









테이터로 알아본, 유튜브 분야 별 성공 공식

 Brightics Studio를 활용해,
유튜브 스타가 되려면 어떻게 해야 성공가능한지 알아보기

브라이틱스 서포터즈 권예진(Becky)-개인 분석 미션

Main Contents

-  주제 선정 이유
-  유튜브 데이터 수집 방법 - Google API 사용
-  세 분야의 유튜브 데이터 확보
-  전반적인 데이터 전처리
-  탐색적 데이터 분석(Exploratory Data Analysis, EDA)
-  유튜브 스타가 되려면? - Text Analysis, ML modeling 으로 궁금증 해결!

주제 선정 이유

성인남녀 10명 중 6명,
유튜버 꿈꾼다!

성인남녀 3,543명 설문조사 [자료 제공: 사람인]

유튜버 도전 의향 **있다 63%**

없다 37%

* 가장 하고 싶은 유튜브 분야

일상 31.1%

게임 13.9%

먹방 10.4%

요리 6.3%

음악 5.9%



사람인



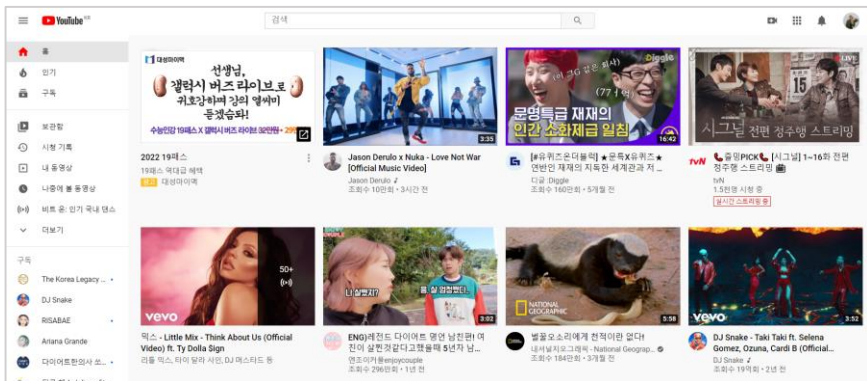
[서포터즈] 베키의 과격한 분석 상담소

Brightics TV

조회수 1.2천회 • 1개월 전

Brightics 를 사용해,
내가 하고 싶은 분석, 가장 “공모전”스럽지 않은 분석을 해보자!

유튜브 데이터 수집 방법



→ Google Developer에서 지원되는 **YouTube API**를 이용!

**Brightics의 Python Script를 사용해
데이터 확보 완료!!**(But, API할당량 문제 발생)



YouTube Data API v3

Google

The YouTube Data API v3 is an API that provides access to YouTube data, such as videos, playlists,...

API 호출과 데이터 저장을 위한
Python코드 작성

```
{
  "kind": "youtube#channelListResponse",
  "etag": "DfO5fBLIH9uEvrbKL2m9InR0wvQ",
  "pageInfo": {
    "resultsPerPage": 1
  },
  "items": [
    {
      "kind": "youtube#channel",
      "etag": "qvGooqsVapIGWfsjCRzyzb0liAo",
      "id": "UCdtRAcd3L_UpV4tMXCw63NQ",
      "statistics": {
        "viewCount": "450299484",
        "commentCount": "0",
        "subscriberCount": "2720000",
        "hiddenSubscriberCount": false,
        "videoCount": "527"
      }
    }
  ]
}
```

json 형식의 raw data

| | title | publishedData | views | likes | dislikes | comments |
|----|--------|------------------|----------|--------|----------|----------|
| 1 | 일반인이 | 2020-07-09T11:00 | 12697762 | 137720 | 5359 | 44921 |
| 2 | 전직 UDT | 2019-11-23T08:00 | 9257188 | 51007 | 1686 | 4915 |
| 3 | 전자가 되 | 2020-07-11T11:00 | 7632456 | 84712 | 2409 | 36467 |
| 4 | 110만 류 | 2019-11-04T08:30 | 6924710 | 43402 | 1156 | 6810 |
| 5 | 민간이 한 | 2020-07-16T11:00 | 6922181 | 85655 | 1804 | 36770 |
| 6 | 넌 30 | 2020-07-18T11:00 | 6022031 | 68991 | 1417 | 34838 |
| 7 | 그들의 하 | 2020-07-30T11:00 | 5305511 | 114845 | 1142 | 38145 |
| 8 | 주위 업 | 2019-01-20T10:30 | 5271933 | 57464 | 1458 | 3348 |
| 9 | 릭 이만 X | 2020-07-23T11:00 | 5201533 | 92045 | 1372 | 48420 |
| 10 | 가짜 사나 | 2020-09-03T11:00 | 5177609 | 72167 | 1469 | 27162 |
| 11 | UDT 선배 | 2020-06-17T11:30 | 4922299 | 33681 | 627 | 3836 |
| 12 | 복한 잠수 | 2019-09-02T09:00 | 4906345 | 49218 | 1254 | 6441 |
| 13 | UDT vs | 2019-09-28T08:00 | 4830644 | 35362 | 912 | 4192 |
| 14 | 사실자 발 | 2020-07-25T11:00 | 4821119 | 63777 | 1127 | 32776 |
| 15 | 50만명어 | 2018-06-04T12:00 | 4571407 | 149533 | 1142 | 3653 |
| 16 | 가짜 사나 | 2020-08-06T11:00 | 4441431 | 59361 | 954 | 10749 |

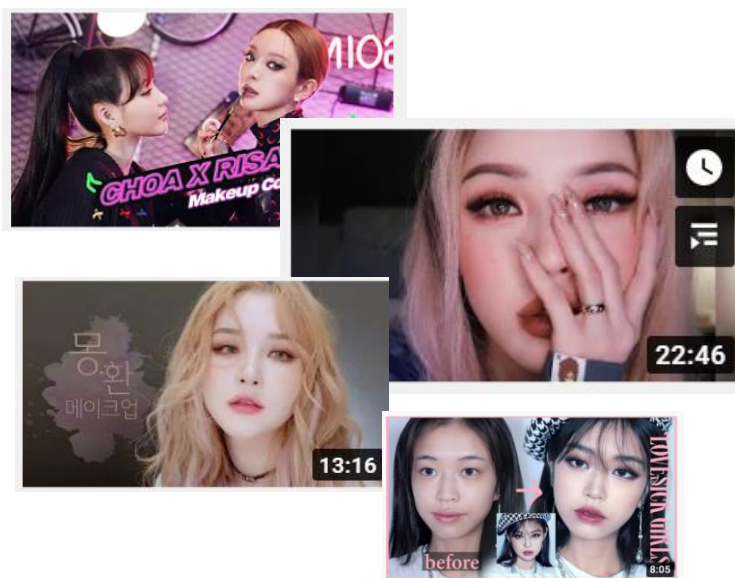
세 분야의 유튜브 데이터 확보

Health & Fitness 채널



구독자 25만명 이상의
16개의 유튜브 채널

Beauty 채널







구독자 55만명 이상의
10개의 유튜브 채널

Video Review 채널



구독자 58만명 이상의
6개의 유튜브 채널

▶ 전반적인 데이터 전처리

Table ▾ **Raw Data**    

| | Unnamed_0 | title | publishedData | views | likes | dislikes | comments | uploadgap | channel | subscribe | views_min_max |
|----|-----------|-----------|---------------|----------|--------|----------|----------|-----------|-------------|-----------|--------------------|
| 12 | 11 | 하루 7분 ... | 2017-11-20 | 1490586 | 19968 | 159 | 5455 | 7 | thankyou... | 2510000 | 0.0864621067994... |
| 13 | 12 | 집에서 하... | 2017-11-27 | 15326041 | 202531 | 2169 | 42169 | 7 | thankyou... | 2510000 | 0.9155383371177... |
| 14 | 13 | 칼로리 소... | 2017-12-04 | 751136 | 19765 | 197 | 1911 | 7 | thankyou... | 2510000 | 0.0772248174654... |
| 15 | 14 | 다리 얇아... | 2017-12-11 | 751136 | 155916 | 1078 | 24885 | 7 | thankyou... | 2510000 | 0.4699075300297... |
| 16 | 15 | 초간단! | 2017-12-18 | 5309231 | 73768 | 850 | 4831 | 7 | thankyou... | 2510000 | 0.3152907070532... |
| 17 | 16 | 무.조.건! | 2018-01-01 | 16735520 | 235389 | 3620 | 28520 | 14 | thankyou... | 2510000 | 0.9999999999999... |
| 18 | 17 | 팔독 다이... | 2018-01-15 | 10187357 | 158993 | 1155 | 27236 | 14 | thankyou... | 2510000 | 0.6076076792885... |
| 19 | 18 | 무.조.건! | 2018-01-23 | 8419534 | 103830 | 1188 | 11349 | 8 | thankyou... | 2510000 | 0.5016725995452... |
| 20 | 19 | 등살 & 허... | 2018-01-29 | 2406322 | 44205 | 385 | 3607 | 6 | thankyou... | 2510000 | 0.1413366989864... |

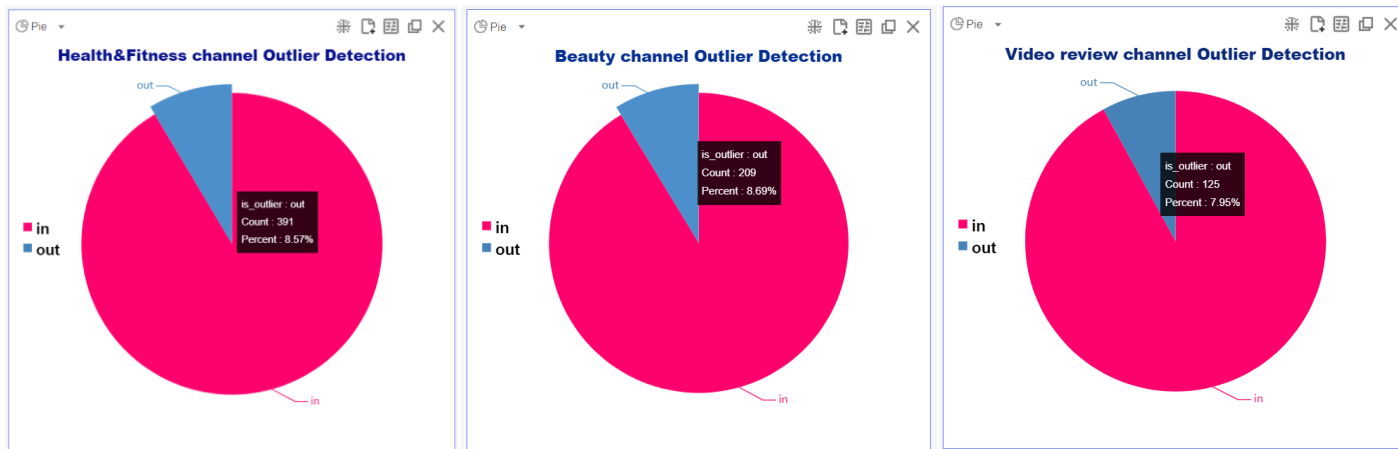
Creating New feature

Brightics Studio의
Python Script → uploadgap 변수 생성
Normalization 사용 → 조회수, 좋아요수,
싫어요수, 댓글수 min_max scaling
Add column 을 사용 → Hot 변수 생성
(분야 별로 전체 데이터의 조회수가 대략 상위 10%에 속하는 영상의 Hot 칼럼은 1)

| likes_min_max | dislikes_min_max | comments_min_max | like_ratio | hot |
|--------------------|----------------------|----------------------|------------|-----|
| 0.0766578880690... | 0.04392265193370166 | 0.08058749925519872 | 125 | 1 |
| 0.7775239747852... | 0.5991712707182321 | 0.6274801882857654 | 93 | 1 |
| 0.0758785635859... | 0.054419889502762435 | 0.027795984031460406 | 100 | 1 |
| 0.5985672714429... | 0.2977900552486188 | 0.3700172793898588 | 144 | 1 |
| 0.2831980712678... | 0.23480662983425415 | 0.07129237919323124 | 86 | 1 |
| 0.903667048010995 | 1 | 0.42416433295596734 | 65 | 1 |
| 0.6103799878686... | 0.31906077348066303 | 0.40503783590538045 | 137 | 1 |
| 0.3986071974263... | 0.32817679558011054 | 0.16838467496871834 | 87 | 1 |
| 0.1697046245038... | 0.106353591160221 | 0.05305964368706429 | 114 | 1 |

탐색적 데이터 분석(EDA)

확보한 모든 유튜브 영상 데이터에는 Nan이 존재하지 않았습니다.



8~9% 정도의 이상치가 존재,

〈Brightics의 Normality test 결과〉
데이터가 정규 분포를 따르지 않는다.
→정규성을 가정으로 하는 통계적 검증 무의미함!

Kolmogorov-Smirnov test result

| data | estimates | p_value |
|------------------|--------------------|---------|
| likes | 0.9778395871819515 | 0.0 |
| dislikes | 0.8572060564417441 | 0.0 |
| comments | 0.9449808796244488 | 0.0 |
| uploadgap | 0.7484641327060019 | 0.0 |
| subscribe | 1.0 | 0.0 |
| views_min_max | 0.5 | 0.0 |
| likes_min_max | 0.5 | 0.0 |
| dislikes_min_max | 0.5 | 0.0 |

탐색적 데이터 분석(EDA)

Statistic Summary를 사용해, 주요 변수들의 통계량을 확인!

업로드 간격이 거의 3~7일인데
이 정도 간격으로는 업로드 해야
유튜브 노출 알고리즘에 영향을
미치는 것 같습니다.

Health & Fitness 채널

| | column_name | max | min | range | sum | avg | stddev | median | iqr |
|---|-------------|----------|-----|----------|-------------|-------------|------------|--------|--------|
| 1 | views | 19429101 | 0 | 19429101 | 16574770... | 363083.6... | 1053319... | 66584 | 234797 |
| 2 | likes | | | | | | | | |
| 3 | dislikes | | | | | | | | |
| 4 | comments | | | | | | | | |
| 5 | uploadgap | | | | | | | | |
| 6 | subscribe | | | | | | | | |
| 7 | hot | | | | | | | | |

Beauty 채널

| | column_name | max | min | sum | avg | stddev | median | iqr |
|---|-------------|----------|-------|-------------|-------------|-------------|----------|-----------|
| 1 | views | 22515524 | 10708 | 12881455... | 535834.2... | 960039.7... | 297814.5 | 407472.25 |
| 2 | likes | 318243 | 0 | 21428903 | 8913.853... | 18621.72... | 4275.5 | 6322 |
| 3 | dislikes | 11822 | 0 | 439535 | 182.8348... | 410.9590... | 100 | 136 |
| 4 | comments | | | | | | | |
| 5 | uploadgap | | | | | | | |
| 6 | subscribe | | | | | | | |
| 7 | HOT | | | | | | | |

Video Review 채널

| | column_name | max | min | sum | avg | stddev | median | iqr |
|---|-------------|----------|--------|-------------|-------------|-------------|----------|------------|
| 1 | views | 32518069 | 2267 | 15042408... | 956896.2... | 1546028... | 511141.5 | 1027651.25 |
| 2 | likes | 54895 | 0 | 10066425 | 6403.578... | 7364.043... | 3961.5 | 6810.25 |
| 3 | dislikes | 17904 | 0 | 498635 | 317.1978... | 603.9752... | 170 | 319 |
| 4 | comments | 21997 | 0 | 1715651 | 1091.381... | 1539.591... | 633 | 1163.5 |
| 5 | uploadgap | 125 | 0 | 7779 | 4.948473... | 5.579567... | 4 | 4 |
| 6 | subscribe | 1150000 | 600000 | 12883200... | 819541.9... | 160532.1... | 860000 | 166000 |
| 7 | hot | 1 | 0 | 185 | 0.117684... | 0.322234... | 0 | 0 |

유튜브 스타가 되려면?

Tokenizer(Korean)로
토큰화하기

| title |
|---------------------------------|
| 단기간 옆구리살 빼는 최고의 운동 BEST5 [집에... |
| 하루 7분! 집에서 하는 유산소운동 BEST5 |
| 3분! 집에서 하는 골반교정 스트레칭 BEST5 |
| 다리살 허벅지살 빼는 최고의 운동 BEST5 |
| 뱃살 빼는 최고의 운동 BEST5 |
| 흰다리 오다리 교정 운동 BEST5 |
| 저희는 운동을 좋아하는 평범한 부부예요! |
| 전신 다이어트 유산소운동 [홈트레이닝] |
| 9분! 초간단 누워서하는 11자 복근운동 |
| 히트! 10분! 디리리리! 사리는 운동 [히버지오더] |

| tokenized_title |
|--|
| ["단기간","옆구리","살","빼다","최고","운동","...] |
| ["하루","분","집","하다","유산소운동"] |
| ["분","집","하다","골반","교정","스트레칭"] |
| ["다리","살","허벅지","살","빼다","최고","운동"] |
| ["뱃살","빼다","최고","운동"] |
| ["흰","다리","오다","교정","운동"] |
| ["저희","운동","좋아하다","평범하다","부부"] |
| ["전신","다이어트","유산소운동","홈","트레이닝"] |
| ["분","초","간단","눕다","서하","자","복근","운동"] |
| ["히트","10분","디리리리","사리는","운동"] |

Stopwords Remover로
불용어 제거

| stopwords_tokenized_title |
|-------------------------------------|
| ["단기간","옆구리","살","빼다","최고","운...] |
| ["하루","분","집","하다","유산소운동"] |
| ["분","집","하다","골반","교정","스트레칭"] |
| ["다리","살","허벅지","살","빼다","최고","...] |
| ["뱃살","빼다","최고","운동"] |
| ["흰","다리","오다","교정","운동"] |
| ["저희","운동","좋아하다","평범하다","부...] |
| ["전신","다이어트","유산소운동","홈","트...] |
| ["분","초","간단","눕다","서하","자","복근...] |
| ["히트","10분","디리리리","사리는","운동"] |

유튜브 스타가 되려면?

제목 속 단어, 조회수, 댓글수, 좋아요수, 싫어요수가 주어졌을 때,
HOT한 영상인지 구분할 수 있는 분류기를 모델링
(Health & Fitness 채널 데이터로 ML 모델링 진행)

Term-document Matrix 를 사용해
모든 단어 라벨 인코딩 해서 칼럼으로 사용



Split Data로 train : test = 8 : 2로 나누기



Naive Bays Classification 진행

Evaluate Classification Result

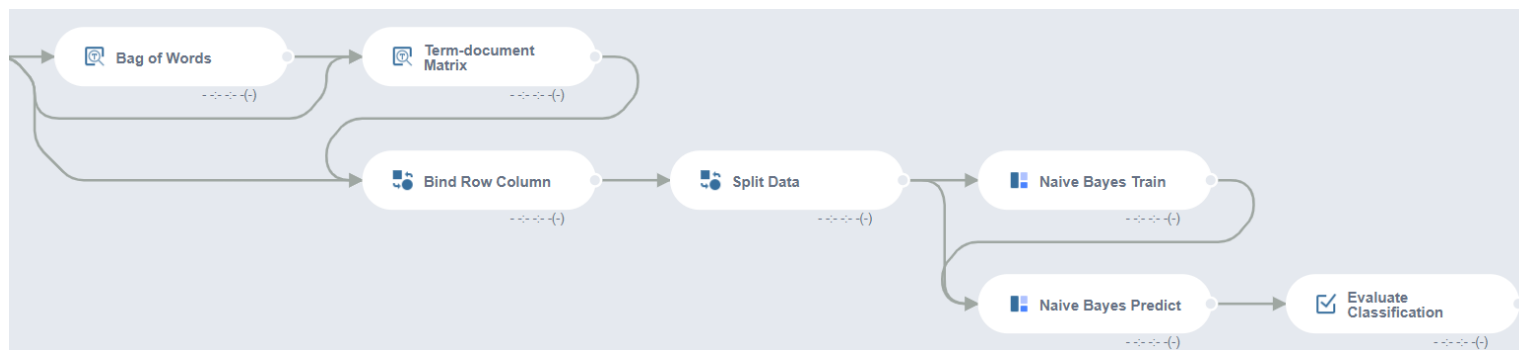
Accuracy : 0.9474260679079957

Metrics

| label | f1 | precision | recall |
|-------|--------------------|--------------------|--------------------|
| 0.0 | 0.9694656488549619 | 0.9947780678851175 | 0.9454094292803971 |
| 1.0 | 0.811023622047244 | 0.7006802721088435 | 0.9626168224299065 |

Table

| | | hot | stopwords_tokenized_title | | 단기간 | 러브 | 배기 | 배다 | 살 | 옆구리 | 운동 | |
|----|-----|-----|---|--------|-----|----|----|----|---|-----|----|--|
| 1 | 18 | 0 | ["단기간", "옆구리", "살", "배다", "최고", "운... | doc_0 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | |
| 2 | 41 | 0 | ["하루", "분", "집", "하다", "유산소운동"] | doc_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 69 | 0 | ["분", "집", "하다", "골반", "교정", "스트레칭"] | doc_2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | 26 | 1 | ["다리", "살", "허벅지", "살", "배다", "최고", "...] | doc_3 | 0 | 0 | 0 | 1 | 2 | 0 | 1 | |
| 5 | 37 | 1 | ["뱃살", "배다", "최고", "운동"] | doc_4 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | |
| 6 | 68 | 0 | ["힘", "다리", "오다", "교정", "운동"] | doc_5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 7 | ull | 0 | ["저희", "운동", "좋아하다", "평범하다", "부... | doc_6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 8 | 54 | 1 | ["전신", "다이어트", "유산소운동", "홍", "트... | doc_7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | 87 | 1 | ["분", "초", "간단", "늘다", "서하", "자", "복근... | doc_8 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 10 | 04 | 1 | ["하루", "땀", "분", "다리", "라인", "살리다", "...] | doc_9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 11 | 98 | 1 | ["상체", "다이어트", "운동", "팔뚝", "살", "겨... | doc_10 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | |
| 12 | 25 | 1 | ["하루", "분", "전신", "운동", "뱃살", "파괴", "...] | doc_11 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |



[결론] 유튜브 스타 되기 _ Health & Fitness 채널

Tip! [업로드 간격] 3~4일

Tip! [제목 속 KEY 단어] 효과, 자극, 단기, 필수

Tip! [HOT한 콘텐츠 소재] 홈 트레이닝, 식단, 신체의 특정 부분 운동

Tip! [목표 조회수] 3개월 마다 300만 이상의 영상 조회수 찍는 콘텐츠 생성

Tip! [50만 구독자를 향한 여정] 근 3년간의 꾸준한 활동

[결론] 유튜브 스타 되기 _ Beauty 채널

Tip! [업로드 간격] 5~7일

Tip! [제목 속 KEY 단어] 최애, 인생템, 꿀팁, 준비

Tip! [HOT한 콘텐츠 소재] 화장품 추천, 계절 별 메이크업, 케어

Tip! [목표 조회수] 6개월 마다 500만 이상의 영상 조회수 찍는 콘텐츠 생성

Tip! [100만 구독자를 향한 여정] 근 5년간의 꾸준한 활동

[결론] 유튜브 스타 되기 _ Video Review 채널

Tip! [업로드 간격] 4~5일

Tip! [제목 속 KEY 단어] 무조건, 떡밥, 최초

Tip! [HOT한 콘텐츠 소재] 마블 코믹스, 내용 총정리, 무비 해석

Tip! [목표 조회수] 2개월 마다 500만 이상의 영상 조회수 찍는 콘텐츠 생성

Tip! 다른 분야의 채널보다 단기간에 구독자를 늘리기 쉬운 분야



Thank you for watching!
Please tell me if you have any questions!