

1번 시각화의 목적과 효과 평가

목적 :

1 변수 간의 상관관계 파악

- 히트맵은 각 변수들 간의 상관관계를 시각적으로 보여준다. 이를 통해 데이터셋의 전반적인 구조와 변수들 간의 관계를 이해하는데 도움을 준다.

2 데이터 탐색 및 인사이트 도출 :

- 히트맵을 통해 강한, 약한 상관관계를 빠르게 파악할 수 있다.

효과 :

1 직관적인 시각화 :

- 색상의 변화를 통해 상관관계 강도를 직관적으로 파악할 수 있다. 더 쉽게 이해할 수 있다 숫자로 된거 보다는

2 데이터의 전반적인 패턴 이해 :

- 데이터셋 전체의 특성과 패턴을 한 눈에 볼 수 있어, 이후의 심화 분석이나 모델링 방향을 설정하는데 유용하다.

개선점 제안

- 상관관계가 높거나 낮은 변수들에 대한 강조 : 0.4 ~ 0.8이상인 경우 강조 표시하여 중요한 관계를 더 쉽게 파악
- 컬러 조정 : 음의 상관관계는 차가운 색 으로 표시하여 더 쉽게 구분

2번 시각화 평가 및 개선

목적 및 효과

목적 : 산점도의 목적은 곡의 기악적 성향과 인기도 간의 관계를 시각적으로 나타낸 것이고 곡의 길이 와 가사의 부적절함 여부에 따라 이 관계가 어떻게 달라지는지 분석합니다.

효과

- 여러 변수를 동시에 시각화하여 데이터의 복잡한 관계를 수비게 이해할 수 있다.
- 각 변수 간의 패턴을 시각적으로 확인하여 새로운 인사이트를 얻을 수 있다.

개선점

1 마커 크기 조정

- 마커 크기 범위를 더 크게 조정하여 'duration_ms' 에 따른 차이를 명확히 표현했습니다

2 색상 팔레트 변경, 투명도

- palette='viridis' 색상 팔레트를 사용하여 'duration_ms'에 따른 차이를 더 잘 구분할 수 있도록 했고 투명도를 추가하여 겹침 현상을 줄였습니다.

3 explicit 변수 표현 방법

- True 는 X, False 는 O로 표현하여 한 눈에 알아볼 수 있게 했습니다.

3번 시각화를 통해 이를 정당화

'explicit' 과 'popularity' 간의 상관 관계가 거의 없다는 결론을 내리기 위해 두 가지 그래프를 사용했습니다 평균 인기도에 대한 막대 그래프와 박스 플롯.

정당화

막대그래프 : 이 그래프는 explicit 여부에 따른 평균 인기도를 보여준다.

막대의 높이 비교 : explicit 인 False인 곡의 평균 popularity 와 True인 곡의 평균 popularity 가 거의 동일합니다 > 평균값이 비슷하다는 것은 explicit 여부가 곡의 인기도에 큰 영향을 미치지 않는다는 것을 의미

박스 플롯

중앙값 : explicit 이 False인 곡과 True인 곡의 popularity 중앙값이 거의 동일

사분위수 범위 :사분위수 범위 도 비슷하다 이는 데이터의 중간 50%가 비슷한 범위 내에 분포하고 있음을 나타낸다

극단값 : 일부 극단값이 있지만 전체적인 분포는 두 그룹 간에 큰 차이가 없다.

4번 내릴 수 있는 결론

인기도와 다른 변수들 간의 상관 관계가 매우 낮아, 특정 변수가 곡의 인기도에 결정적인 영향을 미치지 않는다.