

DS4A / Colombia 4.0:

Practicum Guidelines

This document will provide you with important information about the final project, or Practicum. The Practicum is an essential part of the DS4A program and it provides you with the opportunity to apply what you have learned throughout the program.

| | |
|--|----------|
| Basic Information | 2 |
| Topics | 2 |
| Topic Selection and Investigative Focus | 2 |
| Exploratory Data Analysis | 3 |
| Final Submissions | 4 |
| Final Report | 4 |
| Rubric | 4 |
| Datafolio | 5 |
| Presentation | 5 |
| Code | 5 |
| Timeline of Deliverables | 6 |
| Prizes & Accolades | 6 |

Basic Information

The Practicum allows you to demonstrate the knowledge you have gained from the program in a quasi-practical setting. Teams will consist of 5 - 7 participants each.

The Practicum is meant to incorporate skills from the entire curriculum and is broken up into four main components:

- A final report which details the analysis conducted, the hypotheses tested, and the resulting conclusions and recommendations.
- A final presentation where participants explain their process, key findings, and actions that can be taken based on those findings.
- A Datafolio that highlights the analysis and results of the project in a concise visual way.
- The code you used to conduct your analysis.

We encourage teams to use technologies taught in class while working on their projects, but they are not required to do so.

Topics

Since this is a condensed program, your team can choose between any of the following topics (each with its own respective problem statement to jumpstart your thought process):

- [Lodging and rental industry](#)
- [Residential water quality](#)
- [Infectious diseases](#)
- [E-commerce](#)

Additionally, each of the above topics come with several vetted and cleaned datasets that make working on the problem easier.

If your group wants to work on your own topic, please ask your TA for approval. When asking your TA, you must provide both the custom topic you want to cover along with 3 relevant (and reasonably large) datasets. Your request must be made by **November 23rd** at 11:59 COT to be eligible for approval.

Topic Selection and Investigative Focus

The first step of the project is selecting your topic and an investigative focus. Each topic comes with a number of datasets; you are not expected to use all of the provided datasets, just the ones relevant to your investigative focus.

For example, consider two of the sample focuses provided in the lodging and rental industry topic:

- What trends can you find in the Airbnb rental calendar over time, and how might these be explained by listing-specific and/or neighborhood-level factors?
- Analyze the locational patterns of Airbnb rentals. What relationships can you find between these and the locations of various types of metropolitan area venues?

As you can see, both of these focuses are based on the same general topic, but investigating them relies on the use of different datasets.

For this project submission, in your report, you must state which topic you are working on and what your related investigative focus/question is (you may use a provided sample focus/question if you wish).

In addition, you should start your data wrangling process (where you clean/prepare your datasets for easy use in analytics tools). In your report, you should outline which datasets you plan to use, how you think they are relevant to your focus, and what steps you have taken to clean/prepare your data. You are also welcome (and encouraged) to use additional external datasets in your project work.

This portion of your report must be submitted by 11:59pm COT on **November 28th**, 2020.

Exploratory Data Analysis

The second portion of your report is conducting your exploratory data analysis (EDA). This step is critical to the success of your project as it allows you to ask interesting and pointed questions as you drill down into your data. Generally, the most successful projects are those that ask the most interesting and useful questions, not those with the most advanced analysis techniques.

At this point, you may want to change your investigative focus. You are welcome to do so, just remember to update the investigative focus portion of your report when making this submission.

In your report, you should include all of your EDA. Based on the results of your EDA, you must create hypotheses you would like to test/verify and provide a preliminary plan as to how they can be tested.

This portion of your report must be submitted by 11:59pm COT on **December 5th**, 2020.

Final Submissions

The following submissions are all due by 11:59 COT on **December 19th**, 2020.

Final Report

The final report is a document that catalogues your team's execution of their Practicum.

You are free to structure your report in whatever format you think best communicates your work. However, your report should cover the following:

- **Non-Technical Executive Summary**
 - *Insightfulness of Conclusions.* What is the question that your team set out to answer, and how did you choose it? Are your conclusions precise and nuanced, as opposed to blanket (over)generalizations?
- **Technical Exposition**
 - *Wrangling and Cleaning Process.* Did you conduct proper quality control and handle common error types? How did you transform the datasets to better use them together? What sorts of feature engineering did you perform? Please describe your process in detail within your Report.
 - *Investigative Depth.* How did you conduct your exploratory data analysis (EDA) process? What other hypothesis tests and ad-hoc studies did you perform, and how did you interpret the results of these? What patterns did you notice, and how did you use these to make subsequent decisions?
 - *Analytical & Modeling Rigor.* What assumptions and choices did you make, and what was your justification for them? How did you perform feature selection? If you built models, how did you analyze their performance, and what shortcomings do they exhibit? If you constructed visualizations and/or conducted statistical tests, what was the motivation behind the particular ones you built, and what do they tell you?

Examples: Here are some examples of good past reports for different topics:

- [Example 1](#)
- [Example 2](#)
- [Example 3](#)

Rubric

The following is the rubric used to assess your final report. It outlines the important elements of your report for grading purposes. Teams may use this as a guide for successful final projects.

[Click here to see the grading rubric.](#)

Datafolio

When working on your Datafolio, please consult the [datafolio guidelines](#). Your grading will be based on two key elements: the effectiveness of your *visualizations* and how effectively your data *tells a story*.

For your *visualizations*, consider: How heavily does your datafolio rely on text when graphics, flowcharts, or other visual representations would have been more effective? How well are the charts and graphs interpreted by accompanying text?

For your *data storytelling*, consider: Did you find a compelling narrative that effectively communicates and highlights your analysis? How clear is the phrasing of the question and main insights? Does the layout of the Datafolio complement the flow of the analysis? Please don't just copy and paste text from your report.

Presentation

The presentation will be showcased on the last day of the program. It should discuss the importance of the problem your team posed as well as the key analysis, conclusions, and recommendations made throughout your report.

The most successful presentations usually have a great amount of interactive and meaningful data visualization and very subtle technical exposition.

Format: The presentations will generally:

- Use Powerpoint and other interactive digital tools.
- Be approximately 5 minutes.
- Focus on the overall process (investigative focus selection, methods and tools used, end results, and deliverables) over specific technical details.

Code

You must submit all of the code that contributed to the creation of your final report. This includes the code used for data wrangling, modelling, and visualizations.

Timeline of Deliverables

The key deliverables are summarized below for your convenience.

| Week | Milestone | Deliverable | Deadline |
|------|---|--|----------|
| 1 | Team formation + Decide on Project topic + Start data wrangling | N/A | N/A |
| 2 | Investigative focus determined | Submit report selecting project + outlining investigative focus + initial data wrangling process | Nov 28 |
| 3 | EDA complete + Key hypotheses determined | Submit report outlining EDA and how it was conducted. Based on EDA, explain how key hypotheses to test were decided on | Dec 5 |
| 4 | Conduct hypothesis testing | N/A | N/A |
| 5 | Apply linear modeling + Answer your questions/draw conclusions | Submit final report + Datafolio + Recorded presentation | Dec 19 |

Prizes & Accolades

The top teams will receive prizes, accolades and superior distinction at the end of the program. These teams will be recognized by Correlation One and MinTIC, and will be publicly announced via social media and press releases.