Skrár:

1. **Tags.dat**: UserID :: MovieID :: Tag :: Timestamp. Tag er eitt orð eða stutt setning um ákveðna bíómynd sem ákveðinn notandi setti inn. Held við þurfum þetta ekkert
2. **Movies.dat**: MovieID :: Title :: Genres. Segir sig sjálft. Genres er pipe seperated listi þ.a. ef bíómynd fellur undir 2 flokka, Sci-fi og fantasy, þá er í Genres reit: „Sci-fi|Fantasy“
3. **Ratings.dat:** UserID :: MovieID :: Rating :: Timestamp. Hver lína er einkunn eins notanda á einni bíómynd á skalanum 0-5, gefið í heilum og hálfum. Raðað eftir UserID og subsort er MovieID.

Byrja á því að búa til töflu þar sem við erum með allar bíómyndir í einum dálk og meðaleinkunn þeirra frá notendum í öðrum. Til þess þurfum við að safna öllum einkunnum fyrir hverja mynd frá notendum og reikna meðaleinkuninna.

Búa til töflu með Genres. Það er tæmandi listi í readme.html.

**1: Finnið 10 bestu/versu/vinsælustu/???? Kvikmyndirnar ( þ.e. einn góðan topp-10 lista).**

Okkar lausn: Velur allt að 3 genres úr drop down lista og færð top 10 lista fyrir myndir sem falla undir allt sem valið var. Þurfum að velja myndir sem hafa allar genres sem valið var og raða eftir meðaleinkunn, skila svo 10 hæstu.

**2. Veljið einn einstakling og bíómynd. Setjið fram spá hvaða einkunn hann myndi gefa á þessa bíómynd og berið saman við einkunnina ef hún er til staðar. Rökstyðjið aðferðina.**

http://grouplens.org/recommending-for-new-users-is-surprisingly-difficult/  
http://www.ra.ethz.ch/cdstore/www10/papers/pdf/p519.pdf  
http://en.wikipedia.org/wiki/Netflix\_Prize  
http://www.science20.com/random\_walk/predicting\_movie\_ratings\_math\_won\_netflix\_prize  
http://statweb.stanford.edu/~tibs/ElemStatLearn/  
http://shop.oreilly.com/product/9780596529321.do  
http://webdam.inria.fr/Jorge/html/wdmch19.html

Hugmyndir:

Getum prófað hversu vel líkanið okkar virkar með því að taka t.d. 1000 random línur úr movies.dat þar sem við erum með einkunn frá ákveðnum notanda fyrir ákveðna bíómynd og keyrt líkanið okkar fyrir notandann og bíómyndina, fengið mismun milli raunverulegrar einkunnar og okkar spá og fundið meðalmun.

Ef við erum t.d. með 2 stuðla, annars vegar meðaleinkunn myndar og hinsvegar einkunnagjöf notanda, þá er eini sénsinn að leiða út rétta vigt hvors stuðuls (annað en 50/50) að prófa mismunandi samsetningar og logga útkomuna. Velja samsetningu á vigtum stuðla sem gefur lægsta meðalmun spár og raunverulegra einkunna úrtaks.

t.d. einkunn = stuðull\_1 \* (maxeinkunn/fjöldistuðla) + stuðull\_2 \* (maxeinkunn/fjöldistuðla)

Skoðum hvaða einkunnir hann hefur verið að gefa myndum. Ef einhver gefur aldrei góða einkunn þá er ljóst að hann gefur valinni mynd ekki háa einkunn.

Skoða meðaleinkunn myndar, ef hún er góð/slæm þá er líklegra að hann gefi einkunni betri/verri einkunn.

Nota timestamps í ratings til að leiða út hvernig smekkur aðilans hefur breyst, þ.e. sjá hvaða genres eru að fá góðar einkunnir frá honum nýlega og þannig gera betri spá fyrir hvernig hann muni gefa gefinni bíómynd einkunn. „[Timestamps](http://en.wikipedia.org/wiki/Unix_time) represent seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.“

Geta kallað á imdb upplýsingar fyrir gefna mynd. Title í Movies.dat á að vera staf fyrir staf nákvæmlega eins og titillinn sem kemur fyrir á imdb.  
http://imdbpy.sourceforge.net/  
http://www.omdbapi.com/  
http://www.codecademy.com/courses/python-intermediate-en-6zbLp/0/1  
Id á imdb á að vera það sama og MovieID