Data Analytics Final Project Overview

BU.510.650
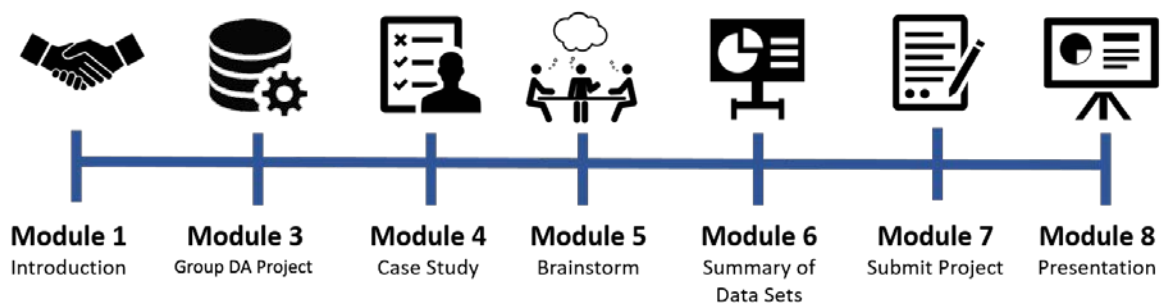**Data Analytics**
Fall 2018
Arnab Bisi

Project Overview

In this course, you will work on a group project that will allow you to identify a problem of interest, locate relevant data, develop a series of research hypotheses based on theory or past empirical evidence, and then apply some techniques to test your data.

In the last modules of this course, you will synthesize everything you've learned and apply this knowledge in an authentic way by presenting your findings to your instructors and peers.

If this sounds slightly overwhelming, don't worry. This project is broken down step-by-step. This Project Overview Document breaks out the assignments in a table that offers you a quick look at how the project progresses throughout the course. And the following Project Assignment Overviews detail what you will be doing, what you will be delivering, and when.

Project Assignments

| Module 1 | Module 3 | Module 4 | Module 5 | Module 6 | Module 7 | Module 8 |
| Introduction | Group DA Project | Case Study | Brainstorm | Summary of Data Sets | Submit Project | Presentation |

Module	Assignment	Description	Deliverable
1	Introduce Yourself	Meet your group and complete the group introduction activity.	Introduction Video and Discussion.
3	Group Assignment	Complete the Group Data Analytics Assignment.	Problem Set Group Discussion and Analysis

4	Analyze a Real-World Case	Read the Vanderbilt case, then discuss and analyze it.	Problem Set Group Discussion and Analysis
5	Brainstorm Ideas and Hypotheses	Meet with your group and generate ideas for your project.	Brainstorming Session Summary
6	Describe Your Data Set	Submit a one-page project summary briefly describing your project and submit the data sets that your group will analyze.	Data Set Description
6	Analyze a Real-World Case	Read the Pilgrim Bank Case and then discuss and analyze it.	Problem Set Group Discussion and Analysis
7	Submit Your Report	Submit.	Analysis and Report
8	Present Your Findings	Use Zoom to record a group presentation of your project, hypothesis, results and supporting data. Post your project and discuss projects among peers.	Presentation Video and Discussion

Data Sets

You will be working together in a team for the duration of your course project. You are encouraged to collect data by yourself and with your team to analyze problems and develop hypotheses. You can collect data from any public sources such as, but not limited to:

(1) UCI Machine Learning Repositor: http://archive.ics.uci.edu/ml/datasets.html

(2) World Bank: http://data.worldbank.org/

(3) U.S. open government data: http://www.data.gov/

(4) U.S. Census Bureau: http://www.census.gov/main/www/access.html

(5) Other sources*

* Be sure that if you are getting data from other sources, you have permission to use that information.

Project Assignment Pages

As with all projects, proper planning, complete engagement, and attention to detail are critical. Each of your project assignment pages will contain very specific details about how, when, and why you should complete each step of the project. On those pages you will find detailed instructions and tools that will help you succeed each step of the way.

Module 1 – M1 Group Assignment - Introduction

Welcome to the first project assignment! Your professor has placed you in a group, and this week you'll introduce yourself to the members of your group, and complete an activity together. This assignment has you

work with your group to create a short presentation that you will submit on **Friday of Module 1** so make sure you contact your group members right away. You will load your presentation to YouTube and one group representative will post the presentation to a discussion board topic. You will then individually watch two presentations and respond to your peers by **Sunday of Module 1**.

The purpose of this activity is to introduce you to the technology for this class (Zoom and YouTube). Your group will become your very own R developer community that will also integrate with your class developer community through the discussion boards. Building and engaging in development communities online and with your peers is essential to becoming proficient in the R language, especially if you are new to programming. Throughout this course, you will be working on several group projects. Your final course project will require a group presentation using Zoom and YouTube and will be evaluated by your peers.

The full project parameters, presentation guidelines, and technical support links are located in the project documents for each activity located in the Assignment pages throughout the course. After reviewing this overview, proceed to the modules and follow the instructions there for the details.

Module 2 – No Course Project Assignment This Week

Module 3 – M3 Group Assignment 1 – Data Analysis

In Module 3 you will complete your first data analysis project. This is a group assignment. You will be given a problem set and some data files and will have to work together to analyze the data and solve the problems. There are a total of 10 problems, worth 5 points each.

You will then post your responses to the questions and discuss your analysis and conclusions in M3 Discussion – Assignment 1 due at the end of the module. You have to post your responses before you can see peer posts.

There are two files to submit for this project: the first is the Problem Set with your answers typed into it. The second is the R script file. The script file will only be used if your professor or TA needs more information about your answers or for verification. Both files should be submitted in the Assignment area.

Module 4 – M4 Group Assignment 2 – Vanderbilt Medical Center

This week you meet again with your group over Zoom to perform a case analysis of Vanderbilt Medical Center. You will work together to analyze the case and utilize data files to complete a problem set. You will then participate in a discussion about your analysis and conclusions.

It's important for everyone in the group to participate equally, and the professor will be monitoring the collaboration to ensure that all members are fully engaged. To ensure that this is in fact the case, it is recommended that you meet as a group and discuss the problem set, then divide the problems among the members. Each group member completes their problems and then you reconvene and discuss the results and findings and collaborate on your final submission. If you find this is not ideal for your group, feel free to determine a process amenable to all group members.

The guidelines and rubrics are in the assignment in Module 4.

Module 5 – Final Project Benchmark 1 – Brainstorm Ideas

This is the first activity as part of the Final Group Project for the course. You will work with your group over Zoom to complete a project that encompasses all course learning objectives, focusing on a real-world scenario.

The project requires each group to identify a problem of interest and locate data that is relevant to the problem. The collected data set should contain at least 200 observations and five variables. The task is to develop a series of research plans and hypotheses, based on theory or past empirical evidence, and then apply some of the techniques covered in class (or not covered) to such data for testing.

Read the project scope and meet with your group members to determine your problem of interest and brainstorm some questions or hypotheses you want to explore with data sets. In Module 6 you will submit a one-page summary and description of the data set. Spend some time researching your potential ideas, hypotheses, and problem of interest to ensure that you can find and analyze sufficient data to thoroughly explore your ideas.

The final project scope for the following activities can be found in the Final Project Scope document located with each benchmark.

Module 6 – Final Project Benchmark 2 - Describe Your Data Set

Last week, you met with your group to choose a problem and explore data sets. This week you should meet to finalize your project plan and work together to develop and submit a one-page summary of the data set and the final problem.

Module 6 – M6 Group Assignment 3 – Pilgrim Bank

During this week, you will also work with your group on a second case analysis, Pilgrim Bank. You will work together to analyze the case and utilize data files to complete a problem set.

You will then participate in a discussion about your analysis and conclusions.

This assignment has the same submission parameters as the Vanderbilt case. You will submit your problem set answer document as well as your output file.

Module 7 – Final Project Report Submission

You've worked closely with your group on data collection, data analysis, and the resulting interpretation. In this week you will submit your final project report, describing the results and conclusion of your analysis. Keep in mind that plots, tables and other visual representations of data are useful in conveying your conclusions. In addition, you may want to include the following parts in your report:

(1) **Questions/Hypotheses**. Write at least one question or hypothesis, more if you wish, that you want to explore with the data sets. After each question, state your expected answers, which may prove different from your conclusions because they are generated prior to your analysis.

(2) **Data Description**. Describe the data sets. What is the data? How was the data collected? What are the variables and results? Briefly summarize the data. Provide the URL link if available.

(3) **Methodologies**. Write a complete, clear description of the analysis you performed. This should be sufficient for someone else to write an R program to reproduce your results. It should also be helpful to people who read your code later. This section should tie your computations to your questions/hypotheses, indicating exactly what results would lead you to what conclusion. You may want to provide the key statistics (e.g., t-statistic, z-statistic, p-values, R2 and the adjusted R2).

(4) **Results and Conclusion**. Discuss your results. Focus in particular on the results that are most interesting, surprising, or important. Discuss the consequences or implications. Interpret the results. If the answers are unexpected, then see whether you can find an explanation for them, such as an external factor that your analysis did not account for.

Include a title for your report and all group members' names.

Each group submits one report. Your submission should include a project report, data set (the URL link is OK if your data set is too large) and a slide deck. The project report should be 4 – 6 pages long (you can use additional pages for figures, graphs, etc.) with 12pt font size and 1.5 or double spacing.

You will submit your project to the Assignment including everything (data set, R scripts in a separate file, report and presentation slides). If your data set is too big or sensitive, a sample subset or URL link will be fine.

Week 8 – Final Project Presentation

Each group should prepare a presentation of their project, the results, and findings. Each presentation should be around 12 minutes. You are encouraged to use slides (e.g., MS PowerPoint, LaTex). The slide deck should summarize the main points of your report, including motivation, research questions, and results.

Your group will submit their presentation to a discussion board, and other students will watch your presentation, ask questions and provide feedback. Asking questions of the other groups and analyzing their results is part of your participation in the project and is included as part of the project grade.

Zoom and YouTube Help Documents
- **Using Zoom to Record Group Presentations**
- **Post a YouTube Video in a Blackboard Discussion**