=====================================================================

**INDIVIDUAL PROJECT**

=====================================================================

Files (5 points possible)

The appropriate files are submitted in the correct formats:

* a report in both PDF and Rmd format
* an R script in R format.

Report (25 points possible)

The report documents the analysis and presents the findings, along with supporting statistics and figures. In order to demonstrate your understanding of course material, please provide thorough explanation or justification for various steps of your project, such as why a specific train/test split (e.g. 50/50 vs 90/10) or algorithm was used. The report must be written in English and uploaded. The report should assume that the reader is not familiar with the project or the dataset. The report must include at least the following sections:

* **An introduction/overview/executive summary section** that describes the dataset and variables, and summarizes the goal of the project and key steps that were performed.
* **A methods/analysis section** that explains the process and techniques used, including data cleaning, data exploration and visualization, any insights gained, and your modeling approach. At least two different models or algorithms must be used, with at least one being more advanced than linear or logistic regression for prediction problems.
* **A results section** that presents the modeling results and discusses the model performance.
* **A conclusion section** that gives a brief summary of the report, its potential impact, its limitations, and future work.
* **A references section** that lists sources for datasets and/or other resources used, if applicable.

0 points: The report is either not uploaded or contains very minimal information AND/OR the report is not written in English AND/OR the report appears to violate the terms of the edX Honor Code.

5 points: One or more required sections of the report are missing.

10 points: The report includes all required sections, but the report is significantly difficult to follow or missing significant supporting detail in multiple sections.

15 points: The report includes all required sections, but the report has flaws: it is difficult to follow and/or missing supporting detail in one section and/or has minor flaws in multiple sections and/or does not demonstrate mastery of the content.

15 points: The report is otherwise fine, but the project is a variation on the MovieLens project.

20 points: The report includes all required sections and is easy to follow, but with minor flaws in one section.

25 points: The report includes all required sections, is easy to follow with good supporting detail throughout, and is insightful and innovative.

Code (20 points)

The code in the R script should run without errors and should be well-commented and easy to follow. It should also use relative file paths and automatically install missing packages. The dataset you use should either be automatically downloaded by your code or provided in your GitHub repo along with the rest of your files (Rmd, PDF, R). If your dataset is provided as a zip file in GitHub, your code should automatically unzip and load it. The R script should contain all of the code and comments for your project.

0 points: Code has major errors that cause it to not run and/or produces many errors and/or code appears to violate the terms of the edX Honor Code.

5 points: Code runs but does not produce output consistent with what is presented in the report and/or there is overtraining (the test set is used for training steps).

10 points: Code runs but is difficult to follow and/or may not produce output entirely consistent with what is presented in the report.

10 points: Code is otherwise fine but the project is a variation on the MovieLens project.

15 points: Code runs, can be followed, is at least mostly consistent with the report, but is lacking (sufficient) comments and explanation OR uses absolute paths instead of relative paths OR does not automatically install missing packages External link OR does not provide easy access to the dataset (either via automatic download or inclusion in a GitHub repository). Code with minor flaws that require small modifications (such as commenting out an inappropriately included line of text) to run may also be awarded this grade.

20 points: Code runs easily, is easy to follow, is consistent with the report, and is well-commented. All file paths are relative and missing packages are automatically installed External link with if(!require) statements.

Have a question about the choose your own project? Want to bounce some ideas for an analysis to do or a dataset to pick off someone else? 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 your 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.