



[Table of Contents](#) > [Final Project](#) > Final Project: Instructions

Final Project: Instructions



Final Project

Purpose & Overview


DS 700 provides foundational knowledge of data science. Data science focuses on utilizing and analyzing data to solve complex business and societal problems, improve decision making, and foment innovation. To achieve these goals, a cohesive understanding of multiple disciplines--including statistics, math, computer science, information technology, business management, and communication--is essential.

The purpose of the final project is to solve a real-world business problem through data science and to demonstrate an understanding of interdisciplinary ideas, concepts, and techniques that constitute data science.

The final project is based on a case study that was designed and written to be used as a group exercise. However, each learner in DS700 will complete the project *individually*. The following sections explain how we will adapt the case study as it is published to allow individual work.


What You'll Need

To get started, you need to:

- purchase the Fargo Health Group: Managing the Demand for Medical Examinations Using Predictive Analytics (<https://hbr.org/product/fargo-health-group-managing-the-demand-for-medical-examinations-using-predictive-analytics/BAB266-PDF-ENG>)  case study.
- download the Dataset.xlsx file (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-9256615>), which is the data set used for the project.
- review this explanation of the data set (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-9602938>).

Changes to the Case Study

To make the case suitable for individual work, please note and follow the following changes to the published case study.

Page	Section	Original information	CHANGES
4	Assignment Details	As written, the case study states that, "each presenter will have anywhere from 3 to 5 minutes for his/her pitch."	You will each create and submit a 10-minute narrated PowerPoint presentation to discuss your work. You can find information on how to add narration to PowerPoint from Microsoft Office Support  .
5	Assignment Details	The case study states that, "there is no set page limit for the written deliverable report."	This is no longer valid. Submit a maximum of 10 pages for the deliverable report.

the written deliverable report.

Pages 4-5 contain a number of questions to answer and points to address in the final presentation and report.

In addition to the questions in the published case study, answer this question: **What are the ethical implications for making forecasts using this case study?**

HINT In answering this question, you may consider the following dimensions:

Context: What was the original purpose of the collection? How close is the new use to its original purpose?

Consent: Was informed consent necessary from affected patients before data collection? If so, did they provide informed consent prior to data collection? Did they have an opportunity to decline?

Reasonability: Is the depth and breadth of the dataset reasonable for the forecast?

Fairness: Will the results be equitable for all parties (patients, Fargo Health, public health agencies, Fargo Health employees, etc.) when your forecasting model is deployed?

Ownership: Who owns the dataset, analysis, and insights gleaned from data analysis? Is there a moral obligation for Fargo Health to act based on the forecasting model?

Accountability: Who is accountable for mistakes and unintended consequences in data collection and analysis? Can the affected parties check the results that affect them?

5, 7	Logistics and Organization, Scheduled Status Updates and Deliverables	These two sections contain information that assumes the case study will be analyzed in groups.	These sections are no longer valid. Each of you will perform data cleaning, create any two time series models (such as ARIMA and Holt's), compare the two selected models (instead of 5 or 6 models mentioned in the case study), create a 10-minute PowerPoint presentation, and answer all questions mentioned in assignment details (pages 4-5).						
8	Grading	The published case study contains a table indicating that the final grade is a weighted composite of four deliverables (presentation, report, received peer evaluation, submitted peer evaluation).	<p>Because you will complete the project individually, there is no peer evaluation. This is the revised grading scheme:</p> <table><tr><th>Grading Component</th><th>Weight</th></tr><tr><td>Presentation: submit a 10-minute narrated PowerPoint presentation</td><td>20%</td></tr><tr><td>Deliverable Report, R-script, and cleaned dataset</td><td>80%</td></tr></table>	Grading Component	Weight	Presentation: submit a 10-minute narrated PowerPoint presentation	20%	Deliverable Report, R-script, and cleaned dataset	80%
Grading Component	Weight								
Presentation: submit a 10-minute narrated PowerPoint presentation	20%								
Deliverable Report, R-script, and cleaned dataset	80%								

Suggestions for Success

The following are some suggestions for the presentation and report that can help you achieve a good score

The following are some suggestions for the presentation and report that can help you achieve a good score.

Your **presentation** should:

- provide a concise summary of the final project, including: the business problem, your data cleaning approach, forecasting models used, comparison of the forecasting models, your recommendations for Fargo Health, and lessons learned.
- be organized in a logical sequence.
- thoughtfully use graphics to explain and reinforce the content.
- be free of misspellings and grammatical errors.
- have clear voice narration to explain the content.

Your **report** should:

- be written in and saved as a Microsoft Word file.
- be formatted with 12-point Times New Roman font.
- be single spaced.
- not exceed 10 pages.
- address **all** questions as stated in the case study (pages 4-5) **and** address the ethics question described previously.
- include citations for references, if any.
- thoughtfully use graphics to explain and reinforce the content.
- be free of misspellings and grammatical errors.
- include your approach to data cleaning. Include necessary plots to show data quality issues and your approach to fixing them.
- include your comparison of the forecasting models used.
- include all R files. Your R programs should work, be free of syntax errors, and use comments to explain what the code is accomplishing.
- include the cleaned dataset, submitted as a separate Excel file.

Submitting Your Work

Create a zip file that includes your:

- 10-minute narrated PowerPoint presentation.
- Deliverable Report written in Microsoft Word.
- complete R script for the time series models.
- cleaned dataset in Microsoft Excel format.

Name this file:

FinalProject_DS700_YourLastName.zip
EXAMPLE Assignment2_DS700_Rossi.zip

Submit your work to the following dropbox:


Final Project Dropbox (/d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-10204350)

Frequently Asked Questions

Q What is the purpose of the final project?

A See the **Purpose & Overview** section above.

Q Where do I get the case study?

A Fargo Health Group: Managing the Demand for Medical Examinations Using Predictive Analytics (<https://hbr.org/product/fargo-health-group-managing-the-demand-for-medical-examinations-using-predictive-analytics/BAB266-PDF-ENG>)  (purchase required).

Q Where is the dataset mentioned in the case study?

A Click here to download the Dataset.xlsx file (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-9256615>). Note that this dataset has intentional data quality issues (missing data, incorrect data, and outliers). Using the details provided in the case study and the explanation of the dataset document (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-9602938>), you will need to clean the dataset before modeling.


Q Is there another dataset available for this case study?

A No. The provided dataset is the only available dataset for this case study.

Q How will I know if my dataset is properly cleaned?

A Ensure that you have accounted for the items mentioned in the explanation of the dataset document (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=content&rcode=UWS1-9602938>). You can examine missing values, invalid values, and outliers using visual analytics. You will need to judiciously impute values for missing and invalid data.

Q What resources will be helpful for learning data cleaning?

A Please review **Data Smart**, Chapter 9; **Practical Data Science with R**, Chapters 3-4; and the Lesson 4 content (Statistics and Machine Learning). The Lesson 5 content (Visual Analytics) will help you create plots helpful in identifying data issues. You may find the Amelia package ([amelia.pdf?_&d2lSessionVal=XNUUnsch3vixaFCJSDIyORxex&ou=3409957](#)) in R helpful for data imputation. The linked document comes from the Amelia program site (<https://cran.r-project.org/web/packages/Amelia/>) .

Q What time series models can I use for this case study?

A You can use any time series model discussed in this course. Whether or not you find any trend and/or seasonality in the data will partly determine your choice. You will create and compare two models. For example, you can create an ARIMA model and a Holt's model. Please review the Lesson 8 content (Forecasting) to learn about time series modeling.

Q How far out do I need to forecast?

A You can forecast the next 12 months using the cleaned dataset.

Q What items do I need to submit for the final project?

A You will submit a written report (maximum of 10 pages), a 10-minute narrated PowerPoint presentation, the cleaned dataset in Excel format, and the R scripts for time series models. The report must address the questions mentioned on pages 4 and 5 of the case study, in addition to the ethics question described in the previous **Changes to the Case Study** section. In the narrated PowerPoint presentation, you need to provide an overview of your work (data cleaning process, models used, model comparison, and any recommendations you may have for Fargo Health).

Q Is there any help available for the final project?

A In addition to the resources mentioned previously, you can post your project questions on the “Final Project” discussion thread of the "Raise Your Hand" discussion forum (</d2l/common/dialogs/quickLink/quickLink.d2l?ou=3409957&type=discuss&rcode=UWS1-9105635>). You can also contact the instructor via e-mail.

Q Will I need to schedule update meetings with the instructor, as mentioned in the case study instructions?

A No. Since this is an individual project, there will not be any status update meetings with the instructor.

 Download

Print

--	--

Activity Details

Task: View this topic

Last Visited Aug 7, 2016 1:46 AM