This file is used to document our procedures through this project.

**Data cleaning and merging:**

To merge tmdb dataset with MovieLens dataset by movie names, we cleaned movie.csv file in R Studio with regular expression. (see clean\_movie\_name.R in the folder)

To narrow down our scope, we decided to keep users with more than 50/100/150 rating records. We are going to take column “rating” for further prediction.