

Python 2. beadandó feladat

ELTE IK PNYFT – 2025-26 1

Spotify-Youtube adatelemzés

Ebben a beadandóban zeneszámok Spotify-Youtube közötti különbségeit fogjuk vizsgálni és vizualizálni.

A beadandót .ipynb formátumú fájlban kell beadni. A Jupyter Notebook formátumú fájlban minden feladat legyen külön kódcellára szétválasztva.

A mellékelt fájlok tartalma

spotify_yt.csv

A mellékelt fájl szerkezete

- **Track:** Zeneszám címe
- **Artist:** Előadó neve
- **Url_spotify:** Előadóhoz vezető Spotify link
- **Album:** Zeneszám albuma
- **Album_type:** Album/egyedülálló lemez/összeállítás része-e a zeneszám
- **Uri:** Zeneszám Spotify linkje
- **Danceability:** Zeneszám táncolhatóságát írja le (0-1 skálán pl. tempó, ritmus alapján: 1-es jelenti, ha lehet rá táncolni)
- **Energy:** Zeneszám intenzitását írja le (0-1 skálán – 1 az energikus)
- **Key:** Zeneszám milyen hangnemben van megírva 0 = C, 1 = C#/D \flat , 2 = D, stb. Ha nem érzékelt a rendszer hangnemet, az érték -1.
- **Loudness:** Zeneszám hangosságát méri decibelben (dB). Általában -60 and 0 dB közötti értékekkel.
- **Speechiness:** Zeneszámban szöveg mennyiséget érzékeli
- **Acousticness:** Zeneszám akusztikusságát méri (1 – akusztikus)
- **Instrumentalness:** Zeneszám vokális/zenei részét méri (1 – inkább csak zene szól, minthogy énekelnének benne)
- **Liveness:** Megmutatja, hogy milyen valószínűséggel volt élőben előadva a zeneszám
- **Valence:** Zeneszám hangulatát adja meg (1 – pozitív hangulatú)
- **Tempo:** Zeneszám becsült tempója (basszus/perc alapján)
- **Duration_ms:** Zeneszám hossza milliszekundumban.
- **Stream:** Spotify streamek száma
- **Url_youtube:** Youtube videó elérése (ha van)
- **Title:** Youtube videó címe
- **Channel:** Youtube videót publikáló csatorna
- **Views:** Youtube videó megtekintései
- **Likes:** Youtube videón lévő lájkok száma
- **Comments:** Youtube videón lévő kommentek száma

- **Description:** Youtube videó leírása
- **Licensed:** Youtube csatorna hivatalos Youtube partner-e (ellenőrzött-e)
- **official_video:** Van-e hivatalos videó a zeneszámról

Feladatok

1. Importáljuk be az adatelemzéshez szükséges könyvtárakat! (1p)
2. Olvassuk be a mellékelt **spotify_yt.csv** fájlt és tároljuk el egy DataFrame-ben a fájl tartalmát! (1p)
3. Írassuk ki az adatkeretre vonatkozó statisztikákat és információkat! (1p)
4. Számoljuk meg, hogy oszloponként mennyi null érték fordul elő! (1p)
5. Töltsük fel a hiányzó adatokat Loudness, Acousticness, Instrumentalness, Energy, Danceability oszlopokban!
Feltöltéskor vegyünk figyelembe, hogy az előadóhoz tartozó számok milyen átlagot alkotnak az adott oszlopokban, és azzal töltsük fel a hiányzó értékeket!
Ellenőrizzük az előző feladat műveletével, hogy sikerült-e a feltöltés! (2p)
6. Írassuk ki azokat a **sorokat**, ahol a Speechiness, Liveness, Valence, Tempo, Key vagy Duration_ms oszlopok értékeiben null érték van!
Kiíratás után dobjuk el ezeket a sorokat helyben, majd ellenőrizzük, hogy sikerült-e a null értékek eltávolítása az adott oszlopokból. (1p)
7. Írassuk ki a DataFrame oszlopainak az adattípusait!
Módosítsuk az adatszerkezetet úgy, hogy a Duration_ms, Key, Views, Likes, Comments, Stream oszlopok adattípusát int-re konvertáljuk, a Licensed, official_video oszlopok értékét boolean típusra konvertáljuk!
Ha az érték null érték lenne, adjunk meg -1-et az inteknél, illetve False-t a boolean adatoknál! (1p)
8. Dobjuk el azokat a sorokat, ahol a Stream oszlopban -1 található!
Dobjuk el az Url_spotify, Uri oszlopokat is! (1p)
9. Mentjük el a tisztított adatokat egy **spotify_yt_cleaned.csv** nevű fájlba! (1p)
10. Számoljuk meg, hogy hány zeneszámmal volt található Youtube videó! (1p)
11. Keressük meg melyik az a 10 előadó, akinek a zenéjére lehetne a leginkább táncolni!
A rangsor állításakor azt vegyük figyelembe, hogy átlagosan milyen Danceability értékei vannak az előadó számainak. (1p)
12. Határozzuk meg és írassuk ki a 10 legtöbbet streamelt előadót a 10 legtöbbet nézett előadóval együtt!
Ábrázoljuk **egyetlen** diagrammon, két tengellyel a 10 legtöbbet streamelt előadót és hasonlítsuk össze oszlopdiaagrammokkal, hogy mennyi Youtube megtekintésük van a Streamek száma mellett!
Figyeljünk a helyes megjelenítésre! Tegyük részben átlátszóvá az előbb álló oszlopot!
Jelenítsük meg jelmagyarázatot legalább az egyik oszlopra! (3p)
13. Ábrázoljuk korrelációs mátrixsal, hogy milyen zenei tulajdonságok járultak hozzá a 10 legnépszerűbb zeneszámmal (streamelések alapján!)
Figyeljünk oda, hogy a számokat tartalmazó oszlopokkal dolgozzunk csak!
Ügyeljünk a jól látható megjelenítésre! (2p)

14. Döntsük el, hogy kinek van a legtöbb streamelése Spotify-on és megtekintése Youtube-on J. Cole, Kendrick Lamar és Drake közül a jelenlegi adathalmazunk alapján!

A szűréskor figyeljünk arra, hogy az említett előadók zeneszámok nevében is feltűnhetnek és hogy ne legyenek duplikátumok az adataink között (tekinthetjük úgy, hogy a zeneszám neve egyedi).

Ábrázoljuk az eredményeket egymás mellett két kördiagrammon, ahol megjelenítjük a neveket a körcikkek mellett, egyedi színeket adunk a körcikkeknek, majd nevet adunk a kördiagrammoknak!

Állítsunk be tetszőleges háttérrel az egész diagrammnak! (3p)

Kérdés esetén a gyakorlatvezetőhöz vagy az előadás vezetőjéhez lehet fordulni.

Az elérhető maximális pontszám: **20 pont**

Jó munkát!