Springboard Data Wrangling Exercise 1

Chinpei Tang May 19, 2016

Introduction

This basic exercise involves cleaning up the data with inconsistent inputs, renaming, separating and consolidating data, and creating new usable data for later analysis. A toy dataset was given, and a few basic tasks were assigned.

Load Data

The given data is in Excel format. The "refine.xlsx" file was first opened in Excel and renamed to "refine original.csv" file, which can be read by RStudio.

As a side note, a quick Google search revealed that there are methods import the xlsx file into RStudio directly, which is out of scope for this exercise.

The csv file is saved in a working directory, and the working directory is set in RStudio. Since the dataset contains header, the dataset is read with header option set true. The dataset is then printed out to analyze.

```
setwd("C:/Users/Chinpei/Documents/GitHub/Springboard_FDS/DW_Ex1")
refine = read.csv("refine_original.csv", header = T)
refine
```

```
##
         company Product.code...number
                                                     address
                                                                city
                                    p-5 Groningensingel 147 arnhem
## 1
        Phillips
## 2
        phillips
                                   p-43 Groningensingel 148 arnhem
## 3
                                    x-3 Groningensingel 149 arnhem
         philips
## 4
         phllips
                                   x-34 Groningensingel 150 arnhem
## 5
         phillps
                                   x-12 Groningensingel 151 arnhem
## 6
                                   p-23 Groningensingel 152 arnhem
        phillipS
## 7
            akzo
                                   v-43
                                           Leeuwardenweg 178 arnhem
## 8
                                           Leeuwardenweg 179 arnhem
            Akzo
                                   v-12
                                    x-5
## 9
            AKZO
                                           Leeuwardenweg 180 arnhem
## 10
            akz0
                                   p-34
                                           Leeuwardenweg 181 arnhem
                                    q-5
## 11
                                           Leeuwardenweg 182 arnhem
           ak zo
## 12
            akzo
                                    q-9
                                           Leeuwardenweg 183 arnhem
## 13
            akzo
                                    x-8
                                           Leeuwardenweg 184 arnhem
## 14
        phillips
                                   p-56
                                           Delfzijlstraat 54 arnhem
## 15
                                           Delfzijlstraat 55 arnhem
         fillips
                                   v-67
## 16
          phlips
                                   v-21
                                           Delfzijlstraat 56 arnhem
## 17 Van Houten
                                   x-45
                                           Delfzijlstraat 57 arnhem
## 18 van Houten
                                   v-56
                                           Delfzijlstraat 58 arnhem
## 19 van houten
                                           Delfzijlstraat 59 arnhem
                                   v-65
## 20 van houten
                                           Delfzijlstraat 60 arnhem
                                   x-21
## 21 Van Houten
                                   p-23
                                           Delfzijlstraat 61 arnhem
## 22
         unilver
                                              Jourestraat 23 arnhem
                                    x-3
## 23
        unilever
                                    q-4
                                              Jourestraat 24 arnhem
## 24
        Unilever
                                    q-6
                                              Jourestraat 25 arnhem
## 25
        unilever
                                              Jourestraat 26 arnhem
                                    q-8
```

```
##
              country
                                  name
## 1
     the netherlands
                         dhr p. jansen
                         dhr p. hansen
## 2
      the netherlands
## 3
     the netherlands
                         dhr j. Gansen
## 4
      the netherlands
                         dhr p. mansen
## 5
     the netherlands
                        dhr p. fransen
     the netherlands dhr p. franssen
## 7
      the netherlands
                         dhr p. bansen
## 8
      the netherlands
                         dhr p. vansen
## 9
      the netherlands
                        dhr p. bransen
## 10 the netherlands
                        dhr p. janssen
## 11 the netherlands mevr 1.
                               rokken
## 12 the netherlands
                      mevr 1.
                                lokken
## 13 the netherlands mevr 1.
                                mokken
## 14 the netherlands mevr 1.
                                mokken
## 15 the netherlands mevr 1.
                                mokken
## 16 the netherlands mevr 1.
                                mokken
## 17 the netherlands mevr 1.
                                sokken
## 18 the netherlands mevr 1.
                                wokken
## 19 the netherlands mevr 1.
                                kokken
## 20 the netherlands mevr 1.
                                Bokken
## 21 the netherlands
                      mevr 1.
## 22 the netherlands mevr 1.
                                gokken
## 23 the netherlands mevr 1.
                               stokken
## 24 the netherlands mevr 1. rokken
## 25 the netherlands mevr 1. rokken
```

Consolidate company names

The first task is to clean up the inconsistent company names. All the company names are printed, and the patterns are recognized.

summary(refine\$company)

```
AKZO
##
                                                Akzo
                                                                      fillips
         ak zo
                      akz0
                                   akzo
##
                                       3
             1
                          1
                                                                             1
##
                               phillipS
      philips
                  phillips
                                           Phillips
                                                         phillps
                                                                       phlips
                          2
##
             1
                                       1
##
      phllips
                  unilever
                               Unilever
                                             unilver
                                                     van houten van Houten
                          2
##
             1
                                       1
                                                   1
                                                                2
                                                                             1
##
   Van Houten
##
             2
```

Since the end results requires all the company names to be converted to lower case, tolower function (with the help of vapply) is applied:

```
refine$company = vapply(refine$company, tolower, character(1))
refine$company
```

```
##
    [1] "phillips"
                       "phillips"
                                     "philips"
                                                    "phllips"
                                                                  "phillps"
    [6] "phillips"
                       "akzo"
                                     "akzo"
                                                    "akzo"
                                                                  "akz0"
## [11] "ak zo"
                                                    "phillips"
                       "akzo"
                                     "akzo"
                                                                  "fillips"
```

```
## [16] "phlips" "van houten" "van houten" "van houten"
## [21] "van houten" "unilever" "unilever" "unilever" "unilever"
```

The pattern recognition grep function is used to identify the patterns of the company names to be consolidated. It can be recognized that all "philips" contains term "ps", "akzo" contains term "ak", "van houten" contains term "van", and "unilever" contains term "uni".

```
refine$company[grep(pattern = "ps", x = refine$company)] = "philips"
refine$company[grep(pattern = "ak", x = refine$company)] = "akzo"
refine$company[grep(pattern = "van", x = refine$company)] = "van houten"
refine$company[grep(pattern = "uni", x = refine$company)] = "unilever"
refine$company
```

```
[1] "philips"
                     "philips"
                                  "philips"
                                               "philips"
                                                            "philips"
## [6] "philips"
                     "akzo"
                                  "akzo"
                                               "akzo"
                                                            "akzo"
## [11] "akzo"
                     "akzo"
                                  "akzo"
                                               "philips"
                                                            "philips"
## [16] "philips"
                     "van houten" "van houten" "van houten" "van houten"
## [21] "van houten" "unilever"
                                               "unilever"
                                                            "unilever"
                                  "unilever"
```

Separate "-" between codes and numbers

Convert the "Product.code...number" into character, and use strsplit to recognize "-" and separate the product codes and number into "product code" and "product number" columns respectively.

```
product_code_number = as.character(refine$Product.code...number)
product_code_number = strsplit(product_code_number, split = "-")
product_code = vapply(product_code_number, function(x){x[1]}, character(1))
product_code

## [1] "p" "p" "x" "x" "x" "p" "v" "v" "x" "p" "q" "q" "x" "p" "v" "v" "x"
## [18] "v" "v" "x" "p" "x" "q" "q" "q"
product_number = vapply(product_code_number, function(x){x[2]}, character(1))
product_number
```

```
## [1] "5" "43" "3" "34" "12" "23" "43" "12" "5" "34" "5" "9" "8" "56" ## [15] "67" "21" "45" "56" "65" "21" "23" "3" "4" "6" "8"
```

Removed the "Product.code...number" column, and substitute with the "product_code" and "product_number" columns.

```
refine = cbind(refine[1], product_code, product_number, refine[3:6])
refine
```

```
##
        company product_code product_number
                                                        address
                                                                 city
## 1
        philips
                                          5 Groningensingel 147 arnhem
                           р
## 2
        philips
                                         43 Groningensingel 148 arnhem
                          p
## 3
        philips
                         X
                                        3 Groningensingel 149 arnhem
                                         34 Groningensingel 150 arnhem
## 4
        philips
                          x
                                         12 Groningensingel 151 arnhem
## 5
        philips
                           х
```

```
## 6
         philips
                                           23 Groningensingel 152 arnhem
                             p
## 7
            akzo
                                           43
                                                Leeuwardenweg 178 arnhem
                             v
## 8
            akzo
                             v
                                           12
                                                 Leeuwardenweg 179 arnhem
## 9
                                            5
                                                Leeuwardenweg 180 arnhem
            akzo
                             х
## 10
            akzo
                                           34
                                                Leeuwardenweg 181 arnhem
                             p
## 11
                                                Leeuwardenweg 182 arnhem
            akzo
                                            5
                             q
## 12
                                            9
                                                Leeuwardenweg 183 arnhem
            akzo
                             q
## 13
            akzo
                             х
                                            8
                                                Leeuwardenweg 184 arnhem
## 14
         philips
                                           56
                                                Delfzijlstraat 54 arnhem
                             р
## 15
         philips
                             V
                                           67
                                                Delfzijlstraat 55 arnhem
## 16
         philips
                             v
                                           21
                                                Delfzijlstraat 56 arnhem
                                           45
                                                Delfzijlstraat 57 arnhem
## 17 van houten
                             X
## 18 van houten
                                           56
                                                Delfzijlstraat 58 arnhem
                             v
## 19 van houten
                             v
                                           65
                                                 Delfzijlstraat 59 arnhem
## 20 van houten
                                           21
                                                Delfzijlstraat 60 arnhem
                             Х
## 21 van houten
                                           23
                                                 Delfzijlstraat 61 arnhem
                             р
## 22
                                            3
                                                    Jourestraat 23 arnhem
        unilever
                             Х
## 23
        unilever
                                            4
                                                    Jourestraat 24 arnhem
                             q
## 24
        unilever
                                            6
                                                    Jourestraat 25 arnhem
                             q
                                                    Jourestraat 26 arnhem
## 25
        unilever
                             q
                                            8
##
              country
                                   name
     the netherlands
                          dhr p. jansen
## 2
      the netherlands
                          dhr p. hansen
      the netherlands
                          dhr j. Gansen
## 4
                          dhr p. mansen
      the netherlands
      the netherlands
                         dhr p. fransen
## 6
      the netherlands
                       dhr p. franssen
      the netherlands
                          dhr p. bansen
## 8
      the netherlands
                          dhr p. vansen
      the netherlands
                         dhr p. bransen
## 10 the netherlands
                         dhr p. janssen
## 11 the netherlands
                       mevr 1.
                                rokken
## 12 the netherlands
                       mevr 1.
                                 lokken
## 13 the netherlands mevr 1. mokken
## 14 the netherlands
                       mevr 1.
                                 mokken
## 15 the netherlands mevr 1.
                                 mokken
## 16 the netherlands mevr 1.
                                mokken
## 17 the netherlands mevr 1.
                                 sokken
## 18 the netherlands mevr 1.
                                 wokken
## 19 the netherlands mevr 1.
                                 kokken
## 20 the netherlands mevr 1.
## 21 the netherlands mevr 1.
                                 dokken
## 22 the netherlands mevr 1.
                                 gokken
## 23 the netherlands mevr 1.
                               stokken
## 24 the netherlands mevr 1. rokken
## 25 the netherlands mevr 1. rokken
```

Rename products

It is then straightforward to convert each product code with the defined product names.

```
refine$product_code = as.character(refine$product_code)
refine$product_code[refine$product_code == "p"] = "Smartphone"
```

```
refine$product_code[refine$product_code == "v"] = "TV"
refine$product_code[refine$product_code == "x"] = "Laptop"
refine$product_code[refine$product_code == "q"] = "Tablet"
refine
```

```
##
         company product_code product_number
                                                           address
                                                                      city
## 1
         philips
                    Smartphone
                                             5 Groningensingel 147 arnhem
## 2
         philips
                                            43 Groningensingel 148 arnhem
                    Smartphone
## 3
         philips
                        Laptop
                                             3 Groningensingel 149 arnhem
## 4
         philips
                        Laptop
                                            34 Groningensingel 150 arnhem
## 5
         philips
                        Laptop
                                            12 Groningensingel 151 arnhem
## 6
                                            23 Groningensingel 152 arnhem
         philips
                    Smartphone
## 7
            akzo
                            TV
                                                 Leeuwardenweg 178 arnhem
                                            43
## 8
            akzo
                            TV
                                            12
                                                 Leeuwardenweg 179 arnhem
## 9
            akzo
                                             5
                                                 Leeuwardenweg 180 arnhem
                        Laptop
## 10
            akzo
                    Smartphone
                                            34
                                                 Leeuwardenweg 181 arnhem
## 11
                        Tablet
                                             5
                                                 Leeuwardenweg 182 arnhem
            akzo
## 12
                        Tablet
                                             9
                                                 Leeuwardenweg 183 arnhem
            akzo
                                             8
## 13
            akzo
                                                 Leeuwardenweg 184 arnhem
                        Laptop
## 14
         philips
                    Smartphone
                                            56
                                                 Delfzijlstraat 54 arnhem
                                                 Delfzijlstraat 55 arnhem
## 15
         philips
                            TV
                                            67
## 16
         philips
                            TV
                                            21
                                                 Delfzijlstraat 56 arnhem
## 17 van houten
                                            45
                                                 Delfzijlstraat 57 arnhem
                        Laptop
## 18 van houten
                            TV
                                            56
                                                 Delfzijlstraat 58 arnhem
                            TV
## 19 van houten
                                            65
                                                 Delfzijlstraat 59 arnhem
## 20 van houten
                        Laptop
                                            21
                                                 Delfzijlstraat 60 arnhem
## 21 van houten
                    Smartphone
                                            23
                                                 Delfzijlstraat 61 arnhem
## 22
        unilever
                        Laptop
                                             3
                                                    Jourestraat 23 arnhem
                                             4
## 23
                                                    Jourestraat 24 arnhem
        unilever
                        Tablet
## 24
        unilever
                        Tablet
                                             6
                                                    Jourestraat 25 arnhem
                                                    Jourestraat 26 arnhem
## 25
        unilever
                        Tablet
                                             8
##
              country
                                   name
      the netherlands
                          dhr p. jansen
## 2
      the netherlands
                          dhr p. hansen
## 3
      the netherlands
                          dhr j. Gansen
## 4
      the netherlands
                          dhr p. mansen
## 5
      the netherlands
                         dhr p. fransen
## 6
      the netherlands
                        dhr p. franssen
      the netherlands
                          dhr p. bansen
## 8
      the netherlands
                          dhr p. vansen
      the netherlands
                         dhr p. bransen
## 10 the netherlands
                         dhr p. janssen
## 11 the netherlands
                       mevr 1.
                                rokken
## 12 the netherlands
                        mevr 1.
                                 lokken
## 13 the netherlands
                       mevr 1.
                                 mokken
## 14 the netherlands
                       mevr 1.
                                 mokken
## 15 the netherlands
                       mevr 1.
                                 mokken
## 16 the netherlands
                       mevr 1.
                                 mokken
## 17 the netherlands mevr 1.
                                 sokken
## 18 the netherlands mevr 1.
                                 wokken
## 19 the netherlands mevr 1.
                                 kokken
## 20 the netherlands mevr 1.
                                 Bokken
## 21 the netherlands mevr 1.
                                 dokken
```

```
## 22 the netherlands mevr 1. gokken
## 23 the netherlands mevr 1. stokken
## 24 the netherlands mevr 1. rokken
## 25 the netherlands mevr 1. rokken
```

Create full address column

Use the paste function to combine the address, city and country, separating with ",", into "full_address" column. Append the column after the "country" column.

```
full_address = paste(refine$address, refine$city, refine$country, sep = ", ")
refine = cbind(refine[1:6], full_address, refine[7])
refine
```

```
##
         company product_code product_number
                                                           address
                                                                      city
## 1
                   Smartphone
                                            5 Groningensingel 147 arnhem
         philips
## 2
         philips
                                           43 Groningensingel 148 arnhem
                   Smartphone
## 3
                                            3 Groningensingel 149 arnhem
         philips
                       Laptop
## 4
         philips
                                           34 Groningensingel 150 arnhem
                       Laptop
                       Laptop
## 5
         philips
                                            12 Groningensingel 151 arnhem
## 6
                                           23 Groningensingel 152 arnhem
         philips
                   Smartphone
## 7
                            TV
                                           43
                                                 Leeuwardenweg 178 arnhem
            akzo
## 8
            akzo
                            TV
                                            12
                                                 Leeuwardenweg 179 arnhem
## 9
            akzo
                        Laptop
                                            5
                                                 Leeuwardenweg 180 arnhem
## 10
            akzo
                   Smartphone
                                            34
                                                 Leeuwardenweg 181 arnhem
## 11
            akzo
                        Tablet
                                            5
                                                 Leeuwardenweg 182 arnhem
                                            9
## 12
            akzo
                       Tablet
                                                 Leeuwardenweg 183 arnhem
## 13
                                            8
            akzo
                                                Leeuwardenweg 184 arnhem
                        Laptop
## 14
         philips
                   Smartphone
                                            56
                                                 Delfzijlstraat 54 arnhem
## 15
                            TV
                                           67
                                                 Delfzijlstraat 55 arnhem
         philips
                            TV
                                                 Delfzijlstraat 56 arnhem
## 16
         philips
                                           21
## 17 van houten
                                           45
                                                 Delfzijlstraat 57 arnhem
                        Laptop
## 18 van houten
                                           56
                                                 Delfzijlstraat 58 arnhem
                            TV
## 19 van houten
                            TV
                                           65
                                                 Delfzijlstraat 59 arnhem
## 20 van houten
                       Laptop
                                           21
                                                 Delfzijlstraat 60 arnhem
## 21 van houten
                                           23
                   Smartphone
                                                 Delfzijlstraat 61 arnhem
## 22
        unilever
                       Laptop
                                            3
                                                    Jourestraat 23 arnhem
## 23
                                            4
                                                    Jourestraat 24 arnhem
        unilever
                        Tablet
## 24
        unilever
                        Tablet
                                            6
                                                    Jourestraat 25 arnhem
## 25
        unilever
                       Tablet
                                            8
                                                    Jourestraat 26 arnhem
##
                                                        full_address
              country
      the netherlands Groningensingel 147, arnhem, the netherlands
      the netherlands Groningensingel 148, arnhem, the netherlands
      the netherlands Groningensingel 149, arnhem, the netherlands
## 4
      the netherlands Groningensingel 150, arnhem, the netherlands
## 5
      the netherlands Groningensingel 151, arnhem, the netherlands
## 6
      the netherlands Groningensingel 152, arnhem, the netherlands
      the netherlands
                         Leeuwardenweg 178, arnhem, the netherlands
      the netherlands
                         Leeuwardenweg 179, arnhem, the netherlands
## 8
## 9
      the netherlands
                         Leeuwardenweg 180, arnhem, the netherlands
## 10 the netherlands
                         Leeuwardenweg 181, arnhem, the netherlands
## 11 the netherlands
                         Leeuwardenweg 182, arnhem, the netherlands
## 12 the netherlands
                         Leeuwardenweg 183, arnhem, the netherlands
```

```
## 13 the netherlands
                        Leeuwardenweg 184, arnhem, the netherlands
## 14 the netherlands
                        Delfzijlstraat 54, arnhem, the netherlands
                        Delfzijlstraat 55, arnhem, the netherlands
## 15 the netherlands
                        Delfzijlstraat 56, arnhem, the netherlands
## 16 the netherlands
## 17 the netherlands
                        Delfzijlstraat 57, arnhem, the netherlands
## 18 the netherlands
                        Delfzijlstraat 58, arnhem, the netherlands
## 19 the netherlands
                        Delfzijlstraat 59, arnhem, the netherlands
                        Delfzijlstraat 60, arnhem, the netherlands
## 20 the netherlands
## 21 the netherlands
                        Delfzijlstraat 61, arnhem, the netherlands
## 22 the netherlands
                           Jourestraat 23, arnhem, the netherlands
## 23 the netherlands
                           Jourestraat 24, arnhem, the netherlands
## 24 the netherlands
                           Jourestraat 25, arnhem, the netherlands
## 25 the netherlands
                           Jourestraat 26, arnhem, the netherlands
##
                  name
## 1
         dhr p. jansen
## 2
         dhr p. hansen
## 3
         dhr j. Gansen
## 4
         dhr p. mansen
## 5
        dhr p. fransen
## 6
       dhr p. franssen
## 7
        dhr p. bansen
## 8
        dhr p. vansen
## 9
       dhr p. bransen
       dhr p. janssen
## 10
## 11
      mevr 1. rokken
## 12
      mevr 1. lokken
## 13
      mevr 1. mokken
## 14
      mevr 1. mokken
## 15
      mevr 1. mokken
## 16
      mevr 1. mokken
## 17
      mevr 1. sokken
## 18
      mevr 1.
               wokken
## 19
      mevr 1. kokken
## 20
      mevr 1. Bokken
## 21
      mevr 1. dokken
## 22 mevr 1. gokken
## 23 mevr 1. stokken
## 24 mevr 1. rokken
## 25 mevr 1. rokken
```

Create binary variables columns for companies

Create binary variable columns of "company_philips", "company_akzo", "company_van_houten", and "company_unilever" respectively for "philips", "akzo", "van houten", and "unilever".

```
refine$company_philips = as.integer(refine$company == "philips")
refine$company_akzo = as.integer(refine$company == "akzo")
refine$company_van_houten = as.integer(refine$company == "van houten")
refine$company_unilever = as.integer(refine$company == "unilever")
refine
```

```
## company product_code product_number address city
## 1 philips Smartphone 5 Groningensingel 147 arnhem
```

```
## 2
         philips
                   Smartphone
                                            43 Groningensingel 148 arnhem
## 3
                        Laptop
         philips
                                            3 Groningensingel 149 arnhem
         philips
                                            34 Groningensingel 150 arnhem
## 4
                        Laptop
## 5
                                            12 Groningensingel 151 arnhem
         philips
                        Laptop
## 6
         philips
                   Smartphone
                                              Groningensingel 152 arnhem
## 7
                                                 Leeuwardenweg 178 arnhem
            akzo
                            TV
                                            43
## 8
            akzo
                            TV
                                            12
                                                 Leeuwardenweg 179 arnhem
## 9
            akzo
                        Laptop
                                             5
                                                 Leeuwardenweg 180 arnhem
## 10
            akzo
                   Smartphone
                                            34
                                                 Leeuwardenweg 181 arnhem
                                             5
## 11
            akzo
                        Tablet
                                                 Leeuwardenweg 182 arnhem
  12
            akzo
                        Tablet
                                             9
                                                 Leeuwardenweg 183 arnhem
  13
                                             8
                                                 Leeuwardenweg 184 arnhem
##
            akzo
                        Laptop
##
   14
                   Smartphone
                                            56
                                                 Delfzijlstraat 54 arnhem
         philips
                                                 Delfzijlstraat 55 arnhem
## 15
         philips
                            TV
                                            67
                            TV
                                            21
## 16
         philips
                                                 Delfzijlstraat 56 arnhem
## 17 van houten
                                            45
                                                 Delfzijlstraat 57 arnhem
                        Laptop
                                            56
## 18 van houten
                            TV
                                                 Delfzijlstraat 58 arnhem
## 19 van houten
                            TV
                                            65
                                                 Delfzijlstraat 59 arnhem
## 20 van houten
                                            21
                                                 Delfzijlstraat 60 arnhem
                        Laptop
## 21 van houten
                   Smartphone
                                            23
                                                 Delfzijlstraat 61 arnhem
##
  22
        unilever
                        Laptop
                                            3
                                                    Jourestraat 23 arnhem
## 23
                                             4
                                                    Jourestraat 24 arnhem
        unilever
                        Tablet
## 24
                                             6
                                                    Jourestraat 25 arnhem
        unilever
                        Tablet
                                             8
                                                    Jourestraat 26 arnhem
## 25
        unilever
                        Tablet
##
              country
                                                        full address
  1
      the netherlands Groningensingel 147, arnhem, the netherlands
      the netherlands Groningensingel 148, arnhem, the netherlands
      the netherlands Groningensingel 149, arnhem, the netherlands
##
   3
## 4
      the netherlands Groningensingel 150, arnhem, the netherlands
      the netherlands Groningensingel 151, arnhem, the netherlands
## 6
      the netherlands Groningensingel 152, arnhem, the netherlands
## 7
      the netherlands
                         Leeuwardenweg 178, arnhem, the netherlands
      the netherlands
                         Leeuwardenweg 179, arnhem, the netherlands
                         Leeuwardenweg 180, arnhem, the netherlands
      the netherlands
## 10 the netherlands
                         Leeuwardenweg 181, arnhem, the netherlands
## 11 the netherlands
                         Leeuwardenweg 182, arnhem, the netherlands
## 12 the netherlands
                         Leeuwardenweg 183, arnhem, the netherlands
## 13 the netherlands
                         Leeuwardenweg 184, arnhem, the netherlands
## 14 the netherlands
                         Delfzijlstraat 54, arnhem, the netherlands
## 15 the netherlands
                         Delfzijlstraat 55, arnhem, the netherlands
## 16 the netherlands
                         Delfzijlstraat 56, arnhem, the netherlands
## 17 the netherlands
                         Delfzijlstraat 57, arnhem, the netherlands
## 18 the netherlands
                         Delfzijlstraat 58, arnhem, the netherlands
## 19 the netherlands
                         Delfzijlstraat 59, arnhem, the netherlands
## 20 the netherlands
                         Delfzijlstraat 60, arnhem, the netherlands
                         Delfzijlstraat 61, arnhem, the netherlands
## 21 the netherlands
## 22 the netherlands
                            Jourestraat 23, arnhem, the netherlands
## 23 the netherlands
                            Jourestraat 24, arnhem, the netherlands
## 24 the netherlands
                            Jourestraat 25, arnhem, the netherlands
##
  25 the netherlands
                            Jourestraat 26, arnhem, the netherlands
##
                  name company_philips company_akzo company_van_houten
## 1
         dhr p. jansen
                                      1
                                                    0
## 2
         dhr p. hansen
                                      1
                                                    0
                                                                        0
## 3
         dhr j. Gansen
                                      1
                                                    0
                                                                        0
```

```
## 4
         dhr p. mansen
                                         1
                                                       0
                                                                             0
        dhr p. fransen
## 5
                                         1
                                                       0
                                                                             0
## 6
       dhr p. franssen
                                         1
                                                       0
                                                                             0
                                         0
                                                                             0
## 7
         dhr p. bansen
                                                       1
## 8
         dhr p. vansen
                                         0
                                                       1
                                                                             0
## 9
                                         0
                                                                             0
        dhr p. bransen
                                                       1
## 10
        dhr p. janssen
                                         0
                                                       1
                                                                             0
                                         0
## 11
       mevr 1.
                rokken
                                                       1
                                                                             0
## 12
       mevr 1.
                 lokken
                                         0
                                                       1
                                                                             0
## 13
                                         0
                                                                             0
       mevr 1.
                 mokken
                                                       1
## 14
       mevr 1.
                 mokken
                                         1
                                                       0
                                                                             0
                                                       0
                                                                             0
## 15
       mevr 1.
                 mokken
                                         1
                                                       0
                                                                             0
## 16
       mevr 1.
                 mokken
                                         1
       mevr 1.
                                         0
                                                       0
## 17
                 sokken
                                                                             1
## 18
       mevr 1.
                 wokken
                                         0
                                                       0
                                                                             1
## 19
       mevr 1.
                 kokken
                                         0
                                                       0
                                                                             1
## 20
                                         0
                                                       0
       mevr 1.
                 Bokken
                                                                             1
## 21
       mevr 1.
                 dokken
                                         0
                                                       0
                                                                             1
## 22
       mevr 1.
                                         0
                                                       0
                                                                             0
                 gokken
## 23 mevr 1.
                stokken
                                         0
                                                       0
                                                                             0
## 24
       mevr 1. rokken
                                         0
                                                       0
                                                                             0
## 25
       mevr 1. rokken
                                                       0
                                                                             0
##
      company_unilever
## 1
                       0
## 2
                       0
## 3
                       0
## 4
                       0
## 5
                       0
## 6
                       0
## 7
                       0
## 8
                       0
## 9
                       0
## 10
                       0
## 11
                       0
## 12
                       0
## 13
                       0
## 14
                       0
## 15
                       0
## 16
                       0
## 17
                       0
## 18
                       0
## 19
                       0
## 20
                       0
## 21
                       0
## 22
                       1
## 23
                       1
## 24
                       1
## 25
                       1
```

Create binary variables columns for product categories

Create binary variable columns of "product_smartphone", "product_tv", "product_laptop", and "product_tablet" respectively for "Smartphone", "TV", "Laptop", and "Tablet".

```
refine$product_smartphone = as.integer(refine$product_code == "Smartphone")
refine$product_tv = as.integer(refine$product_code == "TV")
refine$product_laptop = as.integer(refine$product_code == "Laptop")
refine$product_tablet = as.integer(refine$product_code == "Tablet")
refine
```

```
##
         company product_code product_number
                                                           address
                                                                     city
## 1
         philips
                   Smartphone
                                            5 Groningensingel 147 arnhem
                                           43 Groningensingel 148 arnhem
## 2
         philips
                   Smartphone
## 3
         philips
                        Laptop
                                            3 Groningensingel 149 arnhem
## 4
         philips
                        Laptop
                                           34 Groningensingel 150 arnhem
                                           12 Groningensingel 151 arnhem
## 5
         philips
                       Laptop
## 6
                                           23 Groningensingel 152 arnhem
         philips
                   Smartphone
## 7
            akzo
                                                 Leeuwardenweg 178 arnhem
## 8
                            TV
                                           12
                                                Leeuwardenweg 179 arnhem
            akzo
## 9
            akzo
                        Laptop
                                            5
                                                Leeuwardenweg 180 arnhem
## 10
                   Smartphone
                                           34
                                                Leeuwardenweg 181 arnhem
            akzo
## 11
                        Tablet
                                                Leeuwardenweg 182 arnhem
            akzo
                                            5
## 12
                        Tablet
                                            9
                                                Leeuwardenweg 183 arnhem
            akzo
## 13
                                            8
                                                Leeuwardenweg 184 arnhem
            akzo
                        Laptop
## 14
         philips
                   Smartphone
                                           56
                                                Delfzijlstraat 54 arnhem
## 15
         philips
                            TV
                                           67
                                                Delfzijlstraat 55 arnhem
                            TV
                                           21
                                                Delfzijlstraat 56 arnhem
## 16
         philips
## 17 van houten
                                           45
                                                Delfzijlstraat 57 arnhem
                       Laptop
                                           56
## 18 van houten
                            TV
                                                Delfzijlstraat 58 arnhem
## 19 van houten
                            TV
                                           65
                                                Delfzijlstraat 59 arnhem
## 20 van houten
                       Laptop
                                           21
                                                Delfzijlstraat 60 arnhem
## 21 van houten
                   Smartphone
                                           23
                                                Delfzijlstraat 61 arnhem
## 22
                                            3
        unilever
                        Laptop
                                                    Jourestraat 23 arnhem
## 23
        unilever
                        Tablet
                                            4
                                                    Jourestraat 24 arnhem
                                                    Jourestraat 25 arnhem
## 24
        unilever
                       Tablet
                                            6
                                                    Jourestraat 26 arnhem
## 25
        unilever
                        Tablet
                                            8
##
                                                        full address
              country
      the netherlands Groningensingel 147, arnhem, the netherlands
## 1
      the netherlands Groningensingel 148, arnhem, the netherlands
      the netherlands Groningensingel 149, arnhem, the netherlands
## 3
      the netherlands Groningensingel 150, arnhem, the netherlands
## 5
      the netherlands Groningensingel 151, arnhem, the netherlands
      the netherlands Groningensingel 152, arnhem, the netherlands
## 7
      the netherlands
                        Leeuwardenweg 178, arnhem, the netherlands
      the netherlands
                        Leeuwardenweg 179, arnhem, the netherlands
## 9
      the netherlands
                        Leeuwardenweg 180, arnhem, the netherlands
## 10 the netherlands
                        Leeuwardenweg 181, arnhem, the netherlands
## 11 the netherlands
                        Leeuwardenweg 182, arnhem, the netherlands
## 12 the netherlands
                        Leeuwardenweg 183, arnhem, the netherlands
                        Leeuwardenweg 184, arnhem, the netherlands
## 13 the netherlands
## 14 the netherlands
                        Delfzijlstraat 54, arnhem, the netherlands
## 15 the netherlands
                        Delfzijlstraat 55, arnhem, the netherlands
## 16 the netherlands
                        Delfzijlstraat 56, arnhem, the netherlands
## 17 the netherlands
                        Delfzijlstraat 57, arnhem, the netherlands
## 18 the netherlands
                        Delfzijlstraat 58, arnhem, the netherlands
## 19 the netherlands
                         Delfzijlstraat 59, arnhem, the netherlands
## 20 the netherlands
                        Delfzijlstraat 60, arnhem, the netherlands
```

```
Delfzijlstraat 61, arnhem, the netherlands
## 21 the netherlands
## 22 the netherlands
                              Jourestraat 23, arnhem, the netherlands
## 23 the netherlands
                              Jourestraat 24, arnhem, the netherlands
## 24 the netherlands
                              Jourestraat 25, arnhem, the netherlands
## 25 the netherlands
                              Jourestraat 26, arnhem, the netherlands
##
                   name company_philips company_akzo company_van_houten
          dhr p. jansen
## 1
## 2
                                                       0
                                                                            0
          dhr p. hansen
                                         1
## 3
          dhr j. Gansen
                                         1
                                                       0
                                                                            0
## 4
                                                       0
                                                                            0
          dhr p. mansen
                                         1
## 5
        dhr p. fransen
                                         1
                                                       0
                                                                            0
                                                       0
## 6
                                                                            0
       dhr p. franssen
                                         1
## 7
         dhr p. bansen
                                         0
                                                       1
                                                                            0
## 8
                                         0
                                                                            0
          dhr p. vansen
                                                       1
## 9
        dhr p. bransen
                                         0
                                                       1
                                                                            0
## 10
        dhr p. janssen
                                         0
                                                       1
                                                                            0
## 11
                                         0
                                                       1
                                                                            0
       mevr 1. rokken
                                         0
## 12
       mevr 1.
                 lokken
                                                       1
## 13
       mevr 1.
                 mokken
                                         0
                                                       1
                                                                            0
## 14
       mevr 1.
                 mokken
                                         1
                                                       0
                                                                            0
       mevr 1.
## 15
                 mokken
                                         1
                                                       0
                                                                            0
       mevr 1.
                 mokken
                                                       0
       mevr 1.
## 17
                 sokken
                                         0
                                                       0
                                                                            1
## 18
       mevr 1.
                 wokken
                                         0
                                                       0
                                                                            1
## 19
       mevr 1.
                 kokken
                                         0
                                                       0
                                                                            1
## 20
       mevr 1.
                 Bokken
                                         0
                                                       0
                                                                            1
## 21
       mevr 1.
                 dokken
                                         0
                                                       0
                                                                            1
## 22
       mevr 1.
                                         0
                                                       0
                                                                            0
                 gokken
## 23 mevr 1.
                                         0
                                                       0
                                                                            0
                stokken
                                         0
                                                       0
## 24
       mevr 1. rokken
                                                                            0
## 25
       mevr 1. rokken
                                         0
                                                       0
##
       company_unilever product_smartphone product_tv product_laptop
## 1
                       0
                                            1
                                                        0
## 2
                       0
                                                        0
                                                                         0
                                            1
## 3
                                            0
                       0
                                                        0
                                                                         1
## 4
                       0
                                            0
                                                        0
                                                                         1
## 5
                       0
                                            0
                                                        0
## 6
                       0
                                            1
                                                        0
                                                                         0
## 7
                       0
                                            0
                                                        1
                                                                         0
## 8
                                            0
                       0
                                                                         0
                                                        1
## 9
                       0
                                            0
                                                        0
                                                                         1
## 10
                       0
                                            1
                                                        0
                                                                         0
## 11
                       0
                                            0
                                                        0
                                                                         0
## 12
                       0
                                            0
                                                        0
                                                                         0
## 13
                       0
                                            0
                                                        0
                                                                         1
## 14
                       0
                                            1
                                                        0
                                                                         0
                                            0
## 15
                       0
                                                        1
                                                                         0
## 16
                       0
                                            0
                                                        1
                                                                         0
## 17
                       0
                                            0
                                                        0
                                                                         1
                       0
                                            0
## 18
                                                        1
                                                                         0
## 19
                       0
                                            0
                                                        1
                                                                         0
## 20
                       0
                                            0
                                                        0
                                                                         1
## 21
                       0
                                            1
                                                        0
                                                                         0
                                            0
## 22
                       1
                                                        0
```

```
## 23
                      1
                                           0
                                                                       0
## 24
                      1
                                           0
                                                       0
                                                                       0
                                                       0
                                           0
## 25
##
      product_tablet
## 1
## 2
                    0
## 3
                    0
## 4
                    0
## 5
                    0
                    0
## 6
                    0
## 7
## 8
                    0
## 9
                    0
                    0
## 10
## 11
                    1
## 12
                    1
## 13
                    0
                    0
## 14
## 15
                    0
                    0
## 16
## 17
                    0
## 18
## 19
                    0
## 20
                    0
## 21
## 22
                    0
## 23
                    1
## 24
                    1
## 25
                    1
```

Write to clean file

Finally, write the new dataset into the clean csv file.

```
write.csv(refine, file = "refine_clean.csv")
```