Project\_1.R

stevehaunguyen

Sat May 19 14:34:35 2018

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
  
#0. Load data in RStudio ------------------------------------------------  
  
refine <- read.csv(file = "refine\_original.csv")  
tbl\_df(refine)

## # A tibble: 25 x 6  
## company Product.code...number address city country name   
## <fct> <fct> <fct> <fct> <fct> <fct>  
## 1 Phillips p-5 Groningensingel 147 arnhem the ne~ dhr ~  
## 2 phillips p-43 Groningensingel 148 arnhem the ne~ dhr ~  
## 3 philips x-3 Groningensingel 149 arnhem the ne~ dhr ~  
## 4 phllips x-34 Groningensingel 150 arnhem the ne~ dhr ~  
## 5 phillps x-12 Groningensingel 151 arnhem the ne~ dhr ~  
## 6 phillipS p-23 Groningensingel 152 arnhem the ne~ dhr ~  
## 7 akzo v-43 Leeuwardenweg 178 arnhem the ne~ dhr ~  
## 8 Akzo v-12 Leeuwardenweg 179 arnhem the ne~ dhr ~  
## 9 AKZO x-5 Leeuwardenweg 180 arnhem the ne~ dhr ~  
## 10 akz0 p-34 Leeuwardenweg 181 arnhem the ne~ dhr ~  
## # ... with 15 more rows

#1. Clean up brand names ------------------------------------------------  
refine <- arrange(refine, company)  
  
refine$company[1:7] <- "Akzo"  
refine$company[8:16] <- "philips"  
refine$company[17:20] <- "Unilever"  
refine$company[21:25] <- "Van Houten"  
  
View(refine)  
  
#2. Separate product code and number ------------------------------------  
  
refine$Product\_code <- separate(refine, "Product.code...number", c("Product\_code", "Product\_number"), sep = "-")[ ,2]  
refine$Product\_number <- separate(refine, "Product.code...number", c("Product\_code", "Product\_number"), sep = "-")[ ,3]  
  
#3. Add product categories----------------------------------------------  
  
refine$product\_category <- ifelse(refine$Product\_code == "p", "Smartphone",  
 ifelse(refine$Product\_code == "v", "TV",  
 ifelse(refine$Product\_code == "x", "Laptop",  
 ifelse(refine$Product\_code == "q", "Tablet", NA))))  
  
#4. Add full address for geocoding -----------------------------------------------------------  
  
refine$full\_address <- unite(refine, "full\_address", address, city, country, sep = ",")[ ,3]  
  
#5. Create dummy variables for company and product category -------------  
  
refine$company\_philips <- sample(0:1,nrow(refine), replace=T)  
refine$company\_akzo <- sample(0:1,nrow(refine), replace=T)  
refine$company\_van\_houten <- sample(0:1,nrow(refine), replace=T)  
refine$company\_unilever <- sample(0:1,nrow(refine), replace=T)  
  
refine$product\_smartphone <- sample(0:1,nrow(refine), replace=T)  
refine$product\_tv <- sample(0:1,nrow(refine), replace=T)  
refine$product\_laptop <- sample(0:1,nrow(refine), replace=T)  
refine$product\_tablet <- sample(0:1,nrow(refine), replace=T)  
  
View(refine)