BA Homework #5

Business Analytics - Spring 2020, NYU

INSTRUCTIONS & SUBMISSION

Complete the problems below and submit via NYUClasses. Submit a PDF file with your answers, graphs, and R code.

Problem 1: Progresso Soup Sales

You are provided data for sales of Progresso soup in the U.S. The data are derived from approximately 2000 supermarkets across the country and span 6 years In the file "Progresso.csv" you are provided with the following variables:

IRI_Key	Store ID
Month	1=Jan, 2=Feb, etc.
Region	Region of the US where the store is located (4 regions)
Low_Income	Indicator for low income zip codes
High_Income	Indicator for for high income zip codes
Price.Campbell	Price of Campbell soup
Price.PL	Price of the private label
Price.Progresso	Price of Progresso soup
Sales.Progresso	Sales of Progresso soup
Category_Sales	Total sales of all brands This number is not known before you know the sales of Progresso soup. So it should not be used as a predictor in regression model.

Link to dataset:

https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/Progresso_Soup.csv

Questions:

1. Create a dummy variable for "Winter" months defined as Oct, Nov, Dec, Jan & Feb and answer the following:
    a. What patterns are you seeing in the data?
    b. What patterns are you seeing in sales during the Winter months?
    c. Compute the "Market Share" for Progresso (as percentage of total sales) in the Winter vs. non-Winter months
2. Develop a linear regression model to predict Progresso sales. Explain the results of the regression model (model strength, variable importance, the relationship between the predictors and the dependent variable).

3. Understand your model and give actionable recommendations to the marketing department of Progresso.

Problem 2: Diamond Quotes

A pricing quote for a diamond engagement ring has this characteristic of Price: $3100, Carat Weight: 0.9, Cut: Very Good, Color: J, Clarity: SI2, Polish: Good, Symmetry: Very Good, Certification: GIA. See enclosed characteristics chart

Questions:

1. Run a linear model that predicts price and answer the following questions
    a. Is the quote priced fairly or is the diamond overpriced?
    b. What is the interpretation of each coefficient?
    c. Is this a good model? Justify your answer
2. Drop value of wholesaler #3 and run a second model
    a. What is the impact of dropping wholesaler #3? Why is this needed? Compare both models
    b. Is model 2 better and more correct giving the quote above? Comment on the difference between models terms of "goodness of fit" vs "correctness"

Characteristic	Scale	Comments
Carat		1 carat = 0.2 grams
Color	D-F	Colorless
	G-I J-K	Near colorless Faint yellow
	L-N	Very light yellow
	O-S T-Z	Light yellow Yellow
Cut	Poor	
	Fair Good	
	Very good Excellent	
	Ideal	
Clarity	FL IF	Flawless: No flaws Internally Flawless: No internal flaws
	VVS1	Very, Very Slightly Included: very, very few inclusions at 30×
	VVS2 VS1	Very, Very Slightly Included: very few inclusions at 30× Very Slightly Included: few inclusions at 30×
	VS2	Very, Very Slightly Included: several inclusions at 30×
	SI1	Slightly Included: very, very few inclusions at 10×
	SI2	Slightly Included: very few inclusions at 10×
	SI3 I1 I2	Slightly Included: several inclusions at 10× Included: very few inclusions, but visible to the naked eye Included: few inclusions visible to the naked eye
	I3	Included: several inclusions visible to the naked eye

Link to dataset:

https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/Diamonds.csv