ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ ΣΤΗ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

Ακαδημαϊκό έτος 2020-2021

MOVIE RECOMMENDATION SYSTEM

Ρουμπίνη Σταθοπούλου ΑΜ 1059667

Ελένη Χάιδου ΑΜ 1059627

Ειρήνη Ελένη Τσιαούση ΑΜ 1054273

# Προετοιμασία δεδομένων

Αρχικά μας δίνονται τα παρακάτω δεδομένα σε csv αρχεία:

**movies.csv:** To αρχείο περιέχει στοιχεία για κάθε ταινία.

Οι στήλες που περιέχει είναι:

movie\_id Κωδικός ταινίας

title Τίτλος ταινίας

genres Κατηγορία που ανήκει

**users.csv:** Αφορά τα στοιχεία του χρήστη.

Οι στήλες που περιέχει είναι:

user\_id Κωδικός Χρήστη

gender Φύλο

age Ηλικία

occupation Κωδικός επαγγέλματος χρήστη

zipcode Ταχυδρομικός κώδικας χρήστη

age\_desc Κατηγορία ηλικίας

occ\_desc κατηγορία επαγγέλματος

**rating.csv:** Το αρχείο περιέχει τις βαθμολογίες του κάθε χρήστη για κάθε ταινία.

Οι στήλες που περιέχει είναι:

user\_id Κωδικός του χρήστη

movie\_id Κωδικός ταινίας

rating Βαθμολογία

timestamp Χρονική στιγμή βαθμολογίας

# Περιγραφή μεθόδου

Δημιουργούμε το μέτρο της ομοιότητας των χρηστών όπου παίρνει υπόψη του τα χαρακτηριστικά του φύλου, της ηλικίας , του επαγγέλματος καθώς και της κοινή προτίμησης, δηλαδή πόσες ίδιες ταινίες έχουν δει και έχουν παρόμοια βαθμολογία.

Έτσι για ένα χρήστη παίρνουμε την ομοιότητα με κάθε άλλο χρήστη, στην συνέχεια ταξινομούμε με φθίνουσα σειρά και τελικά εμφανίζουμε τις ταινίες που έχει ο καλύτερος δυνατός χρήστης με βαθμολογία μεγαλύτερη ή ίση του 4.

Ο κώδικας είναι ο εξής:

import pandas as pd

import matplotlib.pyplot as plt

import math

# In[ ]:

#eisagogi stoixeion

#o pinakas twn xristwn

U=pd.read\_csv("users.csv","\t")

#o pinakas twn tainion

M=pd.read\_csv("movies.csv","\t", encoding="ISO-8859-1")

#o pinakas twn vathmologiwn

R=pd.read\_csv("ratings.csv",";", encoding="ISO-8859-1")

# In[ ]:

#synartisi pou vriskei tin omoiotita metaxi dio xristwn

def common(id1,id2):

d=0

#pairnoume tin lista twn xristwn

LU=list(U["user\_id"])

try:

#elegxoume an yparxoun oi dyo xristes

idx1=LU.index(id1)

idx2=LU.index(id2)

#koitame ta koina xaraktiristika diladi ilikia,fulo,epaggelma

if(U["age\_desc"][idx1]==U["age\_desc"][idx2]):

d=d+1

if(U["gender"][idx1]==U["gender"][idx2]):

d=d+1

if(U["occ\_desc"][idx1]==U["occ\_desc"][idx2]):

d=d+1

#an exoume koina metrame poses koines tainies exoun

if(d>1):

#pairnoume tis tainies tou prwtou xristi

R1=R[R["user\_id"]==id1]

#pairnoume tis tainies tou defterou xristi

R2=R[R["user\_id"]==id2]

#vriskoume tis koines

R3=R1[R1["movie\_id"].isin(list(R2["movie\_id"]))]

#an oi tainies exoun mia sxetika koini vatmologia (idia protimisi)

for m in range(R3["movie\_id"]):

rr1=R1[R1["movie\_id"]==m]

rr2=R2[R2["movie\_id"]==m]

if(abs(rr1["rating"].iloc[0]-rr2["rating"].iloc[0])<2):

#tote afksanoume kata mia posotita to d

d=d+0.5

except:

d=0

#epistrefoume tin omoiotita

return d

# In[ ]:

#main

while(True):

#dhmiourgoume to menou

print("1. Recommend a movie for a user")

print("2. Show Plots of Data")

print("3. Exit")

print("----------------------------------------")

choice=input("Give choice:")

#an dosei 1

if(choice=="1"):

try:

#zitame gia poion xristi theloume na paroume apotelesmata

idu=int(input("give user id:"))

#pairnoume tin lista ton xristwn kai psaxnoume

UL=list(U["user\_id"])

#to xristi mas sti lista

inx=UL.index(idu)

#pairnoume ti lista me to dianisma me tis kalyteres tainies

L={}

for i in UL:

if(i!=idu):

L[i]=common(idu,i)

L2=sorted(L.items(), key=lambda x: x[1], reverse=True)

x=L2[0][0]

R1=R[R["user\_id"]==idu]

R2=R[R["user\_id"]==x]

try:

R3=R2[~R2["movie\_id"].isin(list(R1["movie\_id"]))]

for m in list(R3["movie\_id"]):

r=R3[R3["movie\_id"]==m]

if(r["rating"].iloc[0]>=4):

mm=M[M["movie\_id"]==m]

print(mm["title"].iloc[0])

except:

print("error")

except:

print("user not found")

if(choice=="2"):

#provoli stoixeiwn plithismou

plt.hist(U["gender"])

plt.show()

plt.hist(U["age"])

plt.show()

C={}

Mv=[]

for m in range(len(M)):

mm=M["genres"][m]

kat=mm.split("|")

x=[]

x.append(M["movie\_id"][m])

x.append(kat)

Mv.append(x)

for cc in kat:

try:

C[cc]=C[cc]+1

except:

C[cc]=1

plt.bar(range(len(C)), list(C.values()), align='center')

plt.xticks(range(len(C)), list(C.keys()))

if(choice=="3"):

break

# Αποτελέσματα

Τελικά παίρνουμε τα παρακάτω αποτελέσματα:

Αρχικά εμφανίζεται το παρακάτω μενού:

1. Recommend a movie for a user

2. Show Plots of Data

3. Exit

----------------------------------------

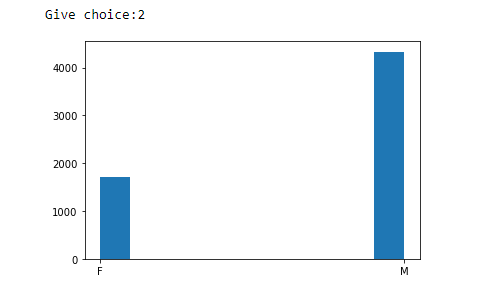
Give choice:

Επιλέγοντας 1 έχουμε:

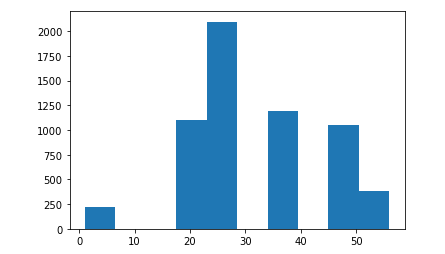
Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Με το 2 παίρνουμε κάποια αποτελέσματα για το δείγμα μας όπως πόσους χρήστες έχουμε και το φύλο τους αντίστοιχα.



Έπειτα παίρνουμε τις ηλικίες των χρηστών.



Και καταλήγουμε με το ποια είναι η κατανομή που έχουμε ανά κατηγορία ταινίας.

