



**ACADEMIA DE STUDII ECONOMICE BUCUREȘTI  
FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**

# **Analiza celor mai redate melodii de pe platforma de streaming Spotify folosind pachete software**

**Profesor coordonator:**

**Asistent de cercetare Științifică**

**NICULAE ANDREEA-MIHAELA**

**Studenti:**

**Militaru Mihaela Felicia**

**Mitu Adrian-Constantin**

**București 2024**

<b>Prezentarea companiei.....</b>	<b>3</b>
<b>Programare Python.....</b>	<b>4</b>
Problema 1: Cele mai populare melodii all-time.....	5
Problema 2: Câta câștigat un anumit artist pe piesele lansate în 2023?.....	6
Problema 3: Colecții de date în Python.....	7
Problema 4: Folosirea structurilor repetitive.....	10
Problema 5: Folosirea iloc și loc.....	11
Problema 6: Statistică despre muzica populară.....	12
Problema 7: Spotify Wrapped.....	12
Problema 8: Spotify Wrapped.....	14
<b>Programare SAS.....</b>	<b>15</b>
Problema 1: Crearea unui set de date SAS din fișiere externe.....	15
Problema 2: Crearea și folosirea de formate definite de utilizator.....	16
Problema 3: Procesarea iterativă și condiționată a datelor.....	17
Problema 4: Utilizarea de funcții SAS.....	18
Problema 5: Crearea de subseturi de date.....	20
Problema 6: Combinarea seturilor de date prin proceduri specifice SAS și SQL.....	21
Problema 7: Folosirea de proceduri statistice.....	21
Problema 8: Generare grafice.....	24

## Prezentarea companiei

**Spotify** este o platformă globală de streaming muzical, fondată în 2006 în Suedia de către **Daniel Ek** și **Martin Lorentzon**. În scurt timp, Spotify a devenit una dintre cele mai populare și influente platforme de streaming audio din lume. Cu peste **345 de milioane de utilizatori activi lunar** în peste **90 de țări**, inclusiv abonați la serviciul premium și utilizatori care utilizează versiunea gratuită suportată de reclame, Spotify s-a impus ca **lider** al industriei muzicale digitale.

Principalul produs al Spotify este **serviciul său de streaming audio**, care oferă utilizatorilor acces la o vastă bibliotecă de milioane de melodii, albume și podcast-uri din întreaga lume. Utilizatorii pot explora și asculta muzică de la artiști consacrați și talente emergente, crea liste de redare personalizate și descoperi noi piese în funcție de preferințele lor muzicale.

Spotify oferă două tipuri principale de servicii: un serviciu **gratuit**, care permite ascultarea muzicii cu reclame între melodiile selectate, și un serviciu **premium**, care elimină reclamele și oferă funcționalități suplimentare, cum ar fi ascultarea offline și calitatea audio îmbunătățită.

În plus față de streamingul muzical, Spotify s-a extins și în domeniul podcast-urilor, devenind una dintre cele mai mari platforme de podcasting din lume. Compania a investit semnificativ în conținut original și parteneriate cu creatorii de conținut pentru a oferi o gamă diversă de podcast-uri pe teme variate, de la știri și politică până la divertisment și educație.

În prezent, Spotify continuă să inoveze și să își extindă oferta de produse și servicii pentru a rămâne în fruntea industriei de streaming audio și pentru a satisface nevoile și preferințele în continuă schimbare ale publicului său global.

## Prezentarea metodologiei proiectului

Acest set de date preluat de pe **Kaggle** conține o listă cuprinzătoare a celor mai cunoscute piese muzicale din 2023, așa cum sunt enumerate pe Spotify. Sunt furnizate informații cuprinzătoare despre fiecare piesă, popularitatea și prezența acesteia pe diverse platforme muzicale. Setul de date include informații precum numele piesei, numele artistului, data lansării, liste de redare și clasamente Spotify, statistici de streaming sau diverse caracteristici audio.

## **Variabilele**

**Denumire:** Numele piesei muzicale.

**Artist:** Numele artistului sau artiștilor care interpretează piesa.

**NrArtisti:** Numărul de artiști implicați în interpretarea piesei.

**AnLansare:** Anul lansării piesei.

**LunaLansare:** Luna lansării piesei.

**ZiLansare:** Ziua lansării piesei.

**SpotifyPlaylists:** Numărul de apariții în playlist-urile Spotify.

**SpotifyTops:** Numărul de apariții în topurile Spotify.

**Streams:** Numărul total de stream-uri.

**ApplePlaylists:** Numărul de apariții în playlist-urile Apple Music.

**AppleTops:** Numărul de apariții în topurile Apple Music.

**DeezerPlaylists:** Numărul de apariții în playlist-urile Deezer.

**DeezerTops:** Numărul de apariții în topurile Deezer.

**ShazamTops:** Numărul de apariții în topurile Shazam.

**Bpm:** Tonica și tonalitatea principală a piesei.

**Cheie:** Tonalitatea principală a piesei.

**Mod:** Modul principal al piesei (Major sau Minor).

**Densabilitate:** Gradul în care o piesă muzicală este potrivită pentru dans, exprimat ca procentaj.

**Valența:** Gradul de pozitivitate sau veselie al unei piese muzicale, exprimat ca procentaj.

**Energie:** Nivelul de energie al unei piese muzicale, exprimat ca procentaj.

**Acusticitate:** Gradul de acusticitate al unei piese muzicale, exprimat ca procentaj.

**Instrumentalitate:** Gradul de instrumentalitate al unei piese muzicale, exprimat ca procentaj.

**Liveness:** Gradul în care o piesă muzicală pare să fie înregistrată live, exprimat ca procentaj.

**Speechiness:** Gradul în care o piesă muzicală conține elemente vorbite sau recitate, cum ar fi rapul, exprimat ca procentaj.

## Programare Python

Înainte de a realiza prelucrarea datelor, acestea au fost importate dintr-un fișier CSV, iar valorile lipsă au fost înlocuite cu media pe coloană a variabilei respective.

```
#importul unei fisier csv sau json în pachetul pandas
data = pd.read_csv("./spotify-2023.csv", encoding='latin-1')
print(data.head(5))

# Identificam coloanele numerice
numeric_cols = data.select_dtypes(include=['number']).columns
# Umplem valorile lipsă numai în coloanele numerice cu media valorilor pe coloana
data[numeric_cols] = data[numeric_cols].fillna(data[numeric_cols].mean())
```

### Problema 1: Cele mai populare melodii all-time

Echipa de promovare pe social media dorește să realizeze o postare pentru a informa următorii conturii Spotify despre **cele mai ascultate melodii all time** dintre cele mai populare ale anului. Scopul este oferirea unei imagini de ansamblu asupra celor mai populare melodii, astfel încât să poată fi **captată** atenția publicului și utilizatorii să fie încurajați să **interacționeze** cu conținutul Spotify. Se vor a fi evidențiate cele mai populare **30 de melodii din toate timpurile**, oferindu-le fanilor ocazia să descopere sau să își reamintească piesele lor preferate.

Astfel, echipa de social media are nevoie de datele celor mai populare melodii pentru a putea realiza implementarea grafică a postărilor.

### Codul folosit:

```
#Problema 1: Analiza celor mai ascultate melodii all time
# Identificarea celor mai populare melodii
top_songs = data.sort_values(by='streams', ascending=False).head(30).reset_index(drop=True)
print("Top 30 songs on spotify all time")
print("-----")
pd.set_option('display.max_colwidth', None)
#print(top_songs[['track_name', 'artist(s)_name', 'streams']])
for index, row in top_songs.iterrows():
    print(f"{index + 1:2}. {row['track_name']:50} {row['artist(s)_name']:30} {row['streams']} streams")
print("-----")
```

### Prezentarea rezultatelor:

Top 30 songs on spotify all time		
1. Blinding Lights	The Weeknd	3703895074 streams
2. Shape of You	Ed Sheeran	3562543890 streams
3. Someone You Loved	Lewis Capaldi	2887241814 streams
4. Dance Monkey	Tones and I	2864791672 streams
5. Sunflower - Spider-Man: Into the Spider-Verse	Post Malone, Swae Lee	2808096550 streams
6. One Dance	Drake, WizKid, Kyla	2713922350 streams
7. STAY (with Justin Bieber)	Justin Bieber, The Kid Laroi	2665343922 streams
8. Believer	Imagine Dragons	2594040133 streams
9. Closer	The Chainsmokers, Halsey	2591224264 streams
10. Starboy	The Weeknd, Daft Punk	2565529693 streams
11. Perfect	Ed Sheeran	2559529074 streams
12. Heat Waves	Glass Animals	2557975762 streams
13. As It Was	Harry Styles	2513188493 streams
14. Senorita	Shawn Mendes, Camila Cabello	2484812918 streams
15. Say You Won't Let Go	James Arthur	2420461338 streams
16. lovely - Bonus Track	Billie Eilish, Khalid	2355719893 streams
17. Watermelon Sugar	Harry Styles	2322580122 streams
18. Don't Start Now	Dua Lipa	2303033973 streams
19. Lucid Dreams	Juice WRLD	2288695111 streams
20. Sweater Weather	The Neighbourhood	2282771485 streams
21. Thinking Out Loud	Ed Sheeran	2280566092 streams
22. Photograph	Ed Sheeran	2236667932 streams
23. Something Just Like This	The Chainsmokers, Coldplay	2204080728 streams
24. Bohemian Rhapsody - Remastered 2011	Queen	2197010679 streams
25. Shallow	Lady Gaga, Bradley Cooper	2159346687 streams
26. Take Me To Church	Hozier	2135158446 streams
27. Circles	Post Malone	2132335812 streams
28. Love Yourself	Justin Bieber	2123309722 streams
29. All of Me	John Legend	2086124197 streams
30. Counting Stars	OneRepublic	2011464183 streams

### Interpretarea rezultatelor:

Se poate observa faptul că, în mare parte, cele mai populare melodii sunt hiturile din ultimii 5-10 ani, populare pe toate platformele de streaming. Cei mai recurenți artiști din acest top sunt **Ed Sheeran**, **Harry Styles**, **The Weeknd** sau **Post Malone**. Totuși, există câteva piese care sunt mai vechi, cum ar fi **Bohemian Rhapsody**, care au reușit să apară în acest clasament. Pe de altă parte, putem observa faptul că primele 2 melodii, **Blinding Lights** și **Shape of You** sunt la o distanță destul de mare (aproape 7-9 miliarde de stream-uri) de următoarele clasate, care se află în range-ul de 20-28 mld. de redări.

### Problema 2: Cât a câștigat un anume artist pe piesele lansate în 2023?

Echipa de analiză a Spotify dorește să dezvolte o funcție în Python pentru a calcula **venitul total** al unui artist pentru melodiile sale populare din 2023. Funcția trebuie să primească **numele artistului**, să caute în dataframe toate melodiile la care apare artistul specificat și care au fost lansate în anul **2023**, să însumeze numărul total de streaming-uri și să returneze **venitul total** al artistului conform unui **tarif pe stream** de 0.05 dolari.

### *Codul folosit:*

```
def get_artist_revenue(artist_name):  
    rate_per_stream = 0.005  
    total_streams = data[data['artist(s)_name'].str.contains(artist_name, case=False)  
        & (data['released_year'] == 2023)]['streams'].sum()  
    total_revenue = total_streams * rate_per_stream  
  
    return total_revenue
```

### *Prezentarea și interpretarea rezultatelor:*

Am luat ca exemplu veniturile melodiilor cele mai populare ale lui Drake:

```
#apel functie  
artist_name = "Drake"  
total_revenue_drake = get_artist_revenue(artist_name)  
print(f"Venitul total al artistului {artist_name} in 2023 este: ${total_revenue_drake:.2f}")
```

**Venitul total al artistului Drake în 2023 este: \$875489.17**

### **Problema 3: Colecții de date în Python**

Un intern dorește să-și consolideze cunoștințele despre **colecțiile de date** în limbajul de programare Python. Pentru a face acest lucru, el decide să dezvolte o mică aplicație Python care să îl ajute să exploreze și să înțeleagă diferitele colecții de date încorporate.

#### *Codul folosit, prezentarea și interpretarea rezultatelor:*

##### **a. Liste**

```
songs_2013 = data[data['released_year'] == 2013]  
  
songs_2013 = songs_2013.sort_values(by='track_name')  
songs_2013 = list(songs_2013['track_name'])  
print("Most listened songs in 2023 from 2013:")  
print(songs_2013)  
  
songs_2013.pop(3) # stergem 'Do I Wanna Know?'  
print(songs_2013)
```

**Most listened songs in 2023 from 2013:  
['All of Me', 'Bound 2', 'Counting Stars', 'Do I Wanna Know?', 'Get Lucky - Radio Edit',**

```
, 'Wake Me Up - Radio Edit', "Why'd You Only Call Me When You're High?"]
```

## b. Tupluri

```
#tupluri
songs_the_weeknd = data[data['artist(s)_name'].str.contains('The Weeknd', case=False)]
songs_the_weeknd = tuple(songs_the_weeknd['track_name'])
print("The Weeknd popular songs:")
print(songs_the_weeknd)
print("After hours position in tuple: " + songs_the_weeknd.index('After Hours'))
```

```
The Weeknd popular songs:
('Popular (with Playboi Carti & Madonna) - The Idol Vol. 1
```

```
'Phantom Regret by Jim', 'Take My Breath')
```

```
After hours position in tuple:
14
```

## c. Dicționare

```
#dicționare
old_songs = data[data['released_year'] < 2005]
dict_old_songs = {}

for index, row in old_songs.iterrows():
    artist = row['artist(s)_name']
    song = row['track_name']
    dict_old_songs[artist] = song

print("\nDictionary: ")
print(dict_old_songs)
```

```
print(dict_old_songs.get("Eminem"))
```

```
Dictionary:
{'Chris Molitor': 'Yellow', 'Vance Joy': 'Riptide', 'Eminem': 'Lose Yourself
```



```
"Guns N' Roses": "Sweet Child O' Mine", 'Coldplay': 'Sparks', 'Southstar': 'Miss You'}
```

```
Lose Yourself - Soundtrack Version
```

#### d. Seturi

```
#seturi  
arist_names = data['artist(s)_name'].unique()  
arist_names = set(arist_names)  
  
print("\nSet:")  
print(arist_names)
```

```
arist_names.clear()  
print(arist_names)
```

```
Set:  
{'The Neighbourhood', 'TAEYANG, Lisa', 'Justin Quiles, Lenny Tavarez, BL',  
'Victor Cibrian', 'Em Beihold', 'GODZZ__-, Zakaria', 'Ana Castela, AgroPlay'}
```

```
set()
```

#### Problema 4: Folosirea structurilor repetitive

Un alt intern, care a făcut o tranziție dintr-o altă arie de programare, dar a vrut să ajute echipa de promovare pe social media oferindu-le 2 **fun fact-uri** pentru o postare, în timp ce recapitulează structurile alternative și repetitive. Acesta face un meniu la consolă în care poate afișa cele 2 informații preluate din setul de date sau poate ieși din meniu.

*Codul folosit și prezentarea rezultatelor:*

```

while(True):
    print("\n-----")
    print("Please tell me what would you like to do: ")
    print("1- Tell you the longest artist name")
    print("2- Tell you the oldest year for a song in this dataframe")
    print("3- You are bored and you want to exit")
    print("-----")
    print("Type your choice: ")
    choice = input()
    if int(choice) == 1:
        longest_artist_name = data['artist(s)_name'].apply(len).idxmax()
        longest_artist = data.loc[longest_artist_name, 'artist(s)_name']
        print("\nThe longest artist name is:", longest_artist)
    elif int(choice) == 2:
        oldest_year = data['released_year'].min()
        print("\nThe oldest year for a song in this dataframe is:", oldest_year)
    elif int(choice) == 3:
        print("\nExiting the program...")
        break
    else:
        print("Invalid choice! Please enter 1, 2, or 3.")

```

```

Please tell me what would you like to do:
1- Tell you the longest artist name
2- Tell you the oldest year for a song in this dataframe
3- You are bored and you want to exit
-----
Type your choice:
1

The longest artist name is: Jordan Fisher, Josh Levi, Finneas O'Connell, 4*TOWN (From Disney and Pixar

```

```

Type your choice:
2

The oldest year for a song in this dataframe is: 1930

```

```

Type your choice:
3

Exiting the program...

```

```
Type your choice:
4
Invalid choice! Please enter 1, 2, or 3.
```

### Problema 5: Folosirea iloc și loc

Echipa de dezvoltare încearcă să dezvolte un mecanism de selectare a melodiilor într-un mod aleatoriu și vrea să îl simuleze mai întâi pe un set de date cu cele mai populare melodii din 2023. Ca exemplu, se va selecta o melodie aleatorie după un număr cuprins între 1 și numărul de melodii și după procentul de dansabilitate.

#### Codul folosit și prezentarea rezultatelor:

```
#Problema 5 iloc si loc:
random_number = 1000
while (random_number > 954):
    print("Spune-mi un numar (mai mic de 954 te rog): ")
    random_number = int(input())
print(data[['track_name', 'artist(s)_name']].iloc[random_number])

random_procent = 101
while (random_procent > 100 or random_procent < 0):
    print("Spune-mi un numar intre 0 si 100: ")
    random_procent = int(input())
print(data[['track_name', 'artist(s)_name']].loc[data['danceability_%'] == random_procent])
```

```
Spune-mi un numar (mai mic de 954 te rog):
54
track_name      Another Love
artist(s)_name   Tom Odell
Name: 54, dtype: object

Spune-mi un numar intre 0 si 100:
45
      track_name  artist(s)_name
54      Another Love      Tom Odell
147  Somewhere Only We Know      Keane
163      Enchanted      Taylor Swift
167  The Night We Met      Lord Huron
292  Call Out My Name      The Weeknd
448  Let It Snow! Let It Snow! Let It Snow!  Dean Martin
493      Too Late      SZA
591  Before You Go      Lewis Capaldi
609  Happier Than Ever - Edit      Billie Eilish
625      Arcade      Duncan Laurence
731  Fuera del mercado      Danny Ocean
756      Golden      Harry Styles
871  Sweet Child O' Mine      Guns N' Roses
```

## Problema 6: Statistică despre muzica populară

În cadrul unui parteneriat pe care îl are cu o universitate de prestigiu din SUA, Spotify s-a angajat să furnizeze anumite date pentru facultățile care studiază muzica și artele adiacente acesteia. Una dintre informațiile pe care acestea o cere este să vadă care este **modul** predominant într-o melodie care a generat mai multe stream-uri.

*Codul folosit, prezentarea rezultatelor și interpretarea rezultatelor:*

```
##Problema 6: Modul predominant într-o melodie care a generat mai multe stream-uri
minor_streams = data['streams'][data['mode']=='Minor'].sum()
print('Numarul total de stream-uri pentru melodiile in modul predominant MINOR: ' + str(f"{minor_streams:,}"))

minor_streams = data['streams'][data['mode']=='Major'].sum()
print('Numarul total de stream-uri pentru melodiile in modul predominant MAJOR: ' + str(f"{minor_streams:,}"))
```

```
Numarul total de stream-uri pentru melodiile in modul predominant MINOR: 195,835,625,001
Numarul total de stream-uri pentru melodiile in modul predominant MAJOR: 293,866,655,154
```

Se poate remarca faptul că, în cadrul **celor mai populare melodii din anul 2023**, modul **major** a generat pe platforma Spotify aproape mai mult de **293 de miliarde** de stream-uri, adică cu aproape **100 de miliarde**, care a generat aproximativ **195 de miliarde**.

## Problema 7: Spotify Wrapped

Spotify dorește să organizeze evenimentul anual "**Spotify Wrapped**", în care le oferă utilizatorilor și următorilor o privire asupra melodiilor și artiștilor pe care i-au ascultat cel mai mult în cursul anului. Pentru a pregăti acest eveniment, echipa de dezvoltare trebuie să genereze un **raport** care să ofere statistici despre melodiile ascultate în anul curent, grupate **pe perioade de timp relevante**.

*Codul folosit, prezentarea rezultatelor și interpretarea rezultatelor:*

```

#Problema 7: pie chart cu numarul de melodii din fiecare an
import matplotlib.pyplot as plt
years = data['released_year'].value_counts().sort_index()

years.index = years.index.astype(str)
print(years)

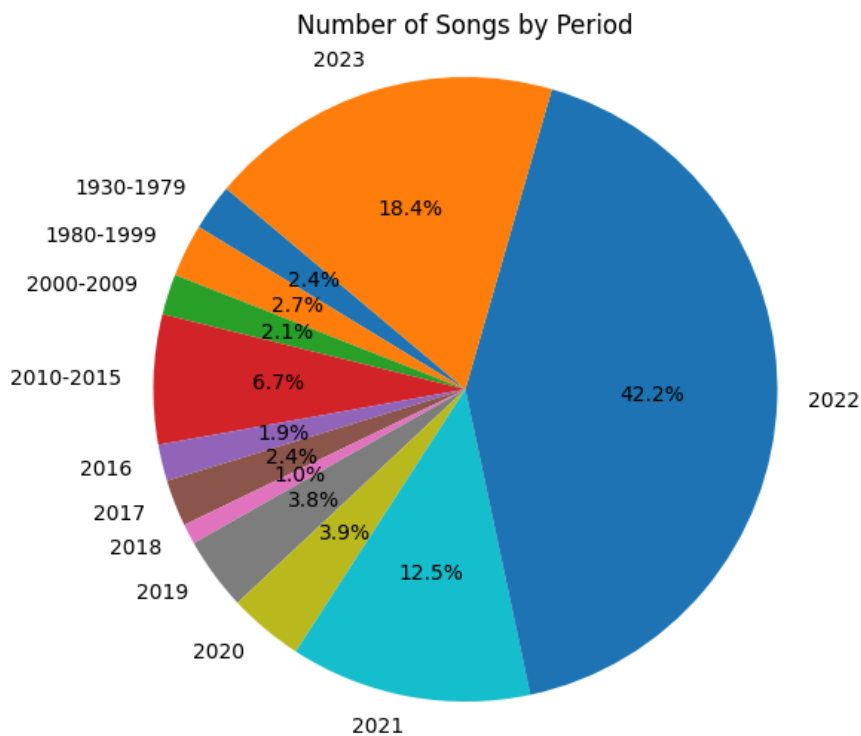
years_df = pd.DataFrame(columns=['Period', 'Number of Songs'])

years_df.loc['1930-1979'] = years.loc['1930':'1979'].sum()
years_df.loc['1980-1999'] = years.loc['1980':'1999'].sum()
years_df.loc['2000-2009'] = years.loc['2000':'2009'].sum()
years_df.loc['2010-2015'] = years.loc['2010':'2015'].sum()

for year in range(2016, 2024):
    years_df.loc[str(year)] = years.loc[str(year)]
print(years_df)

plt.figure(figsize=(10, 6))
plt.pie(years_df['Number of Songs'], labels=years_df.index, autopct='%1.1f%%',
        startangle=140)
plt.title('Number of Songs by Period')
plt.axis('equal') # Asigură că aspectul cercului este păstrat
plt.show()

```



După cum se poate observa, melodiile lansate în **2022** au fost mai ascultate decât cele din anul **2023**. **Cel mai mare procent** de melodii au fost ascultate în anul **2023-2021**, iar din anul 2020 procentele încep să scadă **semnificativ**. Deși există interes pentru piesele vechi și chiar foarte vechi, melodiile **mai noi de 2015** au fost mult mai numeroase între aceste piese populare decât celelalte.

### Problema 8: Spotify Wrapped

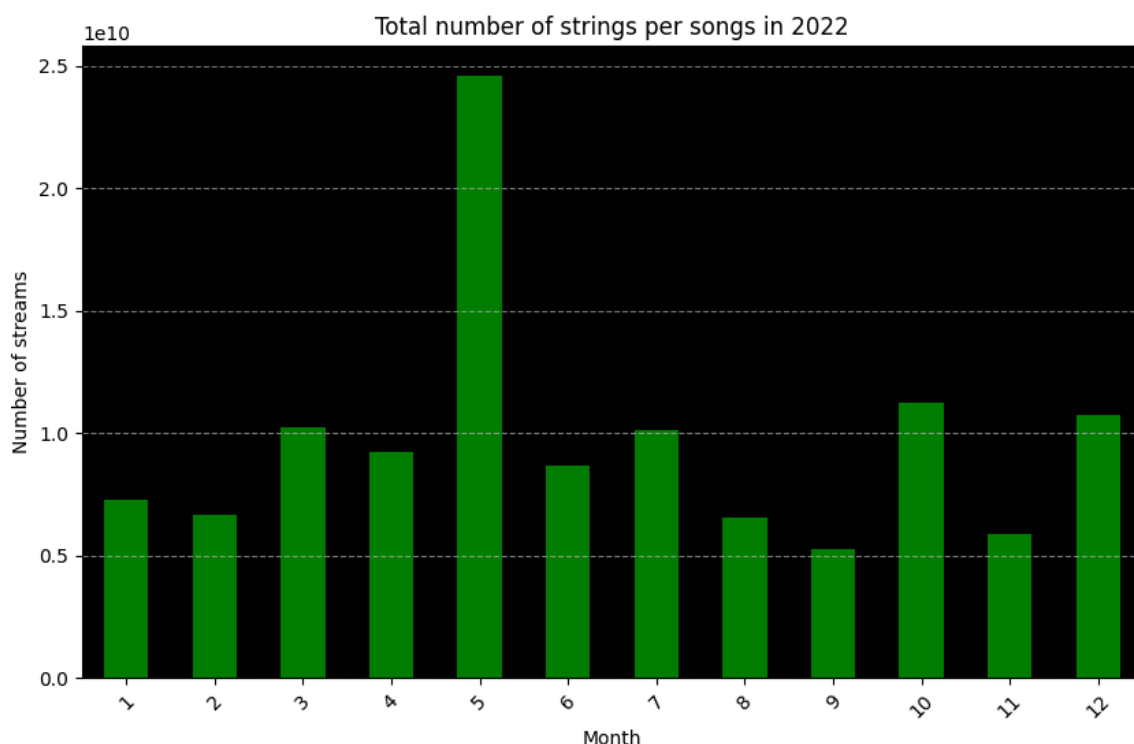
Echipa de **Analiza datelor** trebuie să dezvolte un script Python care să analizeze datele despre **melodiile ascultate în 2022** și să genereze un raport sub formă de **grafic cu bare**. Graficul trebuie să prezinte distribuția numărului total de stream-uri pe lună pentru melodiile ascultate în anul curent.

*Codul folosit, prezentarea rezultatelor și interpretarea rezultatelor:*

```
#Problema 8: Grafic cu bare despre popularitatea melodiilor in functie
songs_2022 = data[data['released_year'] == 2022]

monthly_2022 = songs_2022.groupby('released_month')['streams'].sum()

plt.figure(figsize=(10, 6))
ax = monthly_2022.plot(kind='bar', color='green')
ax.set_facecolor('black')
plt.title('Total number of strings per songs in 2022')
plt.xlabel('Month')
plt.ylabel('Number of streams')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```



Se observă faptul că luna în care o melodie a fost lansată în anul 2022 **nu pare ca a influențat semnificativ** numărul de ascultări pe **Spotify**, cu excepția lunii **mai**, unde numărul de stream-uri este **mai mult decât dublul** numărului de redări ale următoarei luni ca număr de redări. Cel mai probabil, una dintre melodiile sau albumele foarte așteptate sau care a devenit foarte populară a fost lansată atunci și nu are o legătură directă cu preferințele consumatorilor.

# Programare SAS

## Problema 1: Crearea unui set de date SAS din fișiere externe

La finalul anului 2023, Spotify a alcătuit o listă cu cele mai populare melodii incluzând atât informații despre numărul de stream-uri, apariții în playlist-uri și apariții în topuri de pe platforma lor dar și de pe platforme asemănătoare precum Apple Music și Deezer. Totodată, Spotify a adunat și informații tehnice despre fiecare melodie cum ar fi cheia fiecărei melodii, modul dar și “dansabilitatea” și acustica. Pentru a analiza datele, acestea trebuie să fie extrase din fișierul CSV într-un dataset folosind SAS Studio.

### Codul folosit:

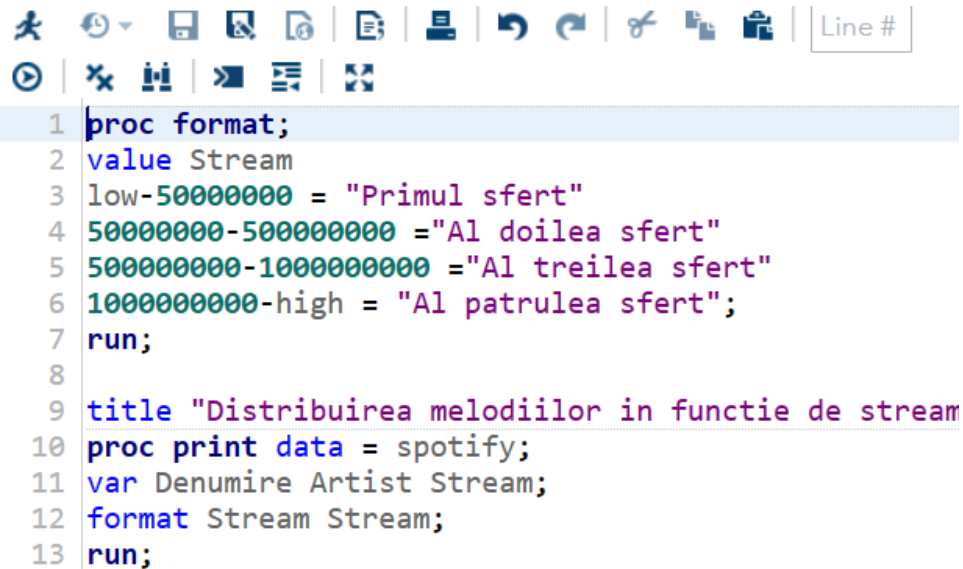
```
1 data spotify;
2 infile '/home/u63854871/Data/spotify-2023.csv' dsd dlm=', ' firstobs=2;
3 length Denumire $50;
4 length Artist $50;
5 input Denumire $ Artist $ NrArtisti AnLansare LunaLansare ZiLansare SpotifyPlaylists SpotifyTops Stream ApplePlaylists AppleTops DeezerPlaylists DeezerTops ShazamTops Bpm Che
6 run;
```

Obs	Denumire	Artist	NrArtisti	AnLansare	LunaLansare	ZiLansare	SpotifyPlaylists	SpotifyTops	Stream	ApplePlaylists	AppleTops	DeezerPlaylists	DeezerTops	ShazamTops	Bpm	Che
1	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	553	147	141381703	43	263	45	10	826	125	B
2	LALA	Myke Towers	1	2023	3	23	1474	48	133716286	48	126	58	14	382	92	C#
3	vampire	Olivia Rodrigo	1	2023	6	30	1397	113	140003974	94	207	91	14	949	138	F
4	Cruel Summer	Taylor Swift	1	2019	8	23	7858	100	800840817	116	207	125	12	548	170	A
5	WHERE SHE GOES	Bad Bunny	1	2023	5	18	3133	50	303236322	84	133	87	15	425	144	A
6	Sprinter	Dave, Central Cee	2	2023	6	1	2186	91	183706234	67	213	88	17	946	141	C#
7	Ella Baila Sola	Eslabon Armado, Peso Pluma	2	2023	3	16	3090	50	725980112	34	222	43	13	418	148	F
8	Columbia	Quevedo	1	2023	7	7	714	43	58148378	25	89	30	13	194	100	F
9	fukumean	Gunna	1	2023	5	15	1096	83	95217315	60	210	48	11	953	130	C#
10	La Bebe - Remix	Peso Pluma, Yng Lvcas	2	2023	3	17	2953	44	553634067	49	110	66	13	339	170	D
11	un x100to	Bad Bunny, Grupo Frontera	2	2023	4	17	2876	40	505671438	41	205	54	12	251	83	F#
12	Super Shy	NewJeans	1	2023	7	7	422	55	58255150	37	202	21	5	168	150	F
13	Flowers	Miley Cyrus	1	2023	1	12	12211	115	1316855716	300	215	745	58	.	118	.
14	Daylight	David Kushner	1	2023	4	14	3528	98	387570742	80	156	182	24	.	130	D
15	As It Was	Harry Styles	1	2022	3	31	23575	130	2513188493	403	198	863	46	.	174	F#
16	Kill Bill	SZA	1	2022	12	8	8109	77	1163093654	183	162	161	12	187	89	G#
17	Cupid - Twin Ver.	Fifty Fifty	1	2023	2	24	2942	77	496795686	91	212	78	6	0	120	B
18	What Was I Made For? (From The Motion Picture "Bar	Billie Eilish	1	2023	7	13	873	104	30546883	80	227	95	24	.	78	.
19	Classy 101	Feid, Young Miko	2	2023	3	31	2610	40	335222234	43	100	54	14	187	100	B
20	Like Crazy	Jimin	1	2023	3	24	596	68	363369738	8	104	23	2	29	120	G
21	LADY GAGA	Gabito Ballesteros, Junior H, Peso Pluma	3	2023	6	22	332	26	8644842	11	163	10	4	0	140	F
22	I Can See You	Taylor Swift	1	2023	7	7	516	38	52135248	73	119	42	1	150	123	F#



## Problema 2: Crearea și folosirea de formate definite de utilizator.

Deoarece cantitatea de date observată este una foarte mare, angajații Spotify doresc să reprezinte într-un tabel distribuția datelor în funcție de numărul de streams înregistrate exclusiv pe platforma lor.



```
1 proc format;
2 value Stream
3 low-50000000 = "Primul sfert"
4 50000000-500000000 = "Al doilea sfert"
5 500000000-1000000000 = "Al treilea sfert"
6 1000000000-high = "Al patrulea sfert";
7 run;
8
9 title "Distribuirea melodiilor in functie de stream";
10 proc print data = spotify;
11 var Denumire Artist Stream;
12 format Stream Stream;
13 run;
```

**Distribuirea melodiilor in functie de streams**

Obs	Denumire	Artist	Stream
1	Seven (f	Latto, J	Al doilea sfert
2	LALA	Myke Tow	Al doilea sfert
3	vampire	Olivia R	Al doilea sfert
4	Cruel Su	Taylor S	Al treilea sfert
5	WHERE SH	Bad Bunn	Al doilea sfert
6	Sprinter	Dave, Ce	Al doilea sfert
7	Ella Bai	Eslabon	Al treilea sfert
8	Columbia	Quevedo	Al doilea sfert
9	fukumean	Gunna	Al doilea sfert
10	La Bebe	Peso Plu	Al treilea sfert
11	un x100t	Bad Bunn	Al treilea sfert
12	Super Sh	NewJeans	Al doilea sfert
13	Flowers	Miley Cy	Al patrulea sfert
14	Daylight	David Ku	Al doilea sfert
15	As It Wa	Harry St	Al patrulea sfert
16	Kill Bil	SZA	Al patrulea sfert
17	Cupid -	Fifty Fi	Al doilea sfert
18	What Was	Billie E	Primul sfert
19	Classy 1	Feid, Yo	Al doilea sfert
20	Like Cra	Jimin	Al doilea sfert
21	LADY GAG	Gabito B	Al doilea sfert
22	I Can Se	Taylor S	Al doilea sfert
23	I Wanna	Arctic M	Al patrulea sfert
24	Peso Plu	Bizarrap	Al doilea sfert
25	Popular	The Week	Al doilea sfert

### Problema 3: Procesarea iterativă și condiționată a datelor

Echipa Spotify dorește să clasifice melodiile în funcție de genul lor, pentru a realiza acest lucru vor lua în considerare variabilele procentuale Densabilitate și Energie realizând o medie între acestea, melodiile cu media peste 70% vor fi clasificate ca și melodii de club, melodiile cu media între 70% și 50% vor fi clasificate în melodii intermediare iar melodiile cu media sub 50% vor fi clasificate ca fiind melodii lente.


```
1 data spotify_clasificare;  
2 set spotify;  
3 Medie_Dens_Energie = (Densabilitate+Energie)/2;  
4 if Medie_Dens_Energie >70 then Clasificare = "Melodii de club";  
5 else if Medie_Dens_Energie <=70 and Medie_Dens_Energie > 50 then Clasificare = "Melodii intermediare";  
6 else Clasificare = "Melodii lente";  
7 run;  
8  
9 proc print data = spotify_clasificare(obs = 20);  
10 var Denumire Artist AnLansare Medie_Dens_Energie Clasificare;  
11 title "Primele 20 de melodii cu clasificarea";  
12 run;
```


Primele 20 de melodii cu clasificarea

Obs	Denumire	Artist	AnLansare	Medie_Dens_Energie	Clasificare
1	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2023	81.5	Melodii de club
2	LALA	Myke Towers	2023	72.5	Melodii de club
3	vampire	Olivia Rodrigo	2023	52.0	Melodii interme
4	Cruel Summer	Taylor Swift	2019	63.5	Melodii interme
5	WHERE SHE GOES	Bad Bunny	2023	72.5	Melodii de club
6	Sprinter	Dave, Central Cee	2023	75.0	Melodii de club
7	Ella Baila Sola	Eslabon Armado, Peso Pluma	2023	71.5	Melodii de club
8	Columbia	Quevedo	2023	69.0	Melodii interme
9	fukumean	Gunna	2023	73.5	Melodii de club
10	La Bebe - Remix	Peso Pluma, Yng Lvcas	2023	64.5	Melodii interme
11	un x100to	Bad Bunny, Grupo Frontera	2023	64.5	Melodii interme
12	Super Shy	NewJeans	2023	80.0	Melodii de club
13	Flowers	Miley Cyrus	2023	69.5	Melodii interme
14	Daylight	David Kushner	2023	47.0	Melodii lente
15	As It Was	Harry Styles	2022	62.5	Melodii interme
16	Kill Bill	SZA	2022	68.5	Melodii interme
17	Cupid - Twin Ver.	Fifty Fifty	2023	68.5	Melodii interme
18	What Was I Made For? [From The Motion Picture "Bar	Billie Eilish	2023	26.5	Melodii lente
19	Classy 101	Feid, Young Miko	2023	76.0	Melodii de club
20	Like Crazy	Jimin	2023	68.0	Melodii interme

### Problema 4: Utilizarea de funcții SAS

În scopul creării unui top al tuturor artiștilor de pe platforma Spotify se cere alcătuirea unui tabel care să afișeze numărul total de streams ale fiecărui artist, adunând numărul de streams de la toate melodiile lor care apar în top.



Line # 

```

1 proc sql;
2 create table TotalStreamsArtist as
3 select Artist,
4 sum(Stream) as TotalStreams
5 from spotify
6 group by Artist
7 order by TotalStreams desc;
8 quit;
9
10 proc print data=TotalStreamsArtist;
11 title "Streams totale pentru fiecare artist Artist";
12 run;


```

**Streams totale pentru fiecare artist Artist**

Obs	Artist	TotalStreams
1	The Weeknd	14185552870
2	Taylor Swift	14023315094
3	Ed Sheeran	13908947204
4	Harry Styles	11608645649
5	Bad Bunny	8733488771
6	Olivia Rodrigo	7442148916
7	Eminem	6183805596
8	Bruno Mars	5846920599
9	Arctic Monkeys	5569806731
10	Imagine Dragons	5272484650
11	Lewis Capaldi	4734698360
12	Doja Cat	4702294655
13	SZA	4557811204
14	Adele	4508746590
15	BTS	4389891591
16	The Neighbourhood	4010009939
17	Coldplay	3825176058
18	Avicii	3426754746
19	Justin Bieber	3290640459
20	Dua Lipa	3227639000

Problema 5: Crearea de subseturi de date

Datorită succesului răsunător cunoscut de artista Taylor Swift pe parcursul anului 2023, angajații Spotify își doresc să analizeze doar melodiile acesteia și datele referitoare la ele înregistrate în cadrul platformei lor.



```
1 Title "Piese Taylor Swift";
2 proc print data=spotify;
3 where Artist contains "Taylor";
4 var Denumire Artist Stream SpotifyTops SpotifyPlaylists;
5 run;
```

Piese Taylor Swift					
Obs	Denumire	Artist	Stream	SpotifyTops	SpotifyPlaylists
4	Cruel Summer	Taylor Swift	800840817	100	7858
22	I Can See You (Taylor's Version) (From The	Taylor Swift	52135248	38	516
34	Anti-Hero	Taylor Swift	999748277	56	9082
38	Blank Space	Taylor Swift	1355959075	53	11434
39	Style	Taylor Swift	786181836	42	7830
60	cardigan	Taylor Swift	812019557	29	7923
67	Karma	Taylor Swift	404562836	23	3818
70	Enchanted (Taylor's Version)	Taylor Swift	39578178	24	148
84	Back To December (Taylor's Version)	Taylor Swift	39228929	17	139
94	Don't Blame Me	Taylor Swift	685032533	23	4875
101	Lover	Taylor Swift	882831184	23	8448
114	Mine (Taylor's Version)	Taylor Swift	36912123	15	99
120	august	Taylor Swift	607123776	22	7324
164	Enchanted	Taylor Swift	621660989	16	4564
177	Shake It Off	Taylor Swift	1113838873	13	21335
192	You Belong With Me (Taylor's Version)	Taylor Swift	350381515	12	2619
206	Hits Different	Taylor Swift	68616963	0	547
218	Karma (feat. Ice Spice)	Taylor Swift, Ice Spice	46142772	0	588
223	Snow On The Beach (feat. More Lana Del Rey)	Lana Del Rey, Taylor Swift	60350538	2	359
236	Lavender Haze	Taylor Swift	488386797	8	3763

## Problema 6: Combinarea seturilor de date prin proceduri specifice SAS și SQL

Din păcate, datorită unei defecțiuni tehnice, toate melodiile lansate în anul 2023, luna a 5-a au înregistrat cu 20% mai puține redări, pentru a realiza o bază de date corespunzătoare și fără greșeli se dorește corectarea acestei erori.

```
1 proc print data = spotify(where=(AnLansare =2023 and LunaLansare =5));
2 var Denumire Artist Stream AnLansare LunaLansare;
3 format Denumire $50. Artist $50.;
4 run;
5
6 proc sql;
7 update spotify set Stream = Stream*1.2 where AnLansare =2023 and LunaLansare =5;
8 quit;
```

Obs	Denumire	Artist	Stream	AnLansare	LunaLansare
5	WHERE SHE GOES	Bad Bunny	366915949.62	2023	5
9	fukumean	Gunna	115212951.15	2023	5
24	Peso Pluma: Bzrp Music Sessions, Vol. 55	Bizarrap, Peso Pluma	242783137.41	2023	5
30	Dance The Night (From Barbie The Album)	Dua Lipa	154164834.34	2023	5
40	TQM	Fuerza Regida	213629705.96	2023	5
61					

Obs	Denumire	Artist	Stream	AnLansare	LunaLansare
5	WHERE SHE GOES	Bad Bunny	440299139.54	2023	5
9	fukumean	Gunna	138255541.38	2023	5
24	Peso Pluma: Bzrp Music Sessions, Vol. 55	Bizarrap, Peso Pluma	291339764.89	2023	5
30	Dance The Night (From Barbie The Album)	Dua Lipa	184997801.21	2023	5
40	TQM	Fuerza Regida	256355647.15	2023	5

## Problema 7: Folosirea de proceduri statistice

Pentru a identifica caracteristicile care creează cele mai bune melodii din punct de vedere a caracteristicilor analizate, echipa Spotify dorește să creeze tabele statistice care să identifice ponderea cea mai mare din fiecare caracteristică pusă la dispoziție de setul de date din anul 2023.

```
1 proc freq data=spotify order=freq;
2 tables Denumire Valenta Energie Acusticitate Instrumentalitate Liveness Speechiness;
3 run;
```

Artist	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Taylor Swift	33	3.51	33	3.51
The Weeknd	22	2.34	55	5.86
SZA	19	2.02	74	7.88
Bad Bunny	18	1.92	92	9.80
Harry Styles	17	1.81	109	11.61
Kendrick Lamar	12	1.28	121	12.89
Morgan Wallen	11	1.17	132	14.06

Valenta	Frequency	Percent	Cumulative Frequency	Cumulative Percent
24	20	2.13	20	2.13
40	20	2.13	40	4.26
55	18	1.92	58	6.18
61	18	1.92	76	8.09
53	17	1.81	93	9.90
59	17	1.81	110	11.71
22	16	1.70	126	13.42
49	16	1.70	142	15.12
42	15	1.60	157	16.72
44	15	1.60	172	18.32

Energie	Frequency	Percent	Cumulative Frequency	Cumulative Percent
74	27	2.88	27	2.88
76	26	2.77	53	5.64
62	25	2.66	78	8.31
66	25	2.66	103	10.97
60	23	2.45	126	13.42
68	23	2.45	149	15.87
67	22	2.34	171	18.21
70	22	2.34	193	20.55
73	22	2.34	215	22.90
79	21	2.24	236	25.13

Liveness	Frequency	Percent	Cumulative Frequency	Cumulative Percent
11	102	10.86	102	10.86
9	90	9.58	192	20.45
10	78	8.31	270	28.75
12	69	7.35	339	36.10
8	55	5.86	394	41.96
13	45	4.79	439	46.75
15	36	3.83	475	50.59
7	35	3.73	510	54.31

Speechiness	Frequency	Percent	Cumulative Frequency	Cumulative Percent
4	169	18.27	169	18.27
3	150	16.22	319	34.49
5	127	13.73	446	48.22
6	75	8.11	521	56.32
7	47	5.08	568	61.41
8	47	5.08	615	66.49

În urma analizării datelor statistice rezultate din crearea unui tabel de frecvență pentru a identifica caracteristicile cel mai des întâlnite în piesele de top ale anului 2023 putem concluziona că o așa numită “melodie perfectă” ar avea următoarele caracteristici:

- Dansabilitate: 4%
- Valență: 24%
- Energie: 74%
- Liveness: 11%
- Speechiness: 4%

Aceasta fiind preferabil cântată de Taylor Swift.

### Problema 8: Generare grafice

Se dorește o vizuale mai amănunțită a distribuției pieselor în funcție de cheia în care sunt cântate, în scopul identificării preferințelor muzicale ale ascultătorilor anului 2023. Astfel, angajații Spotify doresc un generarea unui grafic pentru a ilustra aceste preferințe.



```
1 proc sgplot data = spotify;  
2 vbar Cheie/ datalabel datalabelattrs =(color=black)  
3 barwidth= 0.5;  
4 title "Distributia pieselor in functie de cheia muzicala";  
5 xaxis display =(nolabel);  
6 yaxis grid;  
7 run;
```

