

LAST MODIFED: 061619

AXTRIA INSTITUTE

R102 TRAINING ASSIGNMENT

Copyright © 2019 Axtria Inc. All Rights Reserved.

The information contained herein is subject to change without notice.

# NOTES

* Please place your output for Questions 1 and 2 in different files (CSV/Excel/Text) and submit both your output files and R scripts together.
* Read each sub-question completely before working on it.
* Ideally no for-loops should be used for this assignment. Please take advantage of R's vector-based functions.

# subsetting dataframes

**HINT:**

When subsetting a dataframe you always subset in the form DataFrame[row , column] or DataFrame[row , ] or DataFrame[ , column]

**QUESTIONS:**

* 1. Import training\_data1 and name the dataframe “ims\_trx”.
  2. Create a new column called “TRX\_Round” that is the TRX\_Qty value rounded to the nearest number (integer).
  3. Create another new column called “target\_flag” and from the trgt\_flg column change “Y” into Yes and “N” into No.
  4. Save the strings “ims\_id”, “TRX\_Round”, and “target\_flag” into a vector named “subsetCols”. Use the subsetCols vector to select these columns from the ims\_trx dataframe.
  5. Find all values in the TRX\_Round column greater than 100 and save this to a vector named “subsetRows”. Use the subsetRows vector to select these rows from the ims\_trx dataframe.
  6. Use subsetCols and subsetRows to create a final subset of ims\_trx and save it as “trx\_over\_100”.

# JOINING AND SUMMATIONS

**QUESTIONS:**

* 1. Import the three datasets. Name training\_data1 as “monthlyTRX”, training\_data2 as “territory”, and training\_data3 as “specialtyGroup”. In territory and specialtyGroup replace spaces or periods in column titles with underscores.
  2. Left join territory with specialtyGroup on the Spec and IMS Specialty columns and name this dataframe “terr2”. Keep only the ims\_id, Territory\_ID, and Specialty\_Group columns. For ims\_id's that do not have a specialtyGroup i.e. they are “NA”, set the Specialty\_Group to “OTHER”.
  3. Using the terr2 dataframe create a vector or dataframe named “SpecialtyCount” that is the count of ims\_ids in each specialtyGroup.
  4. Now inner join the terr2 dataframe with the monthlyTRX dataframe on their ims\_ids and save it into a dataframe called “ims\_data”.
  5. Using the ims\_data dataframe sum the TRX\_qty column by ims\_id and save it into a dataframe called trx\_sum\_2012. Order this dataframe from greatest to least.
  6. Using the ims\_data dataframe remove all rows where trgt\_flg is N from the dataset and save as targeted\_physicians
  7. Create a function named topTerritory that takes the targeted\_physicians dataframe and a month as input and outputs the Territory ID with the highest TRX\_Qty in the month.