Aryan Sharma

Problem Statement

Download the file "TelcoChurn.xlsx". This file contains data from a sample of 7043 subscribers of telephone and/or Internet Services for a large telco. We want to create three separate models to understand the predictors of churn of (i) subscribers of telephone services, (ii) subscribers of internet services, and (iii) people who subscribe to both services. Analyze the data carefully (data definitions provided in the second worksheet of the Excel file). Submit your results in the form of a nicely formatted Word (or PDF) file and your R code as two separate files.

1. Describe the process by which you cleaned, processed, and partitioned data as necessary.
2. What predictors do you think contributes to the churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services? Explain the rationale for your answer.
3. Create training and test data sets with a 75:25 split using a random seed of 1024. Use the training data to train three logit models with the variables you identified in Question 2. Combine the outputs of the three modes using stargazer.
4. What are the top three predictors of churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services. Explain using marginal effects how much each predictor contributes to churn occurrence.
5. Fit your models using test data, and compute recall, precision, F1-score, and AUC values for each of your three models. Create a table with these values.