



YouPlace

카프카와 스파크를 활용한
유튜브 영상 속 제주 명소 검색

BOAZ 16기 엔지니어링 고은서 류정화 송경민

INDEX



- ➡ 주제 선정 배경
- ➡ 데이터 파이프라인 소개
- ➡ 데이터 설명 및 전처리 과정
- ➡ 구현 과정 설명
- ➡ 시연 및 개선할 점

Click

YouTube

INDEX



16기 엔지니어링 고은서
중앙대학교
소프트웨어학부



16기 엔지니어링 류정화
성신여자대학교
융합보안공학과



16기 엔지니어링 송경민
국민대학교
소프트웨어학과

[유튜피아] 동영상 골리앗?...이젠 '검색'도 유튜브 시대 밀레니얼 세대, 동영상으로 검색하고 소통하는 시대 주 도

검색 시장에서의 유튜브 영향력도 돌풍 수준이다. 미디어랩 나스미디어가 국내 PC, 모바일 인터넷 이용자의 서비스 행태를 분석한 결과, **조사 대상자 중 60%가 유튜브에서 정보를 검색하는 것으로 조사됐다.** 특히 10대의 경우에는 10명 중 7명(69.6%)이 유튜브를 검색 채널로 이용하고 있는 것으로 나타났다.

텍스트와 이미지를 넘어 자연스럽게 동영상으로 소통하고 동영상으로 검색하는 시대가 이미 현실로 다가온 것이다.

2030세대 67.9% '브이로그' 선호

※ 2030대 성인남녀 1,117명 대상 조사, 자료: 잡코리아×알바몬

Q. 브이로그 콘텐츠 선호 이유 (*복수응답)

- 1위 이미지 보다 생생하게 담아낼 수 있어서, 35.1%
- 2위 동영상 콘텐츠가 더 익숙해서, 34.7%
- 3위 공감하고 소통하기 유용해서, 31.5%
- 4위 다른 사람 일상을 보며 대리만족, 29.3%
- 5위 영상으로 내 일상을 기록, 23.5%



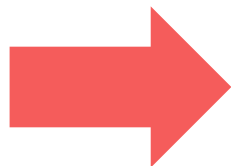
JOBKOREA × albamon

〈관련 뉴스 기사 및 조사결과〉

1. 주제 선정 배경 - 문제점 파악



여행을 준비하기에 앞서
유튜브 영상을 찾아보는
경우가 많음



영상 속의 장소정보들을 확인하려면,
네이버 등 다른 검색엔진 플랫폼을
재서치해야하는 불편함이 있음

유튜브 영상, 장소 세부 설명, 지도 등
필요한 정보들이 여러 매체에
분산되어 있음

Ex 예시

제주 브이로그 시청 -> 영상 속 다양한 장소 정보 획득 -> naver에서 장소 검색 -> 장소의 위치 검색

1. 주제 선정 배경 – 아이디어 좁히기



“ 유튜브 영상의 정보들은 매우 효과적이지만,
필요로 하는 정보들을 한 눈에 파악할 수 있는 플랫폼이 있었으면 좋겠다 ! ”



제주 여행 브이로그
data 활용

WHY

영상 검색을 통해 국내 여행지에 대한 사전 조사를 하는 경우가 많고,
특히 제주 브이로그 콘텐츠에 대한 수요와 공급이 모두 큼
따라서 유튜브를 통한 정보 수집이 수월할 것으로 예상

1. 주제 선정 배경 – 관련 레퍼런스 및 차별점



인스타그램 '지도검색' 기능



인스타그램 '지도검색' 기능 도입..."맛집 탐색 더 쉽게"



“국내 이용자들이 인스타그램을 통해 여러 장소를 적극적으로 탐색하고, 방문한 후기를 공유하는 이용 형태에 주목하여, 더욱 편리하게 장소를 발견하고자하는 바람으로 제작 – 코리아 대표”



1. 인스타그램 속 정보 < 유튜브 속 정보
⇒ 훨씬 다양한 검색 기준 확보 가능
2. 브이로그 영상에 초점
⇒ 유튜브 영상 속 모든 명소를 한 눈에 확인 가능



2. 데이터 파이프라인



유튜브 API(+카카오API) → 카프카 → 스파크 SQL → DB → 스파크 SQL → 장고 (웹)



- 주 활용 데이터

Youtube API, Kakao Map API

- 기술 스택

Cloud AWS EC2, Docker

Engineering Kafka, Spark

DB & Web MYSQL(RDS), Django

3. 데이터 설명 및 전처리 과정 - 데이터 설명



• Youtube API 데이터 수집

Search:list Search API를 통해서 동영상, 채널, 재생목록의 목록 등을 조회 가능

Video:list Search API만으로는 조회수 등과 같은 통계 정보를 받아올 수 없음!

→ Search API에서 얻어온 video ID로 Videos API를 호출하여 자세한 정보 추출

```
{
  "kind": "youtube#searchResult",
  "etag": "etag",
  "id": {
    "kind": "string",
    "videoId": "string",
    "channelId": "string",
    "playlistId": "string"
  },
  "snippet": {
    "publishedAt": "datetime",
    "channelId": "string",
    "title": "string",
    "description": "string",
    "thumbnails": {
      (key): {
        "url": "string",
        "width": "unsigned integer",
        "height": "unsigned integer"
      }
    },
    "channelTitle": "string"
  }
}
```

〈Search:list의 JSON 형태〉

```
{
  "kind": "youtube#video",
  "etag": "etag",
  "id": "string",
  "snippet": {
    "publishedAt": "datetime",
    "channelId": "string",
    "title": "string",
    "description": "string",
    "thumbnails": {
      (key): {
        "url": "string",
        "width": "unsigned integer",
        "height": "unsigned integer"
      }
    },
    "channelTitle": "string",
    "tags": [
      "string"
    ],
    "categoryId": "string"
  },
  "contentDetails": {
    "duration": "string",
    "dimension": "string",
  }
}
```

〈Video:list의 JSON 형태〉

YouTube > Data API

홈	안내	참조	샘플	지원
▶ Channels				
▶ ChannelSections				
▶ Comments				
▶ CommentThreads				
▶ GuideCategories				
▶ PlaylistItems				
▶ Playlists				
▼ Search				
개요				
list				
▶ Subscriptions				
▶ Thumbnails				
▶ VideoAbuseReportReasons				
▶ VideoCategories				
▼ Videos				
개요				
list				
insert				
update				
rate				
getRating				

3. 데이터 설명 및 전처리 과정 - 데이터 설명



- 유튜브 일일 할당량

GCP의 'IAM 및 관리자 메뉴 - 할당량(Auotas)' 에서 할당량을 볼 수 있음 (일간 할당량 : 10,000)

<input type="checkbox"/>	서비스	할당량	측정기준(예: 위치)	한도	현재 사용량 비율 ↓	7일 최고 사용량 비율
<input type="checkbox"/>	YouTube Data API v3	Queries per day		10,000	<div><div></div></div> 2.4%	<div><div></div></div> 2.4%

Quota costs		
search	list	100
subscriptions	list	1
	insert	50
	delete	50
thumbnails	set	50
videoAbuseReportReasons	list	1
videoCategories	list	1
videos	list	1
	insert	1600
	update	50
	rate	50
	getRating	1
	reportAbuse	50
	delete	50

search:list, video:list 모두 maxResult 인자의 max 값: 50개

Search:list → $100(\text{cost}) * 99(\text{호출횟수}) = 9900$

Videos:list → $1(\text{cost}) * 99(\text{호출횟수}) = 99$

$9900 + 99 = 9999 < 10,000$ 이므로

$50 * 99 = \text{약 } 5\text{천개}$

⇒ 일일 총 뽑을 수 있는 영상 : 약 5000개

Kakao API로 장소 데이터 수집

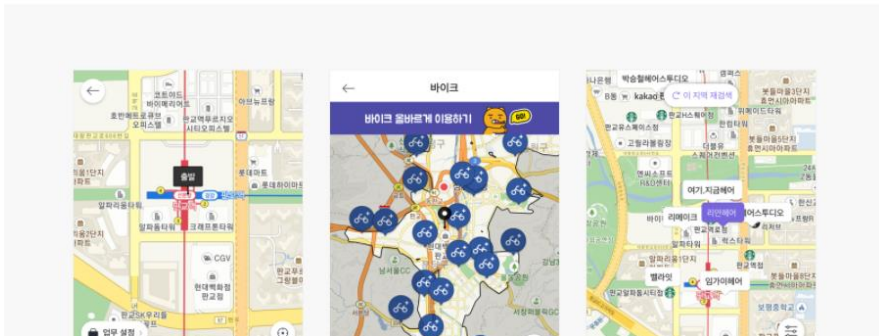
Youtube API를 사용하여 얻은 데이터들에서 추출한 장소를 Kakao API에 넣어 상세한 장소 정보 수집

지도/로컬



주요 특징

지도 SDK는 카카오맵의 기능을 웹과 앱 플랫폼에 구현할 수 있도록 다양한 메서드와 라이브러리를 통해 제공합니다. 로컬 API는 REST API 방식으로 카카오맵의 콘텐츠 및 데이터를 제공합니다. 지도SDK와 로컬API를 통해 다양한 위치 기반 서비스를 만들어보세요.



기능 소개

로컬(local) API는 키워드로 특정 장소 정보를 조회하거나, 좌표를 주소 또는 행정구역으로 변환하는 등 장소에 대한 정보를 제공합니다. 특정 카테고리로 장소를 검색하는 등 폭넓은 활용이 가능하며, 지번 주소와 도로명 주소 체계를 모두 지원합니다.

기능 목록

Name	Description
주소 검색	특정 주소에 해당하는 장소 정보의 좌표 검색
좌표-행정구역정보 변환	좌표 값을 행정동, 법정동으로 변환 맛집, 날씨 등 위치에 맞는 정보 제공 서비스에 활용
좌표-주소 변환	특정 좌표의 지번 주소 및 도로명 주소 제공
좌표계 변환	특정 체계의 좌표 값을 다른 체계의 좌표 값으로 변환 지원되는 좌표계: WGS84, WCONGNAMUL, CONGNAMUL, WTM, TM, KTM, UTM, BESSEL, WKTM, WUTM
키워드 검색	키워드로 관련 장소 및 상세 정보 검색 각 장소 상세 페이지 URL 제공
카테고리 검색	카테고리로 관련 장소 및 상세 정보 검색 각 장소 상세 페이지 URL 제공

3. 데이터 설명 및 전처리 과정 - 전처리 과정



1. Youtube API를 통해 데이터 수집

```
data = {'etag': 't7Yv8zy5-zo7vXOG85uuCixm5_0',
'items': [{ 'contentDetails': { 'caption': 'false',
'contentRating': {},
'definition': 'hd',
'dimension': '2d',
'duration': 'PT7M59S',
'licensedContent': False,
'projection': 'rectangular'},
'etag': 'yB5MCGkvAd2w3ELjIuJYmFoH-QM',
'id': 'Nk9ltFZIojc',
'kind': 'youtube#video',
'snippet': { 'categoryId': '22',
'channelId': 'UCu14PdSURK20iNkTvF01mPQ',
'channelTitle': '훈이네 일상',
'description': '안녕하세요 오랜만에 영상 업로드네요 ㅎㅎ\n'
'이번영상은 친형과 겨울 제주도를\n'
'다녀온 vlog를 만들어 보았어요!\n'
'\n'
'영상을 보면서 제주도 여행을 계획하시는 분들께\n'
'조금이나마 도움이 되었으면 좋겠네요!',
'liveBroadcastContent': 'none',
'localized': { 'description': '안녕하세요 오랜만에 영상 업로드네요 ㅎㅎ\n'
'이번영상은 친형과 겨울 제주도를\n'
'다녀온 vlog를 만들어 보았어요!\n'
'\n'
'영상을 보면서 제주도 여행을 계획하시는\n'
'분들께\n'
'조금이나마 도움이 되었으면 좋겠네요!',
'title': '겨울 제주도여행 | JEJU Vlog | 제주\n'
'브이로그 | 고라니 출연 | 감귤체험 | 비밀의숲 |\n'
'숙성도 | 풍차해변도로'},
'publishedAt': '2022-01-09T14:31:35Z',
```

	id	title	description	publishTime	channelTitle	likeCount	viewCount
0	Nk9ltFZIojc	겨울 제주도여행 JEJU Vlog 제주 브이로그 고라니 출연 감...	안녕하세요 오랜만에 영상 업로드네요 ㅎㅎ 이번 영상은 친형과 겨울 제주도를 다녀온 ...	2022-01-09T14:31:35Z	훈이네 일상	12	184
1	FcvBZLIPcEI	Jeju Island Travel Vlog, Exploring Jeju-do in ...	2021 7월의 제주도 여행 브이로그 WnThanks for watching thi...	2021-07-31T12:15:50Z	ChaeCine - Korean Vlog in Indonesia	59	2374
2	sV2x0VR13rQ	[vlog] 제주도 4박 5일 여행 브이로그 Jeju vlog 첫브이로그 영...	- instagram : @n_m.duri	2022-01-14T11:51:25Z	남돌 NAMDUL	1	18
3	X78E3bp5-0s	[JEJU VLOG] ep02. 제주도 비오는날 갈만한곳 빛의 병커 성산카페 에곤 ...	안녕하세요! 레일라입니다. 이번 영상은 제주도 여행 둘째 날 이야기입니다! Wn통제...	2021-05-18T00:29:20Z	레일라leila	7	250
4	xsZJWeelt7U	[JEJU VLOG] 제주도 서쪽 소품샵투어Top5 (애월읍도드리, 서쪽가게, 베리제주...		2021-11-06T13:47:14Z	림쓰RIMS	9	266

2. title, description에서 영어,숫자 제거 & 특정 단어 삭제('제주', '브이로그' 등) & 띄어쓰기 기준으로 쪼개기

```
[['겨울', '제주도여행', '제주', '브이로그', '고라니', '출연', '감귤체험', '비밀의숲', '숙성도', '풍차해변도로'],
['제주', '여름', '제주', '뚜벅이', '함덕골목', '고요새', '삼양해수욕장', '나이롱책방', '제주당집', '바바호미'],
['히든클리프에서', '호캉스', '제주', '흑돼지', '맛집', '숙성도', '동백', '가득한', '카멜리아힐', '아르떼', '미술관', '제주', '오마카세', '추천', '코하쿠'],
['떠나요', '제주도', '제주도', '브이로그', '제주', '명월국민학교'],
['행복한', '가족의', '제주', '여행']]
```

3. 데이터 설명 및 전처리 과정 - 전처리 과정



3. Kakao API를 활용하여 장소 검색

```
[ '비밀의숲',
  [ { 'address_name': '제주특별자치도 제주시 구좌읍 송당리 2173',
      'category_group_code': 'AT4',
      'category_group_name': '관광명소',
      'category_name': '여행 > 관광,명소 > 숲',
      'distance': '',
      'id': '82389937',
      'phone': '',
      'place_name': '비밀의숲',
      'place_url': 'http://place.map.kakao.com/82389937',
      'road_address_name': '',
      'x': '126.75493209633262',
      'y': '33.45360917744684' },
    { 'address_name': '강원 강릉시 사천면 판교리 1-13',
      'category_group_code': 'AD5',
      'category_group_name': '숙박',
      'category_name': '여행 > 숙박 > 펜션',
      'distance': '',
      'id': '1445063111',
      'phone': '010-6224-4730',
      'place_name': '비밀의숲',
      'place_url': 'http://place.map.kakao.com/1445063111',
      'road_address_name': '강원 강릉시 사천면 이설당길 108-3',
      'x': '128.865325212739',
      'y': '37.8413845352401' },
    { 'address_name': '서울 마포구 서교동 366-8',
      'category_group_code': 'CE7',
      'category_group_name': '카페',
      'category_name': '음식점 > 카페 > 테마카페',
      'distance': '',
      'id': '2048640934',
      'phone': '02-3144-2223',
      'place_name': '비밀의숲 홍대점',
      'place_url': 'http://place.map.kakao.com/2048640934',
      'road_address_name': '서울 마포구 잔다리로6길 30',
      'x': '126.920840923273',
      'y': '37.5522160605285' },
    { 'address_name': '경북 안동시 상아동 423',
      'category_group_code': '',
      'category_group_name': '',
      'category_name': '여행 > 공원',
      'distance': '',
      'id': '36022109',
      'phone': '054-840-3433' }
```

4. 주소가 '제주특별자치도'가 아닌 데이터 삭제 & 검색결과 가장 첫 장소만 남기고 삭제 & 관련 카테고리 데이터만 추출

```
[ [ '감글채름', { 'address_name': '제주특별자치도 서귀포시 남원읍 한남리 480-28', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 관광농원', 'di',
  [ '비밀의숲', { 'address_name': '제주특별자치도 제주시 구좌읍 송당리 2173', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 숲', 'distance': '',
  [ '속성도', { 'address_name': '제주특별자치도 제주시 노형동 941-2', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 한식 > 육류,고기', 'distance': '', 'id',
  [ '풍차해변도로', { 'address_name': '제주특별자치도 제주시 한경면 신창리 1322-1', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 드라이브코스',
  [ '여름', { 'address_name': '제주특별자치도 제주시 구좌읍 세화리 1397-6', 'category_group_code': '', 'category_group_name': '', 'category_name': '가정,생활 > 문구,사무용품', 'distance': '', 'id',
  [ '제주', { 'address_name': '제주특별자치도 서귀포시 서귀동 316-1', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 도보여행 > 둘레길 > 제주둘',
  [ '우박이', { 'address_name': '제주특별자치도 제주시 한경면 고산리 3594-5', 'category_group_code': 'AD5', 'category_group_name': '숙박', 'category_name': '여행 > 숙박 > 게스트하우스', 'distance': ',
  [ '함덕굴뚝', { 'address_name': '제주특별자치도 제주시 조천읍 함덕리 1041-16', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 한식 > 해장국', 'distance': ',
  [ '고요새', { 'address_name': '제주특별자치도 제주시 삼양동 1569-25', 'category_group_code': 'CE7', 'category_group_name': '카페', 'category_name': '음식점 > 카페', 'distance': '', 'id': '166',
  [ '삼양해수욕장', { 'address_name': '제주특별자치도 제주시 삼양동 1960-4', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 해수욕장,해변', 'id',
  [ '니이홍책방', { 'address_name': '제주특별자치도 제주시 삼양동 1646-11', 'category_group_code': '', 'category_group_name': '', 'category_name': '문화,예술 > 도서 > 서점', 'distance': '', 'id',
  [ '제주닭집', { 'address_name': '제주특별자치도 서귀포시 표선면 표선리 600-2', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 치킨', 'distance': '', 'id',
  [ '바바호미', { 'address_name': '제주특별자치도 제주시 조천읍 함덕리 924', 'category_group_code': 'AD5', 'category_group_name': '숙박', 'category_name': '여행 > 숙박 > 게스트하우스', 'distance': ',
  [ '히든클리프에서', { 'address_name': '제주특별자치도 서귀포시 상미동 625', 'category_group_code': 'AD5', 'category_group_name': '숙박', 'category_name': '여행 > 숙박 > 호텔', 'distance': '', 'id',
  [ '흑돼지', { 'address_name': '제주특별자치도 제주시 노형동 2343-3', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 한식 > 육류,고기 > 삼겹살', 'distance',
  [ '맛집', { 'address_name': '제주특별자치도 서귀포시 안덕면 사계리 86', 'category_group_code': 'CE7', 'category_group_name': '카페', 'category_name': '음식점 > 카페', 'distance': '', 'id': '2177',
  [ '속성도', { 'address_name': '제주특별자치도 제주시 노형동 941-2', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 한식 > 육류,고기', 'distance': '', 'id',
  [ '카멜리아힐', { 'address_name': '제주특별자치도 서귀포시 안덕면 상창리 271', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 수목원,식물원', 'id',
  [ '아르떼', { 'address_name': '제주특별자치도 제주시 애월읍 어음리 1503', 'category_group_code': 'CT1', 'category_group_name': '문화시설', 'category_name': '문화,예술 > 문화시설 > 미술관', 'distar',
  [ '미술관', { 'address_name': '제주특별자치도 제주시 애월읍 어음리 1503', 'category_group_code': 'CT1', 'category_group_name': '문화시설', 'category_name': '문화,예술 > 문화시설 > 미술관', 'distar',
  [ '제주', { 'address_name': '제주특별자치도 서귀포시 서귀동 316-1', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 도보여행 > 둘레길 > 제주둘',
  [ '오마카세', { 'address_name': '제주특별자치도 제주시 연동 252-1', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 일식 > 일식집', 'distance': '', 'id',
  [ '코하쿠', { 'address_name': '제주특별자치도 제주시 노형동 138-2', 'category_group_code': 'FD6', 'category_group_name': '음식점', 'category_name': '음식점 > 일식 > 초밥,롤', 'distance': '', 'id',
  [ '제주도', { 'address_name': '제주특별자치도 제주시 오동동 산 182', 'category_group_code': 'AT4', 'category_group_name': '관광명소', 'category_name': '여행 > 관광,명소 > 섬', 'distance': '', 'id',
  [ '명월국민학교', { 'address_name': '제주특별자치도 제주시 한림읍 명월리 1734', 'category_group_code': 'CE7', 'category_group_name': '카페', 'category_name': '음식점 > 카페', 'distance': '', 'id' }
```

5. Youtube API를 통해 얻은 영상정보와 Kakao API를 통해 얻은 장소 정보 합치기

```
[[0,
  {'address_6': '남원,표선,성산(서귀포시 동부)',
   'category': '관광,명소',
   'channelTitle': '홍이네 일상',
   'id': 'Nk9ltFZiojc',
   'likeCount': '12',
   'place_name': ['하례감굴채험농장'],
   'place_url': 'http://place.map.kakao.com/1228554077',
   'publishTime': '2022-01-09T14:31:35Z',
   'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
   'viewCount': '184',
   'x': ['126.681109358351'],
   'y': ['33.29927202034467']}],
 [0,
  {'address_6': '조천,구좌,우도(제주시 동부)',
   'category': '관광,명소',
   'channelTitle': '홍이네 일상',
   'id': 'Nk9ltFZiojc',
   'likeCount': '12',
   'place_name': ['비밀의숲'],
   'place_url': 'http://place.map.kakao.com/82389937',
   'publishTime': '2022-01-09T14:31:35Z',
   'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
   'viewCount': '184',
   'x': ['126.75493209633262'],
   'y': ['33.45360917744684']}],
 [0,
  {'address_6': '제주시내(제주시 중부)',
   'category': '음식점',
   'channelTitle': '홍이네 일상',
   'id': 'Nk9ltFZiojc',
   'likeCount': '12',
   'place_name': ['속성도 노형점'],
   'place_url': 'http://place.map.kakao.com/316010726',
   'publishTime': '2022-01-09T14:31:35Z',
   'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
   'viewCount': '184',
   'x': ['126.48503012387'],
   'y': ['33.4850301883515']}],
 [0,
  {'address_6': '애월,한림,한경(제주시 서부)',
   'category': '관광,명소',
   'channelTitle': '홍이네 일상',
   'id': 'Nk9ltFZiojc',
   'likeCount': '12',
   'place_name': ['신창풍차해안도로'],
   'place_url': 'http://place.map.kakao.com/26239274',
   'publishTime': '2022-01-09T14:31:35Z',
   'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
   'viewCount': '184'}
```

6. 영상 정보의 title, decription을 전처리하고 중복 제거 후 얻은 최종 데이터

```
[{'address_6': '남원,표선,성산(서귀포시 동부)',
  'category': '관광,명소',
  'channelTitle': '레일라leila',
  'id': 'X78E3bp5-0s',
  'likeCount': '7',
  'place_name': ['하례감굴채험농장'],
  'place_url': 'http://place.map.kakao.com/1228554077',
  'publishTime': '2021-05-18T00:29:20Z',
  'title': '[JEJU VLOG] ep02. 제주도 비오는날 갈만한곳 | 빛의 벵거 | 성산카페 에곤 | 제주성산갈만한곳 | 5월 제주도 여행 | 여자혼자여행 | 제주도동쪽여행',
  'viewCount': '250',
  'x': ['126.681109358351'],
  'y': ['33.29927202034467']},
 {'address_6': '조천,구좌,우도(제주시 동부)',
  'category': '관광,명소',
  'channelTitle': '홍이네 일상',
  'id': 'Nk9ltFZiojc',
  'likeCount': '12',
  'place_name': ['비밀의숲'],
  'place_url': 'http://place.map.kakao.com/82389937',
  'publishTime': '2022-01-09T14:31:35Z',
  'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
  'viewCount': '184',
  'x': ['126.75493209633262'],
  'y': ['33.45360917744684']},
 {'address_6': '제주시내(제주시 중부)',
  'category': '음식점',
  'channelTitle': '홍이네 일상',
  'id': 'Nk9ltFZiojc',
  'likeCount': '12',
  'place_name': ['속성도 노형점'],
  'place_url': 'http://place.map.kakao.com/316010726',
  'publishTime': '2022-01-09T14:31:35Z',
  'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
  'viewCount': '184',
  'x': ['126.48503012387'],
  'y': ['33.4850301883515']},
 {'address_6': '애월,한림,한경(제주시 서부)',
  'category': '관광,명소',
  'channelTitle': '홍이네 일상',
  'id': 'Nk9ltFZiojc',
  'likeCount': '12',
  'place_name': ['신창풍차해안도로'],
  'place_url': 'http://place.map.kakao.com/26239274',
  'publishTime': '2022-01-09T14:31:35Z',
  'title': '겨울 제주도여행 | JEJU Vlog | 제주 브이로그 | 고라니 출연 | 감굴채험 | 비밀의숲 | 속성도 | 풍차해변도로',
  'viewCount': '184'}
```

<데이터 구조>

장소의 지역명, 장소 카테고리, 비디오 ID, 좋아요 수, 장소명, 업로드 시간, 조회수, 경도, 위도

YouTube
Data API v3



APACHE
kafka
A distributed streaming platform



Spark SQL



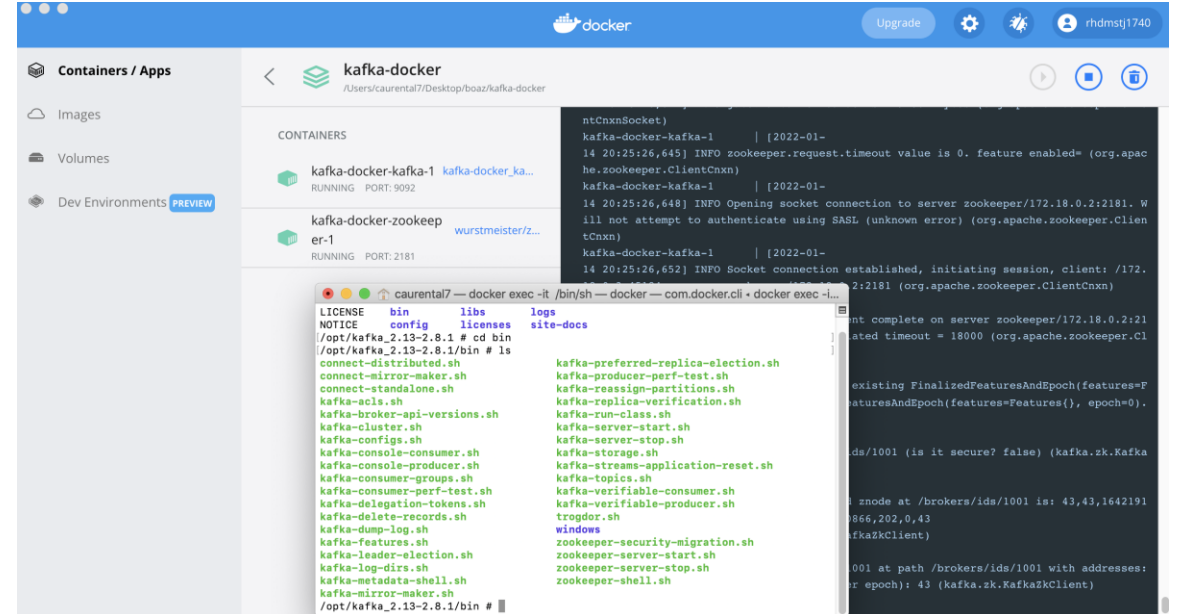
MySQL
DataBase



Spark SQL

• 카프카 서버 구축

- Docker-desktop을 통해 CLI 접속
- Docker Hub에 있는 이미지 활용
- Zookeeper , kafka
- 프로젝트 요구사항에 맞게 config 내 server.properties 수정



Docker desktop에서
Kafka 컨테이너 접속 모습

YouTube
Data API v3



APACHE
kafka
A distributed streaming platform



Spark SQL



MySQL
DataBase



Spark SQL

• Kafka Producer

- Python-kafka 라이브러리 활용
- 프로듀서 매개변수 설정
- 토픽 : 토픽 복제 계수 1 / 파티션 10
- 전처리한 데이터들을 카프카 프로듀서를 통해 send

```
def set_producer(self):
    self.producer = KafkaProducer(
        acks=0,
        compression_type='gzip',
        bootstrap_servers=[self.host],
        value_serializer=lambda x: dumps(x).encode('utf-8'))
```

프로듀서 매개변수 설정

```
{'address_6': '제주 시내 (제주시 중부)',
 'category': '숙박',
 'channelTitle': '해쭈 [HAEJ00]',
 'id': '76oZsi7BhGs',
 'likeCount': '20435',
 'place_name': ['스카이힐비즈니스호텔'],
 'place_url': 'http://place.map.kakao.com/27401069',
 'publishTime': '2020-02-21T14:46:37Z',
 'title': '[VLOG] 제주도 신혼여행인 척하는 4박 5일 동안 먹는 브이로그!',
 'trip!': 'trip!',
 'viewCount': '1865191',
 'x': ['126.52328455661979'],
 'y': ['33.49943653001361']}
```

Kafka Producer에서 다루는
JSON 형식의 데이터

YouTube
Data API v3



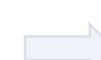
APACHE
kafka
A distributed streaming platform



Spark SQL



MySQL
DataBase



Spark SQL

Kafka Consumer

- Python-kafka 라이브러리 활용
- 컨슈머 매개변수 설정
 - Group_id로 관리
 - Offset_reset = 'latest'
 -> 파티션 별 Offset 저장 가능, 새로 들어온 record만 consume 가능함

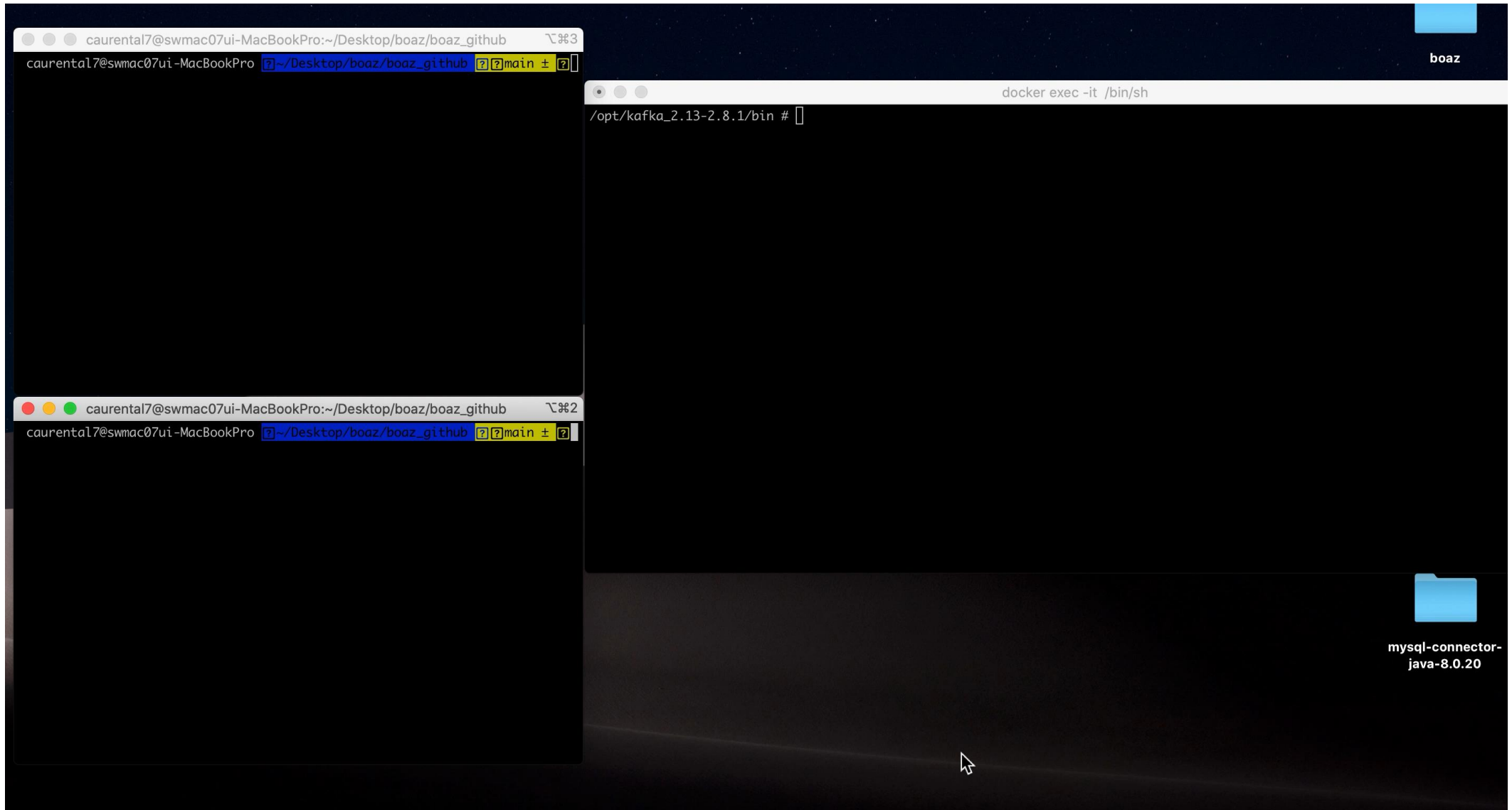
- Poll() & commit()으로 새로 들어온 record 인식

```
bootstrap_servers=[self.host],
group_id=self.group_id,
auto_offset_reset='latest',
enable_auto_commit=True,
value_deserializer=lambda x: x.decode('utf-8'),
consumer_timeout_ms=1000,
```

컨슈머 매개변수 설정

GROUP ID	TOPIC	PARTITION	CURRENT-OFFSET	LOG-END-OFFSET
youplace-consumer	youplace-part	0	66	153
youplace-consumer	youplace-part	1	63	149
youplace-consumer	youplace-part	4	83	175
youplace-consumer	youplace-part	5	82	181
youplace-consumer	youplace-part	2	77	168
youplace-consumer	youplace-part	3	85	179
youplace-consumer	youplace-part	8	72	169
youplace-consumer	youplace-part	9	67	154
youplace-consumer	youplace-part	6	60	159
youplace-consumer	youplace-part	7	89	183

Offset check



mysql-connector-
java-8.0.20

4. 구현 과정 설명 - Spark SQL



YouTube
Data API v3



APACHE
kafka
A distributed streaming platform



Spark SQL



MySQL
DataBase



Spark SQL

- 카프카 Consumer와 Spark **연결**
- Mysql DB 삽입을 위한 **형태 변환 작업 수행**
- **Spark SQL** 라이브러리
 - 정형 데이터 처리를 위한 스파크 모듈
 - 스키마에 대한 표현방법 제공
- 데이터 형태를 Spark Dataframe으로 변환하여 진행
 - 사용자 정의 함수, casting 등을 활용하여 원하는 형태
의 데이터 수집

```
root
|-- address_6: string (nullable = true)
|-- category: string (nullable = true)
|-- channelTitle: string (nullable = true)
|-- id: string (nullable = true)
|-- likeCount: string (nullable = true)
|-- place_name: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- place_url: string (nullable = true)
|-- publishTime: string (nullable = true)
|-- title: string (nullable = true)
|-- viewCount: string (nullable = true)
|-- x: array (nullable = true)
|   |-- element: string (containsNull = t
|-- y: array (nullable = true)
|   |-- element: string (containsNull = t
```

printSchema 전

```
root
|-- address_6: string (nullable = true)
|-- category: string (nullable = true)
|-- channelTitle: string (nullable = true)
|-- id: string (nullable = true)
|-- likeCount: integer (nullable = true)
|-- place_name: string (nullable = true)
|-- place_url: string (nullable = true)
|-- publishTime: timestamp (nullable = true)
|-- title: string (nullable = true)
|-- viewCount: integer (nullable = true)
|-- x: decimal(24,18) (nullable = true)
|-- y: decimal(24,18) (nullable = true)
```

printSchema 후



4. 구현 과정 설명 - MySQL



YouTube
Data API v3



APACHE
kafka
A distributed streaming platform



Spark SQL



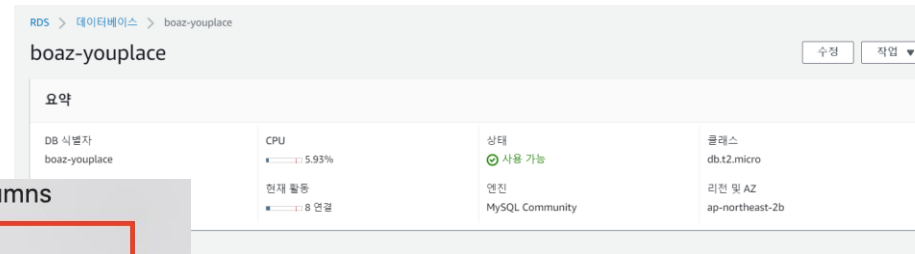
MySQL
DataBase



Spark SQL

- Spark와 mysql 연결
- 다양한 옵션 설정
 - Numpartition 5 : 동시 write 작업

- Mysql RDS 생성



```
g data to a JDBC source

df.write \
  .format("jdbc") \
  .option("url", "jdbc:mysql://boaz-youplace.cai20o") \
  .option("dbtable", "db_youplace.tb_youplace") \
  .option("user", "admin") \
  .option("password", "youplace") \
  .option("numPartitions",5) \
  .option("driver","com.mysql.cj.jdbc.Driver",) \
  .mode('append') \
  .save()
```

Columns

- id
- place_name
- viewCount
- publishTime
- likeCount
- x
- y
- title
- channelTitle
- category
- place_url
- address_6

id	place_name	viewCount	publishTime	likeCount	x	y	title
_KNOe9XaGi4	제라한보쌈	360643	2018-05-08 10:00:05.0	3573	126.470390249071000000	33.483316774940900000	[Eng] 제주도 브이로그 ♥️ 하늘이 추천
_KNOe9XaGi4	착한집	360643	2018-05-08 10:00:05.0	3573	126.507000613623180000	33.499961547893996000	[Eng] 제주도 브이로그 ♥️ 하늘이 추천
-VBZSKXREhQ	다이소 서귀포중문점	221839	2021-03-25 11:31:51.0	1986	126.404518095964990000	33.258900706286180000	[VLOG] 3박4일 제주 여행 브이로그, 3
-VBZSKXREhQ	불랑로세	221839	2021-03-25 11:31:51.0	1986	126.957885292756670000	33.515683602028090000	[VLOG] 3박4일 제주 여행 브이로그, 3
-VBZSKXREhQ	스타벅스 제주애월DT점	221839	2021-03-25 11:31:51.0	1986	126.348763704429120000	33.472752472576900000	[VLOG] 3박4일 제주 여행 브이로그, 3
-VBZSKXREhQ	오논정김밥	221839	2021-03-25 11:31:51.0	1986	126.567621058572000000	33.249595570308300000	[VLOG] 3박4일 제주 여행 브이로그, 3
-VBZSKXREhQ	이촌죽원초고등어찜밥 매일본점	221839	2021-03-25 11:31:51.0	1986	126.418698640025000000	33.488995911539700000	[VLOG] 3박4일 제주 여행 브이로그, 3
0bOh04m7dc	수월불전기자전거	235296	2021-04-20 13:51:23.0	3431	126.167968409641570000	33.298875786487480000	제주 여행 브이로그 서쪽 애월, 협재, 한
2PJ8srQxYKw	출장안마원제주	233833	2021-05-11 12:00:02.0	2978	126.474536865713000000	33.482970182272800000	2박 3일 제주도 출장 브이로그 음식, 회
3afZ3fP5BaQ	비하인드제주출	349095	2019-10-31 11:46:47.0	3499	126.742341835186000000	33.483916822262800000	(Eng) 제주도 여행 브이로그! 10주년 맞이!! 리얼 힐링 (
3lAmfAHoLY	임상다만사	609	2021-01-17 09:00:03.0	20	126.495315788637000000	33.488833945487300000	ENG) 제주도 여행 브이로그! 혼자 떠난
3lAmfAHoLY	출장안마원제주	609	2021-01-17 09:00:03.0	20	126.474536865713000000	33.482970182272800000	ENG) 제주도 여행 브이로그! 혼자 떠난
3Xdx5T9Sik	라마다프리카 제주호텔	686941	2019-12-05 09:52:16.0	5048	126.517879439631000000	33.518780879184500000	[여행Vlog] 힐링하기 급! 제주도 2박3일 (
4GIBK1kgWg	제주에서들고래를만난다면편선	146057	2021-03-12 17:35:26.0	1464	126.870901888628000000	33.528473099471710000	제주브이로그 퇴사 후 제주도 한달살기
6qJnwGnWhSk	동선제안가	103242	2021-03-21 01:00:00.0	1719	126.648458524842000000	33.271969185250500000	제주도 LA 감성으로 여행하기 자매여
6qJnwGnWhSk	사진놀이터	103242	2021-03-21 01:00:00.0	1719	126.441735120785000000	33.456750024214300000	제주도 LA 감성으로 여행하기 자매여
6qJnwGnWhSk	산위일체대성당	103242	2021-03-21 01:00:00.0	1719	126.323808040329000000	33.348630759772600000	제주도 LA 감성으로 여행하기 자매여
6qJnwGnWhSk	자매결혼비	103242	2021-03-21 01:00:00.0	1719	126.381473386322000000	33.306898387127700000	제주도 LA 감성으로 여행하기 자매여
6qJnwGnWhSk	자연과사람들일면	103242	2021-03-21 01:00:00.0	1719	126.661108996110800000	33.460474849268400000	제주도 LA 감성으로 여행하기 자매여
6qJnwGnWhSk	햇살가득돌담집	103242	2021-03-21 01:00:00.0	1719	126.658193770588000000	33.541538533980700000	제주도 LA 감성으로 여행하기 자매여
76oZs7BHGs	시간을달라	1864446	2020-02-21 14:46:37.0	20428	126.843826939342380000	33.529983653065900000	[VLOG] 제주도 신혼여행인 척하는 4박
76oZs7BHGs	안마요시21	1864446	2020-02-21 14:46:37.0	20428	126.769082362794530000	33.206510157443290000	다들 다들 제주도 여행하러 왔는데 4박

최종 DB에 데이터 적재된 모습 (workbench)

4. 구현 과정 설명 - Spark SQL & 웹



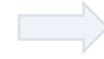
YouTube
Data API v3



APACHE
kafka
A distributed streaming platform



Spark SQL



MySQL
DataBase



Spark SQL

Spark SQL

- Spark sql 라이브러리를 이용한 다양한 통계처리 수행

address_6	count
제주시내 (제주시 중부)	35
서귀포시 (서귀포시 중부)	20
애월, 한림, 한경 (제주시 서부)	17
조천, 구좌, 우도 (제주시 동부)	17
안덕, 대정 (서귀포시 서부)	9
남원, 표선, 성산 (서귀포시 동부)	9

Ex) 구역마다 명소 데이터 개수 count

django

- 최종 데이터가 적재되어 있는 Mysql DB 연동
- KaKao API + Djangotube를 이용한 기능 구현
 - 커스텀 마커 오버레이, 유튜브 영상 재생





시연 영상2



Youplace

x +

← → ↺ ↻

127.0.0.1:8000

📄 ☆ EX ⚙️ 👤 ⋮

앱

구글

네이버

유튜브

GitHub

노션

CAU

포탈

웹백엔드

SQL연습

eclass

YouTube Data API

카슬 원드라이브

벨로그

나동빈이코테

중앙대 소프트웨어학부

»

읽기 목록

Boaz 제 15회 온라인 컨퍼런스 | Web Demo

Youplace

카프카와 스파크를 활용한 유튜브 제주 명소 검색

📍 유튜브 브이로거들이 찾았던 제주 명소들을 지도에서 한 눈에 파악해보세요 😊



애월,한림,한경(제주시 서부)

재생 목록



제주 시내(제주시 중부)

재생 목록



조천,구좌,우도(제주시 동부)

재생 목록





1. 아이디어
 - 제주/여행에 국한 되지 않고 데이터의 범위 확장
2. 스택 활용
 - 높은 처리량을 위해 kafka 서버 옵션을 다양하게 수정 및 테스트
 - 브로커를 여러 개 띄워보지 못한 점
 - Kafka eco system을 다양하게 활용해보지 못한 점
3. 클라우드
 - 더욱 다양한 인프라 활용 (GCP 등)
4. 웹 기능
 - 더욱 다양한 기능 제공 (장소 검색 , 구글 소셜 로그인을 통한 즐겨찾기 기능 등)



Boaz-Youplace



Boaz-Youplace

[Overview](#) [Repositories 2](#) [Packages](#) [People 3](#) [Teams](#) [Projects](#) [Settings](#)

Pinned

Customize your pins

People

[youplace](#) Public

Jupyter Notebook



Invite someone

[Repositories](#)

Find a repository...

Type

Language

Sort

New

[youplace](#) Public

Jupyter Notebook ☆ 0 0 0 0 Updated 2 minutes ago

[kafka-docker](#) Public

Dockerfile for Apache Kafka

Shell ☆ 0 Apache-2.0 2,562 0 0 Updated 15 days ago

Top languages

Shell Jupyter Notebook

(더 자세한 설명은 깃 헵 내 readme와 코드를 참고해주세요 ! 😊)



감사합니다.