

## DESCRIPTION

**Name:** Parinita Badre

**USC ID:** 8804379764

**USC email:** badre@usc.edu

1. Spark Version - 2.3.3 , Python Version - 3.6

2. Task 2 - RMSE and Runtime

A. Model based CF

<b>RMSE</b>	1.06859761038715
<b>Time (sec)</b>	22.9352829456329

B. User based CF

<b>RMSE</b>	1.1631828691011
<b>Time (sec)</b>	65.6079952716827

C. Item based CF

	<b>With LSH</b>	<b>Without LSH</b>
<b>RMSE</b>	1.06093521514426	1.13465312696121
<b>Time (sec)</b>	62.4504079818726	226.741677999496

3. When we implement item based CF without LSH, we generate set of all possible items (businesses) that are similar to the item we want to predict. It's possible that most of the items do not have any corated users and each item has only few corated users. We can save this time by implementing LSH first which uses Jaccard similarity and generate candidate pairs with higher similarity. We use these pairs to generate similarity using Pearson correlation which requires significantly less time.