For code, please visit https://github.com/lumalav/CAP5610/blob/master/HW6/HW6.ipynb

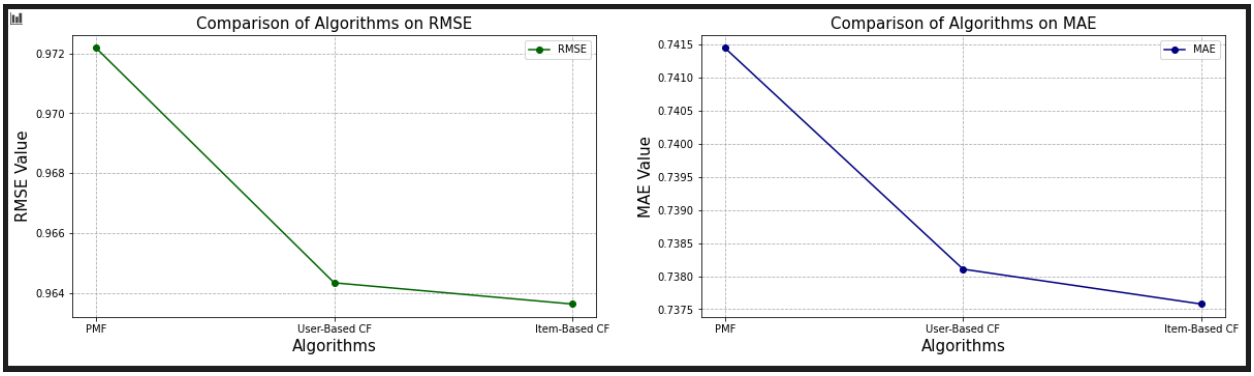
3) a) ratings_small.csv was used because of different memory issues that ratings.csv was causing.

c and d)

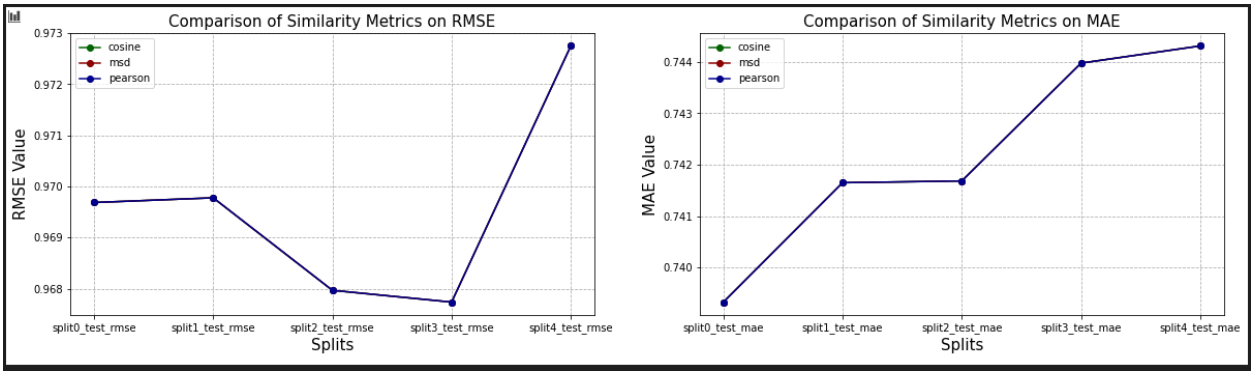
Based on the results, it seems that item-based collaborative filtering is the winner for the ratings_small.csv data.

PMF: 0.972 (RMSE), 0.741 (MAE)
UB: 0.964 (RMSE), 0.738 (MAE)
IB: 0.963 (RMSE), 0.737 (MAE)

```
print(p_m_f.best_score)
{'rmse': 0.9721800346212006, 'mae': 0.7414493302453836}
print(user_based_filtering.best_score)
{'rmse': 0.9643373797589329, 'mae': 0.7381059518572946}
print(item_based_filtering.best_score)
{'rmse': 0.9636279597103437, 'mae': 0.7375786400847663}
```



e) Yes, it is consistent. The same average values of RMSE and MAE were given by the three different metrics. It seems that the greater effect on the performance of the models is caused by the number of neighbors.



f and g)

As K slowly increases, RMSE and MAE decrease until finding the lowest point at 15. After this, RMSE and MAE increase again. For both models, user-based and item-based collaborative filtering, the optimal amount of neighbors is 15.