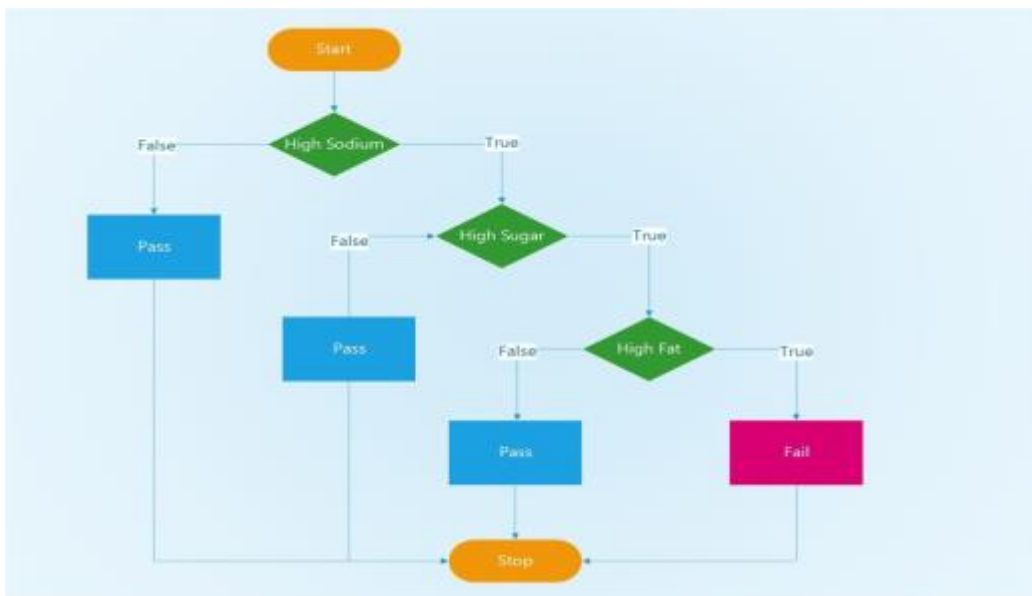


CMTH 642: Advance Methods

Assignment 2

Read the clean csv file you provided in the first assignment (USDAclean_ [your last name]). When you finalize your answers, please save your Rmd file as “R_ [your last name]”, save your final data frame as “USDAclean_ [your last name]”, and send both files to the assignments folder in D2L.

1. Create a visualization to illustrate the distribution of values for Total Fat, Protein and Carbohydrate. (12 p)
2. Create a visualization to illustrate the relationship between a food's Total Fat content and its calorie content. (12 p)
3. Create a logistic regression model, using High Calories as the dependent variable, and Carbohydrate, Protein, Total Fat and Sodium as independent variables. (18 p)
4. Which independent variable is not significant? (10 p)
5. Which independent variable has the strongest positive predictive power in the model? (10 p)
6. Create a script for a HealthCheck function to detect unhealthy foods. Foods that are high in salt, sugar and fat fail the HealthCheck, while all other foods pass. Foods that pass the HealthCheck should be assigned a 1, while foods that fail should be assigned a 0. Use the algorithm flowchart below as a basis for this script. (18 p)



7. Add a new column called HealthCheck to the USDAclean data frame using the output of the function. (10 p)
8. How many foods in the USDAclean data frame fail the HealthCheck? (10 p)