**Code Book**

**#Initial configuration set-up: The following steps are meant to configure the working directory and load the required packages.**

getwd()

setwd("C:/Users/Arvin/Desktop/Coursera\_assignments")

library(dplyr)

library(RCurl)

library(tidyverse)

**#Starting the download of the dataset file: The following steps were taken to download the dataset file from the website which was in .zip format and then unzipped the files to UCI HAR Dataset folder.**

C3datafile <- "Coursera\_Wk3\_dataset.zip"

if (!file.exists(C3datafile)) {

datafileurl <-

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

download.file(datafileurl, C3datafile, method = "curl", mode="wb")

}

#Checking the files and listing them

if (!file.exists("UCI HAR Dataset")) {

unzip(C3datafile)}

list.files ("UCI HAR Dataset", recursive = TRUE)

**#Reading into the dataframes and assigning the variables: The files were read and the column names assigned respectively to get a clean meaningful data with column names.**

data\_features <- read.table("UCI HAR Dataset/features.txt",

col.names = c("n","functions"))

data\_activities <- read.table("UCI HAR Dataset/activity\_labels.txt",

col.names = c("code", "activity"))

x\_train <- read.table("UCI HAR Dataset/train/X\_train.txt",

col.names = data\_features$functions)

y\_train <- read.table("UCI HAR Dataset/train/y\_train.txt",

col.names = "Act\_Code")

subject\_train <- read.table("UCI HAR Dataset/train/subject\_train.txt",

col.names = "Sub\_Code")

x\_test <- read.table("UCI HAR Dataset/test/X\_test.txt",

col.names = data\_features$functions)

y\_test <- read.table("UCI HAR Dataset/test/y\_test.txt",

col.names = "Act\_Code")

subject\_test <- read.table("UCI HAR Dataset/test/subject\_test.txt",

col.names = "Sub\_Code")

**#For objective1:Merge the training and the test datasets into one data set: For creating one dataset initially the train dataset and test datasets were merged column wise to get the activity and subject columns specified with the corresponding data observations. Then for the merging of both the datasets into one dataset, merging was done row wise to have the two datasets merged with corresponding observations.**

merged\_train\_data= cbind(y\_train, subject\_train, x\_train)

merged\_test\_data= cbind(y\_test, subject\_test, x\_test)

## Merging both the above datasets

BothSets\_merged<- rbind(merged\_train\_data, merged\_test\_data)

**#For Objective2: Extract only the measurements on the mean and standard deviation for each measurement: BothSets\_merged dataframe was used to select Sub\_Code and Act\_Code columns denoting subject ID and Activity ID and pipe function used to select the columns containing mean and standard deviation in the names and brought together in the Dataset\_mean\_std dataset.**

Dataset\_mean\_std <- BothSets\_merged %>% select("Sub\_Code", "Act\_Code", contains("mean"), contains("std"))

**#For Objective3&4: Use descriptive activity names to name the activities in the data set and Appropriately labels the data set with descriptive variable names.: using names() the column names were appropriately redefined to clearly assign the descriptive nature.**

Dataset\_mean\_std$Act\_Code<-data\_activities[Dataset\_mean\_std$Act\_Code, 2]

names(Dataset\_mean\_std)[1] = "Subject\_ID"

names(Dataset\_mean\_std)[2] = "Activity"

names(Dataset\_mean\_std)<-gsub("Acc", "Accelerometer", names(Dataset\_mean\_std))

names(Dataset\_mean\_std)<-gsub("Gyro", "Gyrometer", names(Dataset\_mean\_std))

names(Dataset\_mean\_std)<-gsub("Mag", "Magnitude", names(Dataset\_mean\_std))

names(Dataset\_mean\_std)<-gsub("^f", "Frequency", names(Dataset\_mean\_std))

names(Dataset\_mean\_std)<-gsub("-freq()", "Frequency", names(Dataset\_mean\_std))

names(Dataset\_mean\_std)<-gsub("tBody", "TimeBody", names(Dataset\_mean\_std))

**# For Objective5.From the data set in step 4, creates a second, independent tidy data set**

**#with the average of each variable for each activity and each subject: aggregate function used to subset using the column description in Second\_dataset and reodered in terms of the Subject ID in the Third\_dataset and finally the dataset written as text file “Data-revised.txt” as an output file.**

Second\_dataset<- aggregate(.~Subject\_ID + Activity, Dataset\_mean\_std, mean)

Third\_dataset<- Second\_dataset[order(Second\_dataset$Subject\_ID,

Second\_dataset$Activity),]

write.table(Third\_dataset, "Data-revised.txt", row.names=FALSE)

**#End of the script "run\_analysis.R": The R script saved as run\_analysis.R**