

Team_10_Final_Project.Rmd

Team 10

November 16, 2015

```
# Update the working directory for the place on your computer where you store the project
# setwd('C:\\Users\\Aleksey\\Documents\\School\\UW_Data_Science\\UW_Data_Science_450\\Final_Project')
setwd('C:/Users/db345c/Desktop/hw')

# Designate file name for the download data
file.name = "data.csv"

# Usage of the local file named data.csv is preferred. If data is not available, usage of Mac is preferred.
if(!file.exists(file.name)) {
  # works on Mac only. R cannot download HTTPS links on windows any more
  print("Downloading data file")
  download.file("https://www.dropbox.com/s/xnsl4d6u3c3oper/train.csv?dl=0", destfile = file.name, method = "curl")
  print("Finished downloading data file")
  print("Reading data in memory")
  data = read.csv(file.name)
  print("Finished loading data in memory")
} else {
  # read CSV file in memory
  print("Data file exists, reading data file in memory")
  data = read.csv(file.name)
  print("Finished reading file in memory")
}
```

```
## [1] "Data file exists, reading data file in memory"
## [1] "Finished reading file in memory"
```

```
# Extract Complete Cases (removed 4229 records)
print("Removing incomplete records")
```

```
## [1] "Removing incomplete records"
```

```
data = data[complete.cases(data),]
print("Finished removing incomplete records")
```

```
## [1] "Firinshed removing incomplete records"
```

```
# Remove duplicate rows (removed 4443 records)
print("Removing duplicate records")
```

```
## [1] "Removing duplicate records"
```

```
data = data[!duplicated(data), ]
print("Finished removing duplicate records")
```

```
## [1] "Finished removing duplicate records"
```

```
# Add weekend column (0 - workday, 1 - Weekend)
print("Started creating 'Weekend' column" )
```

```
## [1] "Started creating 'Weekend' column"
```

```
data$Weekend = 0
data[data$Weekday=="Saturday",]$Weekend = 1
data[data$Weekday=="Sunday",]$Weekend = 1
print("Finished creating 'Weekend' column")
```

```
## [1] "Finished creating 'Weekend' column"
```

```
# Clean workspace, remove variable that holds file name
rm(file.name)
print("Cleaned up workspace, removed 'file.name' variable")
```

```
## [1] "Cleaned up workspace, removed 'file.name' variable"
```

```
...
```