

Melon Chart Analysis 2005-2022

빅데이터 처리 프로젝트

202044086 이설희

Contents

01. 주제 선정

02. Crawling

03. 시각화

주제 선정

아이돌 뮤직, 케이팝은 여전히 굳건한 인기와 끝없는 경쟁 속에 있고
글로벌 시장이 움직이는 정도로 큰 인기를 유지하고 있습니다.
음원 차트의 트렌드를 살펴보면서 그동안 대중들에게 선택받은 음원은
과연 무엇이며 어떤 변화가 일어났는지 알아보았습니다.



Crawling

크롤링이 가능한 음원 사이트 중에서
멜론 차트를 선택하였습니다.

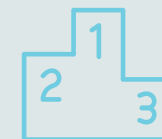
크롤링 데이터 범위는 2005년부터
5년 단위로 현재 연도까지 불러옵니다.

대중들에게 인기 많았던 음원을 분석하기 위해
차트 1위부터 50위까지 불러와 저장을 하였습니다.



멜론차트>시대별차트

시대별 차트 페이지에 들어가서 2005년부터 2020년까지의 데이터를 불러옵니다.



멜론차트>TOP100

현재 2022년도의 음원 순위는 시대별 차트에 포함되지 않기 때문에 TOP100 페이지를 통하여 따로 데이터를 불러 옵니다.



상세정보>장르, 발매일

순위 페이지에는 아티스트, 곡명, 앨범명까지 표시되어 있습니다.
장르와 음원 발매일 데이터도 함께 추출하기 위해 상세정보 페이지에 접근합니다.

시대별 차트 ?

국내

해외

< 2005년도 >

2000년대

2005년
















전체선택

듣기


다운

담기

선물

<input type="checkbox"/> 순위	곡정보	좋아요	유비 다운
<input type="checkbox"/> 1	 <div>죄와 벌 SG 워너비 살다가</div>	49,989	 
<input type="checkbox"/> 2	 <div>갑쟁이 버즈 Buzz Effect</div>	70,918	 
<input type="checkbox"/> 3	 <div>사랑했나봐 윤도현 Difference</div>	111,450	 
<input type="checkbox"/> 4	 <div>살다가 SG 워너비 살다가</div>	72,940	 
<input type="checkbox"/> 5	 <div>광(狂) SG 워너비 살다가</div>	16,408	 

곡 정보



죄와 벌

SG 워너비

앨범

살다가

발매일


2005.03.23


장르

발라드

FLAC

Flac 16bit

 49,989

 곡 다운 >

 FLAC 다운 >

멜론차트>시대별차트

1. 멜론차트>시대별 차트로 가는 URL 저장
2. 시대별 차트에서는 연도를 지정해 줘야 하기 때문에 해당 값에 맞춰서 params 생성
3. 시대별 차트는 '페이지 소스 보기'에서 노래 제목 확인 불가
 - > 개발자 도구로 HXR에서 프리뷰로 Ajax 데이터가 들어오는 걸 확인하고 추출
4. playsong 문자를 포함한 a 태그를 활용하여 songid 추출
5. url + songid를 통하여 상세정보 페이지 이동
6. 순위, 발매일, 곡명, 앨범명, 가수, 장르 데이터 추출

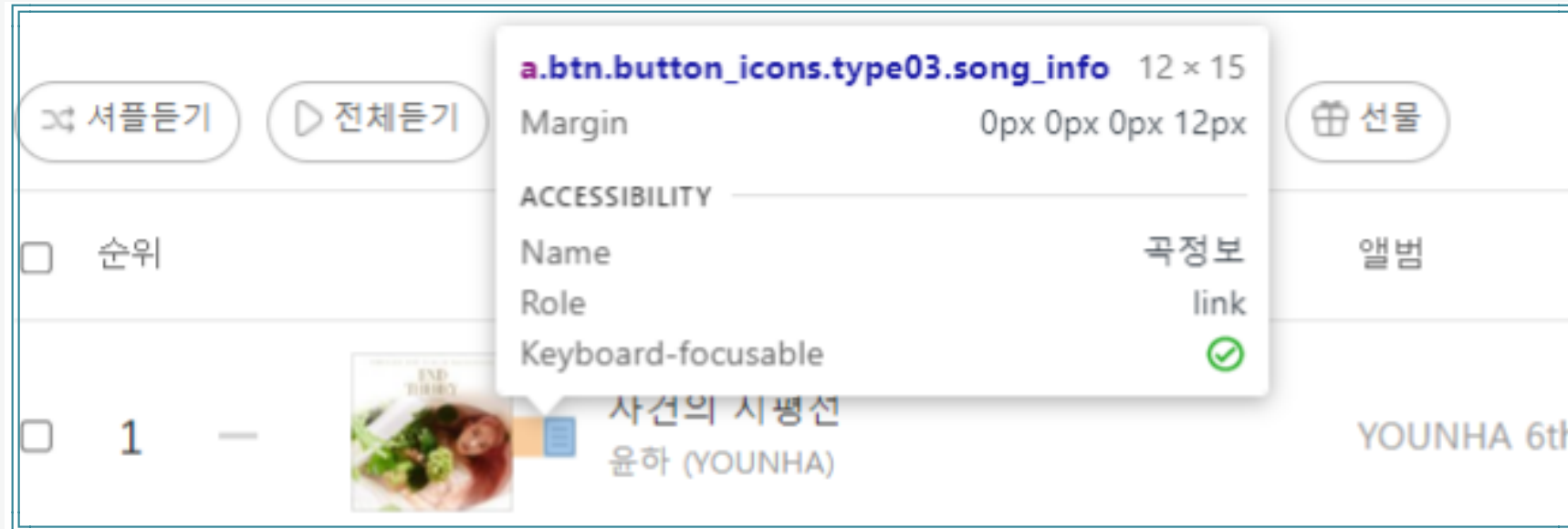
- 연도 리스트 생성하여 리스트 크기만큼 for문 반복
- 연도별 차트 50위 csv 파일 생성
- ex) 2005melon50.csv

멜론차트>TOP100

1. 멜론차트>TOP100으로 가는 URL 저장
2. playsong이 있는 a 태그에 접근
3. a 태그에 정규표현식(re)을 사용하여 songid 번호만 추출
4. url + songid를 통하여 상세정보 페이지 이동
5. 순위, 발매일, 곡명, 앨범명, 가수, 장르 데이터 추출
6. 차트 50위 csv 파일 생성
ex) 2022melon50.csv

멜론 페이지) <https://www.melon.com/chart/index.htm>

앨범 상세정보 페이지) <https://www.melon.com/song/detail.htm?songId=?>



상세정보 페이지 이동을 위한
songid 요소 확인

시각화



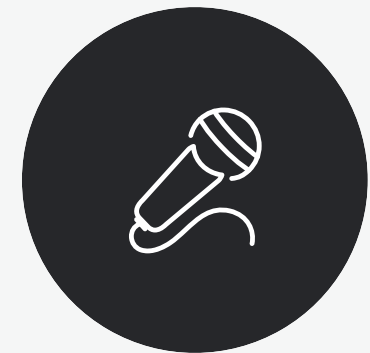
연도별 장르 추세

비율 비교에 효과적인 Matplotlib 파이 차트를 이용하여 과거부터 현재까지의 장르 추세를 확인합니다.



2005년 vs 2022년 장르 변화

2005년과 2022년의 장르 그래프 두 개를 나란히 두어 한눈에 보기 쉽게 표현합니다.



연도별 인기 가수

텍스트 빈도 수 확인하는 워드 클라우드를 이용하여 연도별 많은 인기를 받은 가수를 분석합니다.

vynilist

6:28

연도별 장르 추세

2005년 기준

발라드

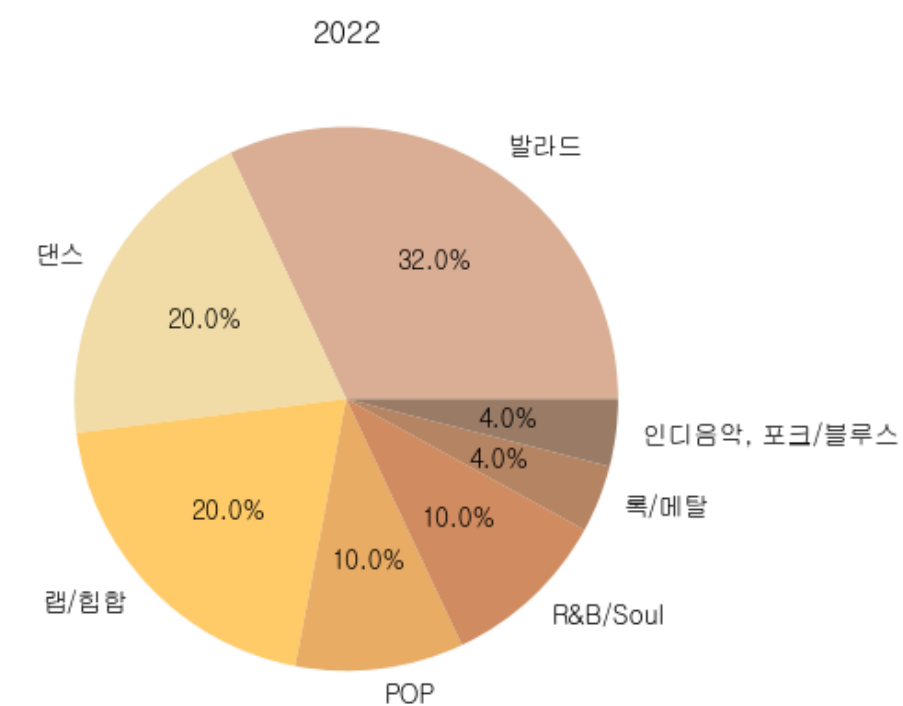
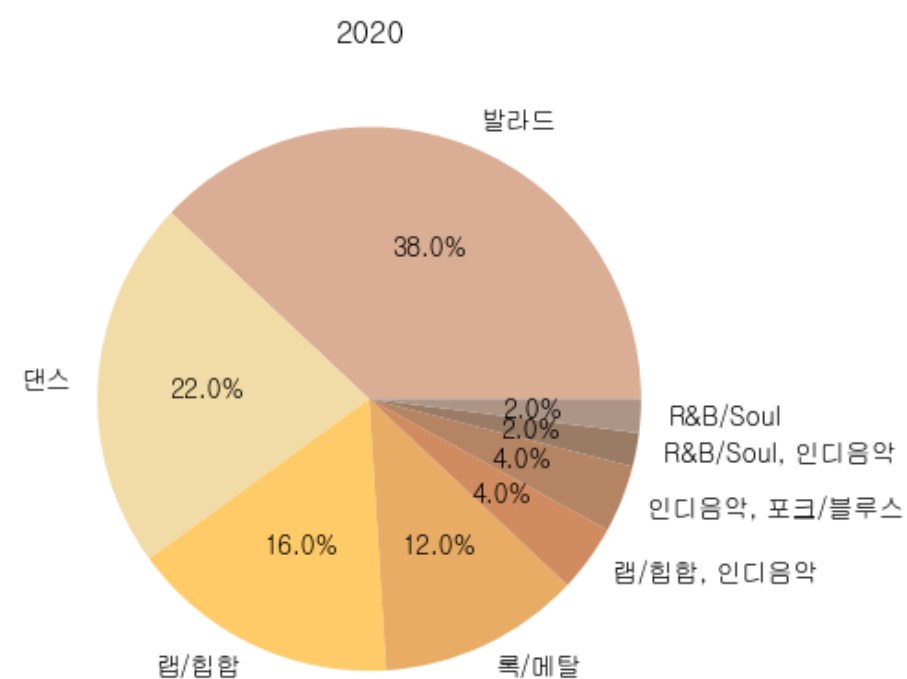
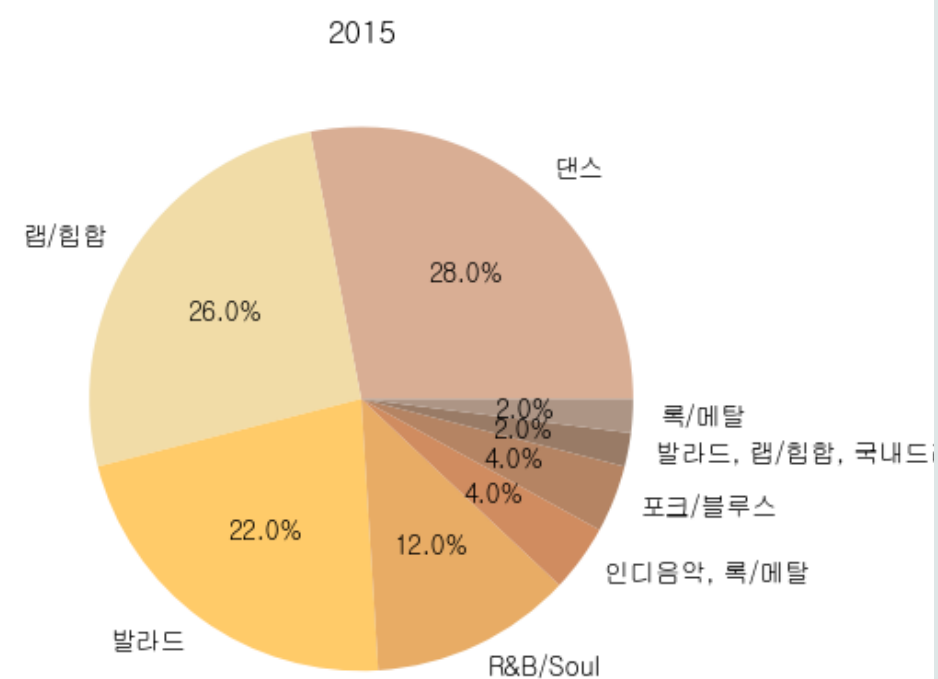
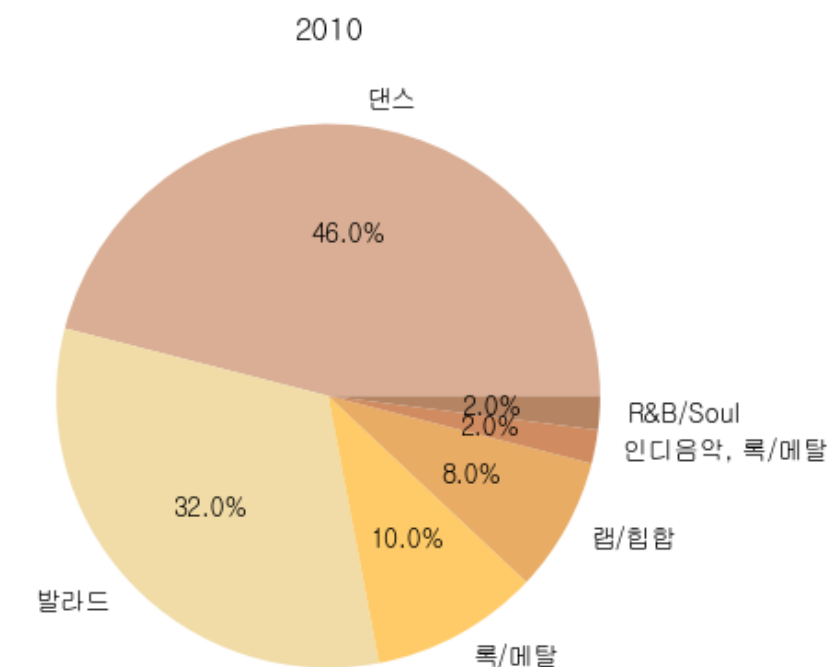
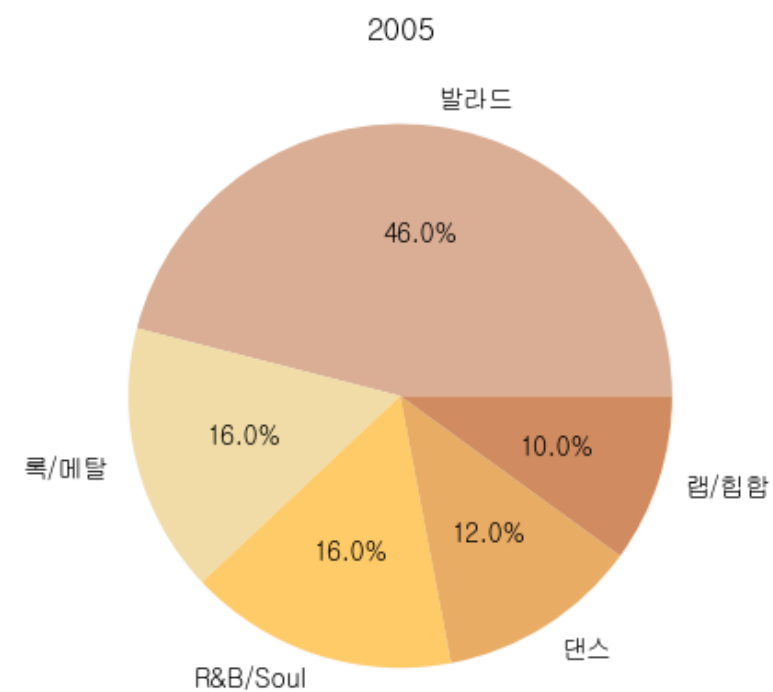
46%

파이 차트는 범주별 구성 비율을 단위 원에 부채꼴로 표현한 그래프입니다. 우선 각 영역의 비율과 이름을 ratio와 labels로 지정해 주고, matplotlib.pyplot 모듈의 pie() 함수에 순서대로 입력합니다.

```
ratio = df['장르'].value_counts(normalize=True)
```

```
labels = ratio.index
```


연도별 장르 추세



2005 vs 2022 장르

그래프 x축 위치 조정														
x축 위치	1	2	3	4	5	6	7	8	9	10	11	12	13	14
라벨	발라드			록/메탈			랩/힙합			댄스			R&B/Soul	
그래프 색상														

2005년 그래프와 2022년 그래프 두 개가 번갈아서 출력됩니다. 이때 간격은 3으로 하고, range형식을 이용합니다.

`df1['장르'].value_counts().sort_index(ascending=False)` 사용하여 장르별 개수를 내림차순으로 가져옵니다.

그리고 `plt.xticks()`를 사용하면 눈금 위치와 라벨을 넣을 수 있습니다.

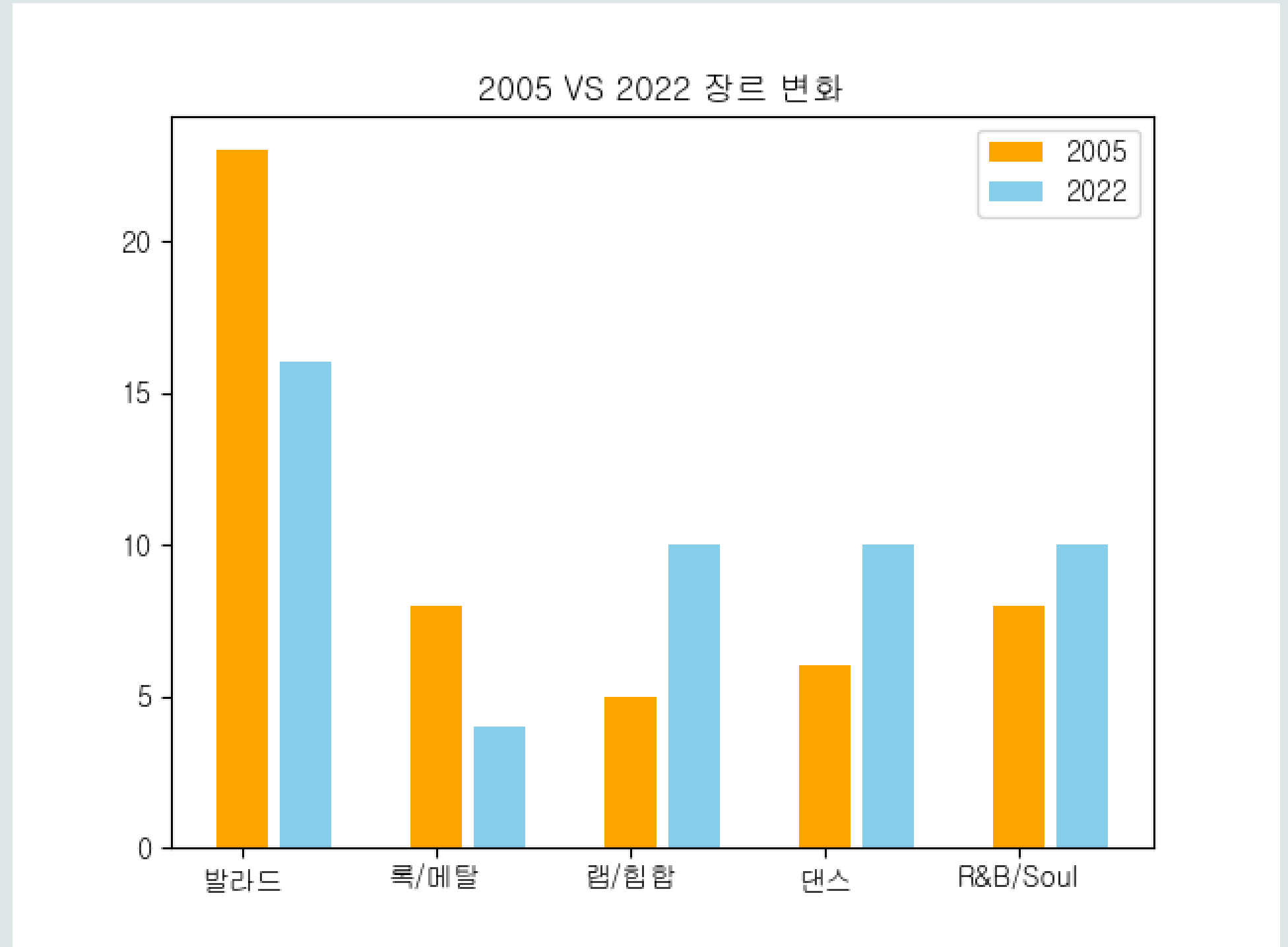
라벨 이름은 리스트로 만들어서 순차적으로 출력되게 합니다.

2005 vs 2022

장르 변화

check point

- ✓ 2005년 발라드 압도적
- ✓ 발라드 장르 꾸준한 인기
- ✓ 시간 지날수록 다양한 장르의 음악 시장



연도별 인기 가수

시각화 도구

Word Cloud

단어를 시각적으로 표현해주는 워드 클라우드를 사용합니다. 템플릿 정보에는 그림 크기 등의 정보를 넣어줍니다.

한 문장으로 변환

Join

csv를 불러와 가수 컬럼 기준으로 만들기 위해 join 함수를 사용하여 한 문장으로 변환합니다.

띄어쓰기 제거

replace

join 함수 사용할 때 공백을 기준으로 합치기 때문에 replace 함수를 통하여 가수 이름 사이에 공백을 제거합니다.

괄호 제거

re 모듈

간혹 가수 이름 뒤에 괄호가 붙은 경우가 있는데, 깔끔한 출력을 위해 정규표현식(re)을 사용해 제거합니다.



감사합니다