ᛦ **eegshou** / **dsc-mod-1-project-v2-1-online-ds-sp-000**

forked from learn-co-students/dsc-mod-1-project-v2-1-online-ds-sp-000

⚖ View license

☆ **0** stars   ᛦ **65** forks

☆ Star	⊙ Watch ⌄

Code	Pull requests	Actions	Projects	Security	Insights	Settings

ᛦ **shou** had recent pushes less than a minute ago

Compare & pull request

ᛦ **shou** ⌄   ···

This branch is 3 commits ahead of learn-co-students:master.

ᛦ Contribute ⌄   ↻ Fetch upstream ⌄

🔳 **eegshou** ···   now   🕘

View code

☰ README.md   ✎

Module 1 Final Project

Introduction

In this lesson, we'll review all of the guidelines and specifications for the final project for Module 1.

Objectives

You will be able to:

- Describe all required aspects of the final project for Module 1

- Describe all required deliverables
- Describe what constitutes a successful project
- Describe what the experience of the project review should be like

Final Project Summary

You've made it all the way through the first module of this course - take a minute to celebrate your awesomeness!

All that remains in Module 1 is to put our newfound data science skills to use with a final project! You should expect this project to take between 20 and 25 hours of solid, focused effort. If you're done way quicker, go back and dig in deeper or try some of the optional "level up" suggestions. If you're worried that you're going to get to 30 hrs and still not even have the data imported, reach out to an instructor in Slack ASAP to get some help!

The Project

Microsoft sees all the big companies creating original video content, and they want to get in on the fun. They have decided to create a new movie studio, but the problem is they don't know anything about creating movies. They have hired you to help them better understand the movie industry. Your team is charged with doing data analysis and creating a presentation that explores what type of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the CEO can use when deciding what type of films they should be creating.

The Dataset

You may scrape or make API calls to get additional data, but included in the repository (in the folder `zippedData`) is some movie-related data from:

- Box Office Mojo
- IMDB
- Rotten Tomatoes
- TheMovieDB.org

The Deliverables

For online students, there will be five deliverables for this project (Note: On-campus students may have different requirements, please speak with your instructor):

1. A well documented **Jupyter Notebook** containing any code you've written for this project and comments explaining it. This work will need to be pushed to your GitHub repository in order to submit your project.

2. An organized **README.md** file in the GitHub repository that describes the contents of the repository. This file should be the source of information for navigating through the repository.

3. A short **Keynote/PowerPoint/Google Slides presentation** (delivered as a PDF export) giving a high-level overview of your methodology and recommendations for non-technical stakeholders. Make sure to also add and commit this pdf of your non-technical presentation to your repository with a file name of presentation.pdf.

4. **A Blog Post**

5. A **Video Walkthrough** of your non-technical presentation. Some common video recording tools used are Zoom, Quicktime, and Nimbus. After you record your presentation, publish it on a service like YouTube or Google Drive, you will need a link to the video to submit your project.

Note: On-campus students may have different requirements, please speak with your instructor.

Jupyter Notebook Must-Haves

For this project, your Jupyter Notebook should meet the following specifications:

Organization/Code Cleanliness

- The notebook should be well organized, easy to follow, and code should be commented where appropriate.
  - Level Up: The notebook contains well-formatted, professional looking markdown cells explaining any substantial code. All functions have docstrings that act as professional-quality documentation
- The notebook is written for technical audiences with a way to both understand your approach and reproduce your results. The target audience for this deliverable is other

data scientists looking to validate your findings.

Visualizations & EDA

- Your project contains at least 4 meaningful data visualizations, with corresponding interpretations. All visualizations are well labeled with axes labels, a title, and a legend (when appropriate)
- You pose at least 3 meaningful questions and answer them through EDA. These questions should be well labeled and easy to identify inside the notebook.
  - **Level Up**: Each question is clearly answered with a visualization that makes the answer easy to understand.
- Your notebook should contain 1 - 2 paragraphs briefly explaining your approach to this project.

Non-Technical Presentation Must-Haves

Another deliverable should be a Keynote, PowerPoint or Google Slides presentation delivered as a pdf file in your fork of this repository with the file name of `presentation.pdf` detailing the results of your project. Your target audience is non-technical people interested in using your findings to make decisions for creating movies.

Your presentation should:

- Contain between 5 - 10 professional-quality slides.
  - **Level Up**: The slides should use visualizations whenever possible, and avoid walls of text.
- Take no more than 5 minutes to present.
- Avoid technical jargon and explain the results in a clear, actionable way for non-technical audiences.

Blog Post Must-Haves

Refer back to the Blogging Guidelines for the technical requirements and blog ideas.

The Process

(Note: On-campus students may have different processes, please speak with your instructor)

1. Getting Started

Please start by reviewing this document. If you have any questions, please ask them in Slack ASAP so (a) we can answer the questions and (b) so we can update this repository to make it clearer.

Be sure to let the instructor team know when you've started working on a project, either by reaching out over Slack or, if you are in a full-time or part-time cohort, by connecting with your Cohort Lead in your weekly 1:1. If you're not sure who to reach out to, post in the #online-ds-sp-000 channel in Slack.

Once you're done with the 10 sections in module 1, please start on the project. Do that by forking this repository, cloning it locally, and working in the `student.ipynb` file. Make sure to also add and commit a pdf of your presentation to the repository with a file name of `presentation.pdf`.

2. The Project Review

Note: On-campus students may have different review processes, please speak with your instructor.

> **When you start on the project, please also reach out to an instructor immediately to schedule your project review** (if you're not sure who to schedule with, please ask in Slack!)

What to expect from the Project Review

Project reviews are focused on preparing you for technical interviews. Treat project reviews as if they were technical interviews, in both attitude and technical presentation *(sometimes technical interviews will feel arbitrary or unfair - if you want to get the job, commenting on that is seldom a good choice).*

The project review is comprised of a 45 minute 1:1 session with one of the instructors. During your project review, be prepared to:

1. Deliver your PDF presentation to a non-technical stakeholder.

In this phase of the review (~10 mins) your instructor will play the part of a non-technical stakeholder that you are presenting your findings to. The presentation should not exceed 5 minutes, giving the "stakeholder" 5 minutes to ask questions.

In the first half of the presentation (2-3 mins), you should summarize your methodology in a way that will be comprehensible to someone with no background in data science and that will increase their confidence in you and your findings. In the second half (the remaining 2-3 mins) you should summarize your findings and be ready to answer a couple of non-technical questions from the audience. The questions might relate to technical topics (sampling bias, confidence, etc) but will be asked in a non-technical way and need to be answered in a way that does not assume a background in statistics or machine learning. You can assume a smart, business stakeholder, with a non-quantitative college degree.

2. Go through the Jupyter Notebook, answering questions about how you made certain decisions. Be ready to explain things like:

```
 * "How did you pick the question(s) that you did?"
 * "Why are these questions important from a business perspective?"
 * "How did you decide on the data cleaning options you performed?"
 * "Why did you choose a given method or library?"
 * "Why did you select those visualizations and what did you learn from each of
 them?"
 * "Why did you pick those features as predictors?"
 * "How would you interpret the results?"
 * "How confident are you in the predictive quality of the results?"
 * "What are some of the things that could cause the results to be wrong?"
```

Think of the first phase of the review (~30 mins) as a technical boss reviewing your work and asking questions about it before green-lighting you to present to the business team. You should practice using the appropriate technical vocabulary to explain yourself. Don't be surprised if the instructor jumps around or sometimes cuts you off - there is a lot of ground to cover, so that may happen.

If any requirements are missing or if significant gaps in understanding are uncovered, be prepared to do one or all of the following:

- Perform additional data cleanup, visualization, feature selection, modeling and/or model validation
- Submit an improved version
- Meet again for another Project Review

What won't happen:

- You won't be yelled at, belittled, or scolded
- You won't be put on the spot without support
- There's nothing you can do to instantly fail or blow it

Please note: We need to receive the URL of your repository at least 24 hours before and please have the project finished at least 3 hours before your review so we can look at your materials in advance.

Submitting your Project

You're almost done! In order to submit your project for review, include the following links to your work in the corresponding fields on the right-hand side of Learn.

1. **GitHub Repo:** Now that you've completed your project in Jupyter Notebooks, push your work to GitHub and paste that link to the right. (If you need help doing so, review the resources here.) *Reminder: Make sure to also add and commit a pdf of your non-technical presentation to the repository with a file name of presentation.pdf.*
2. **Blog Post:** Include a link to your blog post.

3. **Record Walkthrough:** Include a link to your video walkthrough.

Hit "I'm done" to wrap it up. You will receive an email in order to schedule your review with your instructor.

Grading Rubric

Online students can find a PDF of the grading rubric for the project here. On-campus students may have different review processes, please speak with your instructor.

Summary

The end of module projects and project reviews are a critical part of the program. They give you a chance to both bring together all the skills you've learned into realistic projects and to practice key "business judgement" and communication skills that you otherwise might not get as much practice with.

The projects are serious and important. They are not graded, but they can be passed and they can be failed. Take the project seriously, put the time in, ask for help from your peers or instructors early and often if you need it, and treat the review as a job interview and you'll do great. We're rooting for you to succeed and we're only going to ask you to take a review again if we believe that you need to. We'll also provide open and honest feedback so you can improve as quickly and efficiently as possible.

Finally, this is your first project. We don't expect you to remember all of the terms or to get all of the answers right. If in doubt, be honest. If you don't know something, say so. If you can't remember it, just say so. It's very unusual for someone to complete a project review without being asked a question they're unsure of, we know you might be nervous which may affect your performance. Just be as honest, precise and focused as you can be, and you'll do great!

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%