





[New Attempt](#)

- Due Apr 24, 2022 by 11:59pm
- Points 100
- Submitting a file upload

Module 2 Technique Practice



-  **Syllabus** (<https://northeastern.instructure.com/courses/105794/assignments/syllabus>)
-  **Modules** (<https://northeastern.instructure.com/courses/105794/modules>)
-  **Resources** (<https://northeastern.instructure.com/courses/105794/pages/course-resource-page>)
-  **Assignments** (<https://northeastern.instructure.com/courses/105794/assignments>)



Assignments

Technique Practice

Assignment **due** by **Sunday**, 11:59 pm EST.

Overview and Rationale

Data mining is used to reveal hard to see and hidden patterns and relationships in Big Data datasets. Data mining helps to classify data for further examination or create models to predict outcomes for a different set of data. As data miners, you should be able to explain how the code used to mine the data is functioning and be able to analyze and interpret the results of the mining. This allows you to summarize and clarify the results for stakeholders.

Assignment Description

You will looking at a dataset about a company that is launching a new product called, IceCubed. This will be a household ice cream maker similar to that of a Keurig (for coffee). They did a fundraiser to get funding for their project and they learned the following information about their donors. Now that donors have had their time to actually purchase the product, IceCubed wants to know if there is statistical significant differences between donors who purchased and those who did not.



https://northeastern.instructure.com/courses/97871/files/13613560/download?download_frd=1

[Data \(https://northeastern.instructure.com/courses/105794/files/14809385?wrap=1\)](https://northeastern.instructure.com/courses/105794/files/14809385?wrap=1) ↓
https://northeastern.instructure.com/courses/105794/files/14809385/download?download_frd=1

https://northeastern.instructure.com/courses/83414/files/10226782/download?download_frd=1

https://northeastern.instructure.com/courses/68030/files/9834810/download?download_frd=1 -> Who Actually Purchased

You will use logistic regression to build a model that allows us to see which attributes drive purchase and if it is useful.

In this assignment, you will document in a report the results of each step of the mining process, analyze and interpret the results. Suggest the characteristics to use when determining interests of someone that would purchase this product. Make recommendations for additional analysis and variables to examine to build other classifications for customer personas.

Instructions

Your report should include the following:

- Problem Background: Address the problem at hand and the value of solving (such as what can a company learn from understanding their customers).. Check the cleanliness of the data and if needed, clean the data up and provide details from doing EDA.
- Analysis: Based on the model outputs, analyze the data and the relationships revealed about the variables of interest. You should be building a logistic model. Explains the insights provided by the output. Main focus should be on which variables are significant and accuracy of the model. Talk to what the model predicts well/not so well to understand how trustworthy it is. Being able to tie those findings to the problem, will be key to success. Also use visualizations to support your analysis.
- Interpretation and Recommendations: Interpret the results of your analysis and explain what the results mean for the data owner (limit to 2 at the max). Provide recommendations for actions to be taken based on your interpretation. Support those with the data (this should be based on previous discussion in the paper, the conclusion should not have new information).. Explain why and what explicit variables you suggest incorporating.

Your report must be a 5-7 pages Word document. Include an appendix of visualizations and code segments as needed. The report format should follow APA formatting (12pt, Times New Roman font).



Technique Practice



Criteria	Ratings			Pts
Code	35 to >24.5 pts Meets Standard Clearly explains each step of the output and what the code is doing in relation to the data. Goes beyond the comments in the R script file and explains it based on the dataset that is being mined.	24.5 to >21.0 pts Approaching Standard Explains the general steps of the output and what is happening with relation to the data. Explanations reflect the comments in the R script file, but do reference the dataset being mined.	21 to >0 pts Below Standard Provides generic explanation of the output. Does not reference provide explanations with regard to the data set being mined.	35 pts
Analysis	30 to >21.0 pts Meets Standard Provides well-developed insights into results based on output of the R script and supported by clear and compelling visualizations.	21 to >18.0 pts Approaching Standard Provides useful insights into results based on output of the R script and supported by relevant visualizations.	18 to >0 pts Below Standard Provides generic insights into the results that may not be tied directly to the R script output. Provides few, if any, general or trivial visualizations.	30 pts
Interpretations and Recommendations	25 to >17.5 pts Meets Standard Provides a valid interpretation based on the analysis, followed by coherent and detailed recommendations. Suggests specific next steps for further data mining.	17.5 to >15.0 pts Approaching Standard Provides a clear interpretation based on the analysis, followed by rational recommendations. Suggests broad next steps for further data mining	15 to >0 pts Below Standard Provides an interpretation generically tied to the analysis. Recommendations are not consistent with the interpretation or analysis. No suggestions for next steps for further data mining.	25 pts



Criteria	Ratings			Pts
Word/Report: Writing Mechanics, Title Page, & References	10 to >7.0 pts Meets Standard There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references.	7 to >6.0 pts Approaching Standard There are very few errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references.	6 to >0 pts Below Standard There are more than five errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are incomplete.	10 pts
Total Points: 100				

