CHSC 7400 T20
Advanced Topics in Community Health I:
Seminar in Foundations of Disease Analytics

Final Project Requirements
Value: 50% of Final Grade

Relevant Dates
March 3, 2023: Research plan due (submit on UM Learn)
March 31, 2023: Oral report due (in class; submit slides on UM Learn)
April 14, 2023: Written report due (submit on UM Learn)

Study Data
You have been provided with multiple .csv files that contain information about a synthetic sample of >10,000 patients. These .csv files are:

- conditions
- patients
- observations
- careplans
- encounters
- devices
- supplies
- procedures
- medications

Further details about this dataset are provided on UM Learn

Overview
For your final project, you will:

- Construct and validate one predictive model for your data
- Construct and validate a second model for your data, using one of the modeling approaches described below
- Adopt different approaches to visualize your data and modeling results, and critique these visualizations

The predictive model can be constructed using logistic regression, a classification and regression tree, a random forest model, or another suitable choice for predicting a measure in the study dataset.

The second model/modeling approach will be <u>one</u> of the following:

a. **A clustering method:** for grouping individuals (or other units of analysis) so that those in the same group are more similar to each other (i.e., more homogeneous) than to those in other groups. Examples: latent class analysis; k-means clustering
b. **A dimension reduction method:** for keeping only the most relevant variables from a set of data or exploiting the redundancy of the data by finding a smaller set of new variables that may be a combination of the original variables. Examples: principal components analysis; propensity score analysis; discriminant analysis

There are a number of clustering and dimension reduction methods that can be applied to your data. If you have questions, consult with Dr. Lix.

Components of Your Final Project

Component #1: Research Plan

Your research plan will have the following elements:

- **Research Questions and Hypotheses:**
  - Define the research questions and testable hypotheses for your predictive model and second model. In other words, you should have (as a minimum) two research questions and two hypotheses.
- **Flow Diagram of the Proposed Analytic Process:**
  - For each research question, you will prepare a flow diagram that will describe the steps you will take to test the hypothesis and answer your research question. Your flow diagram should identify the key steps in preparing your data for analysis, conducting the analysis, validating your model, and visualizing/reporting your results.
- **Written Summary of the Proposed Analytic Process:**
  - This should support the flow diagram and justify the choice of analytic methods and their appropriateness for your data.
- **Description of Variables:**
  - Identify the variables that you will use for each of your models. Provide descriptive statistics for these variables. For example, you could report frequencies and percentages for categorical variables. Use at least one figure when presenting your descriptive statistics. You may also use tables, as needed, to report the descriptive statistics.
- **Potential Challenges and Mitigation Strategies:**
  - Provide a short description (i.e., about 250 – 350 words) that describes any potential challenges that you anticipate when applying your analytic approaches. For example, are your data highly imbalanced? How might this affect your analysis and what is at least one strategy that you will use to address this potential challenge? As another example, does your dataset have a lot of missing data on a key explanatory variable? What approaches could you will use to address this potential challenge and what are the strengths and limitations of these approaches? Which approach will you choose and why?
- **Format:**
  - Title page with project title (be creative), name, and affiliation
  - Double space the research plan
  - Maximum of 8 pages, **excluding** title page and references and appendices (not required, but may be useful for providing supplementary information). Any figures and tables should be embedded within the report and are included in the 8 page limit
  - References: for justifying proposed analytic methods and potential challenges and mitigation strategies

Component #2: Oral Report

Your oral report will:

- Provide a brief overview of the research topic and its importance
- Provide information about the characteristics or features you selected for your analyses
- List your research questions and hypotheses
- Describe the steps in the models/analyses that you applied to your data, as per your research plan, and note any deviations from the research plan
- Describe and justify the steps in your analyses, considering:
    - the attributes of your data
    - the assumptions that underlie your models
- Summarize modeling results and model validations in both tabular and visual formats (selected results only, keeping in mind the time available for your presentation and the maximum number of slides; you do not need to present all results in the oral presentation)
- Provide one comparison of two different visualizations of the same data (see the written presentation requirements for more information on comparative visualizations) and critique the comparison
- Provide a critique of your analysis, addressing such topics as: What parts of the analysis did not go as expected? What parts of the analysis did go as expected? Were there any analytic challenges that arose because of the attributes of your data? What recommendations would you make to classmates using similar data and/or similar analytic methods?
- Use a maximum of 16 slides (including the title slide). The title slide should include the project title, and your name and affiliation.

You will have 12 minutes to present. You will have five minutes to receive feedback and address questions from the audience. Your presentation will be introduced and timed by one classmate and questions will be moderated by the classmate.

Component #3: Written Report

Your written report is the capstone deliverable for your final project. It will:

- Provide an overview of the research topic and its importance (use references to justify research importance)
- List your research questions and hypotheses
- Summarize the attributes of your data (i.e., descriptive statistics)
- Describe and justify the steps in your analyses (with references), considering:
    - the characteristics of your data
    - the assumptions that underlie your models
- Provide two sets of comparative visualizations that meet the following requirements:
    - **Reconstruction of a Descriptive Visualization:** Use two different visualizations to present descriptive information about your data. For example, if you select a bar chart to describe your data, you might use a violin plot as an alternative visualization. Provide a critique of the visualizations. For example, which one do you prefer and why?
    - **Visualize Model Output:** Using one of the models that you fit to your data, produce two different visualizations that depicts the same output. For example, you might show your model predicted values using two different visualizations. As another example, you might show your model residuals

(i.e., difference between observed and predicted values) using two different visualizations. Provide a critique of the visualizations. For example, which one do you prefer and why?
    - You may use other visualizations of your data, but strive to keep the number manageable
- Contain a minimum of 3 tables:
    - Table 1: descriptive characteristics of your dataset
    - Tables 2 & 3: results of the models that you fit to your data
    - You may use more than three tables to present your data and analyses, but strive to keep the number manageable
- Briefly summarize your analytic results (models and validations) that are presented in tables and visualizations
- Provide an interpretation of the results
- Be a maximum of 17 pages in length
    - This number excludes the title page and reference list
    - This number includes figures and tables, which should be embedded directly within the body of your report
    - The title page should contain the project title, and your name and affiliation
    - Number all pages
    - Double space your report
    - Include an appendix with (note there is no page limit on your appendix):
        - Annotated analytic code (e.g., from SAS, R, SPSS, Python) that shows the key steps in your analysis

Other Notes

- You may use the software of your choosing to conduct the analyses and produce the visualizations

Evaluative Criteria

Your research plan, oral report, and written report will be evaluated by your instructor/teaching assistants. The oral report will also be evaluated by your classmates. The evaluative criteria that will be used for the final project include:

- clarity, logic, and appropriateness of the models
- justification for analytic approaches
- accurate and complete description of the data, the models/analytic approaches, and the results
- accurate and organized reporting of analytic results
- creativity of visualizations
- quality of the critical assessment of your analytic approach and visualizations
- creativity, correctness, validity, and completeness of interpretation
- error-free and well organized reports
- slides support oral presentation
- dynamic and engaging oral presentation
- active participation in discussion/asking questions during oral presentations
- leadership during oral presentations