

---

# DATA MANAGEMENT 1

## INTRODUCTION

This set of exercises will demonstrate how to import, clean, merge, and reshape a dataset.

Our goals for this lab are:

- Import data
- Subset data
- Create appropriate dtypes
- Create new variables as needed
- Merge data

## DATA USED IN THIS EXERCISE:

We will use the following datasets in this exercise:

- data/phishing/Campaign.csv
- data/phishing/Lookup.csv
- data/phishing/User.csv

Concept of the data:

- A company has conducted a phishing campaign to test how susceptible their employees are to falling for a phishing attempt. Multiple campaigns were conducted. There were three types of campaigns – (1) an email requesting users to enter data, (2) an email requesting users to open an attachment, and (3) and an email requesting users to click on a link.
- The User table contains details about all the users (employees) in the system
- The Campaign table contains details about each campaign that was sent, including the time it was sent and the type of campaign it was.
- The lookup table contains the results of each campaign for the users it was sent to.

## QUESTIONS

- 1) Start by writing the import statements for pandas. Also, set up the working directory to point at the folder you have established for this class. Confirm the working directory is what you expect.
- 2) Import the three datasets mentioned above (lookup, user, and campaign):
  - a. First inspect the text file with a text editor
  - b. Import the dataset and save the result
  - c. Check the resulting files: how many rows and columns?
- 3) Write some basic subset and manipulations for these files:
  - a. Print the first 5 rows of each file using `.head()`
  - b. In the user table, subset the table to only the males and print the result. How many rows of 'Males' are there? Repeat this for females.
  - c. Use `.loc[]` to subset to the first 5 rows and the first two columns of the user table
  - d. Use `.iloc[]` to subset to the first 5 rows and the first two columns of the user table