**Springboard Data Science**

**Block 3 – Project 4 – Data Wrangling Exercise 1: Basic Data Manipulation**

# 0: Load the data in RStudio

refine\_original <- read.csv(C:\Users\AnnaVaughan\Documents\refine\_original.csv)

refine\_tidied <- refine\_original

# 1: Clean up brand names

refine\_tidied$company <- tolower(refine\_tidied$company)

refine\_tidied$company <- gsub(pattern = "Phillips|phillips|phllips|phillps|phillipS|fillips|phlips", replacement = "philips", x = refine\_tidied$company)

refine\_tidied$company <- gsub(pattern = "Akzo|AKZO|akz0|ak zo", replacement = "akzo", x = refine\_tidied$company)

refine\_tidied$company <- gsub(pattern = "unilver", replacement = "unilever", x = refine\_tidied$company)

# 2: Separate product code and number

refine\_tidied <- separate(refine\_tidied, Product.code...number, c("product\_code", "product\_number"), sep = "-")

# 3: Add product categories

refine\_tidied <- refine\_tidied %>% group\_by(product\_code) %>% mutate(product\_category = product\_code)

refine\_tidied$company %>% gsub(pattern = “p”, replacement = “Smartphone”)

refine\_tidied$product\_category[refine\_tidied$product\_category == "p"] <- "Smartphone"

refine\_tidied$product\_category[refine\_tidied$product\_category == "v"] <- "TV"

refine\_tidied$product\_category[refine\_tidied$product\_category == "x"] <- "Laptop"

refine\_tidied$product\_category[refine\_tidied$product\_category == "q"] <- "Tablet"

# 4: Add full address for geocoding

refine\_tidied\_new <- refine\_tidied\_new %>%

unite(full\_address, address:country, sep=",")

# 5: Create dummy variables for company and product category

refine\_tidied\_evennewer <- refine\_tidied\_new %>% mutate(company\_philips = as.numeric(company == "philips")) %>% mutate(company\_akzo = as.numeric(company == "akzo")) %>% mutate(company\_van\_houten = as.numeric(company == "van houten")) %>% mutate(company\_unilever = as.numeric(company == "unilever"))

refine\_clean <- refine\_tidied\_evennewer %>% mutate(product\_smartphone = as.numeric(product\_category == "Smarthphone")) %>% mutate(product\_tv = as.numeric(product\_category == "TV")) %>% mutate(product\_laptop = as.numeric(product\_category == "Laptop")) %>% mutate(product\_tablet = as.numeric(product\_category == "Tablet"))

# 6: Submit the project on Github

write.csv(refine\_clean, file = "refine\_clean.csv")