

Assignment 2

Examine Prize-Winning Reports

Learn, firsthand, how top teams tackle analytics problems and extract strategies you can reuse in your case competition. Choose one prize-winning report from the list provided (but make sure no one else in your group uses the same report): Focus on the approach, not the application domain, so you can generalize what you learn to your own case.

- Analyze the Data preparation sections of this report (look for sections titled data preprocessing, data cleaning, data transformation, or feature engineering)
 - a. What specific steps did they take to create, remove, or transform variables?
 - b. Why did they use each step? (Provide the rationale.)
 - c. Which of these steps will you try on your dataset, and why?

Case Competition: Data Pre-processing for Predictive Modeling

Download the data and data description in the Case Competition folder on Blackboard.

1. Each person must contribute to the data cleaning and pre-preprocessing. Provide what role you have and which specific variables will you work with in this process?
2. Please document any changes made for each variable and describe how you will handle each categorical variable.
3. Below are some aspects to consider but the final decisions are up to you.
 - Will you create dummy variables, one hot encoding, will you let the tree-based methods determine where to split the categories, or use another categorical encoder?
 - Will you bin categories together or not?
 - Are there any duplicate rows?
 - How will you handle the date variable?
 - How will you handle NULL or missing values?
 - etc.
4. Each person must submit code for their portion and answers to the above questions.
5. **As a group, implement these steps to create a final dataset, and combine with your group to prepare a final dataset to be checked in class next Tuesday.**