

### Importing External Data and Data Manipulation

Open RStudio and create a new project under your Module 2 folder and call it **Assignment1**. Create a new script called **Mod2Assign1Script** and put the following information at the top of the script using comments (#):

```
# Name: First Last  
# Module 2: Assignment 1
```

Complete the following questions in your script using the example labels below:

```
##Question 1  
R Code for Question 1 goes here  
#Questions and Answers for each part of the code go here  
  
##Question 2  
R Code for Question 2 goes here  
#Questions and Answers for each part of the code go here
```

#### Question 1: Importing an external dataset

- 1) Download the **Marketing.csv** file from Canvas which was originally captured from IBM who provides sample datasets for working with analytical software. Save this file as **Marketing.csv** in your project folder for Assignment1 (Module2 -> Assignment1 folder).
- 2) Import the Marketing.csv file into RStudio using readr which was discussed in the video. Answer the following questions in your script.

```
#How many variables does this dataset contain?  
#What are the number of observations?
```

- 3) Look at the console. Even though you were able to upload the file in RStudio without writing the actual code, R still needs to execute the code in the console. Answer the following question in your script.

```
#What was the first line of code that had to be run to import the file?  
#Why did this line have to be run first before importing the file?
```

### Question 2: Finding and Manipulating Data

- 1) In this module, you were introduced to retrieving data from an atomic vector. The same concepts work for retrieving data from a data frame. Instead of having just a single number to designate the column (from the lecture, `example[4]` returned the fourth entry (or column) in our vector), you will need to also designate the row:

*Example[x,y] – where x is the row and y is the column*

- 2) Within your script, write the code to display the value within the data frame for row 20 in the **SalesinThousands** column. Run this code in your script.
- 3) Answer the question below in your script:

```
#What were the total sales in this row?
```

- 4) Now, using a similar syntax, write the code in your script to change the value of the sales in row 20 to 123.45
- 5) Write the code within your script to doing the following comparisons:
  - a) Check to see if **SalesinThousands** for Location 1 - Week 4 (row 4) > **SalesinThousands** for Location 9 - Week 4 (row36). Run the code and in the script answer the following question:

```
#Were the sales at location 1 greater than location 9 for week 4?
```

- b) Check to see if the **AgeOfStores** in row 389 and row 453 equal to each other. Run the code and in the script answer the following question:

```
#Are the ages of the stores in rows 389 and 453 equal?  
#What would happen if we only use 1 equal sign in the comparison?
```

- 6) You can also search through your data frame to find specific values using the syntax below:

```
Data_frame_name[ which(Data_frame_name$ColumnName=='character_value'), ]  
  
Marketing[ which(Marketing$MarketSize=='Medium'),]
```

In the above example, this would return all the rows in the Marketing data frame that contain a value of Medium in the MarketSize column. Notice that we are using a character to search (i.e., Medium) but numbers can also be used (e.g., `Marketing$MarketID==1`).

- 7) Write the code in your script to search in the Marketing data frame and display all stores that have an age of 22 (Note: this csv file has 4 weeks of data for each store).

## **Module 2: Assignment #1**

8) Run the code and answer the following question in script:

```
#How many stores in the data are 22 years old?
```

9) Upload the **Mod2Assign1Script.R** file into Canvas.