ASSIGNMENT 3

TASK 1:

Jaccard based LSH
Performed Locality Based Sensitive Hashing on Jaccard Similarity as metric.

Hash Function Used:
$f(x) = ((ax + b) \% p) \% m$

Efficiency Table:

Precision	1.00
Recall	0.940993
Duration	30.602 seconds

TASK 2:

Case 1: Model-based CF recommendation system with Spark MLlib:

RMSE	1.15237
DURATION	30.9080 seconds

Case 2: User-based CF recommendation system:

RMSE	1.1097
DURATION	172.153 seconds

Case 3: Item-based CF recommendation system:

RMSE	1.136
DURATION	176.094 seconds

Case 4: Item-based CF recommendation system with Jaccard based LSH:

RMSE	1.075499
DURATION	77.24684 seconds

WHY IS CASE 4 BETTER THAN CASE 3?

Item-based CF recommendation system with Jaccard based LSH is better than : Item-based CF recommendation system because, using Locality Sensitive Hashing (Case 4) we get a set of similar businesses. Hence we calculate weights (Pearson Correlation) on only those similar pairs. In Case 3, we compute Pearson Correlation in all pairs. This takes time. Hence Item Based with LSH for finding similar pairs, is much faster and better. This result decrease in computation.