

# 음원 데이터 분석

219161123 정수용

# 목 차

01

주제 소개

02

데이터 수집

03

데이터 전처리

04

데이터 시각화 및 결론

# 주제 소개

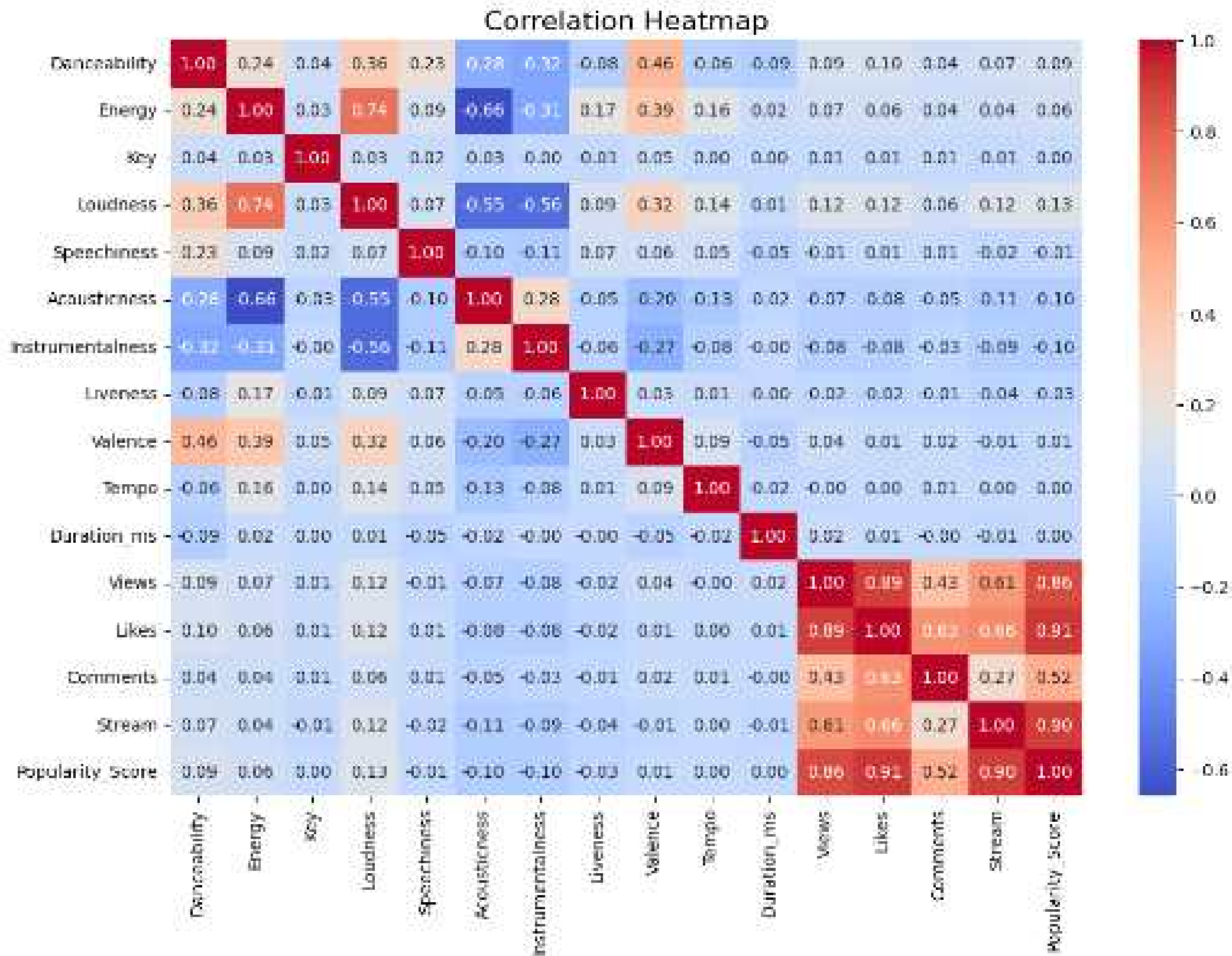
전 세계 다양한 아티스트들의 음원 데이터를 기반으로 음악  
트렌드를 분석하여 청취자들의 선호도에 미치는 영향을 파  
악

# 데이터 수집

kaggle 사이트에서 Spotify와 YouTube의 관한 데이터 수집

- Artist: 아티스트, Loudness: 곡의 음량(데시벨), Energy: 곡의 강렬함과 역동성  
Acousticness: 어쿠스틱 사운드 비율, Instrumentalness: 기악곡 여부 등 다양한 컬럼  
존재

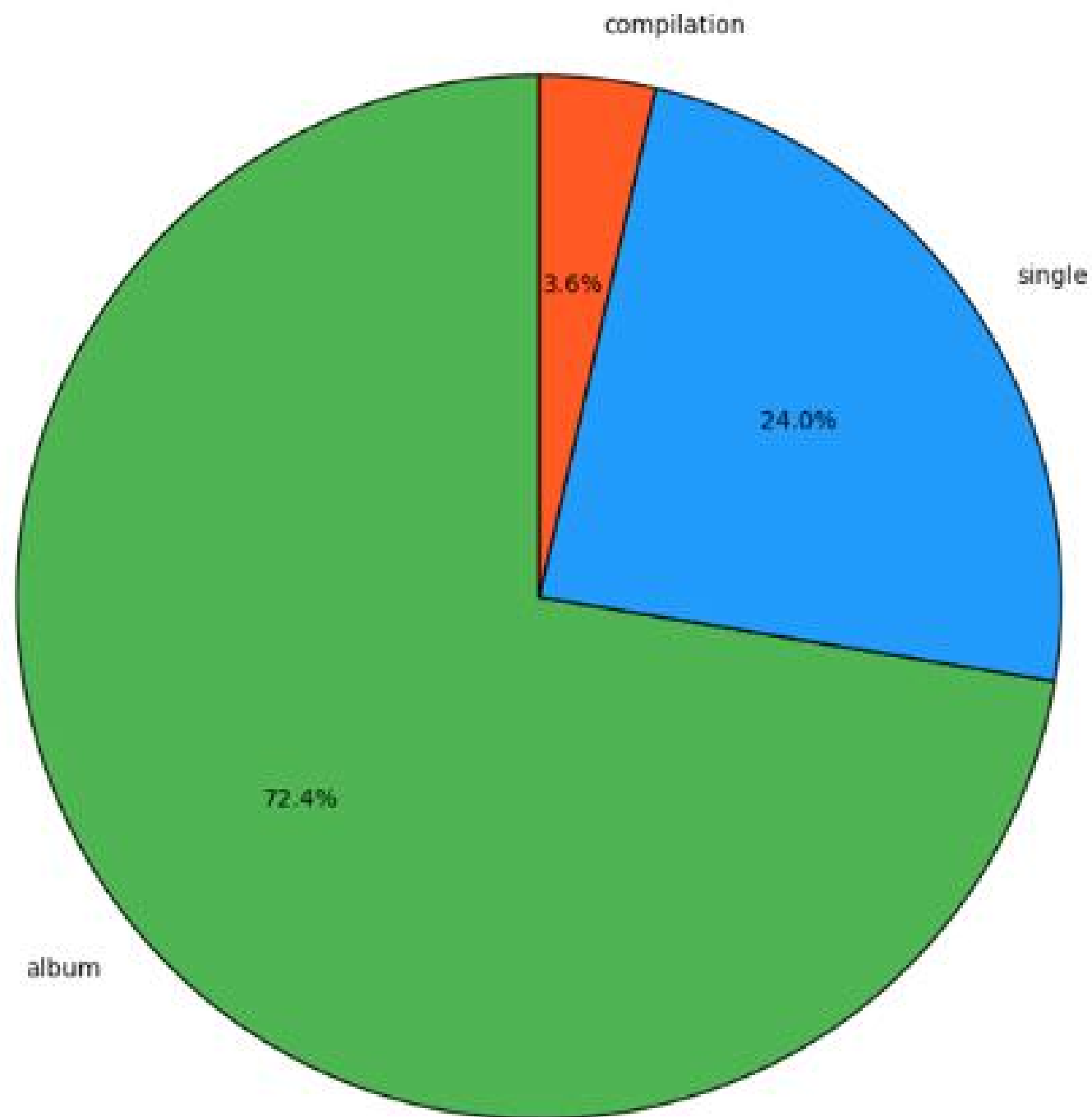
# 데이터 전처리



- 여러가지로 나뉘져 있던 인기 지표를 단일 인기 지표로 만들기 위해 결합

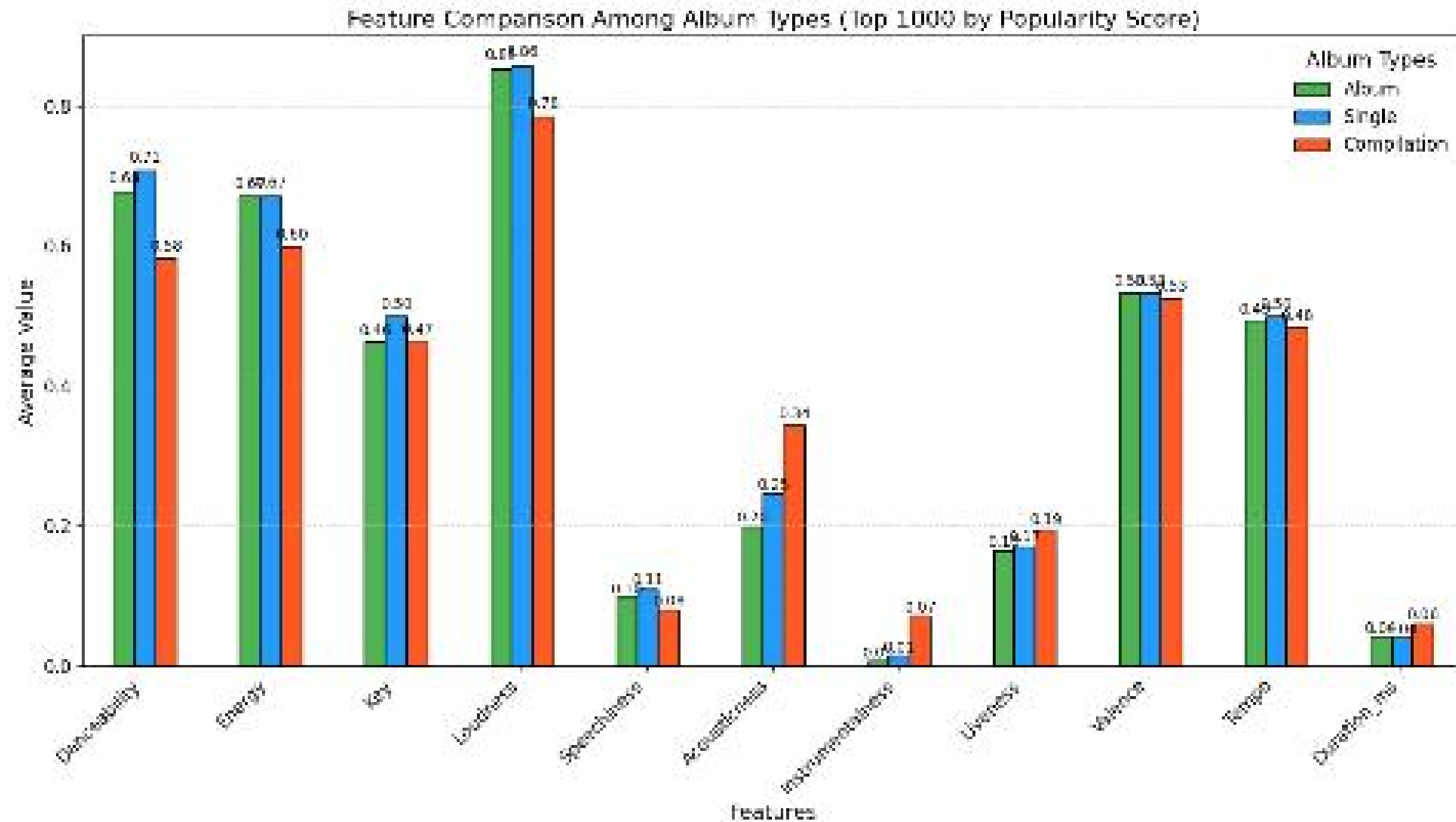
- 피어슨 상관계수 히트맵을 통하여 결합된 컬럼들의 상관관계가 매우 높음을 확인

Frequency of Album Types



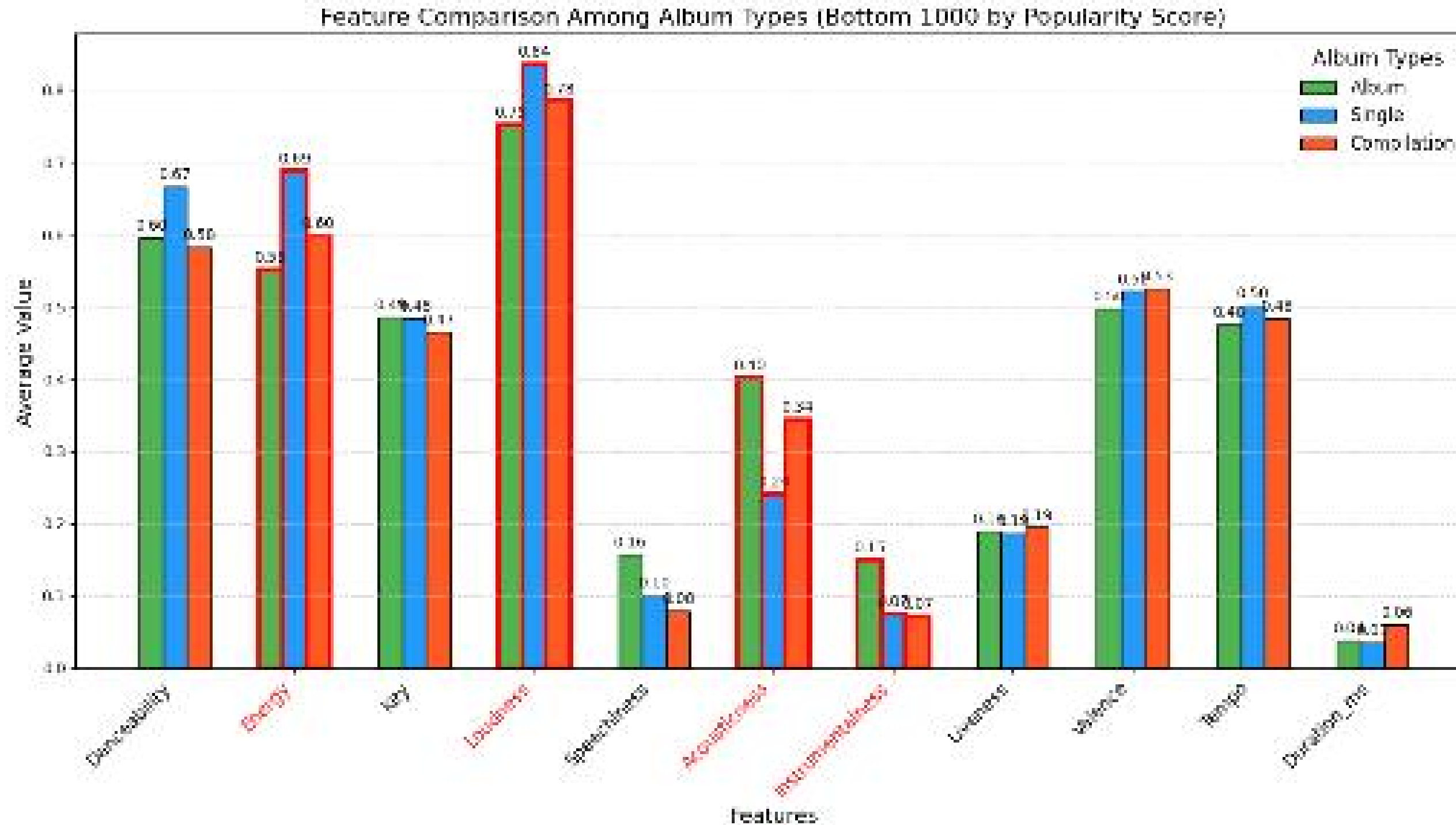
- Album Type 분포도를 확인 하였을때 Album(다수의 곡) 타입이 72.4%를 차지하고 13597개로 이루어져 있음 그 다음으로 Single(소수의곡))이 24%, Compilation(테마곡) 3.6%

# 데이터 시각화



Album\_type(노래 타입) 별로  
인기도를 기준으로 상위 1000  
개 곡을 뽑아 시각화

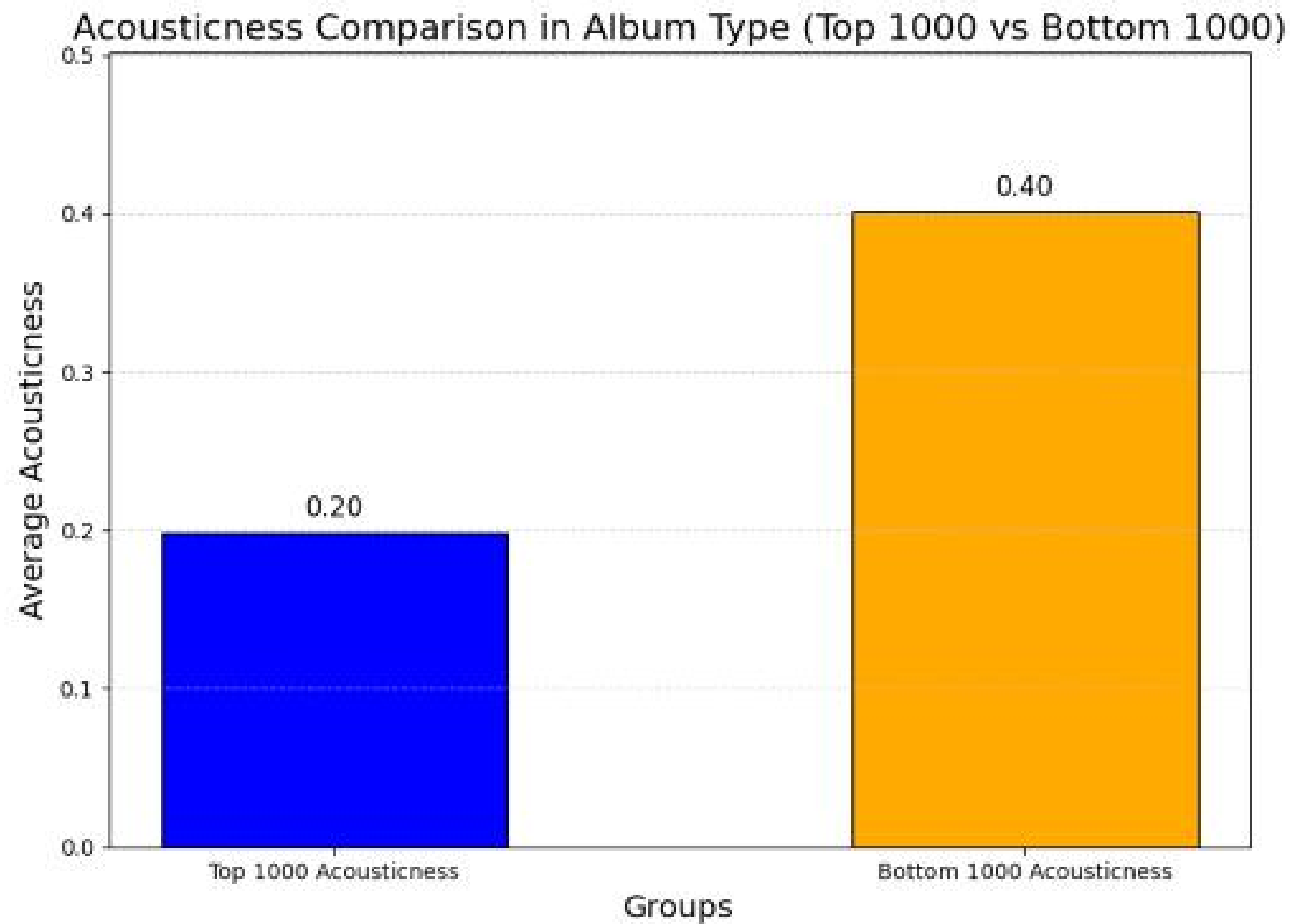
# 데이터 시각화



- 하위 1000개의 곡을 뽑아 시각화 후 비교
- Album Types 중 Album 장르에서 Loudness, Energy, Acousticness, Instrumentalness 컬럼의 차이를 확인



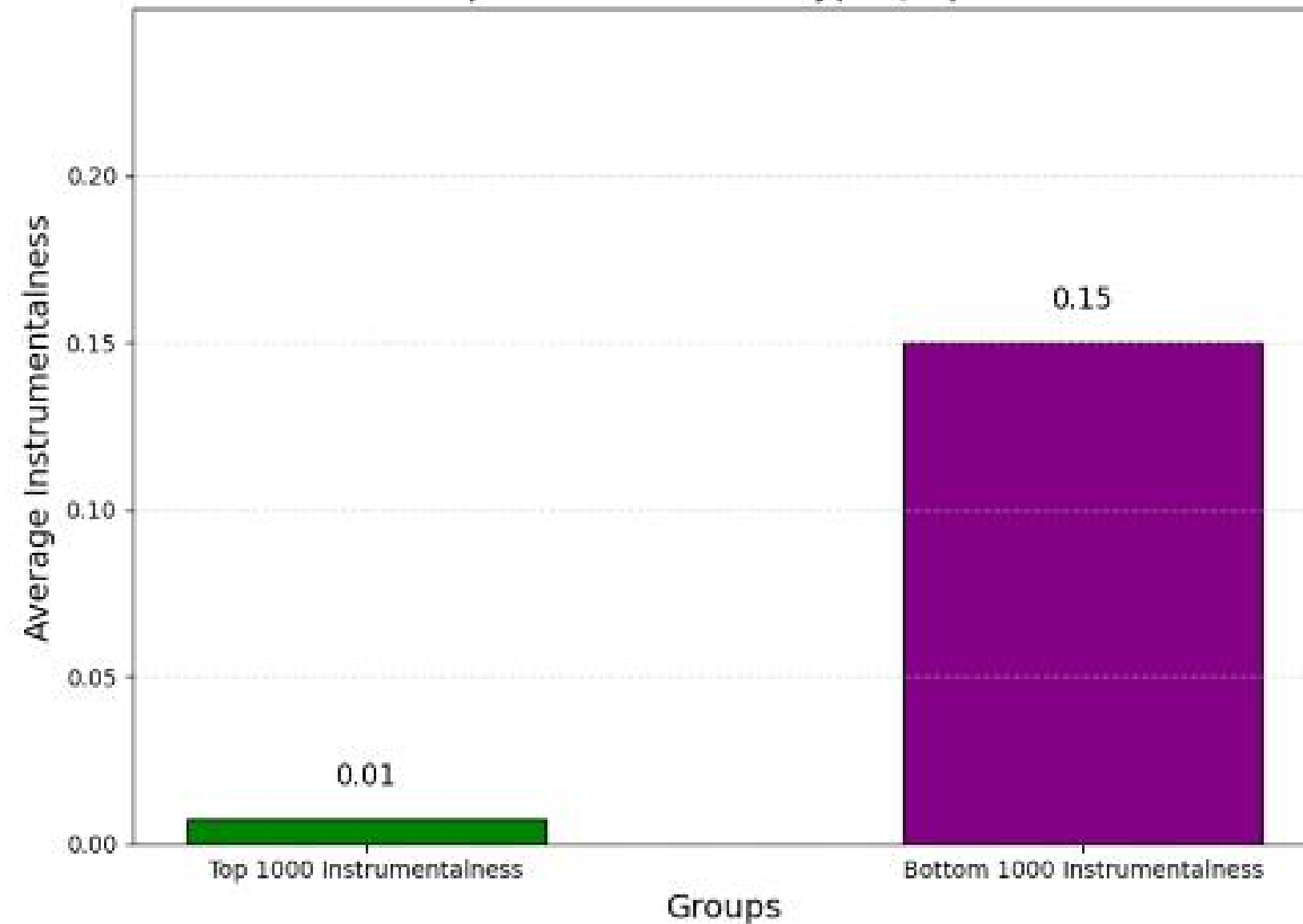
# 데이터 시각화



- Acousticness가 낮을수록 인기도가 높다
- 도출한 결과를 검정하기 위해 T-tset 검정 실시
- p 값이 유의수준을 넘지 못하여 도출한 결과가 **유의미함**을 검정

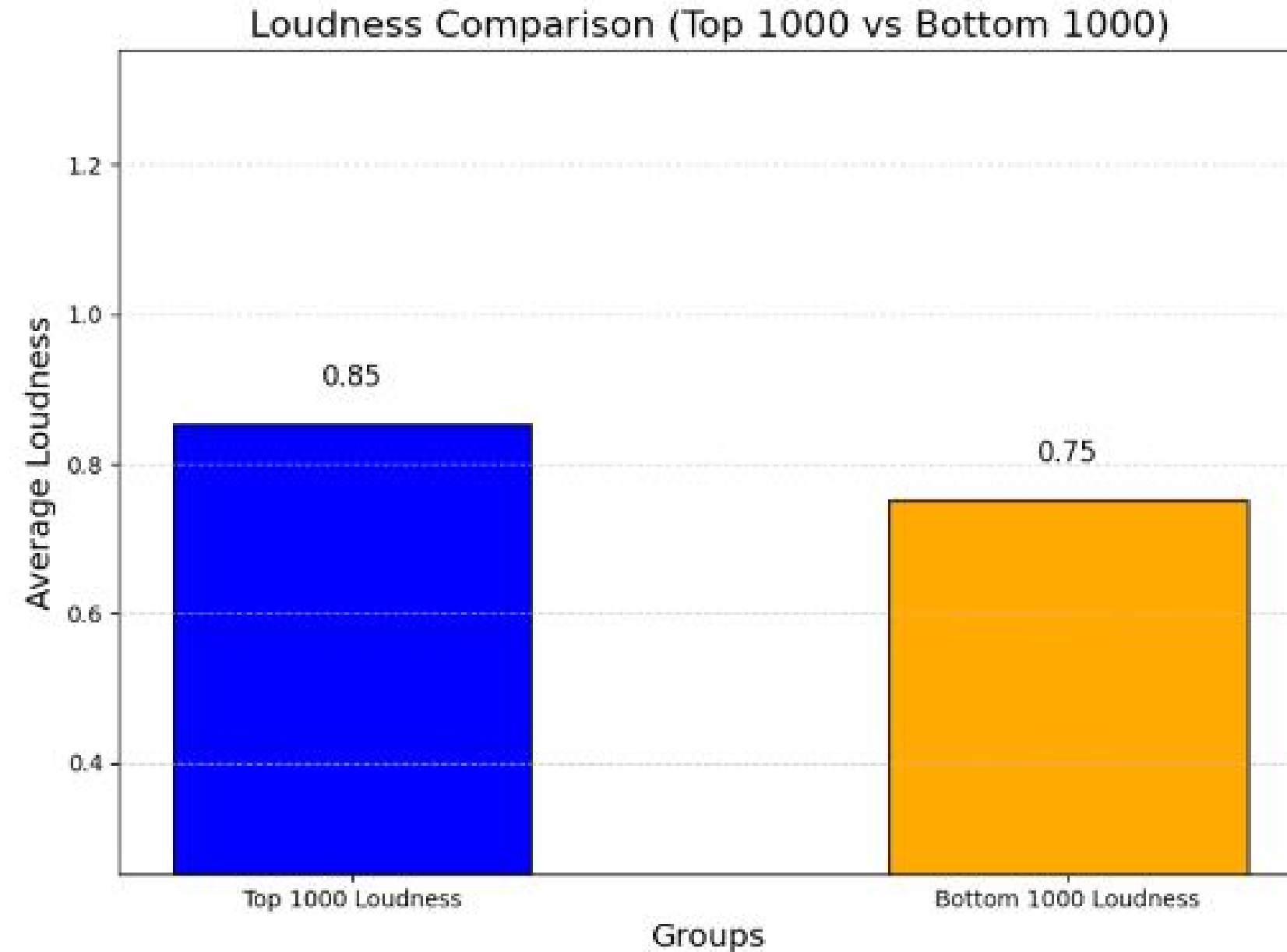
# 데이터 시각화

Instrumentalness Comparison in Album Type (Top 1000 vs Bottom 1000)



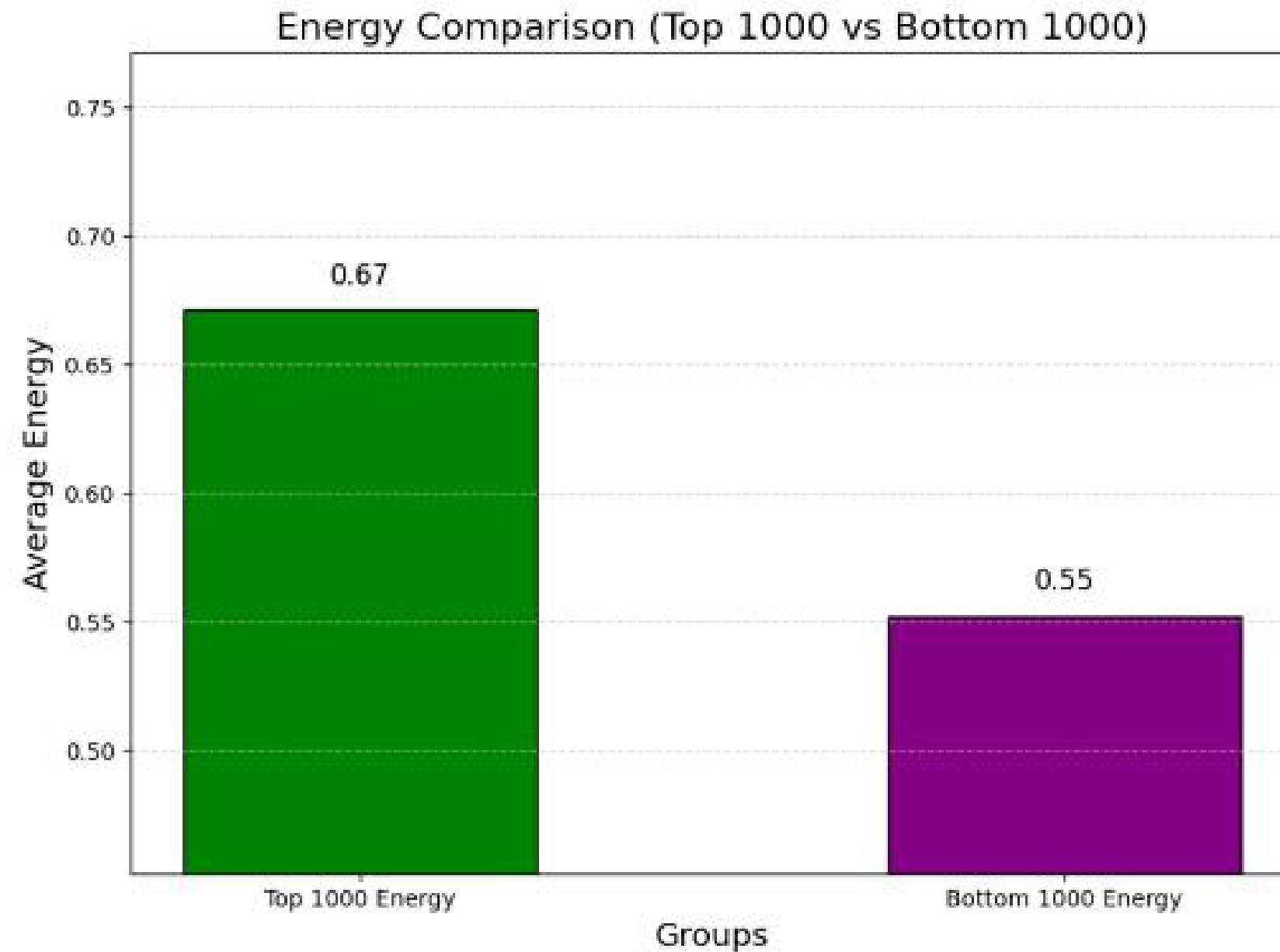
- Instrumentalness가 낮을수록 인기도가 높다
- 도출한 결과를 검정하기 위해 T-tset 검정 실시
- p 값이 유의수준을 넘지 못하여 도출한 결과가 **유의미함**을 검정

# 데이터 시각화



- Loudness가 높을수록 인기도가 높다
- 도출한 결과를 검정하기 위해 T-tset 검정 실시
- p 값이 유의수준을 넘지 못하여 도출한 결과가 **유의미함**을 검정

# 데이터 시각화



- Energy가 높을수록 인기도가 높다
- 도출한 결과를 검정하기 위해 T-tset 검정 실시
- p 값이 유의수준을 넘지 못하여 도출한 결과가 **유의미함**을 검정

# 결론

1. 대부분의 Artist들은 Album\_type중 Album 타입의 곡을 많이 냄
2. T-test 검정을 통해 Acousticness, Instrumentalness 특성은 낮을수록 Loudness, Energy 특성은 높을수록 인기도가 높다는 인사이트를 도출함

# 결론 및 향후 계획

1. 대부분의 Artist들은 Album\_type 중 Album 타입의 곡을 많이 냄
2. T-test 검정을 통해 Acousticness, Instrumentalness 특성은 낮을수록 Loudness, Energy 특성은 높을수록 인기도가 높다는 인사이트를 도출함
3. 현재는 Spotify와 YouTube 데이터 밖에 없지만 다양한 음악 플랫폼 데이터를 수집하여 보강
4. 도출한 인사이트를 토대로 머신러닝을 통해 인기도가 높은 음악 예측

감사합니다