

Melón

# 21세기 대중가요 분석

권순찬 장석원 김의진

# 목차



01

프로젝트 배경

프로젝트 주제 선정 이유

04

모델링

수집한 데이터 모델링 결과

02

활용 데이터 소개

활용 데이터 소개

05

결과 해석

데이터 모델링 결과 분석

03

전처리 과정

데이터 전처리 과정

06

결론 및 보완점

프로젝트 진행 결과



🎵 CHAPTER. 1

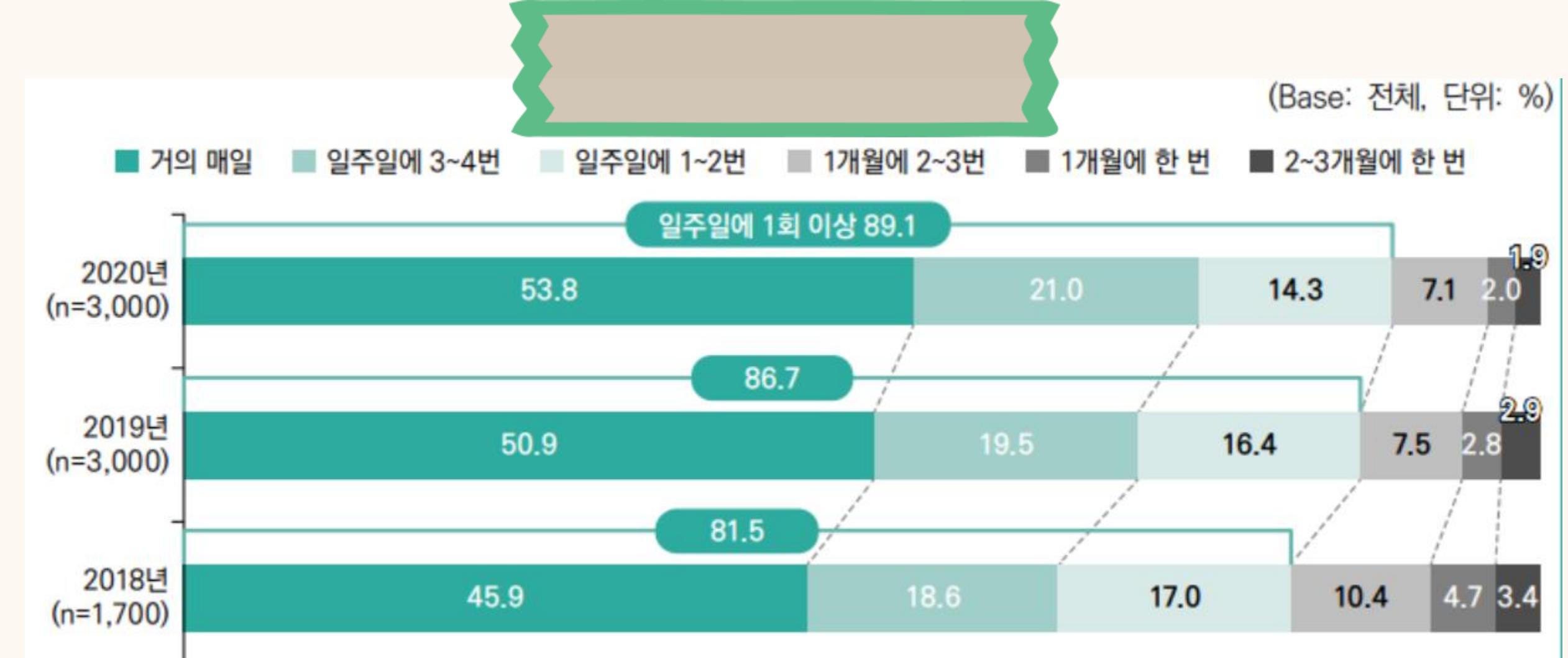
# 프로젝트 배경

- 프로젝트 주제 선정 이유





(Base: 전체, 단위: %)





CHAPTER. 2

# 활용 데이터 소개

- 활용 데이터 소개



# Melon

[Spotlight] 세븐틴에게 루엇이든 물어보세요!



검색

RVNDIT (복디) • 15

"The Seed" 프로젝트 두 번째 주자  
프로듀서 Padi (페디)

멜론차트 최신음악 장르음악 멜론DJ 멜론TV 스타포스트 매거진 뮤직어워드 어학

마이뮤직

TOP100 일간 주간 월간 시대

Q 차트 파인더

## 시대별 차트

국내

해외

&lt; 2010년도 &gt;

2010년대 ▾ 2010년 ▾



2010년도 한국 가요史

아날로그 바람 타고 온 여고생 가수??  
가요계 새로운 대세 '아이유' 전성시대!!'소녀대비' 아이유는 고등학생임에도 나이에 비해 뛰어난 가창력으로 삼촌들의 절대적 지지를 받으며 새로운 국민  
여동생으로 등극했다. 기계음, 댄스 음악에 치친 음악 편들에게 기타 암송으로 새로운 차극체가 되어 주기도  
했다. 가요계 새로운 대세로 떠오른 아이유!! 배야호로 아이유 전성시대가 왔다.

캡처 바로보기 ▾

✓ 전체선택 ▶ 듣기 ▴ 다운 + 담기 ♫ 삽입

순위

곡정보

좋아요

뮤비

다운

<input type="checkbox"/>	1		Bad Girl Good Girl 미쓰에이   Bad But Good	41,192		
<input type="checkbox"/>	2		잔소리 (With 2AM 슬옹) 마이유   잔소리	99,219		
<input type="checkbox"/>	3		죽을 만큼 아파서 (Feat. 멜로우) MC몽   블루브랜드 Trauma Part 2	88,461		

# 기존 21세기 멜론 Top 20

Melon 조류리 '리브 엔!' MV 출연현장 감상평 이벤트 2개월 최대 50% 특가 할인!

멜론차트 최신음악 장르음악 멜론DJ 멜론TV 스타포스트 매거진 뮤직어워드 어학 마이뮤직

TOP100 일간 주간 월간 시대 Q. 차트 파인더

TOP100 ? 2022.06.06 20:00 ~

곡정보 앨범 좋아요 듣기 달기 다운 유비

순위	곡명	앨범	좋아요	듣기	달기	다운	유비
1	That That (prod. & feat. SUG... 싸이 (PSY))	싸다9	♡ 88,606				
2	TOMBOY (여자)아이들	I NEVER DIE	♡ 178,732				

1970년대 Top100



1980년대 Top100



1990년대 Top100



## 데이터 수집

```
[3] df = pd.read_csv('/content/drive/MyDrive/textmining_rawdata.csv')
```

```
[4] df
```

	year	rank	name	genre	lyricist	lyrics
0	1970	1	가을비 우산속	성인가요/트로트	최현	그리움이 눈처럼₩n 쌓인 거리를₩n 나혼자서 걸었네₩n 미련때문에₩n 흐르는 세월따...
1	1970	2	감수광	성인가요/트로트	혜은이	바람부는 제주에는 둘도 많지만₩n 인정많고 마음씨 고운 아가씨도 많지요₩n 감수광 ...
2	1970	3	개구장이	록/메탈	산울림	우리 같이 놀아요₩n 뛸을 뛰며 공을 차며 놀아요₩n 우리 같이 불러요₩n 예쁜 노...
3	1970	4	개 여울	성인가요/트로트	정미조	당신은 무슨 일로 그리합니까₩n 흘로이 개여울에 주저 앉아서₩n 파릇한 풀포기가 들픈...
4	1970	5	고귀한 선물	포크/블루스	장은아	갈매기 날리는 바닷가에도₩n 그대가 없으면 쓸쓸하겠네₩n 그대가 파도가 밀려와 ...
...	...	...	...	...	...	...
751	2022	16	나의 X에게	발라드	이래언, Rayker Dane, 신수윤	우리 다시 만날래₩n 예쁘게 빛나던₩n 모든 주억들이₩n 너무 그리워₩n 너무...
752	2022	17	이제 나만 믿어요	성인가요/트로트	김이나	무얼 믿은 걸까 부족했던 내게서₩n 나조차 못 믿던 내게 여태 머문 사람₩n 무...
753	2022	18	듣고 싶을까	발라드	DrJO, 원슈타인	우리 함께 듣면 그 노랠 듣고 싶을까₩n 듣고 나서 잠시 날 생각은 할까₩n 아...
754	2022	19	무지개	록/메탈	김시온, 멧돼지	오늘 하루 어땠었나요₩n 많이 힘들었나요₩n 쉬지 않고 달려왔던 길에서₩n 나...
755	2022	20	신호등	록/메탈	이무진	이제야 목적지를 정했지만₩n 가려한 날 막아서네 난 갈 길이 먼데₩n 새빨간 얼...

756 rows x 6 columns

## 글자수

```
[10] song = df['lyrics']
```

```
def count_character(song):  
    count = 0  
  
    for i in song:  
        count += len(i)  
  
    return count
```

```
[11] print("공백을 제외한 문자수 : ",count_character(song))
```

공백을 제외한 문자수 : 452939

총 756곡 (452,939개의 글자)

# 데이터 분할



10년 단위

1970년대

1980년대

1990년대

특정별 분할



2000~2006년

2007~2012년

2013~2016년

2017~2020년

2021년~



1970년대



포크송

락

팝계열



1980년대



1990년대



랩/힙합  
락  
아이돌



2000~06년



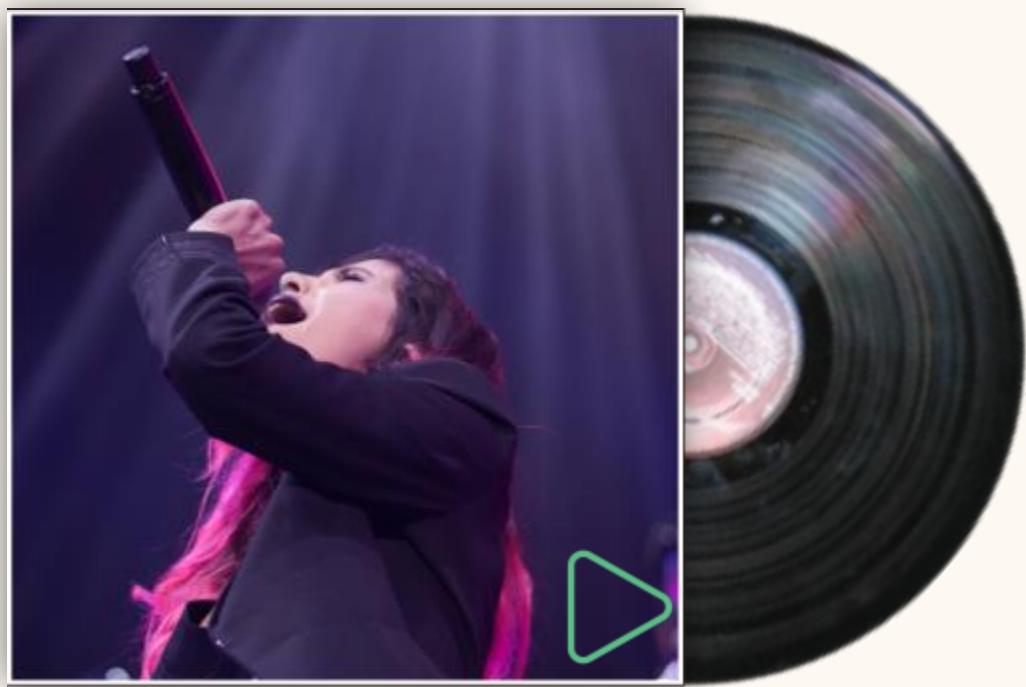
테크노

발라드

R&B



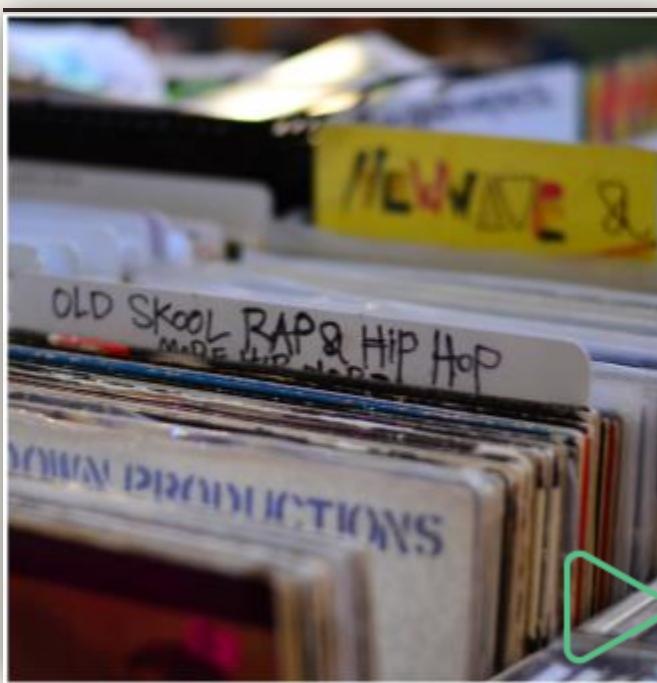
2007~12년



아이돌  
걸그룹  
팝계열



2013~16년



힙합

어반R&B



2017~20년



2021년~



솔로  
아티스트

힙합



## 데이터 분할을 위해 참고한 논문



- 김용학. "한국 대중가요의 의미 연결망 - 1960년대부터 2000년대까지의 변화를 중심으로." *대중서사연구* 21.1 (2015): 145-171.
- 문숙희.(2012).「1970-80년대 서울 관련 대중가요의 두 모습」에 대한 토론문.*한국문학과 예술*,100,206-208.
- 장소원 ( So Won Chang ). "한국 대중가요 가사의 문체 분석." *텍스트언어학* 39.- (2015): 283-311.
- 장유정 ( Eu Jeong Zhang ). "기획논문 : 1970-80년대 한국 대중가요 가사의 특징 -공중파 방송 인기곡을 중심으로-." *공연문화연구* 0.24 (2012): 79-113.
- 장유정 ( Zheng Eu-jeong ). "1990년대와 2000년대 '서울 노래'의 두 모습." *人文科學* 112.- (2018): 7-29.
- 홍성규,이주원, and 서재혁. "1980년대 초기 한·일 헤비메탈음악의 발전 양상과 그 의미에 관한 고찰 — 대표적인 밴드들의 작품 활동을 중심으로 —." *일본학연구* 55.- (2018): 95-120.
- Noh Young-Hae. "Main Themes of Korean Popular Songs in the Last 30 Years." *음악과 문화* 5.- (2001): 149-183.
- Noh Young-Hae. "Main Themes of Korean Popular Songs between 1940 and 1970." *음악과 문화* 7.- (2002): 113-142.

# 데이터 분할 예시



[23] df70

year	rank		name	genre	lyricist	lyrics
0	1970	1	가을비 우산속	성인가요,트로트	최현	그리움이 눈처럼 쌓인 거리를 나 혼자서 걸었네 미련때문에 흐르는 세월따라 잊혀진 그 ...
1	1970	2	감수광	성인가요,트로트	혜은이	바람부는 제주에는 둘도 많지만 인정많고 마음씨 고운 아가씨도 많지요 감수광 감수광 ...
2	1970	3	개구장이	록,메탈	산울림	우리 같이 놀아요 뛴을 뛰며 공을 차며 놀아요 우리 같이 불러요 예쁜 노래 고운 노...
3	1970	4	개 여울	성인가요,트로트	정미조	당신은 무슨 일로 그리합니까 훌로이 개여울에 주저 앉아서 파릇한 풀포기가 돋아 나오...
4	1970	5	고귀한 선물	포크,블루스	장은아	갈매기 날오는 바닷가에도 그대가 없으면 쓸쓸하겠네 파도가 밀려와 속삭여줄때도 그대가...
...	...	...	...	...	...	...
91	1970	96	한사람	포크,블루스	양희은	한사람 여기 또 그곁에 둘이 서로 바라보며 웃네 먼훗날 위해 내미는 손 둘이 서로 ...
92	1970	97	한잔의 주익	포크,블루스	이장희	늦은 밤 쓸쓸히 창가에 앉아 꺼져가는 불빛을 바라보면은 어데선가 날 부르는 소리가 ...
93	1970	98	해뜰 날	성인가요,트로트	송대관	꿈을 안고 왔단다 내가 왔단다 슬픔도 괴로움도 모두 모두 비켜라 안 되는 일 없단다...
94	1970	99	해변의 여인	성인가요,트로트	나훈아	물위에 떠있는 황혼에 종이배 말없이 거니는 해변의 여인아 바람에 휘날리는 머리카락 ...
95	1970	100	휘파람을 부세요	발라드	정미조	제가 보고 싶을 땐 두 눈을 꼭 감고 나즈막히 소리 내어 휘파람을 부세요 외롭다고 ...

96 rows × 6 columns

[24] df17

year	rank		name	genre	lyricist	lyrics
636	2017	1	첫눈처럼 너에게 가겠다	발라드	이미나	널 품기 전 알지 못했다 내 머문 세상 이토록 찬란한 것을 작은 숨결로 달은 사람 ...
637	2017	2	밤편지	발라드	아이유	이 밤 그날의 반딧불을 당신의 창 가까이 보낼게요 음 사랑한다는 말이에요 나 우리의...
638	2017	3	좋니	발라드	윤종신	이제 괜찮니 너무 힘들었잖아 우리 그 마무리가 고작 이별뿐인 건데 우린 참 어려웠어...
639	2017	4	좋다고 말해	인디음악, 포크,블루스	안지영	안녕 오늘도 같은 자리 버스 창가에 기대 앉은 네게 인사를 해 Hi 역시 넌 받아 ...
640	2017	5	Marry Me	발라드	마크툽	비내리는 날엔 우산이 돼주고 어둠이 오면 빛이 돼줄게 주운 겨울이면 난로가 돼주고 ...
...	...	...	...	...	...	...
711	2020	16	살짝 설렜어	댄스	미미	going nonstop 말도 안 되잖아 그치 yayaya 우린 서로가 모르는 ...
712	2020	17	ON	랩,힙합	August Rigo	I can't understand what people are sayin 어느 장단...
713	2020	18	오늘도 빛나는 너에게	R&B,Soul	마크툽	별빛이 내린 밤 그 풍경 속 너와 나 날 새롭게 하는 따뜻하게 만드는 니 눈빛 니 ...
714	2020	19	Downtown Baby	랩,힙합, 인디음악	블루	너는 나의 다운타운 베이비야 너의 눈은 밤하늘에 별이야 매일마다 꾸고 싶은 꿈이야 ...
715	2020	20	마리아	댄스	화사	욕을 하도 먹어 체했어 하도 서러워도 어쩌겠어 I do 모두들 미워하느라 애썼네 날...

80 rows × 6 columns

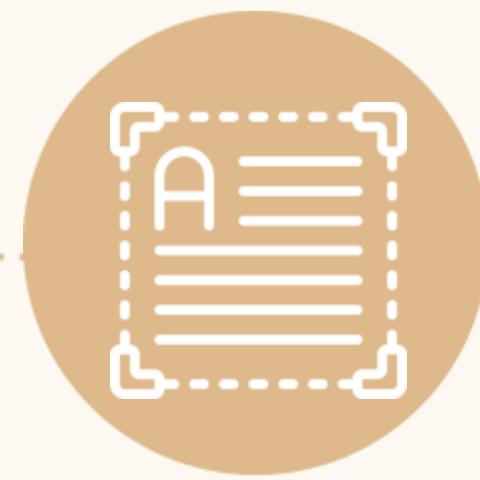
♪ CHAPTER. 3

# 전처리 과정

- 데이터 전처리 과정

3





텍스트 스크래핑



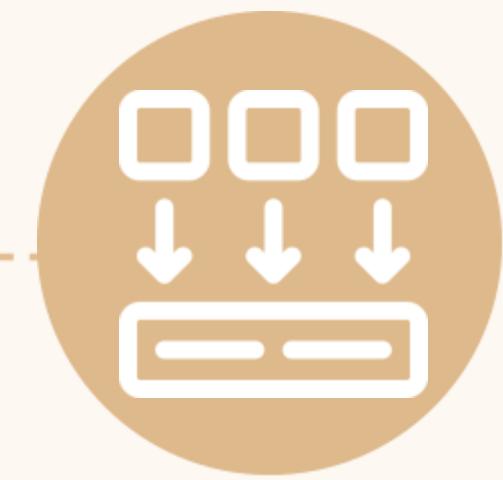
노이즈 캔슬링



토크나이징  
형태소분석/품사태깅

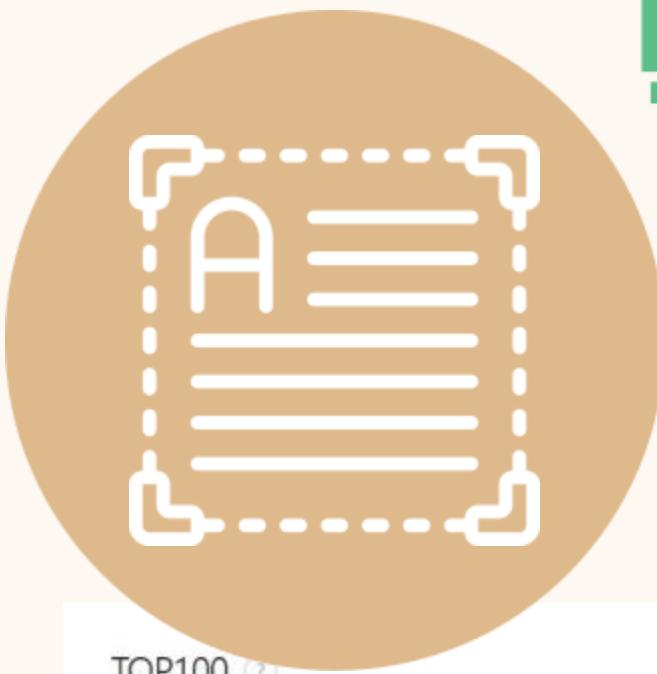


불용어처리



텍스트 벡터화

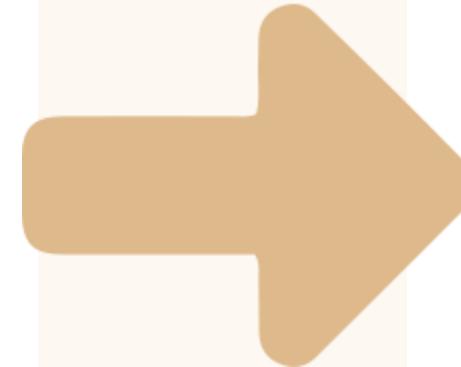
# 텍스트 스크래핑



TOP100

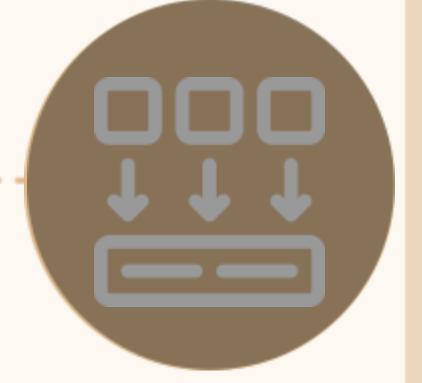
2022.06.06 20:00

순위	곡정보	앨범	좋아요	듣기	달기	다운	큐비
1	That That (prod. & feat. SUGA, 빅히어(PSY))	싸디9	♡ 88,606	▶	+	⬇	□ 0
2	TOMBOY (여자)아이들	I NEVER DIE	♡ 178,732	▶	+	⬇	□ 0
3	LOVE DIVE (여자)아이들	LOVE DIVE	♡ 134,022	▶	+	⬇	□ 0
4	사랑인가 봐 (멜로망스)	사랑인가 봐 (사내맞선 OST)	♡ 123,017	▶	+	⬇	□ 0
5	봄여름가을겨울 (Still Life) BIGBANG (빅뱅)	봄여름가을겨울 (Still Life)	♡ 270,514	▶	+	⬇	□ 0
6	절이라고 하자 (Feat. 10CM) BIG Naughty (서동현)	절이라고 하자	♡ 111,022	▶	+	⬇	□ 0
7	사랑은 늘 도망가 밀영웅	신사와 애가씨 OST Part.2	♡ 162,428	▶	+	⬇	□ 0
8	나의 X에게 경서	나의 X에게	♡ 61,012	▶	+	⬇	□ 0

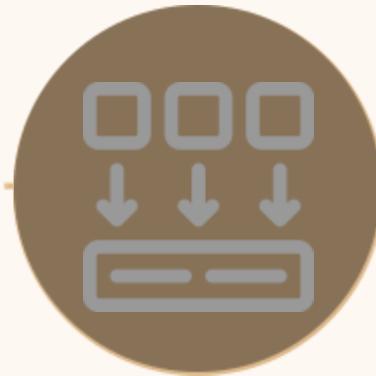
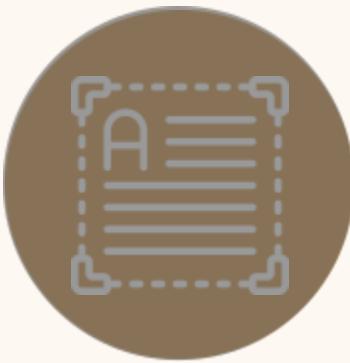


[3]	df = pd.read_csv('/content/drive/MyDrive/textmining_ravdata.csv')
[4]	df
<pre>year rank name genre lyricist lyrics</pre>	
<pre>0 1970 1 가을비 우산속 성인가요/트로트 최현 그리움이 눈처럼 쌓인 거리를hn 나혼자서 걸었네hn 흐르는 세월마...</pre>	
1	1970 2 감수광 성인가요/트로트 혜은이 바람부는 제주에는 들도 많지만hn 인정많고 마음씨 고운 아가씨도 많지요hn 감수광 ...
2	1970 3 개구장이 록/메탈 산울림 우리 같이 놀아요hn 땀을 뛰며 공을 차며 놀아요hn 우리 같이 불러요hn 예쁜 노...
3	1970 4 개 여울 성인가요/트로트 정미조 당신은 무슨 일로 그리합니까hn 풀로이 개여울에 주저 앉아서hn 파릇한 풀포기가 든...
4	1970 5 고귀한 선물 포크/블루스 장은아 갑 얘기 날오는 바닷가에도hn hn 그대가 없으면 쓸쓸하겠네hn hn 파도가 밀려와 ...
...	...
751	2022 16 나의 X에게 발라드 이래연, Rayker Dane, 신수윤 우리 다시 만날래hn 예쁘게 빛나던hn 모든 추억들이hn너무 그리워hn너...
752	2022 17 이제 나만 믿어요 성인가요/트로트 김이나 무얼 믿은 걸까 부족했던 내게서hn나조차 못 믿던 내게 여태 머문 사람hn무...
753	2022 18 듣고 싶을까 발라드 DrJO, 원슈타인 우리 함께 듣면 그 노랠 듣고 싶을까hn듣고 나서 잠시 날 생각은 할까hn아...
754	2022 19 무지개 록/메탈 김시온, 옛돼지 오늘 하루 어땠었나요hn많이 힘들었나요hn쉬지 않고 달려왔던 길에서hn나...
755	2022 20 신호등 록/메탈 이루진 이제야 목적지를 정했지만hn가려한 날 막아서네 난 같 같이 먼데hn새빨간 얼...

756 rows × 6 columns



# 노이즈 캔슬링



Lyrics 칼럼에 있는 개행문자 처리

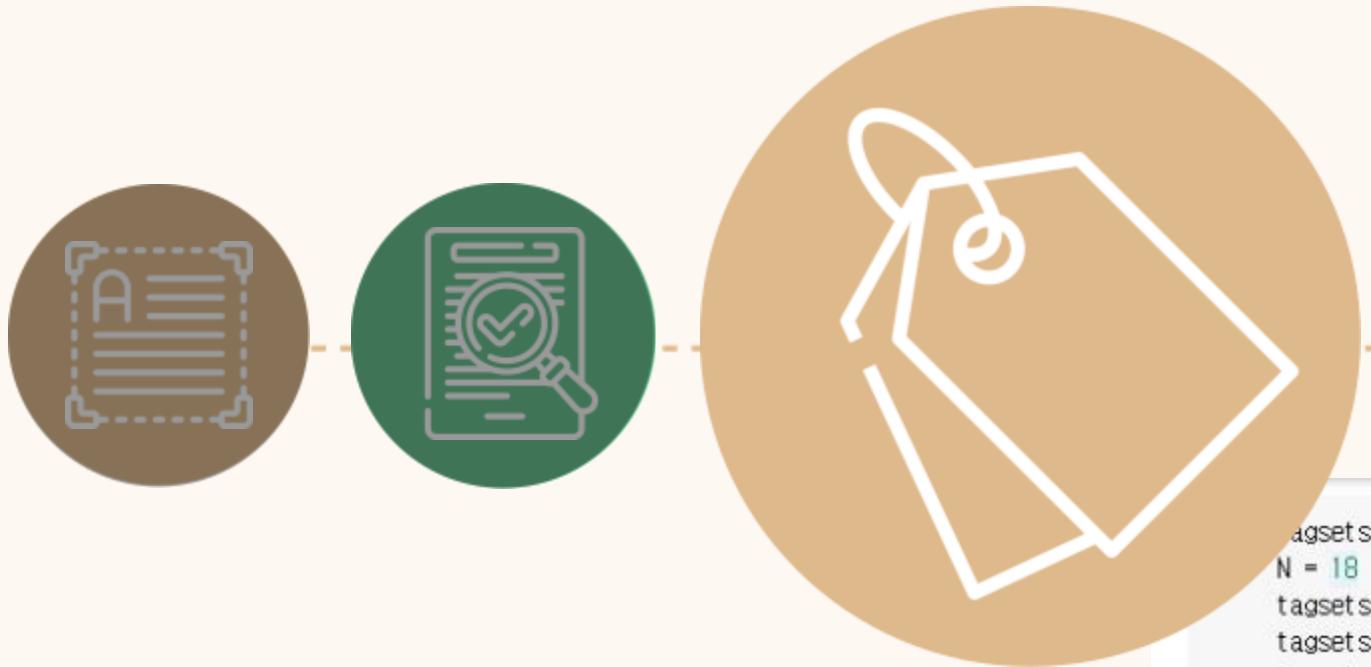
✓ [5] df['lyrics'].replace(r'\s+|\n', ' ', regex=True, inplace=True)  
0초

특수문자 제거

✓ [6] df.replace('<123456789>', '', regex=True, inplace=True)  
0초

# 토크나이징

## 형태소분석/품사태깅



### otk 형태소 분석

[14] pip install konlpy

[15] 1초  
import konlpy  
from konlpy.tag import Okt  
okt = Okt()

[16] okt = Okt()  
words70 = " ".join(df70['lyrics'].tolist())  
words70 = okt.morphs(words70, stem=True)  
words70

[17] 0초  
len(words70)

9722

```
tagsets = pd.DataFrame()
N = 18
tagsets["OKT-품사"] = list(okt.tagset.values()) + list("*" * (N - len(okt.tagset)))
tagsets["OKT-기호"] = list(okt.tagset.keys()) + list("*" * (N - len(okt.tagset)))
tagsets
```

	OKT-품사	OKT-기호	✎
0	형용사	Adjective	
1	부사	Adverb	
2	알파벳	Alpha	
3	접속사	Conjunction	
4	관형사	Determiner	
5	어미	Eomi	
6	감탄사	Exclamation	
7	외국어, 한자 및 기타기호	Foreign	
8	트위터 해쉬태그	Hashtag	
9	조사	Josa	
10	(ex: ㅋㅋ)	KoreanParticle	
11	명사	Noun	
12	숫자	Number	
13	선어말어미	PreEomi	
14	구두점	Punctuation	
15	트위터 아이디	ScreenName	
16	접미사	Suffix	
17	미등록어	Unknown	

### 품사태깅

```
[25] 0초  
words70_tag = " ".join(df70['lyrics'].tolist())
words70_tag = okt.pos(words70_tag, norm=True, stem=True, join=True)
words70_tag
```

'감다/Verb',  
'잊다/Verb',  
'못/VerbPrefix',  
'하다/Verb',  
'그/Noun',  
'사람/Noun',  
'을/Josa',  
'저/Noun',  
'멀리/Noun',  
'저/Noun',  
'멀리/Noun',  
'서/Josa',  
'무지개/Noun',  
'타고/Noun',  
'오다/Verb',  
'</Punctuation',  
'제/Noun',  
'1/Number',  
'절/Noun',  
'>/Punctuation',  
'모두/Noun',  
'들/Suffix',  
'잘들다/Verb',  
'고요한/Noun',  
'이/Determiner',  
'밤/Noun',  
'에/Josa',  
'아이/Noun',  
'하다/Verb',  
'나/Noun',  
'흘로/Noun',  
'자다/Verb'.

# 불용어처리



RANKS NL



```
[37]: with open('/content/drive/MyDrive/stopwords.txt', 'r', encoding='cp949') as f:  
    list_file = f.readlines()  
stopwords = list_file[0].split(",")  
words70_s = [x for x in words70 if x not in stopwords]  
len(words70_s)
```

5650

✓ [38] words70\_s

'그리움',  
'눈',  
'쌓이다',  
'거리',  
'걸다',  
'미련',  
'흐르다',  
'세월',  
'따르다',  
'잊혀지다',  
'얼굴',  
'이다지',  
'속눈썹',  
'또다시',  
'떠오르다',  
'정다웠던',  
'눈길',  
'목소리',  
'어딜',  
'아프다',  
'가슴',  
'달래다',  
'찾다',  
'헤매다',  
'가을비'



CHAPTER. 4

# 모델링

- 수집한 데이터 모델링 결과



빈도분석

빈도 분석

워드클라우드로 시각화

TF-IDF 분석



토픽모델링

LDA 사용

토픽 및 핵심어 설정

▶ 특정 주제 연도별 분석

장르 분석

장소 관련 어휘

작사가별 어휘, 핵심어

영어 사용 비율

길이 분석



# 1970년대 가사 워드클라우드

# 1980년대 가사 워드클라우드

# 1990년대 가사 워드클라우드



# 2000~06년 가사 워드클라우드





# 2007~12년 가사 워드클라우드

# 2013~16년 가사 워드클라우드

# 2017~20년 가사 워드클라우드



# 2021년~ 가사 워드클라우드



순위	단어	TF-IDF
1	사랑	<b>8.469884</b>
2	마음	<b>5.650371</b>
3	사람	<b>5.646274</b>
4	그대	<b>4.956854</b>
5	눈물	<b>2.880280</b>
6	추억	<b>2.753804</b>
7	세월	<b>2.640072</b>
8	생각	<b>2.590047</b>
9	노래	<b>2.440626</b>
10	슬픔	<b>2.425240</b>

## TF-IDF 분석



### 1970년대 분석 결과



[1980년대 분석 결과](#)

[1990년대 분석 결과](#)

[2000 ~ 2006년 분석 결과](#)

[2007 ~ 2012년 분석 결과](#)

[2013 ~ 2016년 분석 결과](#)

[2017 ~ 2020년 분석 결과](#)

[2021년 ~ 분석 결과](#)

순위	단어	TF-IDF
1	사랑	<b>12.880235</b>
2	그대	<b>11.390547</b>
3	마음	<b>9.181263</b>
4	가슴	<b>5.094256</b>
5	사람	<b>5.068631</b>
6	바람	<b>5.063685</b>
7	눈물	<b>4.100261</b>
8	추억	<b>3.391158</b>
9	이별	<b>2.799439</b>
10	생각	<b>2.724399</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과



1990년대 분석 결과

2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>12.076719</b>
2	그대	<b>9.357844</b>
3	마음	<b>5.756596</b>
4	사람	<b>4.923912</b>
5	눈물	<b>4.775339</b>
6	세상	<b>4.175216</b>
7	이별	<b>3.628161</b>
8	기억	<b>3.526527</b>
9	생각	<b>3.491200</b>
10	그녀	<b>3.056032</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

### 1990년대 분석 결과



2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>16.377951</b>
2	그대	<b>11.150462</b>
3	사람	<b>6.021220</b>
4	눈물	<b>5.599424</b>
5	마음	<b>3.889788</b>
6	추억	<b>3.767096</b>
7	가슴	<b>3.487240</b>
8	세상	<b>3.377184</b>
9	생각	<b>3.233919</b>
10	기억	<b>3.217356</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

1990년대 분석 결과

2000 ~ 2006년 분석 결과



2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>8.469884</b>
2	그대	<b>5.650371</b>
3	사람	<b>5.646274</b>
4	눈물	<b>4.956854</b>
5	하루	<b>2.880280</b>
6	세상	<b>2.753804</b>
7	마음	<b>2.640072</b>
8	생각	<b>2.590047</b>
9	가슴	<b>2.440626</b>
10	기억	<b>2.425240</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

1990년대 분석 결과

2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과



2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>7.965553</b>
2	그대	<b>6.697430</b>
3	사람	<b>3.537824</b>
4	눈물	<b>3.250025</b>
5	마음	<b>2.647243</b>
6	나나	<b>2.493731</b>
7	바람	<b>2.243303</b>
8	기억	<b>2.063572</b>
9	수가	<b>2.019069</b>
10	남자	<b>1.993101</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

1990년대 분석 결과

2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과



2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>7.187598</b>
2	그대	<b>3.648808</b>
3	순간	<b>3.158826</b>
4	생각	<b>2.946902</b>
5	사람	<b>2.909701</b>
6	기억	<b>2.678815</b>
7	서로	<b>2.422357</b>
8	노래	<b>2.288480</b>
9	마음	<b>2.209426</b>
10	항상	<b>2.192101</b>

## TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

1990년대 분석 결과

2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과



2021년 ~ 분석 결과

순위	단어	TF-IDF
1	사랑	<b>2.961752</b>
2	그대	<b>2.477299</b>
3	매일	<b>1.992985</b>
4	사람	<b>1.748269</b>
5	생각	<b>1.499952</b>
6	사이	<b>1.277085</b>
7	눈물	<b>1.233891</b>
8	기분	<b>1.173880</b>
9	멀리	<b>1.146118</b>
10	순간	<b>1.107181</b>

## ○ TF-IDF 분석



1970년대 분석 결과

1980년대 분석 결과

1990년대 분석 결과

2000 ~ 2006년 분석 결과

2007 ~ 2012년 분석 결과

2013 ~ 2016년 분석 결과

2017 ~ 2020년 분석 결과

2021년 ~ 분석 결과





## 1) 데이터 불러오기 및 텍스트 데이터 형식 처리

```
Lyrics_data = list(df1970.Lyrics)
Lyrics_data[0:5]
```

```
['그리움이 눈처럼 쌓인 거리를 나 혼자서 걸었네 미련때문에 흐르는 세월따라 잊혀진 그 얼굴이 왜 이다지 솔  
바람부는 제주에는 돌도 많지만 인정많고 마음씨 고운 아가씨도 많지요 감수광 감수광 나어역할령 감수광  
우리 같이 놀아요 힘을 뛰며 공을 차며 놀아요 우리 같이 불러요 예쁜 노래 고운 노래 불러요 이마엔 짬방  
당신은 무슨 일로 그리합니까 홀로이 개여울에 주저 앉아서 파릇한 풀포기가 돌아 나오고 잔물이 봄바람에  
갈매기 날리는 바닷가에도 그대가 없으면 쓸쓸하겠네 파도가 밀려와 속삭여줄때도 그대가 없으면 쓸쓸하겠!
```





## 2) 정규표현식 및 Konlpy 이용한 명사 추출

```
from konlpy.tag import Kkma
Kkma = Kkma()

dataset = []
for i in range(len(Lyrics_data)):
    dataset.append(Kkma.nouns(re.sub('[^가-힣a-zA-Z\w]', '', Lyrics_data[i])))
```





## 3) Genism을 통해 \*Corpus Dictionary 언어모델 형성

```
lyrics_1970 = dataset  
  
lyrics_1970 = [[y for y in x if not len(y)==1]  
                 for x in lyrics_1970]  
dictionary = corpora.Dictionary(lyrics_1970)  
corpus = [dictionary.doc2bow(text) for text in lyrics_1970]
```



\*Corpus Dictionary: 말뭉치 사전



# 토픽 모델링 - LDA

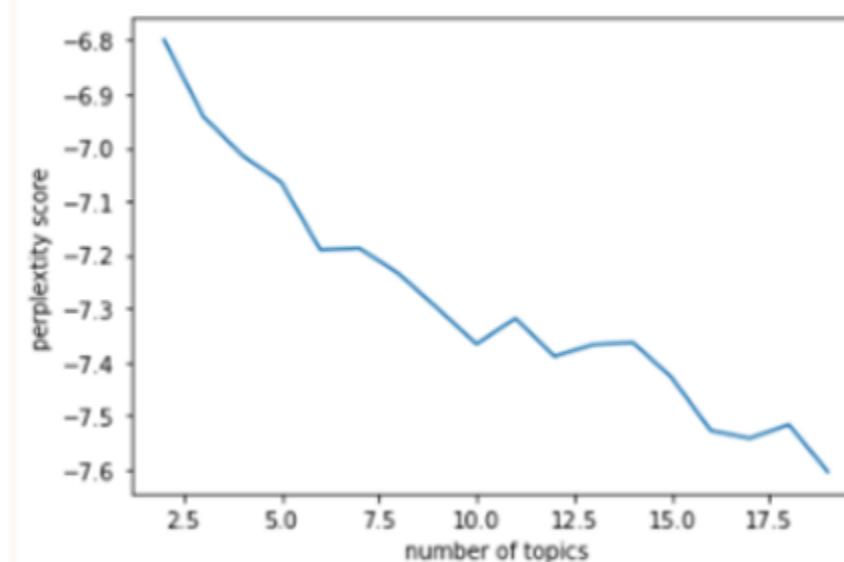
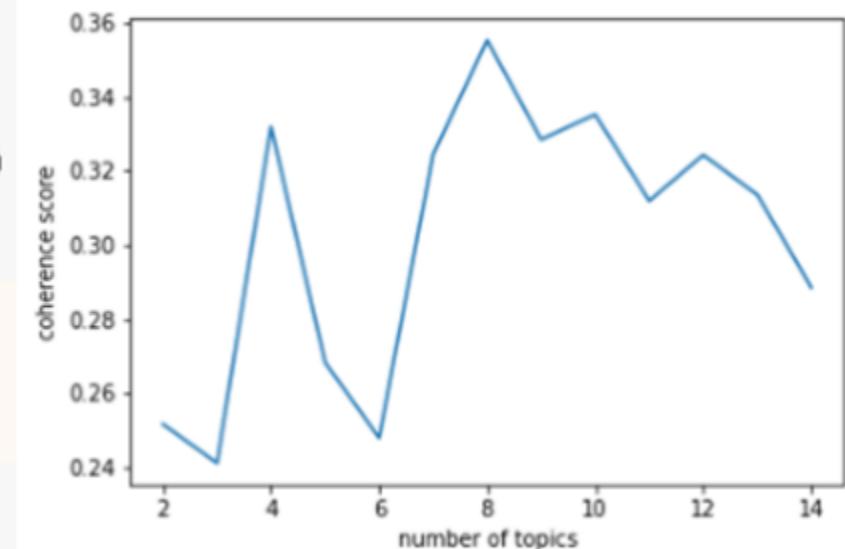


## 4) Perplexity 및 Coherence를 통한 모델 평가 및 토픽 최적화

```
import matplotlib.pyplot as plt
from gensim.models import CoherenceModel

coherence_values=[]
for i in range(2,15):
    ldamodel=gensim.models.LdaModel(corpus, num_topics=i, id2word=dictionary)
    coherence_model_lda=CoherenceModel(model=ldamodel, texts=lyrics_1970, dictionary=dictionary, topn=10)
    coherence_lda=coherence_model_lda.get_coherence()
    coherence_values.append(coherence_lda)

perplexity_values=[]
for i in range(2,20):
    ldamodel=gensim.models.LdaModel(corpus, num_topics=i, id2word=dictionary)
    perplexity_values.append(ldamodel.log_perplexity(corpus))
```



# 토픽 모델링 - LDA



## 5) 하이퍼 파라미터 설정 및 LDA 시작화

```
#LDA 는 Count기반의 Vectorizer만 적용
count_vectorizer = CountVectorizer(max_df=0.1, max_features=1000, min_df=2, ngram_range=(1,2))
    # 2개의 문서 미만으로 등장하는 단어는 제외, 전체의 10% 이상으로 자주 등장하는 단어는 제외
    # bigram도 포함

feat_vect = count_vectorizer.fit_transform(tokenized_list)
print('CountVectorizer Shape:', feat_vect.shape)

lda = LatentDirichletAllocation(n_components=8) # 토픽 수는 8개로 설정
lda.fit(feat_vect)

LatentDirichletAllocation(n_components=8)

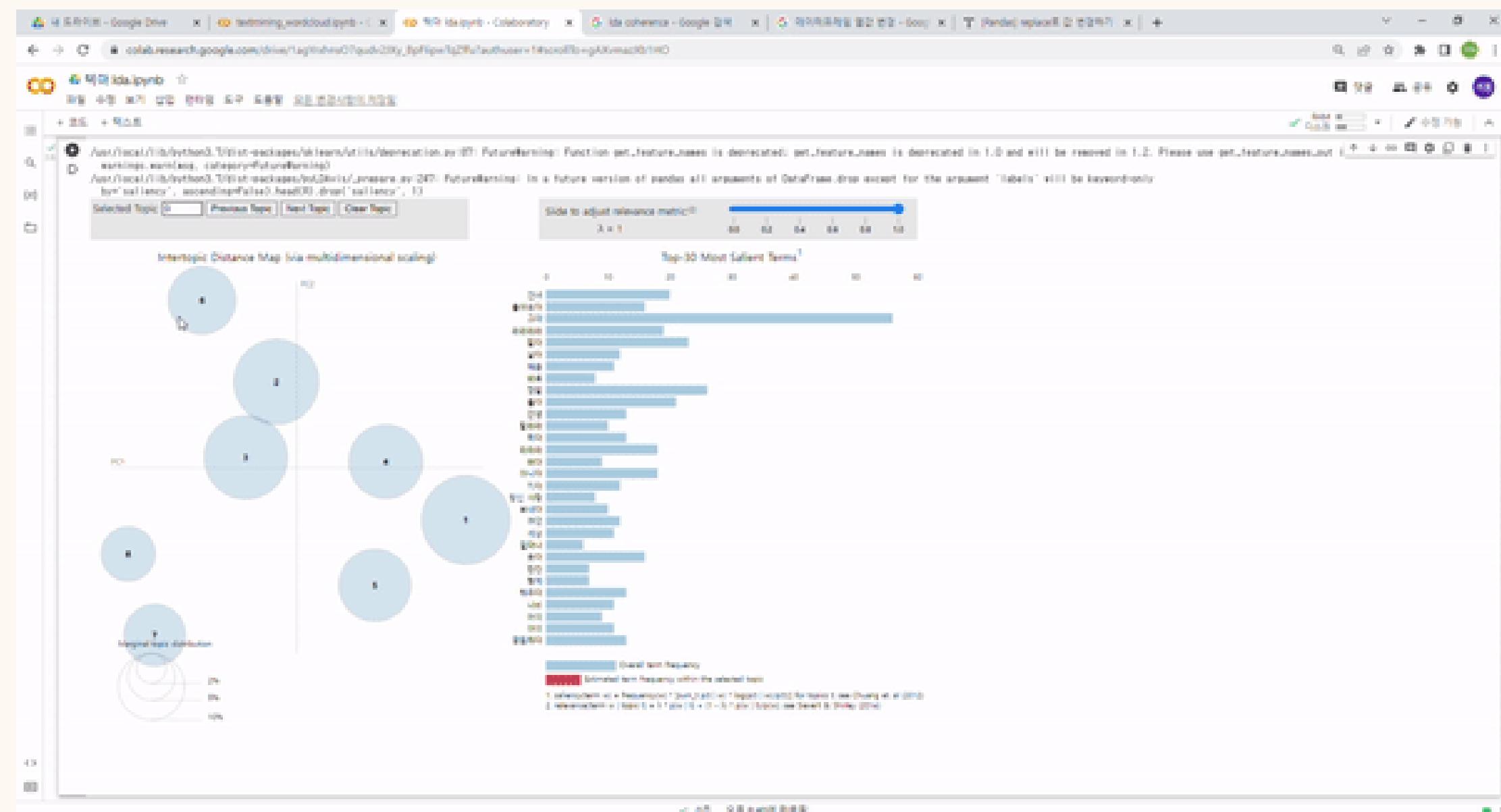
LatentDirichletAllocation(batch_size=128, doc_topic_prior=None,
    evaluate_every=-1, learning_decay=0.7,
    learning_method='batch', learning_offset=10.0,
    max_doc_update_iter=100, max_iter=10, mean_change_tol=0.001,
    n_components=8, n_jobs=None, perp_tol=0.1,
    random_state=None, topic_word_prior=None,
    total_samples=1000000.0, verbose=0)
```



# 토픽 모델링 - LDA



## 5) 하이퍼 파라미터 선정 및 LDA 시각화



# 토픽 모델링 - LDA



## 6) 토픽에 할당된 키워드 및 문서 추출

Doc_Num	Topic	Percentage	Unnamed: 0	Year	Rank	Name	Artist	Genre	Lyrics
0	0	애절한 사랑	0.975680	0	1970	1	가을비 우산속	최현 성인가요/트로트	그리움이 눈처럼 쌓인 거리를 나온자서 걸었네 미련때문에 흐르는 세월따라 잊혀진 그...
1	1	희망/꿈/미래	0.960188	1	1970	2	감수광 혜은이	성인가요/트로트	바람부는 제주에는 둘도 많지만 인정많고 마음씨 고운 아가씨도 많지요 감수광 감수광...
2	2	친구, 형제 위로/한잔	0.967585	2	1970	3	개구장이 산울림	록/에탈	우리 같이 놀아요 펌을 뛰며 공을 차며 놀아요 우리 같이 불러요 예쁜 노래 고운 노...
3	3	과거 회상	0.967583	3	1970	4	개 여울 정미조	성인가요/트로트	당신은 무슨 일로 그리합니까 홍로이 개여울에 주저 앉아서 파릇한 물포기가 돌아 나오...
4	4	다짐/극복/狎사랑	0.610709	4	1970	5	고귀한 선물 장은아	포크/블루스	갈매기 날으는 바닷가에도 그대가 없으면 쓸쓸하겠네 파도가 밀려와 속삭여줄때도 그대가...
...	...	...	...	...	...	...	...	...	...
91	91	애절한 사랑	0.968726	91	1970	96	한사람 양희은	포크/블루스	한사람 여기 또 그곁에 둘이 서로 바라보며 웃네 먼훗날 위해 내미는 손 둘이 서로...
92	92	누군가에 대한 추억	0.987311	92	1970	97	한잔의 추억 이장희	포크/블루스	늦은 밤 쓸쓸히 창가에 앉아 꺼져가는 불빛을 바라보면은 어데선가 날 부르는 소리가...
93	93	회상/그리움	0.974260	93	1970	98	해뜰 날 송대관	성인가요/트로트	꿈을 안고 왔단다 내가 왔단다 슬픔도 괴로움도 모두 모두 비켜라 안 되는 일 없단다...
94	94	희망/꿈/미래	0.968744	94	1970	99	해변의 여인 나훈아	성인가요/트로트	물위에 떠있는 활훈에 중이배 말없이 거니는 해변의 여인아 바람에 휘날리는 머리카락...
95	95	애절한 사랑	0.970822	95	1970	100	튀파랑을 부세요 정미조	발라드	제가 보고 싶을 땐 두 눈을 꼭 감고 나즈막히 소리 내어 튀파랑을 부세요 외롭다고...

96 rows x 10 columns



# 토픽 모델링 - LDA



## 6) 토픽에 할당된 키워드 및 문서 추출



# 장소 관련 어휘 - 1970년대



[('밭', 33),  
('고향', 9),  
('바닷가', 6),  
('해변', 5),  
('동해', 4),  
('서울', 3),  
('남쪽', 2),  
('가로수', 2),  
('골목길', 2),  
('숲', 2),  
('제주', 1),  
('세계', 1),  
('휴전선', 1),  
('대동강', 1),  
('북', 1),  
('지평선', 1)]

## 고향

## 시골

## 전쟁



# 장소 관련 어휘 - 1980년대



[('고향', 10),  
('무리땅', 6),  
('파티', 3),  
('도시', 2),  
('골목길', 1),  
('지하철', 1),  
('경상북도', 1),  
('하와이', 1),  
('미국', 1),  
('숲', 1),  
('자동차', 1)]

## 고향

## 도시화의 시작



# 장소 관련 어휘 - 1990년대



[('전시회장', 4),  
 ('카페', 3),  
 ('골목길', 2),  
 ('학교', 2),  
 ('자동차', 1),  
 ('강', 1),  
 ('지하철', 1)]

## 도시화



# 장소 관련 어휘 - 2000년대



[('비행기', 11),  
('바닷가', 5),  
('카페', 3),  
('자동차', 2),  
('세계', 2),  
('숲', 1),  
('학교', 1),  
('골목길', 1),  
('명동', 1),  
('차이나', 1),  
('도시', 1)]

여가, 여행

세계



# 장소 관련 어휘 - 2010년대



[('우주', 12),  
('여수', 10),  
('클럽', 8),  
('포장마차', 6),  
('양화대교', 5),  
('파티', 4),  
('압구정', 4),  
('은하수', 4),  
('미대', 3),  
('세계', 3),  
('골목길', 3),



[('신촌', 3),  
('동해', 2),  
('지구', 2),  
('도시', 2),  
('강남', 1),  
('강북', 1),  
('이태원', 1),  
('홍대', 1),  
('말리부', 1),  
('해변', 1),  
('지하철', 1),  
('숲', 1)]

유흥

환상, 우주

지역명



♪ CHAPTER. 5

# 결과 해석

- 분석한 데이터 모델링한 결과 분석



1970년대



애절한 사랑  
농촌  
전쟁  
트로트



1980년대



발라드  
이별  
슬픔  
고향



1990년대



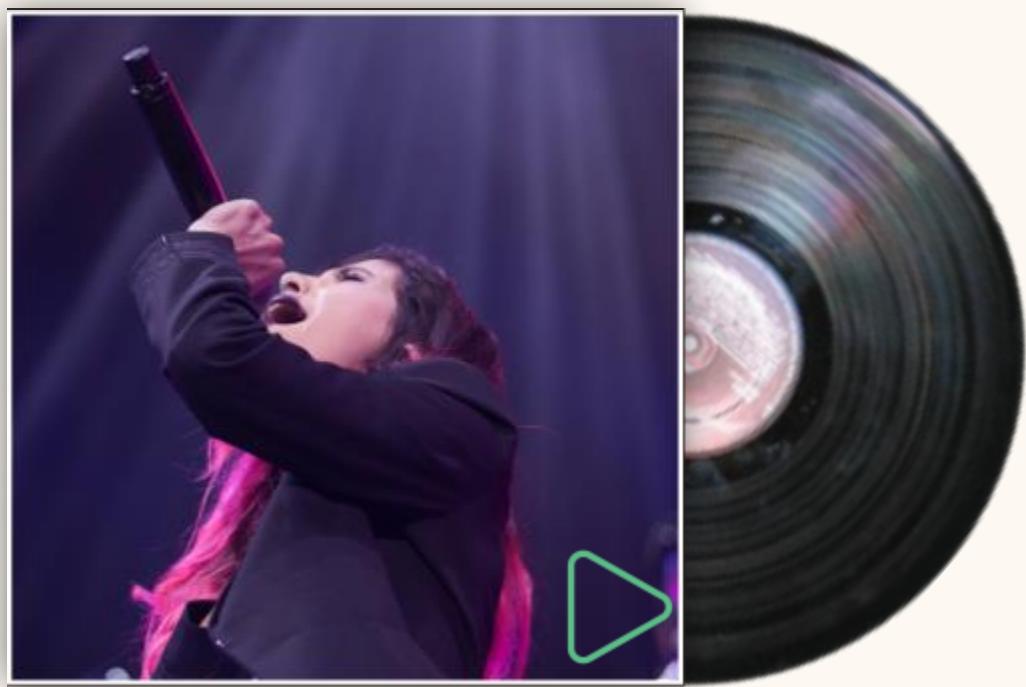
도시화  
세상  
지침  
락  
아이돌



2000~06년



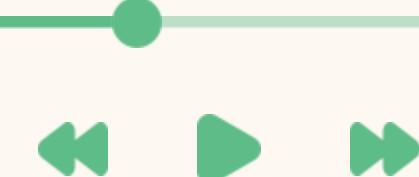
2007~12년



아이돌

후크송

사랑 고백



2013~16년



2017~20년

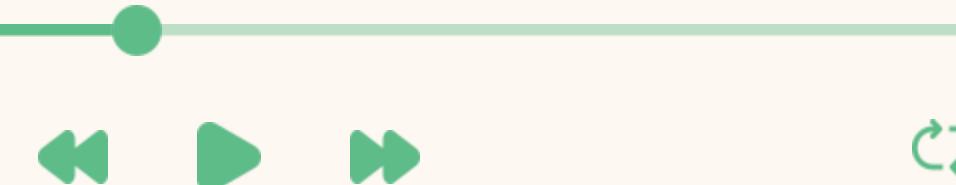


2021년~



솔로  
아티스트

희망적 가사



♪ CHAPTER. 6

# 보완할 점

아쉬웠던 점과 보완할 점

6





## 플랫폼 문제

팬덤 줄 세우기 및 사재기

2005년 이전 순위 정당성

알고리즘 조작

## O 언어적 특성

반복적인 가사

유행어

## ▶ 역량 관련

Attention 기반  
언어모델

빌보드 차트

네트워크 분석

♪ CHAPTER. 6

# 결론

6



감사합니다

