

Review: Final Team Project Introduction



Introduction:


The final project for this course will assess the application of key programming techniques and analytic concepts introduced in the modules. You and your team will import and transform a raw dataset and perform exploratory and descriptive analysis with appropriate summary statistics and visualizations. Then you will determine and apply foundational analytic models to the provided data using *either* R *or* Python. It is an *open-ended* guided project. Enjoy a hands-on final project!

Project Timeline:

- Module 2 (by the end of Week 2): The course instructor will group students into teams of two to three members. The project datasets will be made available for all teams. Canvas, USD Email, or Slack can be used to find prospective team members.
- Module 4 (by the end of Week 4): Each team should select one of the three project datasets and the programming language of their choice by the end of this module. The team representative will need to submit the "Team Project Status Update Form."
- Module 7 (by the end of Week 7): Each team should submit two deliverables for the course project in the final week:
 - One PDF document containing code, explanations, and interpretations.
 - One video recording of the project presentation by all team members. Slides are optional.
- **It is critical to note that no extensions will be given for any of the final projects due dates for any reason, and final projects submitted after the final due date will not be graded.

Project Datasets:

- Three datasets are provided for the final project in Canvas. Each dataset comes with a 'readme' text file that contains information about the origin, size, file type, missing values, column names, and descriptions of the dataset.
 - **Dataset 1** (<https://sandiego.instructure.com/courses/838/files/529235?wrap=1>)  (https://sandiego.instructure.com/courses/838/files/529235/download?download_frd=1) contains data related to the marketing campaigns of a European bank.
 - **Dataset 2** (<https://sandiego.instructure.com/courses/838/files/529197?wrap=1>)  (https://sandiego.instructure.com/courses/838/files/529197/download?download_frd=1) has data related to house sales for a county in Washington state.

- **Dataset 3** (<https://sandiego.instructure.com/courses/838/files/529233?wrap=1>) 
(https://sandiego.instructure.com/courses/838/files/529233/download?download_frd=1)
consists of annual user session data for an e-commerce company.

Requirements:

There are four sections of the final project. You and your team are expected to perform the following tasks within each section to fulfill the project requirements.

1. Data Importing and Pre-processing

- ~~Import dataset and describe characteristics such as dimensions, data types, file types, and import methods used~~
- ~~Clean, wrangle, and handle missing data~~
- ~~Transform data appropriately using techniques such as aggregation, normalization, and feature construction~~
- ~~Reduce redundant data and perform need-based discretization~~

2. Data Analysis and Visualization

- ~~Identify categorical, ordinal, and numerical variables within the data~~
- Provide measures of centrality and distribution with visualizations
- Diagnose for correlations between variables and determine independent and dependent variables
- Perform exploratory analysis in combination with visualization techniques to discover patterns and features of interest

3. Data Analytics

- Determine the need for a supervised or unsupervised learning method and identify dependent and independent variables
- Train, test, and provide accuracy and evaluation metrics for model results


4. Presentation

- In a **5 to 10** minutes video, briefly explain the project workflow from the code and results in your markdown notebook
- State your findings from the data and provide the interpretation of results from your analysis at each stage in the project

Deliverables and Submission Format:

- Implement the project in either Python or R language. Use Jupyter notebook for Python and R Markdown for R language. Within the Python or R notebook, import all packages used for the project in the first cell, use code cells for code and comments, and use markdown cells for

headings and descriptions. Generate a **PDF** document from your notebook for submission with code, comments, and explanations for analysis and results within the notebook. Divide the work equally between the team members for the first three sections; everyone needs to code!

- Prepare a recorded video presentation of your project (slides are optional but recommended) using a screencasting tool, such as Zoom, to record your screen and provide a voice narration. Ensure that the sound quality of your video is good and each member presents an equal portion of the presentation. Export the video file to an **mp4** format.
 - You may use any recording software you wish. View the [Recording Video Presentation and Submission Guidelines for MS-ADS Students](https://sandiego.instructure.com/courses/838/files/529590?wrap=1) (<https://sandiego.instructure.com/courses/838/files/529590?wrap=1>)  (https://sandiego.instructure.com/courses/838/files/529590/download?download_frd=1) guide for additional recording instructions.
- Submit the final PDF document and video file on the final project submission page of Canvas. You will use the naming convention **Final Project-Team Number.pdf** (e.g., Final Project-Team 1.pdf). **Only one member of your team will need to submit these deliverables.**
- Submit the peer evaluation form individually.

NOTE: Team members may not get the same grade on the Final Team Project, depending on each team member's level of contribution.

To understand how your work will be assessed, view the assignment rubric on the [Final Team Project](https://sandiego.instructure.com/courses/838/assignments/73995) (<https://sandiego.instructure.com/courses/838/assignments/73995>) page.