분석에는 'Kaggle'에서 제공하는 ‘Trending YouTube Video Statistics'과 구글 API 통해 추가적으로 유튜브 영상 길이 데이터를 확보했다.

|  |  |
| --- | --- |
| 데이터 출처 | Trending YouTube Video Statistics  (https://www.kaggle.com/datasnaek/youtube-new) |
| 데이터 설명 | 이 데이터 세트에는 일일 인기 급상승 YouTube 동영상에 대한 몇 개월 (및 계산) 데이터가 포함됩니다.  데이터는 미국, GB, DE, CA 및 FR 지역 (각각 미국, 영국, 독일, 캐나다 및 프랑스)에 포함되며 최대 200 개의 트렌드 동영상이 하루에 나열됩니다.  확보한 영상 데이터는 총 202310개 입니다. |

**변수 이름과 설명:**

|  |  |  |  |
| --- | --- | --- | --- |
| **유튜브에 영상 업로드 전**  **알 수 있는 데이터** | | **유튜브에 영상 업로드 후**  **생성되는 데이터** | |
| 변수 이름 | 설명 | 변수 이름 | 설명 |
| title | 동영상의 제목 | video\_id | 유튜브 영상을 구분 짓는 고유한 id |
| channel\_title | 해당 동영상의 유튜브 채널 제목 | Trending\_date | 인기 급상승 동영상으로 선정된 날짜 |
| publish\_time | 동영상을 업로드 한 날짜 | views | 동영상의 조회수 |
| tags | 동영상을 설명하는 짧은 단어, 게시자가 작성한다. | likes | 동영상의 좋아요 수 |
| description | 영상을 설명하는 글 | dislikes | 동영상의 싫어요 수 |
| video\_length | 영상의 길이(시간) | comment\_count | 영상의 댓글 수 |
| category\_id | 동영상의 장르 카테고리를 코드로 표현한 변수  예) 10: 음악 / 17: 스포츠 | +  업로드 전에  알 수 있는 모든 변수 |  |
| **영상을 올리기 전,**  **영상의 조회수를 예측 할 수 있을까?** | | **조회수에 영향을 주는 변수는 뭘까?** | |

그렇다면 흥행이란 뭘 까?