# Module 4 Project Code Book

*activityIDs* – Using the provided "activity_labels.txt" in the UCI HAR Dataset folder, a table is generated that contains all the activity IDs and associated labels.

*subjectTrainID* – Using the "subject_train.txt" in the train folder, a table is generated that contains all the subject IDs in the training set.

*xTrain* – Using the "X_train.txt" in the train folder, a table is generated that contains all the data within the training set.

*yTrain* – Using the "y_train.txt" in the train folder, a table is generated that contains the activity IDs associated with the subject and data training set.

*subjectTestID* – Using the "subject_test.txt" in the test folder, a table is generated that contains all the subject IDs in the test set.

*xTest* – Using the "X_test.txt" in the test folder, a table is generated that contains all the data within the test set.

*yTest* – Using the "y_test.txt" in the test folder, a table is generated that contains all the activity IDs associated with the subject and data in the test set.

*testSet* – Combines all the subject IDs, activity IDs, and training data from the *subjectTrainID*, *yTrain*, and *xTrain* tables respectively. They are coerced into the same order as listed.

*trainSet* – Combines all the subject IDs, activity IDs, and test data from the *subjectTestID*, *yTest*, and *xTest* tables respectively. They are coerced into the same order as listed.

*fullSet* – Stacks the *testSet* and *trainSet* tables into a combined data set.

*subjectIDList* – Pulls the unique subject IDs from the *fullSet*.

*IDcheck* – Initially generates a blank list to be filled in subsequent *for* loops in order to insure that subsequent calculations are not repeated.

*meanDF* – Initially generates a black dataframe with columns named “Subject”, “Activity”, “DataMean”, and “StandardDeviation”. This will be filled in respect to the columns from data generated later.

*“for (subject in subjectIDList)”* – This loop goes through each of the unique Subject IDs found in *subjectIDList* and filters the data found in *fullSet* by each interation into a table called *subjectData*. It then pulls all the unique Activity IDs found in that iteration’s *subjectData* table into a table called *subjectActivity* before performing a secondary *for* loop. After the secondary loop, *IDcheck* is reset to a blank list.

*“for (activity in subjectActivity)”* – This secondary *for* loop goes through each unique activity ID found in *subjectActivity* and checks if the new iterations activity ID is found in the *IDcheck* list. If found, it moves to the next iteration. It then filters the *subjectData* table by the activity ID into *activityData* which is then used to calculate the mean of the data found in the data columns of *activityData* into *activityMean* as well as the standard deviation into *activitySD*. *IDCheck* is then appended with the current iteration’s activity ID and a *newData* dataframe is generated containing the subject ID from the external *for* loops iteration, the activity ID from the secondary *for* loop iteration, the activity mean calculation from the *activityMean* variable, and the activity standard deviation from the *activitySD* variable. This is then stacked into the *meanDF* dataframe outside the loop.

*Results* – Reorders and stores the *meanDF* dataframe first by the Subject column and the Activity column. The Activity column is subsequently overwritten by matching the activity ID numbers in the column with the correct ID name from the *activityIDs* table.