Submit 1

ST

October 30, 2018

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

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)  
library(tidyverse)

## -- Attaching packages -------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ----------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# read refine file  
  
df <- read.csv(file="c:/Users/crish/Desktop/refine\_original.csv", header=TRUE, sep=",")  
# Convert company value to lower case  
df$company <- tolower(df$company)  
# Separate product code and number  
  
data1 <- df %>%   
 separate(Product.code...number, c('product\_code', 'product\_number'), sep = '-')   
# add new columns for Product category, full address and boolean fields for company and product  
data1 <- mutate(data1, productcategory = ' ', full\_address = ' ', company\_philips = ' ', company\_akzo = ' ', company\_van\_houten = ' ', company\_unilever = ' ', product\_smartphone = ' ', product\_tv = ' ', product\_laptop = ' ', product\_tablet = ' ')  
  
### Evaluate full address field using paste function  
data1$full\_address = paste(data1$address, data1$city, data1$country, sep=",")  
  
# loop through data frame and populate necessary values  
for (row in 1:nrow(data1)) {  
  
 if (data1$product\_code[row] == "p") {  
 data1$productcategory[row] <- "Smartphone"  
 data1$product\_smartphone[row] = 1  
 }  
 if (data1$product\_code[row] == "v") {  
 data1$productcategory[row] <- "TV"  
 data1$product\_tv[row] = 1  
 }  
 if (data1$product\_code[row] == "x") {  
 data1$productcategory[row] <- "Laptop"  
 data1$product\_laptop[row] = 1  
 }  
 if (data1$product\_code[row] == "q") {  
 data1$productcategory[row] <- "Tablet"  
 data1$product\_tablet[row] = 1  
 }  
   
}  
data1$company\_philips <- grepl('\*ph\*', data1$company)  
data1$company\_akzo <- grepl('\*ak\*', data1$company)  
data1$company\_van\_houten <- grepl('\*van\*', data1$company)  
data1$company\_unilever <- grepl('\*uni\*', data1$company)