1. GPS information should be used to calculate the length/distance of the trip, and then use this computed trip length as a predictor variable.

To calculate the distance, you will need to create your own function by using the ***function()*** in R. Here, a function ***dist()*** is created for you to ease your work, you just need to apply it in R.

dist <- function (long1, lat1, long2, lat2){

rad <- pi/180

a1 <- lat1 \* rad

a2 <- long1 \* rad

b1 <- lat2 \* rad

b2 <- long2 \* rad

dlon <- b2 - a2

dlat <- b1 - a1

a <- (sin(dlat/2))^2 + cos(a1) \* cos(b1) \* (sin(dlon/2))^2

c <- 2 \* atan2(sqrt(a), sqrt(1 - a))

R <- 6378.145

d <- R \* c

return(d)

}

1. For the variables that include the date and time information, you should extract useful information such as day of week/month/quarter and hour as predictors.

Functions you may use:

as.POSIXlt: convert an object to a Date/Time class

weekdays: give the day of the week

months: give the month name

For more details regarding Date and Time, this following website is useful:

<https://bookdown.org/rdpeng/rprogdatascience/dates-and-times.html>

1. Use both classification tree and logistics regression to fit predictive models.