Your submission for this project is three files:

1. Your report in Rmd format
2. Your report in PDF format (knit from your Rmd file)
3. A script in R format that generates your predicted movie ratings and RMSE score (should contain all code and comments for your project)

You may upload the three files directly to the edX platform or submit a GitHub link in the text response box below.

To upload and submit your files press the "Choose Files" button, select three files at once (using the control key on a Windows machine or command key on a Mac) and press "Choose," type a description for each (PDF, Rmd, R), and then press the "Upload files" button. If uploading files, we recommend also providing a link to a GitHub repository containing the three files above in case there is a problem with the upload process.

Note that when downloading files for peer assessments, R and Rmd files will be downloaded as txt files by default.

**MovieLens Grading Rubric**

The following is the grading rubric your peers will be using to evaluate your project. There are also opportunities for your peers to provide written feedback as well (required for some categories and optional for others). You are encouraged to give thoughtful, specific written feedback to your peers whenever possible (i.e., more than just "good job" or "not enough detail").

Note that to receive full marks on this project, you **may not simply copy code from other courses in the course series and be done with your analysis**. Your work on this project needs to build on code that is already provided.

After you submit your project, **please check immediately after submitting to make sure that all files were correctly uploaded**. Occasionally, there are file upload failures, and it's easiest to fix if these are caught early.

**Files (10 points possible)**

The appropriate files are submitted in the correct formats: a report in both PDF and Rmd format and an R script in R format.

* + 0 points: No files provided AND/OR the files provided appear to violate the [edX Honor Code](https://www.edx.org/edx-terms-service#honor-code).
  + 3 points: Multiple requested files are missing and/or not in the correct formats.
  + 5 points: One file is missing and/or not in the correct format.
  + 10 points: All 3 files were submitted in the requested formats.

**Report (40 points possible)**

The report documents the analysis and presents the findings, along with supporting statistics and figures. The report must be written in English and uploaded. The report should be written assuming that the reader is not familiar with the project or the data. The report must include the RMSE generated. The report must include at least the following sections:

* + 1. an **introduction/overview/executive summary** section that describes the dataset and summarizes the goal of the project and key steps that were performed
    2. a **methods/analysis** section that explains the process and techniques used, including data cleaning, data exploration and visualization, insights gained, and your modeling approach
    3. a **results** section that presents the modeling results and discusses the model performance
    4. a **conclusion** section that gives a brief summary of the report, its limitations and future work
  + 0 points: The report is either not uploaded or contains very minimal information AND/OR the report appears to violate the [edX Honor Code](https://www.edx.org/edx-terms-service#honor-code).
  + 10 points: Multiple required sections of the report are missing.
  + 15 points: The methods/analysis or the results section of the report is missing or missing significant supporting details. Other sections of the report are present.
  + 20 points: The introduction/overview or the conclusion section of the report is missing, not well-presented or not consistent with the content.
  + 20 points: The report includes all required sections, but the report is significantly difficult to follow or missing supporting detail in multiple sections.
  + 25 points: The report includes all required sections, but the report is difficult to follow or missing supporting detail in one section.
  + 30 points: The report includes all required sections and is well-drafted and easy to follow, but with minor flaws in multiple sections.
  + 35 points: The report includes all required sections and is easy to follow, but with minor flaws in one section.
  + 40 points: The report includes all required sections, is easy to follow with good supporting detail throughout, and is insightful and innovative.

**Code (25 points)**

The code in the R script should should be well-commented and easy to follow. The code provided in the R script should contain all of the code and comments for your project. You are **not required** to run the code provided (although you may if you wish), but you should visually inspect it.

* + 0 points: No code provided AND/OR the code appears to violate the [edX Honor Code](https://www.edx.org/edx-terms-service#honor-code).
  + 10 points: Code appears that it would not run/is very difficult to follow or interpret AND/OR is not consistent with the report.
  + 15 points: Code appears that it would run without throwing errors, can be followed, is at least mostly consistent with the report, but has no comments or explanation.
  + 15 points: Code is simply a copy of code provided in previous courses in the series without expanding on it, but is otherwise well-commented.
  + 20 points: Code appears that it would run without throwing errors, can be followed, is largely consistent with the report, but without sufficient comments or explanations.
  + 25 points: Code is easy to follow, is consistent with the report, and is well-commented.

**RMSE (25 points)**

Provide the appropriate score given the reported RMSE. *Please be sure not to use the***validation***set (the final hold-out test set) for training or regularization - you should create an additional partition of training and test sets from the provided***edx***dataset to experiment with multiple parameters or use cross-validation.*

* + 0 points: No RMSE reported AND/OR code used to generate the RMSE appears to violate the [edX Honor Code](https://www.edx.org/edx-terms-service#honor-code).
  + 5 points: RMSE >= 0.90000 AND/OR the reported RMSE is the result of overtraining (validation set - the final hold-out test set - ratings used for **anything** except reporting the final RMSE value) AND/OR the reported RMSE is the result of simply copying and running code provided in previous courses in the series.
  + 10 points: 0.86550 <= RMSE <= 0.89999
  + 15 points: 0.86500 <= RMSE <= 0.86549
  + 20 points: 0.86490 <= RMSE <= 0.86499
  + 25 points: RMSE < 0.86490

Have a question about the MovieLens project? Need some feedback on the best approach to take or some troubleshooting for a snippet of your code? You can ask your questions here!

You are encouraged to discuss **general approaches** to the MovieLens project. It is okay to post **small snippets**of code if you're having trouble getting a particular piece of code to run. However, you **may not post your entire R script** for the project.