In [114]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.neighbors import NearestNeighbors
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
from scipy.spatial.distance import correlation
from surprise.prediction_algorithms.knns import KNNWithMeans
from surprise.model_selection import cross_validate, KFold, train_test_split
from surprise import Dataset, Reader, accuracy
from sklearn import metrics
from collections import Counter
from sklearn.metrics import roc_curve, auc
```

In [389]:
```
df = pd.read_csv("/Users/ryan/Downloads/Synthetic_Movie_Lens/ratings.csv")
ratings_matrix = pd.pivot_table(df, values="rating", columns=["movieId"], index=["userId"])
```

QUESTION 1: Explore the Dataset:

In [3]:
```
print(len(df.userId.unique()))
```

```
610
```

In [4]:
```
print(len(df.movieId.unique()))
```

```
9724
```

A. Compute the sparsity of the movie rating dataset:

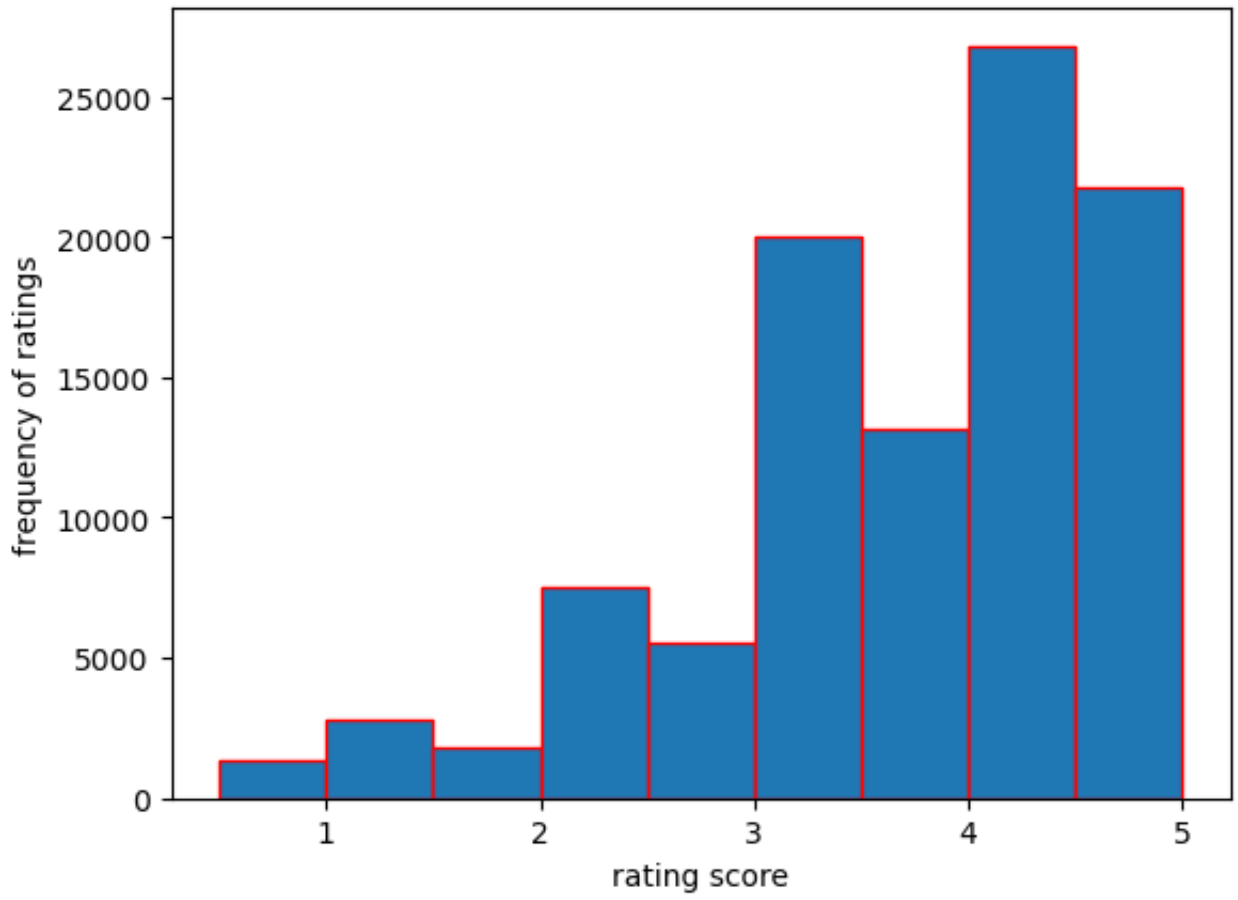
In [383]:
```
user_ID = df.pop('userId').values
movie_ID = df.pop('movieId').values
Rating = df.pop('rating').values
sparsity = len(Rating)/(len(set(movie_ID))*len(set(user_ID)))

print('%.10f'%sparsity)
```

```
0.0169996831
```

B Plot a histogram showing the frequency of the rating values: Bin the raw 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 that fall within each bin and use this count as the height of the vertical axis for that particular bin. Comment on the shape of the histogram.

In [8]:
```
binwidth=0.5
plt.hist(df["rating"],bins = np.arange(min(df.rating), max(df.rating) + binwidth,step = binwidth),edgeco

plt.xlabel('rating score')
plt.ylabel('frequency of ratings')
```

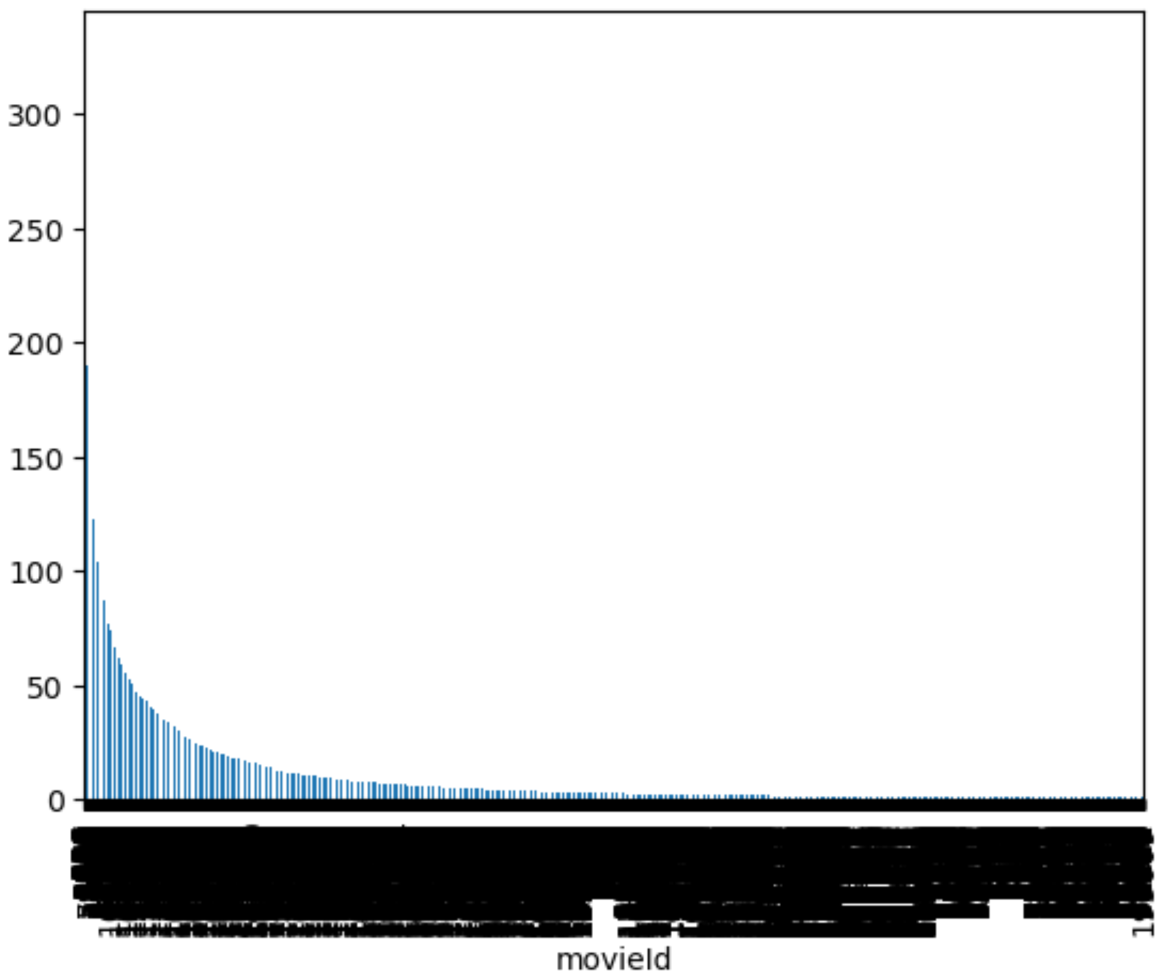
Out[8]: Text(0, 0.5, 'frequency of ratings')



C Plot the distribution of the number of ratings received among movies: 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; ties can broken in any way. A monotonically decreasing trend is expected.

In [9]:
```
movie_count = df.groupby('movieId').size().sort_values(ascending=False)
movie_count.plot(kind='bar',y='number of ratings movie recieved',x='Id')
```

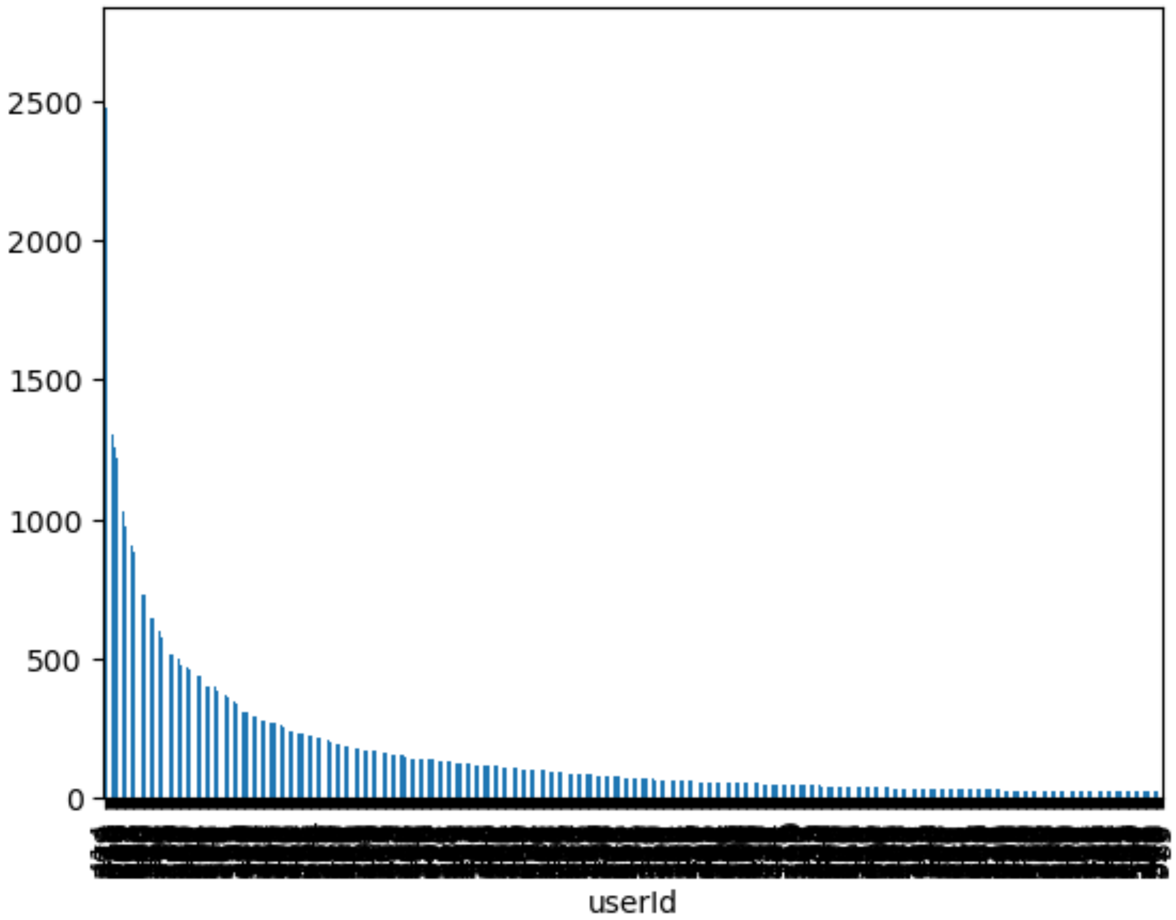
Out[9]: <AxesSubplot:xlabel='movieId'>



D Plot the distribution of ratings among users: The X-axis should be the user index ordered by decreasing frequency and the Y -axis should be the number of movies the user has rated. The requirement of the plot is similar to that in Question C.

In [10]:
```
user_count = df.groupby('userId').size().sort_values(ascending=False)
user_count.plot(kind='bar',y='number of movies rated',x='UserId')
```

Out[10]: <AxesSubplot:xlabel='userId'>



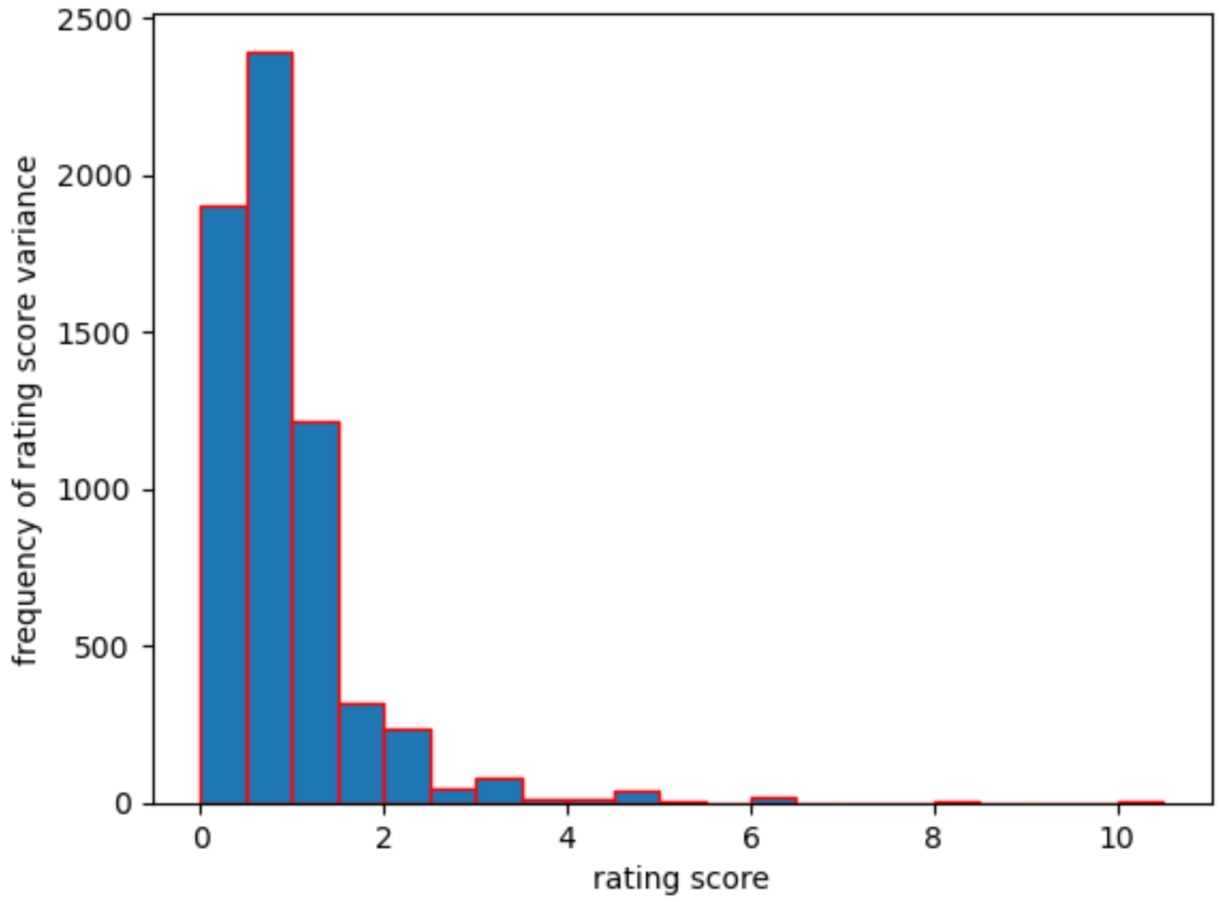
E Discuss the salient features of the distributions from Questions C,D and their implications for the recommendation process.

A: Some movies has recieved many ratings and many user has rated many movies. This means that the movie that recieved many ratings and the users that rated many movies may be important to the recommendation system. And may be we can waive the movies that has little ratings.

F Compute the variance of the rating values received by each movie: 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 resulting histogram

In [11]:
```
var_movie = df.groupby('movieId').var()
binwidth=0.5
plt.hist(var_movie["rating"],bins = np.arange(min(var_movie.rating), max(var_movie.rating) + binwidth,st

plt.xlabel('rating score')
plt.ylabel('frequency of rating score variance')
```

Out[11]: Text(0, 0.5, 'frequency of rating score variance')



QUESTION 2: Understanding the Pearson Correlation Coefficient:

A Write down the formula for μ_u in terms of I_u and r_{uk};

$$\mu_u = \frac{\sum r_{uk}}{\sum I_u}$$

B In plain words, explain the meaning of Iu ∩Iv. Can Iu ∩Iv = ∅? (Hint: Rating matrix R is sparse)

The meaning of Iu ∩ Iv is that the set of items that both user u and v both rated, it can be zero.

QUESTION 3:

Understanding the Prediction function: Can you explain the reason behind mean-centering the raw ratings (rvj −μv) 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.)

A: The Pearson prediction function is a method for making personalized recommendations by finding the correlation between a target user's ratings and the ratings of other users, and using that correlation to predict how the target user would rate items they haven't seen before.

Mean-centering the raw ratings (rvj - μv) is done in the prediction function to account for differences in user rating scales. Without mean-centering, some users who rate all items highly or poorly can have a disproportionate impact on the correlation calculation.

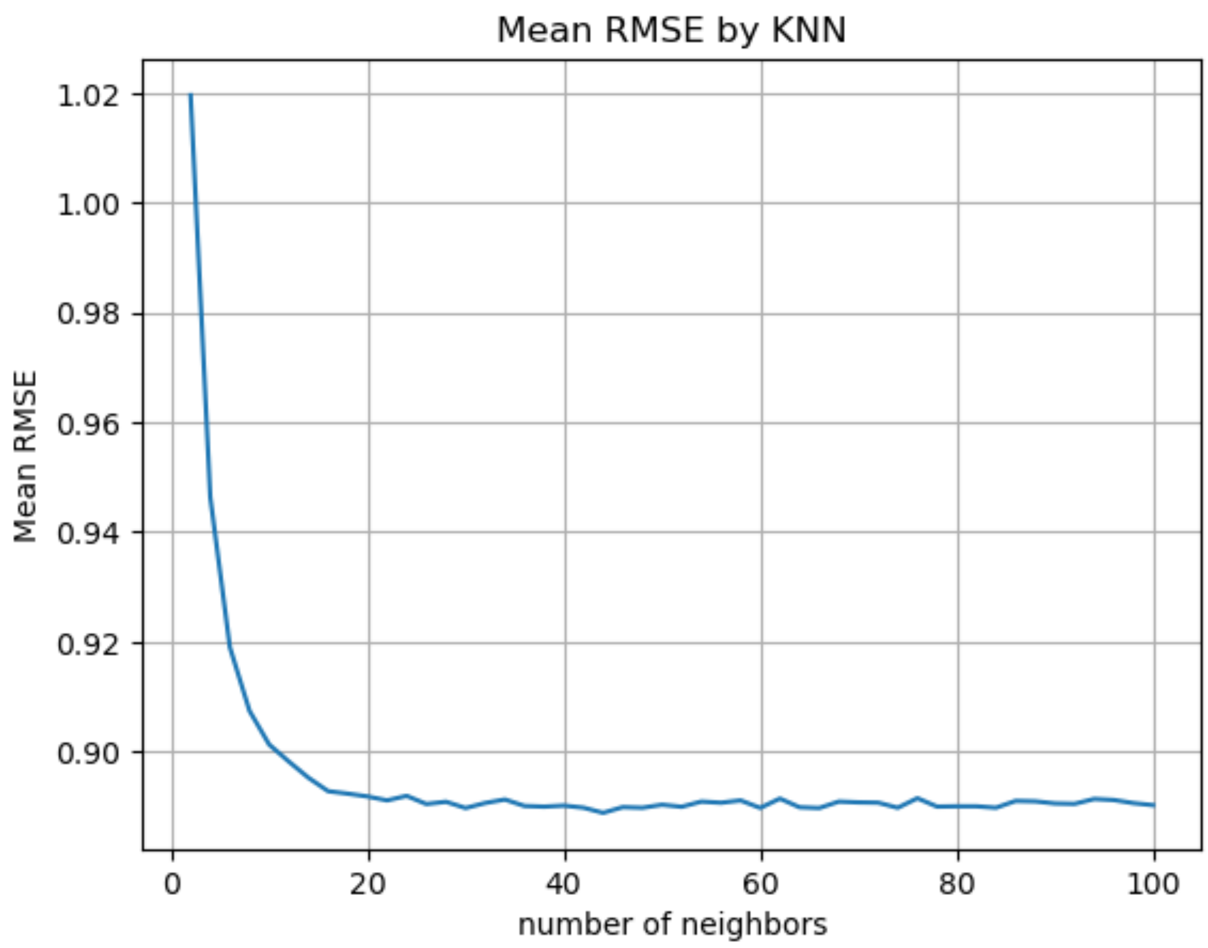
QUESTION 4:

Design a k-NN collaborative filter to predict the ratings of the movies in the original dataset and evaluate its 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)

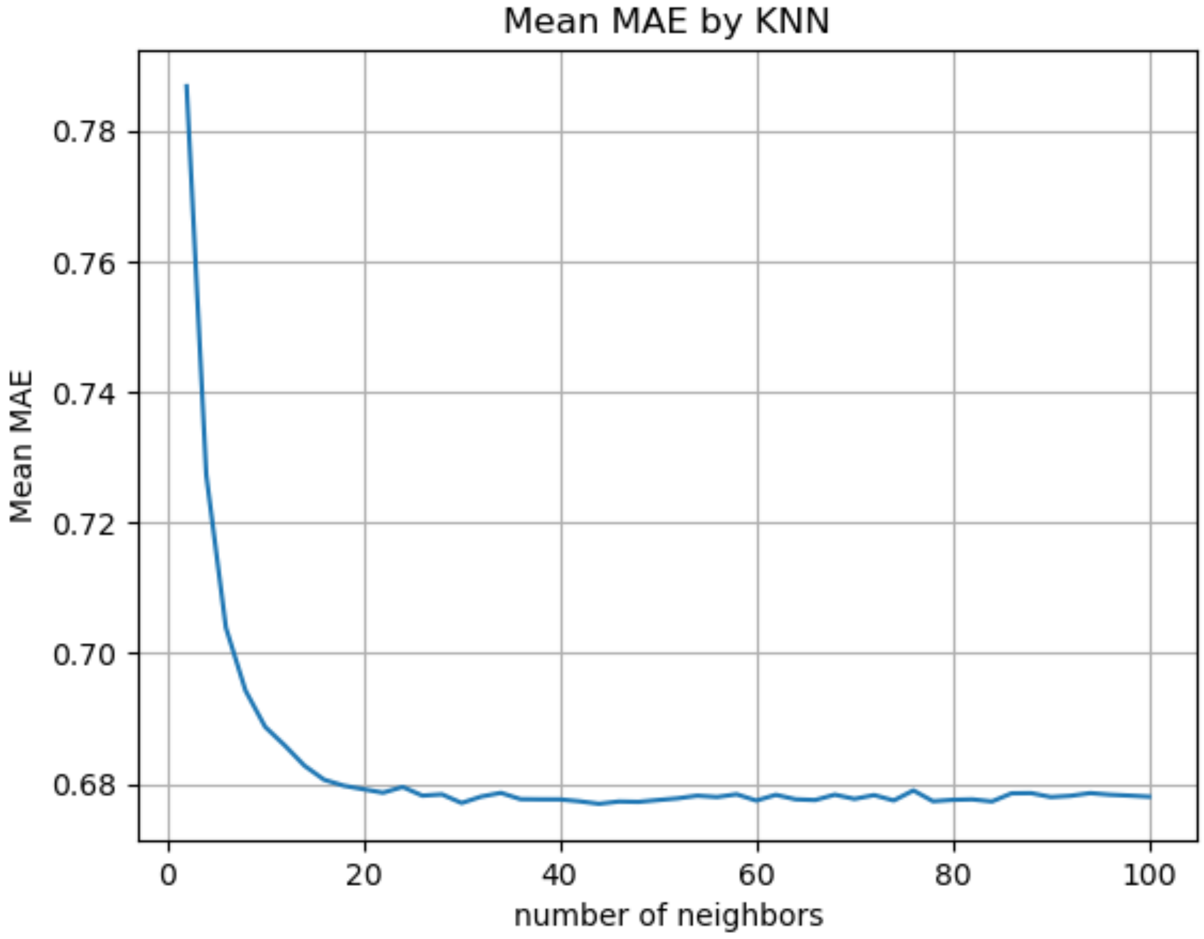
In [15]:
```
from surprise import Reader
from surprise import Dataset

reader = Reader(rating_scale=(0.5, 5))
rating = Dataset.load_from_df(df[['userId','movieId','rating']], reader)
```

In [20]:
```
RMSE = []
MAE = []

for i in range(2,102,2):
   result = cross_validate(KNNWithMeans(k=i, sim_options={'name':'pearson'},verbose=False), data=rating,
   RMSE.append(np.mean(result['test_rmse']))
   MAE.append(np.mean(result['test_mae']))
```

In [22]:
```
KNNrange = np.arange(2, 102, 2)
fig, ax = plt.subplots()
ax.plot(KNNrange, RMSE)

ax.set(xlabel='number of neighbors', ylabel='Mean RMSE',
       title='Mean RMSE by KNN')
ax.grid()


plt.show()
```



In [23]:
```
fig, ax = plt.subplots()
ax.plot(KNNrange, MAE)

ax.set(xlabel='number of neighbors', ylabel='Mean MAE',
       title='Mean MAE by KNN')
ax.grid()


plt.show()
```



QUESTION 5:

Use the plot from question 4, 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.

A: From the graph above, we can observe that both average MAE and RMSE converge when K = 20

In [29]:
```
print("Mean RMSE at K = 20 : ",RMSE[10])
print("Mean MAE at K = 20 : " ,MAE[10] )
```

```
Mean RMSE at K = 20 :  0.8910846957698014
Mean MAE at K = 20 :  0.6786256724713728
```

QUESTION 6:

Within EACH of the 3 trimmed subsets in the dataset, design (train and validate): A k-NN collaborative filter on the ratings of the movies (i.e Popular, Unpopular or High-Variance) and evaluate each of the three models' 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 obtained by averaging the RMSE across all 10 folds. Plot average RMSE (Y-axis) against k (X-axis). Also, report the minimum average RMSE.

Popular movie trimming

In [56]:

```
movie_ratings = df.groupby('movieId').size()
data_df = df[df['movieId'].isin(movie_ratings.index[movie_ratings >= 3])]

# Select the required columns
data_df = data_df[['userId', 'movieId', 'rating']]

# Define the rating scale
reader = Reader(rating_scale=(0.5, 5))

# Load the data from the dataframe
data = Dataset.load_from_df(data_df, reader)
# Define the range of k values to sweep
k_values = range(2, 101, 2)

# Compute the average RMSE for each k using 10-fold cross-validation
avg_rmse_list = []
for k in k_values:
    sim_options = {'name': 'pearson', 'user_based': True}
    algo = KNNWithMeans(k=k, sim_options=sim_options)
    results = cross_validate(algo, data, measures=['RMSE'], cv=10, verbose=False)
    avg_rmse = np.mean(results['test_rmse'])
    avg_rmse_list.append(avg_rmse)

# Plot the average RMSE against k
plt.plot(k_values, avg_rmse_list)
plt.title('Average RMSE vs. k')
plt.xlabel('k')
plt.ylabel('Average RMSE')
plt.show()

# Find the minimum average RMSE and corresponding k value
min_avg_rmse = min(avg_rmse_list)
min_k = k_values[avg_rmse_list.index(min_avg_rmse)]
print(f'Minimum average RMSE: {min_avg_rmse:.4f} (k = {min_k})')
```

```
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
```

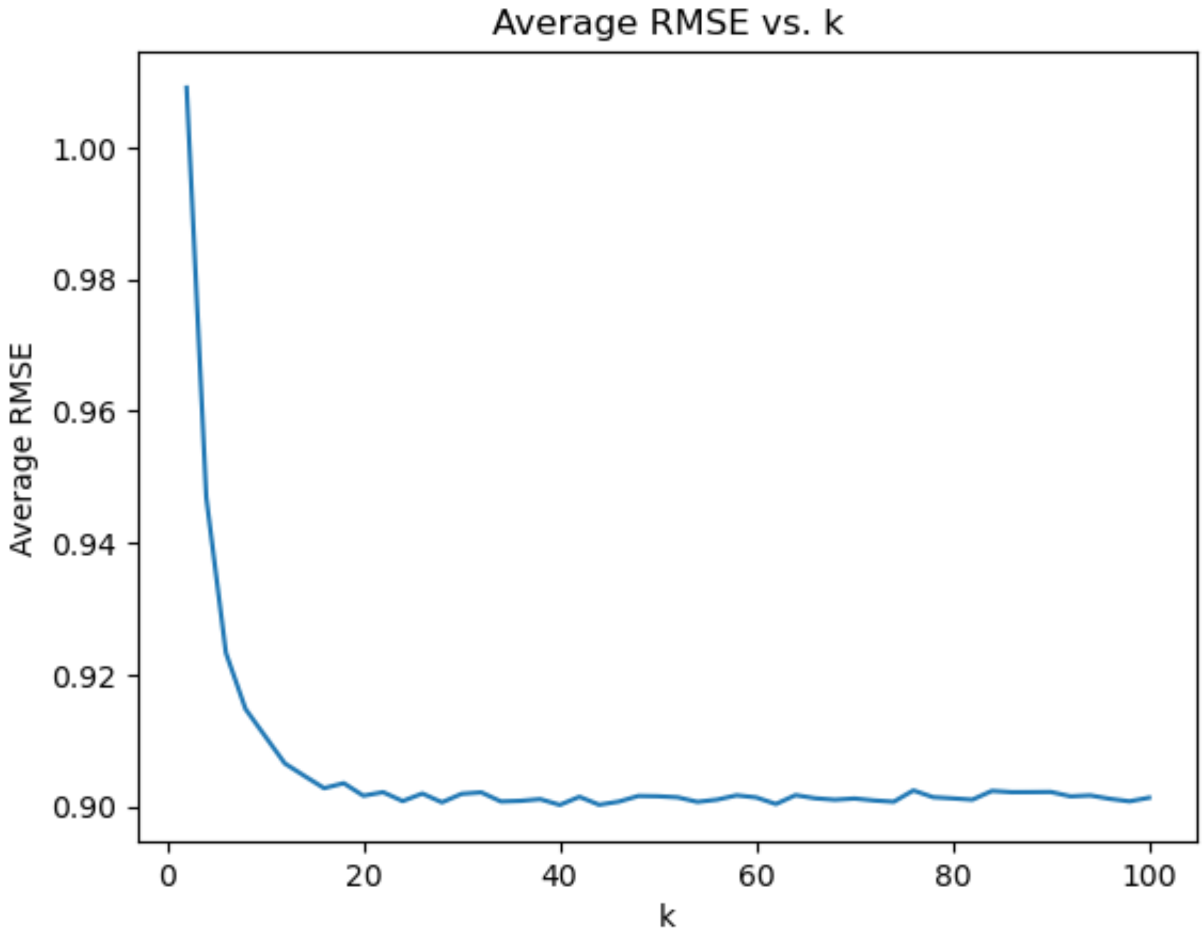
```
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
```

```
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```




```
Minimum average RMSE: 0.9002 (k = 40)
```

Unpopular movie trimming

In [61]:

```
# Unpopular movie trimming
# Load data into Surprise dataset
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)

# Unpopular movie trimming
min_movie_ratings = 3
filter_movies = df['movieId'].value_counts() > min_movie_ratings
filter_movies = filter_movies[filter_movies].index.tolist()

# Apply the filter
df = df[df['movieId'].isin(filter_movies)]

# Load trimmed data into Surprise dataset
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
# Define the range of k values to sweep
k_values = range(2, 101, 2)

# Compute the average RMSE for each k using 10-fold cross-validation
avg_rmse_list = []
for k in k_values:
    sim_options = {'name': 'pearson', 'user_based': True}
    algo = KNNWithMeans(k=k, sim_options=sim_options)
    results = cross_validate(algo, data, measures=['RMSE'], cv=10, verbose=False)
    avg_rmse = np.mean(results['test_rmse'])
    avg_rmse_list.append(avg_rmse)

# Plot the average RMSE against k
plt.plot(k_values, avg_rmse_list)
plt.title('Average RMSE vs. k')
plt.xlabel('k')
plt.ylabel('Average RMSE')
plt.show()

# Find the minimum average RMSE and corresponding k value
min_avg_rmse = min(avg_rmse_list)
min_k = k_values[avg_rmse_list.index(min_avg_rmse)]
print(f'Minimum average RMSE: {min_avg_rmse:.4f} (k = {min_k})')
```

```
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.
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.
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.
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.
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.
```

```
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.

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...
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
```

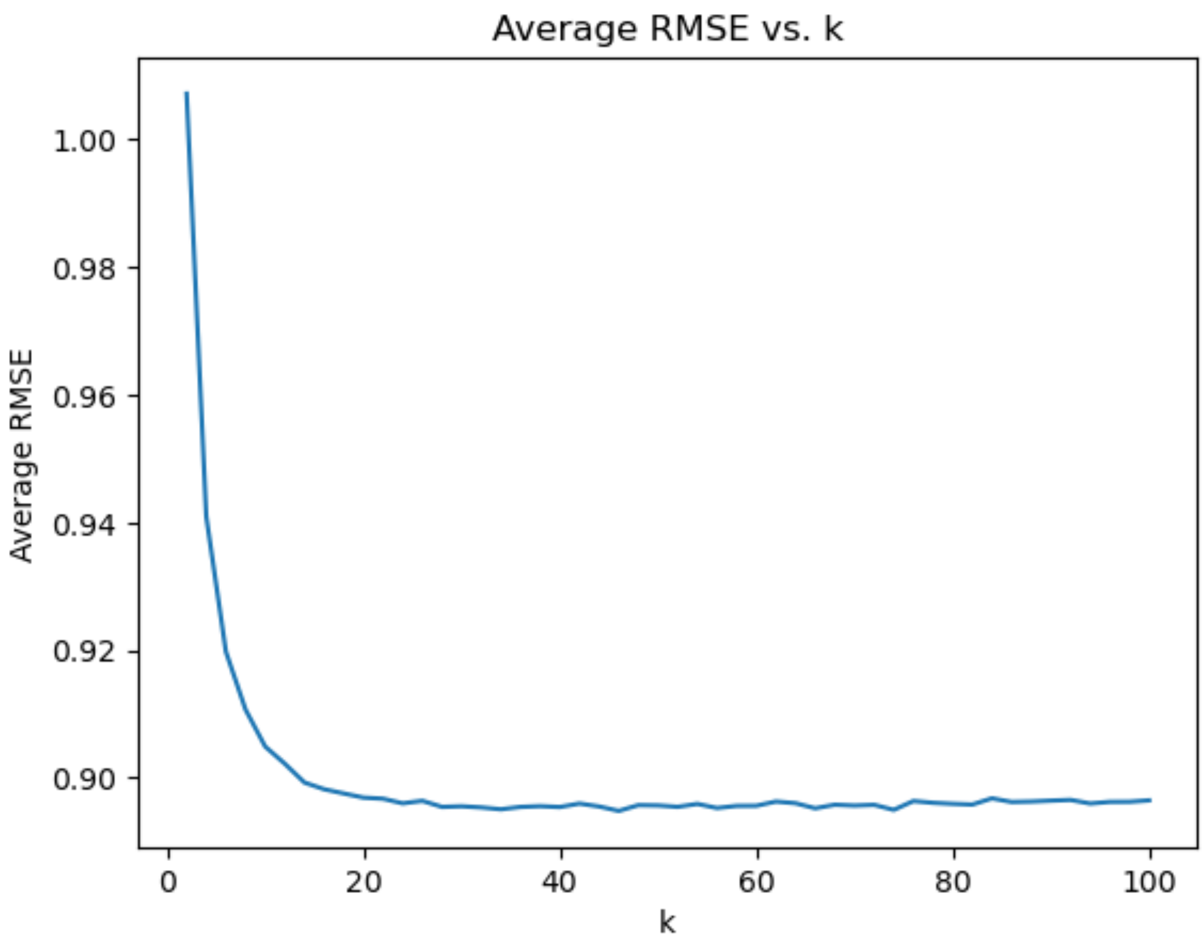
```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.

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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.

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.
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.
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.
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.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```




```
Minimum average RMSE: 0.8948 (k = 46)
```

High variance trimming

In [60]:

```
# Load data from CSV and select only relevant columns
data_df = df[['userId', 'movieId', 'rating']]

# Count the number of ratings per movie
movie_counts = data_df['movieId'].value_counts()

# Compute the variance of ratings for each movie
movie_variances = data_df.groupby('movieId')['rating'].var()

# Trim the dataset to contain only movies with at least 5 ratings and variance >= 2
movie_ids_to_keep = movie_counts[(movie_counts >= 5) & (movie_variances >= 2)].index
trimmed_df = data_df[data_df['movieId'].isin(movie_ids_to_keep)]

# Load data into Surprise Dataset object
reader = Reader()
data = Dataset.load_from_df(trimmed_df, reader)
# Compute the average RMSE for each k using 10-fold cross-validation
avg_rmse_list = []
for k in k_values:
    sim_options = {'name': 'pearson', 'user_based': True}
    algo = KNNWithMeans(k=k, sim_options=sim_options)
    results = cross_validate(algo, data, measures=['RMSE'], cv=10, verbose=False)
    avg_rmse = np.mean(results['test_rmse'])
    avg_rmse_list.append(avg_rmse)

# Plot the average RMSE against k
plt.plot(k_values, avg_rmse_list)
plt.title('Average RMSE vs. k')
plt.xlabel('k')
plt.ylabel('Average RMSE')
plt.show()

# Find the minimum average RMSE and corresponding k value
min_avg_rmse = min(avg_rmse_list)
min_k = k_values[avg_rmse_list.index(min_avg_rmse)]
print(f'Minimum average RMSE: {min_avg_rmse:.4f} (k = {min_k})')
```

```
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.

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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
Computing the pearson similarity matrix...
```

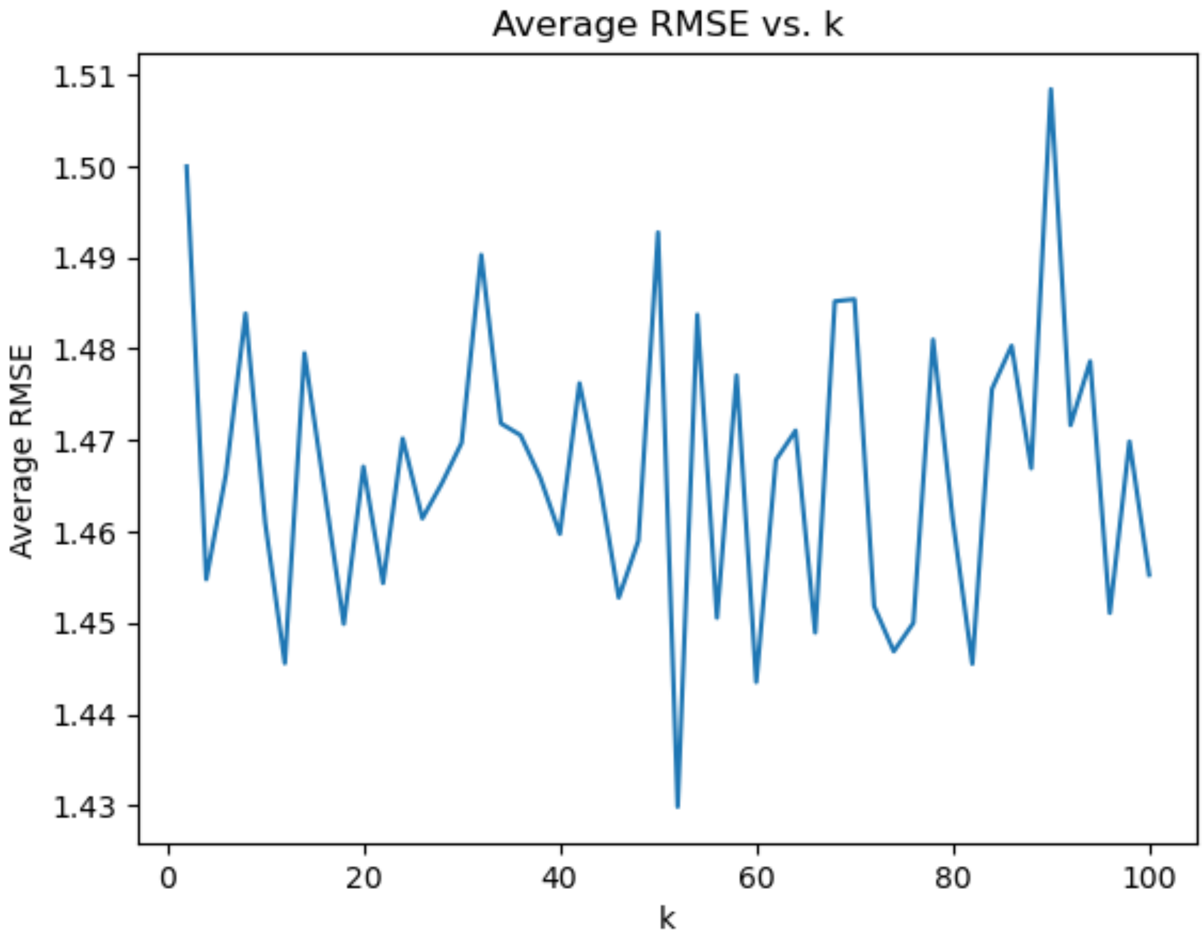
```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
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.
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.
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.
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.
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.
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.
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.
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.
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...
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```




```
Minimum average RMSE: 1.4299 (k = 52)
```

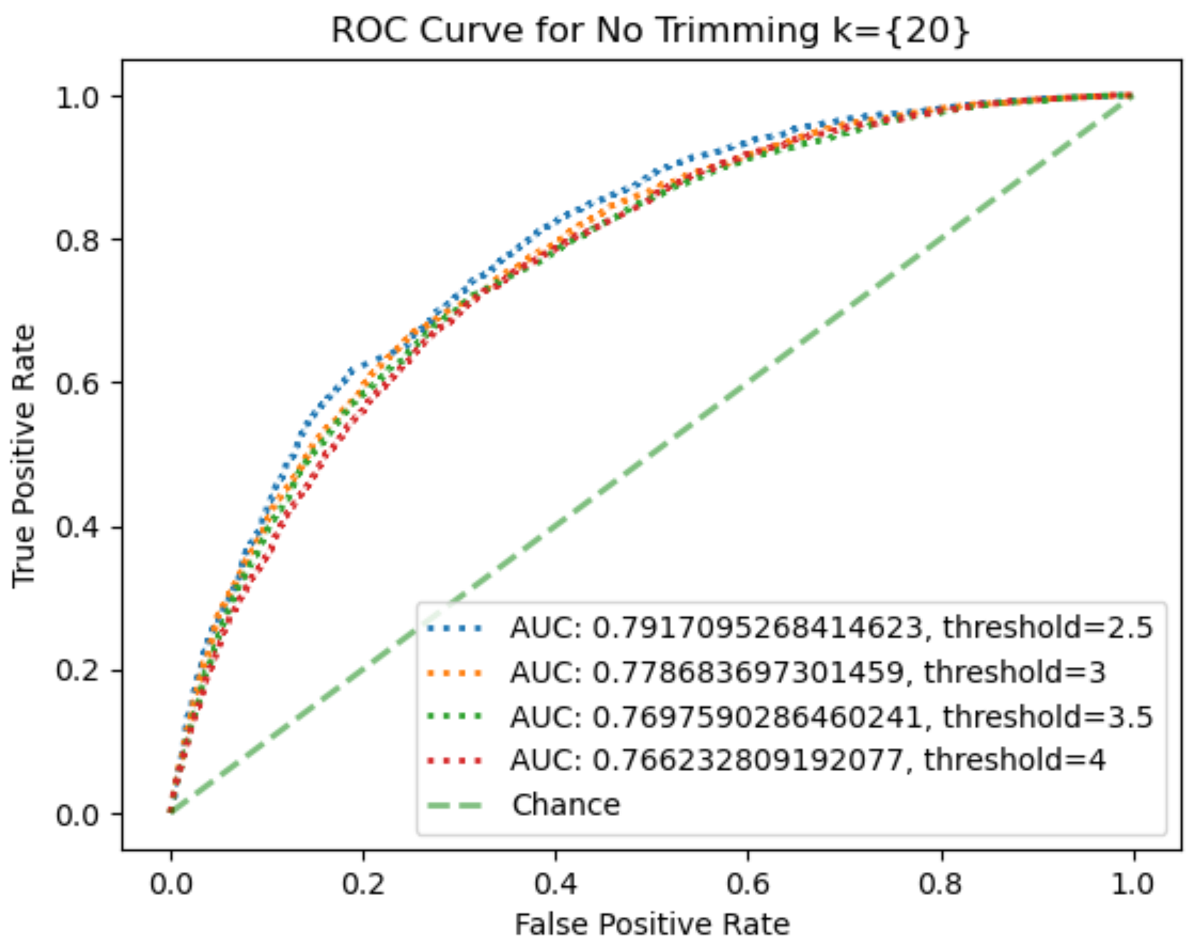
Plot the ROC curves for the k-NN collaborative filters for threshold values [2.5,3,3.5,4]. These thresholds are applied only on the ground truth labels in held-out validation set. For each of the plots, also report the area under the curve (AUC) value. You should have 4 ×4 plots in this section (4 trimming options – including no trimming times 4 thresholds) - all thresholds can be condensed into one plot per trimming option yielding only 4 plots.

In [283]:
```
def popular_trim(testset: List[Tuple[int, int, float]]) -> List[Tuple[int, int, float]]:
    # Count the number of ratings for each movie
    movie_ratings = defaultdict(int)
    for (uid, iid, rating) in testset:
        movie_ratings[iid] += 1

    # Identify popular movies
    popular_movies = set(iid for iid, count in movie_ratings.items() if count > 2)

    # Filter out unpopular movies from the test set
    return [t for t in testset if t[1] in popular_movies]

def unpopular_trim(testset: List[Tuple[int, int, float]]) -> List[Tuple[int, int, float]]:
    # Count the number of ratings for each movie
    movie_ratings = defaultdict(int)
    for (uid, iid, rating) in testset:
        movie_ratings[iid] += 1

    # Identify unpopular movies
    unpopular_movies = set(iid for iid, count in movie_ratings.items() if count <= 2)

    # Filter out popular movies from the test set
    return [t for t in testset if t[1] in unpopular_movies]


def high_var_trim(testset: List[Tuple[int, int, float]]) -> List[Tuple[int, int, float]]:
    # Group ratings by movie
    movie_ratings = defaultdict(list)
    for (uid, iid, rating) in testset:
        movie_ratings[iid].append(rating)

    # Identify high variance movies
    high_variance_movies = set(iid for iid, ratings in movie_ratings.items()
                               if len(ratings) >= 5 and np.var(ratings) >= 2)

    # Filter out low variance movies from the test set
    return [t for t in testset if t[1] in high_variance_movies]
```

In [275]:

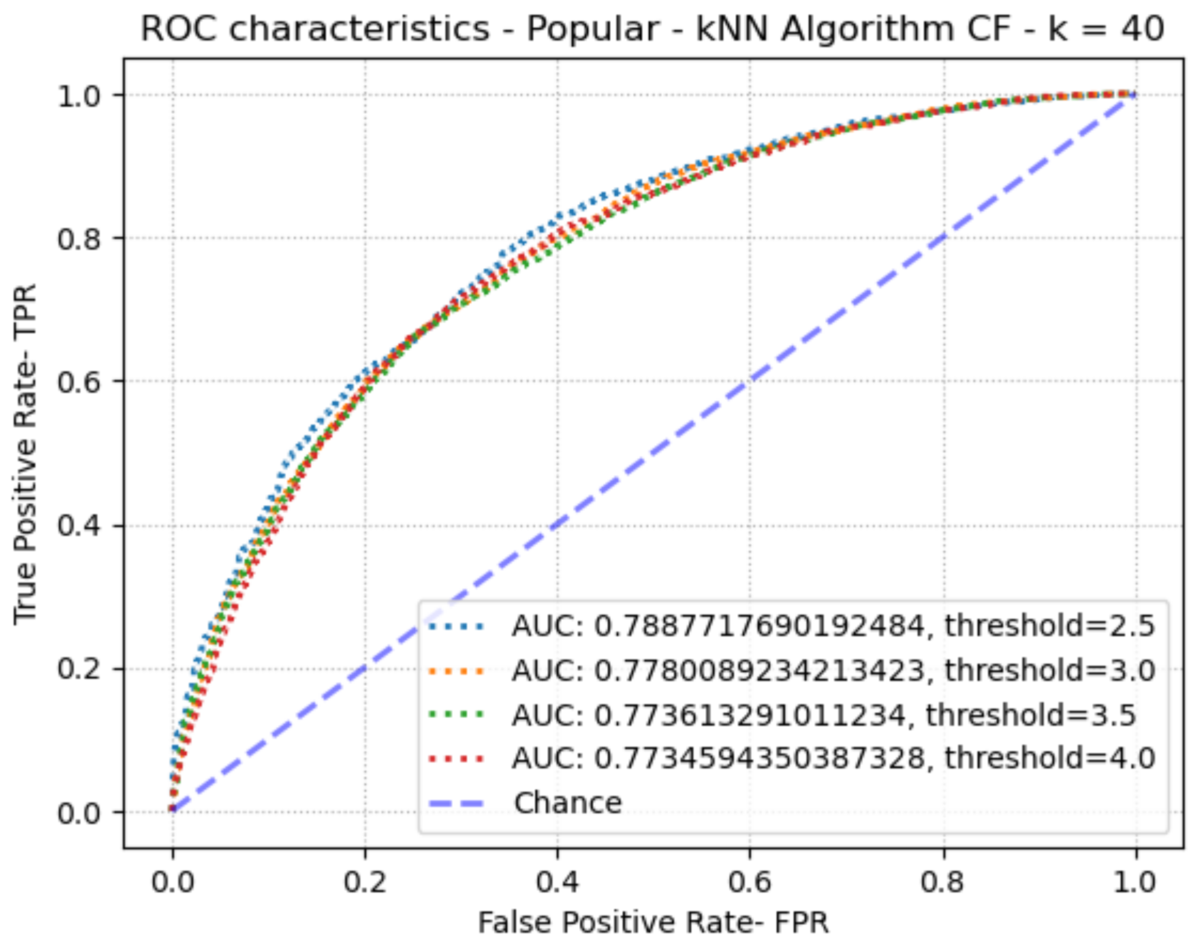
```
# Load the data

reader = Reader(rating_scale=(0.5, 5.0))


# Define the k-NN collaborative filtering algorithm and train the models

thresholds = [2.5, 3, 3.5, 4]

data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
algo = KNNWithMeans(k=20, sim_options={'name': 'pearson', 'user_based': True})
res = algo.fit(trainset).test(testset)
print("Results for No Trimming with k=20")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve for No Trimming k={20}')
plt.legend(loc="lower right")
plt.show()
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Results for No Trimming with k=20
```



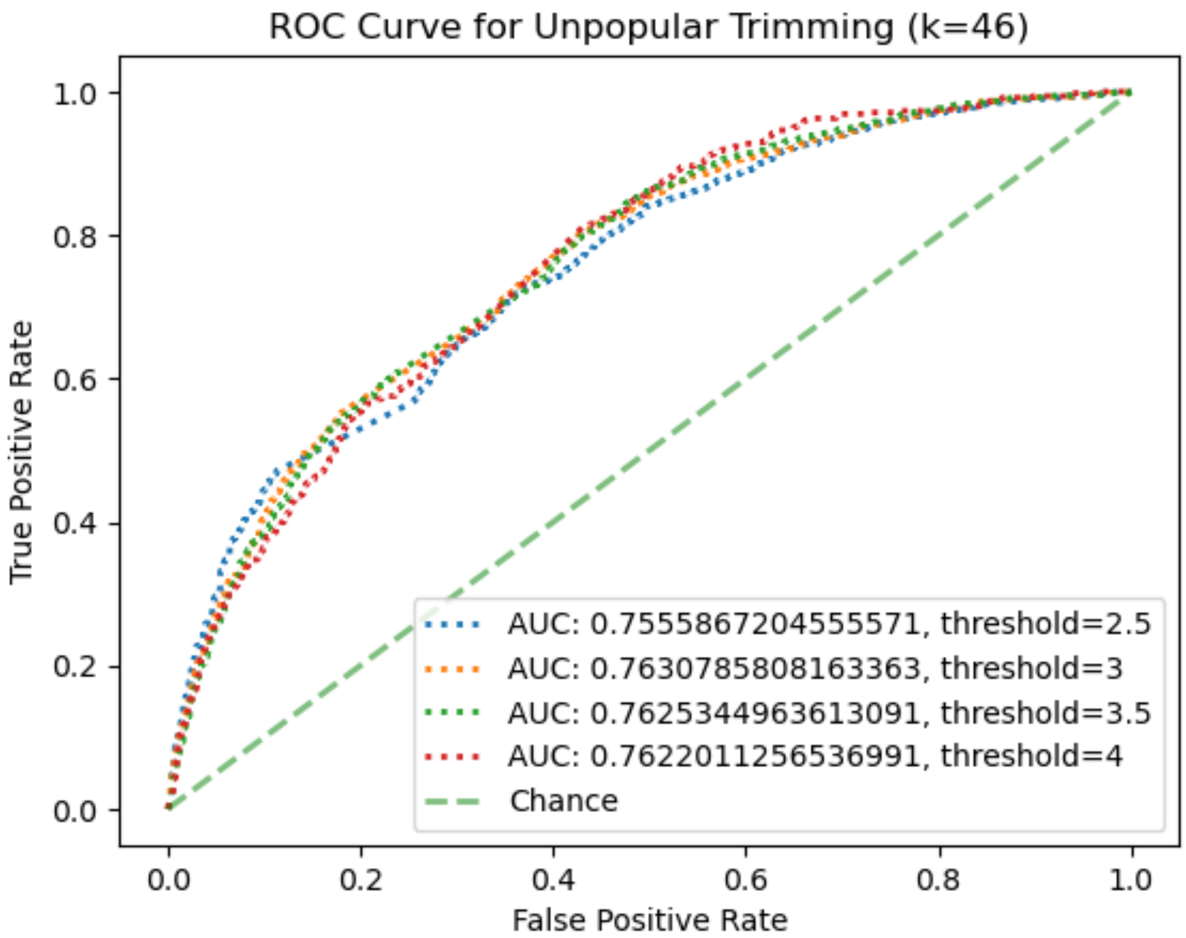
In [272]:

```python
from typing import List, Tuple
from collections import defaultdict

def popular_movie_trimming(testset: List[Tuple[int, int, float]]) -> List[Tuple[int, int, float]]:
    # Count the number of ratings for each movie
    movie_ratings = defaultdict(int)
    for (uid, iid, rating) in testset:
        movie_ratings[iid] += 1

    # Identify popular movies
    popular_movies = set(iid for iid, count in movie_ratings.items() if count > 2)

    # Filter out unpopular movies from the test set
    return [t for t in testset if t[1] in popular_movies]
reader = Reader(rating_scale=(0.5, 5.0))
Dataset_Ratings= Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)



# Define the k-NN collaborative filtering algorithm and train the models

thresholds = [2.5, 3, 3.5, 4]


k = 40
Train_list, Test_list = train_test_split(Dataset_Ratings, test_size=0.1)
Thres_list = [2.5, 3.0, 3.5, 4.0]

Pop_Trimmed_set = popular_movie_trimming(Test_list)
res  = KNNWithMeans(k=k,sim_options={'name':'pearson'},verbose=False).fit(Train_list).test(Pop_Trimmed_s

fig, ax = plt.subplots()
for item in Thres_list:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    FPR, TPR, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(FPR, TPR,lw=2,linestyle=':',label="AUC: "+str(auc(FPR,TPR))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='b', label='Chance', alpha=.5)
plt.legend(loc='best')
plt.grid(linestyle=':')
plt.title('ROC characteristics - Popular - kNN Algorithm CF - k = 40')
plt.ylabel('True Positive Rate- TPR')
plt.xlabel('False Positive Rate- FPR')
plt.savefig('Q6a.png',dpi=350,bbox_inches='tight')
plt.show()
```



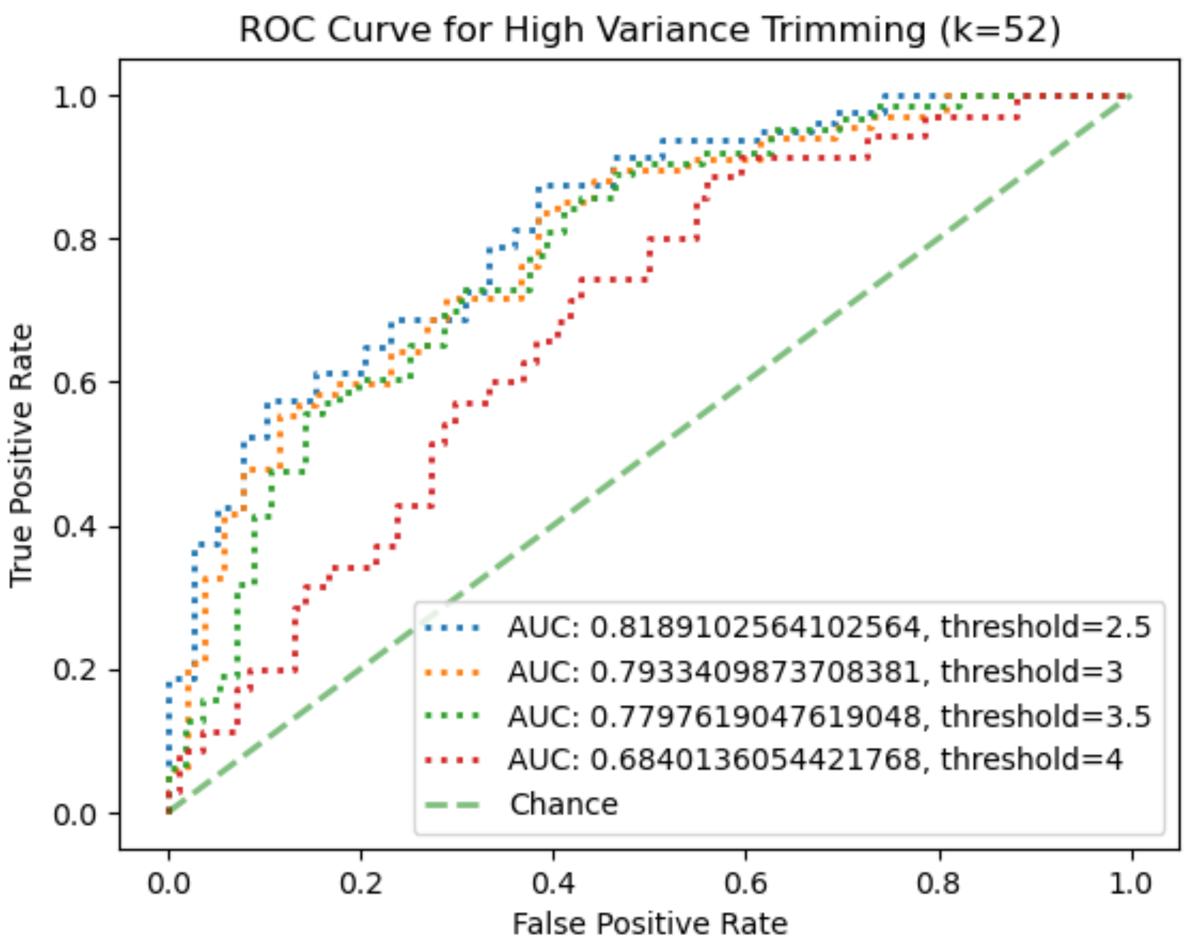
In [277]:
```
#unpopular trimmming

thresholds = [2.5, 3, 3.5, 4]

data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
algo = KNNWithMeans(k=46, sim_options={'name': 'pearson', 'user_based': True})
res = algo.fit(trainset).test(unpopular_trim(testset))
print("Results for unpopular trimming'} with k=46")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve for Unpopular Trimming (k=46)')
plt.legend(loc="lower right")
plt.show()
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Results for unpopular trimming'} with k=46
```



In [280]:
```
#high variance
thresholds = [2.5, 3, 3.5, 4]
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
algo = KNNWithMeans(k=52, sim_options={'name': 'pearson', 'user_based': True})
res = algo.fit(trainset).test(high_var_trim(testset))
print("Results for high vairiance trimming'} with k=52")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve for High Variance Trimming (k=52)')
plt.legend(loc="lower right")
plt.show()
```

```
Computing the pearson similarity matrix...
Done computing similarity matrix.
Results for high vairiance trimming'} with k=52
```



QUESTION 7:

Understanding the NMF cost function: 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

Convexity of Optimization Problem and Least-Squares Formulation

A:

The optimization problem denoted by the equation is not jointly convex with respect to the variables U and V because it is a non-convex optimization problem.

The objective function is a sum of convex functions, each of the form $(r_{ij} - (UV^T)_{ij})^2$, which are convex in either U or V when the other is fixed. However, the product of two convex functions is not necessarily convex. In this case, the product UV^T is a bilinear function of the variables U and V, which is not a convex function.

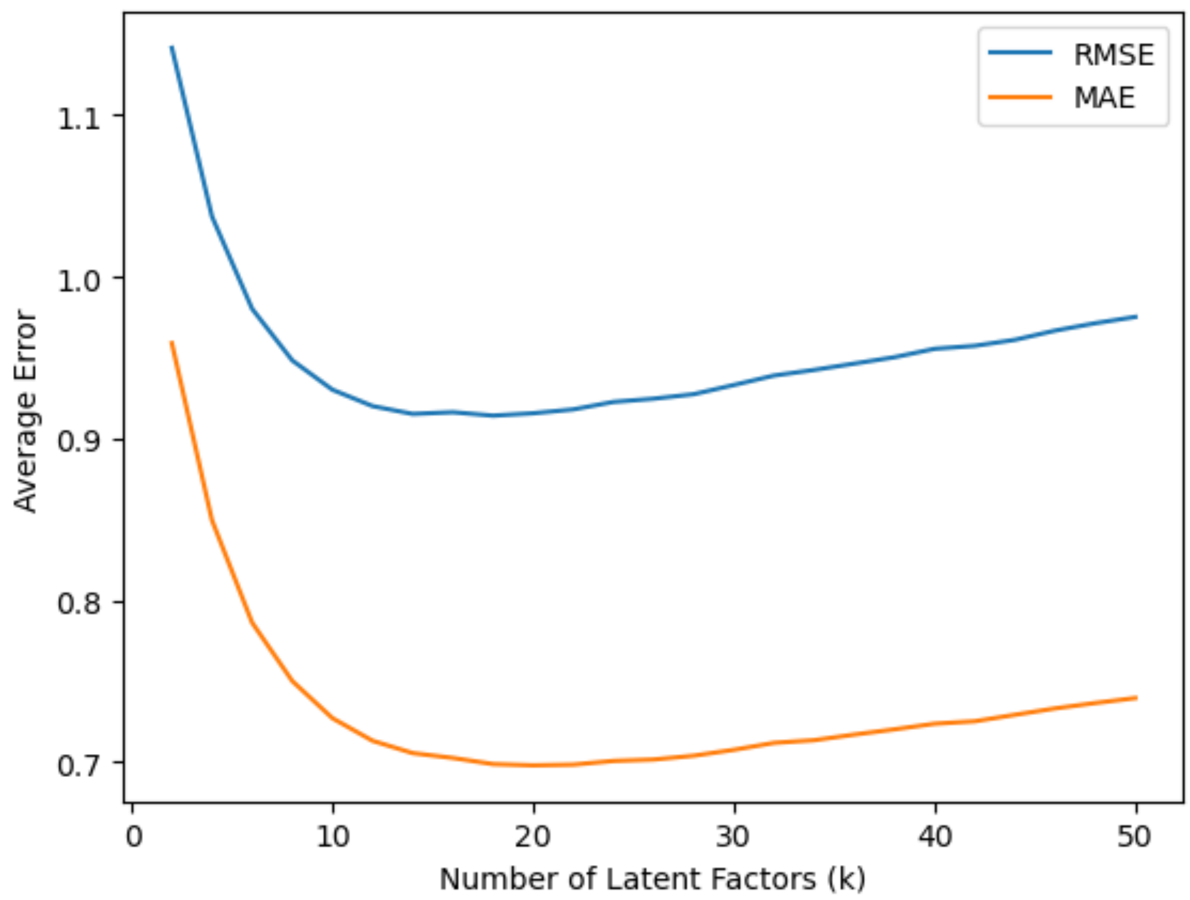
Therefore, the objective function is a non-convex function of both U and V, and the optimization problem is not jointly convex.

Question 8 Designing the NMF Collaborative Filter:

A

Design a NMF-based collaborative filter to predict the ratings of the movies in the original dataset and evaluate its 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. If NMF takes too long, you can increase the step size. Increasing it too much will result in poorer granularity in your results. 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 [163]:
```
from surprise import NMF, Dataset, Reader
from surprise.model_selection import cross_validate
import numpy as np
import matplotlib.pyplot as plt

# load the dataset
reader = Reader(rating_scale=(0.5, 5))
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)

# define the range of k values to test
k_values = np.arange(2, 51, 2)

# perform 10-fold cross-validation with NMF algorithm
rmse_scores = []
mae_scores = []
for k in k_values:
    nmf = NMF(n_factors=k)
    results = cross_validate(nmf, data, measures=['RMSE', 'MAE'], cv=10, verbose=False)
    rmse_scores.append(np.mean(results['test_rmse']))
    mae_scores.append(np.mean(results['test_mae']))

# plot the results
plt.plot(k_values, rmse_scores, label='RMSE')
plt.plot(k_values, mae_scores, label='MAE')
plt.xlabel('Number of Latent Factors (k)')
plt.ylabel('Average Error')
plt.legend()
plt.show()
```



B

Use the plot from the previous part 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?

A: The optimal number of latent factors is 18 for RSME and 20 for MAE, which is really precise to the number of movie genres.

In [184]:

```
# find the optimal number of latent factors (k)
min_rmse = np.min(rmse_scores)
min_mae = np.min(mae_scores)
optimal_k_rmse = k_values[np.argmin(rmse_scores)]
optimal_k_mae = k_values[np.argmin(mae_scores)]
print('Optimal Number of Latent Factors (RMSE):', optimal_k_rmse)
print('Minimum Average RMSE:', min_rmse)
print('Optimal Number of Latent Factors (MAE):', optimal_k_mae)
print('Minimum Average MAE:', min_mae)

# count genre numbers
df_movie = pd.read_csv('/Users/ryan/Downloads/Synthetic_Movie_Lens/movies.csv');
df_movie['genres']

arr_of_genres = []
for i in df_movie['genres']:
  for j in i.split('|'):
    if j not in arr_of_genres:
      arr_of_genres.append(j)

print(arr_of_genres)
print(len(arr_of_genres))
```

```
Optimal Number of Latent Factors (RMSE): 18
Minimum Average RMSE: 0.9142137631676837
Optimal Number of Latent Factors (MAE): 20
Minimum Average MAE: 0.6980593351220484
['Adventure', 'Animation', 'Children', 'Comedy', 'Fantasy', 'Romance', 'Drama', 'Action', 'Crime', 'Thr
iller', 'Horror', 'Mystery', 'Sci-Fi', 'War', 'Musical', 'Documentary', 'IMAX', 'Western', 'Film-Noir',
'(no genres listed)']
20
```

C

Performance on trimmed dataset subsets: For each of Popular, Unpopular and High- Variance subsets - – Design a NMF collaborative filter for each trimmed subset and evaluate its 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 obtained by averaging the RMSE across all 10 folds.

– Plot average RMSE (Y-axis) against k (X-axis); item Report the minimum average RMSE.

In [173]:

```
# load the datasets
popular_df = popular_trim(df)
unpopular_df = unpopular_trim(df)
high_var_df = high_var_trim(df)

# convert the datasets to surprise format
popular_data = Dataset.load_from_df(popular_df[['userId', 'movieId', 'rating']], reader)
unpopular_data = Dataset.load_from_df(unpopular_df[['userId', 'movieId', 'rating']], reader)
high_var_data = Dataset.load_from_df(high_var_df[['userId', 'movieId', 'rating']], reader)
# define the range of k values
ks = range(2, 51, 2)

# define the NMF algorithm with default values


# define dictionaries to store the RMSE values for each subset
popular_rmse = {}
unpopular_rmse = {}
high_var_rmse = {}

# perform 10-fold cross-validation with NMF algorithm for each subset and each k
for k in ks:
    nmf=NMF(n_factors=k)
    popular_results = cross_validate(nmf, popular_data, measures=['RMSE'], cv=10, verbose=True)
    unpopular_results = cross_validate(nmf, unpopular_data, measures=['RMSE'], cv=10, verbose=True)
    high_var_results = cross_validate(nmf, high_var_data, measures=['RMSE'], cv=10, verbose=True)

    # store the average RMSE values for each subset and each k
    popular_rmse[k] = np.mean(popular_results['test_rmse'])
    unpopular_rmse[k] = np.mean(unpopular_results['test_rmse'])
    high_var_rmse[k] = np.mean(high_var_results['test_rmse'])

# plot the average RMSE against k for each subset
plt.plot(ks, list(popular_rmse.values()), label='Popular')
plt.plot(ks, list(unpopular_rmse.values()), label='Unpopular')
plt.plot(ks, list(high_var_rmse.values()), label='High Variance')
plt.xlabel('Number of Latent Factors (k)')
plt.ylabel('Average RMSE')
plt.legend()
plt.show()

# print the minimum average RMSE for each subset
print("Minimum Average RMSE for Popular Subset:", min(popular_rmse.values()))
print("Minimum Average RMSE for Unpopular Subset:", min(unpopular_rmse.values()))
print("Minimum Average RMSE for High Variance Subset:", min(high_var_rmse.values()))
```

```
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.1450  1.1278  1.1451  1.1271  1.1332  1.1368  1.1491  1.1384  1.1447  1.1563  1.140
3  0.0089
Fit time          1.41    1.52    1.49    1.12    1.12    1.13    1.07    1.07    1.06    1.05    1.20
0.18
Test time         0.27    0.26    0.28    0.25    0.23    0.28    0.25    0.24    0.24    0.25    0.25
0.02
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.1438  1.1490  1.1406  1.1449  1.1314  1.1473  1.1444  1.1515  1.1369  1.1436  1.143
3  0.0056
Fit time          1.07    1.17    1.08    1.07    1.08    1.12    1.28    1.06    1.07    1.08    1.11
0.07
Test time         0.24    0.25    0.24    0.24    0.24    0.24    0.24    0.25    0.23    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.8681  1.7544  1.7915  1.6207  1.7553  1.6740  1.6781  1.7573  1.6829  1.5350  1.711
7  0.0892
Fit time          0.02    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
```

```
Std
RMSE (testset)    1.0375  1.0345  1.0292  1.0334  1.0446  1.0453  1.0401  1.0412  1.0416  1.0349  1.038
2  0.0049
Fit time          1.18    1.19    1.20    1.17    1.22    1.21    1.22    1.18    1.20    1.19    1.20
0.02
Test time         0.25    0.24    0.24    0.22    0.24    0.24    0.24    0.24    0.24    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.0333  1.0561  1.0304  1.0311  1.0412  1.0471  1.0412  1.0281  1.0421  1.0405  1.039
1  0.0082
Fit time          1.41    1.20    1.29    1.19    1.26    1.22    1.23    1.20    1.21    1.38    1.26
0.08
Test time         0.25    0.79    0.24    0.24    0.24    0.24    0.24    0.23    0.24    0.25    0.30
0.16
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.6289  1.5829  1.6777  1.7699  1.5747  1.6275  1.6027  1.6221  1.9926  1.6116  1.669
1  0.1202
Fit time          0.02    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9731  0.9762  0.9835  0.9784  0.9969  0.9950  0.9752  0.9797  0.9872  0.9897  0.983
5  0.0080
Fit time          1.35    1.33    1.33    1.36    1.37    1.31    1.34    1.31    1.43    1.32    1.35
0.03
Test time         0.24    0.24    0.24    0.25    0.23    0.25    0.23    0.24    0.23    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9778  0.9929  0.9867  0.9852  0.9712  0.9679  0.9686  0.9853  0.9906  0.9802  0.980
6  0.0086
Fit time          1.36    1.33    1.48    1.33    1.36    1.34    1.39    1.36    1.36    1.35    1.36
0.04
Test time         0.24    0.24    0.24    0.24    0.24    0.24    0.25    0.23    0.25    0.79    0.30
0.17
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.7821  1.5753  1.6807  1.7892  1.8513  1.6044  1.5273  1.5514  1.5265  1.5574  1.644
6  0.1159
Fit time          0.02    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9484  0.9579  0.9566  0.9453  0.9354  0.9487  0.9537  0.9512  0.9565  0.9483  0.950
2  0.0064
Fit time          1.49    1.55    1.48    1.45    1.49    1.62    1.48    1.49    1.47    1.47    1.50
0.05
Test time         0.23    0.24    0.25    0.24    0.24    0.25    0.24    0.24    0.24    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9628  0.9452  0.9399  0.9468  0.9500  0.9497  0.9507  0.9467  0.9493  0.9466  0.948
8  0.0056
Fit time          1.49    1.48    1.50    1.48    1.48    1.47    1.49    1.46    1.67    1.46    1.50
0.06
Test time         0.24    0.24    0.24    0.23    0.25    0.24    0.25    0.25    0.24    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).
```

```
                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.5192  1.6924  1.4742  1.9507  1.5383  1.5281  1.5065  1.7121  1.5022  1.6974  1.612
1  0.1421
Fit time          0.02    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01    0.01
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9326  0.9437  0.9422  0.9254  0.9375  0.9274  0.9107  0.9243  0.9394  0.9166  0.930
0  0.0105
Fit time          1.59    1.63    1.81    1.60    1.61    1.57    1.75    1.59    1.63    1.59    1.64
0.07
Test time         0.24    0.24    0.23    0.23    0.23    0.24    0.26    0.79    0.25    0.24    0.30
0.17
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9326  0.9380  0.9238  0.9182  0.9262  0.9394  0.9294  0.9316  0.9192  0.9256  0.928
4  0.0068
Fit time          1.75    1.60    1.59    1.61    1.77    1.59    1.64    1.58    1.62    1.56    1.63
0.07
Test time         0.24    0.23    0.24    0.24    0.25    0.24    0.24    0.24    0.25    0.25    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.4482  1.7421  1.5210  1.4818  1.6340  1.4943  1.6611  1.6976  1.5257  1.5447  1.575
1  0.0956
Fit time          0.03    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9199  0.9167  0.9215  0.9153  0.9076  0.9168  0.9291  0.9244  0.9271  0.9317  0.921
0  0.0069
Fit time          1.74    1.68    1.76    1.68    1.73    1.74    1.76    1.69    1.75    1.72    1.72
0.03
Test time         0.24    0.24    0.24    0.24    0.24    0.23    0.24    0.24    0.24    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9229  0.9244  0.9211  0.9146  0.9266  0.9169  0.9157  0.9196  0.9239  0.9265  0.921
2  0.0041
Fit time          1.81    1.83    1.77    1.70    1.79    1.70    1.75    1.72    1.89    1.71    1.77
0.06
Test time         0.24    0.26    0.24    0.24    0.25    0.78    0.24    0.24    0.25    0.24    0.30
0.16
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.5578  1.5955  1.7583  1.6636  1.7189  1.6002  1.5457  1.8442  1.3746  1.6616  1.632
0  0.1233
Fit time          0.03    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00

Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9226  0.9156  0.9190  0.9257  0.9146  0.9060  0.9184  0.9087  0.9121  0.9125  0.915
5  0.0058
Fit time          1.84    1.85    2.08    1.84    1.85    1.85    2.10    1.85    1.89    1.80    1.90
0.10
Test time         0.23    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).
```

```
                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9170  0.9163  0.9187  0.9113  0.9210  0.9164  0.9026  0.9232  0.9149  0.9144  0.915
6  0.0054
Fit time          2.00    1.87    1.86    1.88    1.95    1.82    1.87    1.84    1.85    1.86    1.88
0.05
Test time         0.24    0.25    0.24    0.24    0.24    0.25    0.23    0.24    0.23    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.4472  1.6702  1.6226  1.5952  1.8398  1.8410  1.5694  1.4423  1.4443  1.5300  1.600
2  0.1412
Fit time          0.03    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9083  0.9133  0.9237  0.9232  0.9238  0.9107  0.9077  0.9211  0.9054  0.9124  0.915
0  0.0069
Fit time          2.02    1.98    1.99    1.96    1.96    2.10    2.06    1.92    1.98    2.05    2.00
0.05
Test time         0.24    0.24    0.25    0.23    0.24    0.25    0.24    0.24    0.24    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9175  0.8997  0.9207  0.9251  0.9186  0.9104  0.9095  0.9167  0.9123  0.9147  0.914
5  0.0067
Fit time          2.26    1.98    1.97    2.00    2.01    2.02    2.19    2.29    2.06    1.96    2.07
0.12
Test time         0.23    0.24    0.24    0.24    0.24    0.23    0.24    0.25    0.24    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.5015  1.4549  1.4761  1.6296  1.5069  1.6740  1.6231  1.6837  1.4965  1.4281  1.547
4  0.0902
Fit time          0.03    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9086  0.9150  0.9202  0.9176  0.9097  0.9242  0.9148  0.9089  0.9297  0.9020  0.915
1  0.0078
Fit time          2.22    2.10    2.10    2.10    2.11    2.11    2.22    2.32    2.24    2.12    2.16
0.08
Test time         0.24    0.25    0.24    0.24    0.24    0.24    0.23    0.24    0.24    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9178  0.9234  0.9095  0.9141  0.9273  0.8988  0.9012  0.9230  0.9164  0.9184  0.915
0  0.0089
Fit time          2.14    2.14    2.40    2.08    2.32    2.13    2.11    2.13    2.22    2.35    2.20
0.11
Test time         0.24    0.25    0.24    0.24    0.26    0.24    0.23    0.24    0.24    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.8529  1.5900  1.5972  1.6349  1.4911  1.5575  1.4932  1.6194  1.5768  1.3766  1.579
0  0.1169
Fit time          0.03    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.02    0.03
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).
```

```
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    0.9157  0.9122  0.9173  0.9206  0.9082  0.8981  0.9169  0.9316  0.9258  0.9058  0.915
2  0.0093  
Fit time          2.25    2.25    2.22    2.19    2.27    2.34    2.24    2.23    2.27    2.21    2.25    
0.04    
Test time         0.24    0.25    0.24    0.23    0.24    0.23    0.24    0.24    0.24    0.23    0.24    
0.01    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    

Std     
RMSE (testset)    0.9203  0.9079  0.9210  0.8981  0.9213  0.9205  0.9195  0.9251  0.9203  0.9110  0.916
5  0.0078  
Fit time          2.30    2.19    2.39    2.21    2.23    2.25    2.31    2.21    2.31    2.36    2.28    
0.06    
Test time         0.24    0.24    0.24    0.25    0.23    0.24    0.25    0.24    0.24    0.80    0.30    
0.17    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    1.5316  1.5946  1.5747  1.5402  1.5449  1.5760  1.8281  1.4628  1.4669  1.4787  1.559
8  0.0998  
Fit time          0.04    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    
0.00    
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    
0.00    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    0.9288  0.9059  0.9206  0.9142  0.9207  0.9080  0.9219  0.9226  0.9247  0.9205  0.918
8  0.0069  
Fit time          2.32    2.36    2.38    2.34    2.37    2.33    2.48    2.31    2.42    2.55    2.39    
0.07    
Test time         0.23    0.23    0.25    0.24    0.24    0.24    0.24    0.24    0.24    0.25    0.24    
0.01    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    0.9196  0.9253  0.9124  0.9109  0.9248  0.9213  0.9127  0.9311  0.9040  0.9104  0.917
3  0.0080  
Fit time          2.37    2.33    2.37    2.40    2.40    2.36    2.52    2.33    2.42    2.63    2.41    
0.09    
Test time         0.24    0.24    0.24    0.24    0.24    0.24    0.25    0.24    0.24    0.24    0.24    
0.00    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    1.5918  1.5273  1.5146  1.4350  1.6656  1.6257  1.6253  1.6605  1.5135  1.7516  1.591
1  0.0888  
Fit time          0.04    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    
0.00    
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    
0.00    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    0.9307  0.9176  0.9203  0.9387  0.9257  0.9370  0.9214  0.9256  0.9058  0.9096  0.923
2  0.0101  
Fit time          2.51    2.45    2.48    2.56    2.59    2.47    2.61    2.46    2.50    2.45    2.51    
0.05    
Test time         0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.78    0.24    0.25    0.30    
0.16    
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean    
Std     
RMSE (testset)    0.9405  0.9154  0.9113  0.9278  0.9216  0.9193  0.9228  0.9254  0.9214  0.9221  0.922
8  0.0074  
Fit time          2.48    2.48    2.51    2.61    2.57    2.44    2.63    2.47    2.56    2.46    2.52    
0.06    
Test time         0.24    0.24    0.24    0.25    0.24    0.25    0.24    0.24    0.24    0.24    0.24    
0.00
```

```
0.00


Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.4754  1.5721  1.7562  1.6738  1.6077  1.4970  1.5384  1.4660  1.6627  1.3093  1.5
559  0.1215
Fit time          0.04    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.0
3    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9203  0.9295  0.9300  0.9393  0.9198  0.9217  0.9112  0.9285  0.9321  0.9283  0.9
261  0.0075
Fit time          2.56    2.57    2.62    2.64    2.63    2.78    2.97    3.03    2.69    2.81    2.7
3    0.16
Test time         0.25    0.24    0.24    0.24    0.24    0.30    0.26    0.29    0.24    0.25    0.2
5    0.02
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9198  0.9274  0.9449  0.9111  0.9347  0.9206  0.9315  0.9247  0.9174  0.9307  0.9
263  0.0092
Fit time          2.68    2.64    2.78    2.68    2.63    2.95    2.72    2.66    2.66    2.65    2.7
1    0.09
Test time         0.24    0.24    0.25    0.25    0.25    0.81    0.24    0.24    0.24    0.24    0.3
0    0.17
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.4524  1.5049  1.6162  1.6742  1.5176  1.5688  1.6202  1.7118  1.4704  1.6008  1.5
737  0.0820
Fit time          0.04    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.0
3    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9346  0.9378  0.9089  0.9337  0.9133  0.9149  0.9376  0.9159  0.9505  0.9255  0.9
273  0.0129
Fit time          2.80    2.80    2.79    2.74    2.80    2.86    2.82    2.75    2.89    2.77    2.8
0    0.04
Test time         0.24    0.24    0.23    0.24    0.24    0.24    0.24    0.25    0.24    0.24    0.2
4    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9295  0.9176  0.9382  0.9157  0.9229  0.9303  0.9311  0.9282  0.9337  0.9312  0.9
278  0.0067
Fit time          2.83    2.89    2.75    2.81    2.85    2.77    2.77    2.78    2.82    2.80    2.8
1    0.04
Test time         0.24    0.24    0.25    0.24    0.26    0.24    0.24    0.24    0.24    0.24    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.4916  1.4955  1.3239  1.7225  1.7060  1.5008  1.6336  1.5763  1.8391  1.4764  1.5
766  0.1427
Fit time          0.05    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.03    0.0
4    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9293  0.9229  0.9352  0.9347  0.9500  0.9467  0.9265  0.9328  0.9454  0.9178  0.9
341  0.0101
Fit time          2.90    2.86    2.93    2.85    2.88    2.88    2.96    2.83    2.90    2.87    2.8
```

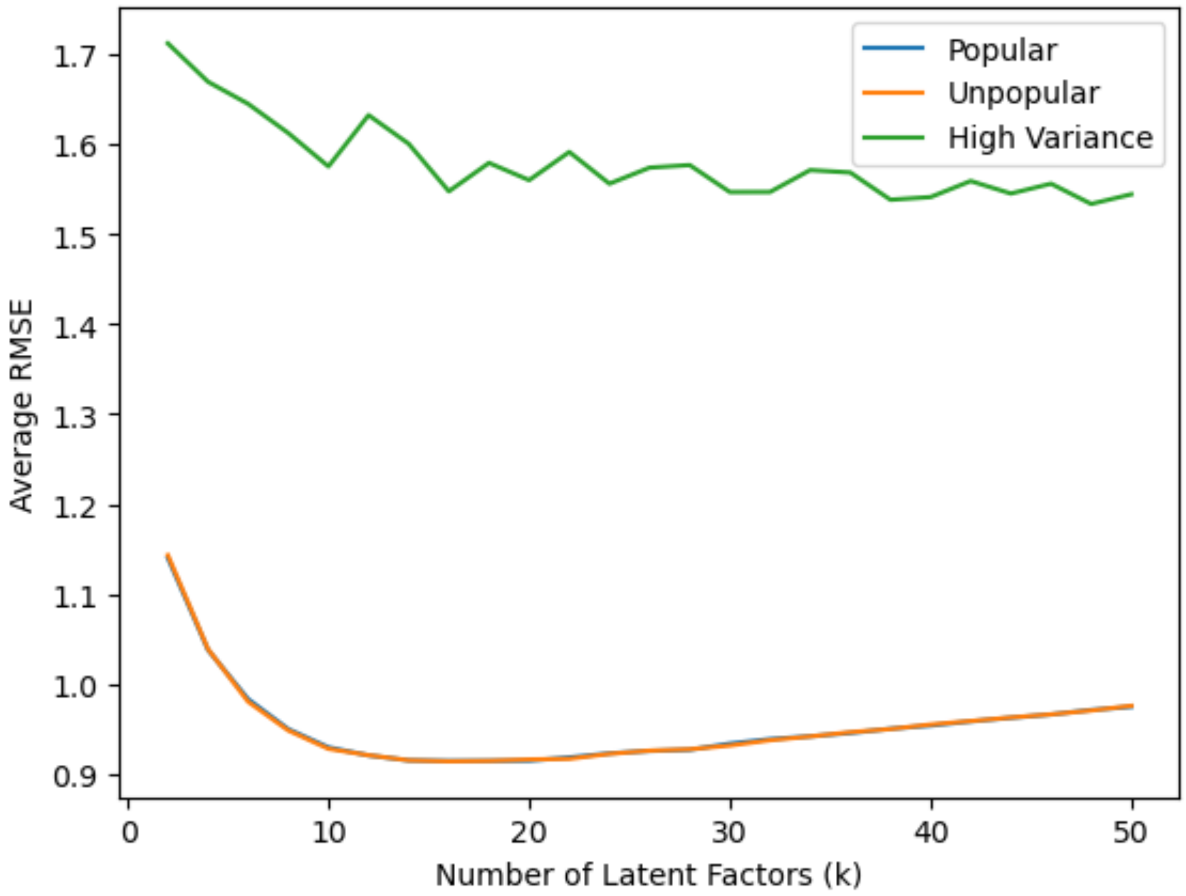
```
9    0.04
Test time         0.24    0.24    0.24    0.79    0.24    0.24    0.23    0.24    0.24    0.25    0.2
9    0.16
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9391  0.9350  0.9378  0.9218  0.9352  0.9330  0.9398  0.9321  0.9241  0.9223  0.9
320  0.0065
Fit time          3.02    2.86    2.98    3.00    2.93    2.87    3.10    2.89    2.93    3.05    2.9
6    0.08
Test time         0.24    0.25    0.26    0.24    0.24    0.24    0.24    0.25    0.24    0.24    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.4498  1.3849  1.5683  1.7016  1.6340  1.6224  1.4664  1.5756  1.3435  1.7196  1.5
466  0.1232
Fit time          0.05    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.0
4    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea

n    Std
RMSE (testset)    0.9371  0.9296  0.9289  0.9401  0.9385  0.9501  0.9622  0.9401  0.9373  0.9273  0.9
391  0.0100
Fit time          3.05    2.99    3.17    3.01    3.05    3.11    3.05    2.99    3.17    3.00    3.0
6    0.07
Test time         0.24    0.24    0.24    0.24    0.23    0.25    0.24    0.25    0.24    0.24    0.2
4    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9344  0.9178  0.9392  0.9512  0.9417  0.9458  0.9422  0.9355  0.9337  0.9377  0.9
379  0.0084
Fit time          3.07    3.20    3.04    3.05    3.12    3.03    3.07    3.08    3.02    3.04    3.0
7    0.05
Test time         0.24    0.79    0.25    0.25    0.25    0.24    0.24    0.24    0.24    0.24    0.3
0    0.16
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.4764  1.5556  1.7335  1.4239  1.5788  1.5680  1.5704  1.6798  1.6016  1.2789  1.5
467  0.1223
Fit time          0.05    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.0
4    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9603  0.9435  0.9315  0.9430  0.9405  0.9512  0.9553  0.9327  0.9355  0.9284  0.9
422  0.0101
Fit time          3.13    3.14    3.20    3.12    3.18    3.12    3.17    3.10    3.17    3.10    3.1
4    0.03
Test time         0.25    0.23    0.24    0.25    0.25    0.24    0.24    0.23    0.24    0.24    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9215  0.9548  0.9398  0.9418  0.9495  0.9669  0.9569  0.9271  0.9327  0.9296  0.9
421  0.0140
Fit time          3.21    3.11    3.17    3.21    3.18    3.16    3.21    3.22    3.22    3.12    3.1
8    0.04
Test time         0.23    0.24    0.24    0.24    0.25    0.25    0.25    0.25    0.24    0.81    0.3
0    0.17
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.5359  1.4853  1.7097  1.5818  1.4119  1.6482  1.3418  1.7503  1.6030  1.6421  1.5
710  0.1225
```

```
Fit time          0.05    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.0
4    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9463  0.9433  0.9471  0.9655  0.9312  0.9362  0.9513  0.9482  0.9473  0.9428  0.9
459  0.0086
Fit time          3.35    3.25    3.28    3.31    3.31    3.29    3.37    3.25    3.30    3.32    3.3
0    0.04
Test time         0.25    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.24    0.2
4    0.00

Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9461  0.9531  0.9536  0.9593  0.9340  0.9465  0.9391  0.9366  0.9398  0.9593  0.946
7  0.0088
Fit time          3.29    3.27    3.33    3.29    3.33    3.27    3.26    3.26    3.40    3.32    3.30
0.04
Test time         0.24    0.24    0.23    0.25    0.24    0.25    0.24    0.25    0.25    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.4633  1.5461  1.5195  1.6522  1.7591  1.4493  1.6995  1.5340  1.3330  1.7271  1.568
3  0.1306
Fit time          0.06    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9457  0.9534  0.9569  0.9594  0.9433  0.9611  0.9385  0.9446  0.9388  0.9646  0.950
6  0.0091
Fit time          3.51    3.44    3.45    3.50    3.39    3.50    3.56    3.38    3.54    3.43    3.47
0.06
Test time         0.25    0.25    0.24    0.24    0.24    0.25    0.24    0.24    0.24    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9436  0.9647  0.9509  0.9467  0.9454  0.9501  0.9397  0.9592  0.9540  0.9490  0.950
3  0.0070
Fit time          3.41    3.46    3.45    3.37    3.56    3.43    3.47    3.49    3.41    3.51    3.46
0.05
Test time         0.24    0.25    0.25    0.24    0.25    0.24    0.25    0.24    0.24    0.25    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.7043  1.5122  1.3585  1.7532  1.3223  1.6088  1.6133  1.4397  1.5100  1.5585  1.538
1  0.1323
Fit time          0.06    0.04    0.04    0.04    0.04    0.05    0.05    0.05    0.05    0.04    0.05
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9620  0.9609  0.9443  0.9519  0.9563  0.9517  0.9389  0.9689  0.9498  0.9570  0.954
2  0.0083
Fit time          3.49    3.49    3.62    3.46    3.55    3.61    3.55    3.51    3.62    3.47    3.54
0.06
Test time         0.24    0.23    0.25    0.25    0.24    0.24    0.24    0.24    0.24    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
```

```
RMSE (testset)    0.9318  0.9595  0.9438  0.9492  0.9647  0.9671  0.9633  0.9403  0.9633  0.9704  0.955
3  0.0125
Fit time          3.53    3.52    3.54    3.53    3.54    3.53    3.71    3.54    3.62    3.61    3.57
0.06
Test time         0.24    0.23    0.24    0.24    0.24    0.24    0.24    0.24    0.25    0.26    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.7300  1.8216  1.4973  1.4407  1.5085  1.5059  1.7472  1.1802  1.4287  1.5500  1.541
0  0.1772
Fit time          0.06    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9452  0.9689  0.9604  0.9598  0.9577  0.9555  0.9636  0.9464  0.9713  0.9598  0.958
9  0.0080
Fit time          3.64    3.65    3.68    3.63    3.79    3.63    3.62    3.71    3.60    3.63    3.66
0.05
Test time         0.24    0.25    0.23    0.24    0.24    0.24    0.25    0.25    0.23    0.24    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9702  0.9515  0.9689  0.9564  0.9563  0.9621  0.9575  0.9524  0.9632  0.9532  0.959
2  0.0063
Fit time          3.66    3.63    3.76    3.64    3.61    3.75    3.71    3.60    3.61    3.62    3.66
0.06
Test time         0.24    0.25    0.24    0.24    0.25    0.24    0.24    0.24    0.24    0.25    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean

Std
RMSE (testset)    1.5774  1.4732  1.5673  1.4949  1.5360  1.5704  1.5291  1.6937  1.6280  1.5186  1.558
9  0.0616
Fit time          0.06    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9695  0.9709  0.9592  0.9677  0.9572  0.9554  0.9657  0.9538  0.9614  0.9673  0.962
8  0.0059
Fit time          3.91    3.76    3.81    3.78    3.80    3.76    3.76    4.01    3.91    3.80    3.83
0.08
Test time         0.24    0.24    0.24    0.24    0.23    0.25    0.24    0.24    0.25    0.24    0.24
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9547  0.9672  0.9524  0.9718  0.9608  0.9628  0.9728  0.9670  0.9571  0.9642  0.963
1  0.0066
Fit time          3.80    3.72    3.83    3.80    3.77    3.78    3.93    3.75    3.79    3.84    3.80
0.05
Test time         0.24    0.24    0.23    0.25    0.24    0.25    0.24    0.25    0.24    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.7013  1.6257  1.4550  1.5717  1.5452  1.4574  1.5342  1.5146  1.6303  1.4140  1.545
0  0.0854
Fit time          0.06    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
```

```
Std
RMSE (testset)    0.9712  0.9631  0.9609  0.9733  0.9651  0.9584  0.9532  0.9760  0.9813  0.9622  0.966
5  0.0083
Fit time          3.84    3.96    3.98    3.86    4.89    3.87    3.96    3.87    3.91    3.87    4.00
0.30
Test time         0.24    0.24    0.23    0.25    0.24    0.23    0.25    0.24    0.24    0.25    0.24
0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    0.9908  0.9671  0.9805  0.9478  0.9589  0.9477  0.9712  0.9777  0.9577  0.9663  0.966
6  0.0133
Fit time          3.93    3.99    3.96    3.86    3.95    3.89    3.96    3.89    4.02    3.97    3.94
0.05
Test time         0.25    0.77    0.24    0.25    0.24    0.25    0.25    0.24    0.24    0.25    0.30
0.16
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mean
Std
RMSE (testset)    1.4979  1.6734  1.4084  1.5879  1.5847  1.7332  1.5586  1.5200  1.5307  1.4640  1.555
9  0.0909
Fit time          0.06    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05    0.05
0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00
0.00

Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9766  0.9637  0.9766  0.9571  0.9813  0.9732  0.9673  0.9698  0.9679  0.9816  0.9
715  0.0074
Fit time          4.06    4.03    4.06    3.95    4.15    4.05    4.02    4.03    4.04    4.04    4.0
4    0.05
Test time         0.24    0.24    0.23    0.24    0.24    0.23    0.24    0.25    0.25    0.24    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9862  0.9639  0.9690  0.9683  0.9708  0.9666  0.9698  0.9653  0.9792  0.9699  0.9
709  0.0064
Fit time          4.07    3.99    3.99    4.03    4.13    4.00    4.05    4.03    4.10    4.11    4.0
5    0.05
Test time         0.24    0.26    0.25    0.23    0.24    0.23    0.24    0.25    0.23    0.24    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.5205  1.4924  1.5286  1.4076  1.5227  1.6194  1.6873  1.6314  1.4136  1.5106  1.5
334  0.0856
Fit time          0.07    0.06    0.06    0.05    0.06    0.06    0.06    0.06    0.06    0.06    0.0
6    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9784  0.9638  0.9867  0.9859  0.9889  0.9684  0.9646  0.9875  0.9619  0.9643  0.9
750  0.0109
Fit time          4.15    4.14    4.14    4.15    4.22    4.09    4.15    4.09    4.10    4.20    4.1
4    0.04
Test time         0.24    0.24    0.23    0.23    0.25    0.24    0.25    0.24    0.25    0.25    0.2
4    0.01
Evaluating RMSE of algorithm NMF on 10 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    0.9645  0.9602  0.9760  0.9793  0.9814  0.9767  0.9718  0.9804  0.9819  0.9884  0.9
761  0.0080
Fit time          4.17    4.16    4.16    4.15    4.29    4.07    4.22    4.12    4.15    4.23    4.1
7    0.06
Test time         0.25    0.25    0.24    0.24    0.25    0.25    0.24    0.24    0.24    0.24    0.2
4    0.00
Evaluating RMSE of algorithm NMF on 10 split(s).
```

```
                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Fold 6  Fold 7  Fold 8  Fold 9  Fold 10 Mea
n    Std
RMSE (testset)    1.5128  1.6112  1.6441  1.4271  1.3871  1.8272  1.5891  1.4654  1.6241  1.3532  1.5
441  0.1360
Fit time          0.07    0.06    0.06    0.06    0.06    0.06    0.06    0.06    0.06    0.06    0.0
6    0.00
Test time         0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.0
0    0.00
```

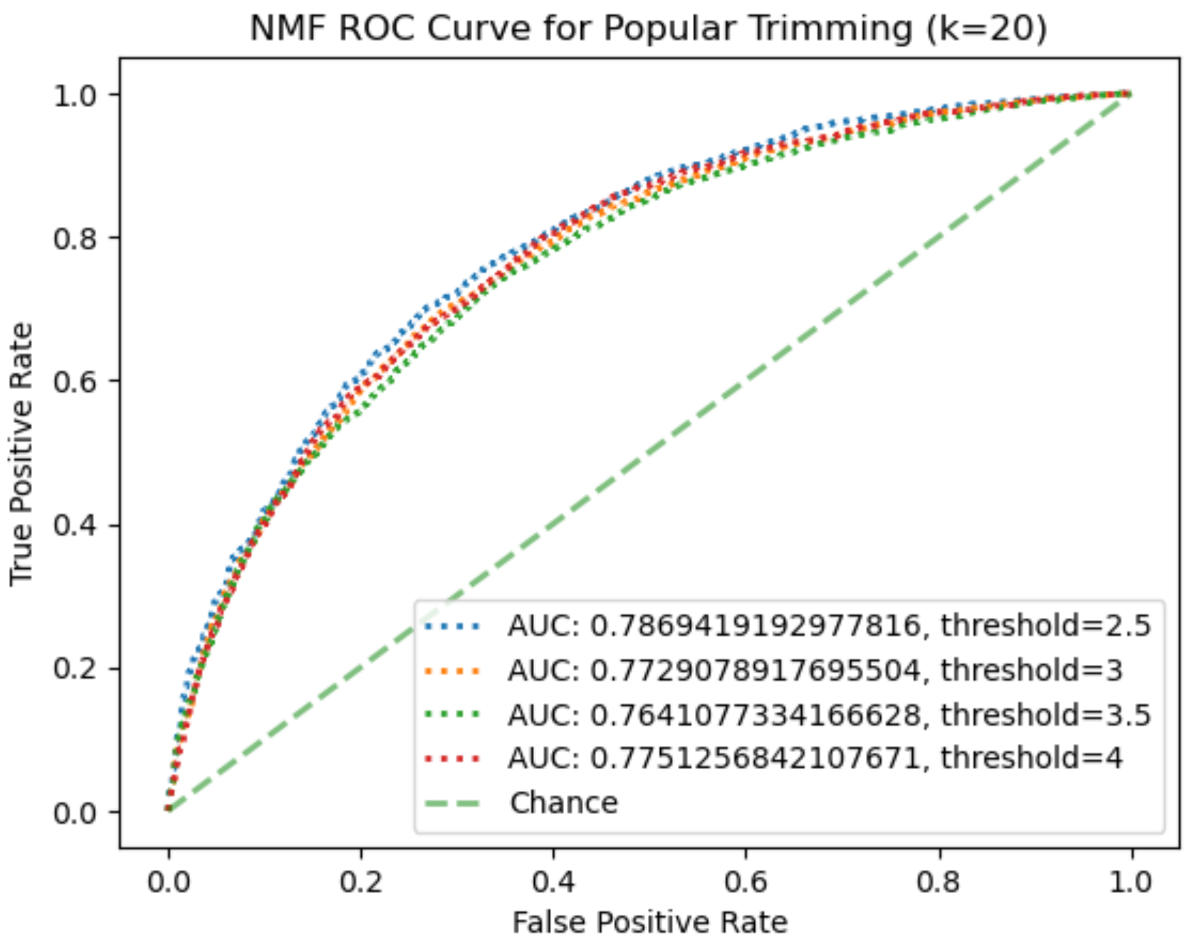



```
Minimum Average RMSE for Popular Subset: 0.9149651798336492
Minimum Average RMSE for Unpopular Subset: 0.9145124802185872
Minimum Average RMSE for High Variance Subset: 1.5334086565329106
```

•Plot the ROC curves for the NMF-based collaborative filter and also report the area under the curve (AUC) value as done in Question 6.

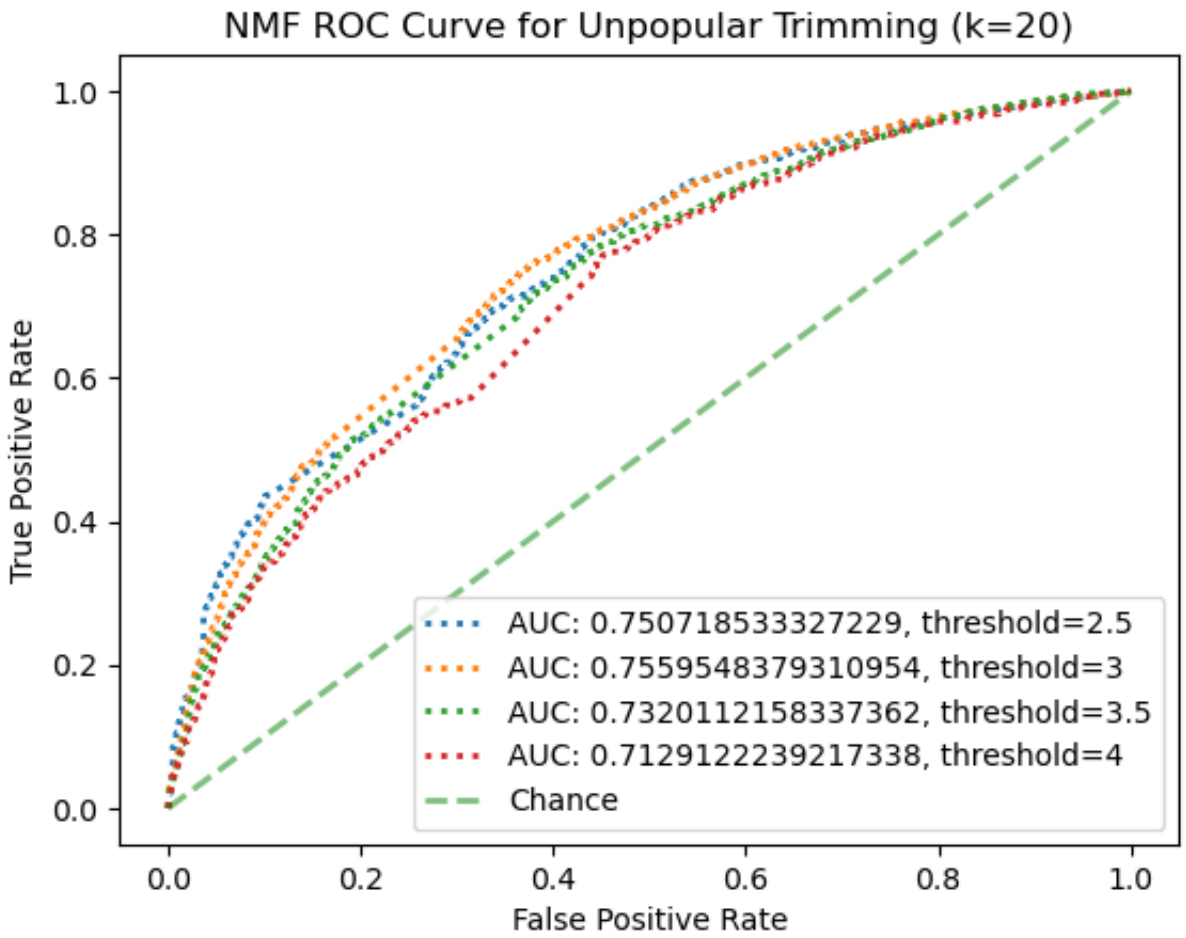
In [284]:
```
#popular trimmming

data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = NMF(n_factors=20).fit(trainset).test(popular_trim(testset))
print("Results for popular trimming'} with k=20")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('NMF ROC Curve for Popular Trimming (k=20)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for popular trimming'} with k=20
```



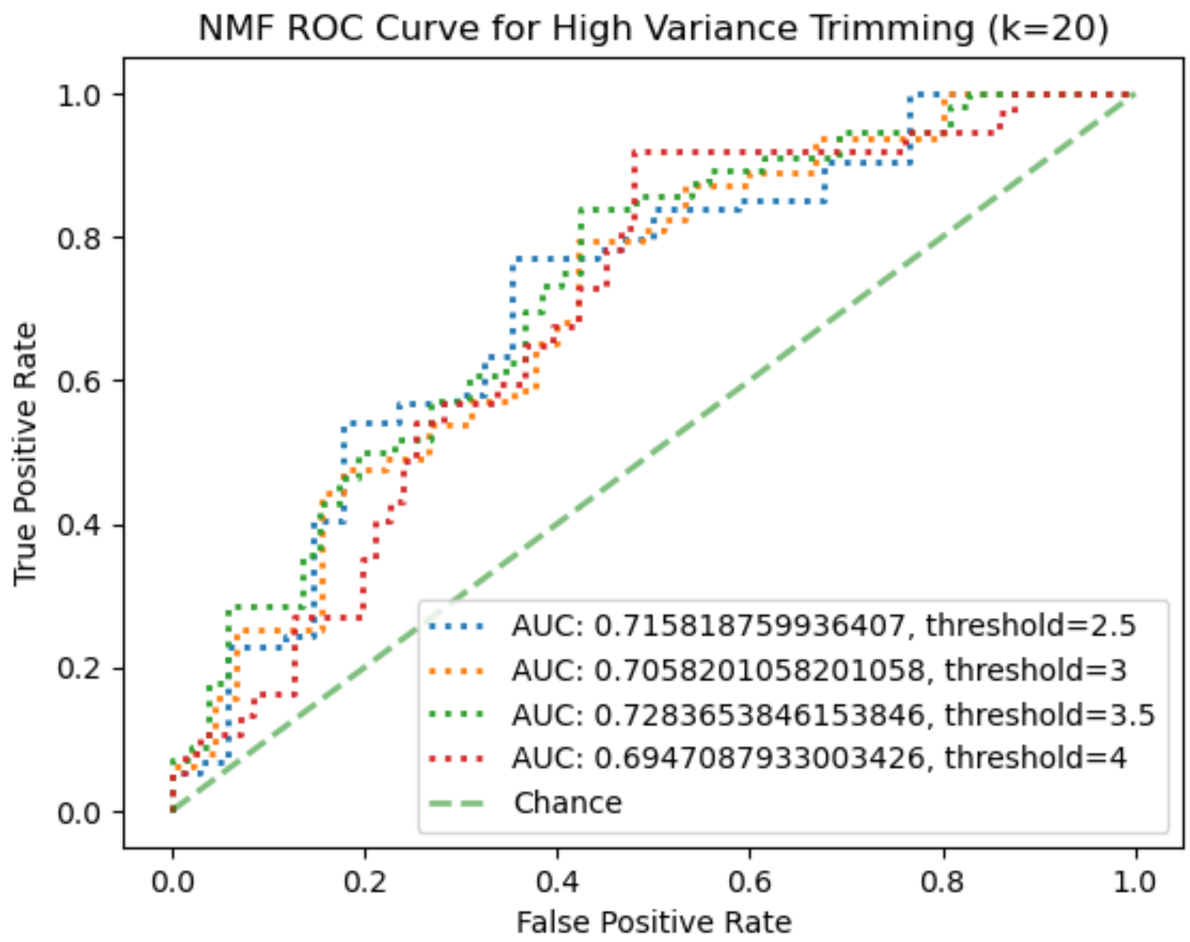
In [286]:
```
#Unpopular Trim
thresholds = [2.5, 3, 3.5, 4]

data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = NMF(n_factors=20).fit(trainset).test(unpopular_trim(testset))
print("Results for unpopular trimming'} with k=20")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('NMF ROC Curve for Unpopular Trimming (k=20)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for unpopular trimming'} with k=20
```



In [287]:
```
#High Variance Trim
thresholds = [2.5, 3, 3.5, 4]
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = NMF(n_factors=20).fit(trainset).test(high_var_trim(testset))
print("Results for high variance trimming'} with k=20")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('NMF ROC Curve for High Variance Trimming (k=20)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for high variance trimming'} with k=20
```



QUESTION 9:

Interpreting the NMF model: 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?

A: The top 10 movies in each column of movie-latent factor tend to belong to a small collection of genres. It is possible that there is a connection between the latent factors and the genres.

In [399]:

```
V = NMF(n_factors=20, verbose=False).fit(train).qi
movie_df =df_movie = pd.read_csv('/Users/ryan/Downloads/Synthetic_Movie_Lens/movies.csv',names=['movieid
for col in range(V.shape[1]):
  movie = V[:, col]
  movie = [(id, rank) for id, rank in enumerate(movie)]
  movie = sorted(movie, key=lambda x : x[1], reverse=True)[:10]
  print(f'column {col}')
  for id, _ in movie:
    print(movie_df['genres'][id])
  print('')
```

```
column 0
Drama
Comedy|Romance|Sci-Fi
Action|Adventure|Fantasy
Action|Adventure|Drama|Mystery|Thriller
Comedy|Drama
Comedy
Adventure|Animation|Children|Fantasy
Mystery|Sci-Fi|Thriller
Horror
Drama|Romance|Western

column 1
Comedy|Drama|Romance
Animation|Drama|Fantasy
Comedy|Drama|Romance
Drama|Romance
Action|Horror|Sci-Fi
Drama
Comedy|Sci-Fi
Comedy
Horror
Comedy

column 2
Crime|Drama|Mystery
Comedy|Drama|Romance
Action|Drama|Romance|War
Action|Horror|Sci-Fi|Thriller
Action|Crime|Drama|Thriller
Drama
Comedy|Romance|Sci-Fi
Comedy
Thriller
Action|Crime|Drama|Thriller

column 3
Crime|Horror|Thriller
Action|Comedy
Adventure|Comedy|War
Comedy|Drama
Documentary|War
Action|Adventure|Crime|Drama|Thriller|War
Comedy|Romance
Drama
Action|Adventure|Fantasy|Romance|IMAX
Drama

column 4
Action|Adventure|Comedy|Western
Adventure|Comedy|War
Action|Adventure|Fantasy
Crime|Drama
Horror|Mystery|Sci-Fi
Sci-Fi
Comedy|Romance
Drama|Thriller|War
Comedy
Documentary

column 5
Action|Thriller
Comedy|Musical
Adventure|Animation|Fantasy|Romance
Action|Drama
Action|Comedy
Comedy|Romance
```

```
Drama
Action|Adventure|Thriller
Action|Adventure|Comedy|Sci-Fi
Drama|Mystery|Romance|Thriller

column 6
Action|Adventure|Fantasy|IMAX
Drama|Horror|Sci-Fi
Drama
Action|Drama
Western
Comedy|Romance
Comedy|Drama|Sci-Fi
Adventure|Drama
Drama|Sci-Fi
Action|Drama

column 7
Crime|Drama
Comedy|Horror
Adventure|Animation|Children|Comedy|Fantasy
Drama
Crime|Mystery
Documentary
Drama
Drama|Romance
Comedy|Sci-Fi
Mystery|Sci-Fi|Thriller

column 8
Action|Comedy
Action|Adventure|Animation|Fantasy|Sci-Fi
Crime|Drama
Action|Comedy|Crime
Drama
Crime|Drama|Thriller
Drama|Western
Comedy|Drama|Romance
Drama|Horror|Mystery|Thriller
Horror

column 9
Documentary
Comedy|Drama|Romance
Comedy
Adventure|Comedy|Crime|Romance
Drama|Romance
Drama|Thriller
Adventure|Animation|Children|Musical|Romance
Action|Crime|Mystery|Sci-Fi|Thriller
Adventure|Crime|Drama
Drama|Horror

column 10
Comedy
Action|Crime|Drama
Comedy|Drama|Romance
Comedy|Drama
Action|Adventure|Crime|Drama
Drama
Thriller
Action|Adventure|Fantasy|Horror
Horror|Thriller
Drama

column 11
Crime|Drama
Comedy
Documentary
Adventure|Comedy
Comedy|Drama|Romance
Drama|Horror
Documentary
Action|Comedy|Crime|Romance
Drama|Romance
Drama|Fantasy

column 12
Comedy|Romance
```

```
Drama|Horror
Action|Sci-Fi
Drama|War
Comedy|Romance
Drama
Comedy
Action|Adventure|Animation|Children|Comedy
Comedy
Children|Comedy|Fantasy

column 13
Comedy|Drama
Action|Adventure|Sci-Fi|Thriller
Comedy
Comedy|Horror
Drama|Thriller
Comedy|Sci-Fi
Action|Adventure|Comedy
Comedy|Crime|Musical
Comedy|Musical
Horror

column 14
Action|Adventure|Drama
Documentary
Comedy
Drama|Horror|Mystery|Thriller
Comedy
Drama|Thriller|War
Drama
Comedy
Comedy|Drama|Romance
Action|Comedy|Romance

column 15
Drama|Romance
Comedy|Drama|Fantasy
Action|Crime|Drama
Comedy
Drama
Drama
Comedy|Drama
Action|Comedy|Drama|Romance
Comedy|Romance
Comedy|Romance

column 16
Drama
Action|Crime|Drama|Thriller
Drama
Drama
Drama|Romance
Crime|Mystery|Thriller
Drama|Romance
Comedy|Romance
Drama|War
Crime|Drama|Thriller

column 17
Drama|Horror
Drama
Drama
Action|Drama|Romance|War
Drama
Comedy|Drama
Action|Adventure|Children|Comedy|Fantasy
Comedy
Adventure|Children|Comedy|Musical
Horror

column 18
Horror
Action|Thriller
Comedy|Crime
Comedy|Drama|Romance
Comedy|Drama|War
Comedy|Fantasy
Action|Adventure|Sci-Fi|Thriller
Adventure|Drama|Romance
```

```
Drama|Romance
Horror|Mystery|Thriller

column 19
Documentary
Action|Comedy
Drama|Horror
Drama|Romance|War
Children|Drama
Comedy|Romance
Action|Comedy|Crime|Drama|Thriller
Drama|Musical
Adventure|Animation|Children|Musical
Animation|Children|Comedy
```

In [193]:

```
reader = Reader(rating_scale=(0.5, 5.0))
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
# create the NMF model with k=20
nmf = NMF(n_factors=20)

# fit the model on the dataset
trainset = data.build_full_trainset()
nmf.fit(trainset)

# Get the V matrix from the trained model
V = nmf.qi


# get the top 3 genres of the top 10 movies for each latent factor based on their scores in the correspo
top_genres = {}
for i in range(V.shape[1]):
    top_movies_idx = np.argsort(V[:,i])[::-1][:10]
    top_movies = df_movie.iloc[top_movies_idx]
    genres = top_movies['genres'].str.split('|', expand=True).stack().value_counts()[:3].index.tolist()
    top_genres[i+1] = genres

# print out the top genres for each latent factor
print("Top genres for each latent factor:")
for k, v in top_genres.items():
    print("Latent factor {}: {}".format(k, ', '.join(v)))
```

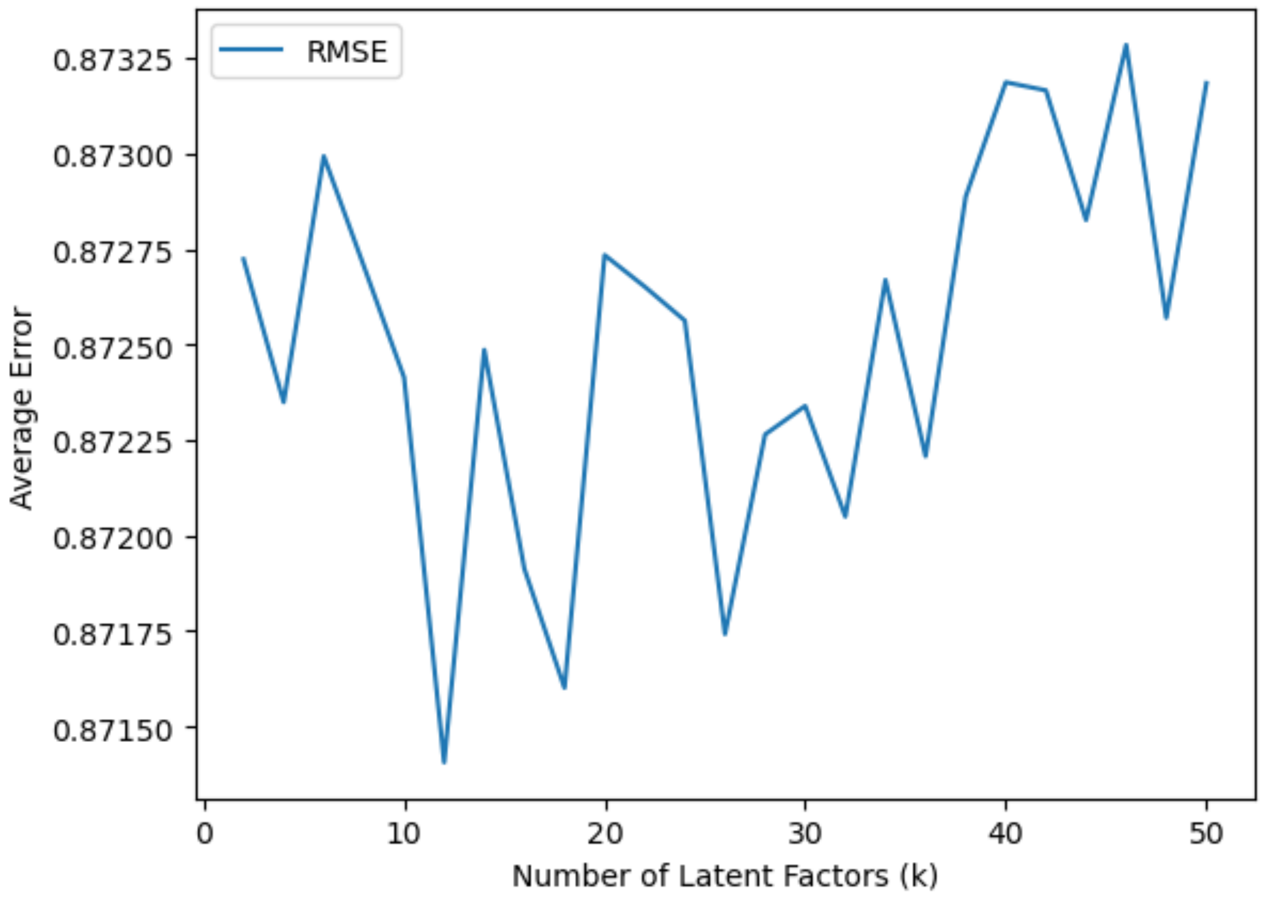
```
Top genres for each latent factor:
Latent factor 1: Drama, Comedy, Crime
Latent factor 2: Drama, Comedy, Romance
Latent factor 3: Comedy, Action, Thriller
Latent factor 4: Comedy, Drama, Romance
Latent factor 5: Drama, Romance, Comedy
Latent factor 6: Drama, Comedy, War
Latent factor 7: Drama, Action, Thriller
Latent factor 8: Comedy, Romance, Drama
Latent factor 9: Action, Drama, Mystery
Latent factor 10: Drama, Action, Romance
Latent factor 11: Drama, Comedy, Sci-Fi
Latent factor 12: Comedy, Drama, Romance
Latent factor 13: Drama, Comedy, Thriller
Latent factor 14: Drama, Romance, Action
Latent factor 15: Drama, Comedy, Horror
Latent factor 16: Drama, Adventure, Western
Latent factor 17: Drama, Adventure, Romance
Latent factor 18: Drama, Comedy, Romance
Latent factor 19: Comedy, Drama, Romance
Latent factor 20: Comedy, Horror, Drama
```

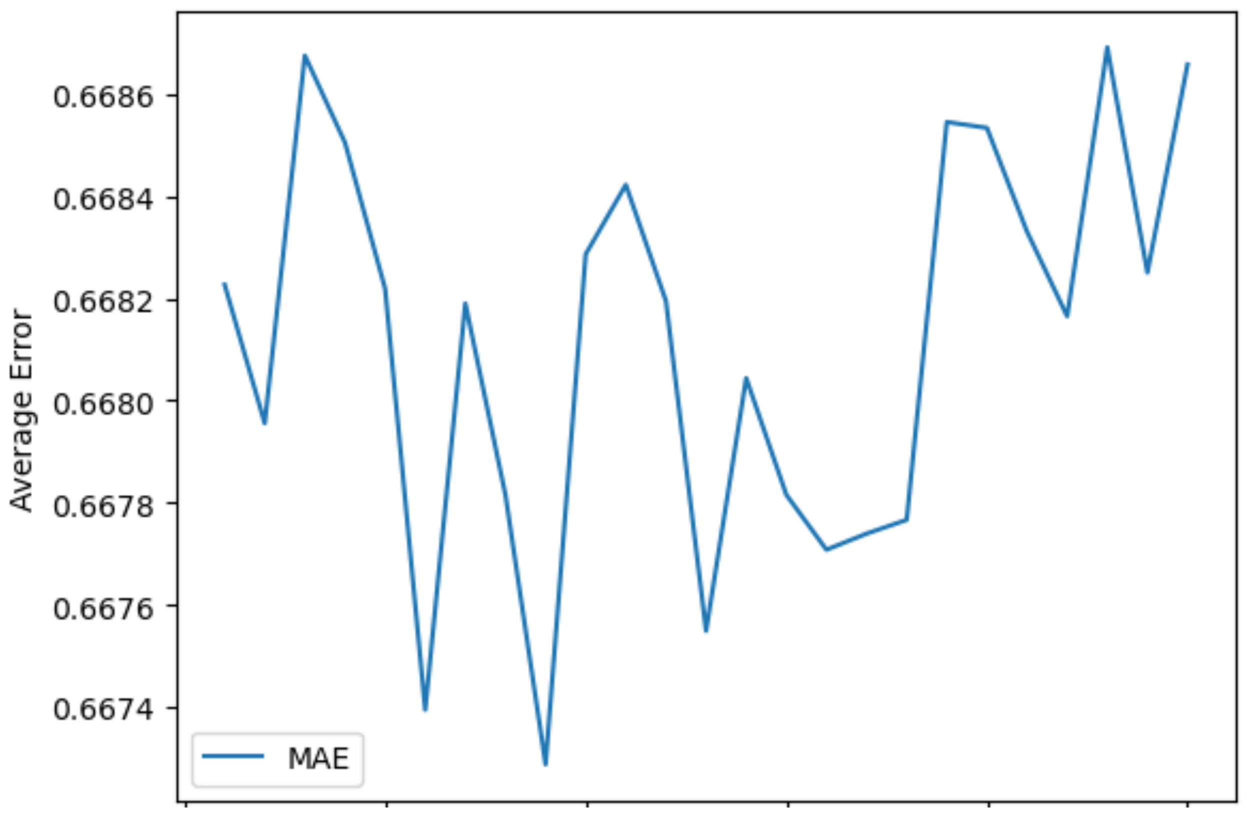
QUESTION 10: Designing the MF Collaborative Filter:

A

Design a MF-based collaborative filter to predict the ratings of the movies in the original 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 [214]:
```
# load the dataset
reader = Reader(rating_scale=(0.5, 5))
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)

# define the range of k values to test
k_values = np.arange(2, 51, 2)


# perform 10-fold cross-validation with MF algorithm for different k values
rmse_scores = []
mae_scores = []
for k in k_values:
    results  = cross_validate(SVD(n_factors=k),
                         measures=['rmse','mae'],data = data, cv=10,  verbose=False)
    rmse_scores.append(np.mean(results['test_rmse']))
    mae_scores.append(np.mean(results['test_mae']))

# plot the results
plt.plot(k_values, rmse_scores, label='RMSE')
plt.xlabel('Number of Latent Factors (k)')
plt.ylabel('Average Error')
plt.legend()
plt.show()

plt.plot(k_values, mae_scores, label='MAE')
plt.xlabel('Number of Latent Factors (k)')
plt.ylabel('Average Error')
plt.legend()
plt.show()
```





B

Use the plot from the previous part 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?

A: The optimal number of latent factors of RMSE for SVD is 12 and of MAE is 18 which is different from the number of movie genres.

In [215]:
```
print("Minimum average value of RMSE for SVD: %f, K Value: %d" % (min(rmse_scores),k_values[[i for i, x
print("Minimum average value of MAE for SVD: %f, K Value: %d" % (min(mae_scores),k_values[[i for i, x in
```

```
Minimum average value of RMSE for SVD: 0.871406, K Value: 12
Minimum average value of MAE for SVD: 0.667288, K Value: 18
```

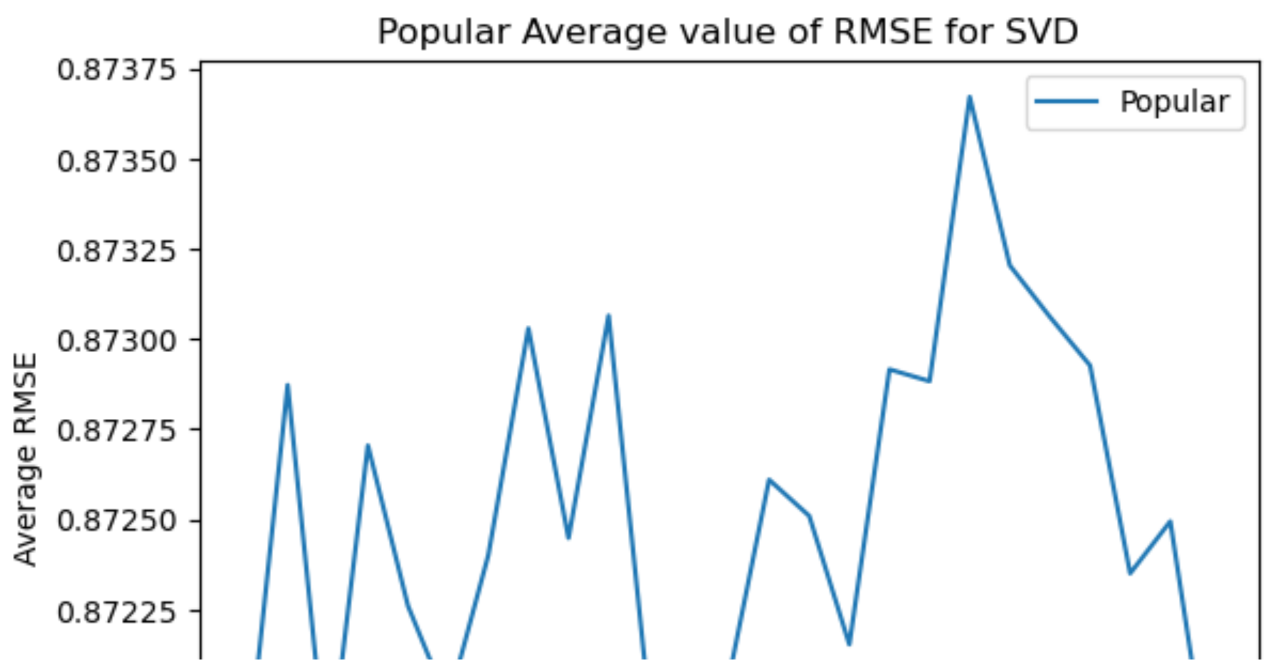
C

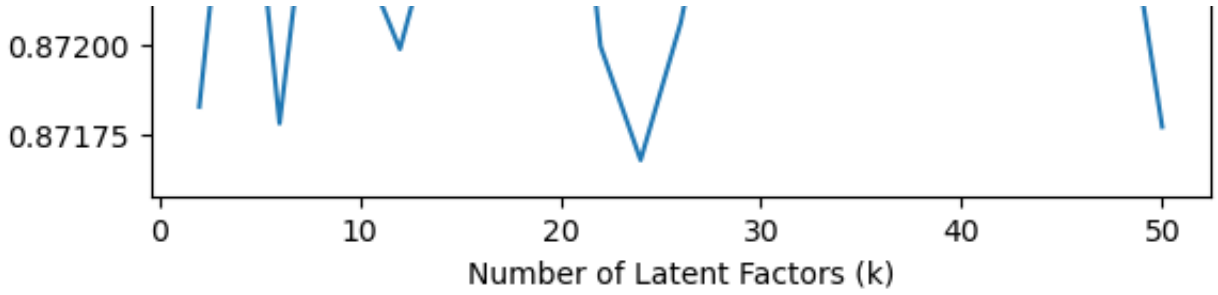
Performance on dataset subsets: For each of Popular, Unpopular and High-Variance subsets

– Design a MF collaborative filter for each trimmed subset and evaluate its 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 obtained by averaging the RMSE across all 10 folds. – Plot average RMSE (Y-axis) against k (X-axis); item Report the minimum average RMSE.

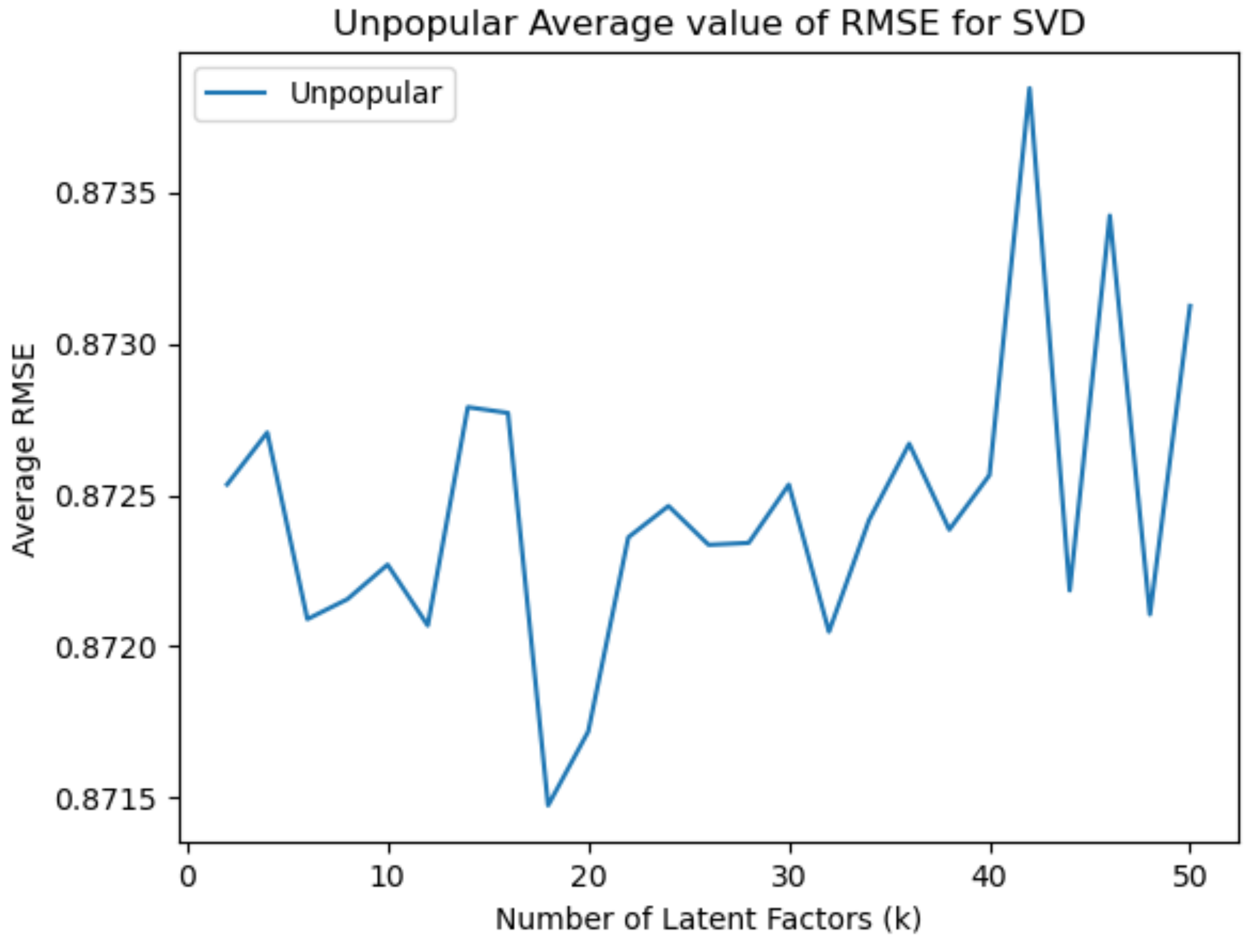
In [206]:
```
# define the range of k values to test
k_values = np.arange(2, 51, 2)

# evaluate performance for each subset
for subset, subset_name in zip([popular_trim, unpopular_trim, high_var_trim], ['Popular', 'Unpopular', '
    print(f"Evaluating {subset_name} subset...")
    subset_data = Dataset.load_from_df(subset(df)[['userId', 'movieId', 'rating']], reader)
    rmse_scores = []
    for k in k_values:
        algo = SVD(n_factors=k)
        results = cross_validate(algo, subset_data, measures=['RMSE'], cv=10, verbose=False)
        rmse_scores.append(np.mean(results['test_rmse']))

    # plot the results
    plt.plot(k_values, rmse_scores, label=subset_name)

    # plot the figure
    plt.title(subset_name +' Average value of RMSE for SVD')
    plt.xlabel('Number of Latent Factors (k)')
    plt.ylabel('Average RMSE')
    plt.legend()
    plt.show()
```

```
Evaluating Popular subset...
```

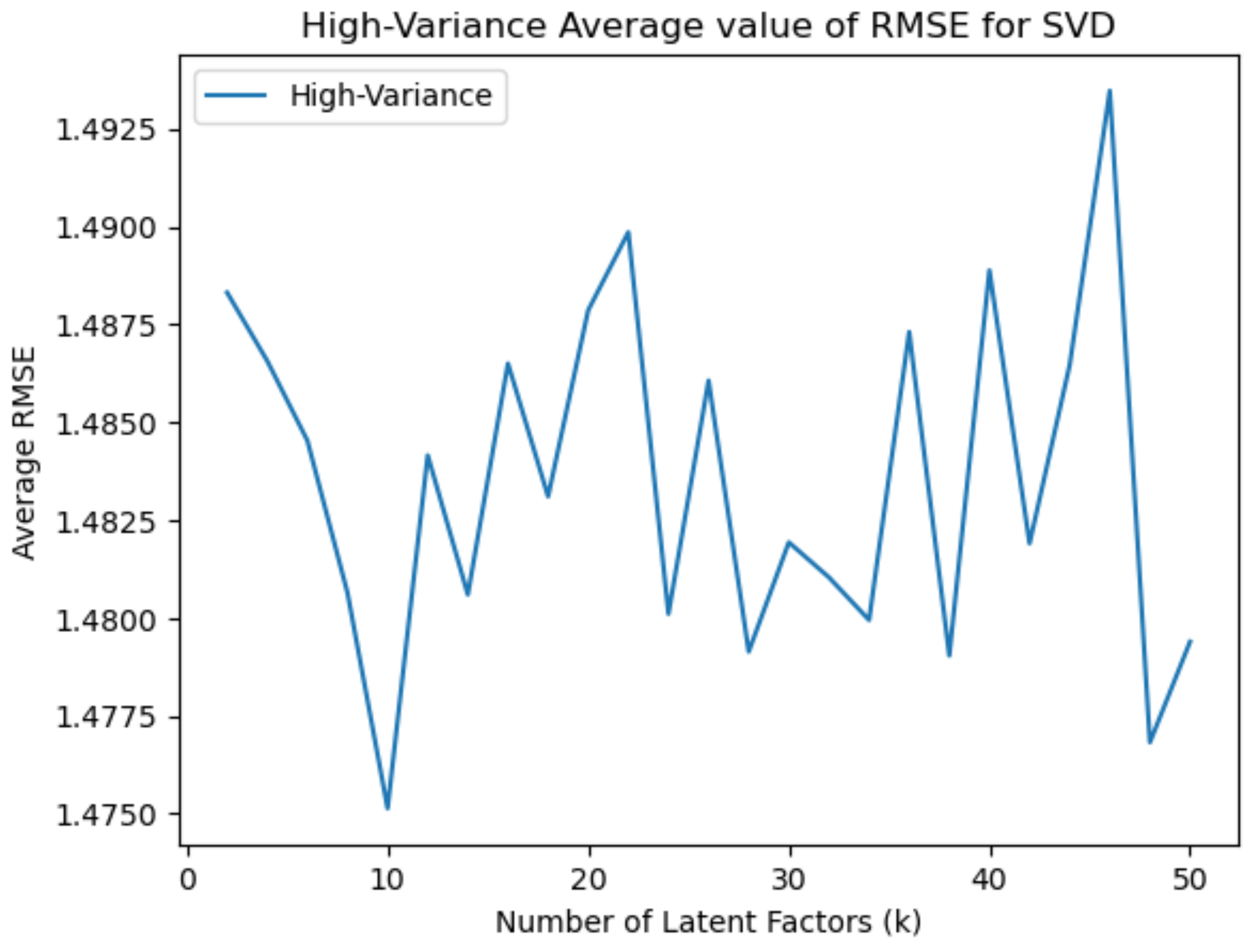





```
Evaluating Unpopular subset...
```



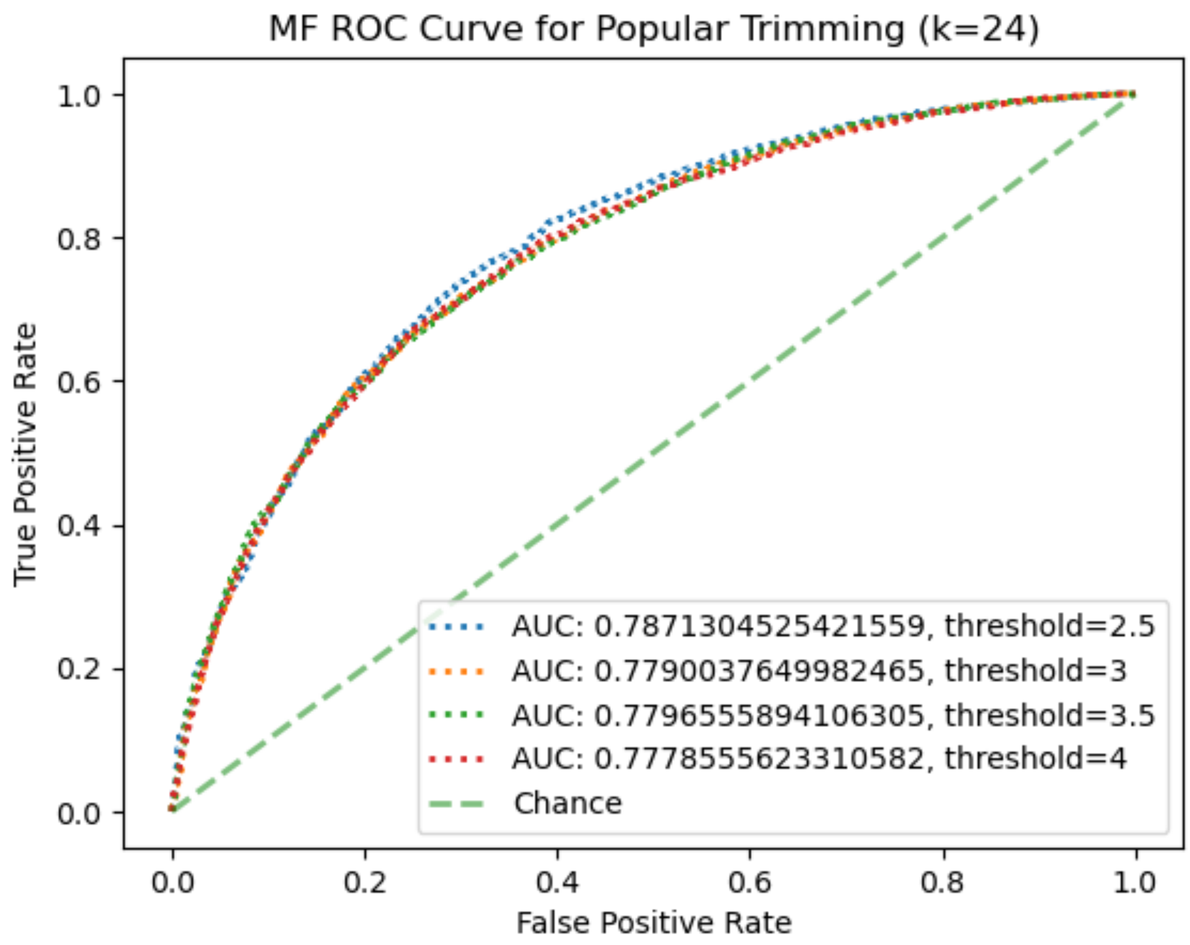

```
Evaluating High-Variance subset...
```



•Plot the ROC curves for the MF-based collaborative filter and also report the area under the curve (AUC) value as done in Question 6.

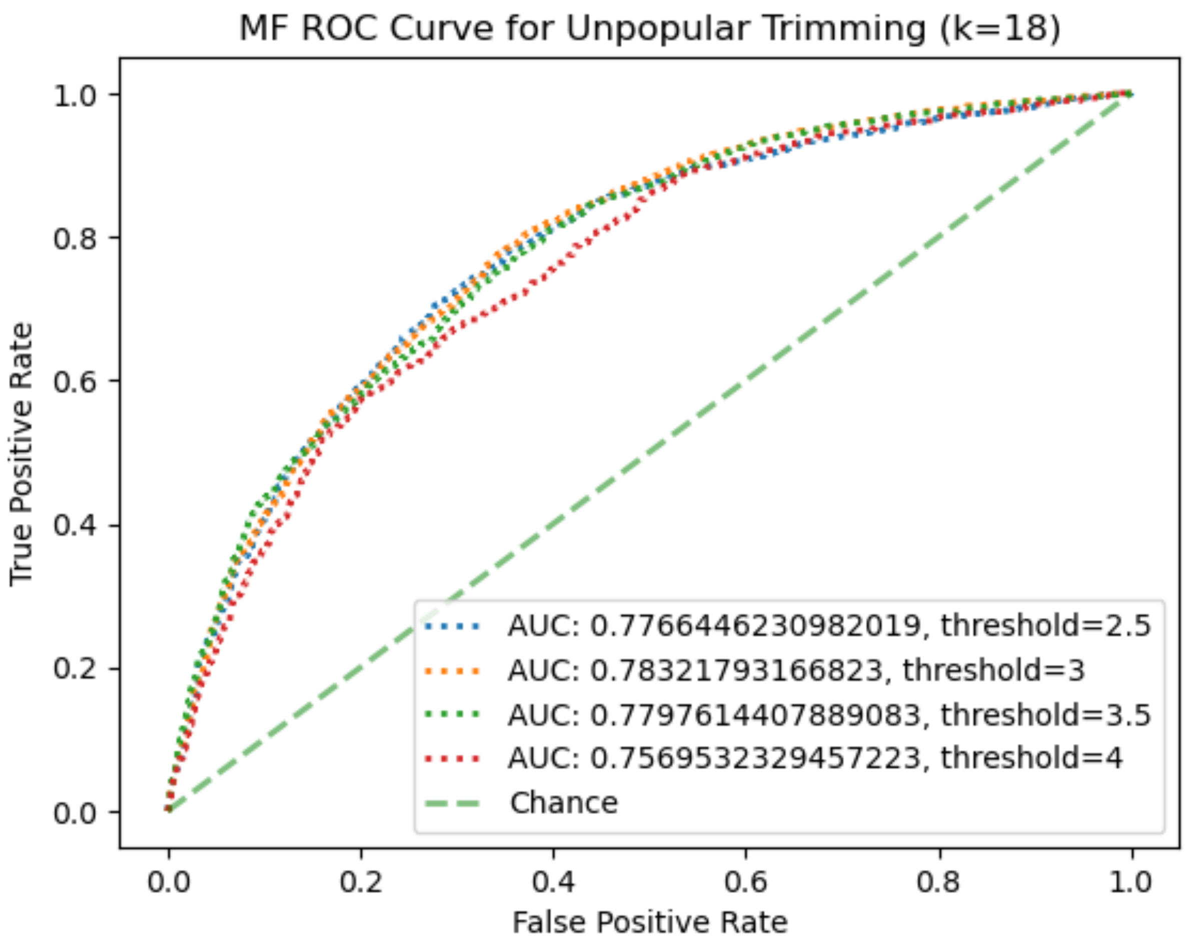
In [292]:
```
def popular_trim(testset: List[Tuple[int, int, float]]) -> List[Tuple[int, int, float]]:
    # Count the number of ratings for each movie
    movie_ratings = defaultdict(int)
    for (uid, iid, rating) in testset:
        movie_ratings[iid] += 1

    # Identify popular movies
    popular_movies = set(iid for iid, count in movie_ratings.items() if count > 2)

    # Filter out unpopular movies from the test set
    return [t for t in testset if len(t) == 3 and t[1] in popular_movies]
threshold_vals = [2.5, 3, 3.5, 4]
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = SVD(n_factors=24).fit(trainset).test(popular_trim(testset))
print("Results for popular trimming with k=24")
fig, ax = plt.subplots()
for threshold_val in threshold_vals:
    thresholded_out = []
    for row in res:
        if row.r_ui > threshold_val:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(threshold_val
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('MF ROC Curve for Popular Trimming (k=24)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for popular trimming with k=24
```



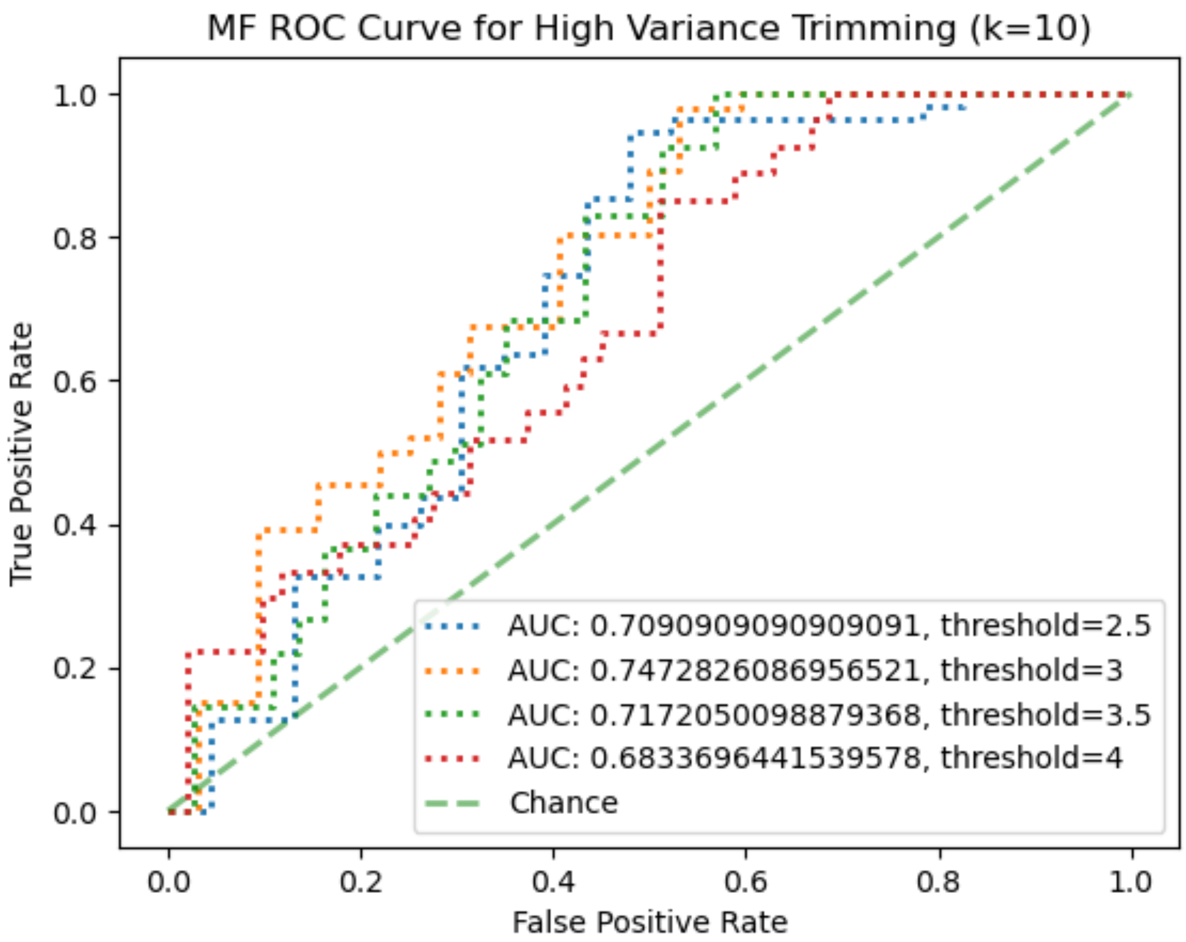
In [293]:
```
#Unpopular Trim
thresholds = [2.5, 3, 3.5, 4]
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = SVD(n_factors=18).fit(trainset).test(unpopular_trim(testset))
print("Results for unpopular trimming'} with k=18")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('MF ROC Curve for Unpopular Trimming (k=18)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for unpopular trimming'} with k=18
```



In [295]:
```
#High Variance Trim
thresholds = [2.5, 3, 3.5, 4]
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
trainset, testset = train_test_split(data, test_size=0.1)
res = SVD(n_factors=10).fit(trainset).test(high_var_trim(testset))
print("Results for high variance trimming'} with k=10")
fig, ax = plt.subplots()
for item in thresholds:
    thresholded_out = []
    for row in res:
        if row.r_ui > item:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label="AUC: "+str(auc(fpr,tpr))+', threshold='+str(item))
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='g', label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('MF ROC Curve for High Variance Trimming (k=10)')
plt.legend(loc="lower right")
plt.show()
```

```
Results for high variance trimming'} with k=10
```



QUESTION 11:

Designing a Na¨ıve Collaborative Filter: •Design a naive collaborative filter to predict the ratings of the movies in the original 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. •Performance on dataset subsets: For each of Popular, Unpopular and High-Variance test subsets - – Design a naive collaborative filter for each trimmed set and evaluate its performance using 10-fold cross validation. – Compute the average RMSE by averaging the RMSE across all 10 folds. Report the average RMSE.

In [343]:
```
from surprise.model_selection import KFold
def popular_trimming(df):
  df = pd.DataFrame(df, columns=['userId', 'movieId', 'rating'])
  filter = df['movieId'].value_counts() > 2
  filter_id = filter[filter].index.to_list()
  return df[df.movieId.isin(filter_id)].to_records(index=False)

def unpopular_trimming(df):
  df = pd.DataFrame(df, columns=['userId', 'movieId', 'rating'])
  filter = df['movieId'].value_counts() <= 2
  filter_id = filter[filter].index.to_list()
  return df[df.movieId.isin(filter_id)].to_records(index=False)

def high_variance_trimming(df):
  df = pd.DataFrame(df, columns=['userId', 'movieId', 'rating'])
  filter = df['movieId'].value_counts() >= 5
  filter_id = filter[filter].index.to_list()
  movie_variances = df.groupby('movieId')['rating'].var(ddof=0)
  low_var = movie_variances[movie_variances < 2].index.to_list()
  filter_id = [id for id in filter_id if id not in low_var]
  return df[df.movieId.isin(filter_id)].to_records(index=False)
def no_trimming(df):
  return df
reader = Reader(rating_scale=(0.5, 5))
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)

rating_record = defaultdict(list)
for user, _, rating, _ in data.raw_ratings:
  rating_record[user].append(rating)
rating = {}
for user, ratings in rating_record.items():
  rating[user] = np.mean(ratings)

for trimming in [no_trimming, popular_trimming, unpopular_trimming, high_variance_trimming]:
  rmse = []
  for _, test in KFold(10).split(data):
    test = trimming(test)
    pred = [rating[user] for user, _, _ in test]
    true = [r for _, _, r in test]
    rmse.append(np.sqrt(metrics.mean_squared_error(pred, true)))
  rmse = np.mean(rmse)
  print(f'Naive Collaborative Filter with {trimming.__name__} {rmse=}')
```

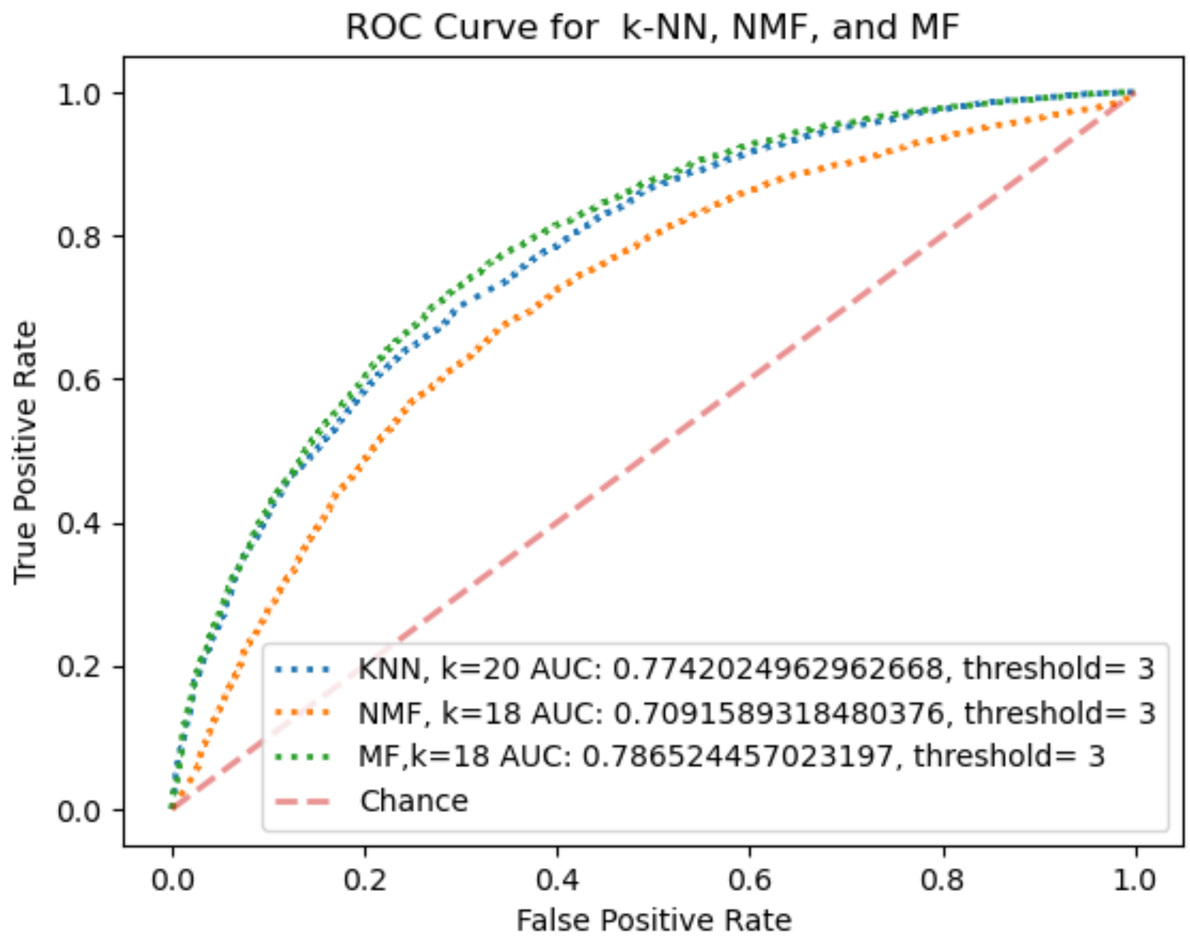
```
Naive Collaborative Filter with no_trimming rmse=0.9346752906105265
Naive Collaborative Filter with popular_trimming rmse=0.9243283649127395
Naive Collaborative Filter with unpopular_trimming rmse=0.9583897415046927
Naive Collaborative Filter with high_variance_trimming rmse=1.4319058156339024
```

QUESTION 12:

Comparing the most performant models across architecture: Plot the best ROC curves (threshold = 3) for the k-NN, NMF, 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 [356]:
```
# Load the data

reader = Reader(rating_scale=(0.5, 5.0))


# Define the k-NN collaborative filtering algorithm and train the models

knn=KNNWithMeans(k=20, sim_options={'name': 'pearson', 'user_based': True})
nmf=NMF(n_factors=18,biased=True)
mf=SVD(n_factors=18, biased=True)

algo = [(knn, "KNN, k=20"),(nmf, "NMF, k=18"),(mf, "MF,k=18")]
print("Results for No Trimming with k=20")
fig, ax = plt.subplots()
for i, algo_name in algo:
    data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)

    trainset, testset = train_test_split(data, test_size=0.1)
    res = i.fit(trainset).test(testset)
    thresholded_out = []
    for row in res:
        if row.r_ui > 3:
            thresholded_out.append(1)
        else:
            thresholded_out.append(0)
    fpr, tpr, thresholds = roc_curve(thresholded_out, [row.est for row in res])
    ax.plot(fpr, tpr,lw=2,linestyle=':',label= algo_name +" AUC: "+str(auc(fpr,tpr))+', threshold= 3')
ax.plot([0, 1], [0, 1], linestyle='--', lw=2, label='Chance', alpha=.5)
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve for  k-NN, NMF, and MF')
plt.legend(loc="lower right")
plt.show()
```

```
Results for No Trimming with k=20
Computing the pearson similarity matrix...
Done computing similarity matrix.
```



QUESTION 13:

Understanding Precision and Recall in the context of Recommender Systems: 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.

A:

Precision is the proportion of recommended items that are relevant to the user's interests. In other words, it is the ratio of the number of correctly recommended items to the total number of recommended items.

Recall is the proportion of relevant items that are recommended by the system. In other words, it is the ratio of the number of correctly recommended items to the total number of relevant items.

QUESTION 14:

Comparing the precision-recall metrics for the different models:

•For each of the three architectures: – Plot average precision (Y-axis) against t (X-axis) for the ranking obtained using the model's predictions.

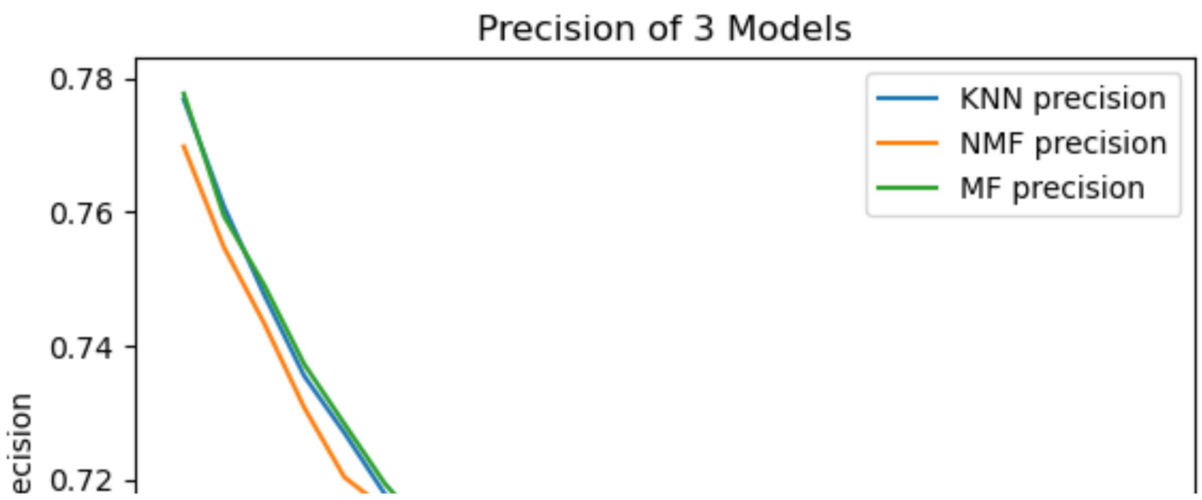
– Plot the average recall (Y-axis) against t (X-axis) and plot the average precision (Y-axis) against average recall (X-axis). – Use the best k found in the previous parts and sweep t from 1 to 25 in step sizes of 1. For each plot, briefly comment on the shape of the plot.

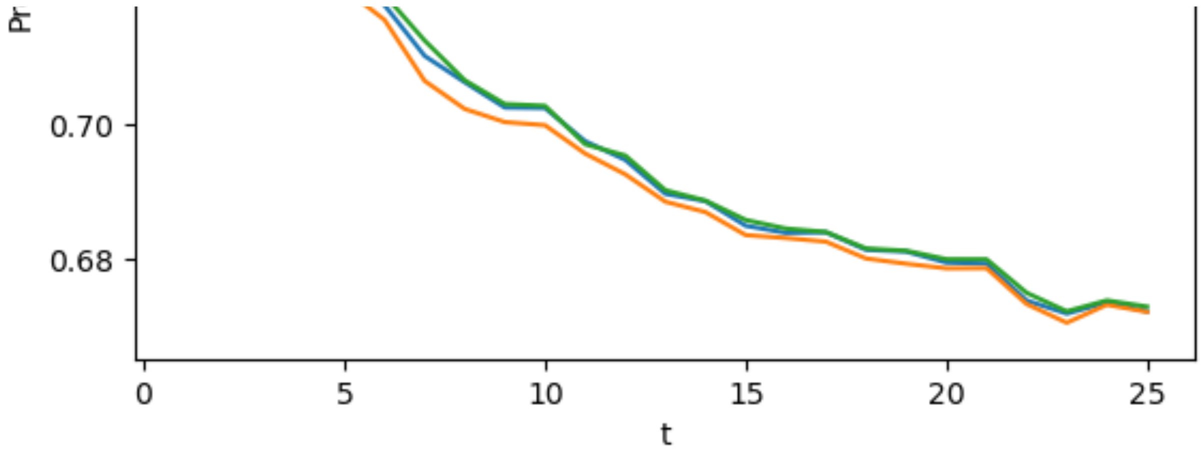
Hints: •Use threshold = 3 for obtaining the set G •Use 10-fold cross-validation to obtain the average precision and recall values for each value of t. To be specific, compute precision and recall for each user using equations 12 and 13 and then average across all the users in the dataset to obtain the precision and recall for this fold. Now repeat the above procedure to compute the precision and recall for all the folds and then take the average across all the 10-folds to obtain the average precision and average recall value for this value of t. •If |G| = 0 for some user in the validation set, then drop this user •If some user in the validation set has rated less than t items, then drop this user.

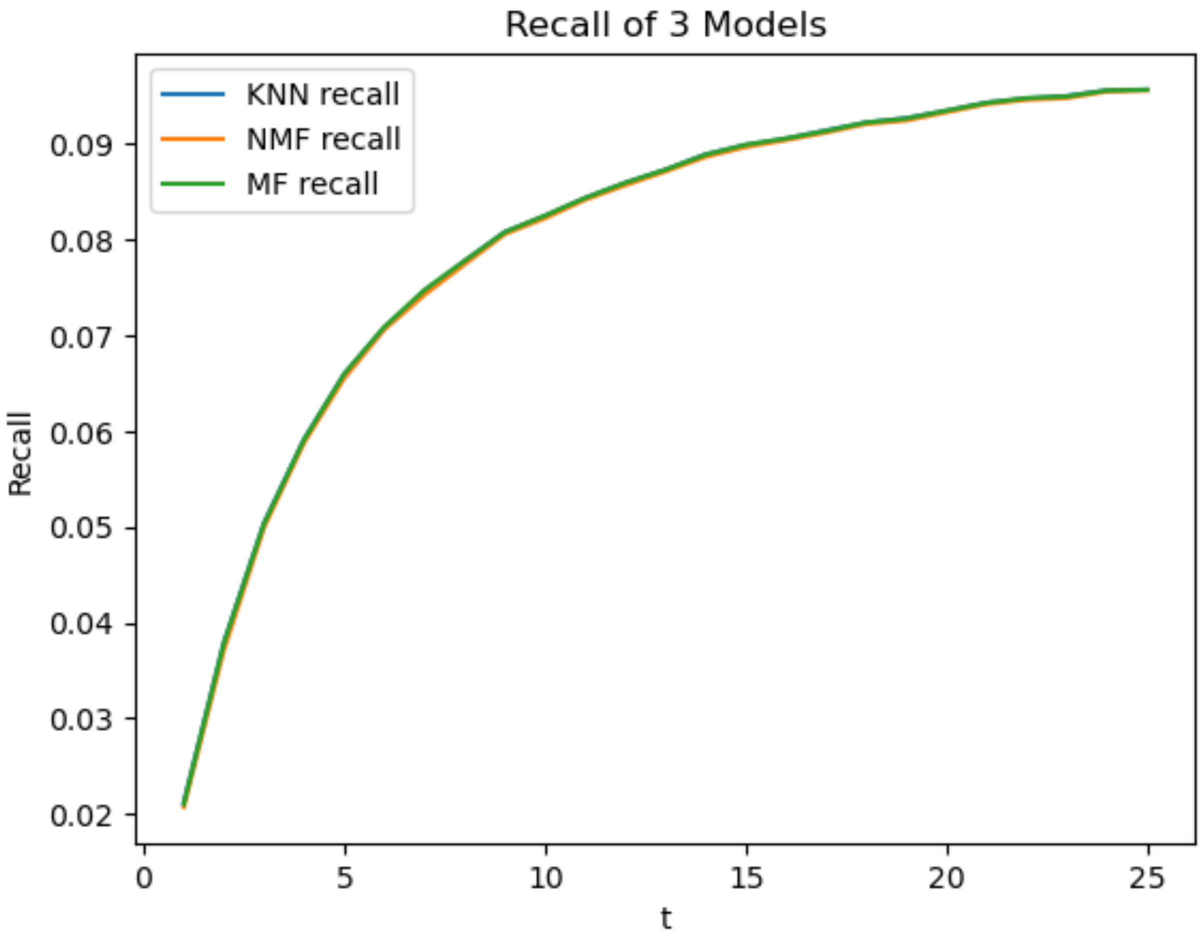
In [382]:

```python
#Plot Average Precision and Recall
from tqdm import tqdm
def get_metrics(pred, G, t, ur):
    S = {}
    for p in pred:
        if p.uid in S.keys():
            continue
        cur_user = set()
        for k in pred:
            if p.uid == k.uid:
                cur_user.add((k.iid, k.est))
        cur_user = sorted(list(cur_user), key=lambda x: x[1], reverse=True)[:t]
        S[p.uid] = [c[0] for c in cur_user]

    precision, recall = [], []
    for i in S:
        g = G[i]
        s = S[i]
        if len(g) == 0 or len(ur[i]) < t: continue
        intersection = list(set(g) & set(s))
        precision.append(len(intersection) / float(len(s)))
        recall.append(len(intersection) / float(len(g)))
    return np.mean(precision), np.mean(recall)
t_list = list(range(1, 26))
kf = KFold(n_splits=10)

G = {}
full_train = data.build_full_trainset()
for u in full_train.ur:
    ratings = full_train.ur[u]
    G[full_train.to_raw_uid(u)] = [full_train.to_raw_iid(i[0]) for i in ratings if i[1] > 3]

res_list = []
for t in tqdm(t_list):
    metrics_matrix = np.zeros((6, 10))
    for idx, (train, test) in enumerate(kf.split(data)):
        knn_pred = KNNWithMeans(k=20, sim_options={'name':'pearson'}, verbose=False).fit(train).test(tes
        nmf_pred = NMF(n_factors=20, verbose=False).fit(train).test(test)
        mf_pred = NMF(n_factors=2, biased=True, verbose=False).fit(train).test(test)

        metrics_matrix[0][idx], metrics_matrix[3][idx] = get_metrics(knn_pred, G, t, full_train.ur)
        metrics_matrix[1][idx], metrics_matrix[4][idx] = get_metrics(nmf_pred, G, t, full_train.ur)
        metrics_matrix[2][idx], metrics_matrix[5][idx] = get_metrics(mf_pred, G, t, full_train.ur)

    res_list.append(list(np.mean(metrics_matrix, axis=1)))
res_list = np.array(res_list)
plt.plot(t_list, res_list[:, 0], label='KNN precision')
plt.plot(t_list, res_list[:, 1], label='NMF precision')
plt.plot(t_list, res_list[:, 2], label='MF precision')
plt.legend()
plt.title("Precision of 3 Models")
plt.xlabel("t")
plt.ylabel("Precision")
plt.show()

plt.plot(t_list, res_list[:, 3], label='KNN recall')
plt.plot(t_list, res_list[:, 4], label='NMF recall')
plt.plot(t_list, res_list[:, 5], label='MF recall')
plt.legend()
plt.title("Recall of 3 Models")
plt.xlabel("t")
plt.ylabel("Recall")
plt.show()
```

```
100%|██████████████████████████████| 25/25 [22:51<00:00, 54.85s/it]
```



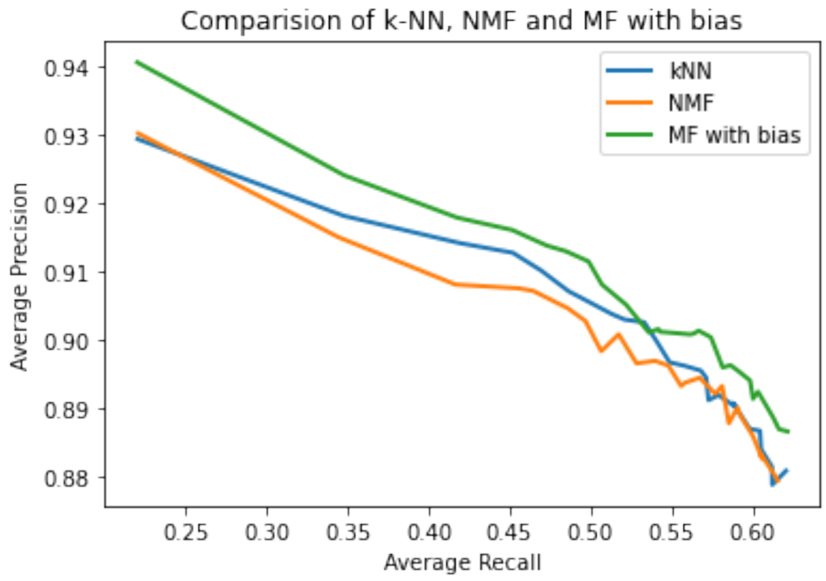




•Plot the best precision-recall curves obtained for the three models (k-NN, NMF, MF) in the same figure. Use this figure to compare the relevance of the recommendation list generated using k-NN, NMF, and MF with bias predictions.

In [ ]:

```
from sklearn import metrics

# Load data and split into train and test sets
data = Dataset.load_from_df(df[['userId', 'movieId', 'rating']], reader)
train, test = train_test_split(data, test_size=0.1)

# Fit models and make predictions on test set
knn_pred = KNNWithMeans(k=20, sim_options={'name':'pearson'}, verbose=False).fit(train).test(test)
nmf_pred = NMF(n_factors=20, verbose=False).fit(train).test(test)
mf_pred = NMF(n_factors=2, biased=True, verbose=False).fit(train).test(test)

# Define a list of (predictions, name) tuples for the three models
pred_list = [(knn_pred, "KNN"), (nmf_pred, "NMF"), (mf_pred, "MF")]

# Plot precision-recall curves for the three models
for pred, name in pred_list:
    y_true = [1 if i.r_ui >= 3 else 0 for i in test]
    score = [i.est for i in pred]
    precision, recall, _ = metrics.precision_recall_curve(y_true, score, pos_label=1)
    plt.plot(recall, precision, label=name)

plt.legend()
plt.title('Precision-recall curve for KNN, NMF and MF')
plt.xlabel('Recall')
plt.ylabel('Precision')
plt.show()
```


Comparision of k-NN, NMF and MF with bias
kNN
NMF
MF with bias
Average Precision
0.94
0.93
0.92
0.91
0.90
0.89
0.88
Average Recall
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60

In [ ]: