



7회차 과제 - 시각화

Date	@2024년 8월 1일 오후 11:59
Tag	과제

[물음1] 1번 시각화의 목적과 효과를 평가하고 개선점을 제안해주세요.

- 목적과 효과 두 가지 이상을 설명해주세요.
 1. 2차원 배열 형태의 데이터 시각화
 - 목적: heatmap은 2차원 배열 형태의 데이터를 색상을 이용해 시각화하기 위해 사용합니다.
 - 효과: 하나의 시각화로 다차원의 데이터를 표현할 수 있습니다.
 2. 데이터 내 변수들 간의 관계 파악
 - 목적: heatmap은 데이터 내 각 변수들 간의 상관관계를 통해 변수 간 관계를 파악하기 위해 사용합니다.
 - 효과: 상관관계가 색상으로 표현됨에 따라 상대적으로 강한 / 약한 상관관계를 가지는 변수들을 효과적으로 파악할 수 있습니다.
- 개선점 두 가지 이상을 설명해주시고 개선점을 반영한 코드를 작성해주세요.
 1. 각 셀에 쓰여진 상관계수의 소수점 자릿수 및 폰트 크기로 인해 상관계수 값의 가독성이 떨어집니다.
 2. 양의 상관계수와 음의 상관계수가 모두 같은 계열의 색상으로 표시됨에 따라 명확하게 구분되지 않습니다.

[물음2] 2번 시각화의 목적과 효과를 평가하고 개선점을 제안해주세요.

- 목적과 효과 한 가지 이상을 설명해주세요.
 - 목적: popularity와 instrumentalness 사이의 관계를 분석하고, 이 관계가 duration_ms 및 explicit에 따라 어떻게 달라지는지를 시각화합니다
 - 효과: popularity와 instrumentalness의 분포를 시각적으로 확인함으로써, instrumentalness가 낮은 쪽에 많은 곡들이 몰려있으며, 이로 인해 popularity와 instrumentalness 사이의 상관관계를 파악하기 어려움을 알 수 있습니다.

- 개선점을 두 가지 이상을 설명해주시고 개선점을 반영한 코드를 작성해주세요.

1. 다른 데이터에 의해 가려져있는 데이터가 많습니다.
2. 각 데이터의 특성에 맞는 범례(hue, size)가 선택되지 않았습니다.

[물음3] explicit이 popularity에 영향을 주는지 주지 않는지 판단하고 시각화를 통해 이를 정당화하세요.

- 판단과 정당화에 대한 설명을 작성해주세요.

explicit에 따른 popularity의 분포를 Box plot을 통해 확인해본 결과, 두 분포의 중위수와 사분위수 범위가 매우 유사하며, 이상치가 하단에 몰려있다는 점에서 두 그룹은 전체적으로 매우 유사한 분포를 보이고 있음을 확인할 수 있습니다. 따라서 explicit은 popularity에 유의미한 영향을 미치지 않음을 보여줍니다.

[물음4] 0725_visualization.ipynb에서 spotify 데이터를 시각화하여 내릴 수 있는 결론을 설명해주세요.

- heatmap에서 상관계수가 0.3 이상인 feature 집합을 파악함으로써 상관관계를 가지는 feature들을 알아낼 수 있습니다.
 - duration_ms와 year
 - explicit과 speechiness
 - danceability와 valence
 - energy와 valence
 - energy, acousticness, loudness