Content based recommender systems

# Importing data

import numpy as np  
import pandas as pd  
import seaborn as sns  
  
columns\_names = ["user\_id","item\_id","rating","timestamp"]  
df = pd.read\_csv("E:/Py-DS-ML-Bootcamp-master/Refactored\_Py\_DS\_ML\_Bootcamp-master/19-Recommender-Systems/u.data",sep="\t",names=columns\_names)  
print(df.head(10))  
print()  
  
movie\_titles = pd.read\_csv("E:/Py-DS-ML-Bootcamp-master/Refactored\_Py\_DS\_ML\_Bootcamp-master/19-Recommender-Systems/Movie\_Id\_Titles")  
print(movie\_titles.head())  
print()  
  
df = pd.merge(df,movie\_titles,on="item\_id")  
print(df.head())  
print()

# Visualising some highest rated movies

We created a new data frame called ratings where we have movies, their rating and number of ratings for that particular movie.

import matplotlib.pyplot as plt  
import seaborn as sns  
sns.set\_style("white")  
  
print(df.groupby("title")["rating"].mean().sort\_values(ascending=False).head(10))  
print()  
  
print(df.groupby("title")["rating"].count().sort\_values(ascending=False).head(10))  
print()  
  
ratings = pd.DataFrame(df.groupby("title")["rating"].mean())  
ratings["num of ratings"] = pd.DataFrame(df.groupby("title")["rating"].count())  
print(ratings.head())  
  
plt.figure(num=1)  
ratings["num of ratings"].hist(bins=70)  
plt.figure(num=2)  
ratings["rating"].hist(bins=70)  
sns.jointplot(x="rating",y="num of ratings",data=ratings,alpha=0.5)

# Using pandas inbuilt corrwith() method to find recommendations

moviemat = df.pivot\_table(index="user\_id",columns="title",values="rating")  
print(moviemat.head())  
print()  
  
starwars\_user\_rating = moviemat["Star Wars (1977)"]  
print(starwars\_user\_rating)  
print()  
liarliar\_user\_rating = moviemat["Liar Liar (1997)"]  
  
similar\_to\_starwars = moviemat.corrwith(starwars\_user\_rating)  
similaar\_to\_liarliar = moviemat.corrwith(liarliar\_user\_rating)  
  
corr\_starwars = pd.DataFrame(similar\_to\_starwars,columns=["Correlation"])  
corr\_starwars.dropna(inplace=True)  
corr\_starwars = corr\_starwars.join(ratings["num of ratings"])  
print(corr\_starwars.head())  
print()  
  
print(corr\_starwars[corr\_starwars["num of ratings"]>100].sort\_values("Correlation",ascending=False).head(10))  
  
plt.show()