

# OTT 로그 분석을 통한 콘텐츠 시청 제고 방안 도출

콘텐츠

Python

## 프로젝트 개요

### 데이터 개요

- OTT 플랫폼 이용자들의 시청 환경(국가, 디바이스 등) 및 콘텐츠 탐색 후 상호작용으로 이어지는 과정에 대한 데이터(자동추천 여부, 검색어, 검색 결과, 탐색에 대한 상호작용 등) 등에 대한 field로 구성됨.

### 분석 목표

- 각 field와 그 value들에 대한 정의가 명확한 데이터셋은 아니었던 바, 인터넷 검색 결과 참고와 약간의 추정을 통해 **데이터 정의 명확화** 작업부터 선행함.
- 정의 명확화 작업 이후 이를 기반으로 null값의 존재가 타당하지 않은 경우를 분류하고 해당 케이스에 속하는 row들을 제거하는 **전처리 작업**을 수행함.
- 전처리한 데이터를 검색 경로, 국가, 시간대 등 **다각적인 측면**에서 살펴봄. 특히 국가별 분석에서는 '(1)국가별 (2)상호작용별 (3)검색 경로'와 같은 구조로 막대그래프를 가지고도 최대한 **다차원적인 분석**을 시도함.

# OTT 로그 분석을 통한 콘텐츠 시청 제고 방안 도출

콘텐츠

Python

## 데이터 전처리

- 1) null값이 존재하는 field는 아래와 같이 2개였음. ([표 1])

① Query Typed [사용자 입력 텍스트]

② Displayed Name [사용자 입력 텍스트 기반 자동완성 텍스트]
- 2) 이들은 Section [검색 결과 생성 경로]이라는 field와 의미상 밀접한 관계가 있어 보였음. ([표 1])
- 3) Section에 속하는 고유값들의 정의가 명확하게 정의되어 있지 않았던 바 각각의 정의를 [표 2]와 같이 명확화한 뒤 분석을 진행함.
- 4) [표 2]의 정의에 따르면 [표 3]의 case들은 논리적으로 성립하지 않기에 제거함.
- 5) 4번 과정 후에도 null값이 모두 제거되지는 않았지만 논리 상 타당한 건들만 남았다고 판단하여<sup>1)</sup> 이렇게 전처리한 데이터를 기반으로 분석을 진행함.
- 1) e.g., 입력어(Query Typed)와 자동완성어(Displayed Name) 모두 null이지만 사전 자동추천 (Section=Prequery Results)으로 분류된 경우는 타당함

[표 1]

Query Typed	플랫폼에서 검색한 텍스트
Displayed Name	검색 결과에 표시된 텍스트
Section	결과가 검색 결과를 기반으로 생성되었는지 자동 생성되었는지 여부

[표 2]

Prequery Results	사용자가 검색어를 타이핑하기 이전에 OTT가 자체적으로 콘텐츠를 추천한 경우 (자동추천)
Suggestion Results	사용자가 검색어를 타이핑하는 도중 입력 중인 텍스트를 기반으로 OTT가 결과를 추천한 경우
Title Results	사용자가 타이핑을 완료하고 ‘검색’ 버튼을 누른 뒤 검색어를 기반으로 OTT가 관련 결과를 표출한 경우

[표 3]

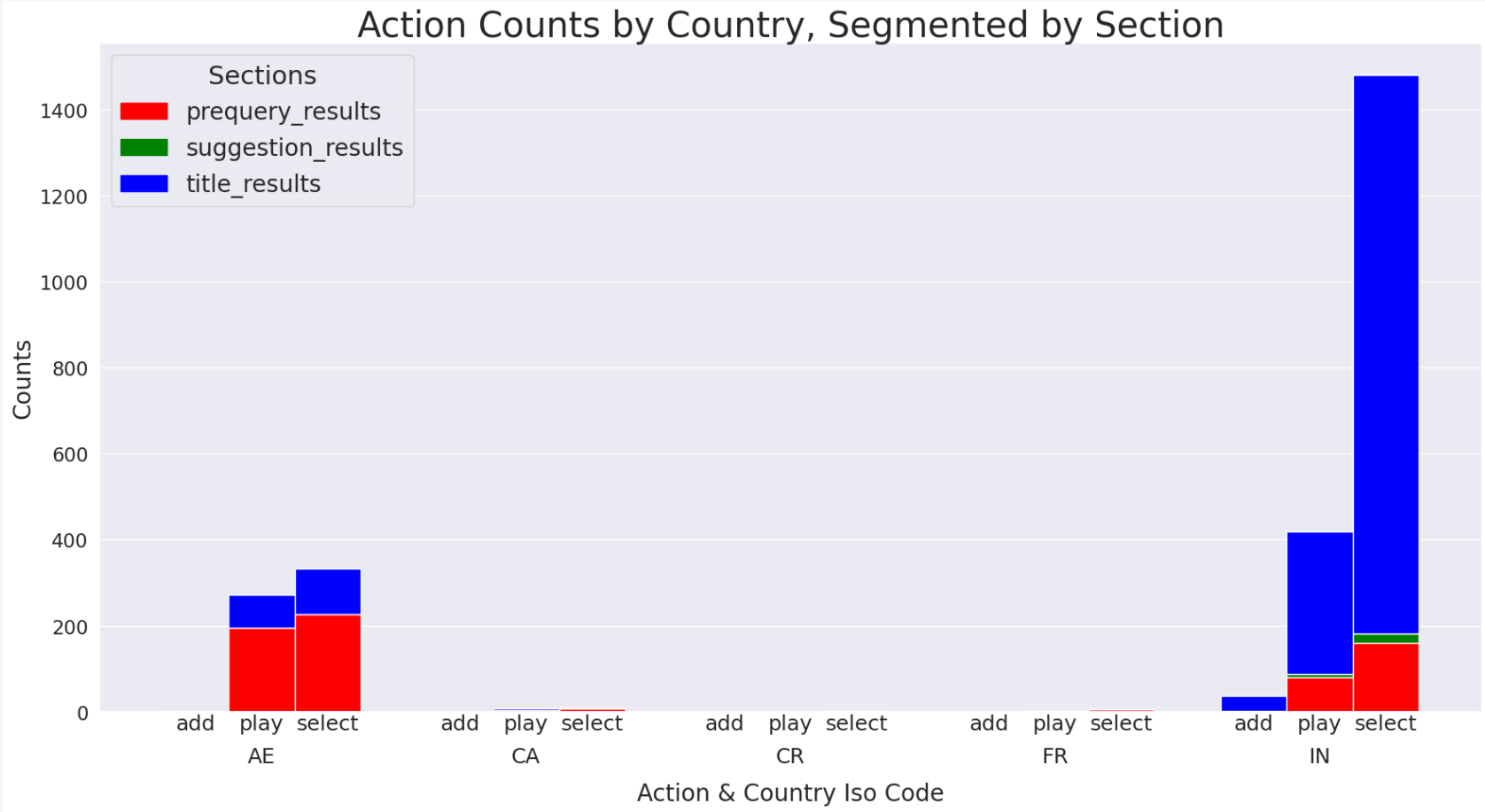
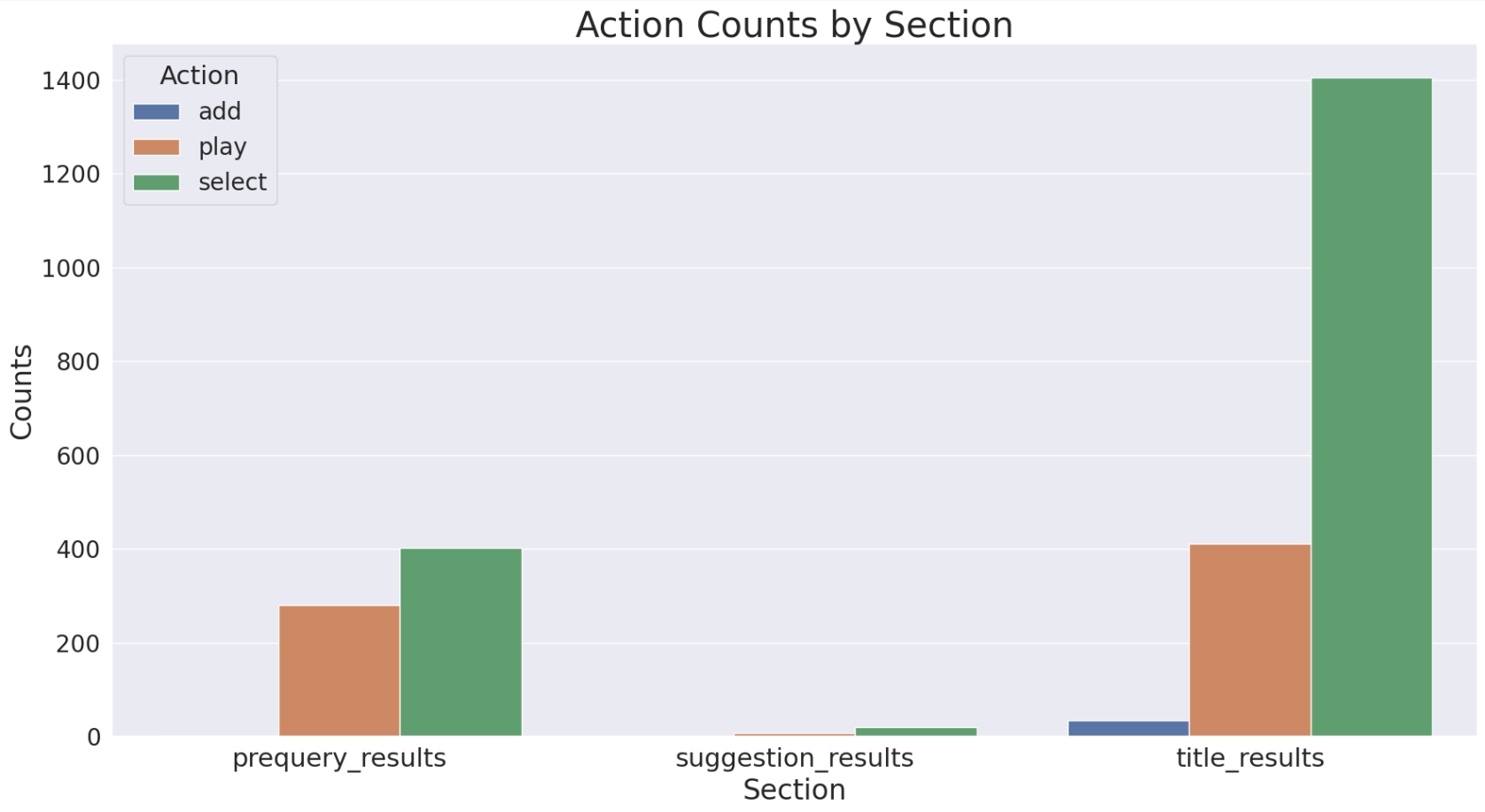
Case	Query Typed	Displayed Name	Section
1	Null	*	Suggestion Results
2	Null	*	Title Results
3	*	Null	Suggestion Results

Case 1 & 2: 사용자가 입력어를 타이핑 하지 않았는데(Query Typed=null) 입력어 기반으로 결과가 생성되었다고(Section) 분류된 경우

Case 3: 자동완성어 기록이 없는데(Displayed Name=null) 자동완성어 기반으로 결과가 생성되었다고(Section) 분류된 경우

# OTT 로그 분석을 통한 시청 제고 방안 도출

## 시각화 및 인사이트 도출



### 분석

예상과 달리 실제 시청(play)으로 이어진 비율은 맞춤 추천(prequery)의 경우가 직접 검색(title)의 경우보다 높음.

### 인사이트

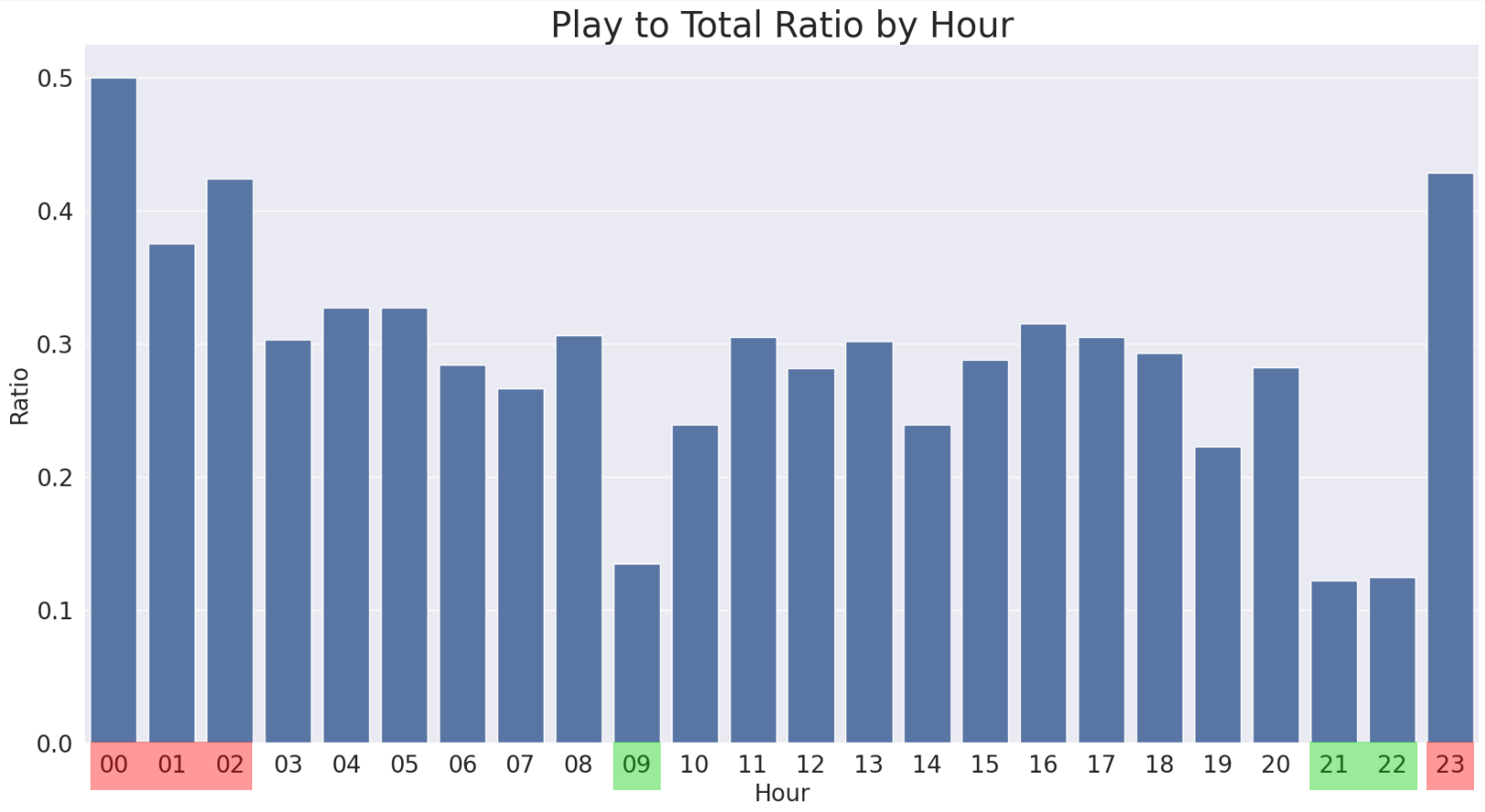
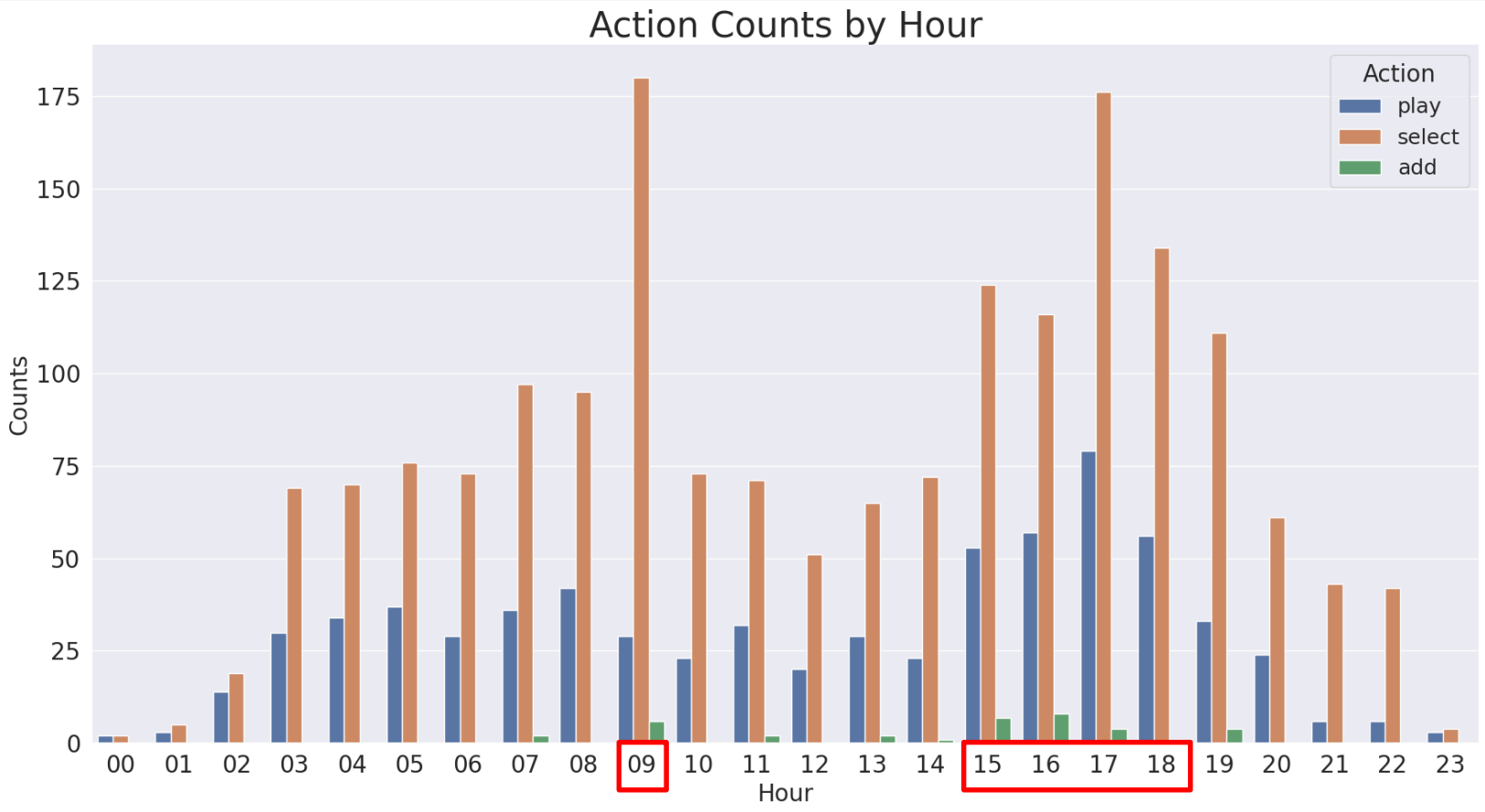
- (1) 해당 플랫폼의 맞춤 추천 알고리즘이 상당히 고도화되어 있을 가능성이 있음.
- (2) 검색어 기반으로 가장 적합한 콘텐츠들을 표출하는 알고리즘에 개선 여지가 존재할 수 있음.
- (3) 혹은 시청자들이 원하는 콘텐츠가 없을 가능성, 즉 작품 포트폴리오 개선이 필요한 상황일 수도 있음.

데이터를 국가별로 살펴 본 결과 나라별로 상호작용 행태가 다름을 알 수 있었음. 시청(play) 건 중 아랍에미리트(AE)는 자동추천(prequery)의 비중이 높고, 인도(IN)는 검색(title)의 비중이 높음.

- (1) 국가별로 다른 전략을 수립할 필요 있음.
- (2) 아랍에미리트: 검색 알고리즘 혹은 콘텐츠 포트폴리오 개선 여지 존재 가능.
- (3) 인도: 추천 알고리즘 개선 여지 존재 가능.

# OTT 로그 분석을 통한 시청 제고 방안 도출

## 시각화 및 인사이트 도출



**분석** 15~18시대에 시청으로 이어진 건수가 비교적 많음. 이 때 선택(select) 역시 활발한 경향을 보임. 09시는 높은 선택 건수와 그에 반해 굉장히 저조한 시청 건수가 눈에 띈다.

- 인사이트**
- (1) 대체로 하루 일과를 마치고 모든 상호작용(Action)이 활발하며 시청 건수 역시 다른 시간대에 비해 높음.
  - (2) 09시의 경우 출근 시간대에 콘텐츠 탐색은 활발하지만 시청까지 이어지기는 어려운 상황에 기인한 현상일 가능성이 있음.

모든 상호작용(play, select, add) 중 시청(play)이 차지하는 비율을 시간대별로 살펴봄. 09, 21, 22시가 눈에 띄게 낮고, 23~02시가 상대적으로 높음.

- (1) 출근 시간대 및 취침 준비 시간대에 콘텐츠를 집중하여 시청하기 어려운 상황에 기인한 현상일 수 있음. 반면 새벽 시간대에는 시청을 결심한 사용자들이 주로 접속한 것에 기인한 현상일 수 있음.
- (2) 앞선 분석들을 보면 추가(add, 관심 콘텐츠 저장) 기능 활용이 매우 저조함. 시청이 어려운 시간대에 추가 기능을 적극적으로 사용하도록 유인하는 전략을 수립하여, 시청이 용이한 시간대에 탐색에 소모되는 시간을 줄이고 시청률을 개선할 수 있을 것으로 보임.