Data Mining HW - 3

1. Spark Version: 2.3.3   Python Version: 3.6

2. Task2 Collaborative Filtering Algorithms:
   a. Model Based CF

RMSE	1.118055136203206
Time(sec)	22.59599995613098

   b. User Based CF

RMSE	1.0988684965031237
Time(sec)	113.89770817756653

   c. Item Based CF with LSH

RMSE	1.0754618108207834
Time(sec)	40.2520010471344

3. Comparison of Item Based CF with LSH and without LSH:

   a. Faster:
      Item Based CF with LSH runs faster than Item Based CF without LSH. With LSH, we compare a business with only similar businesses. Without LSH, we calculate all the businesses the user has rated so far. Obviously, this number will be much larger than the similar business pairs we computed using LSH. Hence with LSH, we get faster results.

   b. Better Recommendation:
      Item Based CF without LSH considers all businesses the user has rated. In this list, few or maybe most businesses do not resemble the business we are trying to predict. Whereas Item Based CF with LSH only considers similar businesses. The key idea is similar businesses will be rated similarly by the same user. Hence, Item Based CF with LSH provides better recommendation compared to Item Based CF without LSH.