**Instructions For 1-5**

**For each of the first five questions, you only need to provide the answer.**

1.Name 3 ways of how you can measure the accuracy of a classification model and the formula of how to do so.

2. Which 3 algorithms did we cover in class that dummy variables would **not** be required for?  
3.True/False, KNN is only for classification problems?

4.When interpreting the output of a multiple linear regression model, which of the following metrics should look at and why, (Coefficients, R-Squared, P-Values)  
5.What is the key advantage/use case for using Support Vectors over Logistic Regression for Binary Classification? (Should be 2-3 sentence explanation)

**Instructions For 6-10**

Use The Dataset In This Module

**Clean missing values first but no exploratory info needed. You need to predict whether the candy is chocolate or not.**

6. Clean the dataset and discuss how you cleaned the variables with missing values.  
7. Find the significant variables using multiple logistic regression. If none are significant, feel free to expand it to .1. Of those that are significant, discuss their coefficients and which variable do you think would be most impactful from a business sense (can someone impact this or not if they are making a candy bar from scratch).*This should be minimum one paragraph with description of the packages uses, technique used to be find the significant variables and a discussion on whether we can give advice/action.*

8. Run decision tree, random forest and gradient boosting (1000 iterations) and address the accuracy of each of the three models *This should be minimum two paragraphs with description of the packages uses, technique used to be find the accuracy measure, what accuracy measure you used (and why) and a discussion on which model you think is best representative of the model.*

9. Build a matrix with all the models that includes at least two measures of accuracy and speed for the models in 7 & 8. Compare the models and give us a reason on which model you think describes accuracy the best.

10. In two paragraphs, based on looking at the variable and model results, give feedback to a candymaker on what you’ve learned from the model. Do you feel one model does a better job than all the others? Does one model gives us a better understanding of what to put in the candy bar? Do we know an ingredient or two that makes the best chocolate? In your two paragraphs, use these questions to frame a speech that you would give if you were presenting to the board of this chocolate company about your findings.

**When you are done, submit a word document and a separate file with your code. You have until 10/28 at Midnight to submit and no later.**