



MapleStory

DATA Mini Project

넥슨 Open API 및 웹 크롤링(Selenium) 활용
'메이플스토리'의 유저 이탈률에 대한 데이터 분석



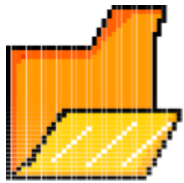
이동연

메이플스토리 하드유저



김민형

메이플 플레이 경험 0회



1 _____ 프로젝트 개요

목적, 필요성

2 _____ 데이터 수집 - 전처리

종류, 특성, 가공

3 _____ 데이터 분석

주요 지표, 변화 관찰

4 _____ 결과 및 인사이트

요약, 원인 도출

5 _____ 최종 결론, 개선방향

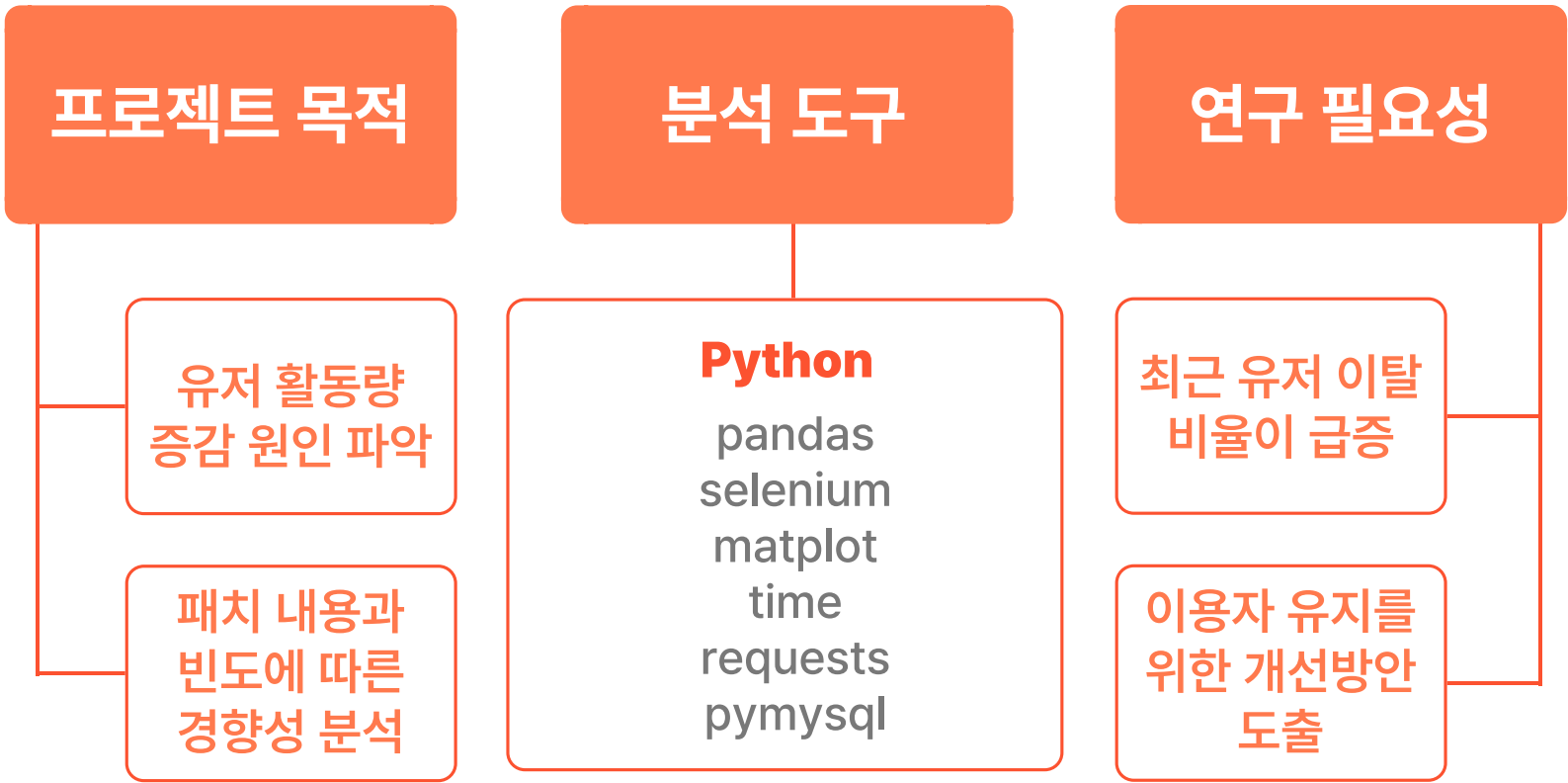
향후 연구, 한계



프로젝트 개요

평소 관심을 가지던 주제인 '게임'에 대한 데이터를 분석했습니다.
메이플스토리는 주기적으로 업데이트, 패치, 라이브 방송 등을 진행합니다.
게임 유저들이 이러한 게임사의 활동에 얼마나, 어떻게 영향을 받을까요?

| | | | | |
|---|---|--|---|---|
| 게임 출시 2D 횡 스크롤 대규모 다중 사용자 온라인 롤플레이밍 게 임(MMORPG)(넥슨) | 만화 '코믹메이플' 출간 어떤 프로젝트를 했는지 이 곳에 설명해주세요 더블클릭으로 수정해봐요 | 북미 시장 진출 총 110여 개의 국가에 출시, 약 1억 7천 명의 유저 확보 | 특수 서버 출시(리부트) 유저간의 거래가 제한되며 캐릭터 성장의 재미가 더욱 강화된 월드 | 확률조작 논란 발생 확률형 아이템의 확률을 소비자에게 불리하게 변경하고도 이를 누락하여 알리지 않고, 거짓으로 알린 행위 |
| 2003 | 2004 | 2006 | 2015 | 2021 |



온라인 언급량을 통한 유저 관심도 측정

메이플스토리는 오랜 역사를 가진 게임이지만, 최근 몇 년간 유저 이탈 문제가 대두되고 있습니다.
이번 프로젝트를 통해 뉴스, 커뮤니티 등의 유저 의견을 파악하고자합니다.

데이터 수집

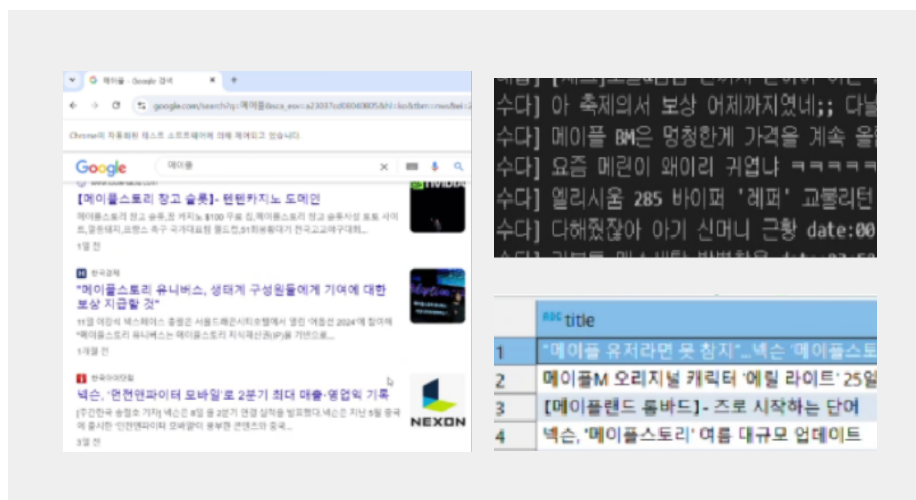
게임 이용자 관심도 측정에는 커뮤니티(메이플 인벤) 크롤링, 넥슨 오픈 API를 사용했습니다.
일반 대중들의 게임에 대한 관심에 대해서는 구글 트렌드 분석 / 구글 뉴스 크롤링을 진행했습니다.

유저 활동 패턴과 언급량

이탈률 => 게임 접속일의 연속성 & 경험치(누적 계) 상승량 비교

01. 웹 크롤링 (Selenium)

chromedriver를 제어하여 화면 조작, SQL 데이터 저장



02. Open API (Nexon)

게임에서 제공하는 오픈 API 를 등록, 관리, 전달



| \ | 언론 보도 정보 데이터 | 커뮤니티 인기글 조회수 | 네임드 유저(랭커) 접속량 |
|-----------|-------------------|---------------------|--------------------|
| 데이터 설명 | 키워드 검색 결과 보도기사 목록 | 올해 추천 30개 이상 인기글 정보 | 활동량, 경험치 누적계, 접속여부 |
| 데이터 수집 방법 | 웹 크롤링 (Google) | 웹 크롤링 (Maple Inven) | Nexon Open API |
| 데이터 수집 목적 | 대상 상상 세계 | 유저 반응 및 관심사 분석 | 고레벨 유저 플레이 패턴 분석 |

데이터 전처리

- 게임이 아닌 불필요한 정보를 별도로 구분
ex. 모바일 메이플, 아파트 단지 메이플 등)
- 날짜 별 구분: 일별 / 월별 / 연도(1월 1일 ~ 현재)
데이터 분류 --- matplotlib으로 시각화

4단계 처리 과정

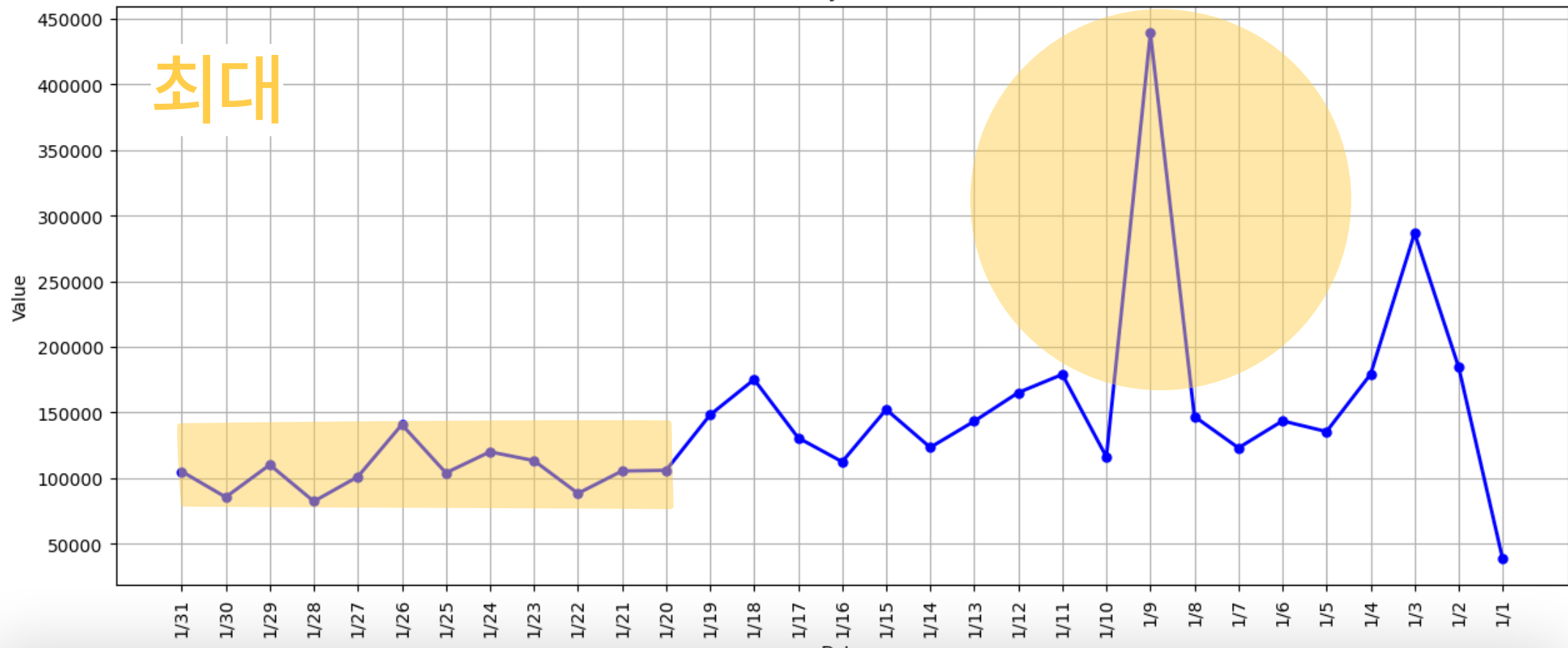
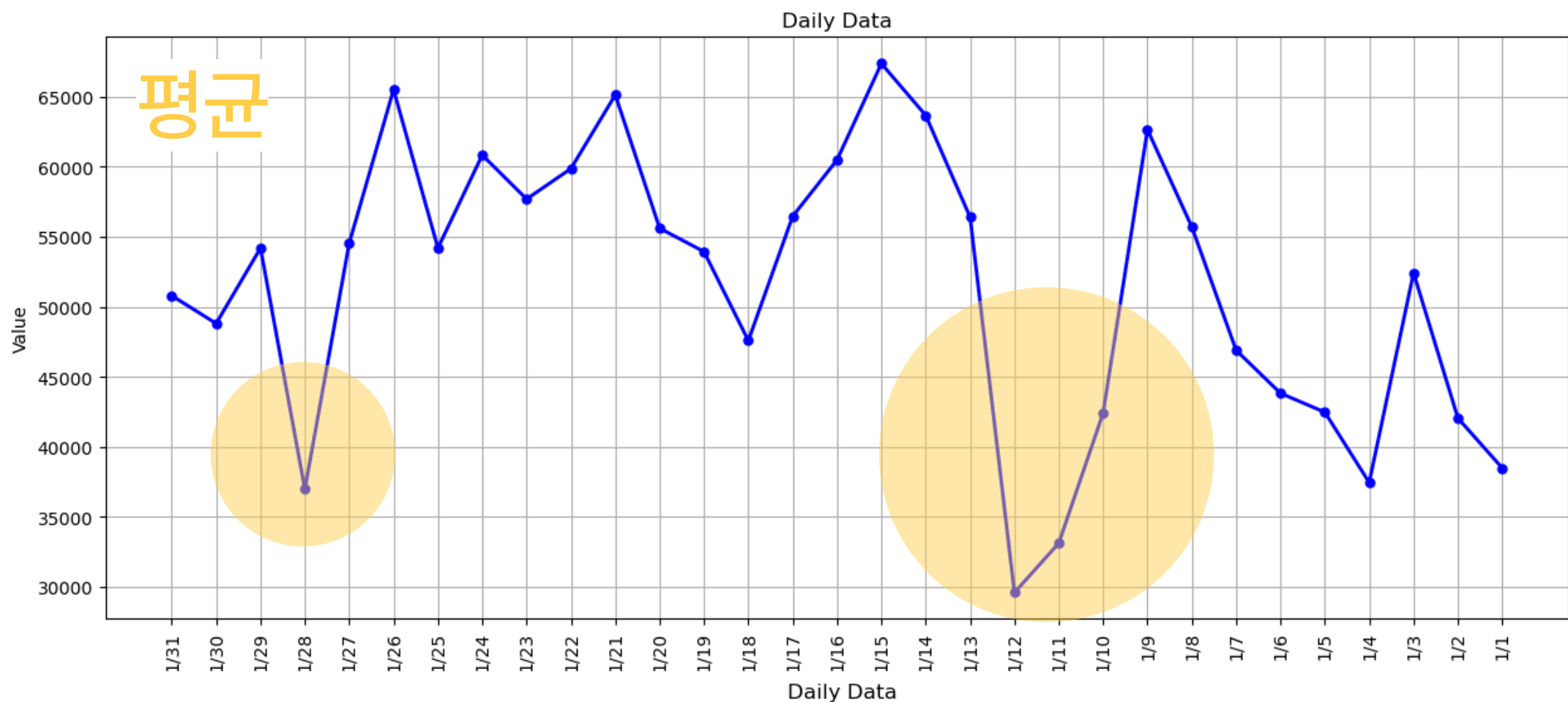
중복 데이터 제거, 평균 대체, 조건별 그룹화 등의 방법을 활용하여 데이터를 정리하고 분석의 정확성 향상



각 서버별로 유저의 활동 로그를 합산하여 집계 --- 결측치는 평균값으로 대체 --- matplotlib을 사용하여 유저 관심도에 대한 시각화

| | 시간 | 비고 |
|-------------|----|----|
| 다 많아 - 조선비즈 | | X |
| | | X |
| 성할 것" | | |
| | | |
| 표 | | |
| "역대 최대 규모" | | |
| 표 | | |

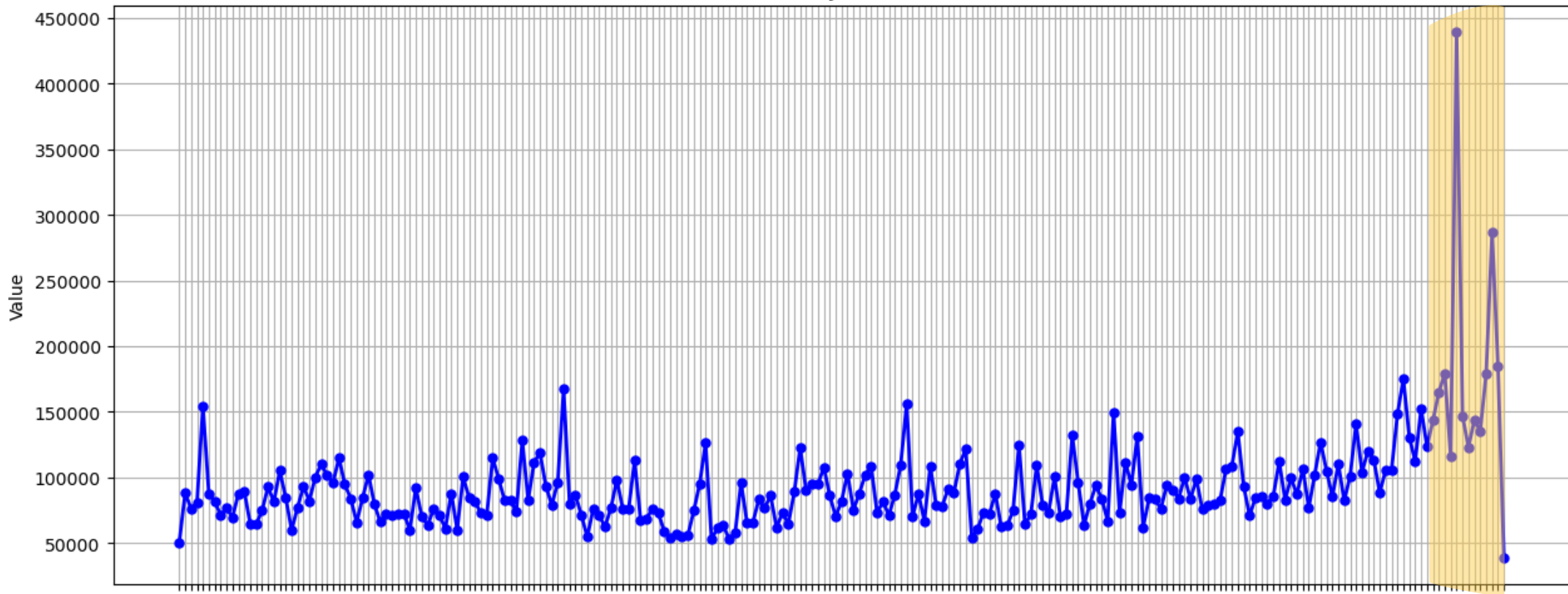
- 데이터의 정확성과 신뢰성을 확보
- 복잡한 데이터를 명확하게 전달할 수 있도록 정리
- 전처리 과정의 체계성 -> 효율성과 품질 향상



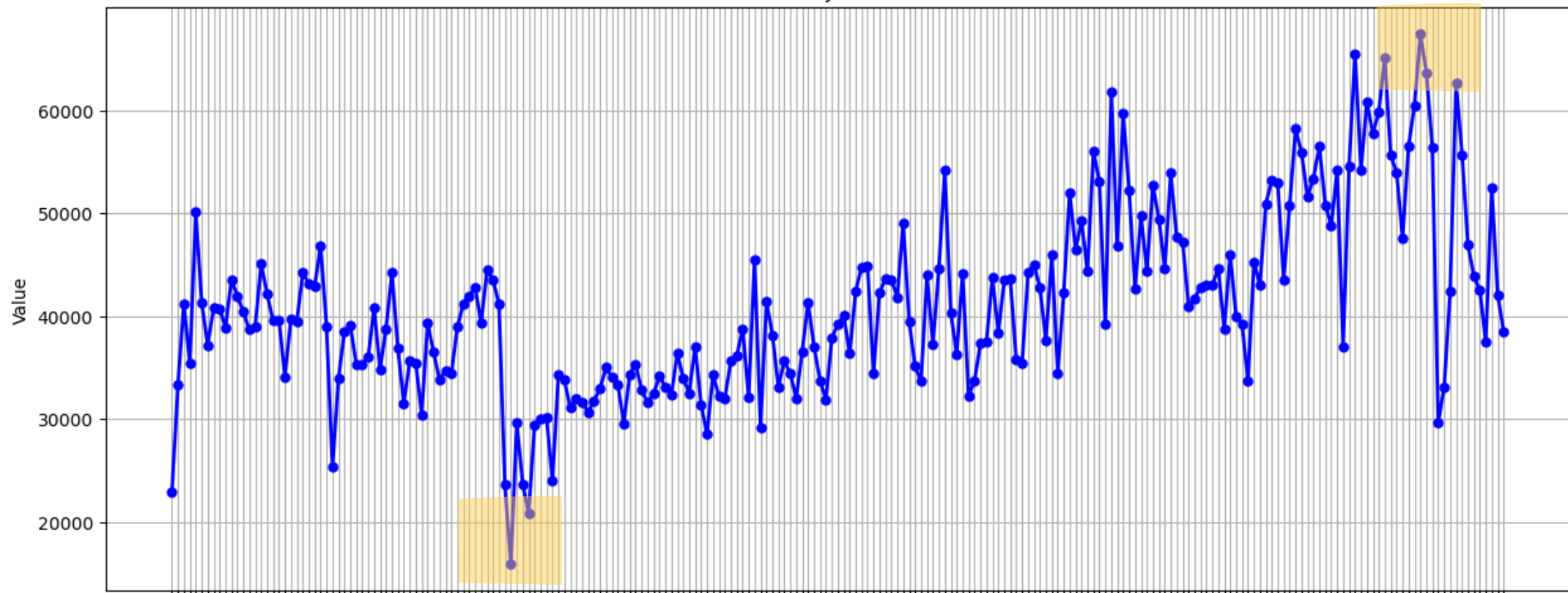
인벤 게시물 조회수

1월의 커뮤니티 인기 게시물 조회수 평균과 최대 수치입니다. 업데이트 및 라이브 방송이 있었던 1월 중순을 전후하여 변화폭이 크게 나타나고 있습니다. 또한 별도의 게임사 측 활동이 없는 시기에는 상대적으로 변화폭이 작습니다.

Daily Data

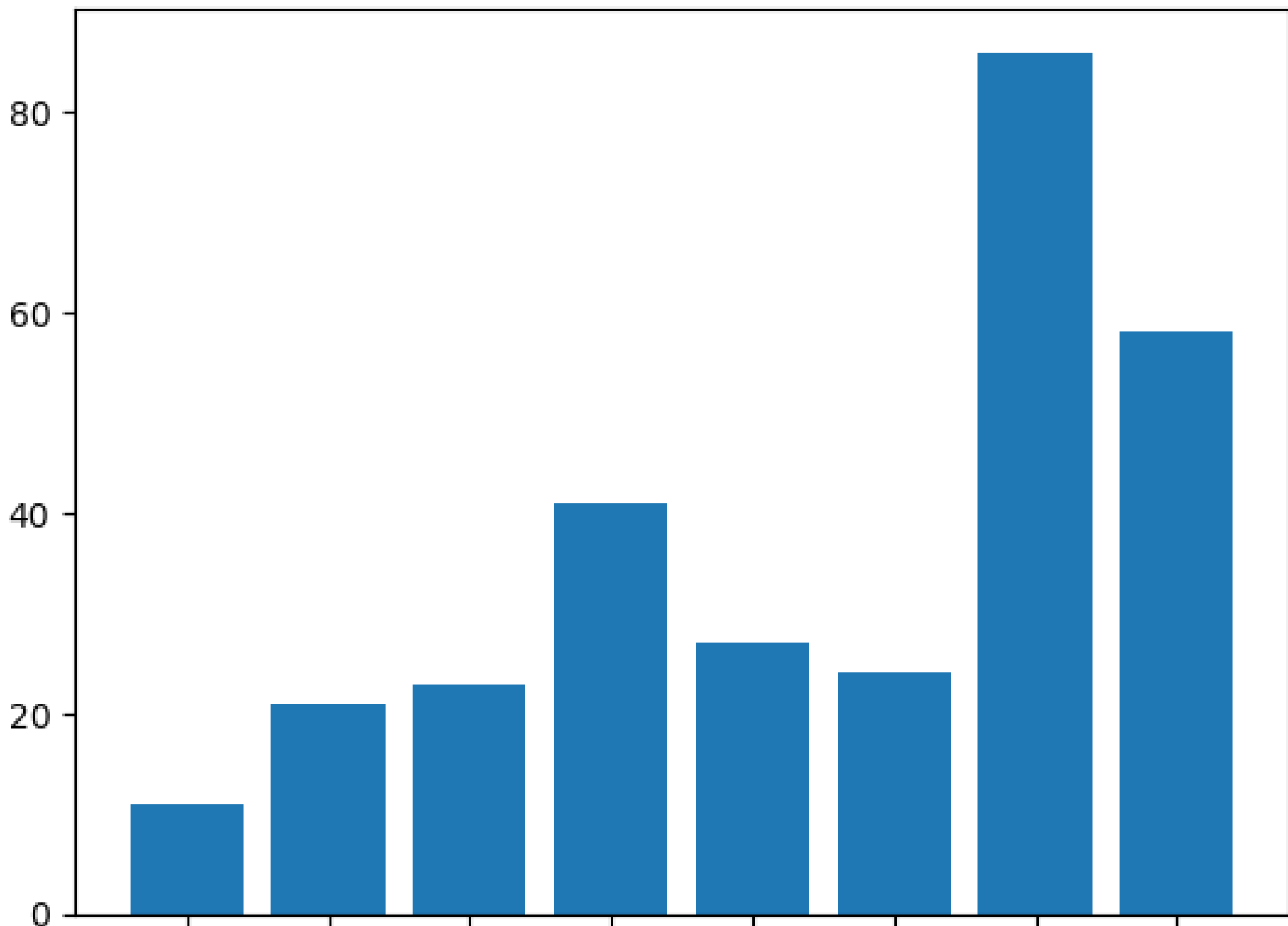


Daily Data



언론보도 언급량

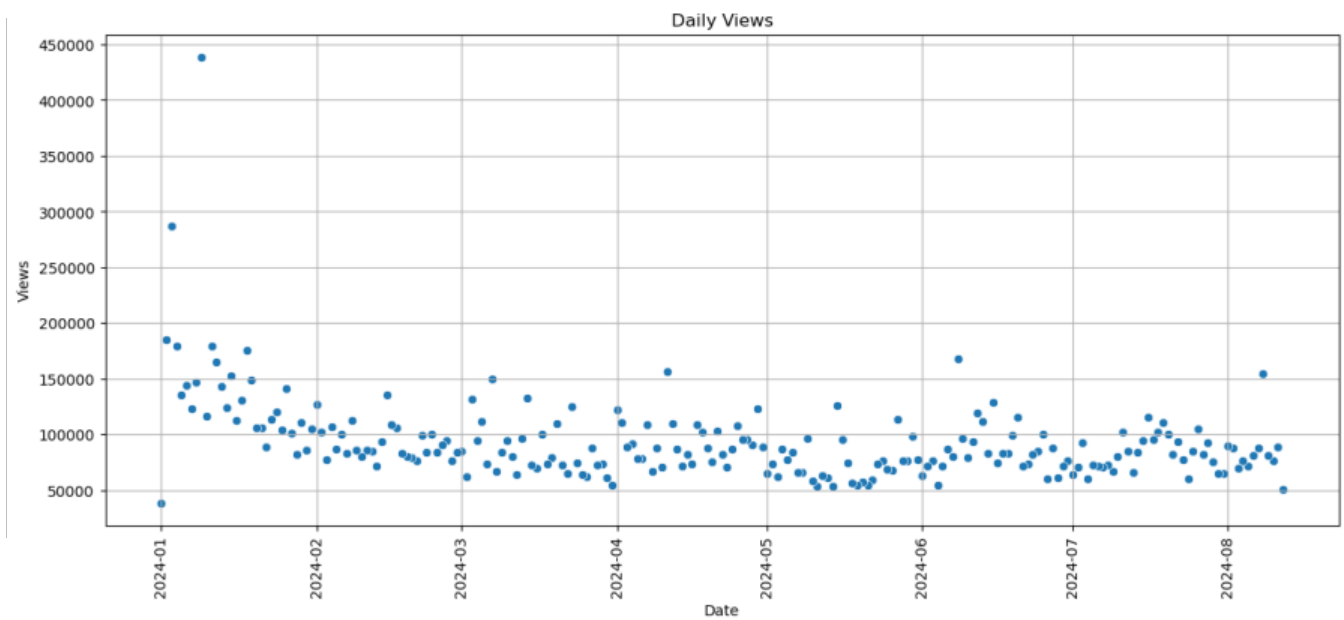
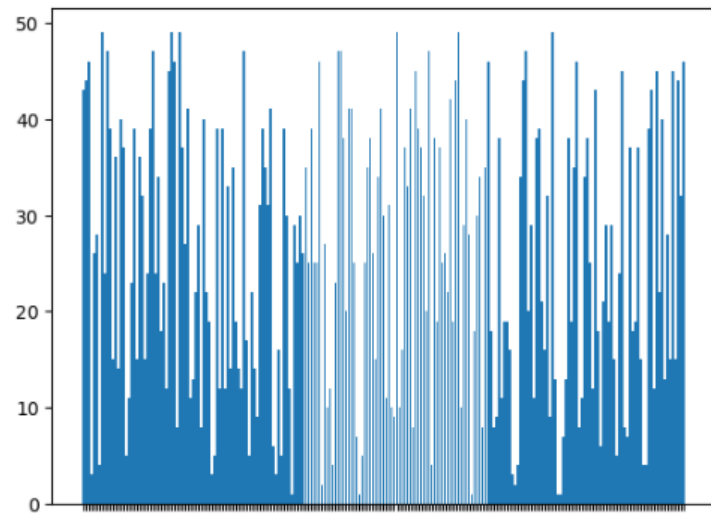
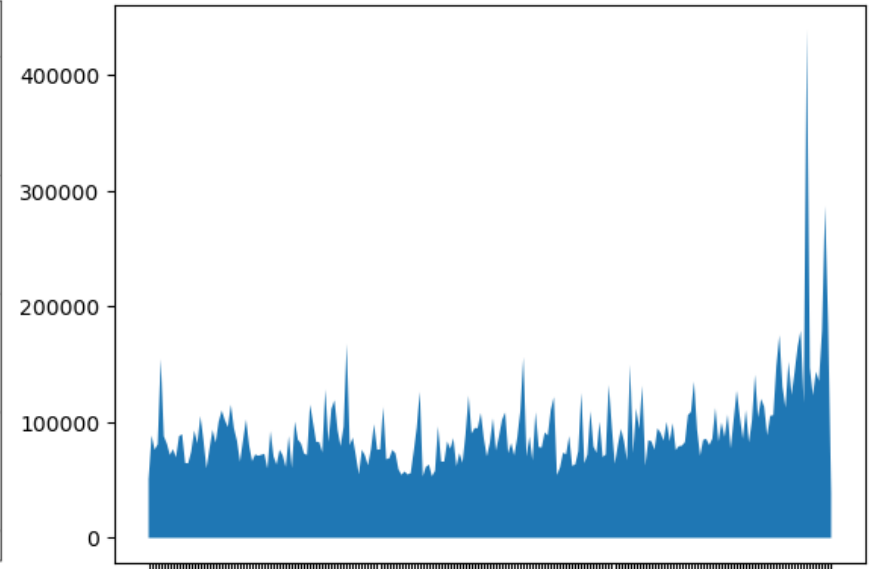
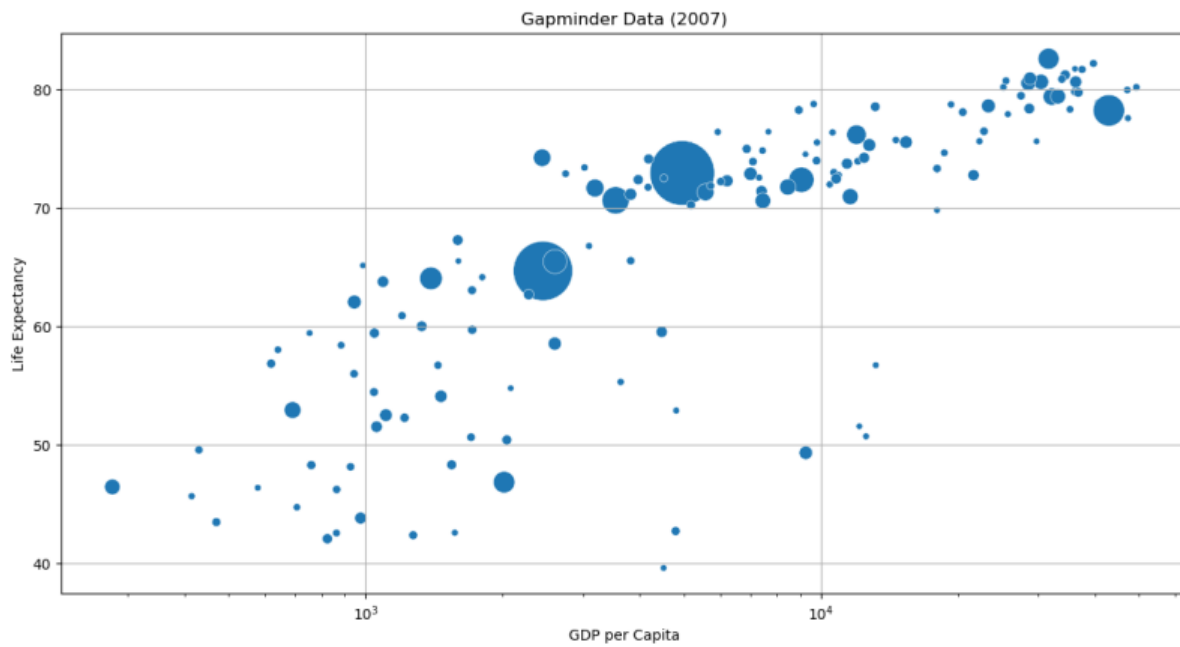
1월과 7월 등 방학 기간에 중점적으로 최고점에 도달함을 확인
업데이트를 알리는 게임사 라이브 방송 후, 유저 불만족으로
인한 게임 화제성 급감



언론보도 개수(월별)

뉴스의 월별 개수를 확인했을 때 업데이트를 앞둔 7월이 가장 많았습니다.

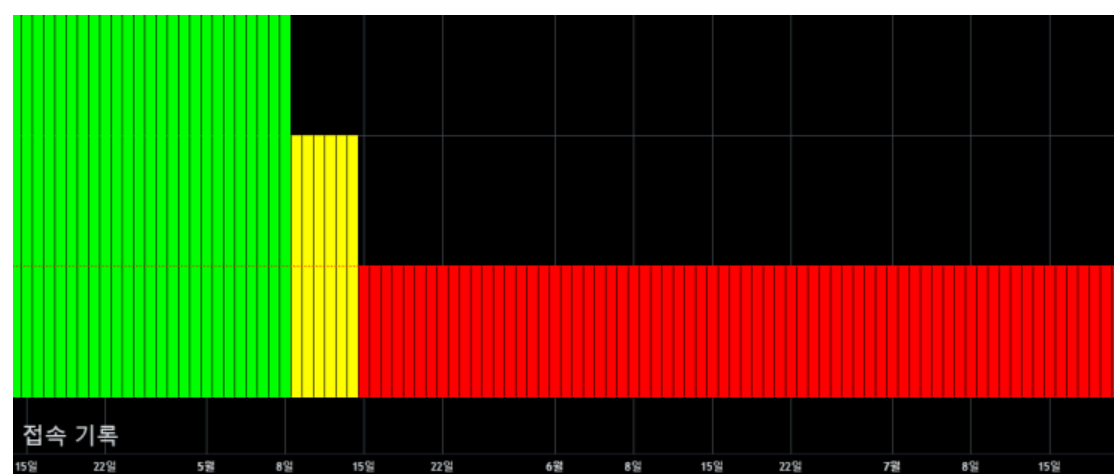
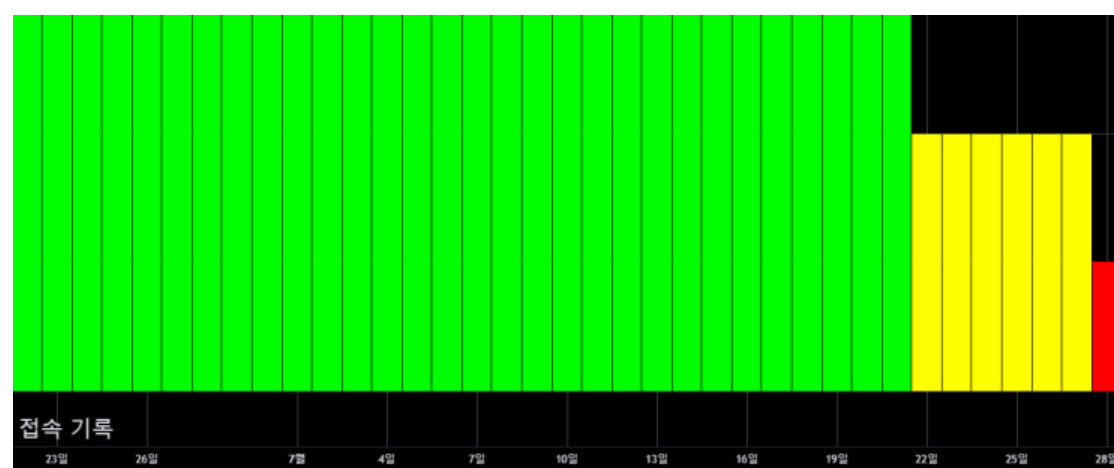
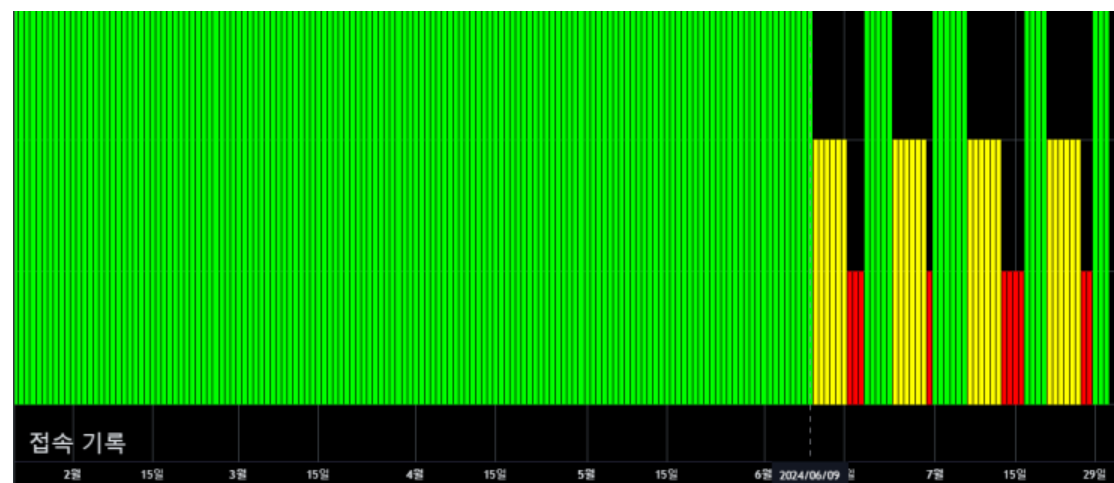
행-뉴스개수 열-월(1월 ~ 8월) / Histogram



그 외 시도해보았던 시각화 종류

Bubble Chart, Area Chart, 기본적인 산포도 등
여러 가지 방식을 적용해보았습니다.

정수 형태(개수, 조회수)의 데이터 이므로 꺾은 선과 히스토그램
방식을 최종적으로 선택했습니다.



랭커 접속빈도

초록색: 접속 O
 노란색: 접속했으나, 접속 시간이 짧아 플레이를 했는지는 불확실
 빨간색: 접속 X

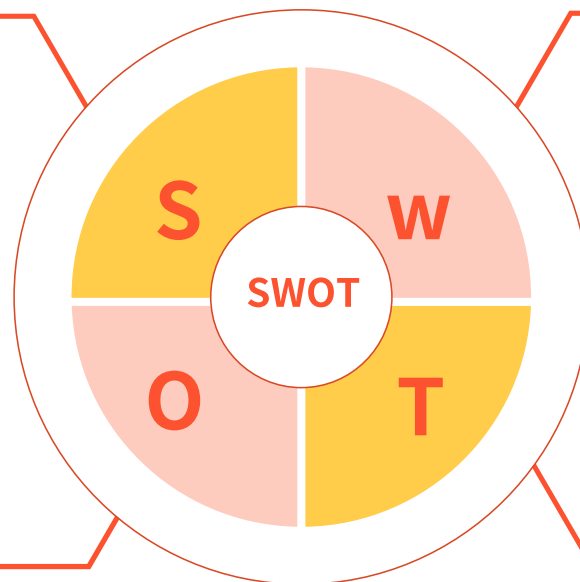
결론 및 향후 연구 방향

게임사의 공지사항, 패치 사항 등 신규 콘텐츠 업데이트는
유저 잔존률, 이탈 여부, 재접속 등에 영향을 유의미하게 줍니다.
하지만 새로운 사항에 대한 유저 만족도에 따라 오히려 언급량
및 게임 접속 빈도가 줄어드는 경우가 존재했습니다.

S strength

- 다양한 자료 수집 방법
- 최신 & 실시간 자료 분석
- SQL 데이터베이스에 Raw Data 저장
→ 정보 보관 및 가공 용이

SWOT CHART



weakness W

- 크롤링 과정 중 수집된 불필요한 정보를 수동으로 제거하는 작업에서 시간적 한계
- 수집된 데이터의 양적 부족 (약 1만 개)
- 제한적인 데이터일 가능성 우려

O opportunity

- 1년 이상, 더 긴 기간 동안의 자료수집 가능성
- 메이플스토리 데이터를 바탕으로 더 많은 종류의 게임의 자료 수집 가능성

threat T

- 이후 웹 크롤링 시 개인정보 침해, 저작권 문제가 발생할 수 있을 가능성 감안

- API, 웹크롤링 이외에도 **소셜 네트워크** 분석 등 그 외의 데이터 수집 방법을 적용할 수 있습니다.
- 이후 배우게 될 **머신러닝과 딥러닝, 시계열 데이터** 분석 이론 등을 향후 활용할 수 있습니다.
- 일정 기간동안만 크롤링해오는 것이 아닌 **실시간 모니터링** 방식으로 새롭게 구축할 수 있습니다.
- 시간적 여유가 있다면 웹페이지에서 확인할 수 있도록 구현하는 방법도 시도해보고 싶습니다.