**#Course Project**

**The Data Source**

Original Data Source:

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

Original Description of the Data Set:

<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

**Data Set Information**

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

**Transformation Details**

You should create one R script called run\_analysis.R that does the following:

1. Merges the training and the test sets to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive activity names.
5. Creates a second, independent tidy data set with the average of each variable for each activity and each subject.

**Steps**

1. On your local drive, create a directory named “UCI HAR Dataset”. Download and unzip the data source into this folder.
2. Create and save “run\_analysis.R” in the parent folder of “UCI HAR Dataset”, then set it as your working directory using setwd() function in RStudio.
3. When "run\_analysis.R" is run it will generate a new file named tiny\_data.txt in your working directory.

**Dependencies**

“run\_analysis.R” file will help you to install the dependencies automatically. It depends on dplyr and data.table libraries.

**What “run\_anakysis.R” does**

1. Loads both test and train data.
2. Loads the features and activity labels.
3. Extracts the mean and standard deviation column names and data.
4. The grep command is used to filter columns that contain Mean or Std. The data\_mean\_std variable is defined that consists of final\_data and subset using its columns.
5. Processes the data and merges it into the final data set.
6. Labels activities and variable names appropriately - Through the use of the gsub command to replace assigned names with more descriptive and friendly names to make the data set easy to use and in preparation for exporting as a tidy data set.

**Output**

A variable is created called tidy\_data\_average as a data set with average for each activity and subject. Then the entries are ordered and then written out as data file tidy\_data.txt that contains the processed data.