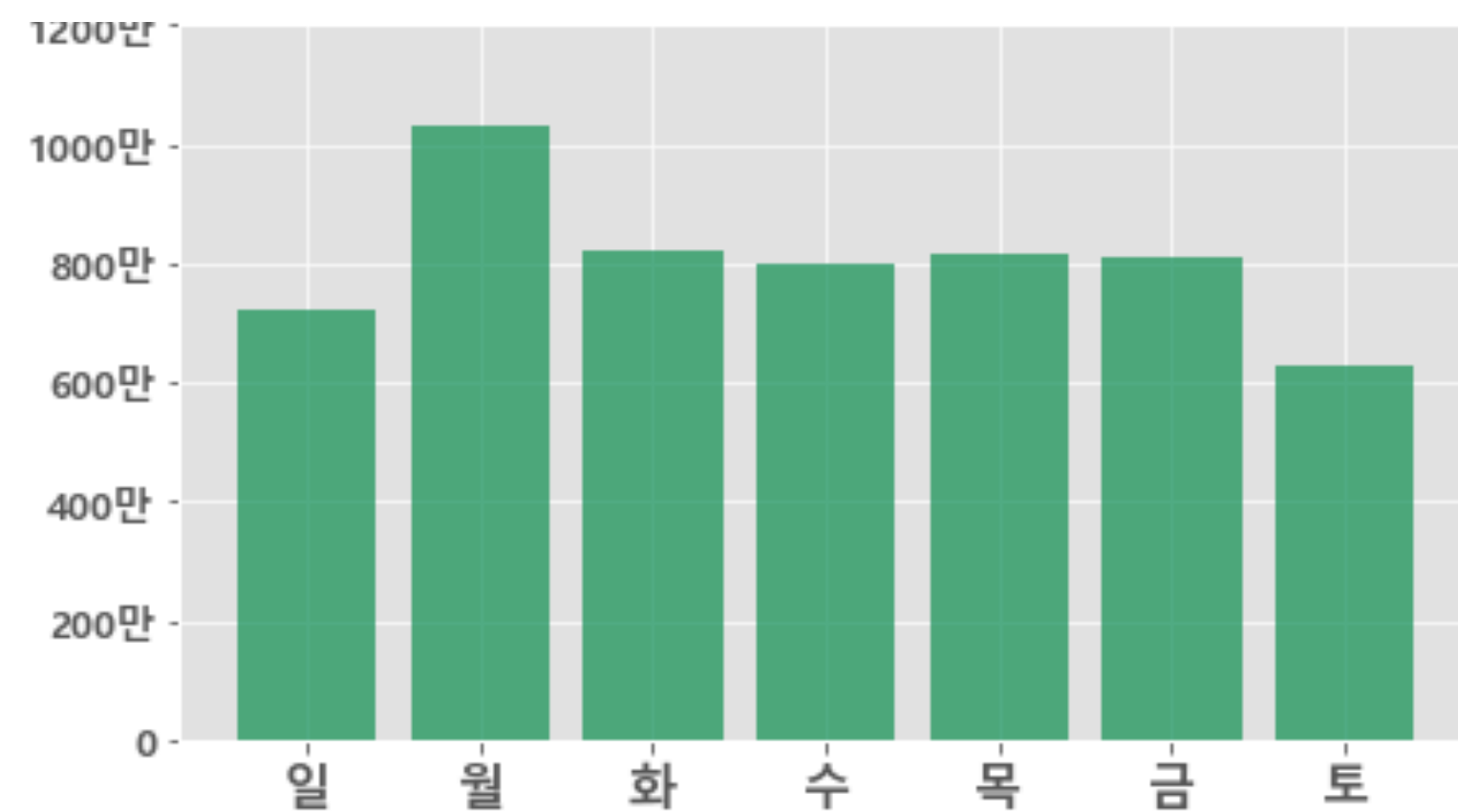
구매 횟수와 총 매출이 가장 많은 달은 **1월**

기본 분석 column별 탐색 : 날짜 (일 단위)

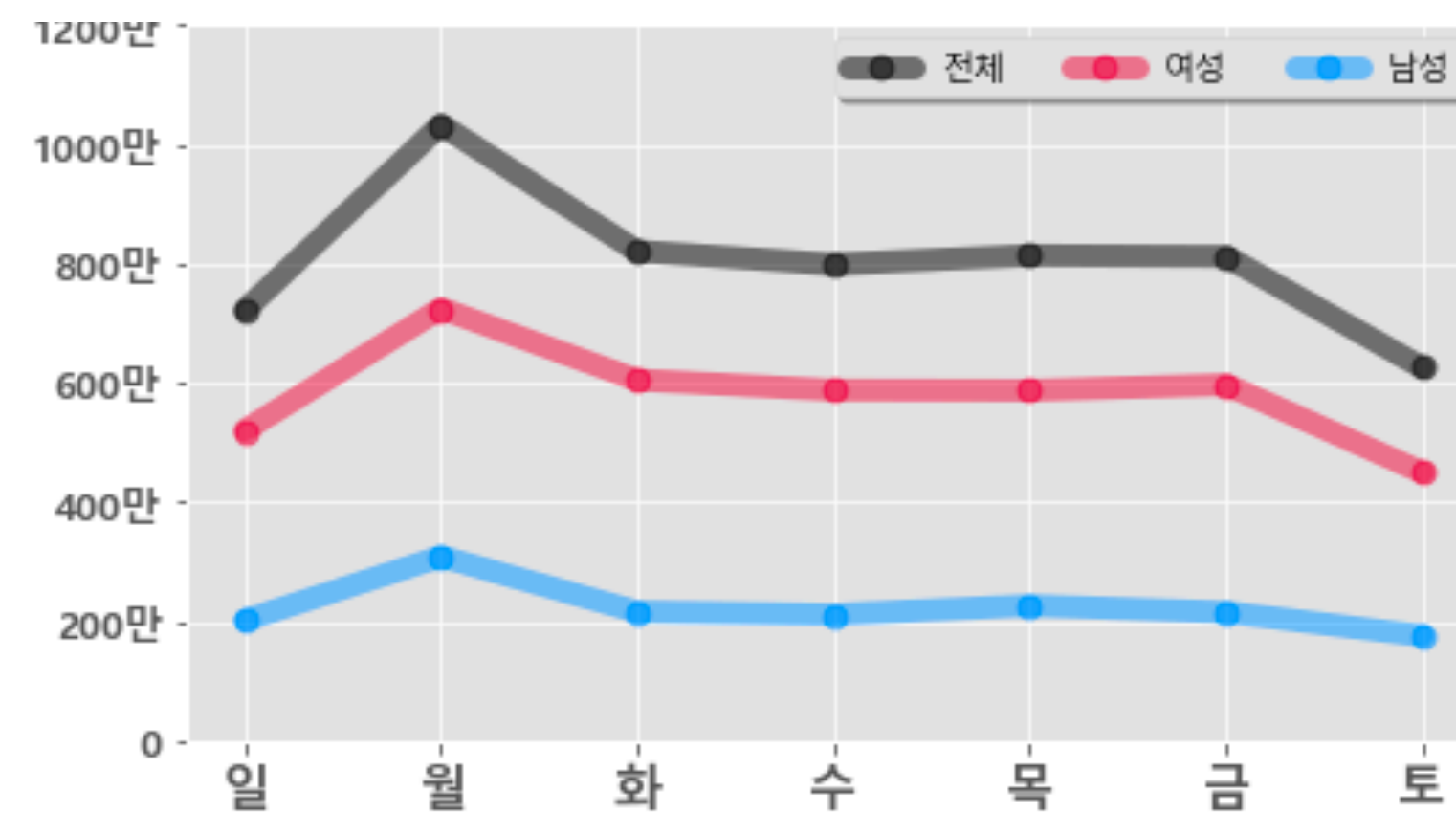


<날짜별 매출액 및 거래 건수>

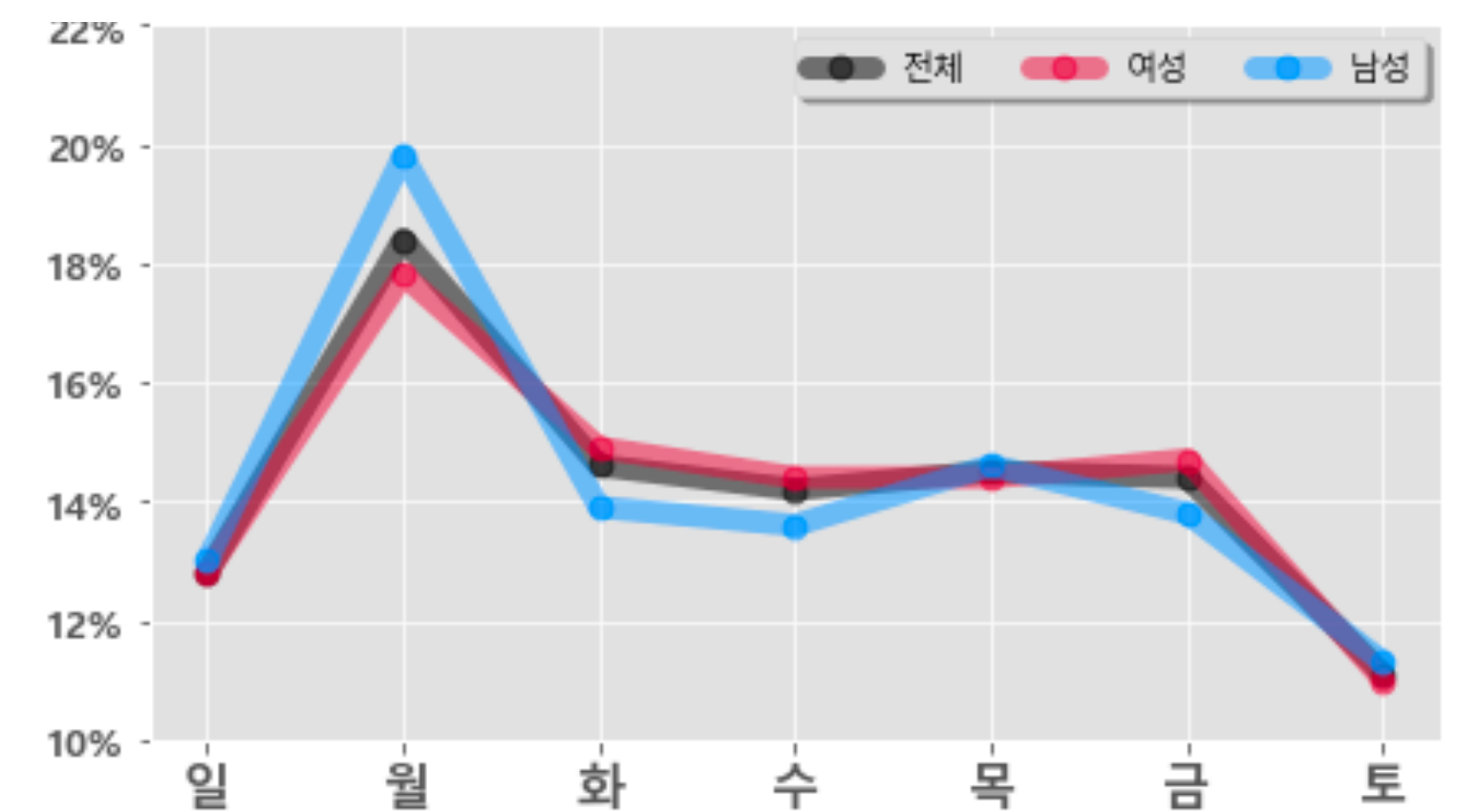
기본 분석 column별 탐색 : 날짜 (요일 단위)



<요일별 거래 건수>



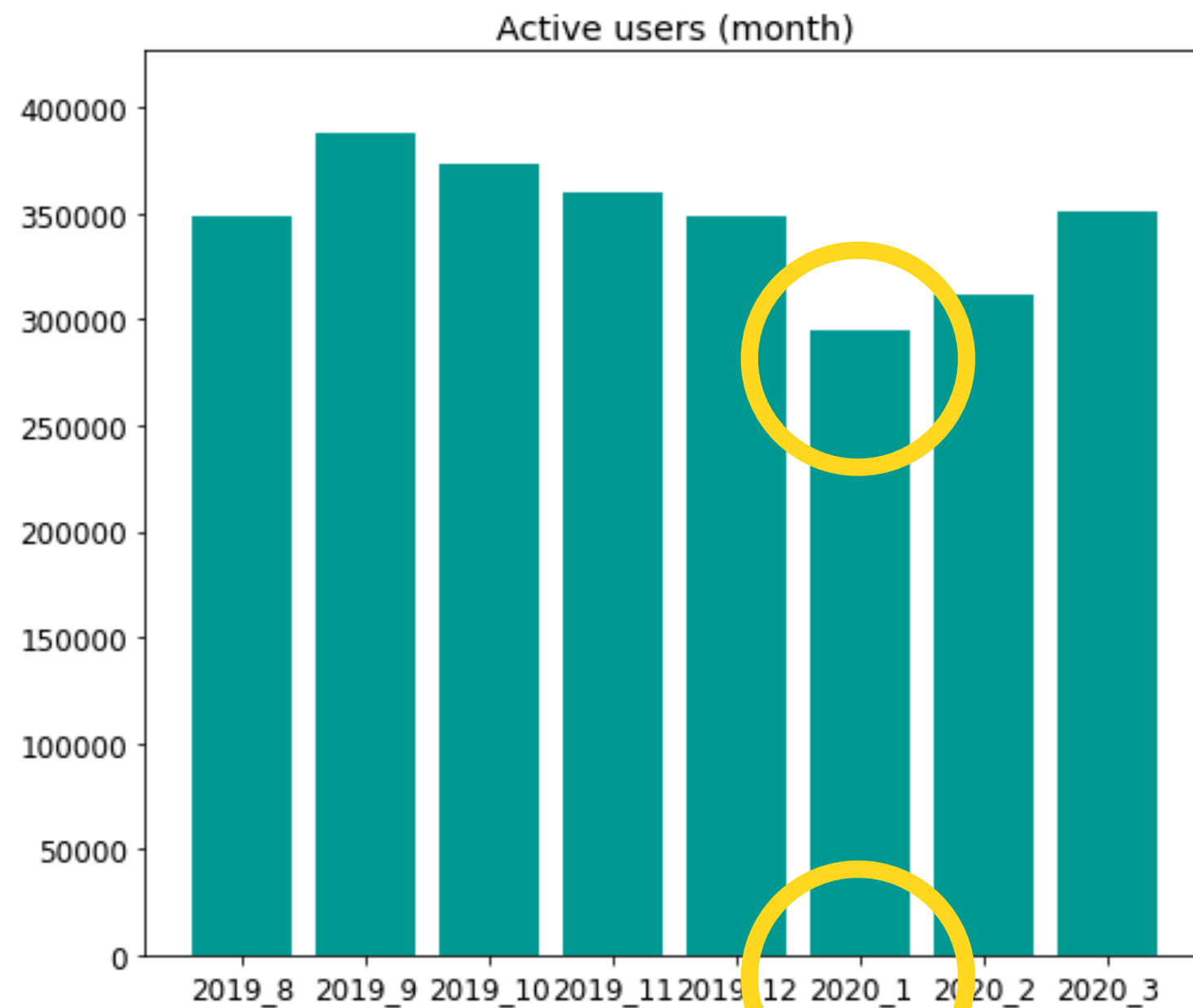
<성별 간 요일 거래 건수>



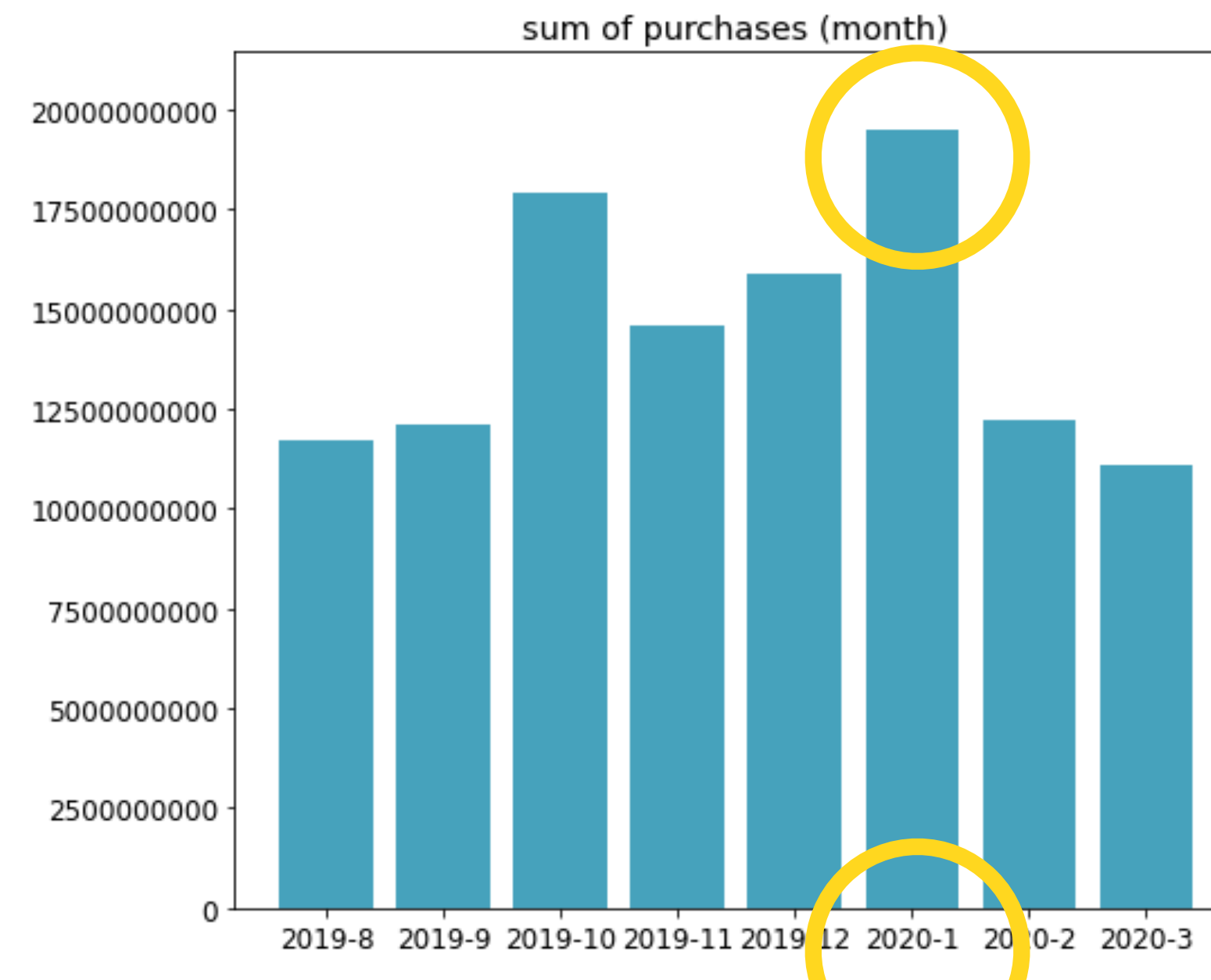
<성별 간 요일 거래 건수 비중>

심화 분석 ⚡

- 활동 유저의 기준 : 각 월별로 거래가 1건 이상인 사용자



<월별 활동 유저의 수>



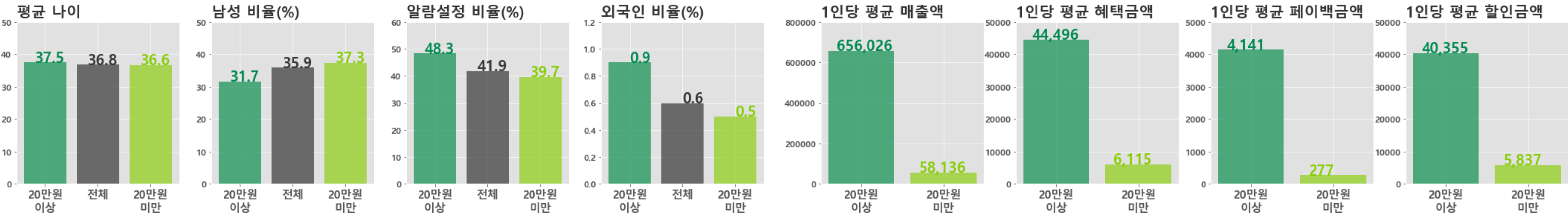
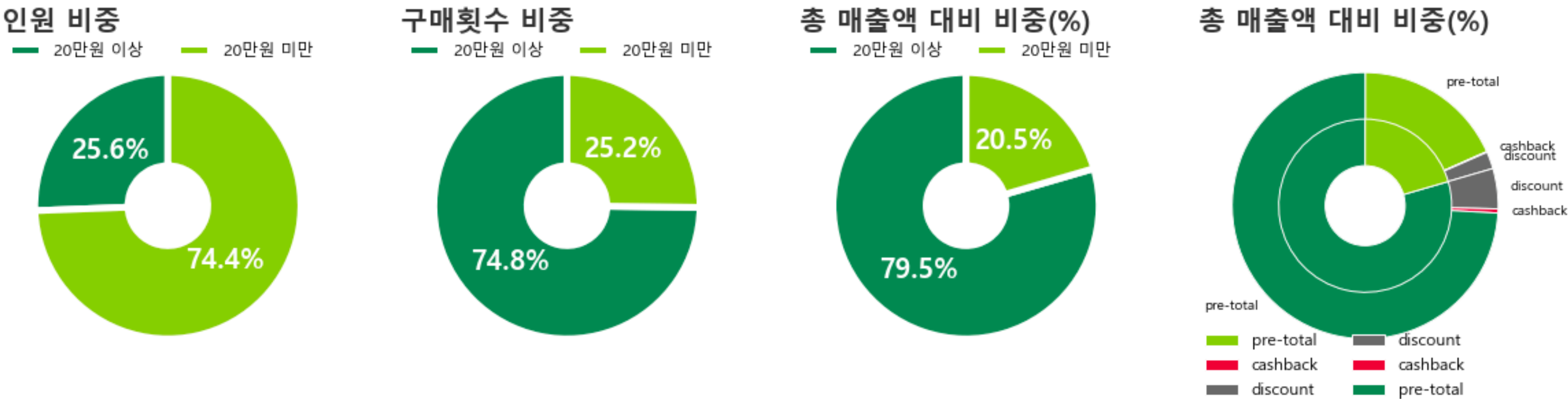
<월별 활동 유저의 거래 건수>

2020-1월은 활동자는 가장 적었지만 매출은 가장 높은 달



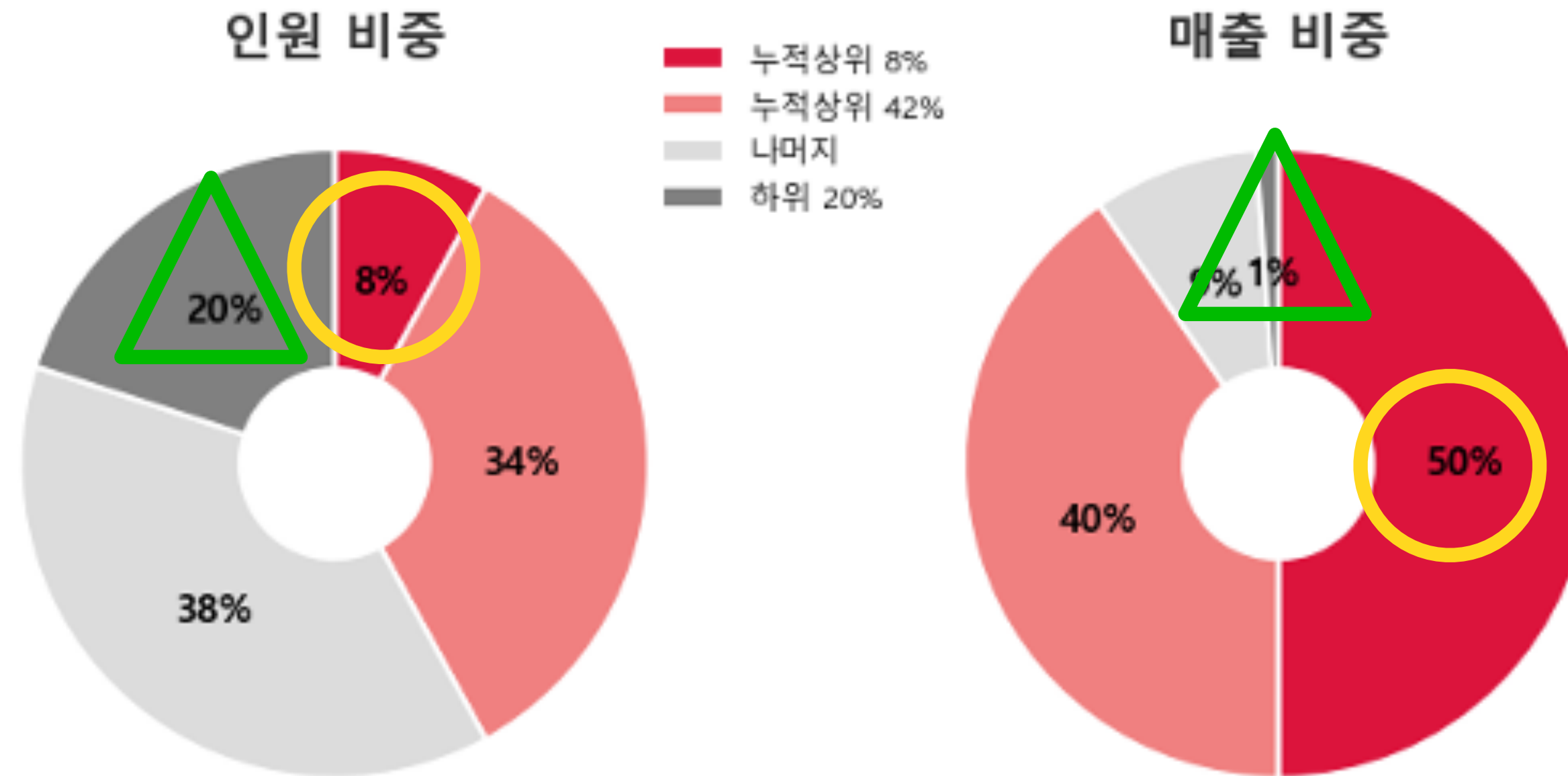
매출은 단순 활동 유저 수에 **비례하지 않음**

심화 분석 구간별 소비 금액 비율 차이 : 20만원으로 분류



매출 상위 구간과 하위 구간을 분류하여 자세하게 탐색할 필요가 있음

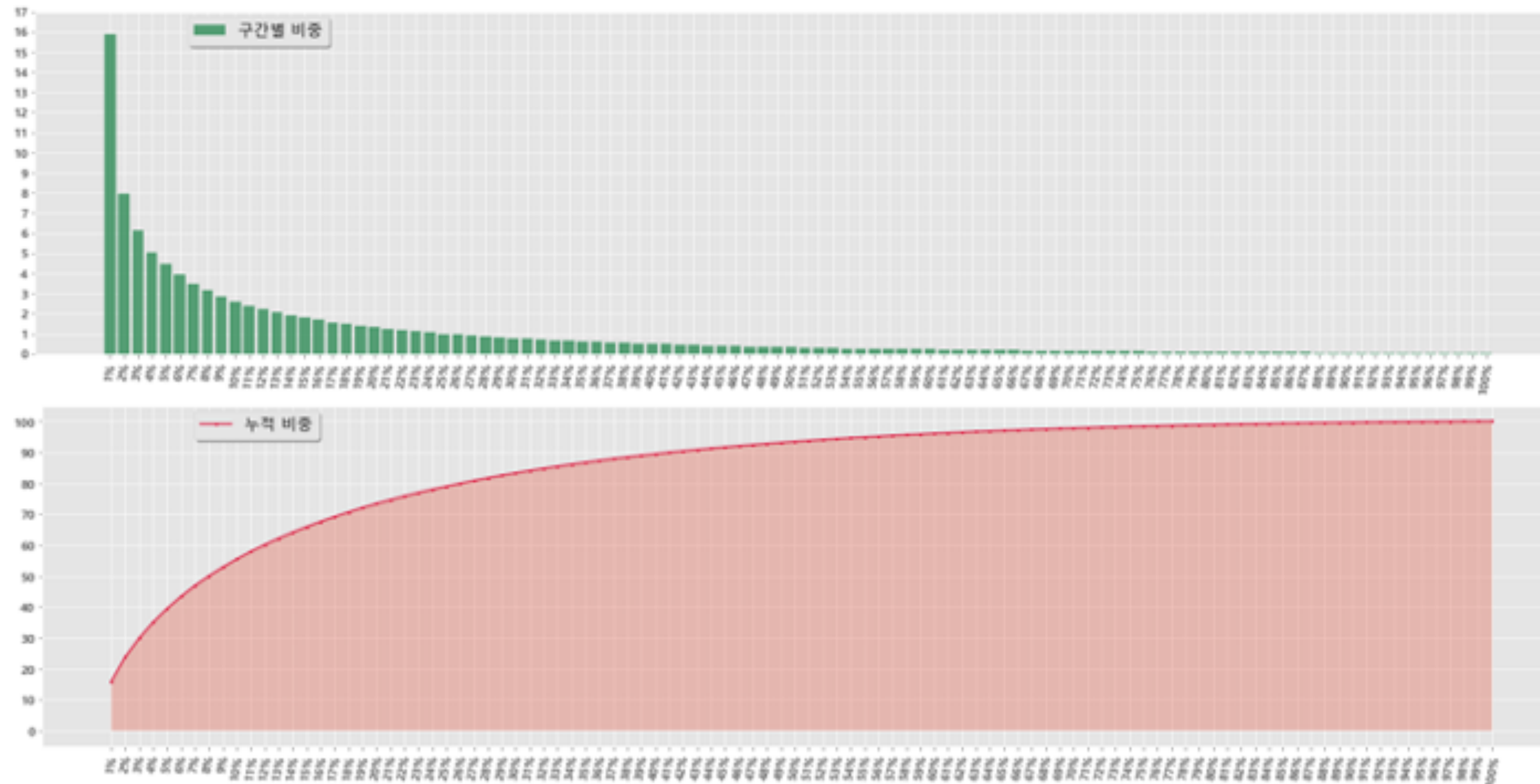
심화 분석 구간별 소비 금액 비율 차이 : 상·하위 그룹 분류



<인원과 매출 비중의 상관관계>

- 상위 8%가 전체 매출의 반 정도를 차지함
- 하위 20%는 전체 매출의 1% 밖에 되지 못함

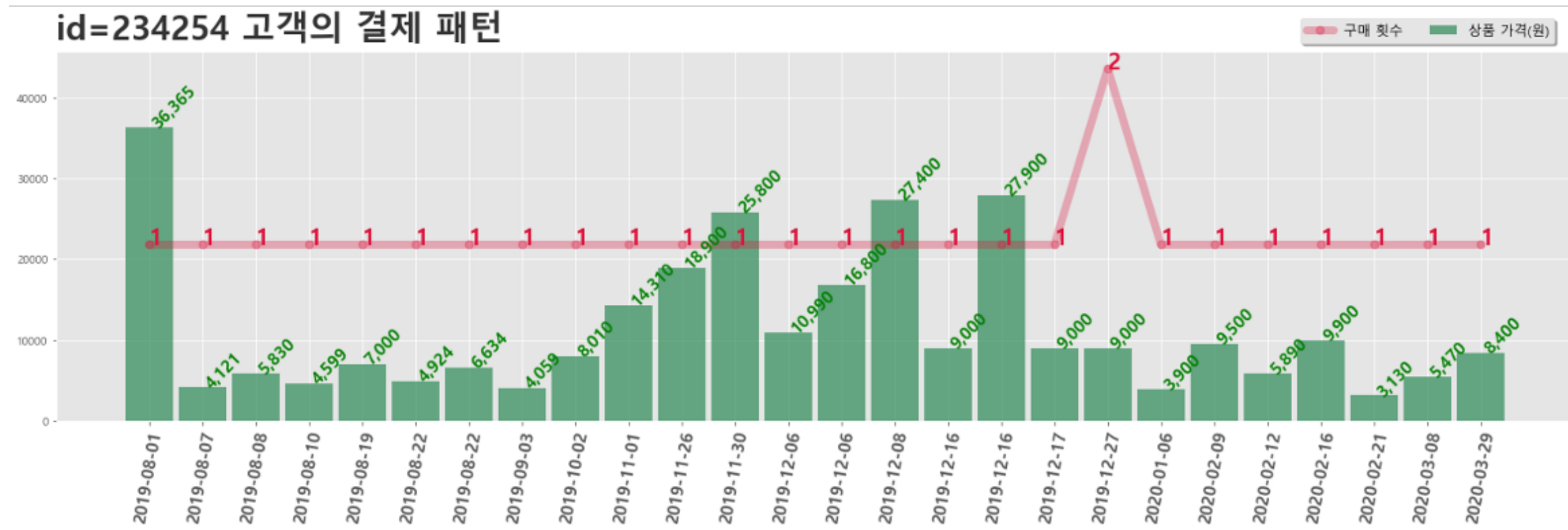
전체 매출액 기준 유저 세분화 : 1% 단위로 쪼개기



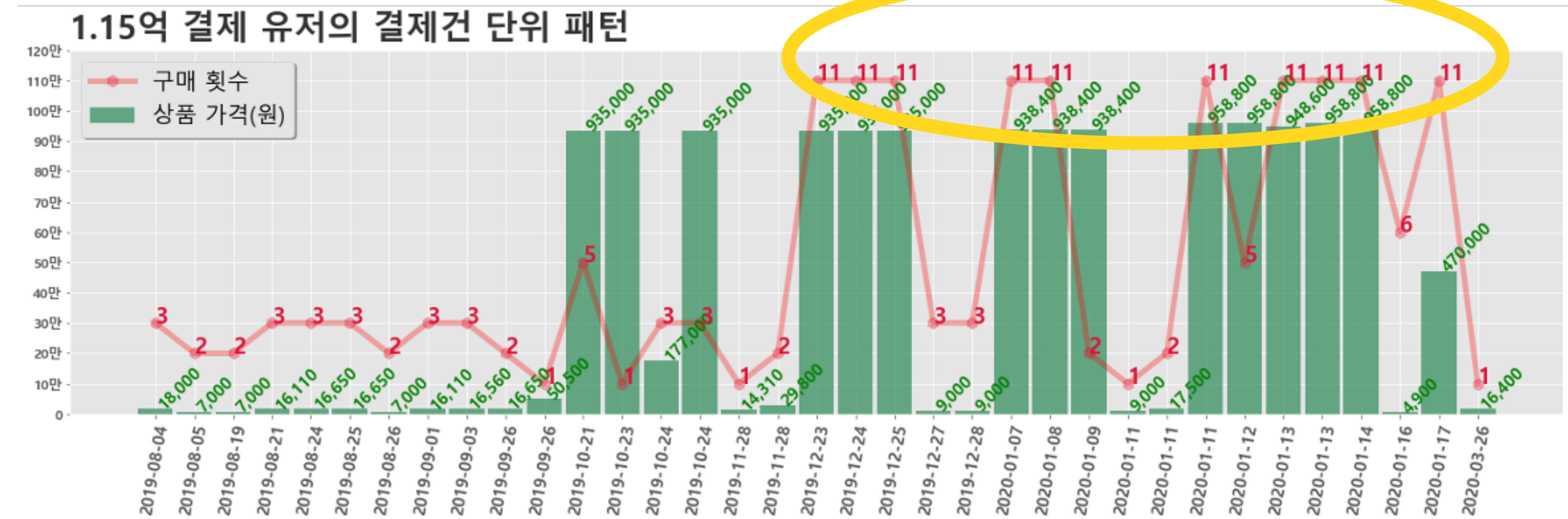
<매출액 기준 전체 고객의 결제금액 및 누적비중>

- 상위 구간이 매출의 대부분을 차지한다.
- 상위 그룹에 대한 조사 필요

심화 분석 상위 매출 고객 : 보따리상



<일반 유저 소비 패턴>



<특정 유저 소비 패턴>

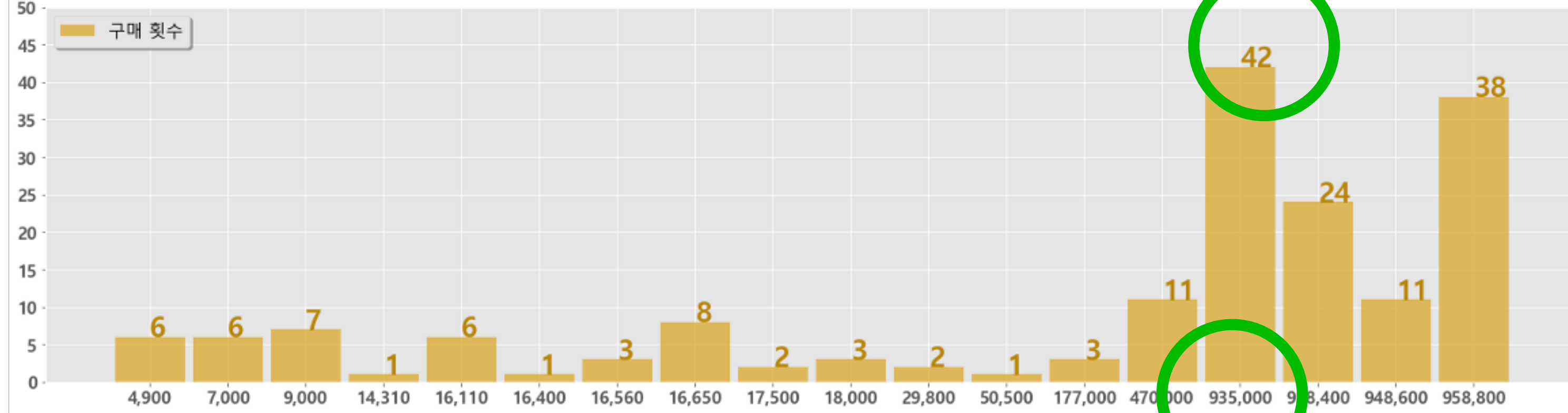
- 일반 유저들은 구매 금액대가 다양하고, 결제 횟수도 많지 않음

✓ 특정 유저에게 **동일** 가격의 제품을 **반복 구매**한 패턴이 존재

- 일반적인 소비 패턴과 다른 유저를 '보따리상'으로 정의하고 소비 패턴을 분석

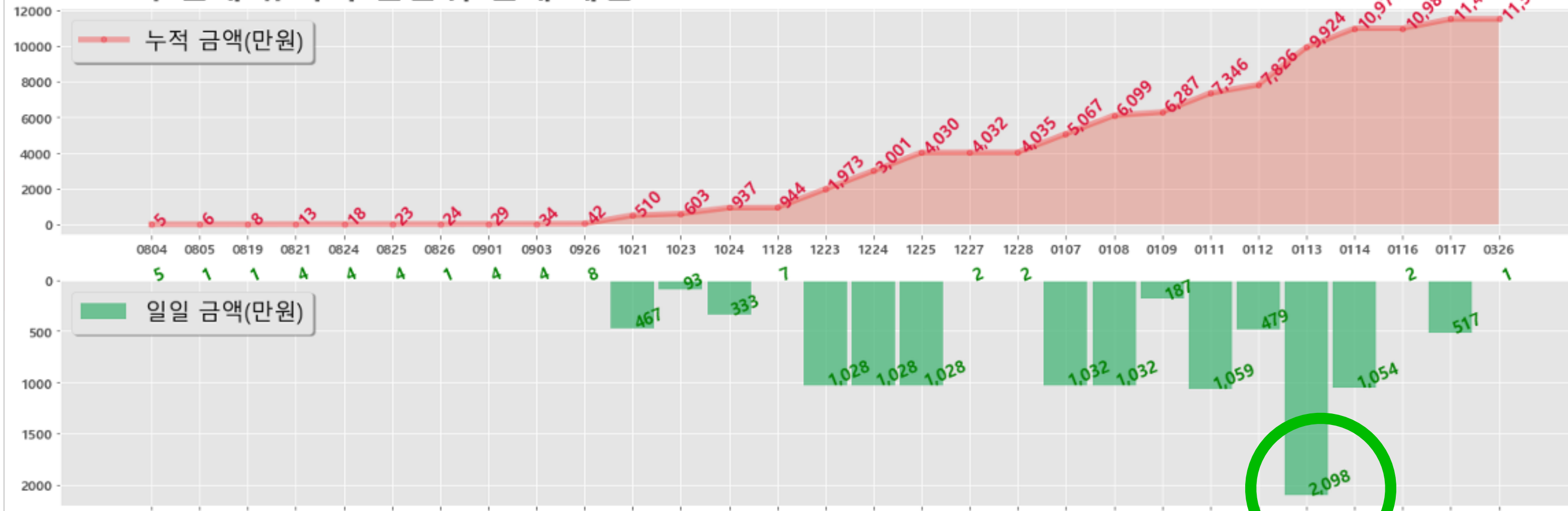
가장 많이 결제한 유저 데이터 : 1.15억

1.15억 결제 유저의 가격별 구매 횟수



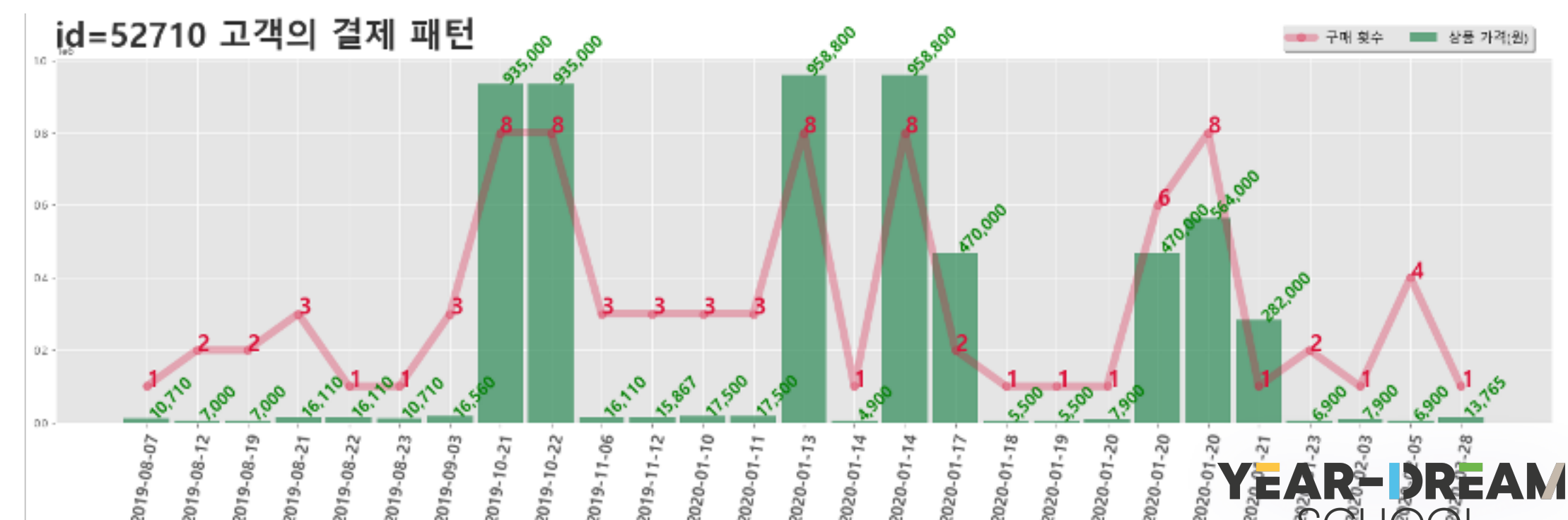
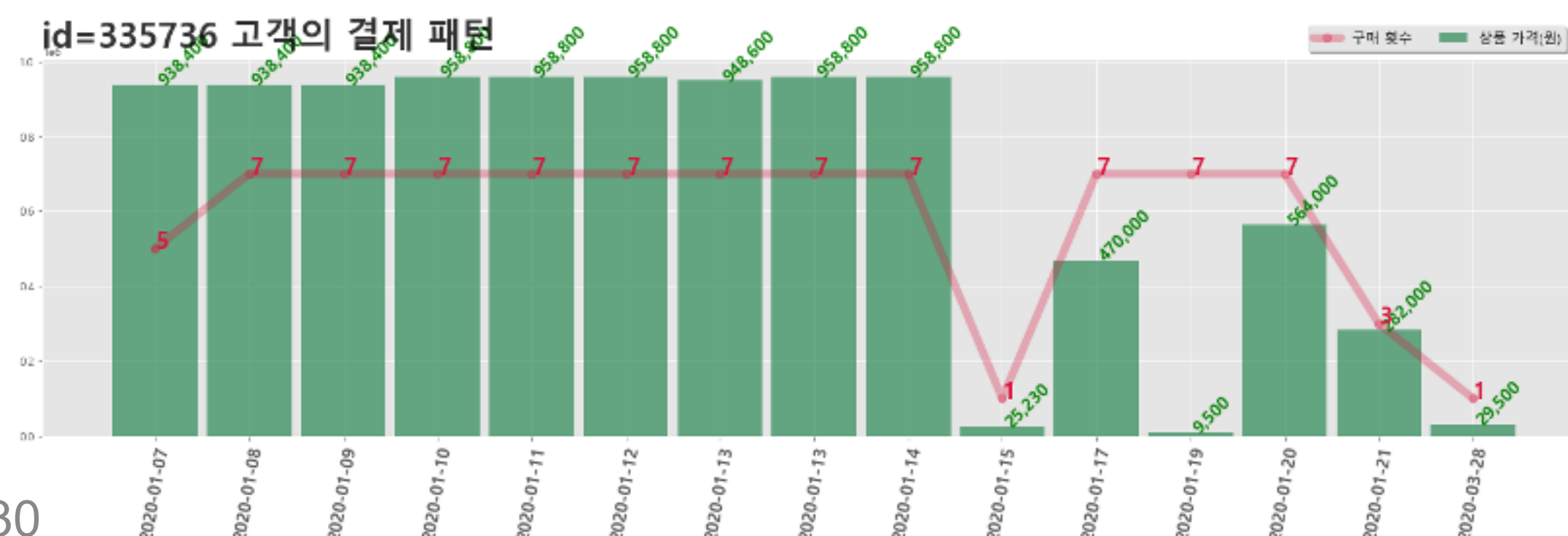
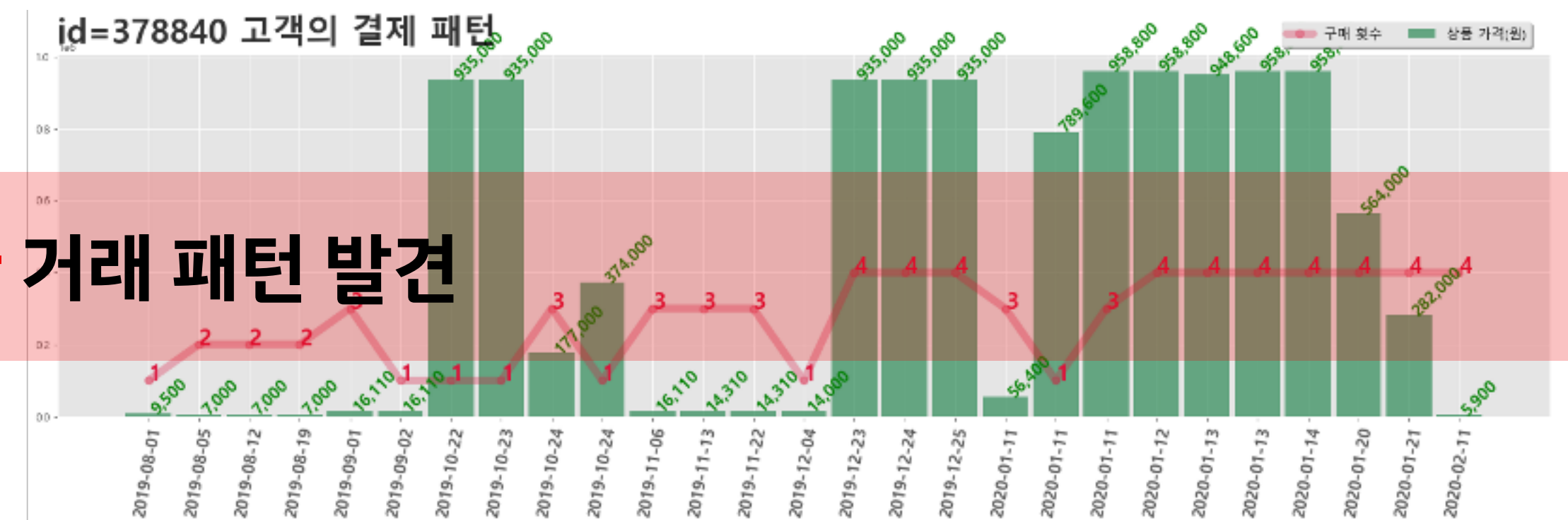
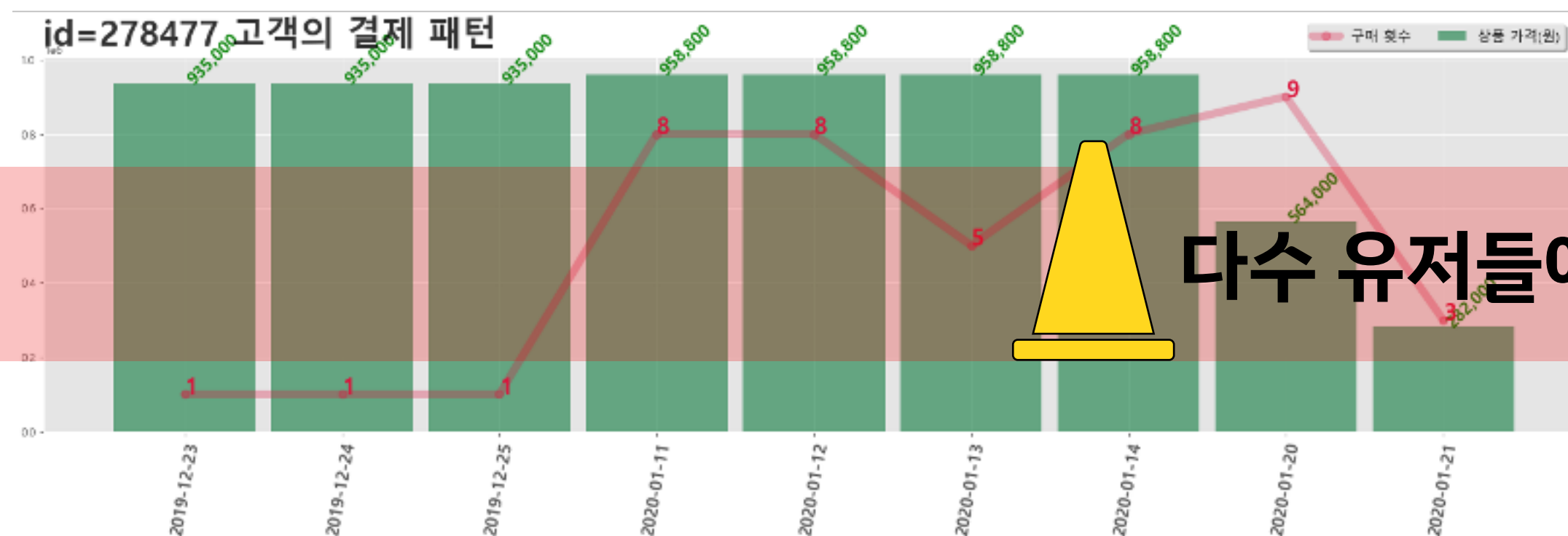
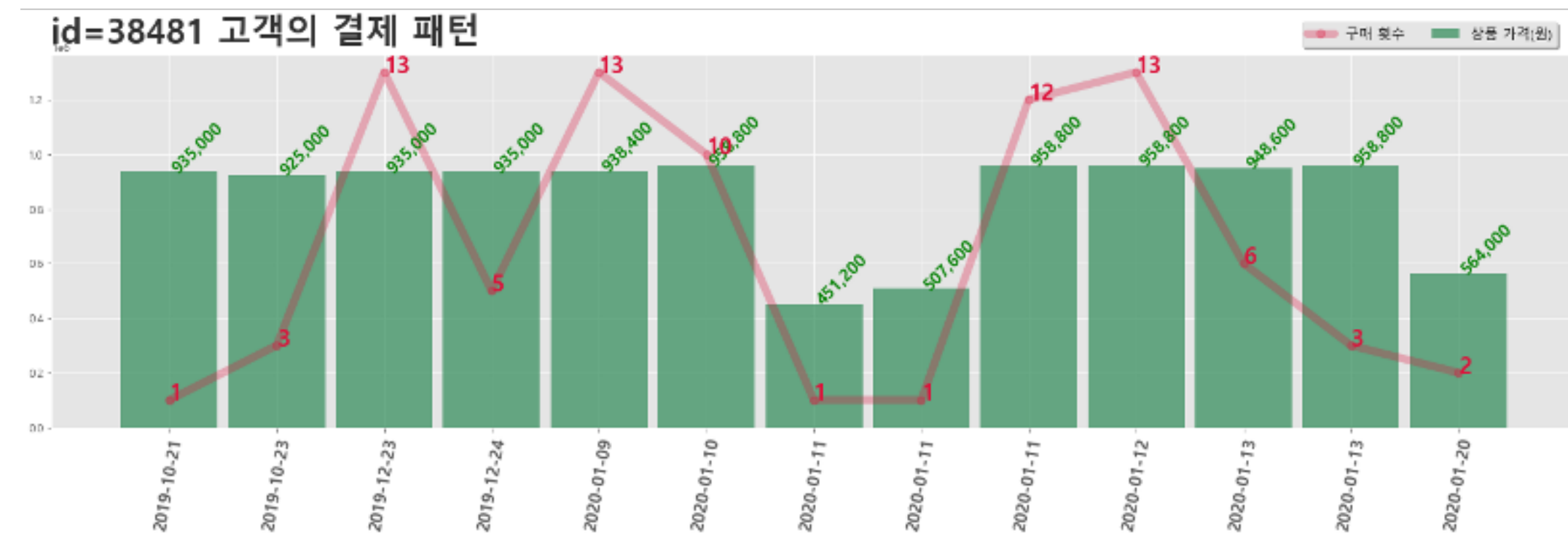
93.5만원 결제가 가장 많다

1.15억 결제 유저의 일단위 결제 패턴

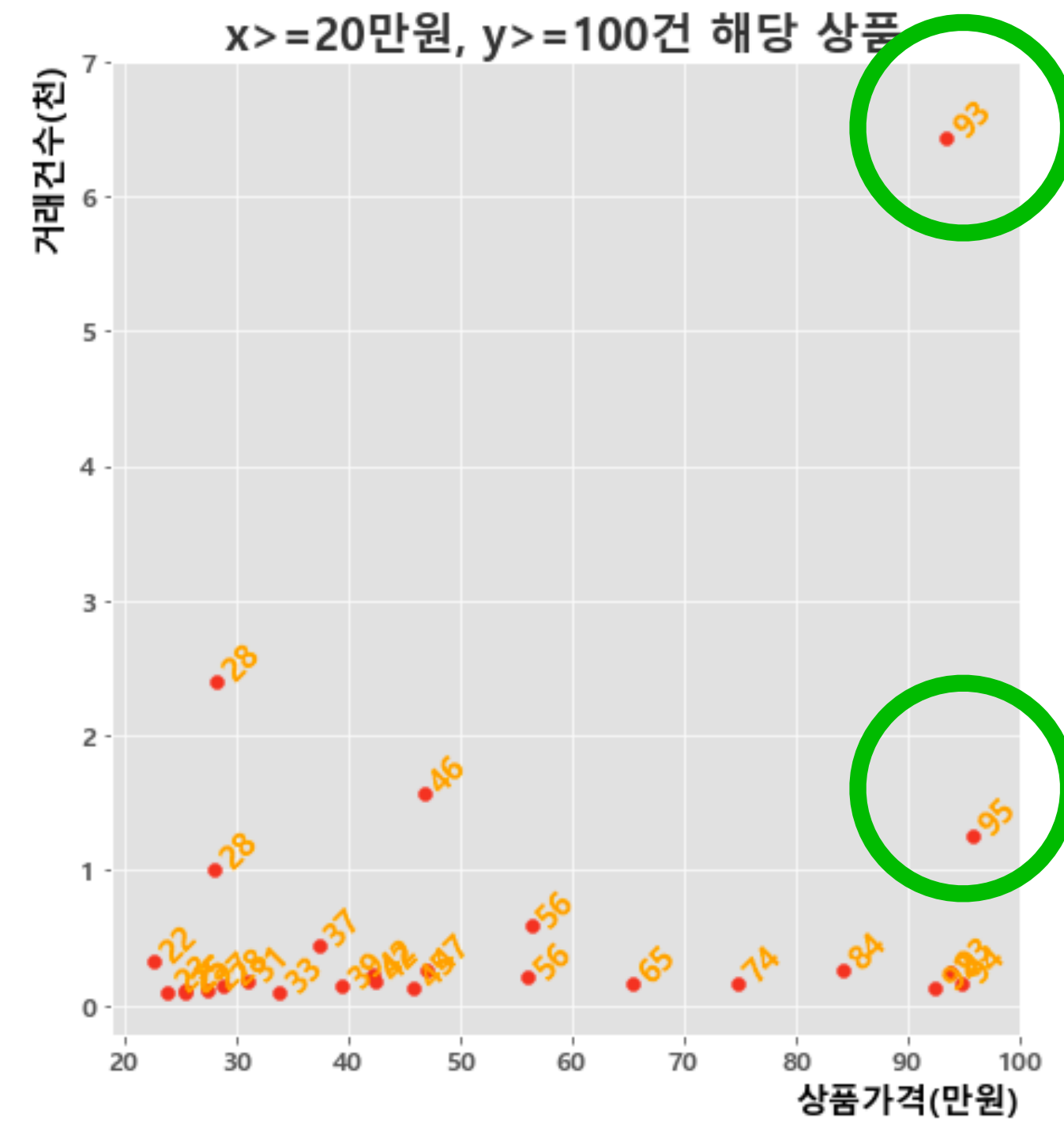
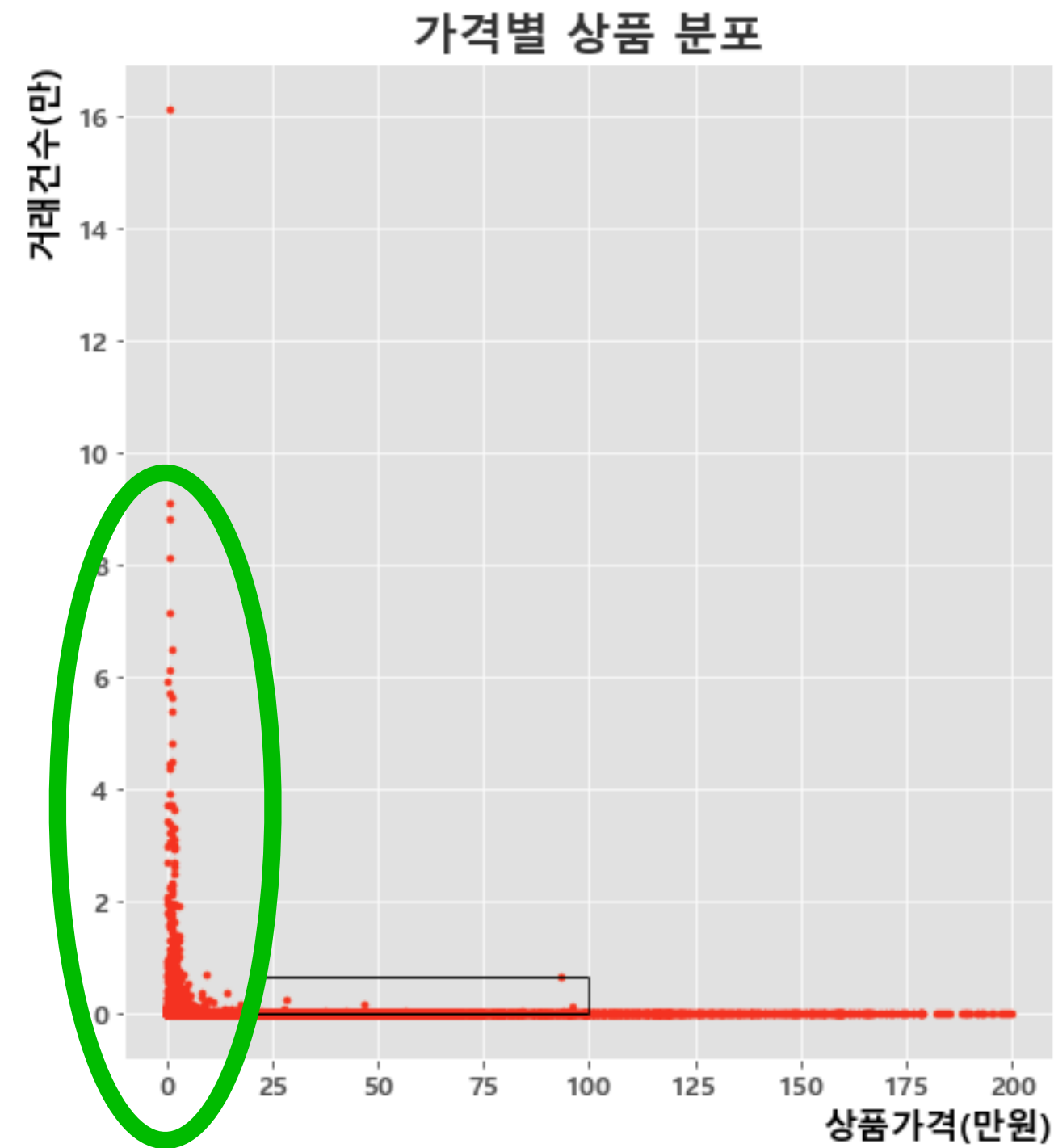


특정 날짜에 93.5만원 상품을 약 12개 구매
1,000만원 - 2,000만원씩 주문

심화 분석 상위 매출 고객 : 보따리상

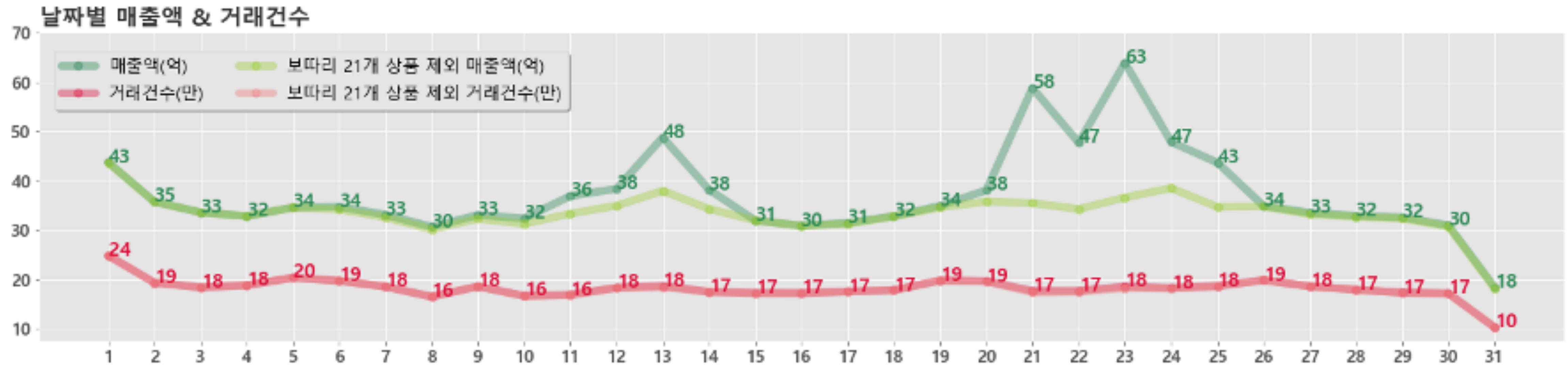


심화 분석 상위 매출 고객 : 보따리상



- 수많은 복수 구매자가 있는데 주로 구매하는 상품 가격이 정해져 있음
- 20만원 이상, 100건 이상 팔린 제품은 21개
- 이 상품을 **보따리 상품**이라 정의하고 분석을 진행

심화 분석 상위 매출 고객 : 보따리상



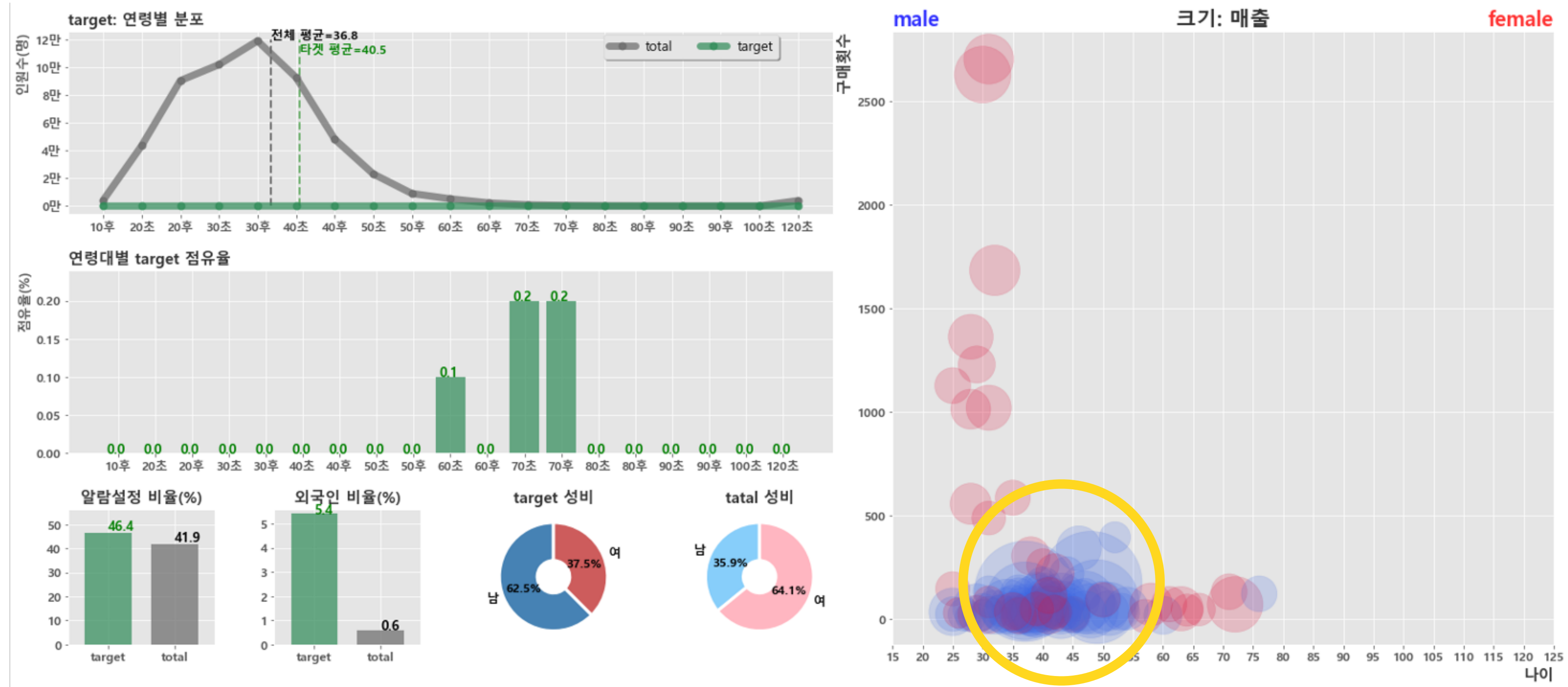
<일자별(1-31) 매출액>



특정한 날짜에 매출이 상승한다.

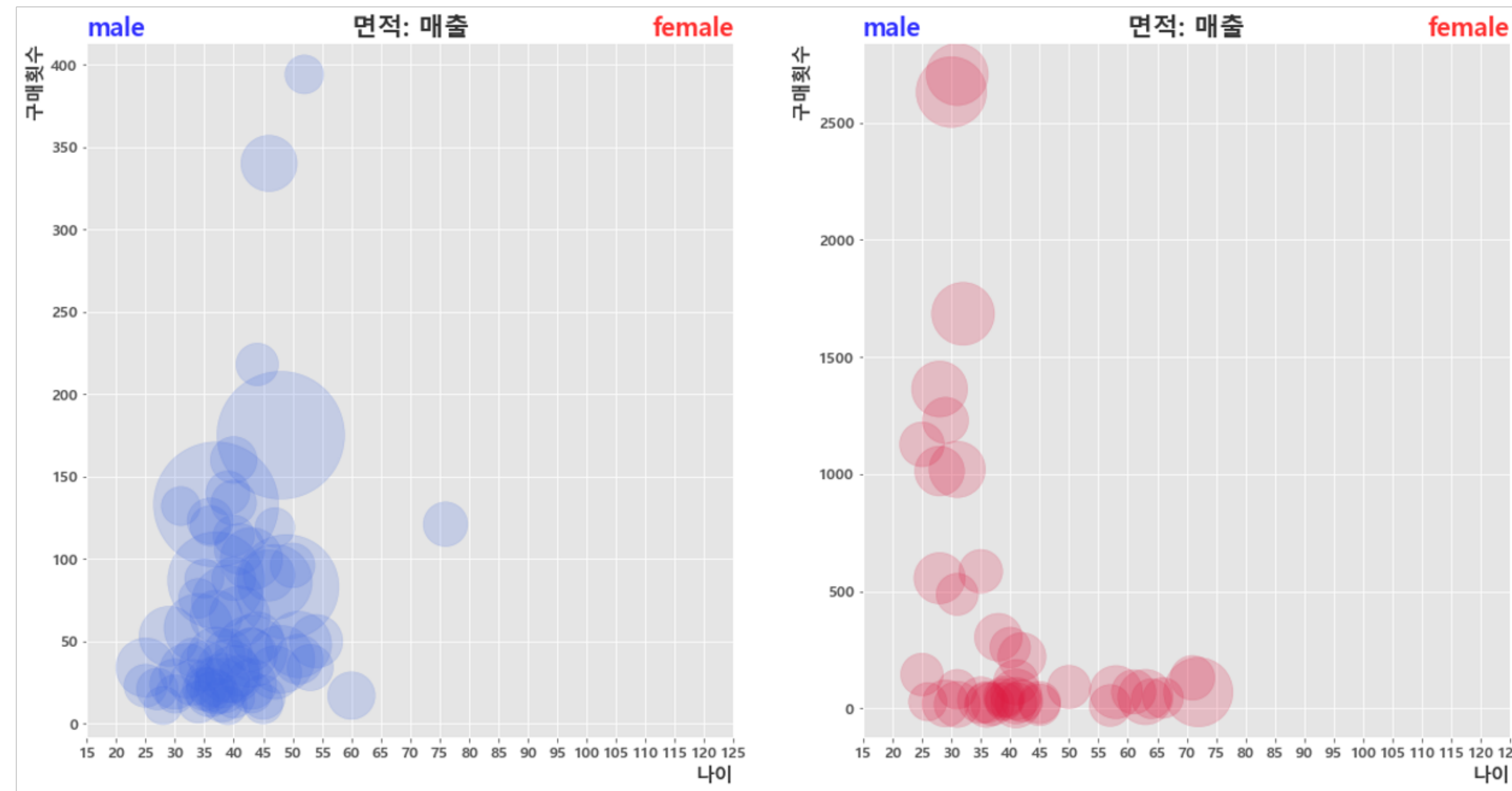
- 보따리 상품(**초록색**)을 제외하면 평소와 비슷한 매출 패턴을 보인다.(**연두색**)

심화 분석 상위 매출 고객 : 보따리상



- 8개월 간 1000만원 이상 구매한 고객 중 남성 비율이 높음
- 구매 횟수는 낮으면서 **고가**의 물품을 **대량** 구매하는 패턴의 고객은 **보따리상**
- 보따리상은 **남성**이 더 많음

1,000만원 이상 구매 유저 데이터

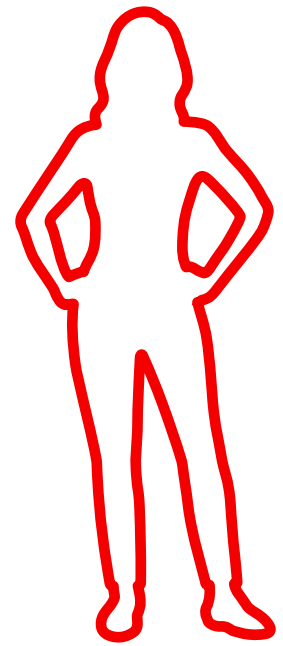


<1,000만원 이상 유저의 연령, 구매횟수, 성별 분포>

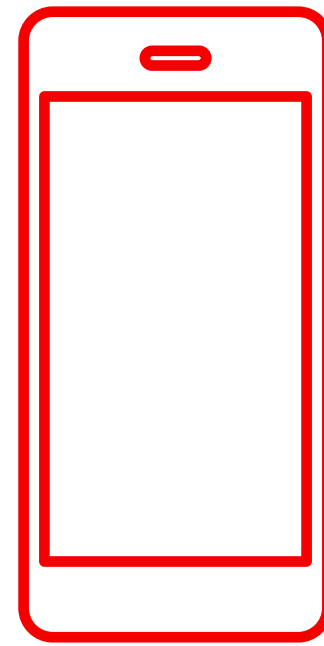
- 각각 남성 70명, 여성 42명
- 남성의 군집이 경향성이 더 있는 것으로 보임

결과 및 결론 ⚡

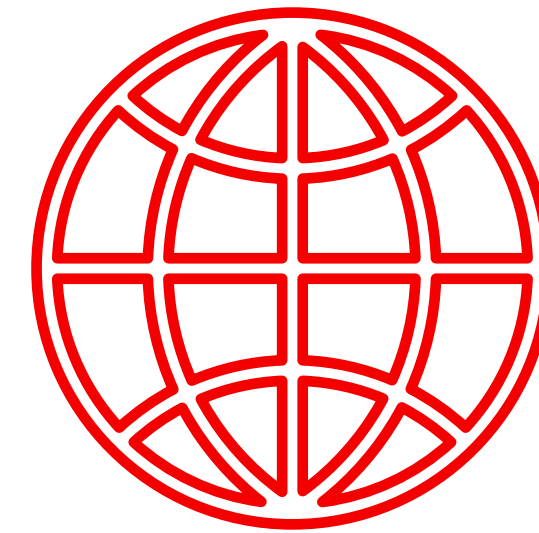
기본 분석



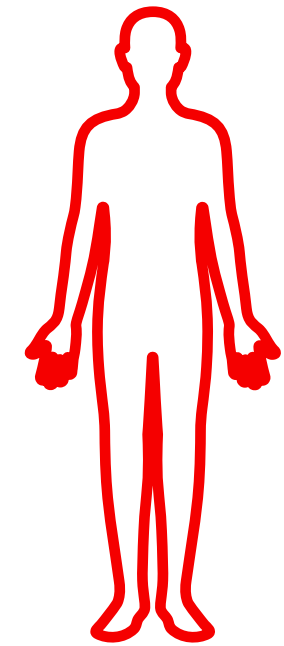
여성 유저의 수가 많고
평균적으로 많이 소비함



Push 알림 설정 유저 수는
적지만 더 많이 소비함

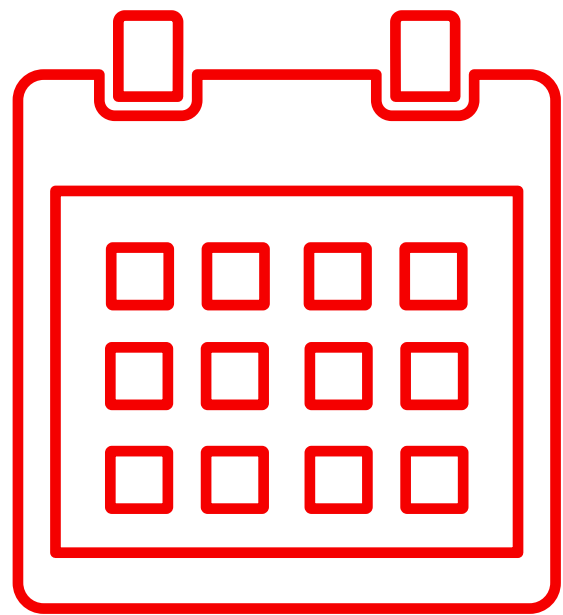


외국인의 소비는 평균보다
높으나 숫자가 매우 적음



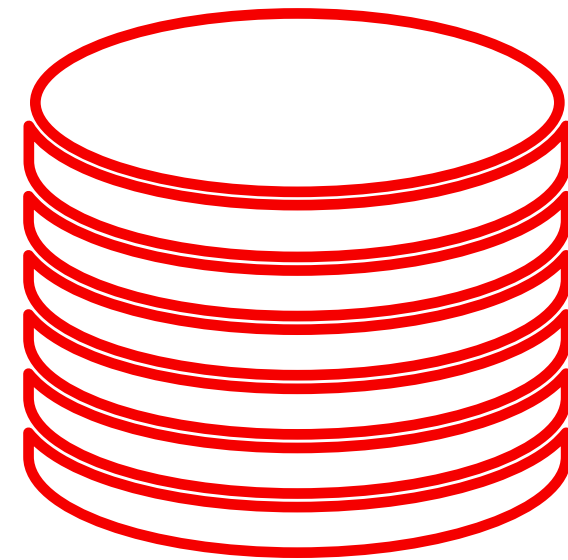
차이 유저의 평균 연령은
약 **25-48세**로 추정
상위 매출 유저의 연령대는
30-60세로 추정

심화 분석



2020년 1월, 활동 유저 수는
가장 적었지만 매출은 가장 높음

따라서 차이의 매출은
단순 활동 유저 수에 비례하지 않음



차이의 매출을 가장 높이는건
상위 8%대 고객들

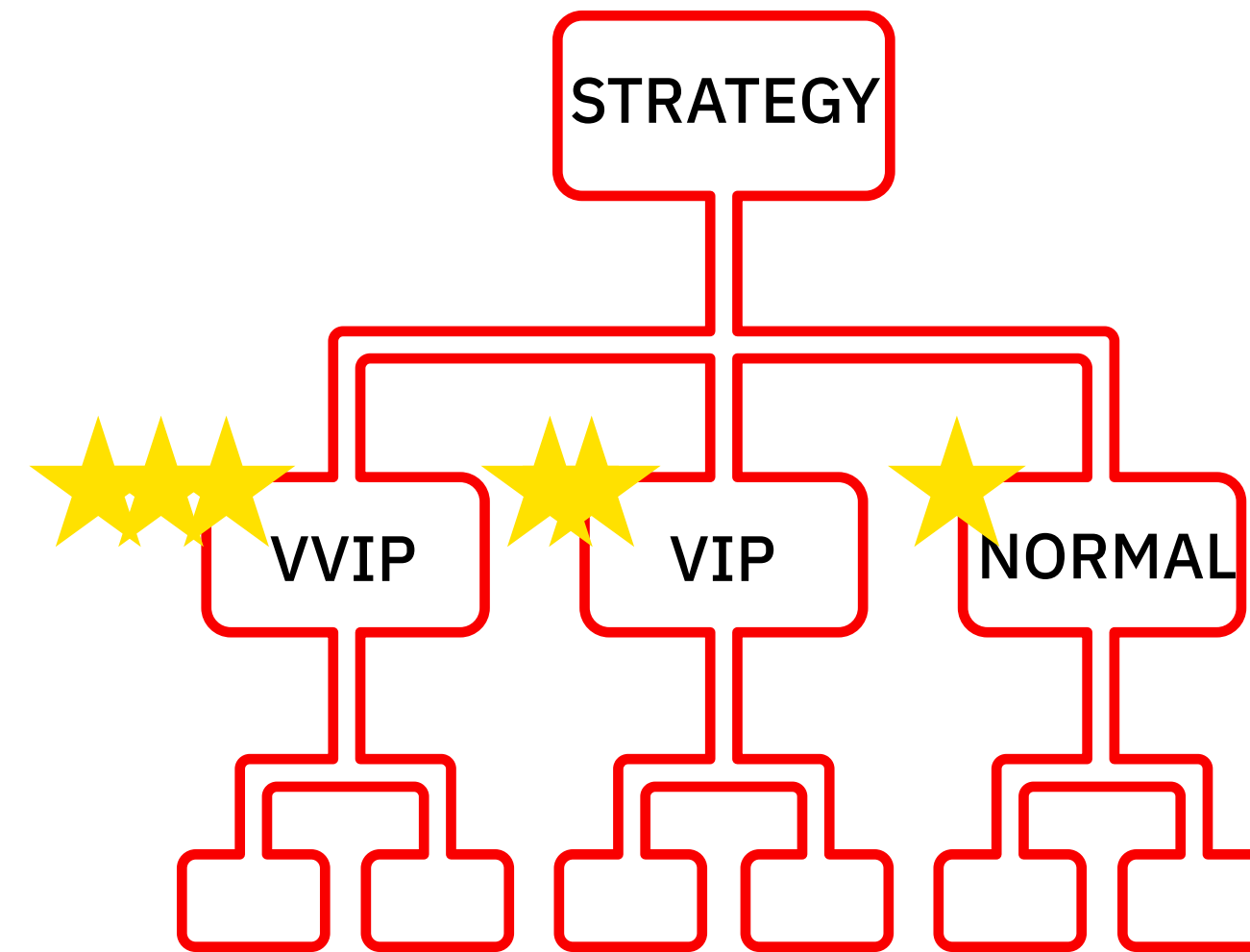
그 중 **상위 1%**가 매출에
가장 큰 영향 (파레토 법칙)



보따리상은 구매 횟수는 낮아도
고가 상품을 복수 구매하는 고객군

주로 구매하는 가격 상품이
있는걸로 보임

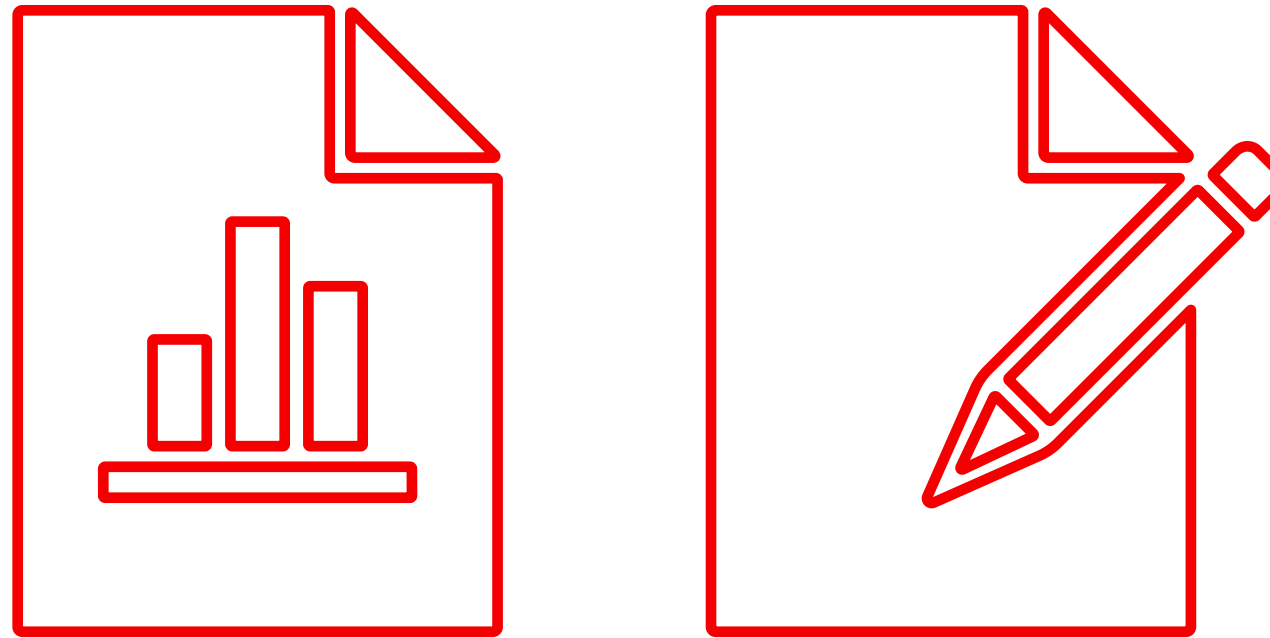
활용 가능성



고객 등급 제도 도입

- 결제 금액, 결제 횟수, 월간 결제 횟수, Push 알림 설정 여부 등 활용하여 점수 및 가중치 부여
- 고객을 Normal, VIP, VVIP로 군집 분류하여 각 분류에 맞는 전략을 수행

활용 가능성



마케팅 활용

- 상위 1% 유저, 보따리상 등 고객군은 특정 거래 패턴은 보이나 표본이 작아 일반화가 어려움
- 결제 대행 서비스에 타겟 마케팅 적용이 적절한 지 검증 필요
- 해당 데이터를 마케팅에 활용한다면 보조적인 수단으로 활용하는 것이 적절함



●● **Thank you for watching!**