

PREDICTIA INCASARILOR PE BAZA ACTORILOR CARE JOACA

IN ACESTA

PROIECT IAPD

Costescu Matei Cristian

Mandru Luca Teodor

INTRODUCERE

1

OVERVIEW

3

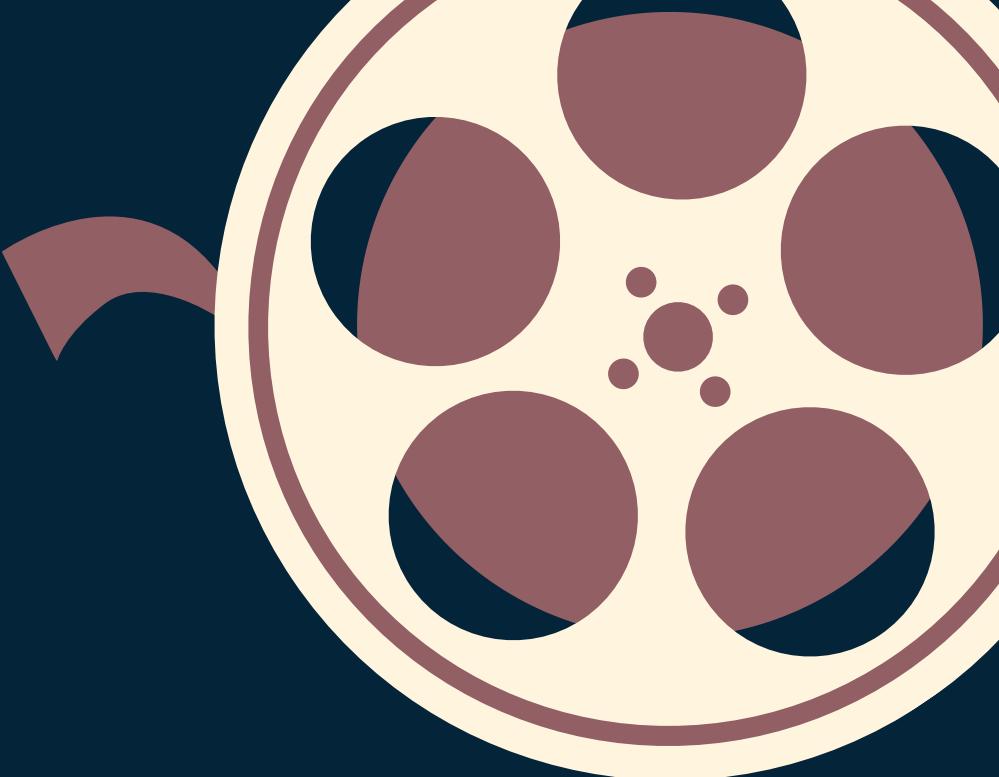
PREPROCESAREA DATELOR

2

EXTRAGEREA DATELOR

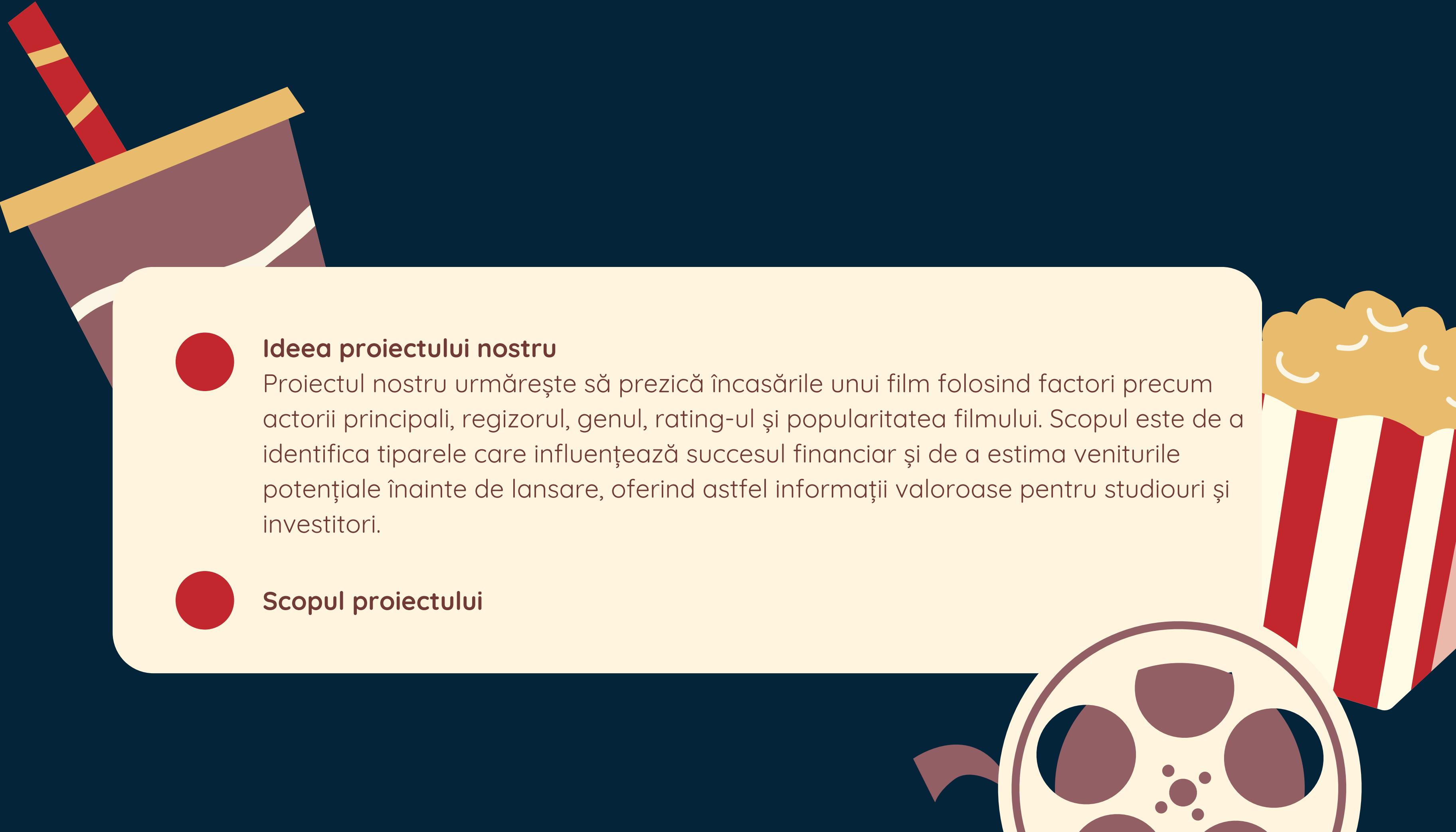
4

ANALIZAREA DATELOR



OVERVIEW





Ideea proiectului nostru

Proiectul nostru urmărește să prezică încasările unui film folosind factori precum actorii principali, regizorul, genul, rating-ul și popularitatea filmului. Scopul este de a identifica tiparele care influențează succesul financiar și de a estima veniturile potențiale înainte de lansare, oferind astfel informații valoroase pentru studiouri și investitori.

Scopul proiectului

EXTRAGEREA DATELOR



PASII URMATI:

1

Am generat o listă de link-uri către filme de pe IMDb folosind un script Python.

2

Am creat un webscraper pentru paginile IMDb ale filmelor, care extrage următoarele informații: titlu, rating, durată, regizori, scenariști, actori principali, genuri, popularitate (un fel de clasament IMDb), buget estimat, câștiguri și URL-ul filmului.

3

Am rulat webscraper-ul pe lista noastră de link-uri și am obținut un set de date cu 9.750 de înregistrări.



PREPROCESAREA DATELOR



PASII URMATI:

1

Având un volum mare de date, am testat două abordări:

1. În prima variantă am reținut doar filmele care aveau toate coloanele complete, rezultând un set de date de aproximativ 2.580 de înregistrări.
2. În a doua variantă, am completat datele lipsă.

2

Pentru coloanele cu valori lipsă, am înlocuit spațiile goale cu valoarea cea mai frecventă din coloana respectivă.

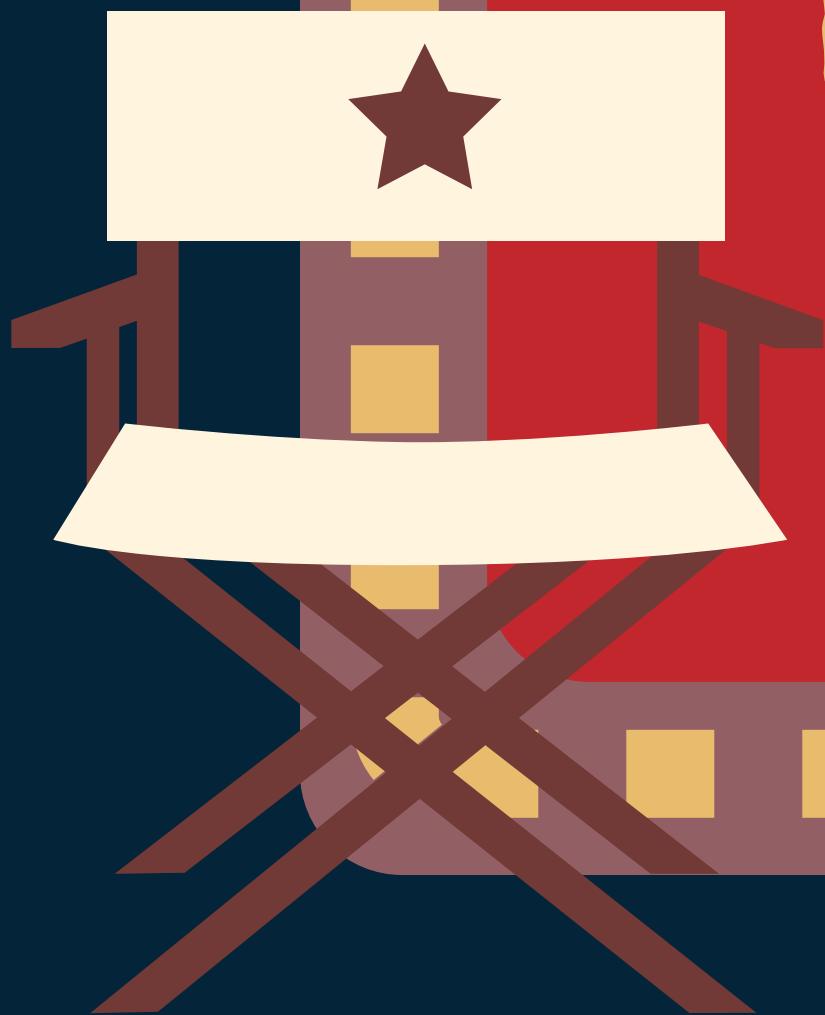
3

Pentru coloanele Budget și Gross Worldwide, am extras moneda asociată fiecărei sume și am convertit toate valorile în dolari, pentru a uniformiza formatul. La final, aceste coloane au fost salvate ca valori numerice.

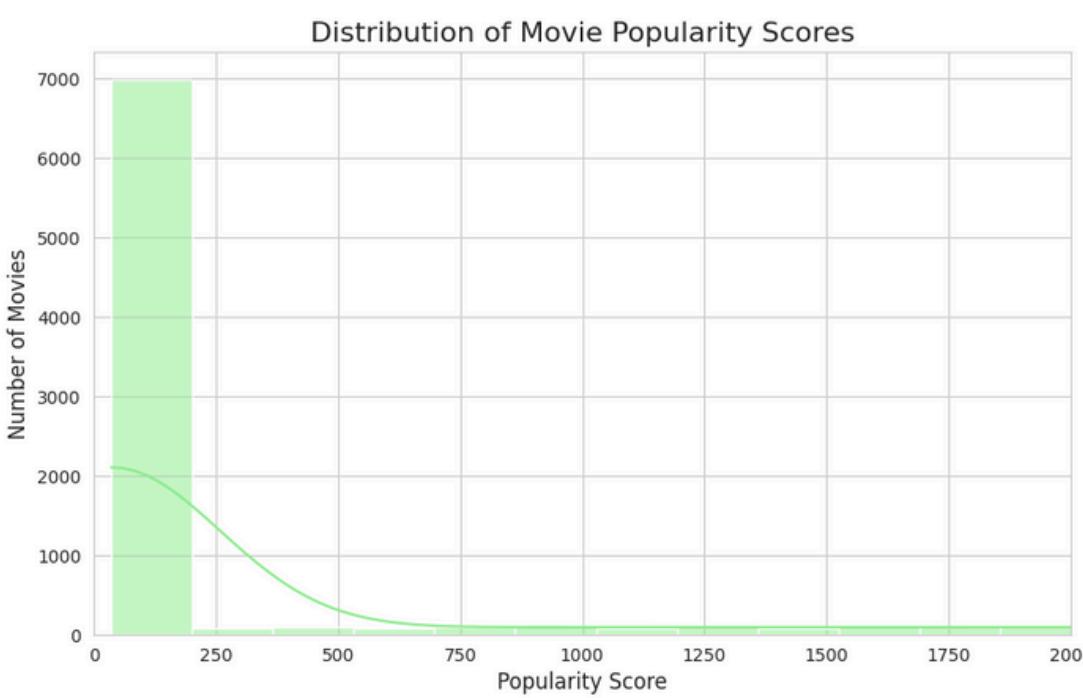
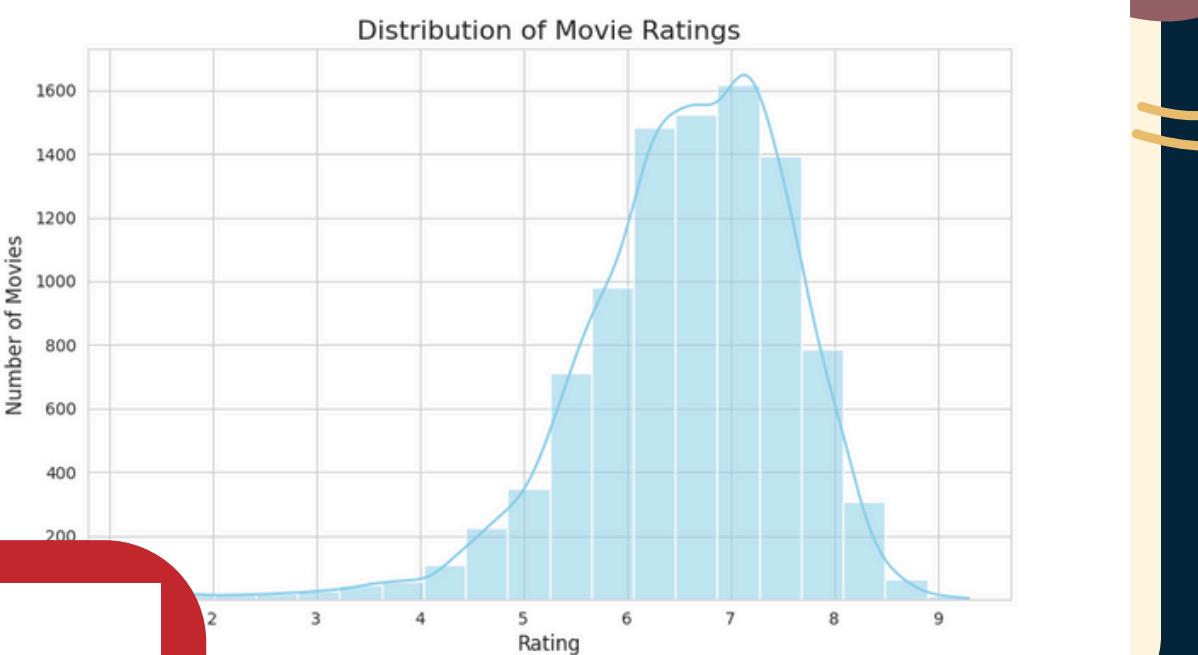
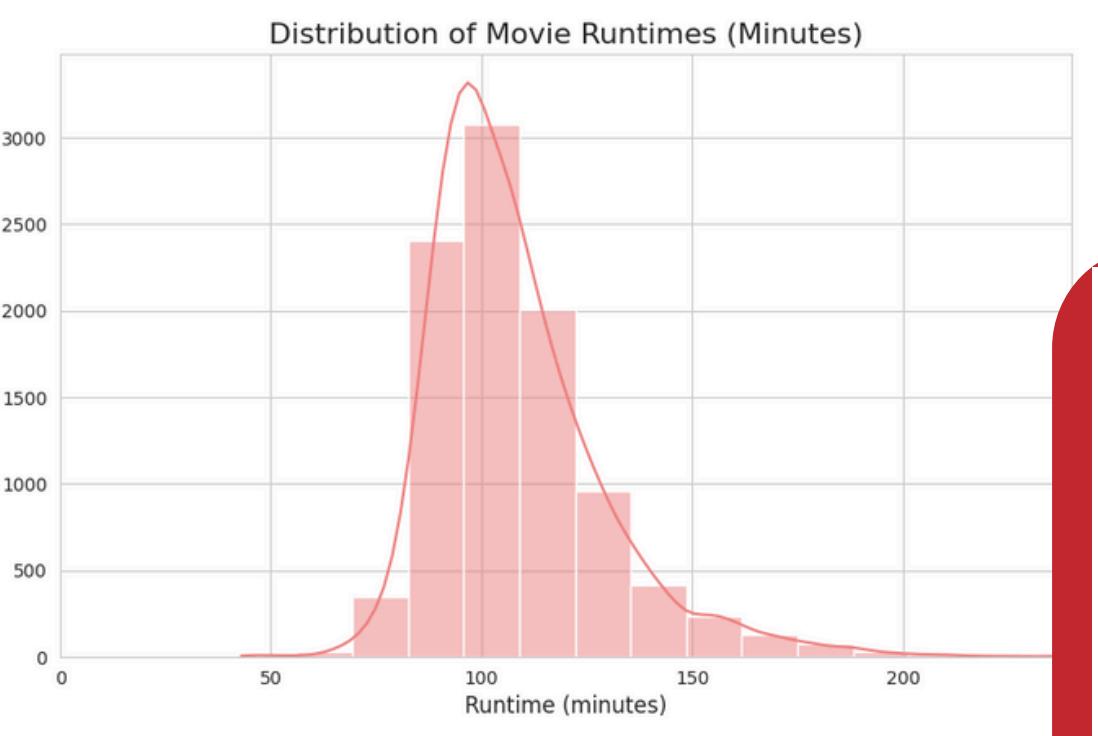
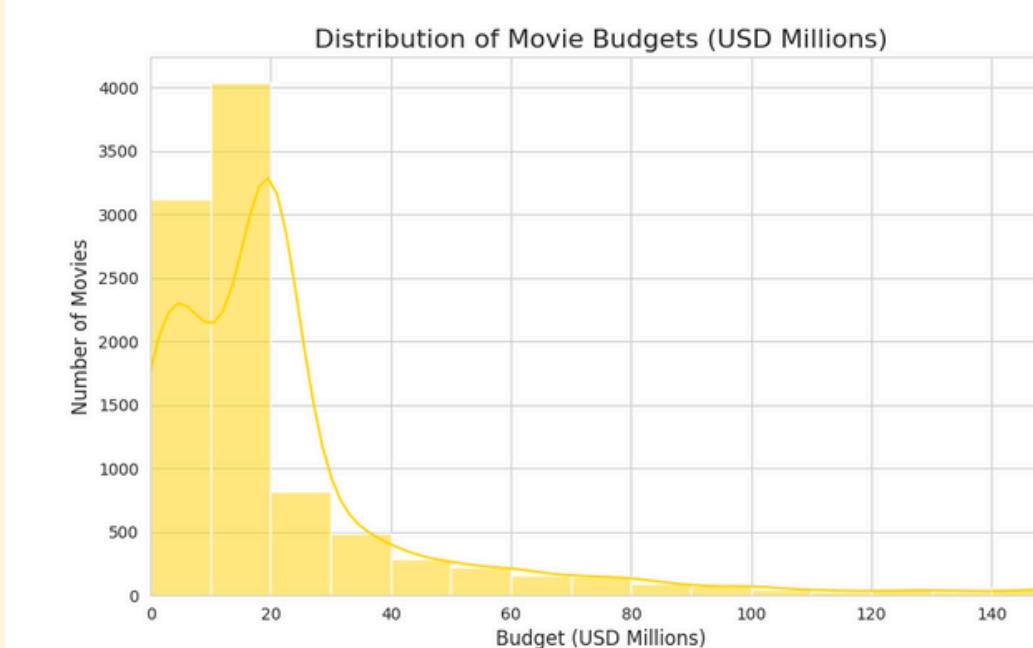
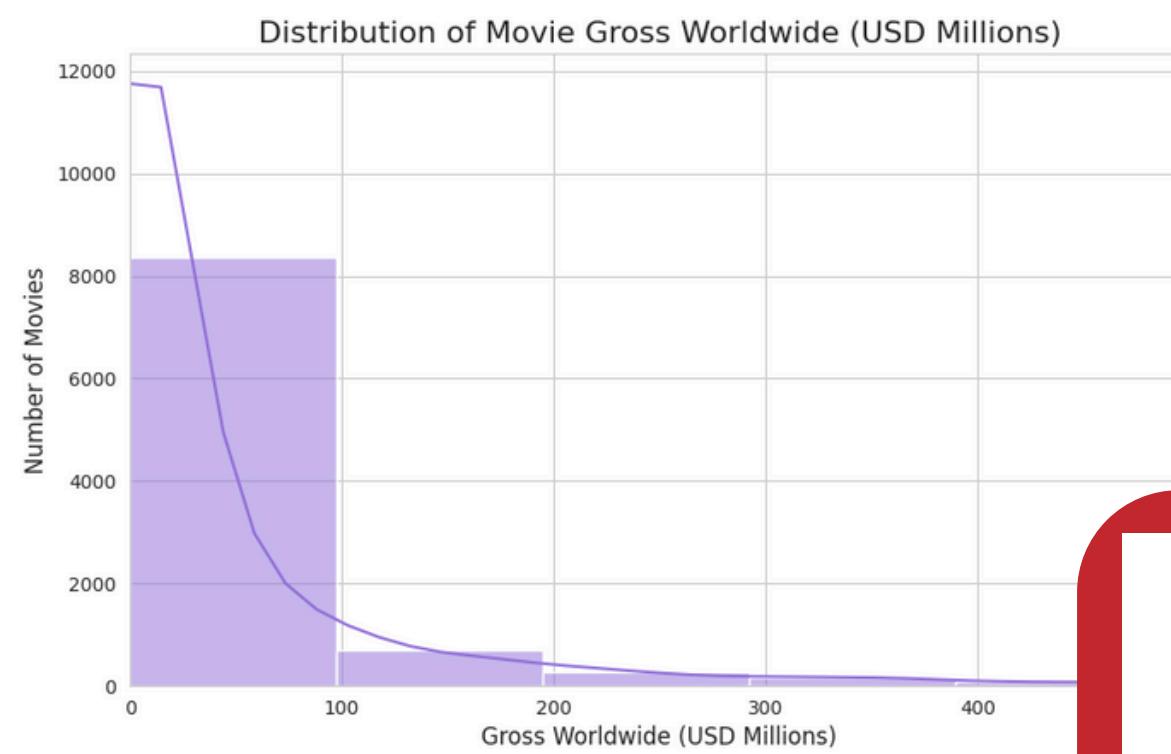




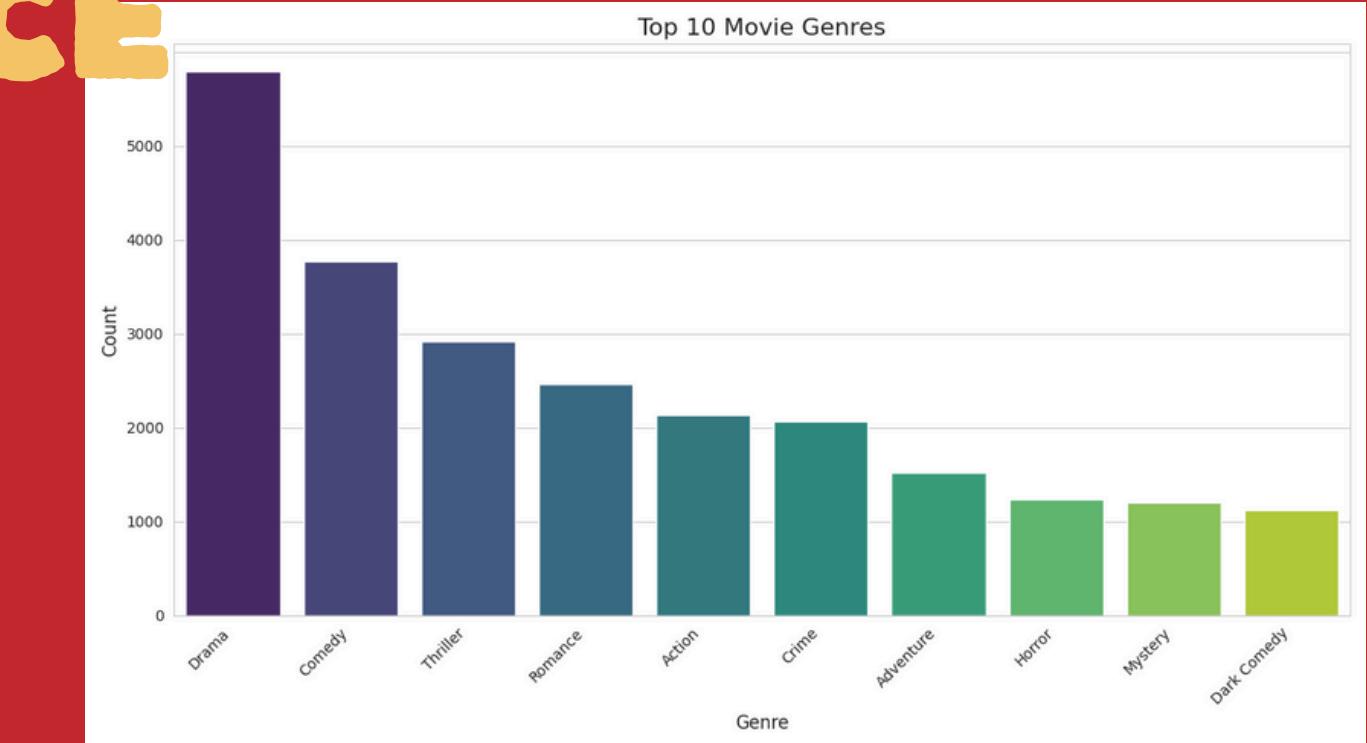
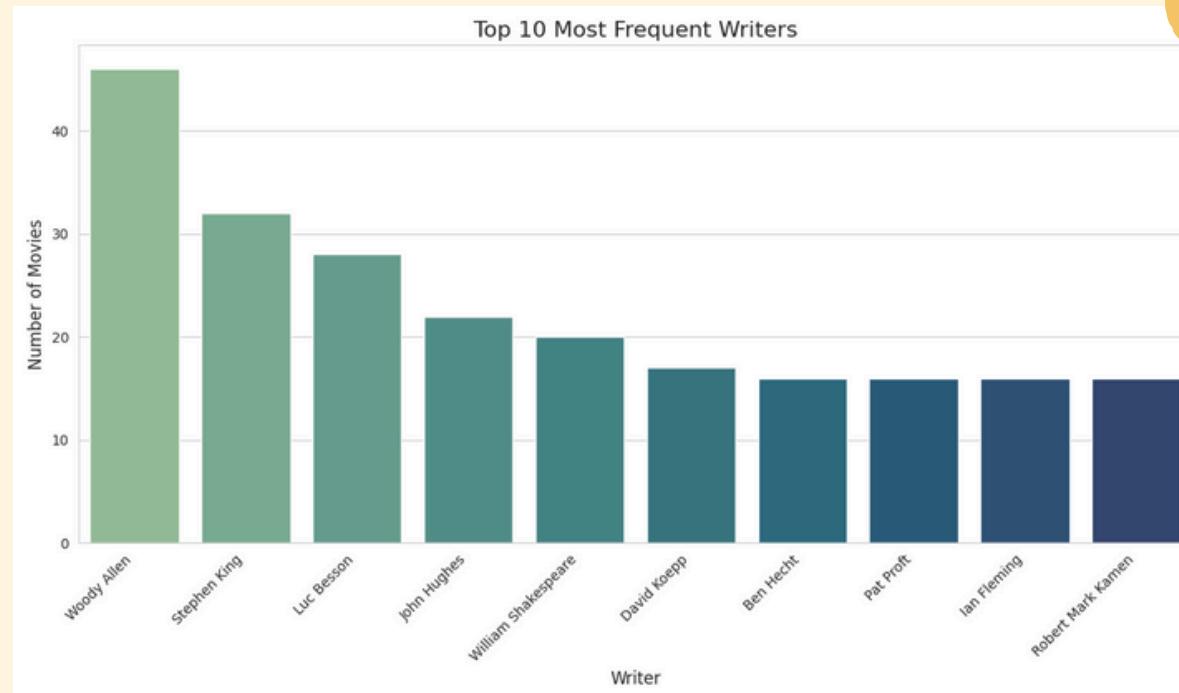
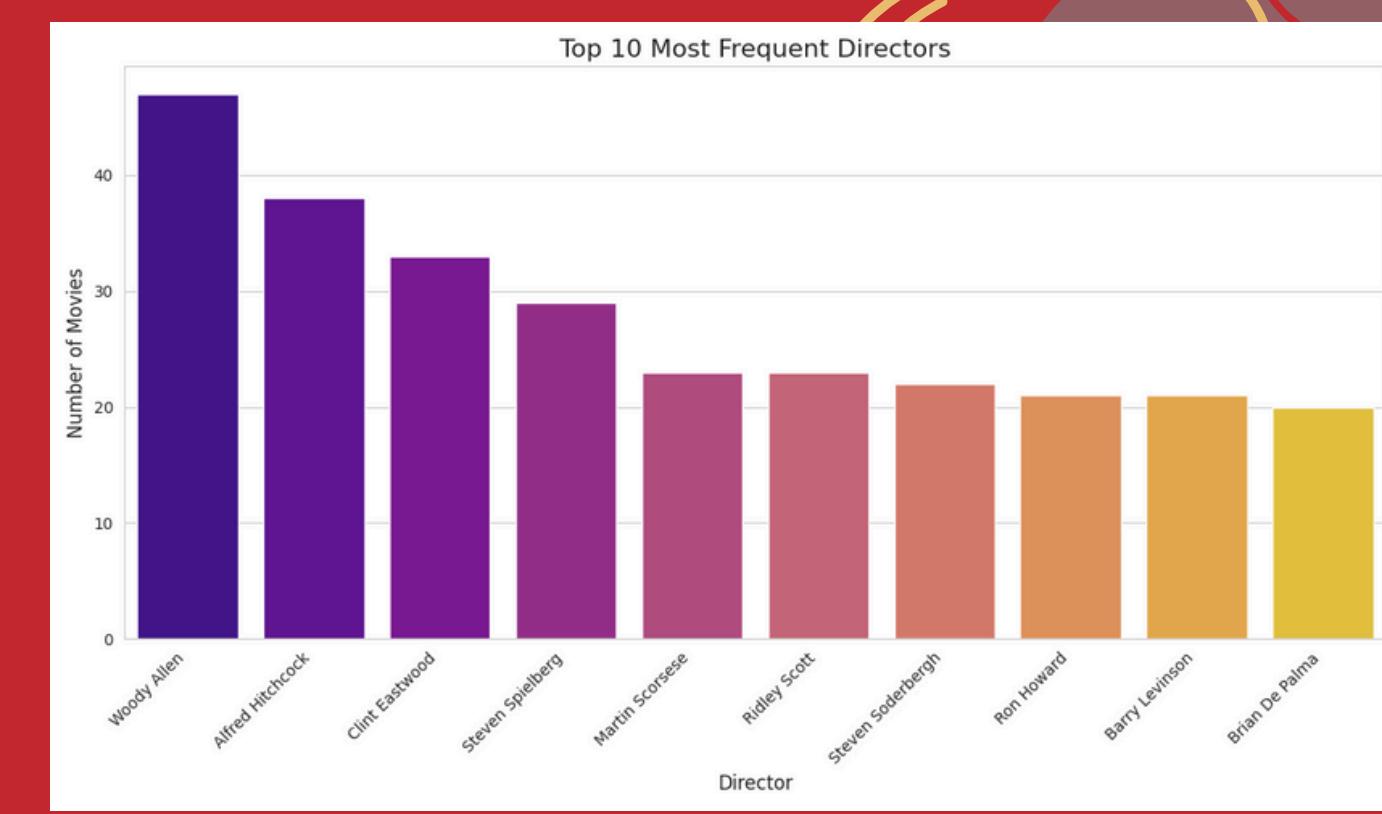
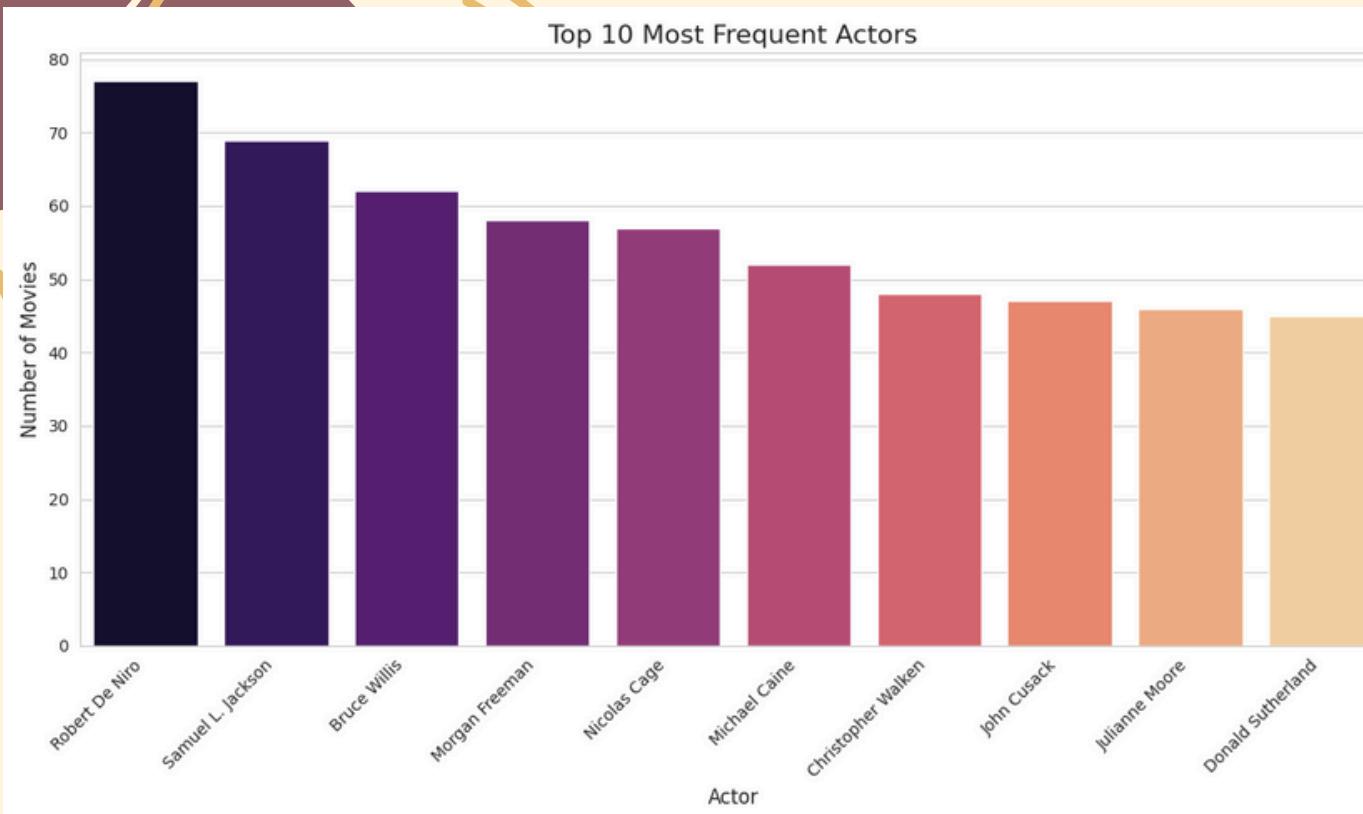
**ANALIZA
DATELOR**



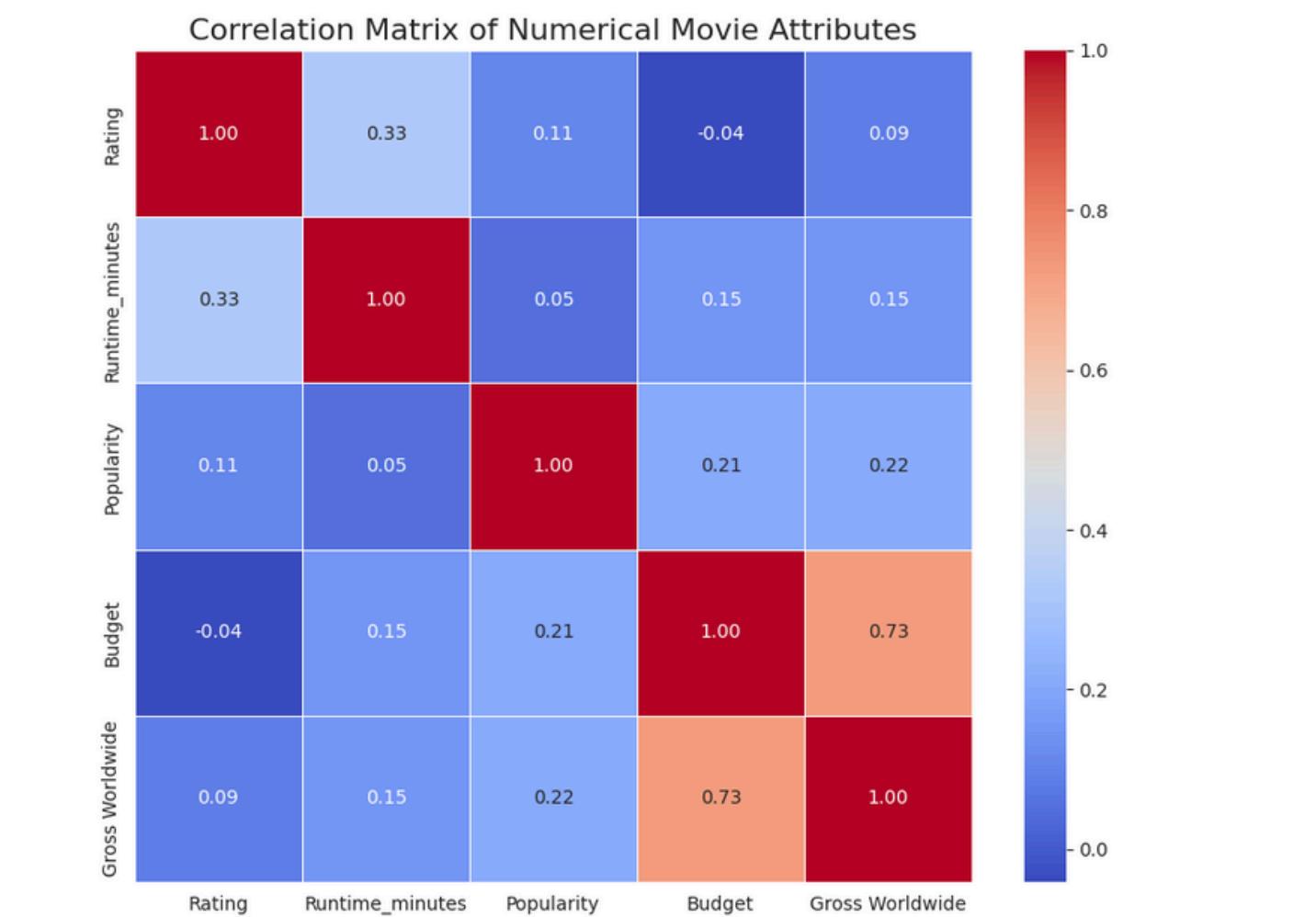
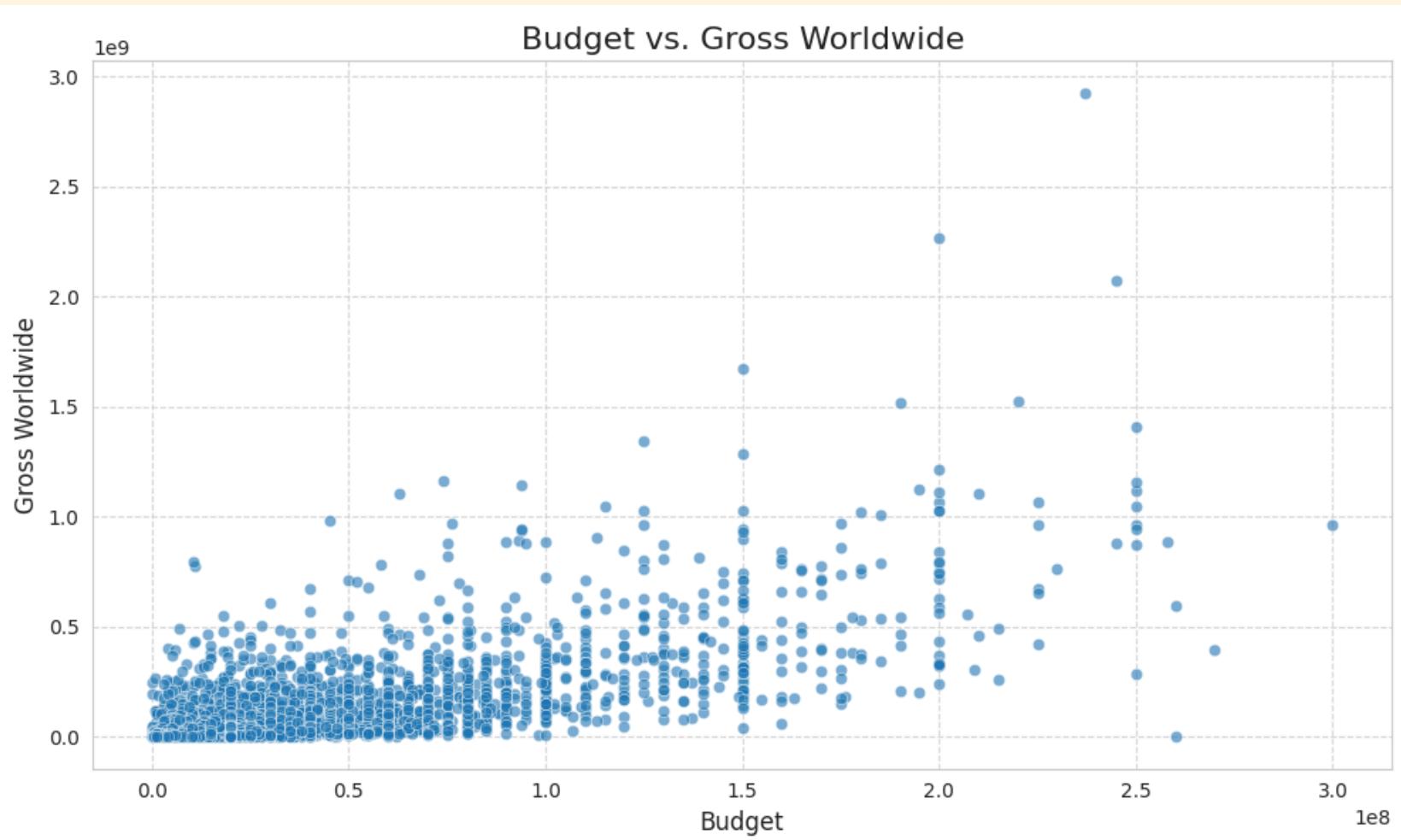
VALORI CONTINUE



VALORI CATEGORICE



CORELAREA ATRIBUTELOR





MULTUMIM
pentru
ATENTIE