

## ee219 project 3 Report

name	UID
Jianwei Zhang	905036358
Xiaopei Zhang	004309991
yutong lu	005027184
hengyu lou	005035476

**Question 1: Compute the sparsity of the movie rating dataset, where sparsity**

```
In [42]: from util import *
import matplotlib.pyplot as plt

num_movies = 163949

num_user = 671

df = readData()
num_movie = np.max(df['mid'])
R = getRatingMatrix(df,num_movie)

C = np.copy(R)
C[C>0] = 1

print ("size of R", R.shape)
print (num_movies*num_user)
print (np.sum(C))
print ("sparsity:", np.sum(C)/num_movies/num_user)

size of R (671, 163949)
110009779
100004.0
sparsity: 0.000909046458497
```

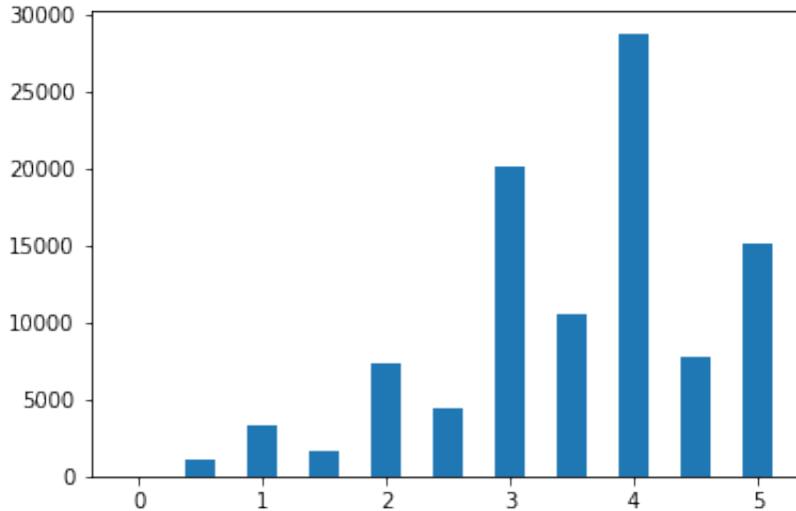
## Question 1 Answer:

There are a total of 163949 movies and 671 users possible number of ratings is 110009779 there are 100004 ratings in the dataset Sparsity is 0.000909046458497

**Question 2: Plot a histogram showing the frequency of the rating values. To be specific, bin the rating values into intervals of width 0.5 and use the binned rating values as the horizontal axis. Count the number of entries in the ratings matrix R with rating values in the binned intervals and use this count as the vertical axis. Briefly comment on the shape of the histogram**

```
In [43]: bins = [-0.25,0.25,0.75,1.25,1.75,2.25,2.75,3.25,3.75,4.25,4.75,5.25]
plt.hist(df['r'],bins=bins,rwidth =0.5)
```

```
Out[43]: (array([      0.,    1101.,    3326.,    1687.,    7271.,    4449.,    20064
      ,
      10538.,    28750.,    7723.,    15095.]),
 array([-0.25,  0.25,  0.75,  1.25,  1.75,  2.25,  2.75,  3.25,  3.75,
        4.25,  4.75,  5.25]),
 <a list of 11 Patch objects>)
```



## Question 2 Answer:

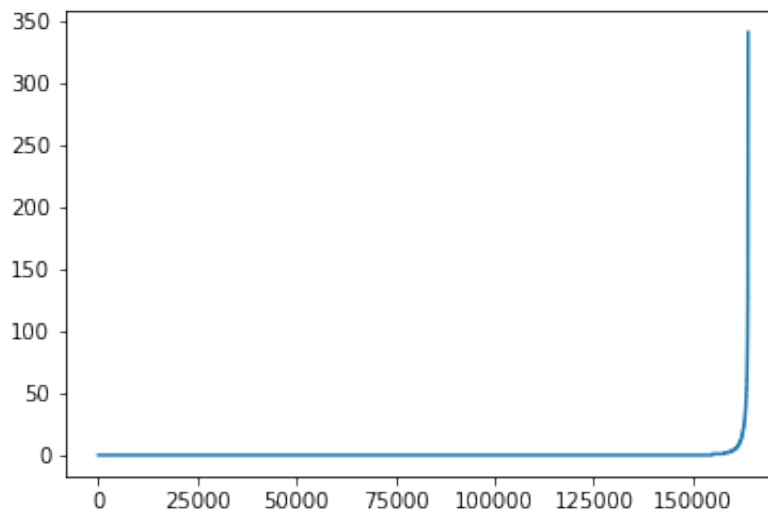
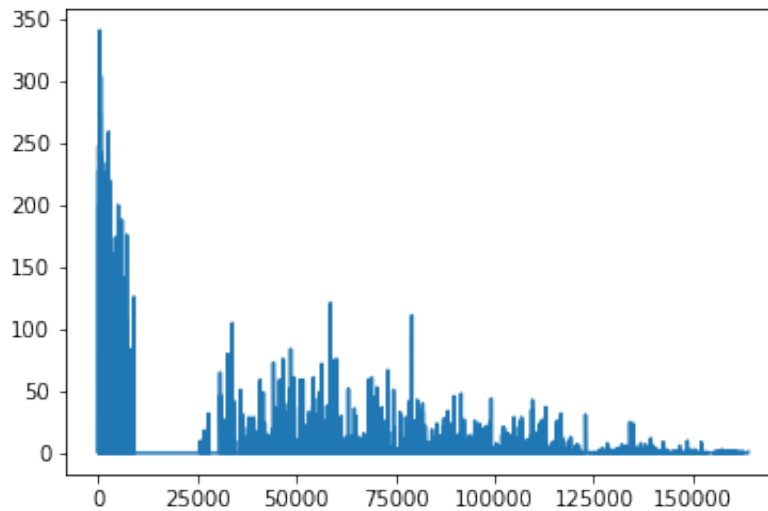
According to the shape of the histogram, most of the ratings are between 3 and 4, only a few ratings are between 0.5 to 2.5

**Question 3: Plot the distribution of ratings among movies. To be specific, the X-axis should be the movie index ordered by decreasing frequency and the Y-axis should be the number of ratings the movie has received.**

```
In [44]: C = np.copy(R)
C[C>0] = 1

num_rating_per_movie = np.sum(C,axis=0)
plt.plot(num_rating_per_movie)
plt.figure()
plt.plot(np.sort(np.trim_zeros(num_rating_per_movie)))
```

Out[44]: [



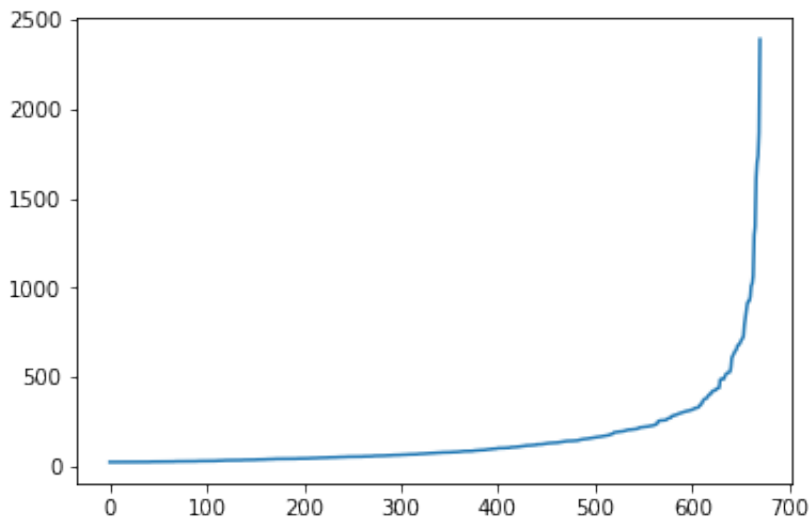
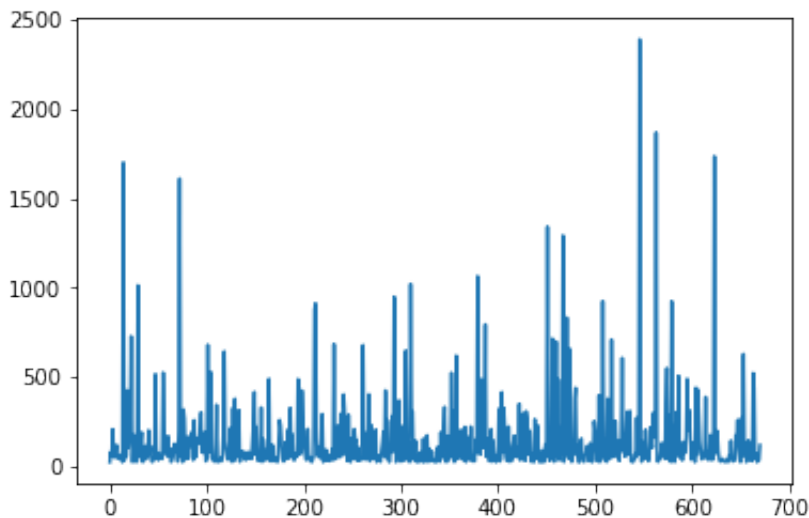
## Question 3 Answer:

The shape of the plot seems to be exponential

**Question 4: Plot the distribution of ratings among users. To be specific, the X-axis should be the user index ordered by decreasing frequency and the Y-axis should be the number of movies the user have rated.**

```
In [45]: num_ratings_per_user = np.sum(C,axis=1)
plt.plot(num_ratings_per_user)
plt.figure()
plt.plot(np.sort(np.trim_zeros(num_ratings_per_user)))
```

Out[45]: [



## Question 4 Answer

The shape of the plot seems to be exponential

**Question 5: Explain the salient features of the distribution found in question 3 and their implications for the recommendation process.**

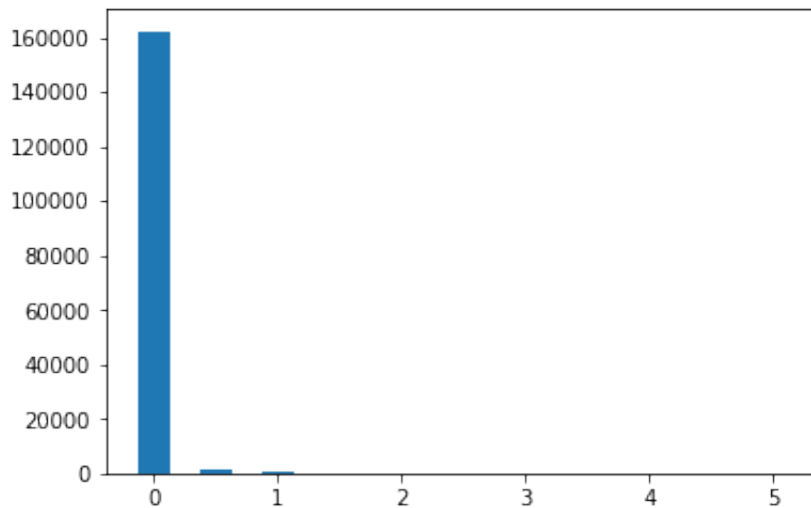
## Question 5 Answer

The number of rating for some movie is very large, indicating that these movies are rated by a lot of people. The exponential shape of the plot shows that the top rated movies are significantly more rated than other less rated movies. Therefore, these top movies make great candidates for our recommendation system.

**Question 6: Compute the variance of the rating values received by each movie. Then, bin the variance values into intervals of width 0.5 and use the binned variance values as the horizontal axis. Count the number of movies with variance values in the binned intervals and use this count as the vertical axis. Briefly comment on the shape of the histogram**

```
In [46]: rating_var_per_movie = np.var(R,axis = 0)
plt.hist(rating_var_per_movie,rwidth =0.5,bins = [-0.25,0.25,0.75,1.25,1.75,2.25,2.75,3.25,3.75,4.25,4.75,5.25])
```

```
Out[46]: (array([ 1.62233000e+05,  1.09900000e+03,  3.23000000e+02,
                  1.20000000e+02,  7.80000000e+01,  4.20000000e+01,
                  2.40000000e+01,  1.00000000e+01,  1.00000000e+01,
                  8.00000000e+00,  2.00000000e+00]),
          array([-0.25,  0.25,  0.75,  1.25,  1.75,  2.25,  2.75,  3.25,  3.75,
                  4.25,  4.75,  5.25]),
          <a list of 11 Patch objects>)
```



## Question 6 Answer:

The graph is not very clear, but from the data we can see that most movies have very small variance while about 500 movies have more than 1 variance.

## Question 7: Write down the formula for $\mu$ in terms of $l_u$ and $r_{uk}$

## Question 7 Answer:

$$\mu_u = \frac{\sum_{i \in I_u} r_{ui}}{|I_u|}$$

We add up all the items' ratings in  $I_u$  and divide by how many items are in  $I_u$ . That is the average rating for user  $u$ .

## Question 8: In plain words, explain the meaning of $I_u \cap I_v$ . Can $I_u \cap I_v = \emptyset$

## Question 8 Answer:

The intersection of  $I_u$  and  $I_v$  is the set of indices of movies that both  $u$  and  $v$  have rated. It is possible to have the intersection to be empty since it's possible for  $u$  and  $v$  to have no common rated movie.

## Question 9: Can you explain the reason behind mean-centering the raw ratings ( $r_{uj} - \mu_v$ ) in the prediction function? (Hint: Consider users who either rate all items highly or rate all items poorly and the impact of these users on the prediction function)

## Question 9 Answer

The zero-centering is a kind of normalization since each time, we have different train and test set, we need the weights to be made under a common standard. And normalization will keep all the parameters at the same scale. With this setup, user data that is extreme will not throw off the entire model.



**Question 10: Design a k-NN collaborative filter to predict the ratings of the movies in the MovieLens dataset and evaluate it's performance using 10-fold cross validation. Sweep k ( number of neighbors) from 2 to 100 in step sizes of 2, and for each k compute the average RMSE and average MAE obtained by averaging the RMSE and MAE across all 10 folds. Plot average RMSE (Y-axis) against k (X-axis) and average MAE (Y-axis) against k (X-axis).**

```
In [47]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

reader = Reader(line_format='user item rating timestamp', sep=',', skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

sim_options = {'name': 'pearson', 'user_base': True}

(knn_rmse, knn_mae) = train_knn(data)

using k = 2
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
```

```
Done computing similarity matrix.
using k = 4
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 6
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 8
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 10
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 12
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
```

Page 12 of 165

```
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 20
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 22
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 24
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 26
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 28
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 30
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 32
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

Page 16 of 165



Page 17 of 165

```
using k = 42
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 44
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 46
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 48
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 50
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 52
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 54
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 56
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 58
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 60
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 62
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
using k = 64
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Done computing similarity matrix.
Done computing similarity matrix.
using k = 66
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 68
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 70
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 72
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 74
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
```



Page 25 of 165

```
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 82
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 84
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 86
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 88
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
```

Page 28 of 165

```
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 96
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 98
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
using k = 100
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Done computing similarity matrix.
Done computing similarity matrix.
k-fold validation finished!
```

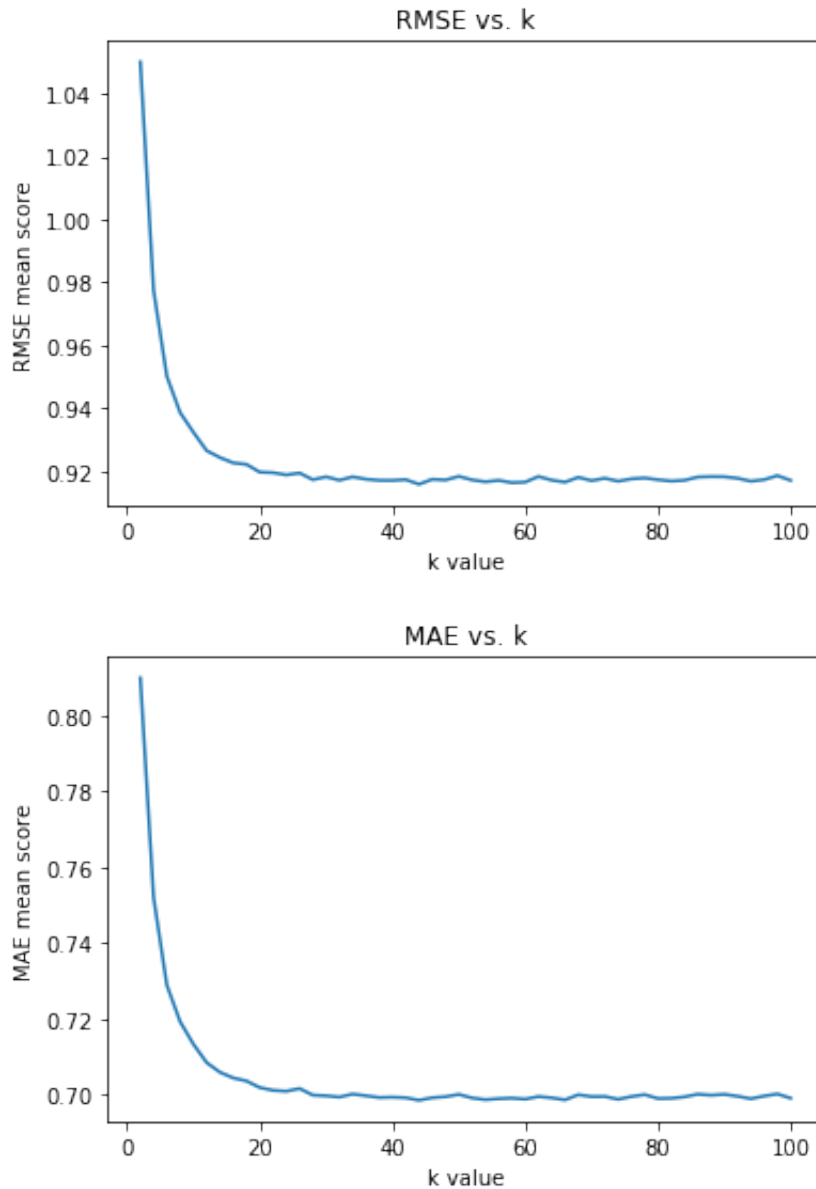
```
In [48]: import matplotlib.pyplot as plt

plt.plot(range(2, 102, 2), knn_rmse)
plt.title('RMSE vs. k')
plt.xlabel('k value')
plt.ylabel('RMSE mean score')

plt.figure()

plt.plot(range(2, 102, 2), knn_mae)
plt.title('MAE vs. k')
plt.xlabel('k value')
plt.ylabel('MAE mean score')
```

```
Out[48]: Text(0,0.5,'MAE mean score')
```



## Question 10 Answer:

The rmse and mae plots are shown above. At around 20, the plot starts to flat out and enter a plain

**Question 11: Use the plot from question 10, to find a 'minimum k'. Note: The term 'minimum k' in this context means that increasing k above the minimum value would not result in a significant decrease in average RMSE or average MAE. If you get the plot correct, then 'minimum k' would correspond to the k value for which average RMSE and average MAE converges to a steady-state value. Please report the steady state values of average RMSE and average MAE**

### **Question 11 Answer:**

The min k from the graph is  $k = 24$ , and the corresponding RMSE is 0.918521258745 and MAE is 0.70097040526

**Question 12,13,14 Trimming test set with popular, popular and high variance**



```
In [49]: # train section

import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

%load_ext autoreload
%autoreload 2

df = readData()
num_movie = 164979
R = getRatingMatrix(df,num_movie)

reader = Reader(line_format='user item rating timestamp', sep=',',skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

sim_options = {'name': 'pearson', 'user_base': True}

knn_trim_rmse_list = train_trim_knn(data,R)
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
using k = 2
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 1.0588
RMSE: 1.0446
RMSE: 1.0789
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 1.0529
RMSE: 1.0140
RMSE: 1.0353
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 1.0639
RMSE: 1.0285
RMSE: 1.0616
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 1.0715
RMSE: 1.0392
RMSE: 1.0067
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 1.0543  
RMSE: 1.0351  
RMSE: 1.0404  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 1.0519  
RMSE: 1.0192  
RMSE: 0.9806  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 1.0772  
RMSE: 1.0480  
RMSE: 1.0781  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 1.0597  
RMSE: 1.0510  
RMSE: 1.0317  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 1.0639  
RMSE: 1.0469  
RMSE: 1.0328  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 1.0596  
RMSE: 1.0649  
RMSE: 1.0581  
using k = 4  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9592  
RMSE: 0.9608  
RMSE: 1.0175  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9934  
RMSE: 0.9646  
RMSE: 0.9522  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9642  
RMSE: 0.9702  
RMSE: 0.9696  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9841  
RMSE: 0.9977  
RMSE: 0.9335
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9778
RMSE: 0.9850
RMSE: 0.9250
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9686
RMSE: 0.9680
RMSE: 0.9058
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9851
RMSE: 0.9795
RMSE: 0.9219
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9719
RMSE: 1.0028
RMSE: 0.9566
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9636
RMSE: 0.9761
RMSE: 0.9832
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9609
RMSE: 0.9684
RMSE: 0.9491
using k = 6
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9452
RMSE: 0.9522
RMSE: 0.9027
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9386
RMSE: 0.9606
RMSE: 0.9149
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9537
RMSE: 0.9537
RMSE: 0.9063
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9456
RMSE: 0.9475
```

```
RMSE: 0.9038
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9311
RMSE: 0.9454
RMSE: 0.8977
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9442
RMSE: 0.9565
RMSE: 0.9533
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9392
RMSE: 0.9744
RMSE: 0.9086
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9463
RMSE: 0.9630
RMSE: 0.9293
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9450
RMSE: 0.9629
RMSE: 0.9426
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9682
RMSE: 0.9672
RMSE: 0.9630
using k = 8
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9217
RMSE: 0.9364
RMSE: 0.8812
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9368
RMSE: 0.9501
RMSE: 0.9105
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9294
RMSE: 0.9561
RMSE: 0.9243
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9284
```

```
RMSE: 0.9471
RMSE: 0.9459
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9559
RMSE: 0.9494
RMSE: 0.9746
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9423
RMSE: 0.9537
RMSE: 0.8926
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9135
RMSE: 0.9621
RMSE: 0.8728
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9460
RMSE: 0.9653
RMSE: 0.9166
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9341
RMSE: 0.9404
RMSE: 0.8475
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9223
RMSE: 0.9434
RMSE: 0.8592
using k = 10
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9327
RMSE: 0.9447
RMSE: 0.9209
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9189
RMSE: 0.9442
RMSE: 0.8528
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9202
RMSE: 0.9492
RMSE: 0.8751
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9139
RMSE: 0.9698
RMSE: 0.8865
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9321
RMSE: 0.9342
RMSE: 0.8941
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9099
RMSE: 0.9442
RMSE: 0.9229
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9141
RMSE: 0.9424
RMSE: 0.8841
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9135
RMSE: 0.9321
RMSE: 0.8407
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9356
RMSE: 0.9547
RMSE: 0.9391
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9208
RMSE: 0.9308
RMSE: 0.9262
using k = 12
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9215
RMSE: 0.9323
RMSE: 0.8839
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9039
RMSE: 0.9455
RMSE: 0.8517
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9159
RMSE: 0.9389
RMSE: 0.9251
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.9191  
RMSE: 0.9480  
RMSE: 0.8925  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9221  
RMSE: 0.9441  
RMSE: 0.8848  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9014  
RMSE: 0.9456  
RMSE: 0.8564  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9254  
RMSE: 0.9424  
RMSE: 0.9372  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9190  
RMSE: 0.9498  
RMSE: 0.8580  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9068  
RMSE: 0.9418  
RMSE: 0.8579  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9288  
RMSE: 0.9372  
RMSE: 0.9108  
using k = 14  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9142  
RMSE: 0.9307  
RMSE: 0.9025  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9125  
RMSE: 0.9310  
RMSE: 0.8903  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8990  
RMSE: 0.9413  
RMSE: 0.8612
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9236
RMSE: 0.9154
RMSE: 0.9141
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9136
RMSE: 0.9534
RMSE: 0.8474
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8903
RMSE: 0.9357
RMSE: 0.7965
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9076
RMSE: 0.9410
RMSE: 0.8712
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9222
RMSE: 0.9587
RMSE: 0.9019
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9272
RMSE: 0.9561
RMSE: 0.9017
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9289
RMSE: 0.9353
RMSE: 0.9346
using k = 16
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9271
RMSE: 0.9531
RMSE: 0.9098
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9150
RMSE: 0.9340
RMSE: 0.8968
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9172
RMSE: 0.9501
```



```
RMSE: 0.8526
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9102
RMSE: 0.9421
RMSE: 0.8307
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9022
RMSE: 0.9412
RMSE: 0.8489
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9151
RMSE: 0.9176
RMSE: 0.8894
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9082
RMSE: 0.9353
RMSE: 0.9169
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9154
RMSE: 0.9352
RMSE: 0.8858
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8927
RMSE: 0.9404
RMSE: 0.8862
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9125
RMSE: 0.9360
RMSE: 0.8718
using k = 18
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9081
RMSE: 0.9188
RMSE: 0.8991
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9106
RMSE: 0.9497
RMSE: 0.8847
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8987
```

```
RMSE: 0.9142
RMSE: 0.8570
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9098
RMSE: 0.9267
RMSE: 0.9194
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9170
RMSE: 0.9544
RMSE: 0.8872
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9091
RMSE: 0.9443
RMSE: 0.8429
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9165
RMSE: 0.9316
RMSE: 0.8694
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9012
RMSE: 0.9497
RMSE: 0.8500
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9038
RMSE: 0.9390
RMSE: 0.8795
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9102
RMSE: 0.9344
RMSE: 0.8740
using k = 20
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9172
RMSE: 0.9416
RMSE: 0.9136
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9208
RMSE: 0.9560
RMSE: 0.8959
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9067
RMSE: 0.9518
RMSE: 0.8676
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9035
RMSE: 0.9311
RMSE: 0.8830
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9179
RMSE: 0.9300
RMSE: 0.9003
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9066
RMSE: 0.9488
RMSE: 0.8357
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9085
RMSE: 0.9519
RMSE: 0.8328
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8933
RMSE: 0.9337
RMSE: 0.8450
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9016
RMSE: 0.9170
RMSE: 0.8996
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9033
RMSE: 0.9208
RMSE: 0.8691
using k = 22
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.9465
RMSE: 0.8997
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9112
RMSE: 0.9366
RMSE: 0.8962
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.9024  
RMSE: 0.9278  
RMSE: 0.8800  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9036  
RMSE: 0.9386  
RMSE: 0.8215  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9069  
RMSE: 0.9287  
RMSE: 0.8753  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8963  
RMSE: 0.9468  
RMSE: 0.8962  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9169  
RMSE: 0.9470  
RMSE: 0.8402  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9021  
RMSE: 0.9390  
RMSE: 0.8758  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9159  
RMSE: 0.9451  
RMSE: 0.8968  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8984  
RMSE: 0.9289  
RMSE: 0.8534  
using k = 24  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8947  
RMSE: 0.9193  
RMSE: 0.8534  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9077  
RMSE: 0.9210  
RMSE: 0.8193
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9238
RMSE: 0.9362
RMSE: 0.9095
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8888
RMSE: 0.9389
RMSE: 0.8287
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9186
RMSE: 0.9495
RMSE: 0.8978
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9078
RMSE: 0.9302
RMSE: 0.8795
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9043
RMSE: 0.9392
RMSE: 0.9012
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8973
RMSE: 0.9251
RMSE: 0.8619
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9636
RMSE: 0.9081
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9331
RMSE: 0.9247
RMSE: 0.8931
using k = 26
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8974
RMSE: 0.9177
RMSE: 0.8805
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9175
RMSE: 0.9226
```

```
RMSE: 0.8702
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9028
RMSE: 0.9424
RMSE: 0.8096
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9001
RMSE: 0.9225
RMSE: 0.8634
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9016
RMSE: 0.9310
RMSE: 0.8738
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8921
RMSE: 0.9346
RMSE: 0.9073
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9134
RMSE: 0.9495
RMSE: 0.8761
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9296
RMSE: 0.9385
RMSE: 0.9115
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9072
RMSE: 0.9402
RMSE: 0.8832
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9001
RMSE: 0.9551
RMSE: 0.8424
using k = 28
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9099
RMSE: 0.9316
RMSE: 0.8210
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9165
```

```
RMSE: 0.9445
RMSE: 0.8861
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9124
RMSE: 0.9429
RMSE: 0.8184
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9113
RMSE: 0.9384
RMSE: 0.9231
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8823
RMSE: 0.9190
RMSE: 0.8524
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8974
RMSE: 0.9281
RMSE: 0.8859
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9023
RMSE: 0.9481
RMSE: 0.9269
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9081
RMSE: 0.9366
RMSE: 0.8325
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9080
RMSE: 0.9461
RMSE: 0.8323
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9106
RMSE: 0.9195
RMSE: 0.9049
using k = 30
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9126
RMSE: 0.9572
RMSE: 0.8645
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9019
RMSE: 0.9354
RMSE: 0.8883
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9015
RMSE: 0.9288
RMSE: 0.8356
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8883
RMSE: 0.9444
RMSE: 0.8307
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9035
RMSE: 0.9361
RMSE: 0.8111
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9169
RMSE: 0.9341
RMSE: 0.8843
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8962
RMSE: 0.9328
RMSE: 0.8791
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9121
RMSE: 0.9298
RMSE: 0.8895
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9149
RMSE: 0.9188
RMSE: 0.9847
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9073
RMSE: 0.9387
RMSE: 0.8395
using k = 32
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9130
RMSE: 0.9369
RMSE: 0.8534
Computing the pearson similarity matrix...
```



```
Done computing similarity matrix.  
RMSE: 0.8970  
RMSE: 0.9252  
RMSE: 0.8631  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9068  
RMSE: 0.9441  
RMSE: 0.8403  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8872  
RMSE: 0.9355  
RMSE: 0.8595  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8856  
RMSE: 0.9260  
RMSE: 0.8642  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9148  
RMSE: 0.9419  
RMSE: 0.9486  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9091  
RMSE: 0.9281  
RMSE: 0.8662  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9129  
RMSE: 0.9302  
RMSE: 0.8872  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9125  
RMSE: 0.9387  
RMSE: 0.8617  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9192  
RMSE: 0.9267  
RMSE: 0.8965  
using k = 34  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9172  
RMSE: 0.9668  
RMSE: 0.8296
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9180
RMSE: 0.9356
RMSE: 0.9206
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9216
RMSE: 0.9388
RMSE: 0.8642
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9011
RMSE: 0.9297
RMSE: 0.8241
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8915
RMSE: 0.9279
RMSE: 0.8961
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9016
RMSE: 0.9332
RMSE: 0.8286
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9020
RMSE: 0.9322
RMSE: 0.8382
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9008
RMSE: 0.9345
RMSE: 0.8777
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8999
RMSE: 0.9604
RMSE: 0.8873
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8963
RMSE: 0.9029
RMSE: 0.9181
using k = 36
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9088
RMSE: 0.9343
```

```
RMSE: 0.8960
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8948
RMSE: 0.9483
RMSE: 0.7782
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8917
RMSE: 0.9405
RMSE: 0.8958
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9152
RMSE: 0.9534
RMSE: 0.8513
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8929
RMSE: 0.9229
RMSE: 0.8151
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8984
RMSE: 0.9209
RMSE: 0.8612
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9124
RMSE: 0.9294
RMSE: 0.8539
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.9279
RMSE: 0.8982
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9005
RMSE: 0.9286
RMSE: 0.8956
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9062
RMSE: 0.9319
RMSE: 0.9251
using k = 38
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8942
```

```
RMSE: 0.9511
RMSE: 0.8419
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8969
RMSE: 0.9257
RMSE: 0.9104
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8880
RMSE: 0.9090
RMSE: 0.8331
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9146
RMSE: 0.9405
RMSE: 0.8809
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8994
RMSE: 0.9360
RMSE: 0.8248
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9111
RMSE: 0.9560
RMSE: 0.9000
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9150
RMSE: 0.9320
RMSE: 0.8656
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8945
RMSE: 0.9308
RMSE: 0.8839
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9155
RMSE: 0.9441
RMSE: 0.9248
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9097
RMSE: 0.9175
RMSE: 0.8491
using k = 40
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9159
RMSE: 0.9404
RMSE: 0.8680
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9104
RMSE: 0.9461
RMSE: 0.8382
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8982
RMSE: 0.9497
RMSE: 0.9029
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8983
RMSE: 0.9107
RMSE: 0.9115
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.9469
RMSE: 0.8641
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9072
RMSE: 0.9316
RMSE: 0.8646
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8941
RMSE: 0.9402
RMSE: 0.8598
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9105
RMSE: 0.9476
RMSE: 0.8423
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9019
RMSE: 0.9273
RMSE: 0.8591
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8959
RMSE: 0.9098
RMSE: 0.8738
using k = 42
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.
RMSE: 0.9005
RMSE: 0.9337
RMSE: 0.8411
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.9299
RMSE: 0.9128
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9012
RMSE: 0.9192
RMSE: 0.9045
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9124
RMSE: 0.9254
RMSE: 0.8772
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8950
RMSE: 0.9219
RMSE: 0.8970
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9035
RMSE: 0.9331
RMSE: 0.7913
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9035
RMSE: 0.9478
RMSE: 0.8720
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8980
RMSE: 0.9616
RMSE: 0.8682
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8997
RMSE: 0.9242
RMSE: 0.8483
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9031
RMSE: 0.9532
RMSE: 0.8897
using k = 44
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9070
RMSE: 0.9504
RMSE: 0.8864
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8880
RMSE: 0.9329
RMSE: 0.8837
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9063
RMSE: 0.9255
RMSE: 0.8965
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8986
RMSE: 0.9228
RMSE: 0.8446
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9341
RMSE: 0.9471
RMSE: 0.9323
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8919
RMSE: 0.9295
RMSE: 0.8403
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9062
RMSE: 0.9419
RMSE: 0.8514
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9009
RMSE: 0.9272
RMSE: 0.8622
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8887
RMSE: 0.9572
RMSE: 0.8318
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9189
RMSE: 0.9238
RMSE: 0.8473
```

```
using k = 46
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9022
RMSE: 0.9198
RMSE: 0.9088
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9325
RMSE: 0.9196
RMSE: 0.9138
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9046
RMSE: 0.9597
RMSE: 0.8598
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9184
RMSE: 0.9060
RMSE: 0.8396
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9204
RMSE: 0.9383
RMSE: 0.8519
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8969
RMSE: 0.9312
RMSE: 0.8426
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9072
RMSE: 0.9304
RMSE: 0.8898
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8996
RMSE: 0.9464
RMSE: 0.8410
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8977
RMSE: 0.9447
RMSE: 0.8631
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8835
RMSE: 0.9499
```



```
RMSE: 0.8481
using k = 48
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9153
RMSE: 0.9250
RMSE: 0.8868
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9012
RMSE: 0.9447
RMSE: 0.9293
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9020
RMSE: 0.9240
RMSE: 0.8551
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9006
RMSE: 0.9365
RMSE: 0.9121
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8983
RMSE: 0.9278
RMSE: 0.8577
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9047
RMSE: 0.9376
RMSE: 0.8430
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9039
RMSE: 0.9545
RMSE: 0.9041
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9103
RMSE: 0.9317
RMSE: 0.7860
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9055
RMSE: 0.9497
RMSE: 0.8359
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8932
```

```
RMSE: 0.9250
RMSE: 0.8538
using k = 50
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9214
RMSE: 0.9419
RMSE: 0.8951
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8976
RMSE: 0.9523
RMSE: 0.8808
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9175
RMSE: 0.9425
RMSE: 0.8984
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9026
RMSE: 0.9621
RMSE: 0.8287
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8861
RMSE: 0.9141
RMSE: 0.8367
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9067
RMSE: 0.9295
RMSE: 0.8204
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9082
RMSE: 0.9163
RMSE: 0.8648
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8986
RMSE: 0.9343
RMSE: 0.8877
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8999
RMSE: 0.9225
RMSE: 0.9356
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9011
RMSE: 0.9227
RMSE: 0.8310
using k = 52
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9032
RMSE: 0.9055
RMSE: 0.9163
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9034
RMSE: 0.9347
RMSE: 0.8699
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8937
RMSE: 0.9151
RMSE: 0.8620
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9085
RMSE: 0.9421
RMSE: 0.8925
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9306
RMSE: 0.9557
RMSE: 0.9030
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9078
RMSE: 0.9351
RMSE: 0.8665
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8974
RMSE: 0.9346
RMSE: 0.8348
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9182
RMSE: 0.9209
RMSE: 0.8796
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8816
RMSE: 0.9672
RMSE: 0.8274
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.8971  
RMSE: 0.9553  
RMSE: 0.8179  
using k = 54  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9150  
RMSE: 0.9455  
RMSE: 0.9017  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8936  
RMSE: 0.9428  
RMSE: 0.8293  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9098  
RMSE: 0.9440  
RMSE: 0.8829  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9013  
RMSE: 0.9482  
RMSE: 0.8438  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8990  
RMSE: 0.9323  
RMSE: 0.8616  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8940  
RMSE: 0.9156  
RMSE: 0.8372  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9076  
RMSE: 0.9181  
RMSE: 0.8532  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9067  
RMSE: 0.9233  
RMSE: 0.9410  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9143  
RMSE: 0.9377  
RMSE: 0.8286
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9036
RMSE: 0.9431
RMSE: 0.9091
using k = 56
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8972
RMSE: 0.9447
RMSE: 0.8968
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9039
RMSE: 0.9310
RMSE: 0.8380
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9102
RMSE: 0.9207
RMSE: 0.8258
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8943
RMSE: 0.9293
RMSE: 0.8705
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9074
RMSE: 0.9478
RMSE: 0.9175
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9244
RMSE: 0.9671
RMSE: 0.9649
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9088
RMSE: 0.9130
RMSE: 0.9017
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8925
RMSE: 0.9263
RMSE: 0.8122
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9088
RMSE: 0.9503
```

```
RMSE: 0.8175
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8971
RMSE: 0.9170
RMSE: 0.8299
using k = 58
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9180
RMSE: 0.9270
RMSE: 0.8841
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9090
RMSE: 0.9117
RMSE: 0.9168
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9043
RMSE: 0.9205
RMSE: 0.8495
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9148
RMSE: 0.9424
RMSE: 0.8433
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9126
RMSE: 0.9373
RMSE: 0.8720
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8951
RMSE: 0.9582
RMSE: 0.8135
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8923
RMSE: 0.9209
RMSE: 0.8967
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9077
RMSE: 0.9469
RMSE: 0.8983
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8968
```

```
RMSE: 0.9383
RMSE: 0.7948
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9005
RMSE: 0.9317
RMSE: 0.9077
using k = 60
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9199
RMSE: 0.9362
RMSE: 0.8910
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9279
RMSE: 0.8162
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9063
RMSE: 0.9306
RMSE: 0.8215
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9150
RMSE: 0.9484
RMSE: 0.8537
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9061
RMSE: 0.9399
RMSE: 0.9391
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9120
RMSE: 0.9370
RMSE: 0.7934
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8876
RMSE: 0.9442
RMSE: 0.8927
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8953
RMSE: 0.9113
RMSE: 0.8165
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9171
RMSE: 0.9443
RMSE: 0.9295
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8883
RMSE: 0.9344
RMSE: 0.9231
using k = 62
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9051
RMSE: 0.9496
RMSE: 0.8314
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8958
RMSE: 0.9378
RMSE: 0.8798
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9067
RMSE: 0.9405
RMSE: 0.8619
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9114
RMSE: 0.9420
RMSE: 0.8254
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9025
RMSE: 0.9381
RMSE: 0.9127
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.8987
RMSE: 0.8536
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9046
RMSE: 0.9330
RMSE: 0.8301
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9056
RMSE: 0.9383
RMSE: 0.8617
Computing the pearson similarity matrix...
```



```
Done computing similarity matrix.  
RMSE: 0.8925  
RMSE: 0.9324  
RMSE: 0.8794  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9031  
RMSE: 0.9380  
RMSE: 0.9567  
using k = 64  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8966  
RMSE: 0.9491  
RMSE: 0.8493  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9070  
RMSE: 0.9320  
RMSE: 0.8396  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9038  
RMSE: 0.9549  
RMSE: 0.9077  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9085  
RMSE: 0.9198  
RMSE: 0.8964  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8955  
RMSE: 0.9552  
RMSE: 0.8825  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9058  
RMSE: 0.9278  
RMSE: 0.8756  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9200  
RMSE: 0.9463  
RMSE: 0.8866  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9036  
RMSE: 0.9177  
RMSE: 0.8465
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8911
RMSE: 0.9305
RMSE: 0.8203
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9113
RMSE: 0.9278
RMSE: 0.8655
using k = 66
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8975
RMSE: 0.9472
RMSE: 0.8498
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9068
RMSE: 0.9231
RMSE: 0.8741
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8888
RMSE: 0.9442
RMSE: 0.8393
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9022
RMSE: 0.9336
RMSE: 0.8460
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9046
RMSE: 0.9191
RMSE: 0.8693
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9197
RMSE: 0.9423
RMSE: 0.9302
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9190
RMSE: 0.9234
RMSE: 0.8902
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9093
RMSE: 0.9315
```

```
RMSE: 0.8977
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9028
RMSE: 0.9355
RMSE: 0.8286
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8917
RMSE: 0.9339
RMSE: 0.8364
using k = 68
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8943
RMSE: 0.9278
RMSE: 0.9095
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9001
RMSE: 0.9396
RMSE: 0.9223
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9150
RMSE: 0.9452
RMSE: 0.9367
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8862
RMSE: 0.9317
RMSE: 0.8123
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9209
RMSE: 0.9175
RMSE: 0.8642
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8999
RMSE: 0.9429
RMSE: 0.8231
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9195
RMSE: 0.9392
RMSE: 0.9416
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8992
```

```
RMSE: 0.9397
RMSE: 0.8004
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8894
RMSE: 0.9382
RMSE: 0.8000
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9149
RMSE: 0.9268
RMSE: 0.8616
using k = 70
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9037
RMSE: 0.9214
RMSE: 0.8865
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8848
RMSE: 0.9515
RMSE: 0.7951
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9006
RMSE: 0.9229
RMSE: 0.8979
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9131
RMSE: 0.9303
RMSE: 0.8752
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9126
RMSE: 0.9250
RMSE: 0.8939
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9144
RMSE: 0.9365
RMSE: 0.8860
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8969
RMSE: 0.9330
RMSE: 0.7978
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9180
RMSE: 0.9423
RMSE: 0.8793
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9003
RMSE: 0.9412
RMSE: 0.8785
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9113
RMSE: 0.9488
RMSE: 0.8787
using k = 72
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9129
RMSE: 0.9284
RMSE: 0.9817
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8876
RMSE: 0.9453
RMSE: 0.8790
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8977
RMSE: 0.9181
RMSE: 0.8677
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9054
RMSE: 0.9444
RMSE: 0.8159
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9129
RMSE: 0.9452
RMSE: 0.8603
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8998
RMSE: 0.9307
RMSE: 0.8499
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9042
RMSE: 0.9493
RMSE: 0.8107
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.9046  
RMSE: 0.9253  
RMSE: 0.8985  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9047  
RMSE: 0.9181  
RMSE: 0.8520  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9050  
RMSE: 0.9327  
RMSE: 0.8751  
using k = 74  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9058  
RMSE: 0.9382  
RMSE: 0.9184  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9132  
RMSE: 0.9343  
RMSE: 0.9110  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9056  
RMSE: 0.9323  
RMSE: 0.9129  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8827  
RMSE: 0.9411  
RMSE: 0.8499  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9091  
RMSE: 0.9361  
RMSE: 0.8696  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9261  
RMSE: 0.9133  
RMSE: 0.9252  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9083  
RMSE: 0.9213  
RMSE: 0.8772
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9151
RMSE: 0.9383
RMSE: 0.8524
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8954
RMSE: 0.9395
RMSE: 0.8337
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8932
RMSE: 0.9518
RMSE: 0.7503
using k = 76
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9302
RMSE: 0.9177
RMSE: 0.8849
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9027
RMSE: 0.9198
RMSE: 0.8759
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8788
RMSE: 0.9325
RMSE: 0.8283
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8968
RMSE: 0.9356
RMSE: 0.8895
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9096
RMSE: 0.9351
RMSE: 0.9196
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9058
RMSE: 0.9652
RMSE: 0.8698
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8947
RMSE: 0.9455
```

```
RMSE: 0.8839
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9130
RMSE: 0.9349
RMSE: 0.8555
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9148
RMSE: 0.9486
RMSE: 0.8116
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9053
RMSE: 0.9237
RMSE: 0.8434
using k = 78
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9116
RMSE: 0.9132
RMSE: 0.9206
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8978
RMSE: 0.9323
RMSE: 0.8036
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9133
RMSE: 0.9346
RMSE: 0.8714
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8963
RMSE: 0.9527
RMSE: 0.9069
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9066
RMSE: 0.9525
RMSE: 0.8601
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9026
RMSE: 0.9551
RMSE: 0.9076
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9220
```



```
RMSE: 0.9463
RMSE: 0.8539
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9100
RMSE: 0.8706
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9027
RMSE: 0.9268
RMSE: 0.8443
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9020
RMSE: 0.9301
RMSE: 0.8773
using k = 80
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9174
RMSE: 0.9343
RMSE: 0.9236
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8971
RMSE: 0.9486
RMSE: 0.9132
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9090
RMSE: 0.9359
RMSE: 0.8139
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8884
RMSE: 0.9259
RMSE: 0.8820
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8947
RMSE: 0.9404
RMSE: 0.8422
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9141
RMSE: 0.9246
RMSE: 0.7902
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9062
RMSE: 0.9372
RMSE: 0.8524
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9101
RMSE: 0.9219
RMSE: 0.8613
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9133
RMSE: 0.9437
RMSE: 0.9017
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8992
RMSE: 0.9462
RMSE: 0.9146
using k = 82
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9144
RMSE: 0.9465
RMSE: 0.9242
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9001
RMSE: 0.9383
RMSE: 0.8554
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9098
RMSE: 0.9291
RMSE: 0.8713
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9261
RMSE: 0.9350
RMSE: 0.9046
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9471
RMSE: 0.8218
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8980
RMSE: 0.9624
RMSE: 0.8879
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.8997  
RMSE: 0.9345  
RMSE: 0.8809  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9038  
RMSE: 0.9233  
RMSE: 0.8541  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8846  
RMSE: 0.9373  
RMSE: 0.8270  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9096  
RMSE: 0.9170  
RMSE: 0.8762  
using k = 84  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9035  
RMSE: 0.9429  
RMSE: 0.8633  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9046  
RMSE: 0.9243  
RMSE: 0.8856  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9053  
RMSE: 0.9292  
RMSE: 0.8549  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9131  
RMSE: 0.9402  
RMSE: 0.8887  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8902  
RMSE: 0.9357  
RMSE: 0.8535  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9005  
RMSE: 0.9382  
RMSE: 0.8196
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9083
RMSE: 0.9377
RMSE: 0.9162
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9275
RMSE: 0.9308
RMSE: 0.9388
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9099
RMSE: 0.9498
RMSE: 0.8523
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8945
RMSE: 0.9239
RMSE: 0.8267
using k = 86
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8900
RMSE: 0.9231
RMSE: 0.8906
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9148
RMSE: 0.9373
RMSE: 0.7929
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8908
RMSE: 0.9241
RMSE: 0.8671
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8866
RMSE: 0.9343
RMSE: 0.8645
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9299
RMSE: 0.9450
RMSE: 0.8952
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8958
RMSE: 0.9229
```

```
RMSE: 0.8315
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9022
RMSE: 0.9452
RMSE: 0.9025
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9067
RMSE: 0.9434
RMSE: 0.8908
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9189
RMSE: 0.9442
RMSE: 0.8911
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9169
RMSE: 0.9341
RMSE: 0.8523
using k = 88
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9098
RMSE: 0.9528
RMSE: 0.9080
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9111
RMSE: 0.9548
RMSE: 0.8479
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9113
RMSE: 0.9440
RMSE: 0.8364
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9071
RMSE: 0.9237
RMSE: 0.9123
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8976
RMSE: 0.9477
RMSE: 0.8525
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9037
```

```
RMSE: 0.9254
RMSE: 0.8889
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8957
RMSE: 0.9228
RMSE: 0.8239
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9112
RMSE: 0.9146
RMSE: 0.9136
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8909
RMSE: 0.9368
RMSE: 0.8008
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9127
RMSE: 0.9182
RMSE: 0.9126
using k = 90
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9237
RMSE: 0.9553
RMSE: 0.8245
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8943
RMSE: 0.9283
RMSE: 0.8576
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9163
RMSE: 0.9464
RMSE: 0.9249
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8994
RMSE: 0.9226
RMSE: 0.8376
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8944
RMSE: 0.9374
RMSE: 0.8827
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
RMSE: 0.9068
RMSE: 0.9159
RMSE: 0.8617
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8795
RMSE: 0.9311
RMSE: 0.8516
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8970
RMSE: 0.9324
RMSE: 0.9034
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9241
RMSE: 0.9686
RMSE: 0.8681
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9009
RMSE: 0.9248
RMSE: 0.8487
using k = 92
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9085
RMSE: 0.9229
RMSE: 0.8285
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9360
RMSE: 0.8767
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8946
RMSE: 0.9346
RMSE: 0.8550
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9158
RMSE: 0.9263
RMSE: 0.9296
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8907
RMSE: 0.9372
RMSE: 0.7985
Computing the pearson similarity matrix...
```

```
Done computing similarity matrix.  
RMSE: 0.9208  
RMSE: 0.9347  
RMSE: 0.8329  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9174  
RMSE: 0.9510  
RMSE: 0.9200  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8925  
RMSE: 0.9282  
RMSE: 0.8588  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9215  
RMSE: 0.9451  
RMSE: 0.8858  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8895  
RMSE: 0.9524  
RMSE: 0.8702  
using k = 94  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9061  
RMSE: 0.9460  
RMSE: 0.8813  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8917  
RMSE: 0.9335  
RMSE: 0.8759  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.8988  
RMSE: 0.9257  
RMSE: 0.8484  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9129  
RMSE: 0.9263  
RMSE: 0.9000  
Computing the pearson similarity matrix...  
Done computing similarity matrix.  
RMSE: 0.9081  
RMSE: 0.9305  
RMSE: 0.8520
```



```
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9169
RMSE: 0.9371
RMSE: 0.8242
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8930
RMSE: 0.9448
RMSE: 0.8848
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9055
RMSE: 0.9357
RMSE: 0.8347
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9088
RMSE: 0.9400
RMSE: 0.9127
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9076
RMSE: 0.9198
RMSE: 0.8496
using k = 96
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9131
RMSE: 0.9269
RMSE: 0.9194
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8991
RMSE: 0.9334
RMSE: 0.8804
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9081
RMSE: 0.9347
RMSE: 0.8407
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8924
RMSE: 0.9448
RMSE: 0.8421
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9030
RMSE: 0.9543
```

```
RMSE: 0.8694
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8933
RMSE: 0.9381
RMSE: 0.8850
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9126
RMSE: 0.9348
RMSE: 0.8634
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8933
RMSE: 0.9252
RMSE: 0.8761
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9264
RMSE: 0.9256
RMSE: 0.8983
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9102
RMSE: 0.9475
RMSE: 0.8333
using k = 98
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9213
RMSE: 0.9217
RMSE: 0.8190
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9244
RMSE: 0.9453
RMSE: 0.8431
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9048
RMSE: 0.9480
RMSE: 0.8802
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8901
RMSE: 0.9274
RMSE: 0.8186
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9100
```

```
RMSE: 0.9174
RMSE: 0.9199
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8946
RMSE: 0.9380
RMSE: 0.8944
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9120
RMSE: 0.9124
RMSE: 0.9119
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9052
RMSE: 0.9452
RMSE: 0.8476
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8949
RMSE: 0.9381
RMSE: 0.8634
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8992
RMSE: 0.9608
RMSE: 0.8771
using k = 100
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9134
RMSE: 0.9302
RMSE: 0.8948
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8974
RMSE: 0.9147
RMSE: 0.8454
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9002
RMSE: 0.9365
RMSE: 0.8652
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9107
RMSE: 0.9679
RMSE: 0.8599
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```

RMSE: 0.8957
RMSE: 0.9278
RMSE: 0.8552
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9148
RMSE: 0.9374
RMSE: 0.8903
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.8952
RMSE: 0.9188
RMSE: 0.7852
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9035
RMSE: 0.9430
RMSE: 0.8957
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9186
RMSE: 0.9365
RMSE: 0.9071
Computing the pearson similarity matrix...
Done computing similarity matrix.
RMSE: 0.9048
RMSE: 0.9507
RMSE: 0.8653
KNN with trim is finished!!

```

In [50]: *# plot section*

```

import matplotlib.pyplot as plt

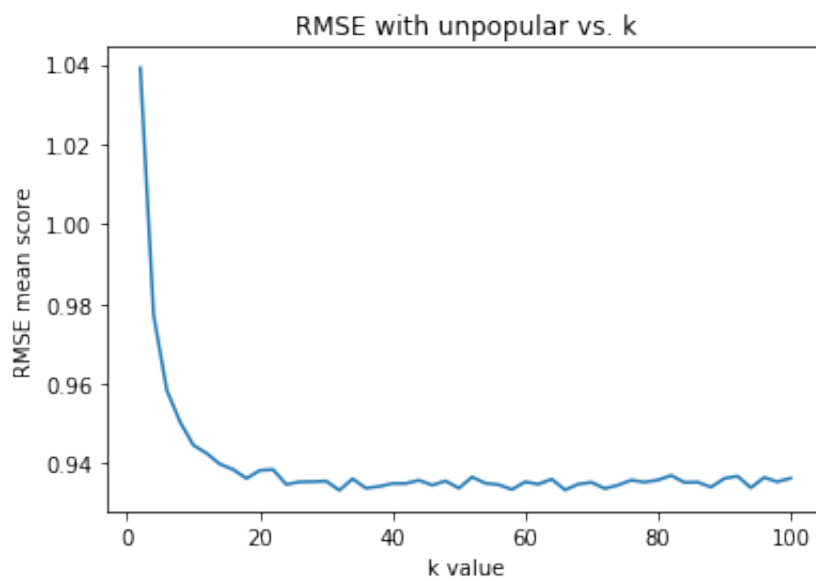
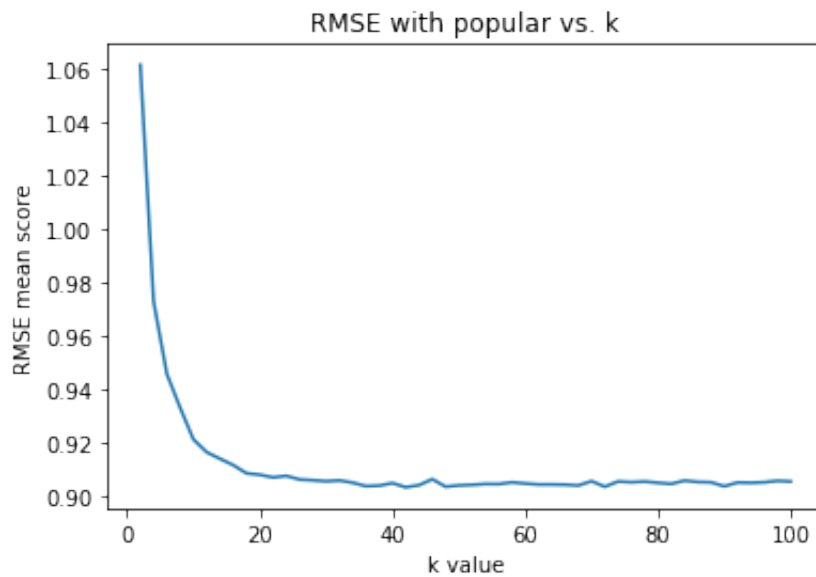
trim_name = ['popular', 'unpopular', 'high var']

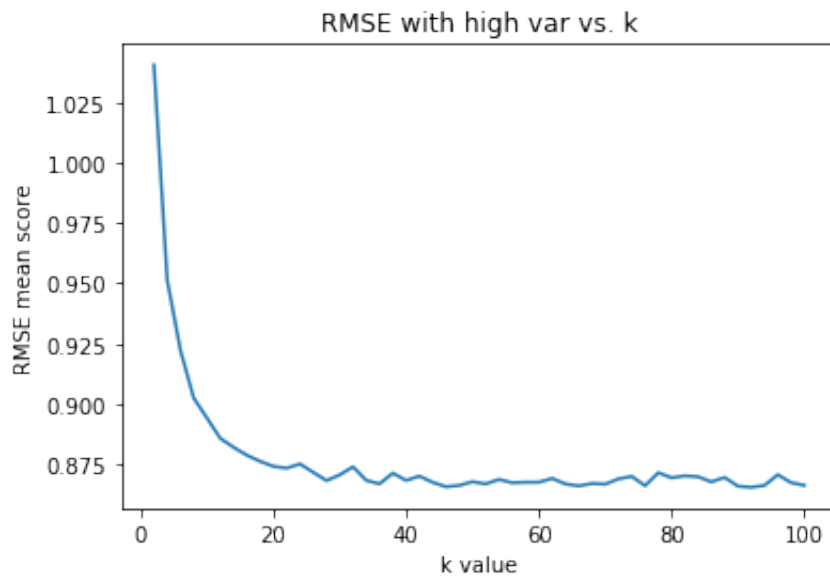
for i in range(3):
    plt.figure()
    plt.plot(range(2, 102, 2), knn_trim_rmse_list[i])
    plt.title('RMSE with '+trim_name[i]+' vs. k')
    plt.xlabel('k value')
    plt.ylabel('RMSE mean score')

    print("min RMSE for " + trim_name[i] + " trimming = ", np.min(knn_t
rim_rmse_list[i]))

```

```
min RMSE for popular trimming = 0.903268800933  
min RMSE for unpopular trimming = 0.933316858228  
min RMSE for high var trimming = 0.865605555229
```





## Question 12,13,14 Answer:

min RMSE for popular trimming = 0.903268800933 min RMSE for unpopular trimming = 0.933316858228

min RMSE for high var trimming = 0.865605555229

**Question 15: Plot the ROC curves for the k-NN collaborative filter designed in question 10 for threshold values [2.5, 3, 3.5, 4]. For the ROC plotting use the k found in question 11. For each of the plots, also report the area under the curve (AUC) value.**

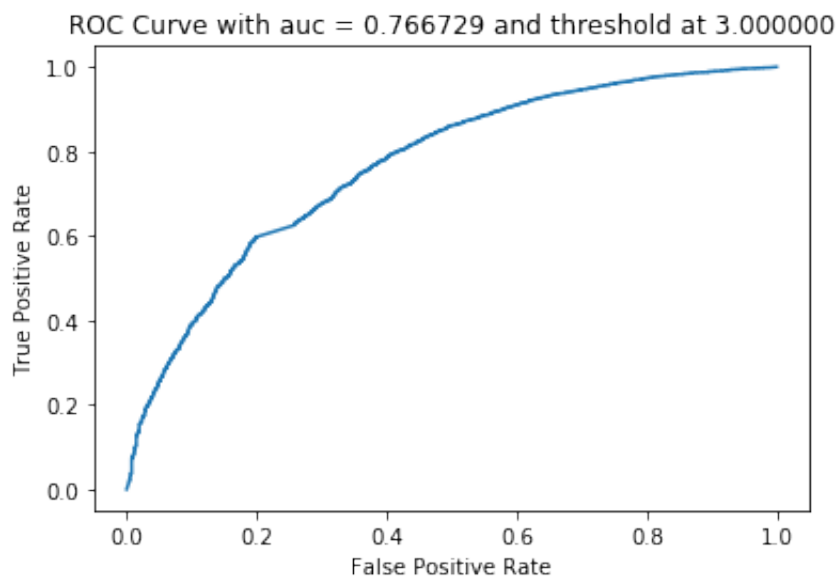
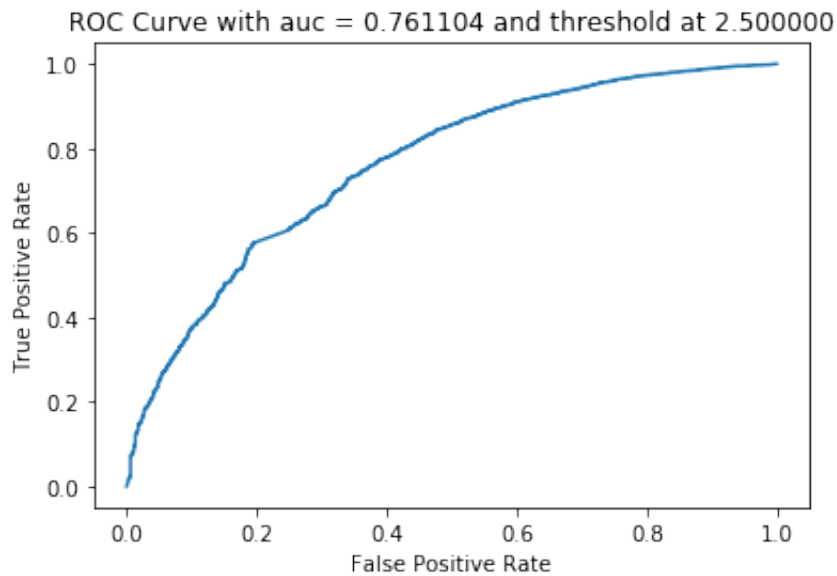
```
In [51]: from surprise.prediction_algorithms.knns import KNNWithMeans
from surprise.model_selection import KFold
from sklearn import metrics
```

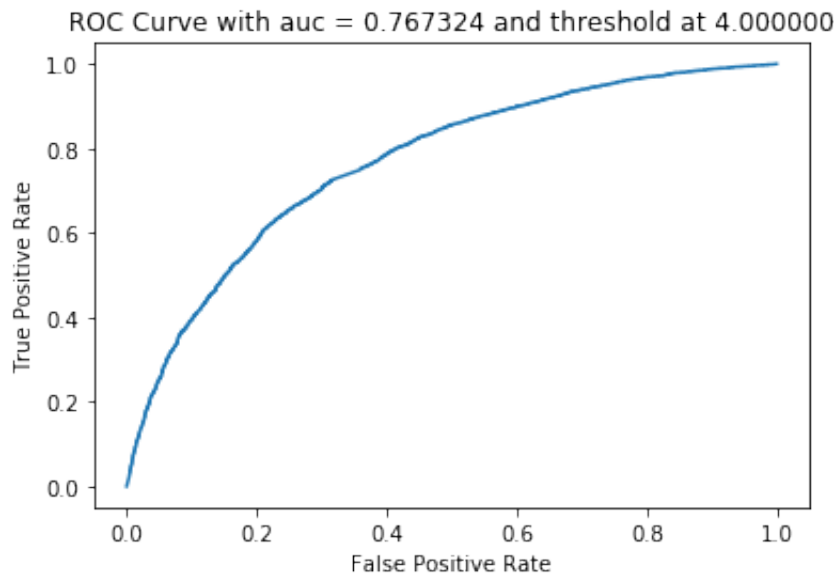
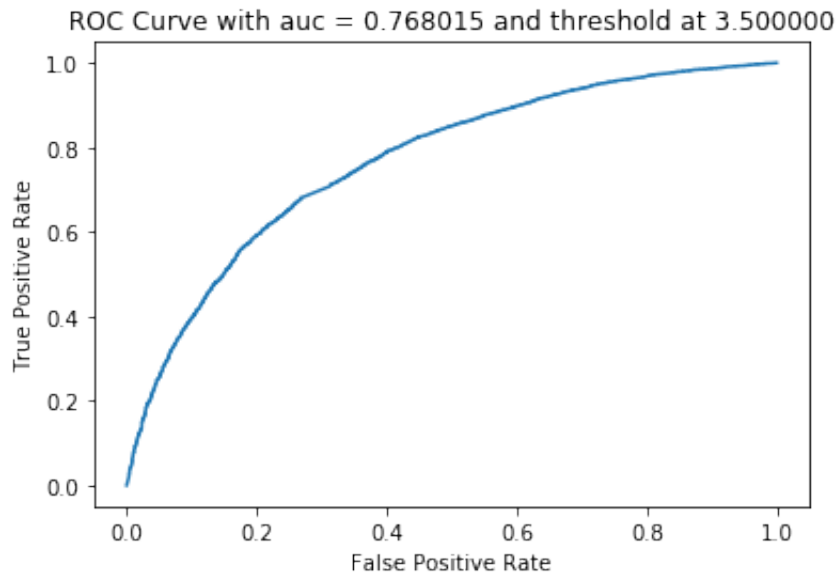
```
thresholds = [2.5, 3, 3.5, 4]
min_k = 24
kf = KFold(n_splits = 10)
sim_options = {'name': 'pearson'}
knn = KNNWithMeans(k = min_k, sim_options = sim_options)
for trainset, testset in kf.split(data):
    pass
knn.fit(trainset)
pred=knn.test(testset)
```

Computing the pearson similarity matrix...

Done computing similarity matrix.

```
In [52]: for i in thresholds:
          (rs,rs_pred) = ([],[])
          for j in pred:
              if j[2] >= i: rs.append(1)
              else: rs.append(0)
              rs_pred.append(j[3])
          fpr, tpr, _ = metrics.roc_curve(y_true = rs, y_score = rs_pred, pos_label=1)
          auc = metrics.roc_auc_score(y_true=rs, y_score=rs_pred)
          plt.figure()
          plt.plot(fpr,tpr)
          plt.xlabel('False Positive Rate')
          plt.ylabel('True Positive Rate')
          plt.title('ROC Curve with auc = %f and threshold at %f'%(auc,i))
```





## Question 15 Answer:

The auc value is shown in each ROC graph title, we can see that when the threshold is at 3.0, the auc score is highest.

**Question 16: Is the optimization problem given by equation 5 convex? Consider the optimization problem given by equation 5. For  $U$  fixed, formulate it as a least-squares problem.**



## Question 16 Answer:

If  $U$  is fixed, we can take the second derivative of equation 5 with  $V$ . Then we get a matrix that is semi-positive definite when means the problem is convex.

**Question 17: Design a NMF-based collaborative filter to predict the ratings of the movies in the MovieLens dataset and evaluate it's performance using 10-fold cross-validation. Sweep  $k$  (number of latent factors) from 2 to 50 in step sizes of 2, and for each  $k$  compute the average RMSE and average MAE obtained by averaging the RMSE and MAE across all 10 folds. Plot the average RMSE (Y-axis) against  $k$  (X-axis) and the average MAE (Y-axis) against  $k$  (X-axis). For solving this question, use the default value for the regularization parameter.**

```
In [53]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

reader = Reader(line_format='user item rating timestamp', sep=',', skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

(nmf_rmse, nmf_mae) = train_nmf(data)
```

```
using k = 2
using k = 4
using k = 6
using k = 8
using k = 10
using k = 12
using k = 14
using k = 16
using k = 18
using k = 20
using k = 22
using k = 24
using k = 26
using k = 28
using k = 30
using k = 32
using k = 34
using k = 36
using k = 38
using k = 40
using k = 42
using k = 44
using k = 46
using k = 48
using k = 50
k-fold validation finished!
```

```
In [54]: import matplotlib.pyplot as plt

plt.plot(range(2, 52, 2),nmf_rmse)
plt.title('RMSE vs. k')
plt.xlabel('k value')
plt.ylabel('RMSE mean score')

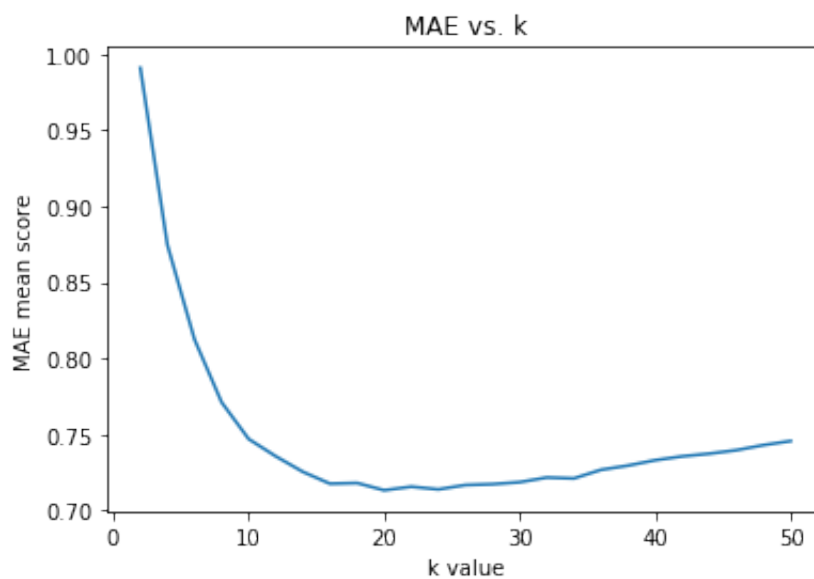
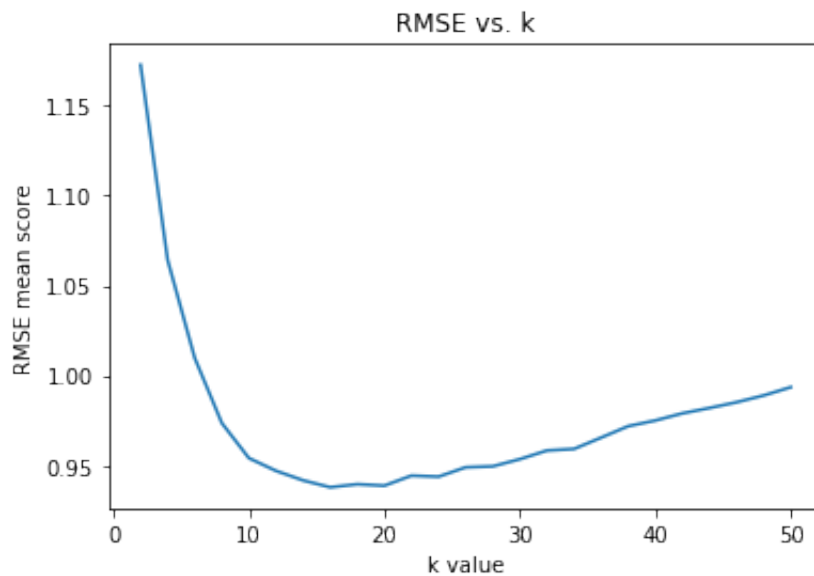
plt.figure()

plt.plot(range(2, 52, 2),nmf_mae)
plt.title('MAE vs. k')
plt.xlabel('k value')
plt.ylabel('MAE mean score')

print("min RMSE = %f when k = %d" % (np.min(nmf_rmse),np.argmin(nmf_rmse)*2+2))
print("min MAE = %f when k = %d" % (np.min(nmf_mae),np.argmin(nmf_mae)*2+2))
```

min RMSE = 0.937829 when k = 16

min MAE = 0.713261 when k = 20



**Question 18: Use the plot from question 17, to find the optimal number of latent factors. Optimal number of latent factors is the value of k that gives the minimum average RMSE or the minimum average MAE. Please report the minimum average RMSE and MAE. Is the optimal number of latent factors same as the number of movie genres?**

## Question 18 Answer:

min RMSE = 0.937829 when k = 16 min MAE = 0.713261 when k = 20

we can choose min K to be 20. There are 19 genres in the data, the latent factors are pretty good classifications.

## Question 19,20,21, run nmf with trimming

```
In [55]: # train section

import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

%load_ext autoreload
%autoreload 2

df = readData()
num_movie = 164979
R = getRatingMatrix(df,num_movie)

reader = Reader(line_format='user item rating timestamp', sep=',',skip
_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=re
ader)

nmf_trim_rmse_list = train_trim_nmf(data,R)
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
using k = 2
RMSE: 1.1625
RMSE: 1.1715
RMSE: 1.1623
RMSE: 1.1731
RMSE: 1.1880
RMSE: 1.1977
RMSE: 1.1820
RMSE: 1.1795
RMSE: 1.1712
RMSE: 1.1755
```

```
RMSE: 1.1680
RMSE: 1.1440
RMSE: 1.1881
RMSE: 1.1674
RMSE: 1.1971
RMSE: 1.1532
RMSE: 1.1857
RMSE: 1.1436
RMSE: 1.1785
RMSE: 1.2186
RMSE: 1.1997
RMSE: 1.1829
RMSE: 1.1824
RMSE: 1.1702
RMSE: 1.1759
RMSE: 1.1740
RMSE: 1.2199
RMSE: 1.1726
RMSE: 1.1705
RMSE: 1.1883
using k = 4
RMSE: 1.0555
RMSE: 1.0979
RMSE: 1.0419
RMSE: 1.0632
RMSE: 1.0970
RMSE: 1.0261
RMSE: 1.0412
RMSE: 1.0857
RMSE: 1.0090
RMSE: 1.0637
RMSE: 1.0846
RMSE: 1.0288
RMSE: 1.0711
RMSE: 1.0800
RMSE: 1.0342
RMSE: 1.0551
RMSE: 1.0886
RMSE: 1.0387
RMSE: 1.0599
RMSE: 1.0581
RMSE: 1.1242
RMSE: 1.0568
RMSE: 1.1004
RMSE: 1.0263
RMSE: 1.0522
RMSE: 1.0565
RMSE: 1.0994
RMSE: 1.0699
RMSE: 1.0888
```

```
RMSE: 1.0396
using k = 6
RMSE: 0.9940
RMSE: 1.0408
RMSE: 0.9841
RMSE: 1.0029
RMSE: 1.0301
RMSE: 0.9714
RMSE: 0.9795
RMSE: 1.0306
RMSE: 0.9748
RMSE: 0.9903
RMSE: 1.0149
RMSE: 1.0208
RMSE: 0.9943
RMSE: 1.0210
RMSE: 0.9779
RMSE: 1.0085
RMSE: 1.0348
RMSE: 0.9597
RMSE: 0.9851
RMSE: 1.0242
RMSE: 0.9270
RMSE: 1.0066
RMSE: 1.0109
RMSE: 0.9856
RMSE: 1.0043
RMSE: 1.0347
RMSE: 0.9651
RMSE: 0.9803
RMSE: 1.0412
RMSE: 0.9219
using k = 8
RMSE: 0.9557
RMSE: 0.9956
RMSE: 0.9074
RMSE: 0.9556
RMSE: 1.0007
RMSE: 0.8664
RMSE: 0.9739
RMSE: 1.0140
RMSE: 0.9032
RMSE: 0.9726
RMSE: 0.9899
RMSE: 0.9671
RMSE: 0.9546
RMSE: 0.9946
RMSE: 0.9894
RMSE: 0.9622
RMSE: 0.9982
```

```
RMSE: 0.9629
RMSE: 0.9524
RMSE: 0.9941
RMSE: 0.9563
RMSE: 0.9638
RMSE: 0.9869
RMSE: 0.9170
RMSE: 0.9692
RMSE: 0.9929
RMSE: 0.9056
RMSE: 0.9587
RMSE: 0.9933
RMSE: 0.9409
using k = 10
RMSE: 0.9461
RMSE: 0.9847
RMSE: 0.9130
RMSE: 0.9185
RMSE: 0.9670
RMSE: 0.8981
RMSE: 0.9354
RMSE: 0.9866
RMSE: 0.9664
RMSE: 0.9682
RMSE: 0.9930
RMSE: 0.8938
RMSE: 0.9507
RMSE: 0.9713
RMSE: 0.8953
RMSE: 0.9374
RMSE: 1.0220
RMSE: 0.8992
RMSE: 0.9365
RMSE: 0.9717
RMSE: 0.8917
RMSE: 0.9339
RMSE: 0.9805
RMSE: 0.9035
RMSE: 0.9329
RMSE: 0.9585
RMSE: 0.9068
RMSE: 0.9283
RMSE: 0.9875
RMSE: 0.8700
using k = 12
RMSE: 0.9473
RMSE: 0.9676
RMSE: 0.8849
RMSE: 0.9350
RMSE: 0.9760
```

```
RMSE: 0.8373
RMSE: 0.9357
RMSE: 0.9745
RMSE: 0.8305
RMSE: 0.9162
RMSE: 0.9652
RMSE: 0.8759
RMSE: 0.9138
RMSE: 0.9796
RMSE: 0.9125
RMSE: 0.9179
RMSE: 0.9529
RMSE: 0.8682
RMSE: 0.9256
RMSE: 0.9713
RMSE: 0.9144
RMSE: 0.9322
RMSE: 0.9665
RMSE: 0.9058
RMSE: 0.9169
RMSE: 0.9611
RMSE: 0.8733
RMSE: 0.9341
RMSE: 0.9786
RMSE: 0.8994
using k = 14
RMSE: 0.9335
RMSE: 0.9800
RMSE: 0.9232
RMSE: 0.9221
RMSE: 0.9489
RMSE: 0.8421
RMSE: 0.9225
RMSE: 0.9535
RMSE: 0.8781
RMSE: 0.9301
RMSE: 0.9777
RMSE: 0.9146
RMSE: 0.9235
RMSE: 0.9427
RMSE: 0.8510
RMSE: 0.9304
RMSE: 0.9556
RMSE: 0.9222
RMSE: 0.8950
RMSE: 0.9616
RMSE: 0.8539
RMSE: 0.9178
RMSE: 0.9624
RMSE: 0.8426
```



```
RMSE: 0.9199
RMSE: 0.9774
RMSE: 0.8656
RMSE: 0.9307
RMSE: 0.9588
RMSE: 0.9574
using k = 16
RMSE: 0.9270
RMSE: 0.9359
RMSE: 0.8234
RMSE: 0.9178
RMSE: 0.9659
RMSE: 0.8715
RMSE: 0.9153
RMSE: 0.9685
RMSE: 0.8982
RMSE: 0.9162
RMSE: 0.9624
RMSE: 0.9019
RMSE: 0.9438
RMSE: 0.9597
RMSE: 0.8586
RMSE: 0.9177
RMSE: 0.9539
RMSE: 0.9059
RMSE: 0.9280
RMSE: 0.9616
RMSE: 0.9370
RMSE: 0.9226
RMSE: 0.9449
RMSE: 0.9016
RMSE: 0.9152
RMSE: 0.9713
RMSE: 0.8351
RMSE: 0.9173
RMSE: 0.9681
RMSE: 0.9269
using k = 18
RMSE: 0.9073
RMSE: 0.9621
RMSE: 0.7772
RMSE: 0.9180
RMSE: 0.9621
RMSE: 0.8917
RMSE: 0.9274
RMSE: 0.9553
RMSE: 0.9368
RMSE: 0.9256
RMSE: 0.9573
RMSE: 0.8656
```

```
RMSE: 0.9014
RMSE: 0.9698
RMSE: 0.8792
RMSE: 0.9196
RMSE: 0.9615
RMSE: 0.8403
RMSE: 0.9291
RMSE: 0.9622
RMSE: 0.8990
RMSE: 0.9271
RMSE: 0.9580
RMSE: 0.9086
RMSE: 0.9236
RMSE: 0.9497
RMSE: 0.9090
RMSE: 0.9330
RMSE: 0.9725
RMSE: 0.9422
using k = 20
RMSE: 0.9201
RMSE: 0.9478
RMSE: 0.8535
RMSE: 0.9235
RMSE: 0.9636
RMSE: 0.8992
RMSE: 0.9250
RMSE: 0.9847
RMSE: 0.8526
RMSE: 0.9443
RMSE: 0.9738
RMSE: 0.9183
RMSE: 0.9319
RMSE: 0.9672
RMSE: 0.8698
RMSE: 0.9241
RMSE: 0.9479
RMSE: 0.8660
RMSE: 0.9240
RMSE: 0.9577
RMSE: 0.9061
RMSE: 0.9283
RMSE: 0.9324
RMSE: 0.9333
RMSE: 0.9187
RMSE: 0.9647
RMSE: 0.8916
RMSE: 0.9187
RMSE: 0.9662
RMSE: 0.8812
using k = 22
```

RMSE: 0.9424  
RMSE: 0.9779  
RMSE: 0.9712  
RMSE: 0.9163  
RMSE: 0.9823  
RMSE: 0.8991  
RMSE: 0.9314  
RMSE: 0.9618  
RMSE: 0.8349  
RMSE: 0.9150  
RMSE: 0.9571  
RMSE: 0.8169  
RMSE: 0.9316  
RMSE: 0.9648  
RMSE: 0.8898  
RMSE: 0.9270  
RMSE: 0.9504  
RMSE: 0.8797  
RMSE: 0.9065  
RMSE: 0.9365  
RMSE: 0.8899  
RMSE: 0.9124  
RMSE: 0.9572  
RMSE: 0.8633  
RMSE: 0.9406  
RMSE: 0.9629  
RMSE: 0.9029  
RMSE: 0.9613  
RMSE: 0.9752  
RMSE: 0.9222  
using k = 24  
RMSE: 0.9299  
RMSE: 0.9750  
RMSE: 0.8449  
RMSE: 0.9192  
RMSE: 0.9605  
RMSE: 0.9003  
RMSE: 0.9176  
RMSE: 0.9587  
RMSE: 0.8598  
RMSE: 0.9345  
RMSE: 0.9725  
RMSE: 0.9298  
RMSE: 0.9555  
RMSE: 0.9709  
RMSE: 0.9101  
RMSE: 0.9297  
RMSE: 0.9463  
RMSE: 0.8679  
RMSE: 0.9217

```
RMSE: 0.9552
RMSE: 0.8455
RMSE: 0.9439
RMSE: 0.9704
RMSE: 0.8940
RMSE: 0.9176
RMSE: 0.9526
RMSE: 0.8916
RMSE: 0.9267
RMSE: 0.9742
RMSE: 1.0064
using k = 26
RMSE: 0.9174
RMSE: 0.9649
RMSE: 0.8785
RMSE: 0.9532
RMSE: 0.9499
RMSE: 0.9277
RMSE: 0.9240
RMSE: 0.9608
RMSE: 0.8738
RMSE: 0.9200
RMSE: 0.9846
RMSE: 0.8826
RMSE: 0.9344
RMSE: 0.9537
RMSE: 0.9050
RMSE: 0.9423
RMSE: 0.9673
RMSE: 0.8993
RMSE: 0.9235
RMSE: 0.9846
RMSE: 0.8650
RMSE: 0.9438
RMSE: 0.9758
RMSE: 0.8738
RMSE: 0.9468
RMSE: 0.9711
RMSE: 0.9489
RMSE: 0.9601
RMSE: 0.9495
RMSE: 0.9384
using k = 28
RMSE: 0.9408
RMSE: 0.9758
RMSE: 0.9258
RMSE: 0.9452
RMSE: 0.9476
RMSE: 0.9302
RMSE: 0.9319
```

```
RMSE: 0.9631
RMSE: 0.9117
RMSE: 0.9377
RMSE: 0.9417
RMSE: 0.8629
RMSE: 0.9509
RMSE: 0.9716
RMSE: 0.8731
RMSE: 0.9442
RMSE: 0.9744
RMSE: 0.9283
RMSE: 0.9390
RMSE: 0.9857
RMSE: 0.8679
RMSE: 0.9249
RMSE: 0.9827
RMSE: 0.9824
RMSE: 0.9367
RMSE: 0.9601
RMSE: 0.8338
RMSE: 0.9373
RMSE: 0.9643
RMSE: 0.9190
using k = 30
RMSE: 0.9319
RMSE: 0.9708
RMSE: 0.9070
RMSE: 0.9431
RMSE: 0.9829
RMSE: 0.9388
RMSE: 0.9388
RMSE: 0.9741
RMSE: 0.8454
RMSE: 0.9460
RMSE: 0.9684
RMSE: 0.9032
RMSE: 0.9382
RMSE: 0.9820
RMSE: 0.8927
RMSE: 0.9559
RMSE: 0.9472
RMSE: 0.9094
RMSE: 0.9488
RMSE: 0.9881
RMSE: 0.9309
RMSE: 0.9560
RMSE: 0.9602
RMSE: 0.8745
RMSE: 0.9314
RMSE: 0.9836
```

```
RMSE: 0.8639
RMSE: 0.9491
RMSE: 0.9789
RMSE: 0.9894
using k = 32
RMSE: 0.9480
RMSE: 0.9620
RMSE: 0.9535
RMSE: 0.9586
RMSE: 0.9838
RMSE: 0.9252
RMSE: 0.9419
RMSE: 0.9812
RMSE: 0.9100
RMSE: 0.9545
RMSE: 0.9895
RMSE: 0.9424
RMSE: 0.9331
RMSE: 0.9795
RMSE: 0.8836
RMSE: 0.9359
RMSE: 0.9798
RMSE: 0.8467
RMSE: 0.9401
RMSE: 0.9616
RMSE: 0.9387
RMSE: 0.9560
RMSE: 0.9742
RMSE: 0.9189
RMSE: 0.9441
RMSE: 0.9780
RMSE: 0.8848
RMSE: 0.9463
RMSE: 0.9708
RMSE: 0.9328
using k = 34
RMSE: 0.9660
RMSE: 0.9694
RMSE: 0.9181
RMSE: 0.9517
RMSE: 0.9789
RMSE: 0.9098
RMSE: 0.9610
RMSE: 0.9900
RMSE: 0.9719
RMSE: 0.9691
RMSE: 0.9829
RMSE: 0.9058
RMSE: 0.9372
RMSE: 0.9846
```

```
RMSE: 0.8982
RMSE: 0.9466
RMSE: 0.9682
RMSE: 0.8635
RMSE: 0.9469
RMSE: 0.9616
RMSE: 0.9077
RMSE: 0.9459
RMSE: 0.9782
RMSE: 0.8681
RMSE: 0.9418
RMSE: 1.0038
RMSE: 0.9427
RMSE: 0.9515
RMSE: 0.9594
RMSE: 0.9088
using k = 36
RMSE: 0.9496
RMSE: 0.9706
RMSE: 0.8988
RMSE: 0.9416
RMSE: 0.9910
RMSE: 0.9364
RMSE: 0.9504
RMSE: 0.9757
RMSE: 0.9038
RMSE: 0.9635
RMSE: 0.9706
RMSE: 0.8825
RMSE: 0.9575
RMSE: 0.9838
RMSE: 0.9892
RMSE: 0.9554
RMSE: 0.9748
RMSE: 0.9350
RMSE: 0.9519
RMSE: 0.9941
RMSE: 0.9038
RMSE: 0.9547
RMSE: 0.9878
RMSE: 0.9095
RMSE: 0.9755
RMSE: 0.9806
RMSE: 0.9378
RMSE: 0.9416
RMSE: 0.9634
RMSE: 0.9189
using k = 38
RMSE: 0.9661
RMSE: 1.0059
```

```
RMSE: 0.9990
RMSE: 0.9521
RMSE: 0.9851
RMSE: 0.8882
RMSE: 0.9541
RMSE: 0.9808
RMSE: 0.9113
RMSE: 0.9442
RMSE: 0.9577
RMSE: 0.9460
RMSE: 0.9500
RMSE: 0.9834
RMSE: 0.8824
RMSE: 0.9678
RMSE: 1.0074
RMSE: 0.8885
RMSE: 0.9594
RMSE: 0.9764
RMSE: 0.9362
RMSE: 0.9593
RMSE: 0.9791
RMSE: 0.9525
RMSE: 0.9613
RMSE: 0.9910
RMSE: 0.9499
RMSE: 0.9966
RMSE: 0.9837
RMSE: 0.9654
using k = 40
RMSE: 0.9504
RMSE: 0.9982
RMSE: 0.8299
RMSE: 0.9660
RMSE: 0.9899
RMSE: 0.9606
RMSE: 0.9813
RMSE: 0.9949
RMSE: 0.9599
RMSE: 0.9557
RMSE: 0.9853
RMSE: 0.8985
RMSE: 0.9488
RMSE: 0.9871
RMSE: 0.9427
RMSE: 0.9653
RMSE: 0.9822
RMSE: 0.9613
RMSE: 0.9832
RMSE: 0.9671
RMSE: 0.9715
```



```
RMSE: 0.9751
RMSE: 1.0078
RMSE: 0.9331
RMSE: 0.9502
RMSE: 0.9761
RMSE: 0.9235
RMSE: 0.9618
RMSE: 0.9824
RMSE: 0.9508
using k = 42
RMSE: 0.9753
RMSE: 0.9905
RMSE: 1.0184
RMSE: 0.9725
RMSE: 0.9829
RMSE: 0.9666
RMSE: 0.9654
RMSE: 0.9885
RMSE: 0.9153
RMSE: 0.9628
RMSE: 1.0205
RMSE: 0.9059
RMSE: 0.9838
RMSE: 0.9838
RMSE: 0.9940
RMSE: 0.9825
RMSE: 0.9999
RMSE: 0.9171
RMSE: 0.9645
RMSE: 0.9861
RMSE: 0.9356
RMSE: 0.9500
RMSE: 1.0165
RMSE: 0.8695
RMSE: 0.9568
RMSE: 0.9916
RMSE: 0.9269
RMSE: 0.9763
RMSE: 0.9872
RMSE: 0.9618
using k = 44
RMSE: 0.9648
RMSE: 1.0044
RMSE: 0.9133
RMSE: 0.9708
RMSE: 1.0054
RMSE: 0.9078
RMSE: 0.9837
RMSE: 1.0160
RMSE: 0.9176
```

```
RMSE: 0.9863
RMSE: 0.9728
RMSE: 0.9753
RMSE: 0.9643
RMSE: 0.9825
RMSE: 0.9145
RMSE: 0.9757
RMSE: 0.9808
RMSE: 0.9498
RMSE: 0.9763
RMSE: 0.9844
RMSE: 0.9535
RMSE: 0.9670
RMSE: 0.9998
RMSE: 0.9322
RMSE: 0.9570
RMSE: 0.9926
RMSE: 0.9736
RMSE: 0.9774
RMSE: 0.9863
RMSE: 0.9479
using k = 46
RMSE: 0.9960
RMSE: 0.9975
RMSE: 0.9985
RMSE: 0.9721
RMSE: 1.0077
RMSE: 0.9328
RMSE: 0.9707
RMSE: 0.9887
RMSE: 0.9703
RMSE: 0.9683
RMSE: 0.9923
RMSE: 0.9516
RMSE: 0.9941
RMSE: 1.0178
RMSE: 0.9066
RMSE: 0.9732
RMSE: 0.9901
RMSE: 1.0023
RMSE: 0.9597
RMSE: 1.0178
RMSE: 0.8745
RMSE: 0.9869
RMSE: 1.0059
RMSE: 0.9944
RMSE: 0.9697
RMSE: 0.9710
RMSE: 0.9487
RMSE: 0.9625
```

```
RMSE: 0.9808
RMSE: 0.8692
using k = 48
RMSE: 0.9917
RMSE: 1.0025
RMSE: 0.9633
RMSE: 0.9841
RMSE: 1.0140
RMSE: 0.9426
RMSE: 0.9815
RMSE: 1.0010
RMSE: 0.9971
RMSE: 0.9861
RMSE: 0.9775
RMSE: 0.9408
RMSE: 0.9881
RMSE: 0.9996
RMSE: 0.9832
RMSE: 0.9812
RMSE: 1.0237
RMSE: 0.9979
RMSE: 0.9920
RMSE: 0.9928
RMSE: 0.9992
RMSE: 0.9723
RMSE: 0.9824
RMSE: 0.8810
RMSE: 0.9769
RMSE: 0.9922
RMSE: 0.8901
RMSE: 0.9681
RMSE: 1.0058
RMSE: 0.8916
using k = 50
RMSE: 1.0001
RMSE: 0.9945
RMSE: 0.9829
RMSE: 0.9792
RMSE: 1.0195
RMSE: 0.9210
RMSE: 0.9727
RMSE: 0.9940
RMSE: 0.9700
RMSE: 0.9813
RMSE: 0.9914
RMSE: 0.9501
RMSE: 0.9929
RMSE: 0.9971
RMSE: 0.9974
RMSE: 0.9931
```

```
RMSE: 1.0127
RMSE: 0.9287
RMSE: 0.9843
RMSE: 1.0089
RMSE: 0.9506
RMSE: 0.9812
RMSE: 0.9954
RMSE: 0.9090
RMSE: 0.9830
RMSE: 1.0021
RMSE: 0.9920
RMSE: 0.9898
RMSE: 1.0028
RMSE: 1.0207
NMF with trim is finished!!
```

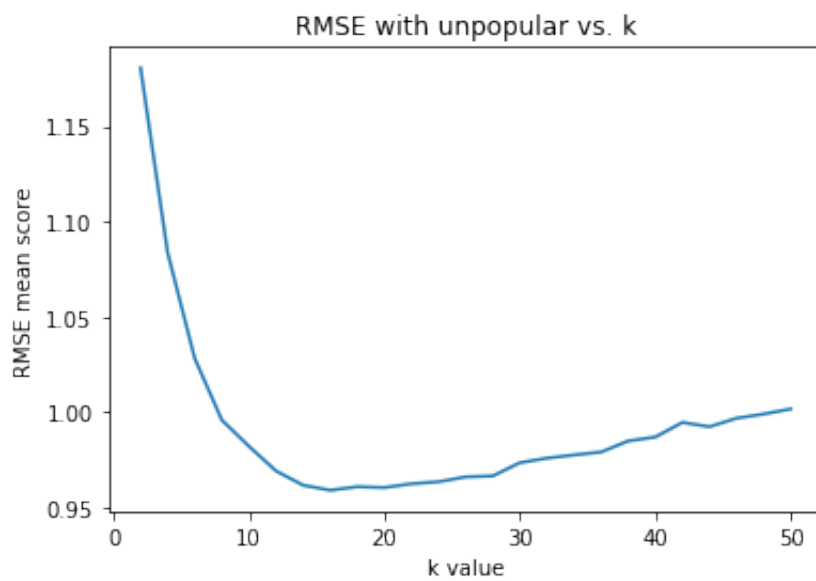
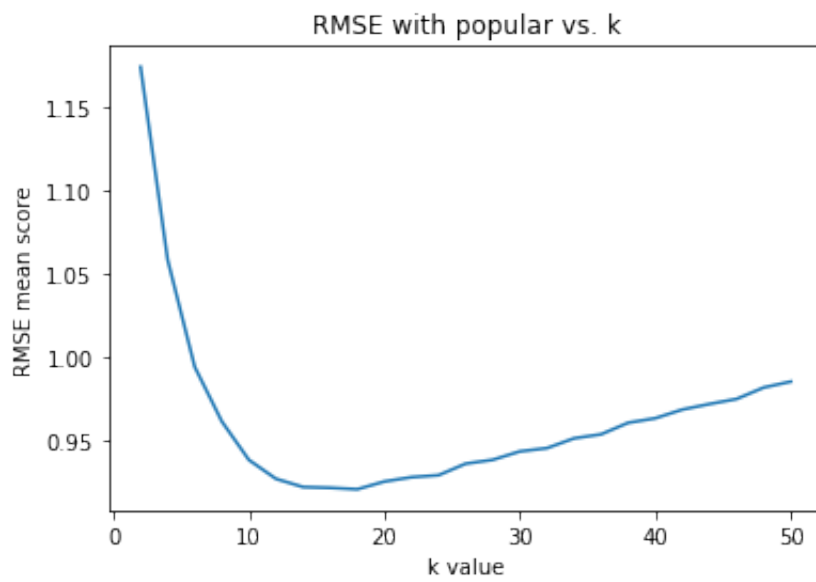
In [56]: *# plot section*

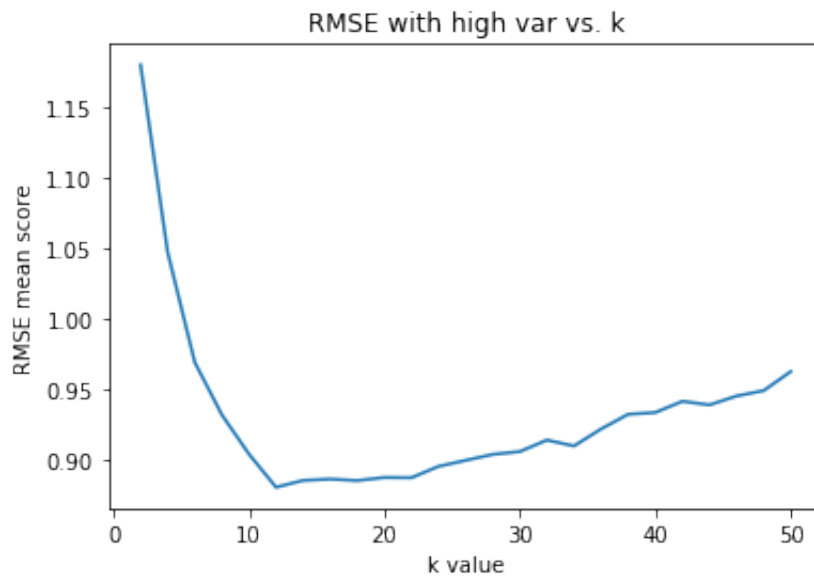
```
import matplotlib.pyplot as plt

trim_name = ['popular','unpopular','high var']

for i in range(3):
    plt.figure()
    plt.plot(range(2, 52, 2),nmf_trim_rmse_list[i])
    plt.title('RMSE with '+trim_name[i]+' vs. k')
    plt.xlabel('k value')
    plt.ylabel('RMSE mean score')
    print("min RMSE for " + trim_name[i]+ " trimming = ", np.min(nmf_t
rim_rmse_list[i]))
```

```
min RMSE for popular trimming = 0.921220686066
min RMSE for unpopular trimming = 0.959210950026
min RMSE for high var trimming = 0.880233016554
```





## Question 19,20,21 Answer:

min RMSE for popular trimming = 0.921220686066 min RMSE for unpopular trimming = 0.959210950026

min RMSE for high var trimming = 0.880233016554

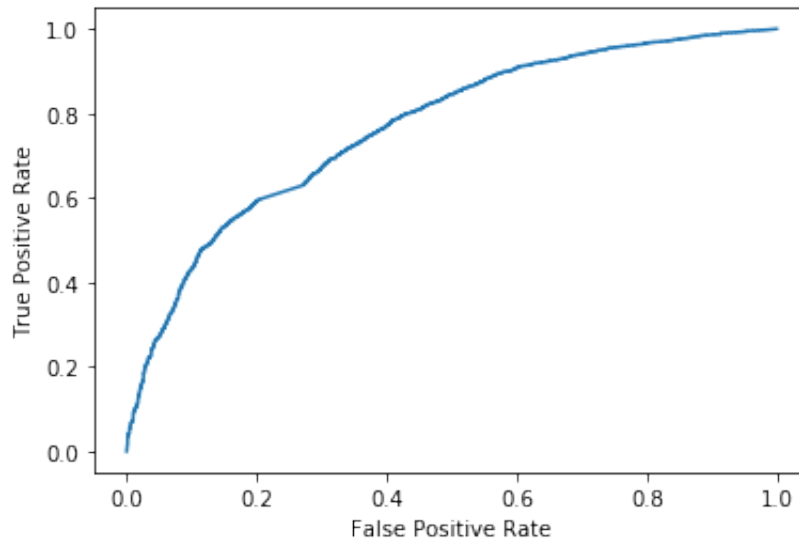
**Question 22: Plot the ROC curves for the NMF-based collaborative filter designed in question 17 for threshold values [2.5,3,3.5,4]. For the ROC plot- ting use the optimal number of latent factors found in question 18. For each of the plots, also report the area under the curve (AUC) value.**

```
In [57]: from surprise.prediction_algorithms.knns import KNNWithMeans
from surprise.model_selection import KFold
from sklearn import metrics
from surprise.prediction_algorithms.matrix_factorization import NMF, SVD

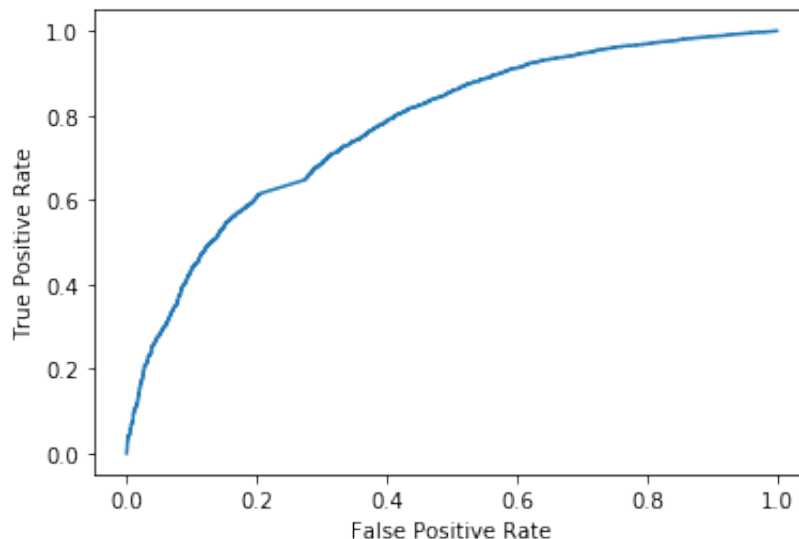
thresholds = [2.5, 3, 3.5, 4]
nmf_min_k = 20
kf = KFold(n_splits = 10)
nmf = NMF(n_factors = nmf_min_k)
for trainset, testset in kf.split(data):
    pass
nmf.fit(trainset)
pred=nmf.test(testset)
```

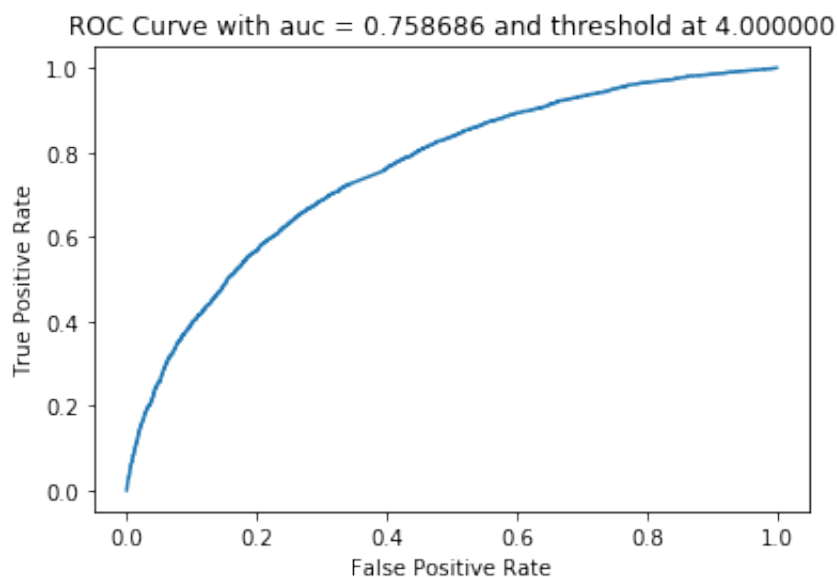
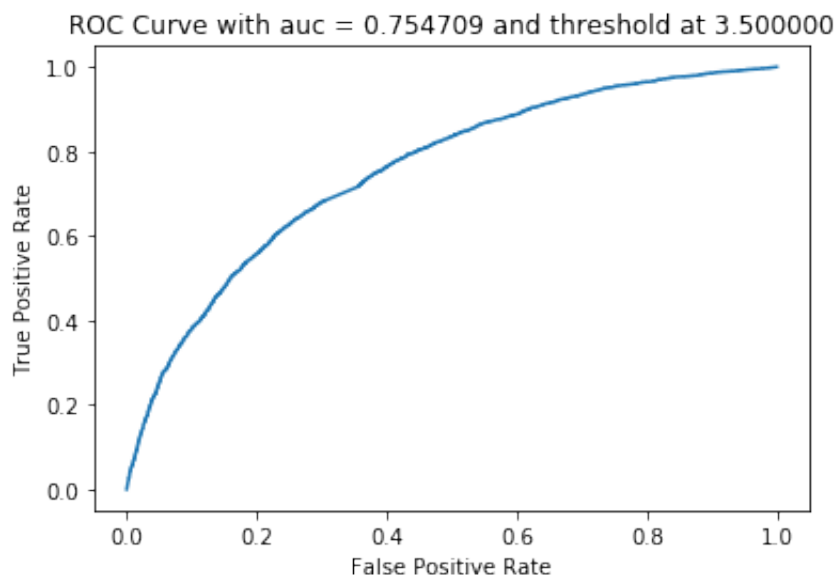
```
In [58]: for i in thresholds:
          (rs,rs_pred) = ([],[])
          for j in pred:
              if j[2] >= i: rs.append(1)
              else: rs.append(0)
              rs_pred.append(j[3])
          fpr, tpr, _ = metrics.roc_curve(y_true = rs, y_score = rs_pred, pos_label=1)
          auc = metrics.roc_auc_score(y_true=rs, y_score=rs_pred)
          plt.figure()
          plt.plot(fpr,tpr)
          plt.title('ROC Curve with auc = %f and threshold at %f'%(auc,i))
          plt.xlabel('False Positive Rate')
          plt.ylabel('True Positive Rate')
```

ROC Curve with auc = 0.766565 and threshold at 2.500000



ROC Curve with auc = 0.774405 and threshold at 3.000000





## Question 22 Answer:

The auc value is shown in each ROC graph title, we can see that when the threshold is at 3.0, the auc score is highest.



**Question 23: Perform Non-negative matrix factorization on the ratings matrix  $R$  to obtain the factor matrices  $U$  and  $V$ , where  $U$  represents the user-latent factors interaction and  $V$  represents the movie-latent factors interaction (use  $k = 20$ ). For each column of  $V$ , sort the movies in descending order and report the genres of the top 10 movies. Do the top 10 movies belong to a particular or a small collection of genre? Is there a connection between the latent factors and the movie genres?**

```
In [59]: from surprise import NMF
nmf = NMF(n_factors=20)
nmf.fit(data.build_full_trainset())
mov = pd.read_csv('ml-latest-small/movies.csv', names=['mid', 'title',
'genres'], header=0)

V = nmf.qi

top_num = 10
top_mov_ids = []
for i in range(20):
    col = np.argsort(V[:,i])
    top_mov_ids.append(col)
genres = mov['genres']
```

```
In [60]: for i in range(20):
print("genres for col %d" % i)
for j in range(10):
    print ("-"*20)
    print (genres[top_mov_ids[i][j]])
    print ("-"*20)
```

```
genres for col 0
-----
Horror|Mystery|Thriller
-----
-----
Comedy
-----
-----
Action|Adventure|Children
-----
-----
```

Adventure|Children|Fantasy  
-----  
-----  
Comedy|Drama  
-----  
-----  
Action|Adventure|Animation|Comedy  
-----  
-----  
Action|Comedy|Crime|Thriller  
-----  
-----  
Action|Adventure|Sci-Fi|Thriller  
-----  
-----  
Drama|Romance  
-----  
-----  
Action|Crime|Drama  
-----  
genres for col 1  
-----  
Comedy  
-----  
-----  
Adventure|Children|Comedy|Drama  
-----  
-----  
Drama  
-----  
-----  
Action|Horror|Sci-Fi  
-----  
-----  
Adventure|Children|Drama  
-----  
-----  
Action|Adventure|Animation|Comedy  
-----  
-----  
Adventure|Animation|Children|Comedy  
-----  
-----  
Comedy|Romance  
-----  
-----  
Horror|Mystery|Thriller  
-----  
-----  
Action|Adventure|Drama|Sci-Fi

-----  
genres for col 2  
-----

Comedy|Drama|Thriller  
-----

-----  
Adventure|Children|Drama|Fantasy|IMAX  
-----

-----  
Comedy|Romance  
-----

-----  
Horror|Mystery|Thriller  
-----

-----  
Action|Crime|Fantasy|Sci-Fi|Thriller  
-----

-----  
Children|Comedy|Fantasy  
-----

-----  
Drama  
-----

-----  
Action|Animation|Comedy|Crime|Drama|Romance|Thriller  
-----

-----  
Documentary  
-----

-----  
Comedy|Musical  
-----

genres for col 3  
-----

-----  
Horror|Sci-Fi|Thriller  
-----

-----  
Drama|War  
-----

-----  
Action|Adventure|Drama|Sci-Fi  
-----

-----  
Documentary  
-----

-----  
Fantasy|Sci-Fi  
-----

-----  
Documentary  
-----

```
-----
-----
Horror|Mystery|Thriller
-----
-----
Comedy|Fantasy|Mystery
-----
-----
Adventure|Children|Comedy|Drama
-----
-----
Drama|War
-----
genres for col 4
-----
Western
-----
-----
Drama|War
-----
-----
Drama|Fantasy|Romance|Sci-Fi
-----
-----
Crime|Drama|Film-Noir
-----
-----
Action|Adventure|Animation|Comedy
-----
-----
Comedy|Romance
-----
-----
Drama
-----
-----
Action|Crime|Drama|Thriller
-----
-----
Comedy
-----
-----
Documentary
-----
genres for col 5
-----
Comedy|Musical|Sci-Fi
-----
-----
Adventure|Drama|Mystery|Thriller
```

```
-----  
-----  
Comedy | Romance  
-----  
-----  
Fantasy | Sci-Fi  
-----  
-----  
Drama  
-----  
-----  
Comedy | Drama | Romance  
-----  
-----  
Drama  
-----  
-----  
Comedy  
-----  
-----  
Action | Adventure | Drama | War  
-----  
-----  
Comedy | Drama | Fantasy | Romance  
-----  
genres for col 6  
-----  
Adventure | Children | Fantasy  
-----  
-----  
Action | Romance | War | Western  
-----  
-----  
Drama | Romance  
-----  
-----  
Action | Drama | Sci-Fi  
-----  
-----  
Children | Comedy  
-----  
-----  
Comedy | Documentary  
-----  
-----  
Horror | Thriller  
-----  
-----  
Action | Adventure | Comedy | Fantasy  
-----
```

-----  
Action  
-----

-----  
Action|Mystery|Sci-Fi|Thriller  
-----

genres for col 7  
-----

Adventure|Children|Fantasy  
-----

-----  
Horror|Mystery|Thriller  
-----

-----  
Action|Adventure|Sci-Fi|Thriller|IMAX  
-----

-----  
Documentary  
-----

-----  
Horror|Sci-Fi  
-----

-----  
Drama  
-----

-----  
Action|Adventure|Children  
-----

-----  
Action|Comedy  
-----

-----  
Action|Crime|Horror  
-----

-----  
Comedy|Drama  
-----

genres for col 8  
-----

-----  
Comedy|Drama  
-----

-----  
Horror|Mystery|Thriller  
-----

-----  
Action|Adventure|Sci-Fi|Thriller|IMAX  
-----

-----  
Action|Comedy  
-----

-----  
Comedy | Drama  
-----

-----  
Adventure | Comedy | Mystery | Romance  
-----

-----  
Comedy  
-----

-----  
Documentary  
-----

-----  
Comedy | Romance  
-----

-----  
Comedy | Fantasy  
-----

-----  
genres for col 9  
-----

-----  
Mystery | Thriller  
-----

-----  
Action | Adventure | Crime | Thriller  
-----

-----  
Comedy | Drama  
-----

-----  
Drama | War  
-----

-----  
Action | Comedy  
-----

-----  
Drama | Sci-Fi | War  
-----

-----  
Children | Drama  
-----

-----  
Action | Sci-Fi  
-----

-----  
Action | Crime | Drama  
-----

-----  
Action | Drama | War  
-----

-----  
genres for col 10  
-----

-----  
Action | Adventure | Comedy | Fantasy  
-----

-----  
Comedy | Mystery  
-----

-----  
Action | Adventure | Drama | War  
-----

-----  
Drama | Romance  
-----

-----  
Action | Drama | Romance  
-----

-----  
Action | Crime | Drama  
-----

-----  
Drama | Thriller  
-----

-----  
Crime | Drama  
-----

-----  
Comedy | Romance  
-----

-----  
Documentary | Musical  
-----

-----  
genres for col 11  
-----

-----  
Crime | Drama  
-----

-----  
Action | Adventure | Drama | Sci-Fi  
-----

-----  
Children | Comedy | Romance  
-----

-----  
Comedy  
-----

-----  
Crime | Drama | Thriller  
-----

-----  
Adventure | Fantasy | Sci-Fi  
-----  
-----



Action | Romance | War | Western  
-----  
-----  
Drama | Romance  
-----  
-----  
Comedy | Drama | Romance  
-----  
-----  
Crime | Drama  
-----  
genres for col 12  
-----  
Comedy | Romance  
-----  
-----  
Comedy | Drama | Thriller  
-----  
-----  
Children | Comedy  
-----  
-----  
Documentary | Musical  
-----  
-----  
Crime | Drama  
-----  
-----  
Horror | Sci-Fi | Thriller  
-----  
-----  
Action | Romance | War | Western  
-----  
-----  
Crime | Drama  
-----  
-----  
Adventure | Comedy | Mystery | Romance  
-----  
-----  
Comedy  
-----  
genres for col 13  
-----  
Horror | Mystery | Thriller  
-----  
-----  
Crime | Drama  
-----  
-----

Horror | Mystery | Thriller  
-----  
-----  
Action | Adventure | Drama | War  
-----  
-----  
Comedy | Musical  
-----  
-----  
Action | Adventure | Drama | Sci-Fi  
-----  
-----  
Drama  
-----  
-----  
Drama  
-----  
-----  
Action | Adventure | Comedy | Sci-Fi | Thriller  
-----  
-----  
Action | Adventure | Children  
-----  
genres for col 14  
-----  
Drama | War  
-----  
-----  
Children | Drama  
-----  
-----  
Action | Adventure | Crime | Thriller  
-----  
-----  
Action | Romance | War | Western  
-----  
-----  
Crime | Drama | Film-Noir  
-----  
-----  
Comedy | Drama  
-----  
-----  
Children | Comedy | Fantasy  
-----  
-----  
Children | Comedy  
-----  
-----  
Documentary

```
-----  
-----  
Documentary  
-----  
genres for col 15  
-----  
Action|Adventure|Animation|Comedy  
-----  
-----  
Adventure|Children|Fantasy  
-----  
-----  
Action|Mystery|Sci-Fi|Thriller  
-----  
-----  
Drama|Romance  
-----  
-----  
Action|Drama|Sci-Fi  
-----  
-----  
Comedy  
-----  
-----  
Horror|Mystery|Thriller  
-----  
-----  
Comedy|Drama|Fantasy|Romance  
-----  
-----  
Action|Adventure|Drama|War  
-----  
-----  
Drama|Thriller  
-----  
genres for col 16  
-----  
Action|Adventure|Drama|Sci-Fi  
-----  
-----  
Adventure|Fantasy|Sci-Fi  
-----  
-----  
Horror|Mystery|Thriller  
-----  
-----  
Fantasy|Sci-Fi  
-----  
-----  
Action|Drama|War
```

```
-----  
-----  
Comedy | Drama | Fantasy | Romance  
-----  
-----  
Comedy | Romance  
-----  
-----  
Comedy | Drama | Romance  
-----  
-----  
Adventure | Children | Fantasy  
-----  
-----  
Comedy | Drama | Thriller  
-----  
genres for col 17  
-----  
Drama | Romance  
-----  
-----  
Action | Drama | War  
-----  
-----  
Crime | Drama | Film-Noir  
-----  
-----  
Drama  
-----  
-----  
Comedy  
-----  
-----  
Action | Drama | Sci-Fi  
-----  
-----  
Action  
-----  
-----  
Drama | War  
-----  
-----  
Comedy | Documentary  
-----  
-----  
Documentary  
-----  
genres for col 18  
-----  
Adventure | Drama | Mystery | Thriller
```

```
-----  
-----  
Comedy | Drama | Thriller  
-----  
-----  
Adventure | Children | Drama | Fantasy | IMAX  
-----  
-----  
Drama | Horror  
-----  
-----  
Comedy | Romance  
-----  
-----  
Drama | Horror | Thriller  
-----  
-----  
Comedy | Musical  
-----  
-----  
Comedy | Romance  
-----  
-----  
Comedy  
-----  
-----  
Action | Animation | Comedy | Crime | Drama | Romance | Thriller  
-----  
genres for col 19  
-----  
Horror | Mystery | Thriller  
-----  
-----  
Comedy | Drama | Thriller  
-----  
-----  
Adventure | Romance | Thriller  
-----  
-----  
Drama  
-----  
-----  
Comedy | Documentary  
-----  
-----  
Adventure | Children | Drama | Fantasy | IMAX  
-----  
-----  
Comedy | Romance  
-----
```

```
-----  
Action|Comedy  
-----  
-----  
Comedy|Drama  
-----  
-----  
Horror|Mystery|Thriller  
-----
```

## Question 23 Answer:

From above output, we can see that the top 10 movies' genres are in a small group. For example, col 0, the genres are mostly action/adventure, which means the the first latent factor is focused on picking out action related movies. These latent factors will classify movies of similar genres into group.

**Question 24: Design a MF with bias collaborative filter to predict the ratings of the movies in the MovieLens dataset and evaluate it's performance using 10-fold cross-validation. Sweep k (number of latent factors) from 2 to 50 in step sizes of 2, and for each k compute the average RMSE and average MAE obtained by averaging the RMSE and MAE across all 10 folds. Plot the average RMSE (Y-axis) against k (X-axis) and the average MAE (Y-axis) against k (X-axis). For solving this question, use the default value for the regularization parameter.**

```
In [73]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

reader = Reader(line_format='user item rating timestamp', sep=',', skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

(svd_rmse,svd_mae) = train_svd(data)

using k = 2
using k = 4
using k = 6
using k = 8
using k = 10
using k = 12
using k = 14
using k = 16
using k = 18
using k = 20
using k = 22
using k = 24
using k = 26
using k = 28
using k = 30
using k = 32
using k = 34
using k = 36
using k = 38
using k = 40
using k = 42
using k = 44
using k = 46
using k = 48
using k = 50
k-fold validation finished!
```

```
In [74]: import matplotlib.pyplot as plt

plt.plot(range(2, 52, 2),svd_rmse)
plt.title('RMSE vs. k')
plt.xlabel('k value')
plt.ylabel('RMSE mean score')

plt.figure()

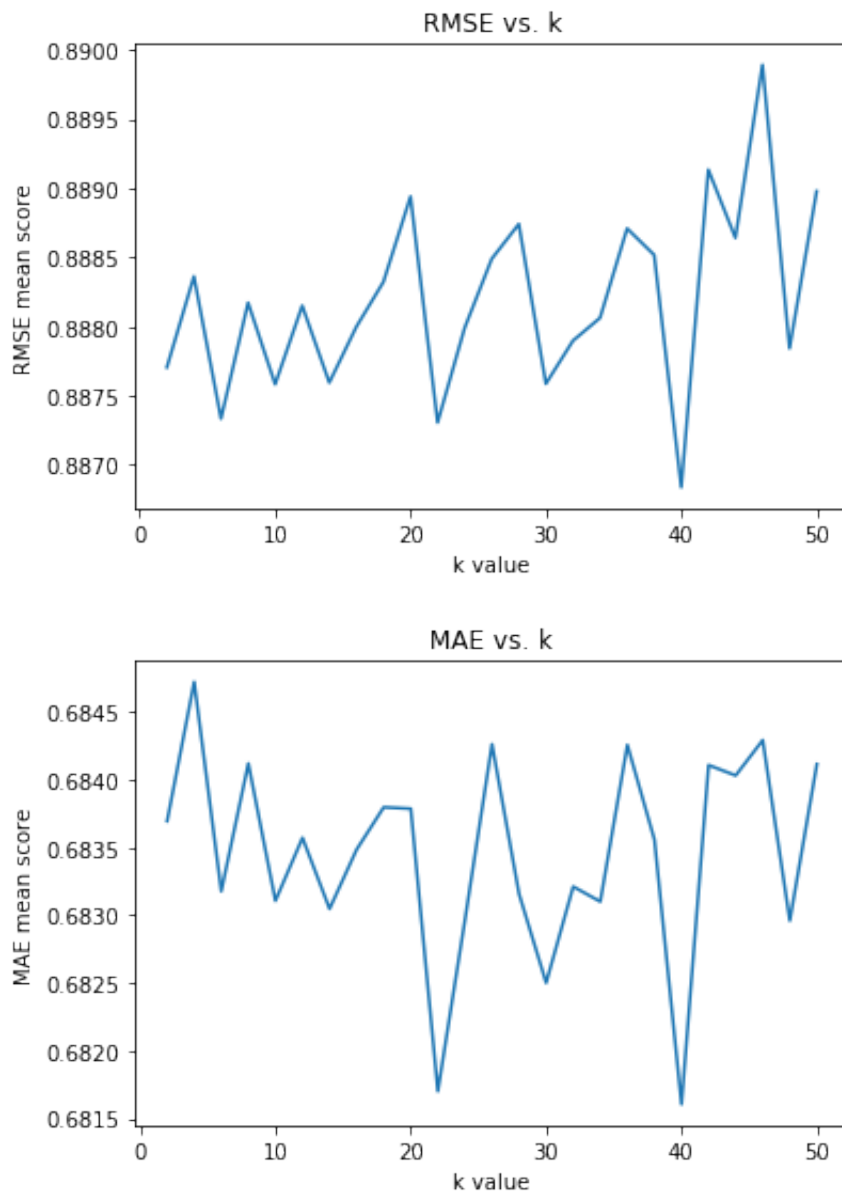
plt.plot(range(2, 52, 2),svd_mae)
plt.title('MAE vs. k')
plt.xlabel('k value')
plt.ylabel('MAE mean score')

print("min RMSE = %f when k = %d" % (np.min(svd_rmse),np.argmin(svd_rmse)*2+2))
print("min MAE = %f when k = %d" % (np.min(svd_mae),np.argmin(svd_mae)*2+2))
```



min RMSE = 0.886838 when k = 40

min MAE = 0.681604 when k = 40



## Question 24 Answer:

The plots are shown above

**Question 25: Use the plot from question 24, to find the optimal number of latent factors. Optimal number of latent factors is the value of k that gives the minimum average RMSE or the minimum average MAE. Please report the minimum average RMSE and MAE.**

## Question 25 Answer:

min RMSE = 0.886838 when k = 40

min MAE = 0.681604 when k = 40

We choose k = 16 as min k.

## Question 26,27,28

```
In [63]: # train section

import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

%load_ext autoreload
%autoreload 2

df = readData()
num_movie = 164979
R = getRatingMatrix(df,num_movie)

reader = Reader(line_format='user item rating timestamp', sep=',',skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

svd_trim_rmse_list = train_trim_svd(data,R)
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
using k = 2
```

```
RMSE: 0.9000
RMSE: 0.9055
RMSE: 0.8979
RMSE: 0.8825
RMSE: 0.8847
RMSE: 0.8625
RMSE: 0.8895
RMSE: 0.9057
RMSE: 0.8123
RMSE: 0.8744
RMSE: 0.8960
RMSE: 0.8235
RMSE: 0.8875
RMSE: 0.8797
RMSE: 0.8426
RMSE: 0.8787
RMSE: 0.8866
RMSE: 0.8373
RMSE: 0.8925
RMSE: 0.8638
RMSE: 0.8800
RMSE: 0.8918
RMSE: 0.8850
RMSE: 0.8971
RMSE: 0.8872
RMSE: 0.8924
RMSE: 0.8012
RMSE: 0.8759
RMSE: 0.8976
RMSE: 0.9084
using k = 4
RMSE: 0.8777
RMSE: 0.8919
RMSE: 0.8602
RMSE: 0.8869
RMSE: 0.8881
RMSE: 0.8078
RMSE: 0.8940
RMSE: 0.8851
RMSE: 0.8438
RMSE: 0.8840
RMSE: 0.8970
RMSE: 0.8523
RMSE: 0.8844
RMSE: 0.8885
RMSE: 0.8531
RMSE: 0.8803
RMSE: 0.8851
RMSE: 0.8497
RMSE: 0.8876
```

```
RMSE: 0.8948
RMSE: 0.8075
RMSE: 0.9075
RMSE: 0.8811
RMSE: 0.9609
RMSE: 0.8805
RMSE: 0.8979
RMSE: 0.8746
RMSE: 0.8771
RMSE: 0.8887
RMSE: 0.8586
using k = 6
RMSE: 0.8836
RMSE: 0.8900
RMSE: 0.8655
RMSE: 0.8902
RMSE: 0.8908
RMSE: 0.8756
RMSE: 0.9003
RMSE: 0.8820
RMSE: 0.9118
RMSE: 0.8746
RMSE: 0.8979
RMSE: 0.8452
RMSE: 0.8909
RMSE: 0.8980
RMSE: 0.8160
RMSE: 0.8878
RMSE: 0.8977
RMSE: 0.8543
RMSE: 0.8961
RMSE: 0.8897
RMSE: 0.8480
RMSE: 0.8748
RMSE: 0.8982
RMSE: 0.8434
RMSE: 0.8677
RMSE: 0.8755
RMSE: 0.8584
RMSE: 0.8872
RMSE: 0.8876
RMSE: 0.8755
using k = 8
RMSE: 0.8920
RMSE: 0.8788
RMSE: 0.8469
RMSE: 0.8870
RMSE: 0.8884
RMSE: 0.8344
RMSE: 0.8780
```

```
RMSE: 0.8921
RMSE: 0.8516
RMSE: 0.8816
RMSE: 0.8902
RMSE: 0.8502
RMSE: 0.8902
RMSE: 0.9029
RMSE: 0.9047
RMSE: 0.8624
RMSE: 0.8871
RMSE: 0.8246
RMSE: 0.8788
RMSE: 0.8896
RMSE: 0.8877
RMSE: 0.8853
RMSE: 0.9002
RMSE: 0.8608
RMSE: 0.9001
RMSE: 0.8853
RMSE: 0.8146
RMSE: 0.8924
RMSE: 0.8793
RMSE: 0.8738
using k = 10
RMSE: 0.8889
RMSE: 0.8867
RMSE: 0.8392
RMSE: 0.8895
RMSE: 0.8873
RMSE: 0.8732
RMSE: 0.8891
RMSE: 0.8858
RMSE: 0.9180
RMSE: 0.8903
RMSE: 0.8742
RMSE: 0.8898
RMSE: 0.8794
RMSE: 0.8944
RMSE: 0.8086
RMSE: 0.8990
RMSE: 0.9005
RMSE: 0.8659
RMSE: 0.8818
RMSE: 0.8958
RMSE: 0.8937
RMSE: 0.8958
RMSE: 0.8826
RMSE: 0.8486
RMSE: 0.8878
RMSE: 0.9100
```

```
RMSE: 0.8252
RMSE: 0.8638
RMSE: 0.8909
RMSE: 0.8355
using k = 12
RMSE: 0.8751
RMSE: 0.8900
RMSE: 0.8684
RMSE: 0.8874
RMSE: 0.8847
RMSE: 0.8217
RMSE: 0.8799
RMSE: 0.8910
RMSE: 0.8379
RMSE: 0.8909
RMSE: 0.8848
RMSE: 0.8764
RMSE: 0.8886
RMSE: 0.9082
RMSE: 0.8496
RMSE: 0.8901
RMSE: 0.8817
RMSE: 0.9109
RMSE: 0.8730
RMSE: 0.8897
RMSE: 0.8572
RMSE: 0.8851
RMSE: 0.8864
RMSE: 0.8563
RMSE: 0.8959
RMSE: 0.8828
RMSE: 0.8428
RMSE: 0.8858
RMSE: 0.9110
RMSE: 0.8520
using k = 14
RMSE: 0.8880
RMSE: 0.9014
RMSE: 0.8146
RMSE: 0.8898
RMSE: 0.9025
RMSE: 0.8369
RMSE: 0.8829
RMSE: 0.8793
RMSE: 0.9382
RMSE: 0.8867
RMSE: 0.8722
RMSE: 0.8597
RMSE: 0.8957
RMSE: 0.8819
```

```
RMSE: 0.7956
RMSE: 0.8940
RMSE: 0.9134
RMSE: 0.9361
RMSE: 0.8714
RMSE: 0.8936
RMSE: 0.8664
RMSE: 0.8977
RMSE: 0.8708
RMSE: 0.8492
RMSE: 0.8734
RMSE: 0.9135
RMSE: 0.8033
RMSE: 0.8834
RMSE: 0.8705
RMSE: 0.8615
using k = 16
RMSE: 0.8803
RMSE: 0.8814
RMSE: 0.8157
RMSE: 0.8852
RMSE: 0.8908
RMSE: 0.8978
RMSE: 0.8822
RMSE: 0.8888
RMSE: 0.7950
RMSE: 0.8935
RMSE: 0.8952
RMSE: 0.8229
RMSE: 0.8728
RMSE: 0.9055
RMSE: 0.8383
RMSE: 0.9048
RMSE: 0.8733
RMSE: 0.8630
RMSE: 0.8909
RMSE: 0.9059
RMSE: 0.8880
RMSE: 0.8866
RMSE: 0.8971
RMSE: 0.8938
RMSE: 0.8914
RMSE: 0.8926
RMSE: 0.9286
RMSE: 0.8810
RMSE: 0.8799
RMSE: 0.8470
using k = 18
RMSE: 0.8830
RMSE: 0.8797
```

```
RMSE: 0.8926
RMSE: 0.8834
RMSE: 0.9008
RMSE: 0.8796
RMSE: 0.8808
RMSE: 0.9096
RMSE: 0.8083
RMSE: 0.8784
RMSE: 0.9015
RMSE: 0.8528
RMSE: 0.8834
RMSE: 0.8724
RMSE: 0.8274
RMSE: 0.8915
RMSE: 0.8853
RMSE: 0.8809
RMSE: 0.8833
RMSE: 0.8918
RMSE: 0.8107
RMSE: 0.8891
RMSE: 0.8855
RMSE: 0.8841
RMSE: 0.8942
RMSE: 0.8943
RMSE: 0.8581
RMSE: 0.8961
RMSE: 0.8885
RMSE: 0.8277
using k = 20
RMSE: 0.8944
RMSE: 0.8724
RMSE: 0.8362
RMSE: 0.8917
RMSE: 0.8859
RMSE: 0.8933
RMSE: 0.8958
RMSE: 0.8812
RMSE: 0.8298
RMSE: 0.8953
RMSE: 0.9092
RMSE: 0.8769
RMSE: 0.8804
RMSE: 0.8922
RMSE: 0.8294
RMSE: 0.8824
RMSE: 0.8907
RMSE: 0.8166
RMSE: 0.8911
RMSE: 0.8931
RMSE: 0.8711
```



```
RMSE: 0.8740
RMSE: 0.8902
RMSE: 0.9142
RMSE: 0.8816
RMSE: 0.8935
RMSE: 0.8365
RMSE: 0.8751
RMSE: 0.8840
RMSE: 0.8514
using k = 22
RMSE: 0.8802
RMSE: 0.8616
RMSE: 0.8110
RMSE: 0.8890
RMSE: 0.9063
RMSE: 0.8189
RMSE: 0.8830
RMSE: 0.9060
RMSE: 0.8970
RMSE: 0.8757
RMSE: 0.8999
RMSE: 0.8304
RMSE: 0.8775
RMSE: 0.8910
RMSE: 0.8891
RMSE: 0.8947
RMSE: 0.8748
RMSE: 0.8524
RMSE: 0.8852
RMSE: 0.8899
RMSE: 0.8325
RMSE: 0.8915
RMSE: 0.8687
RMSE: 0.8693
RMSE: 0.8797
RMSE: 0.8918
RMSE: 0.8783
RMSE: 0.9047
RMSE: 0.9145
RMSE: 0.8746
using k = 24
RMSE: 0.8888
RMSE: 0.9048
RMSE: 0.8139
RMSE: 0.8980
RMSE: 0.8970
RMSE: 0.8177
RMSE: 0.8825
RMSE: 0.9082
RMSE: 0.8068
```

```
RMSE: 0.8803
RMSE: 0.8940
RMSE: 0.8496
RMSE: 0.8961
RMSE: 0.9014
RMSE: 0.8652
RMSE: 0.8833
RMSE: 0.8783
RMSE: 0.8345
RMSE: 0.8780
RMSE: 0.8896
RMSE: 0.8880
RMSE: 0.8968
RMSE: 0.8647
RMSE: 0.8049
RMSE: 0.8915
RMSE: 0.8932
RMSE: 0.9406
RMSE: 0.8662
RMSE: 0.8849
RMSE: 0.9091
using k = 26
RMSE: 0.8835
RMSE: 0.8975
RMSE: 0.8683
RMSE: 0.8882
RMSE: 0.8884
RMSE: 0.8049
RMSE: 0.8705
RMSE: 0.8970
RMSE: 0.7997
RMSE: 0.8856
RMSE: 0.9093
RMSE: 0.9090
RMSE: 0.8859
RMSE: 0.8861
RMSE: 0.8659
RMSE: 0.8889
RMSE: 0.8902
RMSE: 0.8948
RMSE: 0.8792
RMSE: 0.8773
RMSE: 0.8441
RMSE: 0.8936
RMSE: 0.8789
RMSE: 0.8428
RMSE: 0.8806
RMSE: 0.8883
RMSE: 0.8500
RMSE: 0.8941
```

```
RMSE: 0.8897
RMSE: 0.8718
using k = 28
RMSE: 0.8817
RMSE: 0.9098
RMSE: 0.8041
RMSE: 0.8836
RMSE: 0.8821
RMSE: 0.7935
RMSE: 0.8945
RMSE: 0.9026
RMSE: 0.9033
RMSE: 0.8931
RMSE: 0.8704
RMSE: 0.8903
RMSE: 0.8962
RMSE: 0.8875
RMSE: 0.8141
RMSE: 0.8995
RMSE: 0.8931
RMSE: 0.8602
RMSE: 0.8722
RMSE: 0.8861
RMSE: 0.8796
RMSE: 0.8759
RMSE: 0.8925
RMSE: 0.8904
RMSE: 0.8832
RMSE: 0.8854
RMSE: 0.8780
RMSE: 0.8810
RMSE: 0.9063
RMSE: 0.8422
using k = 30
RMSE: 0.8908
RMSE: 0.8958
RMSE: 0.8759
RMSE: 0.8960
RMSE: 0.8876
RMSE: 0.8640
RMSE: 0.8798
RMSE: 0.8960
RMSE: 0.8738
RMSE: 0.8841
RMSE: 0.8967
RMSE: 0.8074
RMSE: 0.8670
RMSE: 0.8856
RMSE: 0.8912
RMSE: 0.8908
```

```
RMSE: 0.8924
RMSE: 0.9160
RMSE: 0.8839
RMSE: 0.8928
RMSE: 0.7934
RMSE: 0.8771
RMSE: 0.8920
RMSE: 0.8248
RMSE: 0.8858
RMSE: 0.8697
RMSE: 0.8268
RMSE: 0.9063
RMSE: 0.9020
RMSE: 0.8765
using k = 32
RMSE: 0.8616
RMSE: 0.8762
RMSE: 0.7633
RMSE: 0.8891
RMSE: 0.8817
RMSE: 0.8767
RMSE: 0.8930
RMSE: 0.8951
RMSE: 0.7786
RMSE: 0.8868
RMSE: 0.9082
RMSE: 0.8973
RMSE: 0.8958
RMSE: 0.8936
RMSE: 0.8398
RMSE: 0.8809
RMSE: 0.9003
RMSE: 0.8740
RMSE: 0.8841
RMSE: 0.9016
RMSE: 0.9075
RMSE: 0.8874
RMSE: 0.8810
RMSE: 0.8956
RMSE: 0.8843
RMSE: 0.8900
RMSE: 0.8117
RMSE: 0.8859
RMSE: 0.8987
RMSE: 0.8270
using k = 34
RMSE: 0.8705
RMSE: 0.8936
RMSE: 0.8670
RMSE: 0.8785
```

```
RMSE: 0.9018
RMSE: 0.8465
RMSE: 0.8909
RMSE: 0.8803
RMSE: 0.8268
RMSE: 0.8850
RMSE: 0.9054
RMSE: 0.8540
RMSE: 0.8926
RMSE: 0.8915
RMSE: 0.8568
RMSE: 0.8912
RMSE: 0.9051
RMSE: 0.8254
RMSE: 0.9025
RMSE: 0.9051
RMSE: 0.8862
RMSE: 0.8806
RMSE: 0.8751
RMSE: 0.9098
RMSE: 0.8886
RMSE: 0.8783
RMSE: 0.8263
RMSE: 0.8800
RMSE: 0.8831
RMSE: 0.8604
using k = 36
RMSE: 0.8955
RMSE: 0.8721
RMSE: 0.8089
RMSE: 0.8770
RMSE: 0.8921
RMSE: 0.8751
RMSE: 0.8708
RMSE: 0.8980
RMSE: 0.8568
RMSE: 0.8805
RMSE: 0.8985
RMSE: 0.8393
RMSE: 0.8898
RMSE: 0.8829
RMSE: 0.8375
RMSE: 0.8771
RMSE: 0.8921
RMSE: 0.8530
RMSE: 0.8835
RMSE: 0.9076
RMSE: 0.8234
RMSE: 0.8955
RMSE: 0.9014
```

```
RMSE: 0.8472
RMSE: 0.9022
RMSE: 0.8954
RMSE: 0.8727
RMSE: 0.8810
RMSE: 0.8820
RMSE: 0.8858
using k = 38
RMSE: 0.8794
RMSE: 0.8987
RMSE: 0.8714
RMSE: 0.8942
RMSE: 0.8812
RMSE: 0.8671
RMSE: 0.8788
RMSE: 0.8958
RMSE: 0.8911
RMSE: 0.8807
RMSE: 0.9021
RMSE: 0.8062
RMSE: 0.8959
RMSE: 0.8877
RMSE: 0.8725
RMSE: 0.8838
RMSE: 0.8911
RMSE: 0.8490
RMSE: 0.8824
RMSE: 0.8986
RMSE: 0.8409
RMSE: 0.8909
RMSE: 0.8710
RMSE: 0.8877
RMSE: 0.8815
RMSE: 0.8963
RMSE: 0.8279
RMSE: 0.9015
RMSE: 0.8917
RMSE: 0.8071
using k = 40
RMSE: 0.8763
RMSE: 0.8938
RMSE: 0.8229
RMSE: 0.8898
RMSE: 0.8998
RMSE: 0.8268
RMSE: 0.8804
RMSE: 0.9192
RMSE: 0.8519
RMSE: 0.8748
RMSE: 0.8872
```

```
RMSE: 0.8590
RMSE: 0.8969
RMSE: 0.8959
RMSE: 0.8679
RMSE: 0.9123
RMSE: 0.8858
RMSE: 0.8837
RMSE: 0.8940
RMSE: 0.8772
RMSE: 0.8352
RMSE: 0.8789
RMSE: 0.8969
RMSE: 0.8775
RMSE: 0.8652
RMSE: 0.8788
RMSE: 0.8641
RMSE: 0.8926
RMSE: 0.8818
RMSE: 0.8975
using k = 42
RMSE: 0.8906
RMSE: 0.9021
RMSE: 0.8593
RMSE: 0.8872
RMSE: 0.8692
RMSE: 0.9122
RMSE: 0.8822
RMSE: 0.8883
RMSE: 0.7990
RMSE: 0.8931
RMSE: 0.8960
RMSE: 0.9096
RMSE: 0.8735
RMSE: 0.8865
RMSE: 0.8018
RMSE: 0.8921
RMSE: 0.9034
RMSE: 0.8842
RMSE: 0.8986
RMSE: 0.8802
RMSE: 0.8837
RMSE: 0.8864
RMSE: 0.8909
RMSE: 0.8852
RMSE: 0.8741
RMSE: 0.9031
RMSE: 0.8108
RMSE: 0.8945
RMSE: 0.8961
RMSE: 0.8335
```

```
using k = 44
RMSE: 0.8732
RMSE: 0.9053
RMSE: 0.7943
RMSE: 0.8970
RMSE: 0.8900
RMSE: 0.8715
RMSE: 0.8721
RMSE: 0.8901
RMSE: 0.8432
RMSE: 0.8845
RMSE: 0.8738
RMSE: 0.8466
RMSE: 0.9074
RMSE: 0.9023
RMSE: 0.8743
RMSE: 0.8851
RMSE: 0.8947
RMSE: 0.8978
RMSE: 0.8882
RMSE: 0.8932
RMSE: 0.8371
RMSE: 0.8942
RMSE: 0.9001
RMSE: 0.8837
RMSE: 0.8866
RMSE: 0.8829
RMSE: 0.8429
RMSE: 0.8785
RMSE: 0.8847
RMSE: 0.8924
using k = 46
RMSE: 0.8783
RMSE: 0.8999
RMSE: 0.8620
RMSE: 0.8919
RMSE: 0.8910
RMSE: 0.8764
RMSE: 0.8840
RMSE: 0.9084
RMSE: 0.8576
RMSE: 0.8795
RMSE: 0.9036
RMSE: 0.8408
RMSE: 0.8880
RMSE: 0.8896
RMSE: 0.8451
RMSE: 0.8788
RMSE: 0.9047
RMSE: 0.8261
```



```
RMSE: 0.8895
RMSE: 0.8721
RMSE: 0.8164
RMSE: 0.8822
RMSE: 0.8925
RMSE: 0.8610
RMSE: 0.8882
RMSE: 0.8839
RMSE: 0.9104
RMSE: 0.8964
RMSE: 0.8711
RMSE: 0.8775
using k = 48
RMSE: 0.8844
RMSE: 0.8920
RMSE: 0.8434
RMSE: 0.8943
RMSE: 0.8821
RMSE: 0.8964
RMSE: 0.8751
RMSE: 0.8867
RMSE: 0.8119
RMSE: 0.8976
RMSE: 0.8698
RMSE: 0.8677
RMSE: 0.8861
RMSE: 0.8993
RMSE: 0.8583
RMSE: 0.9016
RMSE: 0.8928
RMSE: 0.8539
RMSE: 0.8753
RMSE: 0.9048
RMSE: 0.8395
RMSE: 0.8927
RMSE: 0.8955
RMSE: 0.8665
RMSE: 0.9016
RMSE: 0.9099
RMSE: 0.8696
RMSE: 0.8569
RMSE: 0.8935
RMSE: 0.8064
using k = 50
RMSE: 0.9003
RMSE: 0.8796
RMSE: 0.8467
RMSE: 0.8711
RMSE: 0.8952
RMSE: 0.8778
```

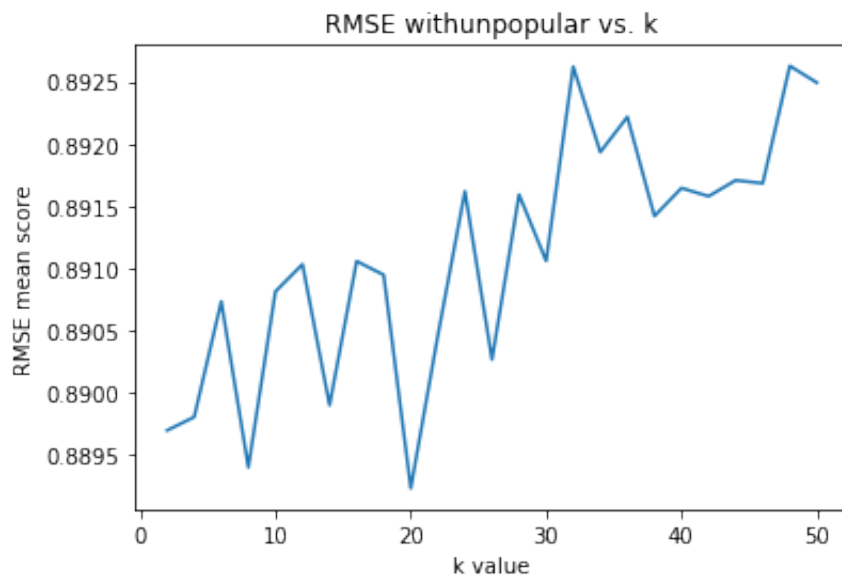
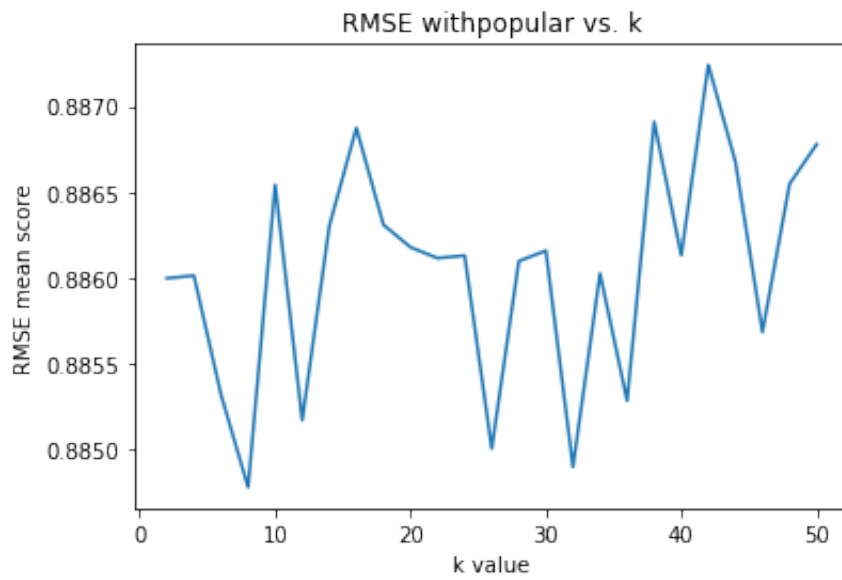
```
RMSE: 0.8937
RMSE: 0.8943
RMSE: 0.8377
RMSE: 0.8986
RMSE: 0.8789
RMSE: 0.8618
RMSE: 0.8885
RMSE: 0.9111
RMSE: 0.7468
RMSE: 0.9024
RMSE: 0.8805
RMSE: 0.8935
RMSE: 0.8876
RMSE: 0.9095
RMSE: 0.8672
RMSE: 0.8775
RMSE: 0.8948
RMSE: 0.8503
RMSE: 0.8710
RMSE: 0.8928
RMSE: 0.8416
RMSE: 0.8771
RMSE: 0.8882
RMSE: 0.8767
SVD with trim is finished!!
```

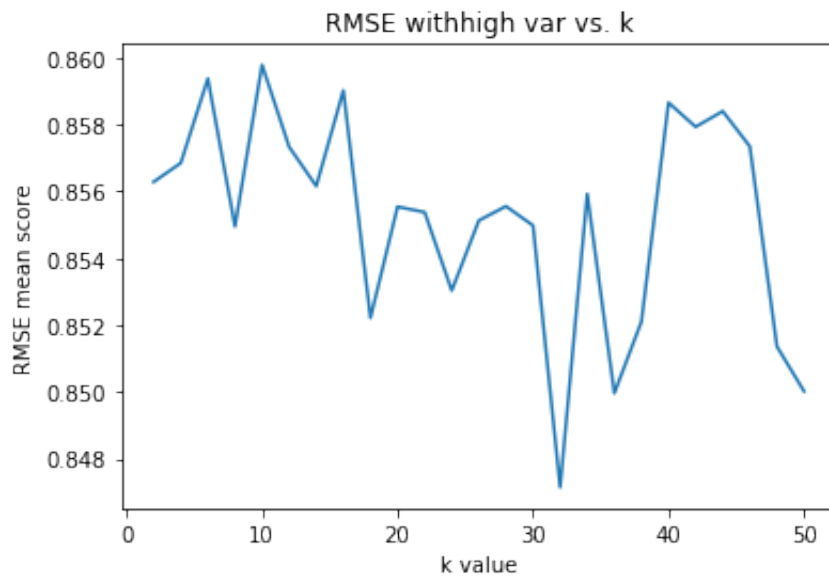
```
In [64]: import matplotlib.pyplot as plt

trim_name = ['popular', 'unpopular', 'high var']

for i in range(3):
    plt.figure()
    plt.plot(range(2, 52, 2), svd_trim_rmse_list[i])
    plt.title('RMSE with'+trim_name[i]+' vs. k')
    plt.xlabel('k value')
    plt.ylabel('RMSE mean score')
    print("min RMSE for " + trim_name[i]+ " trimming = ", np.min(svd_t
rim_rmse_list[i]))

min RMSE for popular trimming = 0.884778688883
min RMSE for unpopular trimming = 0.889232703409
min RMSE for high var trimming = 0.847153331712
```





### Question 26,27,28 Answer:

min RMSE for popular trimming = 0.884778688883 min RMSE for unpopular trimming = 0.889232703409

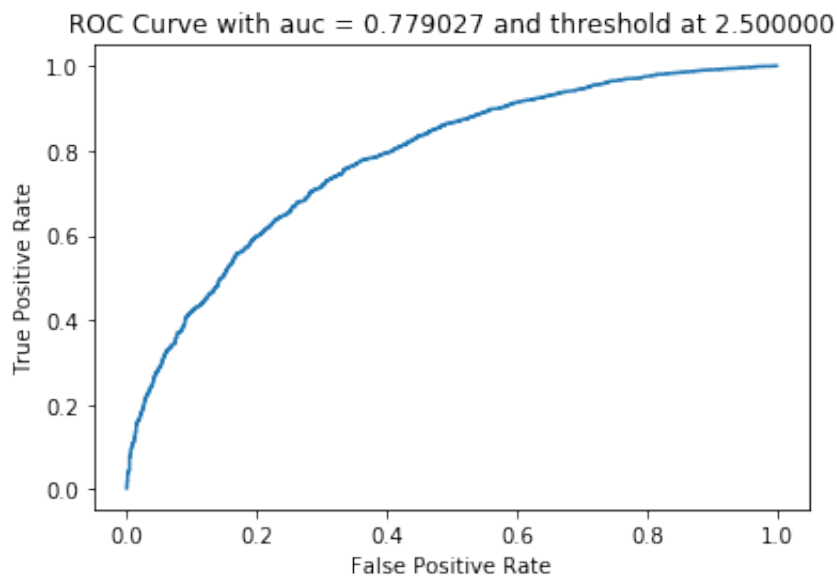
min RMSE for high var trimming = 0.847153331712

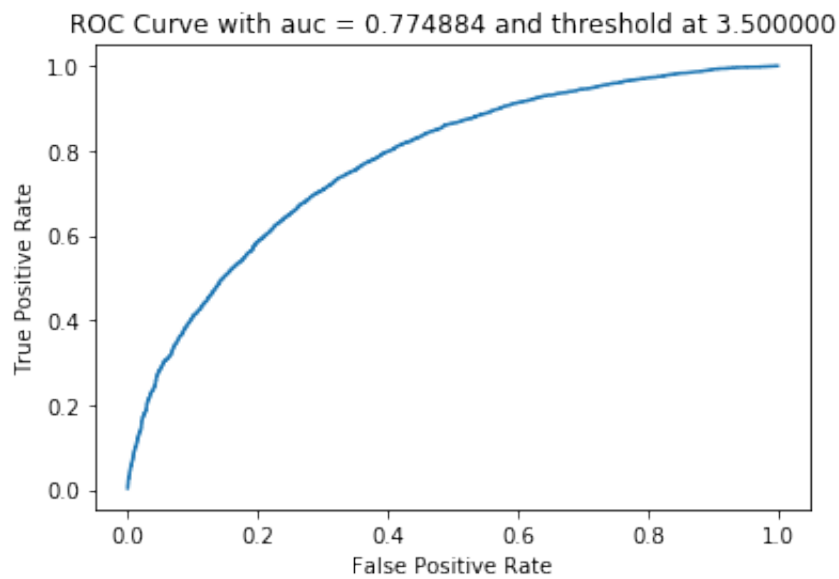
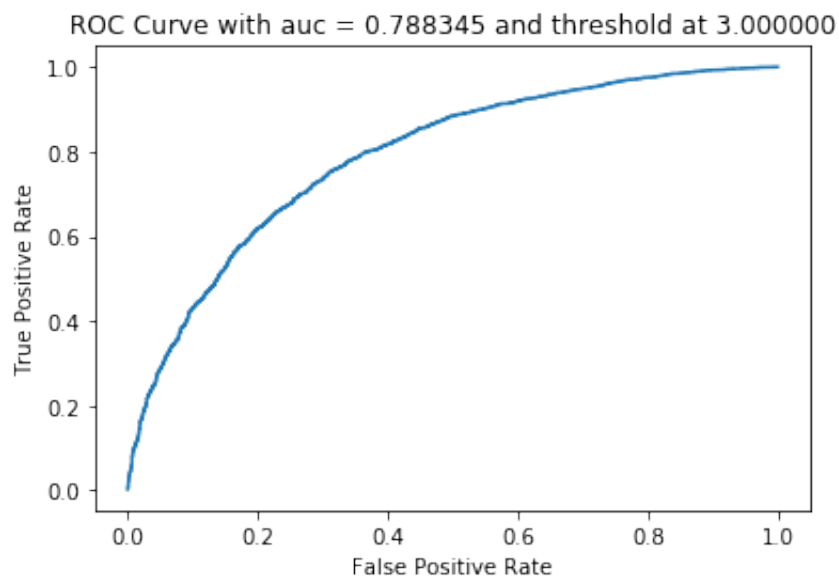
**Question 29: Plot the ROC curves for the MF with bias collaborative filter designed in question 24 for threshold values [2.5,3,3.5,4]. For the ROC plot-ting use the optimal number of latent factors found in question 25. For each of the plots, also report the area under the curve (AUC) value.**

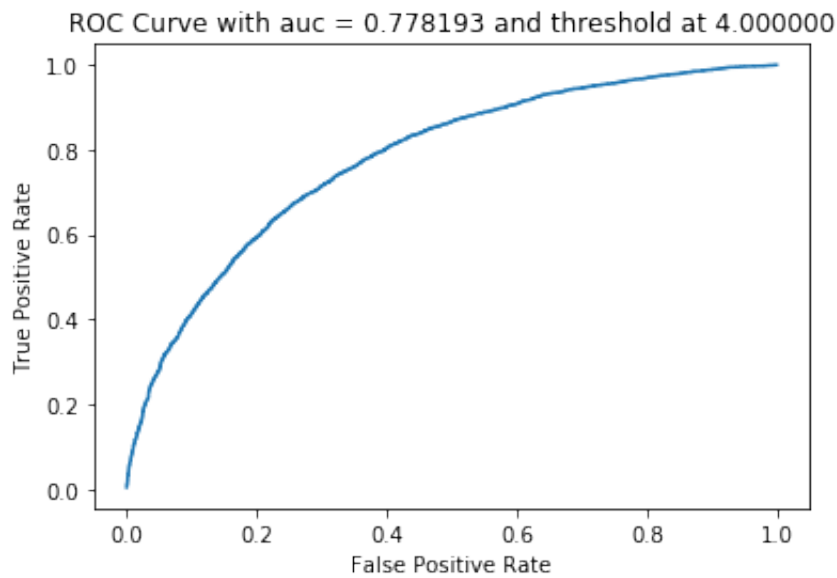
```
In [75]: from surprise.prediction_algorithms.knns import KNNWithMeans
from surprise.model_selection import KFold
from sklearn import metrics
from surprise.prediction_algorithms.matrix_factorization import NMF, SVD

thresholds = [2.5, 3, 3.5, 4]
svd_min_k = 40
kf = KFold(n_splits = 10)
svd = SVD(n_factors = svd_min_k)
for trainset, testset in kf.split(data):
    pass
svd.fit(trainset)
pred = svd.test(testset)

for i in thresholds:
    (rs, rs_pred) = ([], [])
    for j in pred:
        if j[2] >= i: rs.append(1)
        else: rs.append(0)
        rs_pred.append(j[3])
    fpr, tpr, _ = metrics.roc_curve(y_true = rs, y_score = rs_pred, pos_label=1)
    auc = metrics.roc_auc_score(y_true=rs, y_score=rs_pred)
    plt.figure()
    plt.plot(fpr, tpr)
    plt.title('ROC Curve with auc = %f and threshold at %f'%(auc, i))
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
```







## Question 29 Answer:

The auc value is shown in each ROC graph title, we can see that when the threshold is at 3.0, the auc score is highest.

**Question 30: Design a naive collaborative filter to predict the ratings of the movies in the MovieLens dataset and evaluate it's performance using 10-fold cross validation. Compute the average RMSE by averaging the RMSE across all 10 folds. Report the average RMSE.**

```
In [66]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

df = readData()
num_movie = 164979
R = getRatingMatrix(df,num_movie)

reader = Reader(line_format='user item rating timestamp', sep=',',skip
_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=re
ader)

print(train_naive(data,R))

3.68869926255
```

## Question 30 Answer:

the average Rmse = 3.68869926255

## Question 31,32,33



```
In [67]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

%load_ext autoreload
%autoreload 2

df = readData()
num_movie = 164979
R = getRatingMatrix(df,num_movie)

reader = Reader(line_format='user item rating timestamp', sep=',',skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

print(train_trim_naive(data,R))
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
(3.7136303466017195, 3.6532010216772335, 3.9180587411431818)
```

## Question 31,32,33 Answer:

the average RMSE for popular, unpopular, high var trimming are (3.7136303466017195, 3.6532010216772335, 3.9180587411431818)

**Question 34: Plot the ROC curves (threshold = 3) for the k-NN, NNMF, and MF with bias based collaborative filters in the same figure. Use the figure to compare the performance of the filters in predicting the ratings of the movies.**

```

In [68]: threshold = 3

for trainset, testset in kf.split(data):
    pass
print ("doing knn")
min_k = 24
kf = KFold(n_splits = 10)
sim_options = {'name': 'pearson'}
knn = KNNWithMeans(k = min_k, sim_options = sim_options)
knn.fit(trainset)
pred_knn=knn.test(testset)

print("doing nmf")
nmf_min_k = 20
nmf = NMF(n_factors = nmf_min_k)
nmf.fit(trainset)
pred_nmf=nmf.test(testset)

print("doing MF")
svd_min_k = 40
svd = SVD(n_factors = svd_min_k)
svd.fit(trainset)
pred_svd=svd.test(testset)

preds = [pred_knn, pred_nmf, pred_svd]
fprs = []
tprs = []
aucs = []
for i in range(3):
    pred = preds[i]
    rs = []
    rs_pred = []
    for j in pred:
        if j[2] >= threshold: rs.append(1)
        else: rs.append(0)
        rs_pred.append(j[3])
    fpr, tpr, _ = metrics.roc_curve(y_true = rs, y_score = rs_pred, pos_label=1)
    aucs.append(metrics.roc_auc_score(y_true=rs, y_score=rs_pred))
    fprs.append(fpr)
    tprs.append(tpr)

```

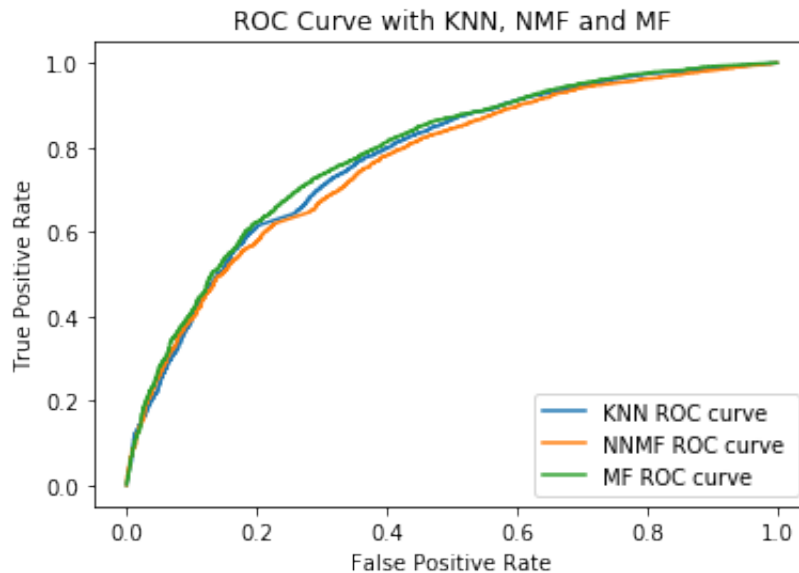
```

doing knn
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing MF

```

```
In [69]: plt.figure()  
plt.plot(fprs[0],tprs[0],label='KNN ROC curve ')  
plt.plot(fprs[1],tprs[1],label='NNMF ROC curve ')  
plt.plot(fprs[2],tprs[2],label='MF ROC curve ')  
plt.title('ROC Curve with KNN, NMF and MF')  
plt.xlabel('False Positive Rate')  
plt.ylabel('True Positive Rate')  
plt.legend(loc="lower right")  
  
print ("KNN AUC =", aucs[0])  
print ("NMF AUC =", aucs[1])  
print ("MF AUC =", aucs[2])
```

```
KNN AUC = 0.773815574657  
NMF AUC = 0.762299167603  
MF AUC = 0.785260659566
```



## Question 34 Answer:

From the graph, we can see that the MF curve is above the other 2 plot. And the auc score of MF is higher than others. This means MF has the best performance here.

```
KNN AUC = 0.773815574657  
NMF AUC = 0.762299167603  
MF AUC = 0.785260659566
```

**Question 35: Precision and Recall are defined by the mathematical expressions given by equations 12 and 13 respectively. Please explain the meaning of precision and recall in your own words.**

### **Question 35 Answer:**

precision is defined as ratio of correct prediction(recommendation) to all recommendation, this measures how precise our recommendation is for a user.

recall is defined as the ratio of correct prediction(recommendation) to the num of items that the user actually liked. This measures how many predictions are inside the user's previously recorded data.

### **Question 36, 37, 38**

```

In [70]: import numpy as np
from surprise.dataset import Dataset
from surprise.reader import Reader
from util import *

reader = Reader(line_format='user item rating timestamp', sep=',', skip_lines=1)

data = Dataset.load_from_file('ml-latest-small/ratings.csv', reader=reader)

threshold = 3

min_k = 24
sim_options = {'name': 'pearson'}
knn = KNNWithMeans(k = min_k, sim_options = sim_options)

nmf_min_k = 20
nmf = NMF(n_factors = nmf_min_k)

svd_min_k = 40
svd = SVD(n_factors = svd_min_k)

models = [knn,nmf,svd]

precisions,recalls = calculate_precision_recall(models,threshold,data)

Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd

```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
doing nmf
doing svd
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
In [71]: import matplotlib.pyplot as plt

for i in range(len(model_names)):
    ts = range(1, 26)

    plt.figure()

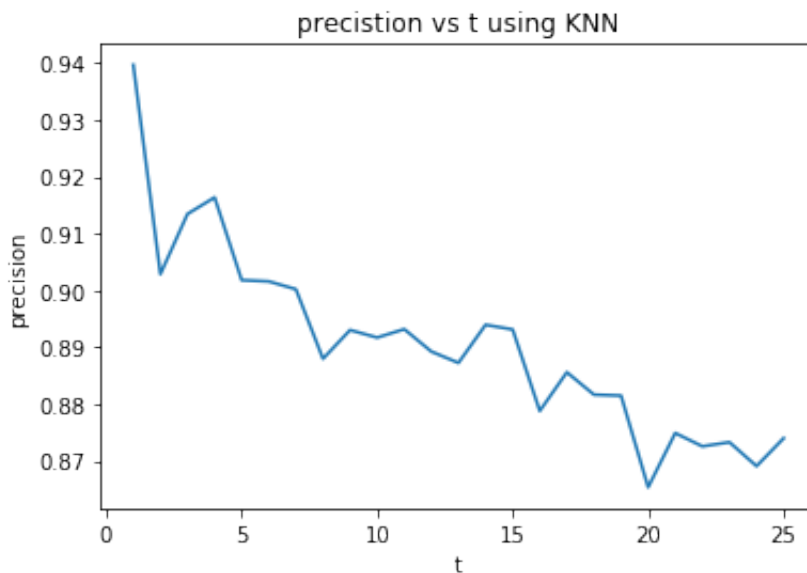
    plt.plot(range(1, 26), precisions[i])
    plt.title("precision vs t using "+model_names[i])
    plt.xlabel('t')
    plt.ylabel("precision")

    plt.figure()

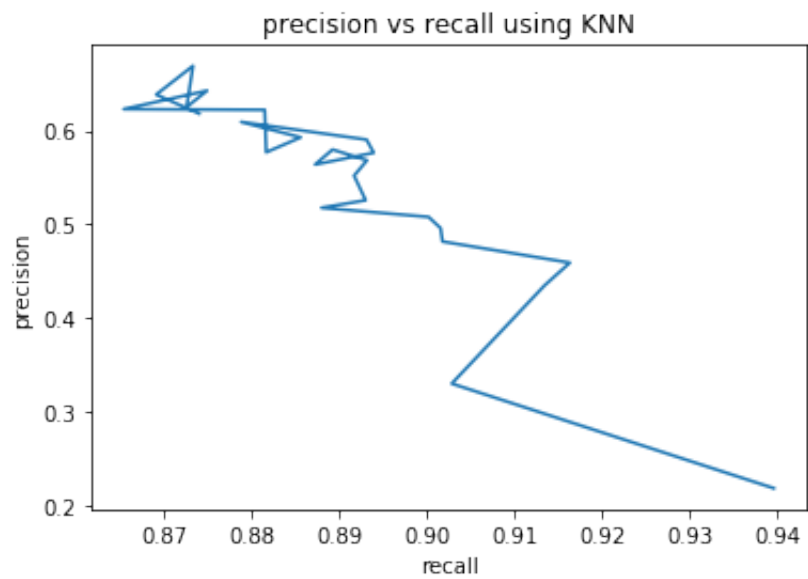
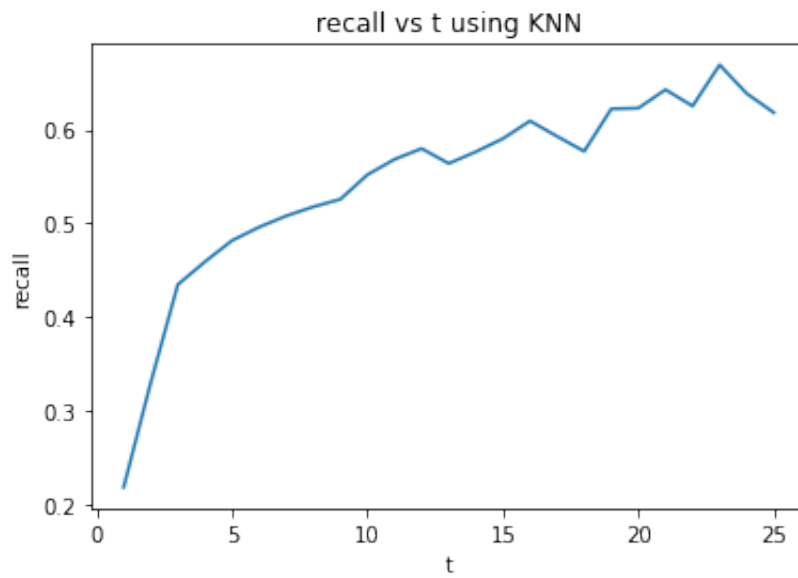
    plt.plot(range(1, 26), recalls[i])
    plt.title("recall vs t using "+model_names[i])
    plt.xlabel('t')
    plt.ylabel("recall")

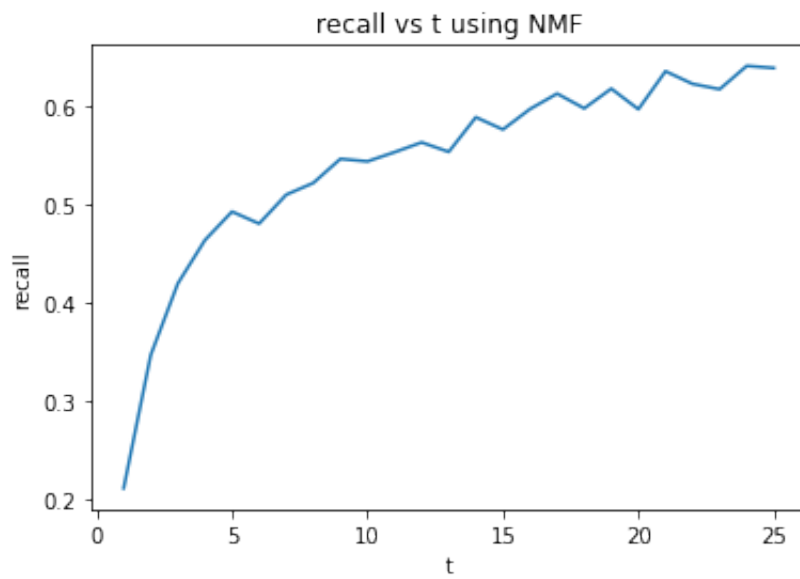
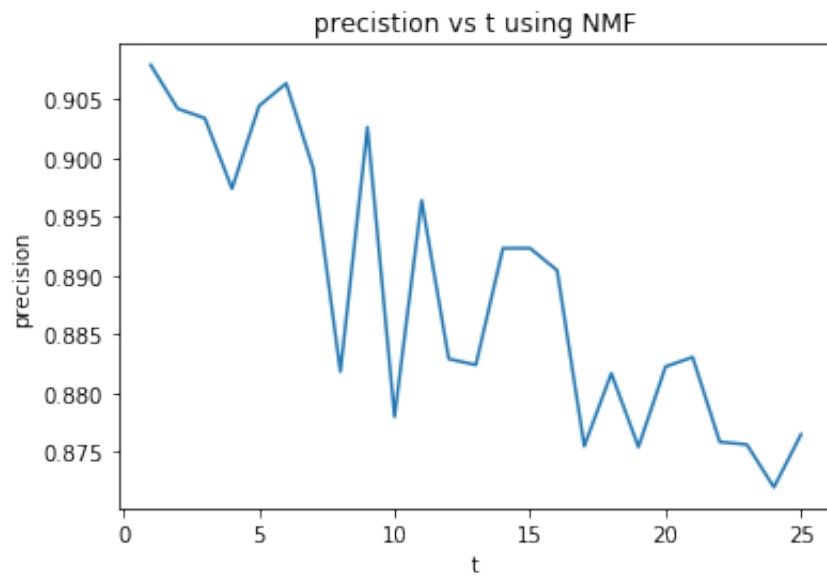
    plt.figure()

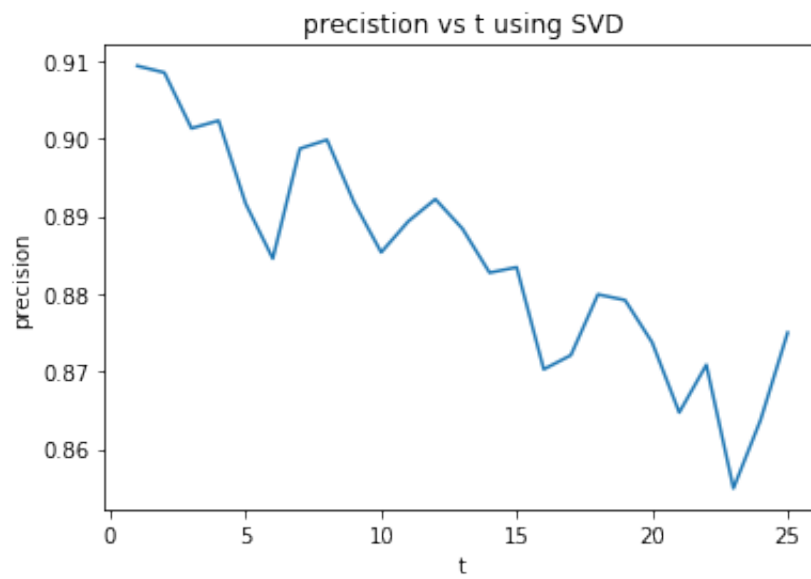
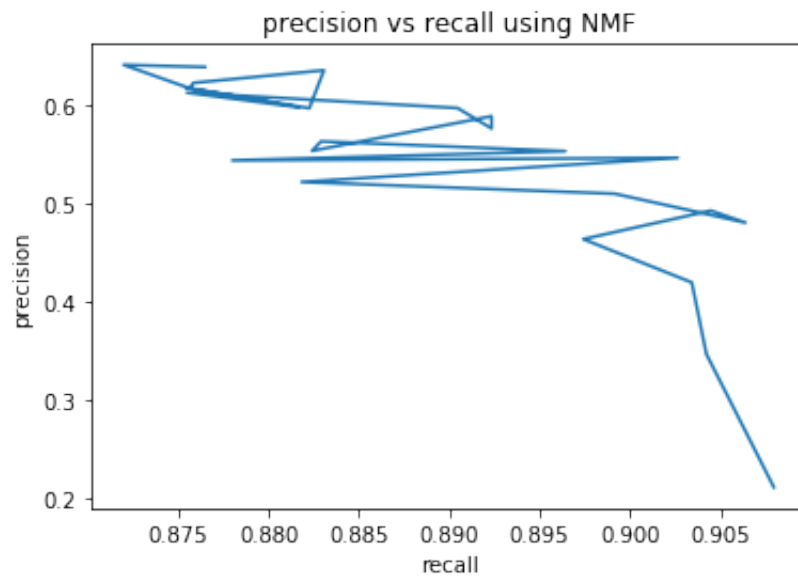
    plt.plot(precisions[i], recalls[i])
    plt.title("precision vs recall using "+model_names[i])
    plt.xlabel('recall')
    plt.ylabel("precision")
```

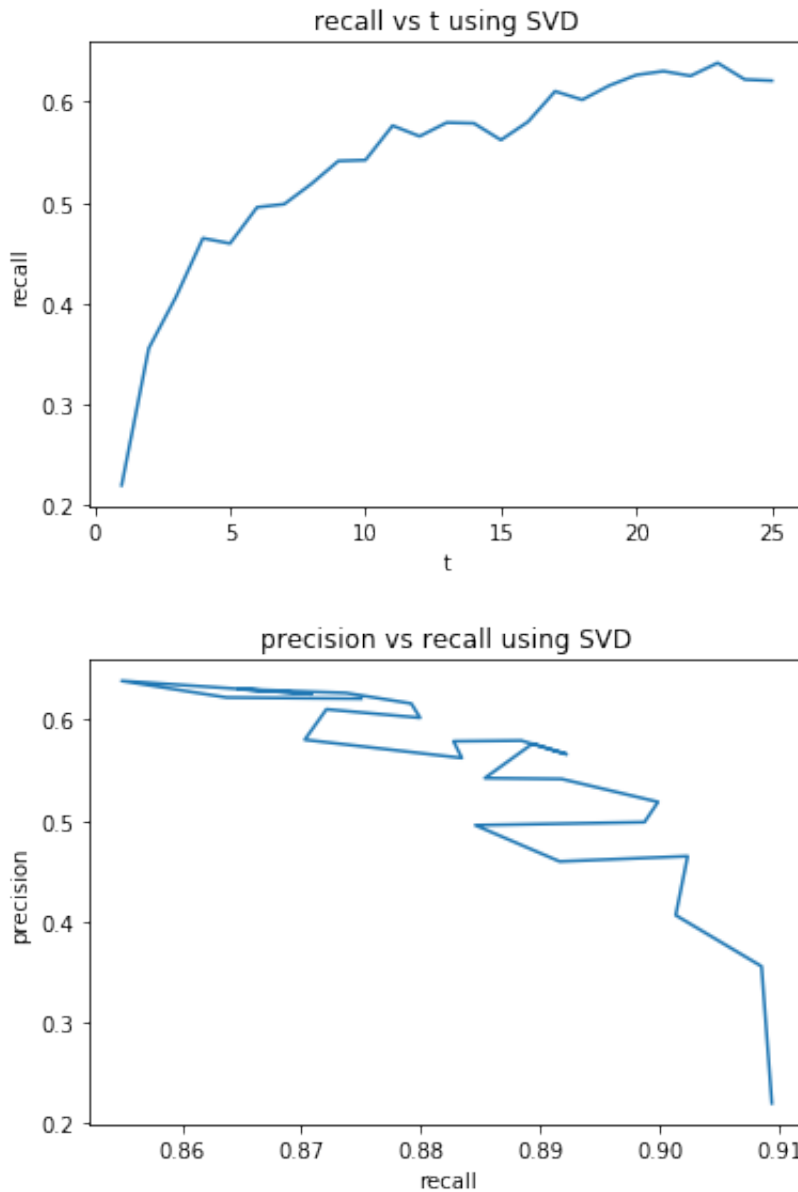












## Question 36,37,38 Answer:

For the precision vs t plot of KNN, NMF and MF, we can see a general decreasing of precision as t increases. This should indicate using the few top recommendation results in a better performance.

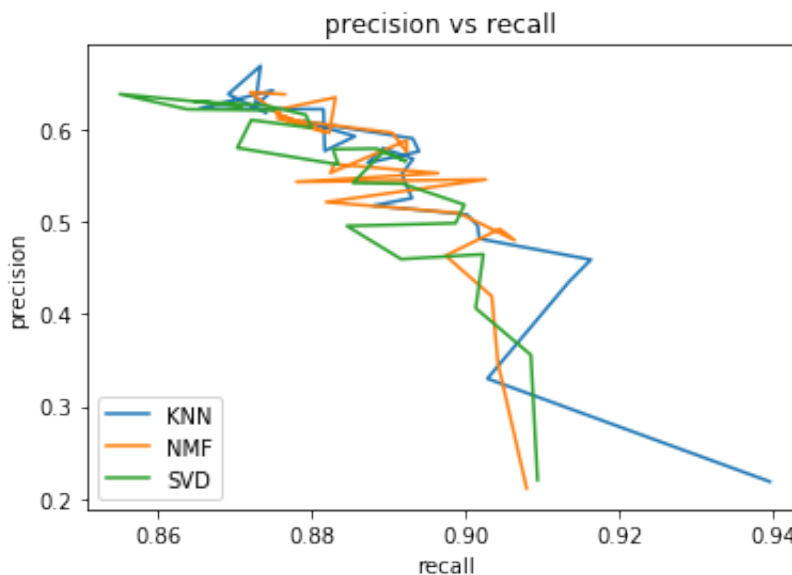
For the recall vs t plot of KNN, NMF and MF, we see a increase in recall as t increases. This indicates that as t increase, we have get more and more recommendation that user actually liked.

For the precision vs recall plot of KNN, NMF and MF, we see decrease of precision as recall increases, which means as we give out more recommendations, we are covering more and more what user actually liked, but the prediction is poor.

**Question 39: Plot the precision-recall curve obtained in questions 36,37, and 38 in the same figure. Use this figure to compare the relevance of the recommendation list generated using k-NN, NMF, and MF with bias predictions.**

```
In [72]: for i in range(len(model_names)):
          plt.plot(precisions[i], recalls[i], label = model_names[i])
          plt.title("precision vs recall ")
          plt.xlabel('recall')
          plt.ylabel("precision")
          plt.legend(loc="lower left")
```

Out[72]: <matplotlib.legend.Legend at 0x11521ff60>



## Question 39 Answer:

At the end of the plot, we can see that KNN has better recall performance, and at the start of the performance, SVD has better precisions. The general trend is decreasing percision with increased recall. The result of 3 models show same trend, which indicates that when t is small, percision is good, but rather low recall. And at high t, percision is low, but better recall.