**Coursera-DS3-SamsungProject**

**Coursera Data Science specialization 3 - Final Project**

R E A D M E

The script 'run\_analysis.R' performs the following actions on the UCI HAR Dataset. The dataset is divided into

Traing and Test data sets. The script first merges and then summarizes the data.

**LIBRARIES**

The script loads the following libraries:

library(readr)

library(data.table)

library(plyr)

library(dplyr)

**MERGE TRAINING AND TEST DATA SETS**

The following operations are performed:

* Merge the raw data sets. This is broken down into the following discrete actions:
  + Load data from the files './train/X\_train.txt' and './test/X\_test.txt'
  + Read the data into two variables 'SamsungDF1' and 'SamsungDF2', as fixed width using a prepared list of widths 'Width\_txt', a vector of 561 column widths of 16 characters each.
  + rbind these two data sets into one new data set 'SamsungDFAll'.
* Convert the raw data into a meaningful data set through the following actions:
  + Apply meaningful column names by reading the list of 561 feature names from file './features.txt' and add the names as a new column in dataset 'SamsungDFAll'.
  + Add meaningful activity labels:
    - Read the activity IDs for the training and test data sets from files './train/y\_train.txt' and './test/y\_test.txt', giving both lists the column name 'ActivityID'.
    - Combine these activity data sets into a new data set 'ActivitiesAll' using rbind.
    - Read the meaningful activity labels into the data table 'ActivityNames' from the file './activity\_labels.txt', applying the name 'ActivityID' to the index column, and 'Activity' to the labels column.
    - Add the 'Activity' labels to the 'ActivitiesAll' data set using the join function to preserve the same order as the original training and test data sets, and 'ActivityID' as the common variable .
    - Add the activity labels to dataset 'SamsungDFAll' using cbind.
  + Add subject IDs:
    - Load the subject IDs for the training and test data sets from the files './train/subject\_train.txt' and './test/subject\_test.txt', applying the column name 'SubjectID'.
    - Combine these subject sets into a new data set 'SubjectsAll' usinf rbind.
    - Add the Subjects to dataset 'SamsungDFAll' using cbind.
* Output the combined data set as csv file 'Merged.csv'

**SUMMARIZE DATA**

The following operations are performed:

* Generate a subset of 'SamsungDFAll', 'FilteredDF', containing only the columns 'SubjectID', 'Activity', and any column containing the text strings 'Mean', 'mean', or 'std' (standard deviation).
* Summarize the filtered data by 'Activity' and 'SubjectID' and apply the mean function to the remaining variables. Group by 'Activity' first then 'SubjectID' to sort the data set by subject ID.
* Apply meaningful names to the group by columns, 'Activity' and 'SubjectID'.
* Move the column 'SubjectID' to the start of the data set.
* Output the sumarized data set as csv file 'Summary.csv'

**END OF SCRIPT**

**Example output from Merged.csv**

SubjectID Activity tBodyAcc tBodyAcc tBodyAcc  
 -mean()-X -mean()-Y -mean()-Z

1 1 STANDING 0.28858451 -0.020294171 -0.13290514  
 2 1 STANDING 0.27841883 -0.016410568 -0.12352019  
 3 1 STANDING 0.27965306 -0.019467156 -0.11346169

**Example output from Summary.csv**

SubjectID Activity tBodyAcc tBodyAcc tBodyAcc  
 .mean...X .mean...Y .mean...Z

1 1 LAYING 0.2215982 -0.040513953 -0.11320355  
 2 1 SITTING 0.2612376 -0.001308288 -0.10454418  
 3 1 STANDING 0.2789176 -0.01613759 -0.11060182