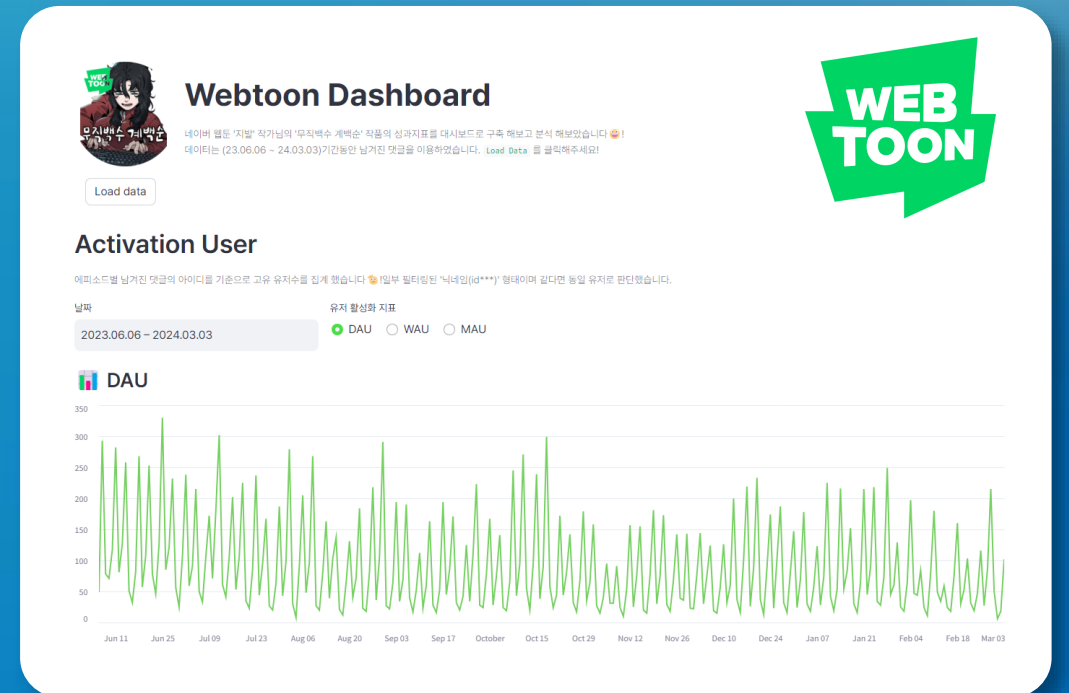


WEBTOON KPI 대시보드 및 분석

평소에 즐겨 보던 네이버 웹툰 '무직백수 계백순' 작품을 하나의 프로덕트로 보고
핵심성과지표 대시보드를 만들어 데이터를 분석해 보았습니다.



프로젝트 개요

웹툰의 수익구조를 파악하고 웹툰을 연재하는 작가의 입장에서 어떤 지표들을 확인하고 분석하면 좋을지 고민해보았습니다.

웹툰의 수익모델? 🤔

수익성 분배 PPS
(Page Profit Share)

웹툰 하단의 이미지 광고,
IP기반 비즈니스를 통한 수익창출

부분 유료화

'쿠키'를 결제하여 미리보기
유료 판매 수익

웹툰 조회수 \propto 광고 노출수
쿠키 결제량 \propto 수익

“ 독자들의 관심도가 중요 ”

“ 독자들의 관심도를 높이고 ”
“ 독자들이 작품(서비스)을 꾸준히 보고 있는지 ”
관리할 수 있도록 모니터링 및 분석이 필요하다

대시보드

활성화지표
Stickiness
RFM, LTV
...

분석

독자들의 특징
결제가 높은 에피소드의 특징
가설 검정
....

평소에 즐겨 보던 무직 백수 계백순 이라는 작품을 하나의 프로덕트로 보고
독자들이 꾸준히 작품을 이용하고 있는지 확인하고
인사이트를 찾아 보려고 합니다!

데이터 수집

지표 선정

독자들의 고유 key 필요

웹툰에 남겨진 댓글의 닉네임(ID***)
을 이용하여 고유 KEY 수집
(아이디, 댓글 작성 날짜)

에피소드에 관한 지표

제목, 좋아요, 업로드 날짜



Python Selenium

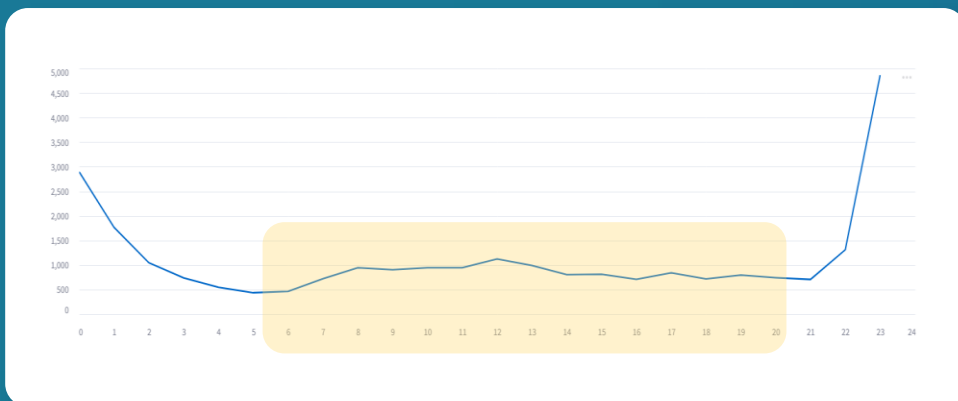
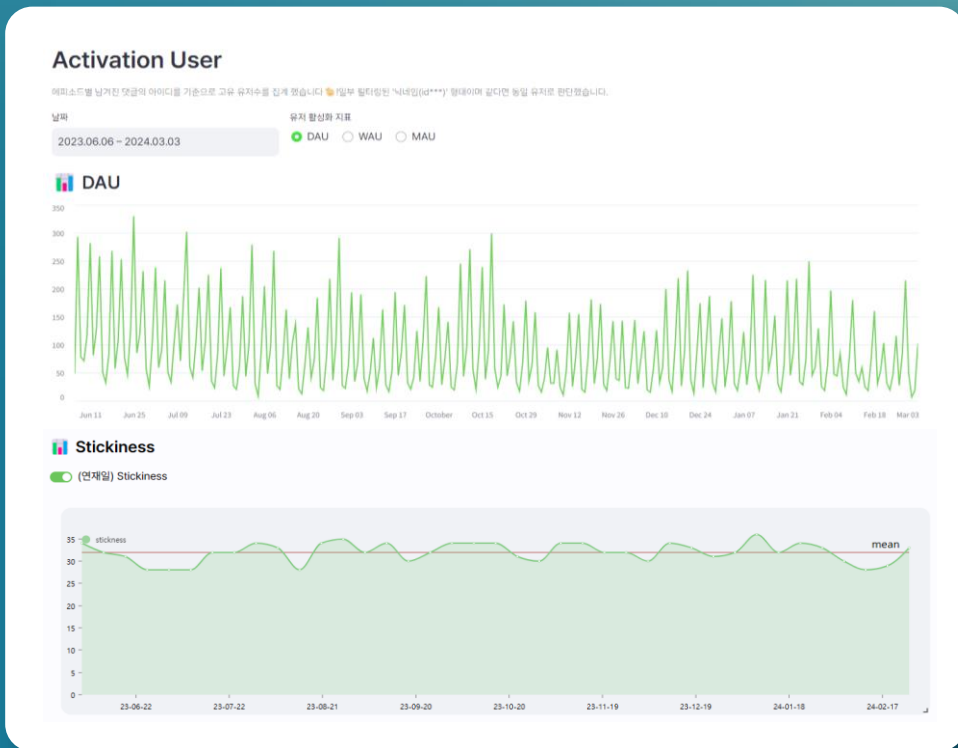
episode	↓ user_id	comment_date
13화 : OGMA 컴퍼니 (5)	겨울이 (wint****)	2023-06-20 22:50:00
12화 : OGMA 컴퍼니 (4)	겨울이 (wint****)	2023-06-17 22:54:00
10화 : OGMA 컴퍼니 (2)	겨울이 (wint****)	2023-06-13 22:56:00
9화 : OGMA 컴퍼니 (1)	겨울이 (wint****)	2023-06-13 22:54:00
9화 : OGMA 컴퍼니 (1)	겨울이 (wint****)	2023-06-13 22:54:00
8화 : 습격	겨울이 (wint****)	2023-06-10 22:35:00
8화 : 습격	겨울이 (wint****)	2023-06-11 02:11:00

'닉네임(ID***)' 형태가 같은 경우
동일한 독자로 판단했습니다.

기본 성과 지표

댓글 데이터를 통해 구한 독자 활성화 지표와 웹툰 이용 시간대에 대한 내용이 들어있습니다.

기본 성과 지표



1. 활성화 지표, Stickiness

활성화 독자의 기준을 “ 댓글을 남겼다 ” 라는 행동으로 정하고 계산했습니다.
댓글은 웹툰을 보고 난 뒤, 서비스를 이용했다는 가장 확실한 흔적이라고 볼 수 있기 때문입니다.

높은 DAU 변동성

정해진날에 연재되는 웹툰 서비스 특성상
연재일에만 들어오는 독자들이 많았습니다.

평균 고착도(DAU/WAU) 32%

주 2회 연재일 기준
100명중 32명은 꾸준히 댓글을 남기고 있습니다.

2. 웹툰 이용시간대

독자가 남긴 댓글 작성시간을 통해 집계하였습니다.

Mission

상대적으로 적은 오전/낮 시간대 UV를 발견할 수 있었는데요!
해당 시간대에 독자들을 효율적으로 끌어오려면 어떻게 해야 할까요?

Insight

RFM^{*}, LTV^{*} 산출하기

상대적으로 적은 낮, 오전의 UV를 확보하기 위한 프로모션을 고민해보고
해당 프로모션을 진행 예산을 효율적으로 사용하기 위해 반응할 확률이 높은 우선 독자들을 선정해보았습니다.

RFM

최근에, 자주, 얼마큼의 금액을 사용하여 서비스를 이용했는지를 통해 고객을 세분화하는 지표

LTV (고객 생애 가치)

고객이 평생 우리 서비스에 얼마큼 이익을 줄 것인지를 측정하는 지표

프로모션 설정해보기 🎁

오전/낮 시간에도 많은 UV 확보하기 위해
'등교/출근, 점심 시간대에 맞춘 n분 무료보기 및 쿠키 조조할인 혜택 및 광고'
프로모션을 진행하려고 합니다!

해당 프로모션을
효율적으로 진행하려면?

반응할 확률이 높은 독자들을
우선 선별하는게 좋지 않을까?



독자들의 특징 분석이
필요하다!

독자들의 특징을 구하고
세그먼트로 나눠보자



고객별 LTV
RFM
이용 시간대
...

지표들을 구하여
독자들을 세그먼트로 나누고
프로모션 우선 진행 독자를
선별 해보려고 합니다!

파생변수 추가

① 독자별 Monetary를 구하기 위해 Price를 정의하고 계산해야 했습니다.

① 댓글 작성수 + 받은 좋아요

독자들의 참여도를 이끄는 대표적인 지표로
금액의 가치가 있다고 생각했습니다.

(※ 댓글 작성수 당 500원, 받은 좋아요당 1원 으로 환산하여 집계)



② * 쿠키(유료결제)사용 금액계산 🍪

웹툰이 게시된 날짜 > 댓글이 작성된 날짜 인 경
우
웹툰을 미리 보았다는 증거 이므로
쿠키를 사용한 독자로 파악했습니다.

(※ 실제 쿠키의 기점을 해당 120원 이지만 그점의 표현을 위해 1200원으로 조정)

episode	upload_at	user_id	comment_date
56화 : 무서운 에어컨 (2)	2023-12-16	두부 (pok2****)	2023-11-19
29화 : 물골에 관하여 (2)	2023-09-12	핑핑이 (kkan****)	2023-08-08
42화 : 노동의 맛 (2)	2023-10-28	나야나 (pjhn****)	2023-09-24

(table) 미리보기 결제를 이용한 독자 일부

쿠키?

네이버 웹툰에서 유료 회차 감상(미리보기)을 위해 사용되는 화폐, 결제 수단입니다.

파생변수 추가

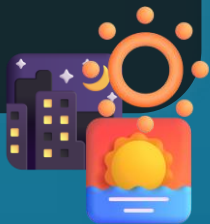
② 주로 이용하는 시간대 변수를 추가했습니다. 프로 모션을 진행하는데 파악이 필요한 변수 입니다.

댓글 작성 시간^에 따라
이용시간대를 구분했습니다.

Morning
오전 6시 ~ 12시

Day
오후 12시 ~ 18시

Night
나머지



모델링

Python 패키지의 Lifetimes 를 이용하여 독자들의 RFM과 LTV를 산출했습니다.



Lifetimes

BG/NBD
독자별 예상 구매횟수

GammaGamma
독자별 예상 구매금액

해당 모델 사용이유

RFM 정보를 쉽게 산출
+
이를 기반으로 미래가치를
계산하고 고객단위 타겟팅이
가능하다는 장점

모델링 과정 요약

RFMT 계산

summary_data_from_transaction_data 함수를 이용하여 독자별 RFMT 를 계산합니다.

Data split

데이터 크기는 약 8개월치 입니다. Train(8) test(2) 의 비율로 분리하기 위해 holdout(test)를 48일로 설정했습니다.

L2 페널티

데이터 크기가 작은 경우 매개변수 추정 값이 커질 수 있으므로 L2 페널티를 추가하여 최적화 합니다. (측정 기준 MSE)

모델링

BG/NBD 👉 독자별 예상 구매횟수
GammaGamma 👉 독자별 예상 구매금액
향후 8개월 간의 LTV 산출

모델링 결과

예측 구매 횟수 MSE
 ± 0.364 일

예측 구매 금액의 MSE
 ± 24085.169 원

LTV 등급 설정

독자별 LTV 분포

파레토 법칙

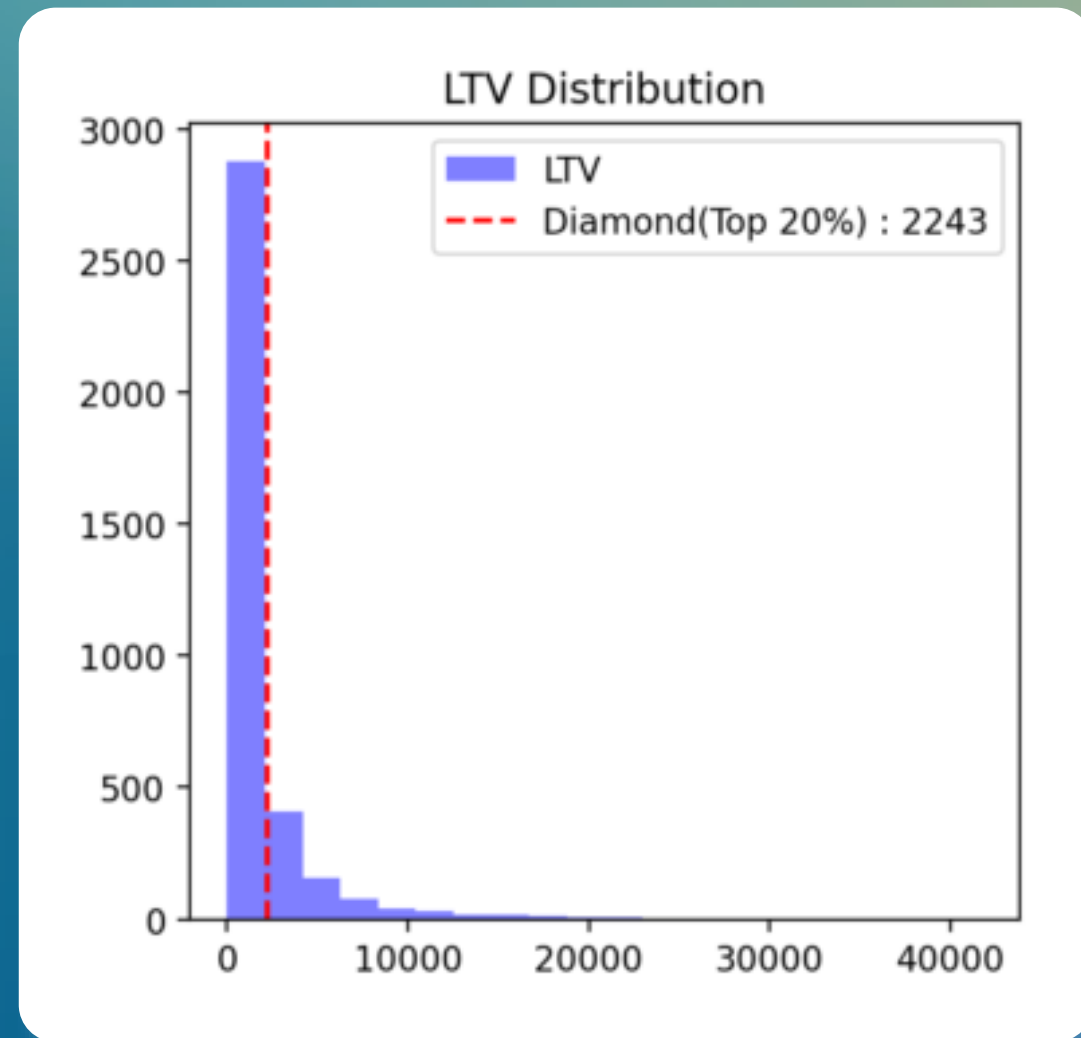
산출된 LTV의 값이 크게 나뉘지는 구간이면서
* 파레토 법칙이 적용 되는 상위 20%에 해당하는 독자들을
가장 높은 등급으로 선정하는 것이 적합해 보였습니다.

LTV 세그먼트 기준

(상위 20%) diamond
(20~40) platinum
(40~60) gold
(나머지) silver

파레토 법칙이란?

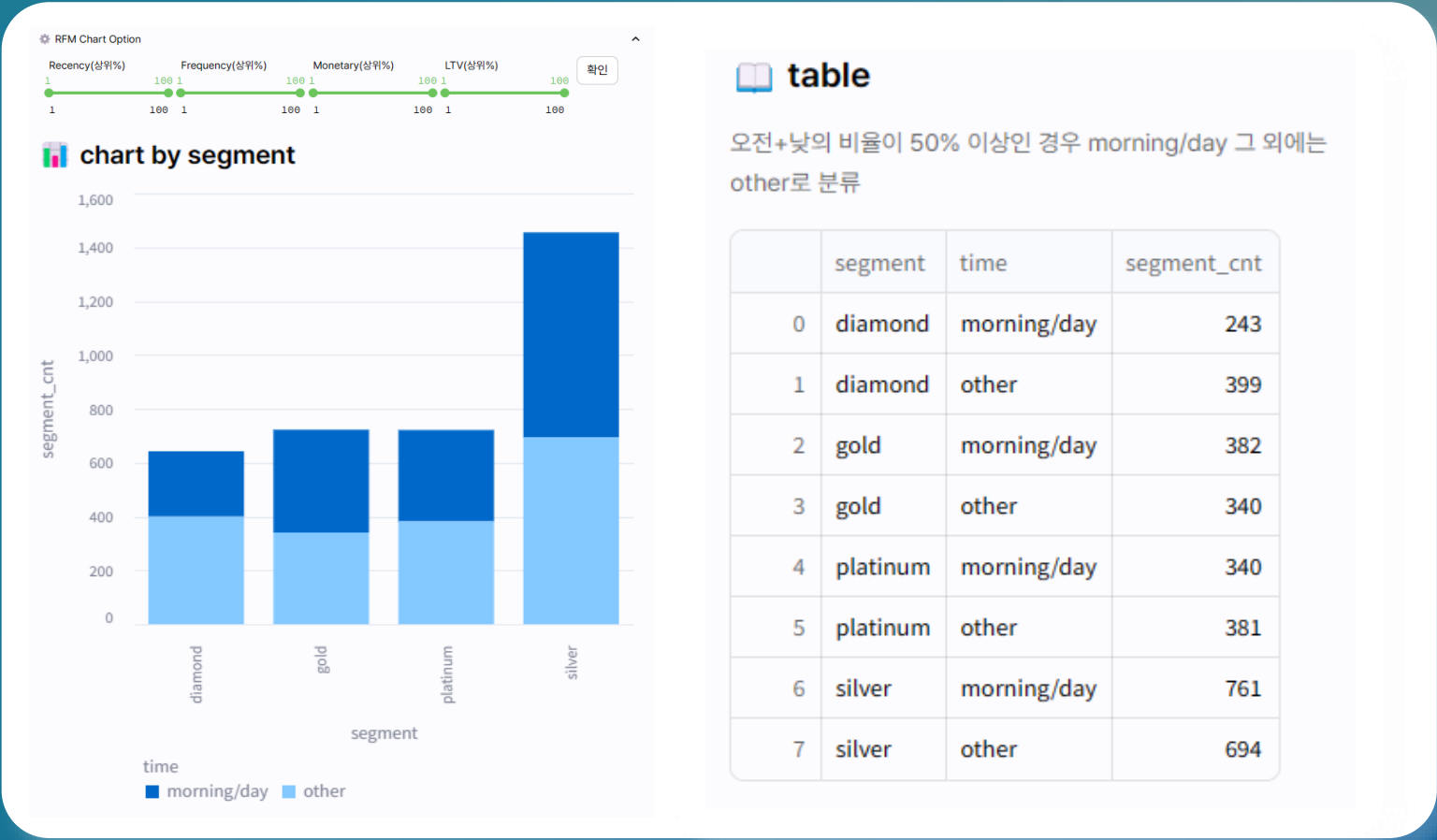
20%의 원인에서 80%의 결과가 나온다는 현상으로 비즈니스 활동의 우선순위를 정하는데 이용됩니다.



타겟 독자 설정

RFM LTV Chart

좀 더 디테일한 독자들의 분포를 보기 위해 산출된 RFM , LTV의 범위(상위%)를 조절하여 세그먼트와 이용시간대 별로 확인했습니다.
이용시간대 비율의 기준은 오전 + 점심 이용비율이 50% 보다 높은 독자의 경우 morning/day 그 외는 other로 분류하여 세그먼트를 나눴습니다.



table

오전+낮의 비율이 50% 이상인 경우 morning/day 그 외에는 other로 분류

	segment	time	segment_cnt
0	diamond	morning/day	243
1	diamond	other	399
2	gold	morning/day	382
3	gold	other	340
4	platinum	morning/day	340
5	platinum	other	381
6	silver	morning/day	761
7	silver	other	694

타겟 독자 설정

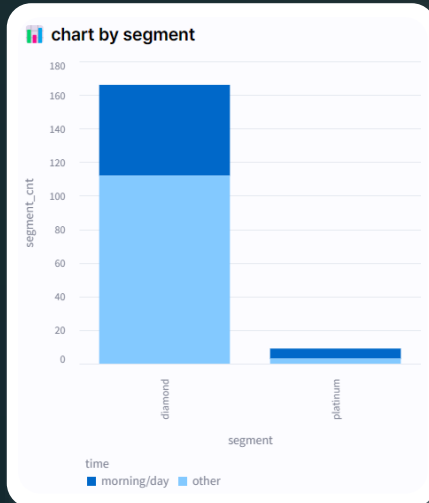
이를 통해 구해본 프로모션에 반응할 확률이 높은 독자들은 크게 3가지 입니다!

열혈 독자 🧒

chart option
RFM(1~30%) LTV(1~30%)

description
RFM, LTV 모두 준수한 독자
(+α 오전/낮 이용 비율이 50% 이상인 독자)

- 미래 가치가 높은 독자
- 꾸준히 이용하고 있는 고정독자층
- 특히, 아침/낮 독자의 경우 해당 프로모션에 가장 잘 반응할 독자라고 생각

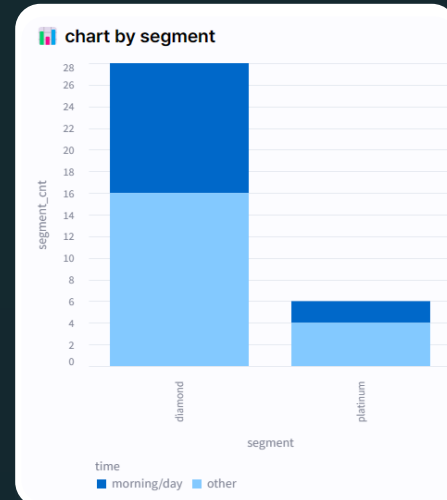


잠든 열혈 독자 🧒

chart option
LTV(1~30%) R(50~100%) FM(1~30%)

description
생애 가치가 높고 자주 많이 사용한 고객이지만
최근성이 떨어지는 독자
(+α 오전/낮 이용 비율이 50% 이상인 독자)

- 과거 높은 참여도의 이력
- 프로모션에 따라 복귀할 확률이 높음
- 특히, 미래 생애 가치와 지출금액이 높기 때문에 붙잡아야 하는 독자



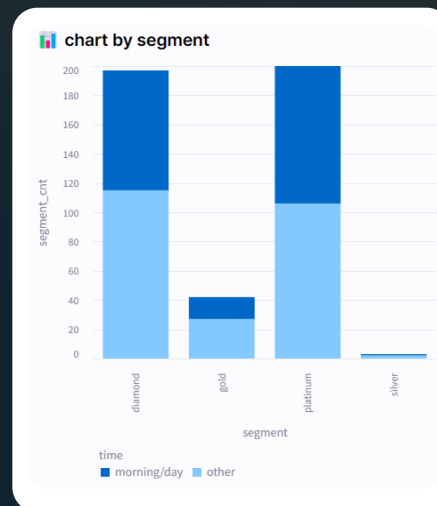
잠재 독자 🧒

chart option
LTV(1~30%) RF(1~30%) M(30~100%)

description
지출 금액은 적지만 최근에 많이 활동한 독자
(+α 오전/낮 이용 비율이 50% 이상인 독자)

why?

- 현재 해당 웹툰에 긍정적이고 참여도가 높은 독자
- 더 나아가 혜택을 받기 위해 프로모션에 관심을 갖고 있는 독자라고 생각



가설검정

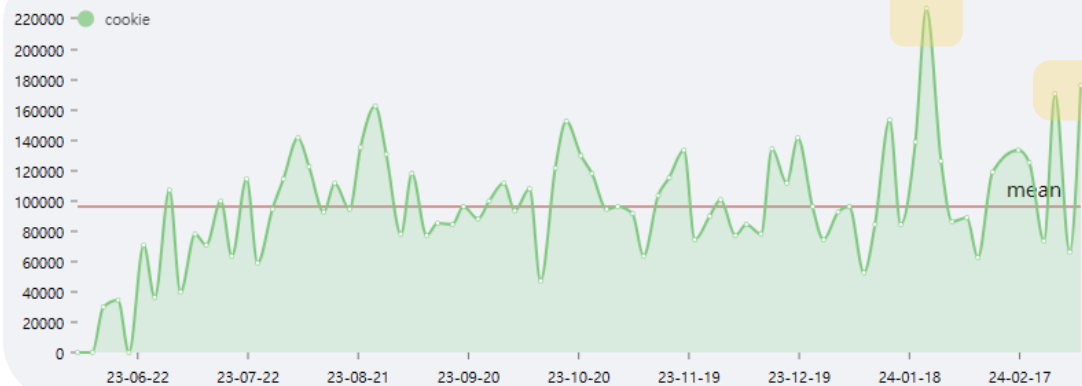
쿠키사용량이 높았던 에피소드의 특징은 무엇일까요? 에피소드에 따라 쿠키 사용량의 차이가 있을까요?
해당 웹툰의 에피소드별 쿠키사용량을 시각화하고 발견한 인사이트로 가설을 세우고 검정 해보려고 합니다!



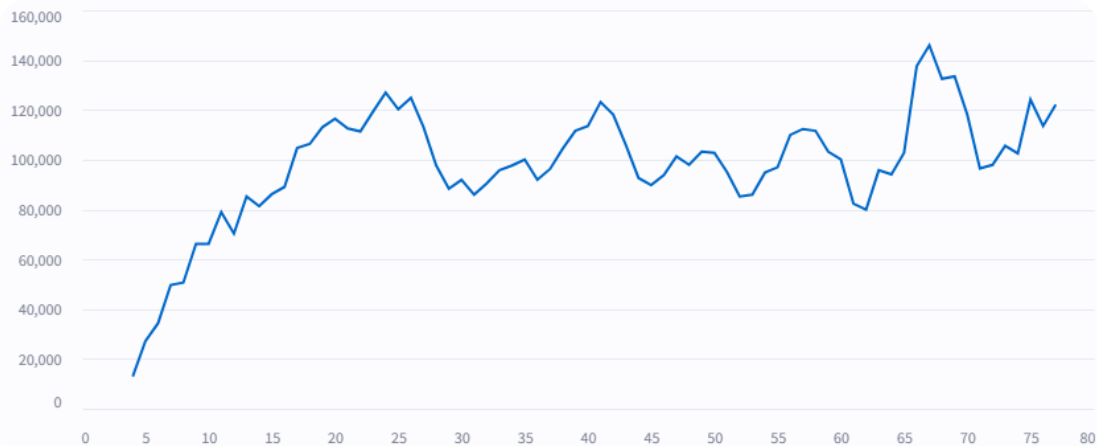
Episode by Cookie

해당 웹툰의 에피소드별 쿠키 사용량은 꾸준히 증가하는 추세를 갖고 있었는데요! 🤖

x: 업로드 날짜



업로드 날짜별 쿠키 사용량



쿠키량이 높았던 에피소드 TOP3

insight

그 중 눈에 띄었던
쿠키량이 많은 에피소드의 공통점은

모두 "시리즈" 형태의
에피소드 라는 점 이었습니다.

실험 가설

그렇다면..

“ 시리즈의 형태의 에피소드 일 수록 쿠키 사용량이 높은 걸까?”

하루분량으로 끝나는 '단편 스토리' 와 반대로 계속해서 이어지는 '시리즈' 형태(1~N)의 에피소드 의 경우 독자들의 입장에서 뒷 내용에 대한 궁금증이 더 유발되고 이에 따라 쿠키를 사용하지 않을까? 생각이 들었습니다.

검정 과정

데이터 전처리

- 초기에 미리보기 할 웹툰이 없다는 점을 고려하여 초기 에피소드는 제외
- 2개 이상의 에피소드(A) 단편 에피소드(B) 두개의 집단으로 분리하여 평균 쿠키 집계

정규성 등분산성 확인

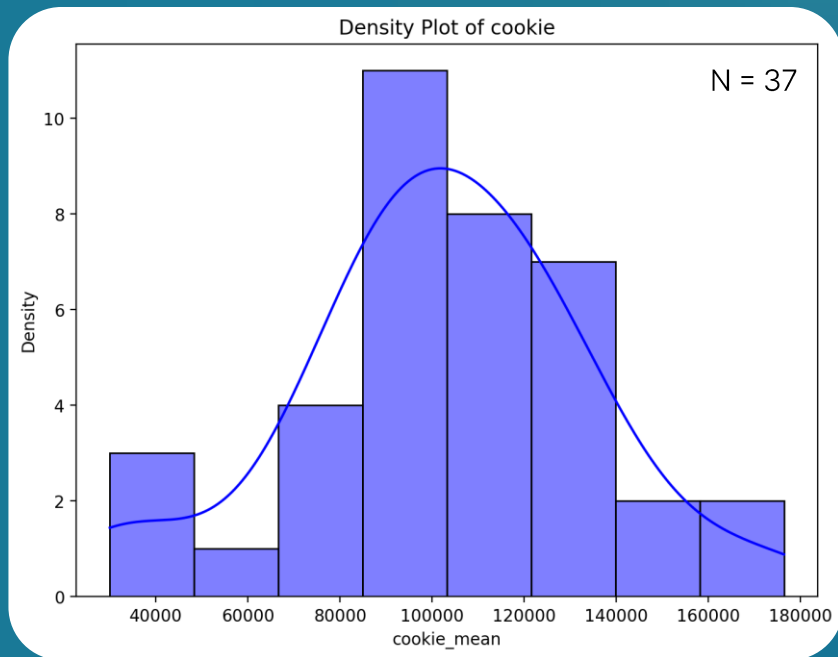
- 정규분포 그래프 확인
- Levene 등분산성 검정

독립 표본 T 검정

검정 결과

① 정규성 및 등분산성 확인

왜도가 심하지 않은 정규분포의 형태를 띄고 있었지만 Levenes p-value 값이 0.01 으로 등분산성을 만족하지 않았습니다.



이분산 검정 Welch's t-test 결정

② 독립표본 t 검정 (Welch's t-test)

Welch's t-test를 검정결과 시리즈 형태의 에피소드라고 쿠키 사용량이 높다고 볼 수는 없었습니다.

(T-statistic : -0.439 , P-value : 0.664)

아쉬웠던 점

단순히 에피소드의 형태 보다 '재미도', '장르반전', '서비스신', '시즌 마무리' 등 여러 요소들이 적용할 수 있다고 생각이 들었습니다.

에피소드 별 특징들을 더 구해 지표로 추가하여 분석을 진행하면 좋았을 것 같다는 아쉬움이 남았습니다.

웹툰 성과지표 분석 및 대시보드

2024.02 ~ 2023.03

Tool

Python

- Selenium
- Pandas
- Matplotlib, Seaborn
- Lifetimes

Streamlit

 아이콘 클릭 시 이동합니다.

Detail



블로그



깃허브



대시보드

Summary

평소에 즐겨보던 웹툰을 하나의 프로덕트로보고 성과 지표를 분석하여 대시보드로 구축 했습니다.

- 고객 활성화 지표의 기준을 '댓글'로 정하고 DAU, WAU, MAU, Stickiness, LTV를 산출했습니다.
- 밤에 비해 적은 오전,낮 시간대의 UV를 늘리기 위한 캠페인을 세워 보고 진행할 독자들을 선정했습니다.
- 유료결제가 많았던 에피소드의 특징이 무엇인지 가설을 세우고 분석했습니다.

Learnd

- CRM 데이터 분석에 대한 전반적인 이해
- Selenium 을 이용한 웹스크래핑
- 직접 가설을 세우고 검증하는 과정

