Assignment 1: Can Data Help a Friend?

BUAD 6951 – Spring 2021

---

1. **Objective:**
   The purpose of this assignment is to build regression models using R and to visualize the data using Tableau. The assignment also allows you to work on your teamwork and presentation skills.

2. **What You Will Need**
   - Access to a computer with R and R studio
   - Access to a computer with Tableau
   - Screen capturing software (https://zoom.us/, https://screencast-o-matic.com/,https://support.apple.com/downloads/quicktime)

3. **What You Will Hand In**
   - A Tableau file with your visualizations. You should name this file *Assignment1_Group#.twb*, where *Group#* is the number of your group.
   - An R file with your code. You should name this file *Assignment1_Group#.R*, where *Group#* is the number of your group.
   - Each person must submit the PresentaionRubric.xlsx, that rates the other teams.
   - A 15-min video presenting your recommendation, model, results and limitations. This video should be uploaded to the Panopto folder in Blackboard.

4. **Due Date**
   Presentations: 03/18/2021 at 6:00PM EST.
   Peer Evaluation: 03/19/2021 at 11:59PM EST.

5. **Note on Collaboration**
   This is a *Category B* assignment. Specifically, your group may not receive help from anyone outside your group. All questions concerning this assignment should be addressed to your professor. It is an honor code offense to give help to other groups and individuals or receive assistance from other groups and individuals

Assignment Instructions

Teams:
Join/form a group of up to four students (minimum three students). Please add your team members in the google spreadsheet.
https://docs.google.com/spreadsheets/d/1NE_0JwUUlay-_x74-HC-uHsX65ru84hemlUjCVg2wGQ/edit?usp=sharing

The Data:

Speed Dating
Data file Dating.csv. The data was gathered from college participants in experimental speed dating events from 2002-2004. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end of their four minutes, participants were asked if they would like to see their date again. They were also asked to rate their date on six attributes: Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests.

The dataset also includes questionnaire data gathered from participants at different points in the process. These fields include: demographics, dating habits, self-perception across key attributes, beliefs on what others find valuable in a mate, and lifestyle information. *Source: Kaggle.*

The "dec" variable is the response variable, and all other variables may be used to predict "dec". Do not use the "Match" variable as a predictor.

Objectives:

- Using the Dating.csv data set, build a regression model for a single wave, and provide a recommendation following the results of your model and descriptive statistics.
- Using the Dating.csv data set, create a visualization in Tableau that provides insight to the results found in your regression analysis.
- Provide a recommendation of how your "friend" should approach speed dating to successfully find a match. This recommendation should be based on the data and any analysis that you have performed.

Presentation:

The presentation must include:

- Motivation/problem that the project addresses.
- Methods used to tackle the problem.
- Data overview and descriptive statistics.
- Recommendation and limitations.

As your team works in the project try to answer the following questions: What is the problem/motivation? Why is it relevant? How can we resolve the problem (methods, plan of attack)? How can the data help us? What are the limitations of the data, methods and analysis that we use? What should be our final recommendation?

Workflow:

- Get familiar with the data. For this, you might want to explore the data dictionary (Dating.doc file).
- Make sure the variables are coded correctly. If you are dealing with a categorical variable make sure it is a factor, if you are dealing with a magnitude make sure you it is numeric.
- Take some time to visualize the data. Graph some relationships between your response variable and your independent variables.
- Think about what makes a person decide favorably on selecting a date and build a model that explains the decisions on average.
- Check measures of fit AIC, AICc and BIC to choose the best model.
- Make note of the variables that are statistically significant in your model.
- Run some descriptive statistics and graphs. Does your model align with the findings in these? See how you can reconcile your findings, sometimes you realize you have made mistakes.
- Once your descriptive statistics and model match, create a visualization in tableau that highlights your recommendation, insight, or overall story. This visualization should not contradict your model and it should be related to your findings. The visualization should help you make your recommendation.