KJ Morgan

Data Exercise 1

refine\_original <- read.csv("refine\_original.csv")

refine\_clean <- refine\_original

refine\_clean$company[grepl("p", refine\_clean$company, ignore.case = TRUE)] <- "phillips"

refine\_clean$company[grepl("k", refine\_clean$company, ignore.case = TRUE)] <- "akzo"

refine\_clean$company[grepl("t", refine\_clean$company, ignore.case = TRUE)] <- "van houten"

refine\_clean$company[grepl("r", refine\_clean$company, ignore.case = TRUE)] <- "unilever"

head(refine\_clean)

refine\_clean$product\_code <- substr(refine\_clean$Product.code...number, 1,1)

refine\_clean$product\_number <- substr(refine\_clean$Product.code...number, 3,4)

head(refine\_clean)

refine\_clean$product\_category <- ifelse(refine\_clean$product\_code == "p", "Smartphone",

ifelse(refine\_clean$product\_code == "v", "TV",

ifelse(refine\_clean$product\_code == "x", "Laptop", "Tablet")))

head(refine\_clean)

refine\_clean$full\_address <- paste(refine\_clean$address, refine\_clean$city, refine\_clean$country, sep = ", ")

head(refine\_clean)

refine\_clean$company\_phillips <- ifelse(refine\_clean$company == "phillips", 1, 0)

refine\_clean$company\_akzo <- ifelse(refine\_clean$company == "akzo", 1, 0)

refine\_clean$company\_van\_houten <- ifelse(refine\_clean$company == "van houten", 1, 0)

refine\_clean$company\_unilever <- ifelse(refine\_clean$company == "unilever", 1, 0)

refine\_clean$product\_smartphone <- ifelse(refine\_clean$product\_category == "Smartphone", 1, 0)

refine\_clean$product\_tv <- ifelse(refine\_clean$product\_category == "TV", 1, 0)

refine\_clean$product\_laptop <- ifelse(refine\_clean$product\_category == "Laptop", 1, 0)

refine\_clean$product\_tablet <- ifelse(refine\_clean$product\_category == "Tablet", 1, 0)

head(refine\_clean)

refine\_clean