**SpringBoard Data Science Exercise 1**

**Daniel Teodorescu**

**#install necessary packages**

install.packages("tidyr")

install.packages("dplyr")

**#Read in the Excel File**

library(readxl)

refine <- read\_excel("SpringBoard DS Course/refine\_original.csv.xlsx")

refine

**#1. Clean up brand names**

refine[1:6,1] <- "phillips"

refine[14:16,1] <- "phillips"

refine[7:13,1] <- "akzo"

refine[17:21,1] <- "van houten"

refine[22:25,1] <-"unilever"

refine

**#2. Separate product code and number**

library(tidyr)

refine <-tbl\_df(refine)

refine <- refine %>%

separate("Product code/ number", c("Product\_code", "Product\_number"), sep = "-")

**#3. Add product category**

categories<-c(p = "Smartphone", v = "TV", x = "Laptop", q = "Tablet")

refine <- refine %>% mutate(category = categories[product\_code])

refine

**#4: Add full address for geocoding**

refine <- refine %>% mutate(full\_address = paste(address, city, country, sep = ","))

refine

**#5: Create dummy variables for company and product category**

refine <- refine %>%

mutate(company\_philips = ifelse(grepl("philips", company), 1, 0)) %>%

mutate(company\_akzo = ifelse(grepl("akzo", company), 1, 0)) %>%

mutate(company\_van\_houten = ifelse(grepl("van houten", company), 1, 0)) %>%

mutate(company\_unilever = ifelse(grepl("unilever", company), 1, 0)) %>%

mutate(product\_smartphone = ifelse(grepl("Smartphone", category), 1, 0)) %>%

mutate(product\_tv = ifelse(grepl("TV", category), 1, 0)) %>%

mutate(product\_laptop = ifelse(grepl("Laptop", category), 1, 0)) %>%

mutate(product\_tablet = ifelse(grepl("Tablet", category), 1, 0))

refine