**1: Clean up brand names**

> refine$company<-c("philips", "philips", "philips", "philips", "philips", "philips","akzo", "akzo", "akzo", "akzo", "akzo", "akzo", "akzo", "philips", "philips", "philips", "Van Houten", "Van Houten", "Van Houten", "Van Houten", "Van Houten","Unilever", "Unilever", "Unilever", "Unilever")

**2: Separate product code and number**

refine<-separate(refine, code, c("code", "number"), sep = "-")

**3: Add product categorie**s

**4: Add full address for geocoding**

> library(tidyr)

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

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

**Company\_philips**

refine$company\_akzo<-refine$company=="akzo"

as.numeric(refine)

as.numeric(refine$company\_akzo)

[1] 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0

> refine$company\_akzo<-as.numeric(refine$company\_akzo)

**Company\_akzo**

refine$company\_akzo<-refine$company=="akzo"

> as.numeric(refine)

> as.numeric(refine$company\_akzo)

[1] 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0

> refine$company\_akzo<-as.numeric(refine$company\_akzo)

**company\_van\_houten**

refine$company\_van\_houten<-refine$company=="Van Houten"

> as.numeric(refine)

> as.numeric(refine$company\_van\_houten)

[1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0

> refine$company\_van\_houten<-as.numeric(refine$company\_van\_houten)

**company\_unilever**

refine$company\_unilever<-refine$company=="Unilever"

> as.numeric(refine)

> as.numeric(refine$company\_unilever)

[1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1

> refine$company\_unilever<-as.numeric(refine$company\_unilever)

**3: Add product categorie**s

refine$product\_category<-c("smartphone", "smartphone", "Laptop", "Laptop", "Laptop", "smartphone", "TV", "TV", "Laptop", "smartphone", "Tablet", "Tablet", "Laptop", "smartphone", "TV", "TV", "Laptop", "TV", "TV", "Laptop", "smartphone", "Laptop", "Tablet", "Tablet", "Tablet")

https://github.com/prinklek/Capstone-project.git