You will answer the following questions that are related to the concepts in the Chapter 3 of the textbook (Larose and Larose).

*Hands-on analysis:*

For the following exercises, build upon your EM project resulting from this week’s EM hands-on tutorial. Read this week’s chapter of the textbook and review the slides before you start the homework. Have the tutorial handy and use it as reference. For your reference, a table of the complete list of variables and definitions is appended in the end of this assignment.

#1. Develop appropriate EDA procedures, analyze and report the association between the target variable, *Churn*, and each of the following nominal/binary variables and interval variables:

* Nominal/binary variables
  + VMail Plan
  + Area Code
  + State
* Interval variables:
  + Account Length
  + VMail Message
  + Day Mins
  + Day Calls
  + Eve Mins
  + Eve Calls
  + Night Mins
  + Night Calls
  + Intl Mins
  + Intl Calls

In your reports, for each variable, answer the following and provide your rationale.

1. Does that variable appear to be associated with the target variable?
2. Should you include it as an input variable in a supervised DM method?

#2. Explain why you need not conduct the analyses above for the following variables:

*Day Charge, Eve Charge, Night Charge, Intl Charge*?

Points breakdown:

|  |  |
| --- | --- |
| # | Points |
| 1 | 2.0 |
| 2 | 1.0 |

This homework is worth 3 points (out of 100 points).

Due date: see syllabus section **Tentative Lecture Outline**

Requirements (Important! Read carefully):

* Use MS Word to build your answers; submit a Word file to Blackboard.
* For non-hands-on questions, elaborate your answers as much as you can. Include and elaborate your rationale if asked, and the main steps taken if relevant.
* **For hands-on questions, you are required to use SAS EM**. Do not only provide the short answer to the question. Besides the short answer, you are required to paste the appropriate screenshots of the relevant steps in SAS EM, including but not limited to, diagram, interface, results output, plot/chart, etc., wherever appropriate. You are advised to err on the side completeness in including screenshots, typically the more the better, as they will allow the Professor or TA to evaluate your competence with SAS.
  + You will receive substantial deduction in points if you only include a short answer, when it is indeed necessary to see relevant screenshots of your EM project and/or process for your Professor or TA to evaluate how you obtained the result in SAS EM.

|  |  |
| --- | --- |
| **Variable** | **Definition** (*obvious variables not explained*) |
| Churn | (binary) your target variable, 1 is your primary outcome |
| Account Length | (continuous) how long account has been active |
| Area Code | (categorical) |
| CustServ Calls | (continuous) number of calls to customer service department |
| Int’l Plan | (binary) International plan |
| Phone | (categorical) unique phone number of the customer; good for an ID field |
| State | (categorical) |
| VMail Plan | (binary) voice mail plan |
| VMail Message | (continuous) number of voice mail messages |
| Day Mins | (continuous) total day minutes |
| Day Calls | (continuous) total day calls |
| Day Charge | (continuous) total day charge |
| Night Mins | (continuous) total evening minutes |
| Night Calls | (continuous) total evening calls |
| Night Charge | (continuous) total evening charge |
| Eve Mins | (continuous) total night minutes |
| Eve Calls | (continuous) total night calls |
| Eve Charge | (continuous) total night charge |
| Intl Mins | (continuous) total international minutes |
| Intl Calls | (continuous) total international calls |
| Intl Charge | (continuous) total international charge |