Movie Recommendation

What we have -

1. u.data } Both are
2. u.item } Data files

what we do -

Algorithm

Step-1- Reading data from Data Set.

first we read u.data file and named the Columns.

user_id, item_id, rating, timestamp.

Next reading another data file "u.item" and named the columns item_id & title.

Step-2:- Now merging the both the data fame. and we get a single data file and we merged on Column "item_id".

Data Analysis.

Step-3:- Exploratory

Here we will do some analysis such as if any movie ^Avg rating is 5 and what if only 10-20 people watched that movie and rate that movie 5, which is not a good case to involve

for that we will plot the histogram of item_id and number of rating.

So we do some data sorting along with the number of rating and sort them descending.

Now we plot the histogram.

Here we Conclude that many movies have rated between 2.5 to 4.5 by large number of people but few movies are rated 1 to 2.5 and 4.5 to 5 which is not good to include.

Next we use seaborn to check the clear relation between rating and number of rating and we found that as the rating increases the number of rating also increases and

the [illegible]

some point of time vice versa.

Step-4:- Now we have user_id which rated the movies so, what we will do is we make a matrix of type -

But the problem is not each user rated all the movies, we will get lot of NaN.

So, first we create a matrix. using pivot_table using user_id and title. and values = rating

Step-5:- Now extracting similar movies from matrix by correlation with that movie rating which we are given.

Hence we get correl
movies but we get NaN which is need
to drop. So we drop NaN.
the higest correlation is +1 and lowest
is -1

Now we want only those movies in recomme-
ndation which is rated by large number
of people such as above 100 people.
So we extract those movies and sort them
in descending order.

Therefore we got the result.

∴ This is just a doc to read code
is available at

https://github.com/praveen101295/movie_recomendation_python