

1. 전처리

1. 데이터 타입 변경

browser_id, session_id, event_type, datetime -> text
purchase_amount -> int

2. 데이터 Import

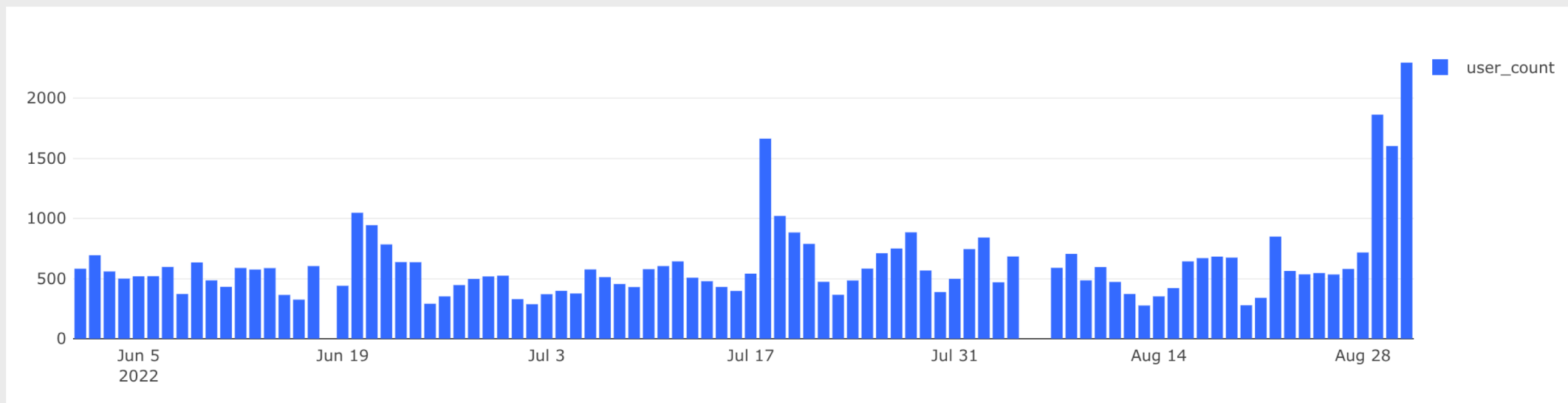
1. purchase_data.csv -> purchase_data 테이블

3. 데이터 이해

1. 2022-06-01부터 2022-08-31까지 54402명 고객의 일/시별 구매 데이터

전체 사용자들의 거래의 최근성(Recency) 분포 분석

Recency

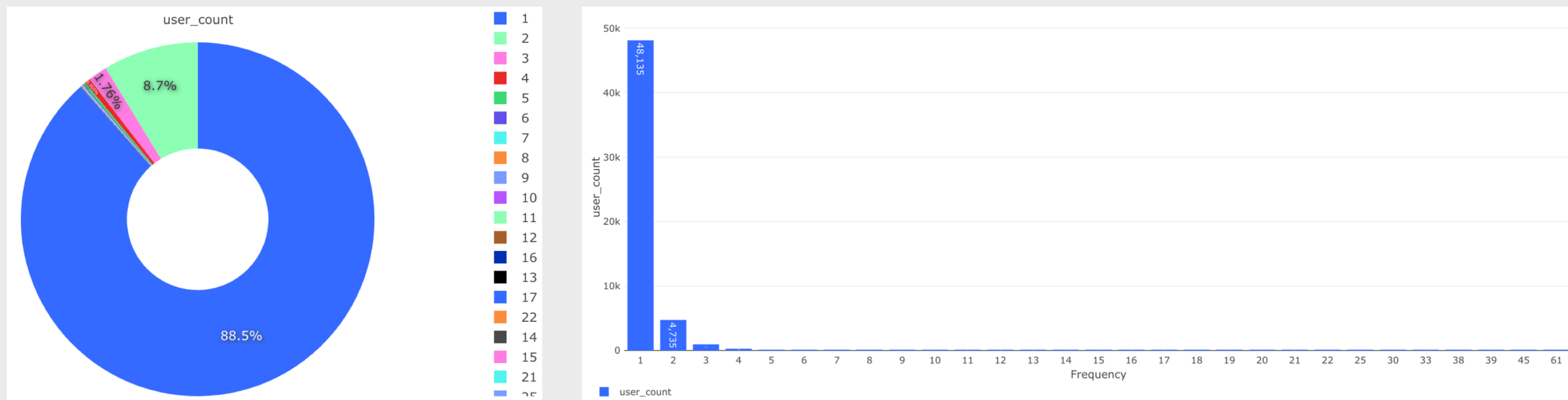


전체 데이터의 가장 최근 3일 (2022-08-29, 30, 31)에 가장 최신/마지막 구매가 이루어진 고객이 1500명 이상으로 가장 많으며, 그 이전 기간에 마지막으로 구매한 고객은 약 500명/1일로 고르게 분포한다.

특징적으로 보이는 점은 7월 18일에 마지막 구매를 한 고객이 1,663명으로 많은 것인데 이는 해당일 큰 할인행사 등으로 반짝 유입된 고객이 많았던 것으로 추정된다. 또한 6월 18일, 8월 5-6일에 마지막 구매 고객이 한 명도 없다는 점이 두드러진다. 해당 일자에 대한 분석이 필요해 보인다.

전체 사용자들의 거래의 빈도(Frequency) 분포 분석

Frequency

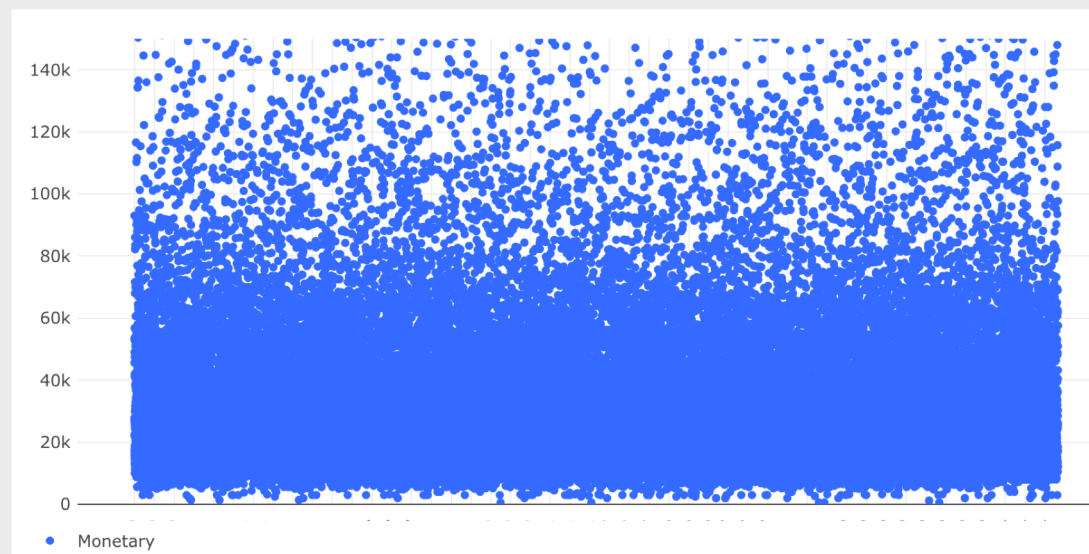
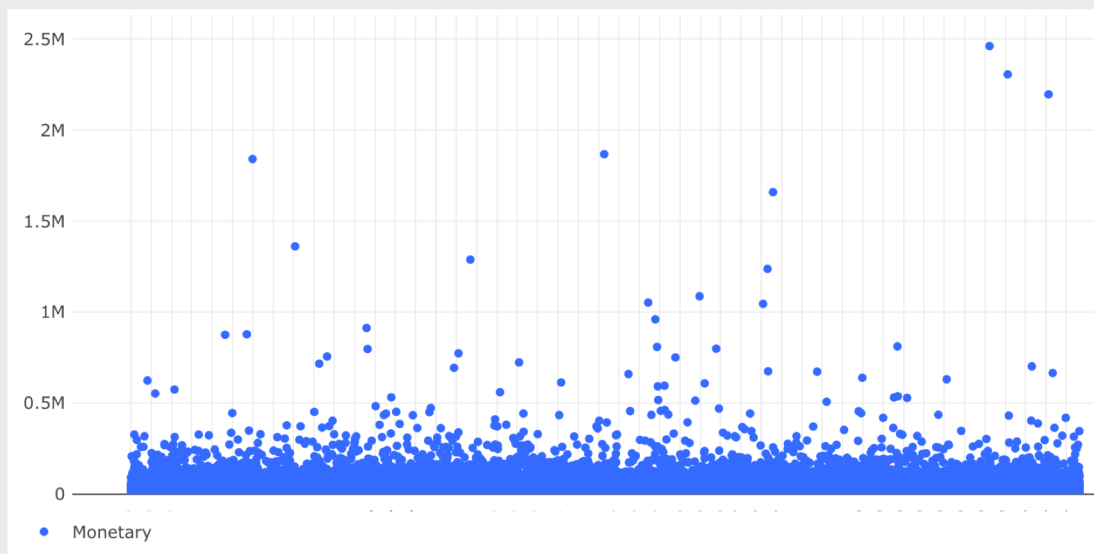


위의 두 그래프에서 볼 수 있듯, 구매 빈도(Frequency)가 1인 고객이 48,135명 전체 88.5%로 가장 많고, 구매 빈도가 2인 고객이 4,735명 8.7%, 3인 고객이 1.76% 등 구매 빈도 1과 2인 고객이 전체의 97.2%를 차지하고 있다.

따라서 RFM 등급 계산시 빈도 1에서 61까지의 구매 빈도를 단순히 구간으로 나누는 것이 아닌, 구매 빈도 1,2를 1점(가장 낮은 점수)으로 매기고 3~5를 2점, 6~61을 3점으로 매기는 등 구체적인 범위를 나누는 등의 전처리가 필요하다.

전체 사용자들의 거래의 거래금액(Monetary value) 분포 분석

Monetary value



왼쪽 그래프에서 볼 수 있듯이 전체 데이터에서 총 거래금액은 600원부터 2,460,900원까지 매우 다양하게 분포되어 있다.

오른쪽 그래프는 거래금액의 최댓값을 150,000원으로 설정하여 본 것인데, 구매 고객 대부분의 총 거래금액이 80,000원 이하임을 확인할 수 있다.

따라서 전체적으로 데이터가 500,000원 이하에 분포되어 있는 반면 500,000원 이상의 거래 금액은 이상치처럼 보이므로 거래금액의 스케일을 맞춰 주거나 거래금액대의 범위를 나누는 등의 전처리가 필요하다.

RFM 기법을 활용해 각 사용자의 R, F, M 등급(점수)을 계산하고 나눈 세그먼트에 대한 분석

```
SELECT DISTINCT browser_id,
  MAX(DATE_FORMAT(datetime, '%y-%m-%d')) as 'Recency',
  COUNT(datetime) as 'Frequency',
  SUM(purchase_amount) as 'Monetary',
  NTILE(3) OVER (ORDER BY max(DATE_FORMAT(datetime, '%y-%m-%d')) ASC) as R,
  NTILE(3) OVER (ORDER BY count(datetime) ASC) as F,
  NTILE(3) OVER (ORDER BY sum(purchase_amount) ASC) as M
FROM purchase_data
GROUP BY browser_id;
```

browser_id	Recency	Frequency	Monetary	R	F	M
f7a4011e5b7dac1cf9ad3b4a79599a9af8195ad5a941b58700888da749e8d7bf	22-08-31	61	2,195,810.00	3	3	3
9aae35c51dfae3c78eb2c4aebf1bc5b62bad9121566cc8b5b9dee15522a60e49	22-08-31	8	609,300.00	3	3	3
9014075fe38506fa1a7d0d31bf667c73cc1cdb5749da2b48f52ade8294560704	22-08-31	16	596,070.00	3	3	3
8e229f3c4361caa7083701b4119749fc4e832bcc3fe19d3bda45f0bd8f13013	22-08-31	2	592,500.00	3	3	3
46a2d374d9ef453cf2b0a52f16b0db591d4a02d4c13070e2e624738ee399e302	22-08-31	16	532,720.00	3	3	3
8e53314a97eeb651d1cd6d1bc520e0ac12dd7f7f5995d061fbf07da5f200447c	22-08-31	10	517,300.00	3	3	3
c3a932a6d063b9d18cc4794049c01389da5c58df63bca2e4f19e84ff128e0ef7	22-08-31	2	456,200.00	3	3	3
a6e663993c09731d446c5051391ffbcb47f8383b5d257cbe54f98d89eff1fdf6	22-08-31	11	443,300.00	3	3	3
ca3012dcb46bcd043e1a1d0c846c773c1f25d19a43d803faa36dffdeb6fbf70d	22-08-31	2	420,200.00	3	3	3
624c2353047392523f1e2f64caec0ac9f73990b2b1bc1ce98545025aeb6348bd	22-08-31	8	411,260.00	3	3	3
36d2e5f9e5f9e736c7ba43d2cadf2c9286403b902cce414a077358876e35d8f0	22-08-31	8	403,600.00	3	3	3
7e19fe974dd65c00892bcd8bda341e821c9c51e4c4973c1816901d788d052d866	22-08-31	4	362,000.00	3	3	3
a73394e572034d7234bca095ba782c3eb05d380a99f774c629eedb092078d625	22-08-31	6	347,500.00	3	3	3
5893854f2eaa9d3eb6d4f47951e827a8d5be368703a82e63c1381cb1a1d35816	22-08-31	2	340,400.00	3	3	3

전처리 없이 NTILE로 3그룹 분할

R, F, M이 모두 3점인 고객 그룹은 VIP 그룹으로 지정하여 떠나지 않을 마케팅 전략을 세워볼 수 있다. 구매 금액이 커지거나 더 짧은 주기 이내로 구입할 경우 혜택을 주는 등의 방식을 떠올려볼 수 있다.

R, F, M에서 1, 2, 3점이 섞여있는 고객 그룹은 이탈 가능성이 있지만, 가장 붙잡아야 할 고객으로 R이 낮다면 오랜만에 접속시 할인 쿠폰을 주고, M이 낮다면 5만원 이상 구매시 10% 할인 등의 일정 금액 이상 구입시 할인 쿠폰을 제공하는 등의 방법을 제시해볼 수 있다.