Phishing Campaign Drive – who falls for phishing attempts?

This is a simulated dataset. There is a (fictious) company with several thousand employees. The company is testing which employees fall for a phishing attempt. The company is testing three different phishing campaign approaches using different modalities. The modalities are:

1. an email requesting users to enter data
2. an email requesting users to open an attachment
3. an email requesting users to click on a link

The company is repeatedly sending out a marketing campaign using one of the modalities above to a random sample of their employees over the course of several months. Employees can be sent multiple different phishing attempts.

Data Structure:

- Lookup Phishing table:
    - This has one row for each user outreach attempt. If the user failed the phishing attempt, then the datetime of the event is recorded (opened email, clicked link, submitted data, or opened attachment). If not it is blank.
    - This table can be linked to the employee national table (on EmployeeID) and the Campaign table (on Campaign ID).
- EmployeeNational table:
    - One row per employee – columns are self-explanatory
- CampaignPhishing
    - One row per phishing campaign

Variables you should create for analysis:

- Age as of campaign date
- Tenure with company as of campaign date (time since hire date)

Intentional data issues to be cleaned:

- Some rows have lower case education level
- Some rows have date of birth == 1900-01-01

Some notes on the patterns in the data:

- Time of day the campaign was sent has an effect on user response
- Tenure with the company has an effect on user response
- Education level with the company has an effect on user response
- The effect of age on user response depends on the type of campaign
- The effect of age on user response depends on the type of campaign

    ○ Age effect is not always linear
 - The effect of department on user response depends on the type of campaign (text, email, IVR)

Here are the actual simulated relationships:

```
if (Campaign[i,CampaignType == 'Text_Message']){
    beta_c <-
      (-0.5
      -  0.4*as.numeric(scale(0.18*t_dt_user[,age]**6))
      + .3*t_dt_user[,ifelse(Department=="Marketing",1,0)]
      - .4*t_dt_user[,ifelse(Department=="Operations",1,0)]
      +  1.5*t_dt_user[,event_offset]
      +  0.8*t_dt_user[,state_offset]
      + .4*t_dt_user[,as.numeric(scale(tenure,center = F))]**0.3
      - .4*t_dt_user[,ifelse(as.numeric(Campaign[i,substr(format(DateTimeSent),12,13)]) > 13,
                       1,0)]
      +  .3*t_dt_user[,IsManager]
      + .2*t_dt_user[,ifelse(EducationLevel=="College",1,0)]
      + .4*t_dt_user[,ifelse(EducationLevel=="Grad School",1,0)])
if (Campaign[i,CampaignType == 'Email']){
    beta_c <-
      (-1
      +  0.4*as.numeric(scale(0.18*t_dt_user[,age]))
      #- .6*t_dt_user[,ifelse(Department=="Marketing",1,0)]
      - .4*t_dt_user[,ifelse(Department=="Operations",1,0)]
      +  1.5*t_dt_user[,event_offset]
      +  0.8*t_dt_user[,state_offset]
      + .4*t_dt_user[,as.numeric(scale(tenure,center = F))]**0.3
      - .4*t_dt_user[,ifelse(as.numeric(Campaign[i,substr(format(DateTimeSent),12,13)]) > 13,
                       1,0)]
      +  .3*t_dt_user[,IsManager]
      + .2*t_dt_user[,ifelse(EducationLevel=="College",1,0)]
      + .4*t_dt_user[,ifelse(EducationLevel=="Grad School",1,0)])

if (Campaign[i,CampaignType == 'IVR_Phone']){
    beta_c <-
      (-1
      +  0.3*as.numeric(scale(0.18*t_dt_user[,age]**10))
      - .6*t_dt_user[,ifelse(Department=="Marketing",1,0)]
      + .4*t_dt_user[,ifelse(Department=="Operations",1,0)]
      +  1.5*t_dt_user[,event_offset]
      +  0.8*t_dt_user[,state_offset]
      + .4*t_dt_user[,as.numeric(scale(tenure,center = F))]**0.3
      - .4*t_dt_user[,ifelse(as.numeric(Campaign[i,substr(format(DateTimeSent),12,13)]) > 13,
                       1,0)]
      +  .3*t_dt_user[,IsManager]
      + .2*t_dt_user[,ifelse(EducationLevel=="College",1,0)]
      + .4*t_dt_user[,ifelse(EducationLevel=="Grad School",1,0)])
```