Group Assignment 4

Write the name of all your group members at the top of the report.

Attach all pieces of the code to the respective part of the question.

Background

Neiman Marcus, a luxury department store, has struggled to shift its high-value customers toward online shopping. These customers, typically older and wealthier, strongly prefer the in-store shopping experience. However, the company's online store has successfully attracted a new customer base, generating 11% of total revenue. This presents an opportunity to expand Neiman Marcus' digital presence. To improve online sales, the company formed a dedicated e-commerce team, including Angela Wong, who was tasked with optimizing customer emails. Wong focused on two key challenges: determining the right email frequency and personalizing email content.

Finding the Right Email Frequency

Initially, Neiman Marcus had no limits on email frequency, and each department sent promotional emails independently. This led to an overwhelming 11 emails per week for an average customer. Wong implemented a 5-email-per-week limit, ensuring customers received messages from only one department at a time.

However, Wong suspected 5 emails per week was still excessive, leading to potential unsubscribes, reducing future engagement. To determine the optimal frequency, she randomly assigned 30,000 customers into four groups, receiving 1 to 4 emails per week. After six months, she measured:

- Average monthly revenue per customer
- Unsubscribe rates

Personalizing Emails

While reducing email volume was an easy win, the next challenge was targeting emails more effectively. Wong argued that email content should be personalized based on customer preferences. However, Neiman Marcus' department managers opposed this, wanting to send emails to all customers. Instead of running a costly experiment, Wong realized she could use existing data—since emails were randomly assigned to customers, she could analyze which emails led to higher engagement and purchases. By modeling “next product to buy,” she aimed to tailor emails at the individual level, rather than broad customer segments.

Wong’s plan was to estimate a “next product to buy” model which would allow her to figure out which department was the best option for each customer. Wong asked her analytics team to pull the following data:

- Which department sent the most recent email to each customer
- The basic demographic information that the company had about each customer
- Historical information about customer purchases — specifically, how many items the customer had purchased from each of the different departments in the past year
- Outcome variables: whether the customer clicked on the email and made a purchase, and how much they spent during that purchase occasion
- Which departments the customer purchased from when they clicked on the email, and how much they spent in each department.

In this context, buyers were defined as people who clicked on the email link and then went on to make a purchase from the Neiman Marcus online store.

Question 1) Assume that the cost of goods sold is 60% and that the monthly discount rate is 1%. Looking at the values in Exhibit 1, note that the average monthly revenue for a subscribed customer rises as the company sends more emails. In addition, the average monthly revenue for an unsubscribed customer also rises as the company sends more emails. What could explain both of these patterns? Additionally, calculate the 6-month LTV for each of the four tested email frequencies. Given the LTV analysis, how many emails-per-week should Neiman Marcus be sending to its customers? This email frequency should apply to all customers; Neiman Marcus doesn’t want to implement a different email frequency for different kinds of people.

Note that you can use Excel for calculating LTV if you prefer using that software.

Exhibit 1

	1 email per week	2 emails per week	3 emails per week	4 emails per week
Avg. monthly unsubscription rate	2.4%	3.4%	9.7%	17.1%
Avg. monthly revenue for a subscribed customer	$9.97	$14.21	$14.56	$14.89
Avg. monthly revenue for an unsubscribed customer	$6.32	$6.36	$6.70	$7.11

Question 2) Run a series of logistic regressions of "click_and_purch" dependent variable based on different messages (note that we practiced something similar in the class when deciding which book offer type to send to customers). Based on the prediction probabilities, calculate which message should be sent to each customer. Show the percentage of customers who will receive each of the seven types of messages. Now, use the profitability instead of probability to guide your message targeting decisions. Show the percentage of customers who will receive each of the seven types of messages if profitability is the objective instead of just probability.

Note that when getting predictions using the predict.glm() or predict command, if you get data.frame error, you can replace "data = data" part with simply "data" to resolve it.

Question 3) Compare the average expected profit if we send each customer the targeted message based on the highest profitability (not probability criteria) against the case of sending customers different messages at random. How much more profit can we gain compared to the random case if the entire customer base is 1,000,000 customers?

Variable Name	Variable Description
user_id	Unique ID number for each customer
click_and_purch	Did the customer click on the email and complete a purchase within two days of receiving the email? (1 = yes, 0 = no)
os_total	Total order size (in dollars) conditional on the customer clicking on the email and making a purchase (0 otherwise)
message	The most recent message that the customer received
age	Age group of the customer (coded in 4 buckets)
female	The customer's gender (1 = female, 0 = not female)
income	Average income at the neighborhood level of the customer, rounded to the nearest $5000
education	Percent of college graduates at the neighborhood level of the customer, coded from 0-100
children	Average number of children at the neighborhood level of the customer
region	Which region of the country the customer lives in (coded in 7 regions)
numitems_X	The number of items that the customer bought from department X over the last year
os_X	The department-specific order size (in dollars). This is a breakdown of the os_total variable. How much did the customer spent in each department X if they clicked on the email and completed a purchase within two days of receiving the email? (0 otherwise)

Description of variables