MKTG 6620: Project 1
Submission and Presentation Due Date: June 13, 2017

In order to complete the project successfully, please keep the following in mind:

1. This is a group project
2. The project description has two parts i) the details of the situation and ii) the format in which the written report needs to be submitted.

Business situation

This is your first day as data scientist at a grocery store chain. You look at your schedule and find that you have a meeting with both the brand manager and sales manager. At the meeting they tell you that they want to know how to make the Orange Juice category perform better than what it does currently. You learn that the grocery store chain sells two brands of orange juice Citrus Hill (CH) and Minute Maid (MM). CH gets higher margins than MM[1].

Listening to their description of the situation and their requirements it seems like the Brand Manager is interested in finding out what variables influence a person's probability of buying MM. In other words, what can he do to increase customers' probability of buying MM? On the other hand, the sales manager is interested in having a predictive model where he can simply predict the probability of a customer purchasing MM.

They look to you for providing a solution to their specific problems. You tell them that although both relate to MM the nature of questions they need answered makes the analysis process very different. They tell you to use different analyses if needed but inform them what you used to answer their question, why this method is better, and importantly what are your specific recommendations. They ask you to present a written report in which you clearly explain the method and your recommendations in a user-friendly manner.

You are given a dataset that contains 1070 purchases where the customer either purchased Citrus Hill (CH) or Minute Maid (MM) Orange Juice [see appendix for details of this dataset].

Considering the different goals of the brand and sales manager, you first need to figure out whether using the same method can help you answer both their questions or would you have to use different methods. Although you would prefer to use the same method, you realize that answering their specific queries as completely as possible is more important. To understand the problem better you write down the questions asked by each manager.

[1] Higher margin means they make more money on per unit sale of CH than MM

Brand manager

1. What predictor variables influence the sales of MM?
2. Are all the variables in the dataset effective or are some more effective than others?
3. How confident are you in your recommendations?
4. Based on your analysis what are specific recommendations you have for the brand manager?

Sales manager

1. Can you provide him a predictive model that can tell him the probability of customers buying MM?
2. How good is the model in its predictions?
3. How confident are you in your recommendations?

Instructions to complete the written report

Page limit:
Font: Times New Roman Font size: 12

The written submission should consist of the following sections

Cover Page: Group members names and a title for your report.

Define the problem

In this section describe what problems are being faced by sales manager and the brand manager. Importantly what is your objective? What do they expect from you(this should not exceed 1 page).

Define the method

Which method(s) of analysis are you planning on using and why? Defend your choice by explaining how the method(s) will help you answer the specific questions that have been asked. Would one method suffice or would you need more than one method[2].

- Did you scale/standardize variables?
- What efforts did you make to reduce overfitting (i.e., train test split, validation etc.)?
- Describe the data and variables that you used in your analyses.
- What happens when you include all variables?
- Is it better to not include some variables? Why? Provide methodological rationale for your responses and criteria you used to include or exclude variables.

[2] Hint: Use methods discussed in class so far.

- Explain in detail the analyses you conducted and any assumptions you made.

Include the R code as an appendix so that your analyses can be replicated.

Results and Conclusion

In this section you would be using the output from the analyses to generate your recommendation. Explain in detail what your analyses suggest. The output from the method(s) that you used would help you in your responses. Include the output of your analyses and along with explaining how each of the parameter values or test informed you. It is important to explain the R output in an understandable manner.

Provide your recommendation to each of the managers. If you are unable to answer any specific question of theirs, be upfront about it and explain clearly why. Don't blame the data. If you want to include results graphically feel free to do so. However, remember to explain the graphs clearly.

Important points to keep in mind for the report

1. Please clearly mark each section and subsection. Merging everything into long, unmarked paragraphs is unhelpful.
2. Ensure that you are providing proper reference and including the reference in a reference section.
3. Table and graphs should be clearly numbered with titles and should be self-explanatory. If you are adding them at the end of the report please reference them appropriately.
4. Please edit the analyses output you include in the appendix so that it is understandable to any reader. Merely pasting it would not suffice.
5. Please use language to communicate clearly and succinctly. Check for spelling and grammatical mistakes. A methodologically sound report riddled with language mistakes suffers overall.
6. Revise, edit and read the report carefully before submitting.

APPENDIX

Use the following code to download data in R

```
OJ<-read.csv(url("http://data.mishra.us/files/OJ.csv"))
```

OJ is a data frame with 1070 observations on the following 18 variables.

Purchase

A factor with levels CH and MM indicating whether the customer purchased Citrus Hill or Minute Maid Orange Juice

WeekofPurchase

Week of purchase

StoreID

Store ID

PriceCH

Price charged for CH

PriceMM

Price charged for MM

DiscCH

Discount offered for CH

DiscMM

Discount offered for MM

SpecialCH

Indicator of special on CH

SpecialMM

Indicator of special on MM

LoyalCH

Customer brand loyalty for CH

SalePriceMM

Sale price for MM

SalePriceCH

Sale price for CH

PriceDiff

Sale price of MM less sale price of CH

Store7

A factor with levels No and Yes indicating whether the sale is at Store 7

PctDiscMM

Percentage discount for MM

PctDiscCH

Percentage discount for CH

ListPriceDiff

List price of MM less list price of CH

STORE

Which of 5 possible stores the sale occured at