

Untitled

SIMON

04/08/2021

OPTIMISEZ LA GESTION DES DONNEES D'UNE BOUTIQUE AVEC R OU PYTHON

LIBRARIES

```
library(gdata)
```

```
## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location of a
## gdata: valid perl intrprerer.
## gdata:
## gdata: (To avoid display of this message in the future, please ensure
## gdata: perl is installed and available on the executable search path.)
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.
```

```
##
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
```

```
##
```

```
## gdata: Run the function 'installXLSXsupport()'  
## gdata: to automatically download and install the perl  
## gdata: libraries needed to support Excel XLS and XLSX formats.
```

```
##  
## Attaching package: 'gdata'
```

```
## The following object is masked from 'package:stats':  
##  
##     nobs
```

```
## The following object is masked from 'package:utils':  
##  
##     object.size
```

```
## The following object is masked from 'package:base':  
##  
##    startsWith
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(dplyr, warn.conflicts = FALSE)  
library(readxl)
```

IMPORTATION DES DONNEES

```
setwd("C:/Users/stars/onedrive")
mydir = "Documents"
myfiles = list.files(path= mydir, pattern="*.xl", full.names=TRUE)
myfiles
```

```
## [1] "Documents/~$LOGEMENTPQ.xlsx"           "Documents/~$MUTATIONPQ.xlsx"
## [3] "Documents/ADRESSE TEST EXCEL.xlsx"      "Documents/BASE_OFF_P6.xlsx"
## [5] "Documents/basefin1.xlsx"                 "Documents/DONNEE-PANEL-FINALE.xlsx"
## [7] "Documents/erp (1).xlsx"                  "Documents/liaison (1).xlsx"
## [9] "Documents/liaison.xlsx"                 "Documents/P3_SAGNA_BASEDONNEES.xlsx"
## [11] "Documents/P3_SAGNA_MUTATION.xlsx"        "Documents/rapprochement.xlsx"
## [13] "Documents/web (1).xlsx"
```

```
liaison <- read_excel("Documents/liaison (1).xlsx")
erp<- read_excel("Documents/erp (1).xlsx")
web<- read_excel("Documents/web (1).xlsx")
```

FONCTION

```
Cle <- function(df, colonnes){if(nrow(df) == nrow(unique(df[colonnes]))){
  print("La clé n'est pas présente plusieurs fois dans le dataframe.")
  print("Elle peut donc être utilisée comme clé primaire.")
}else {
  print("La clé est présente plusieurs fois dans le dataframe.")
  print("Elle ne peut donc pas être utilisée comme clé primaire.")
}

print(sprintf("Le dataframe est de la forme : (%s) (lignes, colonnes)", paste0(dim(df), collapse=",")))}
```

PREPARATION DU FICHIER LIAISON

```
print(dim(liaison))
```

```
## [1] 825 2
```

```
print(str(liaison))
```

```
## # A tibble: 825 x 2 (S3:tbl_df/tbl/data.frame)
##   $ product_id: num [1:825] 3847 3849 3850 4032 4039 ...
##   $ id_web    : chr [1:825] "15298" "15296" "15300" "19814" ...
## # ... NULL
```

```
head(liaison)
```

```
## # A tibble: 6 x 2
##   product_id id_web
##       <dbl> <chr>
## 1      3847 15298
## 2      3849 15296
## 3      3850 15300
## 4      4032 19814
## 5      4039 19815
## 6      4040 15303
```

```
Cle(liaison, c('product_id'))
```

```
## [1] "La clé n'est pas présente plusieurs fois dans le dataframe."
## [1] "Elle peut donc être utilisée comme clé primaire."
## [1] "Le dataframe est de la forme : (825,2) (lignes, colonnes)"
```

PREPARATION DU FICHIER ERP

```
print(dim(erp))
```

```
## [1] 825 5
```

```
print(str(erp))
```

```
## # tibble [825 x 5] (S3:tbl_df/tbl/data.frame)
## $ product_id : num [1:825] 3847 3849 3850 4032 4039 ...
## $ onsale_web : num [1:825] 1 1 1 1 1 1 1 1 1 1 ...
## $ price      : num [1:825] 24.2 34.3 20.8 14.1 46 34.3 32.7 31.2 60 42.6 ...
## $ stock_quantity: num [1:825] 0 0 0 0 0 0 15 34 12 66 ...
## $ stock_status : chr [1:825] "outofstock" "outofstock" "outofstock" "outofstock" ...
## NULL
```

```
head(erp)
```

```
## # A tibble: 6 x 5
##   product_id onsale_web price stock_quantity stock_status
##       <dbl>        <dbl>    <dbl>        <dbl>     <chr>
## 1     3847         1    24.2          0 outofstock
## 2     3849         1    34.3          0 outofstock
## 3     3850         1    20.8          0 outofstock
## 4     4032         1    14.1          0 outofstock
## 5     4039         1     46           0 outofstock
## 6     4040         1    34.3          0 outofstock
```

```
Cle(erp, c('product_id'))
```

```
## [1] "La clé n'est pas présente plusieurs fois dans le dataframe."
## [1] "Elle peut donc être utilisée comme clé primaire."
## [1] "Le dataframe est de la forme : (825,5) (lignes, colonnes)"
```

PREPARATION DU FICHIER WEB

```
print(dim(web))
```

```
## [1] 1513    28
```

```
print(str(web))
```

```

## tibble [1,513 x 28] (S3: tbl_df/tbl/data.frame)
## $ sku : chr [1:1513] "bon-cadeau-25-euros" "15298" "15296" "15300" ...
## $ virtual : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ downloadable : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ rating_count : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ average_rating : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ total_sales : num [1:1513] 10 6 0 0 3 0 0 0 7 3 ...
## $ tax_status : chr [1:1513] "taxable" "taxable" "taxable" "taxable" ...
## $ tax_class : logi [1:1513] NA NA NA NA NA ...
## $ post_author : num [1:1513] 1 2 2 2 2 2 2 2 2 ...
## $ post_date : POSIXct[1:1513], format: "2018-06-01 13:53:46" "2018-02-08 12:58:52" ...
## $ post_date_gmt : POSIXct[1:1513], format: "2018-06-01 11:53:46" "2018-02-08 11:58:52" ...
## $ post_content : logi [1:1513] NA NA NA NA NA ...
## $ post_title : chr [1:1513] "Bon cadeau de 25\200" "Pierre Jean Villa Saint-Joseph Préface 2018" "Pierre Jean Vil
la Saint-Joseph Rouge Tildé 2017" "Pierre Jean Villa Crozes-Hermitage Accroche Coeur 2018" ...
## $ post_excerpt : chr [1:1513] "<span style=\"color: #a85253;\"><strong>Parlons peu, parlons vin.</strong>\r\n<stron
g>Parlons cadeau !</strong>"| __truncated__ "Il exhale un nez subtil, racé avec du poivre fin et de la tourbe. La bouche est un
e caresse grâce à des tanins "| __truncated__ "Les vieilles vignes lui apportent une rare profondeur. L'attaque affiche de l'él
équence. La bouche est portée pa"| __truncated__ "Dentelle de fruit de jeunes syrah, aux tanins légers et épics. Hyper digestive.
Un délice." ...
## $ post_status : chr [1:1513] "publish" "publish" "publish" "publish" ...
## $ comment_status : chr [1:1513] "closed" "closed" "closed" "closed" ...
## $ ping_status : chr [1:1513] "closed" "closed" "closed" "closed" ...
## $ post_password : logi [1:1513] NA NA NA NA NA ...
## $ post_name : chr [1:1513] "bon-cadeau-de-25-euros" "pierre-jean-villa-saint-joseph-preface-2018" "pierre-jean-v
illa-saint-joseph-tilde-2017" "pierre-jean-villa-croze-hermitage-accroche-coeur-2018" ...
## $ post_modified : POSIXct[1:1513], format: "2018-06-01 14:13:57" "2019-12-30 09:30:29" ...
## $ post_modified_gmt : POSIXct[1:1513], format: "2018-06-01 12:13:57" "2019-12-30 08:30:29" ...
## $ post_content_filtered: logi [1:1513] NA NA NA NA NA ...
## $ post_parent : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ guid : chr [1:1513] "https://www.bottle-neck.fr/?post_type=product&p=4954" "https://www.bottle-neck.
fr/?post_type=product&p=3847" "https://www.bottle-neck.fr/?post_type=product&p=3849" "https://www.bottle-neck.fr/?pos
t_type=product&p=3850" ...
## $ menu_order : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## $ post_type : chr [1:1513] "product" "product" "product" "product" ...
## $ post_mime_type : chr [1:1513] NA NA NA NA ...
## $ comment_count : num [1:1513] 0 0 0 0 0 0 0 0 0 ...
## NULL

```

```
head(web)
```

```
## # A tibble: 6 x 28
##   sku      virtual downloadable rating_count average_rating total_sales tax_status
##   <chr>     <dbl>        <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 bon-c~     0            0            0            0            10 taxable
## 2 15298      0            0            0            0             6 taxable
## 3 15296      0            0            0            0             0 taxable
## 4 15300      0            0            0            0             0 taxable
## 5 19814      0            0            0            0             3 taxable
## 6 19815      0            0            0            0             0 taxable
## # ... with 21 more variables: tax_class <lgl>, post_author <dbl>,
## #   post_date <dttm>, post_date_gmt <dttm>, post_content <lgl>,
## #   post_title <chr>, post_excerpt <chr>, post_status <chr>,
## #   comment_status <chr>, ping_status <chr>, post_password <lgl>,
## #   post_name <chr>, post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

```
Cle(web, c('sku'))
```

```
## [1] "La clé est présente plusieurs fois dans le dataframe."
## [1] "Elle ne peut donc pas être utilisée comme clé primaire."
## [1] "Le dataframe est de la forme : (1513,28) (lignes, colonnes)"
```

Après observation de la base de données web on peut remarquer que nous avons la présence de valeurs nulles. il serait important de les identifier.

Recherche de valeurs nulles

```
nul= web[is.na(web$sku),]
print(dim(nul))
```

```
## [1] 85 28
```

```
print(str(nul))
```

```
## # tibble [85 x 28] (S3: tbl_df/tbl/data.frame)
## # $ sku : chr [1:85] NA NA NA NA ...
## # $ virtual : num [1:85] 0 0 0 0 0 0 0 0 0 ...
## # $ downloadable : num [1:85] 0 0 0 0 0 0 0 0 0 ...
## # $ rating_count : num [1:85] 0 0 0 0 0 0 0 0 0 ...
## # $ average_rating : num [1:85] NA NA NA NA NA NA NA NA ...
## # $ total_sales : num [1:85] NA NA NA NA NA NA NA NA ...
## # $ tax_status : chr [1:85] NA NA NA NA ...
## # $ tax_class : logi [1:85] NA NA NA NA NA ...
## # $ post_author : num [1:85] NA NA NA NA NA NA NA NA ...
## # $ post_date : POSIXct[1:85], format: NA NA ...
## # $ post_date_gmt : POSIXct[1:85], format: NA NA ...
## # $ post_content : logi [1:85] NA NA NA NA NA ...
## # $ post_title : chr [1:85] NA NA NA NA ...
## # $ post_excerpt : chr [1:85] NA NA NA NA ...
## # $ post_status : chr [1:85] NA NA NA NA ...
## # $ comment_status : chr [1:85] NA NA NA NA ...
## # $ ping_status : chr [1:85] NA NA NA NA ...
## # $ post_password : logi [1:85] NA NA NA NA NA ...
## # $ post_name : chr [1:85] NA NA NA NA ...
## # $ post_modified : POSIXct[1:85], format: NA NA ...
## # $ post_modified_gmt : POSIXct[1:85], format: NA NA ...
## # $ post_content_filtered: logi [1:85] NA NA NA NA NA ...
## # $ post_parent : num [1:85] NA NA NA NA NA NA NA NA ...
## # $ guid : chr [1:85] NA NA NA NA ...
## # $ menu_order : num [1:85] NA NA NA NA NA NA NA NA ...
## # $ post_type : chr [1:85] NA NA NA NA ...
## # $ post_mime_type : chr [1:85] NA NA NA NA ...
## # $ comment_count : num [1:85] NA NA NA NA NA NA NA NA ...
## # NULL
```

```
head(nul)
```

```
## # A tibble: 6 x 28
##   sku    virtual downloadable rating_count average_rating total_sales tax_status
##   <chr>    <dbl>        <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 <NA>      0            0            0           NA           NA <NA>
## 2 <NA>      0            0            0           NA           NA <NA>
## 3 <NA>      0            0            0           NA           NA <NA>
## 4 <NA>      0            0            0           NA           NA <NA>
## 5 <NA>      0            0            0           NA           NA <NA>
## 6 <NA>      0            0            0           NA           NA <NA>
## # ... with 21 more variables: tax_class <lgl>, post_author <dbl>,
## #   post_date <dttm>, post_date_gmt <dttm>, post_content <lgl>,
## #   post_title <chr>, post_excerpt <chr>, post_status <chr>,
## #   comment_status <chr>, ping_status <chr>, post_password <lgl>,
## #   post_name <chr>, post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

On constate sur cette base 'nul' nous avons deux lignes qui présentent des valeurs non nulles. Dans un premier temps, je vais les identifier à partir de la colonnes post_names qui me servira de base.

Présentation des valeurs non nulles

```
valeurs_non_nulles = nul[!is.na(nul$post_name),]
print(dim(valeurs_non_nulles))
```

```
## [1] 2 28
```

```
head(valeurs_non_nulles)
```

```
## # A tibble: 2 x 28
##   sku    virtual downloadable rating_count average_rating total_sales tax_status
##   <chr>    <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 <NA>      0            0            0            0    taxable
## 2 <NA>      0            0            0            0    taxable
## # ... with 21 more variables: tax_class <lgl>, post_author <dbl>,
## #   post_date <dttm>, post_date_gmt <dttm>, post_content <lgl>,
## #   post_title <chr>, post_excerpt <chr>, post_status <chr>,
## #   comment_status <chr>, ping_status <chr>, post_password <lgl>,
## #   post_name <chr>, post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

ces deux lignes contiennent de véritables produits. Il serait nécessaire de mener une investigation afin de résoudre ce problème.

suppression des lignes NA

```
non_null = web[!is.na(web$sku),]
print(dim(non_null))
```

```
## [1] 1428 28
```

```
print(str(non_null))
```

```

## tibble [1,428 x 28] (S3: tbl_df/tbl/data.frame)
## $ sku : chr [1:1428] "bon-cadeau-25-euros" "15298" "15296" "15300" ...
## $ virtual : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ downloadable : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ rating_count : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ average_rating : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ total_sales : num [1:1428] 10 6 0 0 3 0 0 0 7 3 ...
## $ tax_status : chr [1:1428] "taxable" "taxable" "taxable" "taxable" ...
## $ tax_class : logi [1:1428] NA NA NA NA NA ...
## $ post_author : num [1:1428] 1 2 2 2 2 2 2 2 2 ...
## $ post_date : POSIXct[1:1428], format: "2018-06-01 13:53:46" "2018-02-08 12:58:52" ...
## $ post_date_gmt : POSIXct[1:1428], format: "2018-06-01 11:53:46" "2018-02-08 11:58:52" ...
## $ post_content : logi [1:1428] NA NA NA NA NA ...
## $ post_title : chr [1:1428] "Bon cadeau de 25\200" "Pierre Jean Villa Saint-Joseph Préface 2018" "Pierre Jean Vil
la Saint-Joseph Rouge Tildé 2017" "Pierre Jean Villa Crozes-Hermitage Accroche Coeur 2018" ...
## $ post_excerpt : chr [1:1428] "<span style=\"color: #a85253;\"><strong>Parlons peu, parlons vin.</strong>\r\n<stron
g>Parlons cadeau !</strong>"| __truncated__ "Il exhale un nez subtil, racé avec du poivre fin et de la tourbe. La bouche est un
e caresse grâce à des tanins "| __truncated__ "Les vieilles vignes lui apportent une rare profondeur. L'attaque affiche de l'él
équence. La bouche est portée pa"| __truncated__ "Dentelle de fruit de jeunes syrah, aux tanins légers et épics. Hyper digestive.
Un délice." ...
## $ post_status : chr [1:1428] "publish" "publish" "publish" "publish" ...
## $ comment_status : chr [1:1428] "closed" "closed" "closed" "closed" ...
## $ ping_status : chr [1:1428] "closed" "closed" "closed" "closed" ...
## $ post_password : logi [1:1428] NA NA NA NA NA ...
## $ post_name : chr [1:1428] "bon-cadeau-de-25-euros" "pierre-jean-villa-saint-joseph-preface-2018" "pierre-jean-v
illa-saint-joseph-tilde-2017" "pierre-jean-villa-croze-hermitage-accroche-coeur-2018" ...
## $ post_modified : POSIXct[1:1428], format: "2018-06-01 14:13:57" "2019-12-30 09:30:29" ...
## $ post_modified_gmt : POSIXct[1:1428], format: "2018-06-01 12:13:57" "2019-12-30 08:30:29" ...
## $ post_content_filtered: logi [1:1428] NA NA NA NA NA ...
## $ post_parent : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ guid : chr [1:1428] "https://www.bottle-neck.fr/?post_type=product&p=4954" "https://www.bottle-neck.
fr/?post_type=product&p=3847" "https://www.bottle-neck.fr/?post_type=product&p=3849" "https://www.bottle-neck.fr/?pos
t_type=product&p=3850" ...
## $ menu_order : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## $ post_type : chr [1:1428] "product" "product" "product" "product" ...
## $ post_mime_type : chr [1:1428] NA NA NA NA ...
## $ comment_count : num [1:1428] 0 0 0 0 0 0 0 0 0 ...
## NULL

```

```
head(non_null)
```

```
## # A tibble: 6 x 28
##   sku      virtual downloadable rating_count average_rating total_sales tax_status
##   <chr>     <dbl>        <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 bon-c~     0           0           0           0           10 taxable
## 2 15298      0           0           0           0           6 taxable
## 3 15296      0           0           0           0           0 taxable
## 4 15300      0           0           0           0           0 taxable
## 5 19814      0           0           0           0           3 taxable
## 6 19815      0           0           0           0           0 taxable
## # ... with 21 more variables: tax_class <lgl>, post_author <dbl>,
## #   post_date <dttm>, post_date_gmt <dttm>, post_content <lgl>,
## #   post_title <chr>, post_excerpt <chr>, post_status <chr>,
## #   comment_status <chr>, ping_status <chr>, post_password <lgl>,
## #   post_name <chr>, post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

```
Cle(non_null, c('sku'))
```

```
## [1] "La clé est présente plusieurs fois dans le dataframe."
## [1] "Elle ne peut donc pas être utilisée comme clé primaire."
## [1] "Le dataframe est de la forme : (1428,28) (lignes, colonnes)"
```

la colonnes post_types a deux types : product et attachment.On peut supprimer l'un des deux.J'ai choisi les attachments qui sont des liens vers les visuels d'illustration du produit. Ces derniers n'ont pas une grande importance.

Conservation du type product

```
new_web = non_null[non_null$post_type=='product',]
print(dim(new_web))
```

```
## [1] 714 28
```

```
print(str(new_web))
```

```

## # tibble [714 x 28] (S3: tbl_df/tbl/data.frame)
## # $ sku : chr [1:714] "bon-cadeau-25-euros" "15298" "15296" "15300" ...
## # $ virtual : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ downloadable : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ rating_count : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ average_rating : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ total_sales : num [1:714] 10 6 0 0 3 0 0 0 7 3 ...
## # $ tax_status : chr [1:714] "taxable" "taxable" "taxable" "taxable" ...
## # $ tax_class : logi [1:714] NA NA NA NA NA ...
## # $ post_author : num [1:714] 1 2 2 2 2 2 2 2 2 ...
## # $ post_date : POSIXct[1:714], format: "2018-06-01 13:53:46" "2018-02-08 12:58:52" ...
## # $ post_date_gmt : POSIXct[1:714], format: "2018-06-01 11:53:46" "2018-02-08 11:58:52" ...
## # $ post_content : logi [1:714] NA NA NA NA NA ...
## # $ post_title : chr [1:714] "Bon cadeau de 25\200" "Pierre Jean Villa Saint-Joseph Préface 2018" "Pierre Jean Vill
a Saint-Joseph Rouge Tildé 2017" "Pierre Jean Villa Crozes-Hermitage Accroche Coeur 2018" ...
## # $ post_excerpt : chr [1:714] "<span style=\"color: #a85253;\"><strong>Parlons peu, parlons vin.</strong>\r\n<strong>Parlons cadeau !</strong>"| __truncated__ "Il exhale un nez subtil, racé avec du poivre fin et de la tourbe. La bouche est une
caresse grâce à des tanins "| __truncated__ "Les vieilles vignes lui apportent une rare profondeur. L'attaque affiche de l'élég
ance. La bouche est portée pa"| __truncated__ "Dentelle de fruit de jeunes syrah, aux tanins légers et épics. Hyper digeste. U
n délice." ...
## # $ post_status : chr [1:714] "publish" "publish" "publish" "publish" ...
## # $ comment_status : chr [1:714] "closed" "closed" "closed" "closed" ...
## # $ ping_status : chr [1:714] "closed" "closed" "closed" "closed" ...
## # $ post_password : logi [1:714] NA NA NA NA NA ...
## # $ post_name : chr [1:714] "bon-cadeau-de-25-euros" "pierre-jean-villa-saint-joseph-preface-2018" "pierre-jean-vi
lla-saint-joseph-tilde-2017" "pierre-jean-villa-croze-hermitage-accroche-coeur-2018" ...
## # $ post_modified : POSIXct[1:714], format: "2018-06-01 14:13:57" "2019-12-30 09:30:29" ...
## # $ post_modified_gmt : POSIXct[1:714], format: "2018-06-01 12:13:57" "2019-12-30 08:30:29" ...
## # $ post_content_filtered: logi [1:714] NA NA NA NA NA ...
## # $ post_parent : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ guid : chr [1:714] "https://www.bottle-neck.fr/?post_type=product&p=4954" "https://www.bottle-neck.f
r/?post_type=product&p=3847" "https://www.bottle-neck.fr/?post_type=product&p=3849" "https://www.bottle-neck.fr/?post
_type=product&p=3850" ...
## # $ menu_order : num [1:714] 0 0 0 0 0 0 0 0 ...
## # $ post_type : chr [1:714] "product" "product" "product" "product" ...
## # $ post_mime_type : chr [1:714] NA NA NA NA ...
## # $ comment_count : num [1:714] 0 0 0 0 0 0 0 0 ...
## # NULL

```

```
head(new_web)
```

```
## # A tibble: 6 x 28
##   sku      virtual downloadable rating_count average_rating total_sales tax_status
##   <chr>     <dbl>        <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 bon-c~     0            0            0            0            10 taxable
## 2 15298      0            0            0            0             6 taxable
## 3 15296      0            0            0            0             0 taxable
## 4 15300      0            0            0            0             0 taxable
## 5 19814      0            0            0            0             3 taxable
## 6 19815      0            0            0            0             0 taxable
## # ... with 21 more variables: tax_class <lgl>, post_author <dbl>,
## #   post_date <dttm>, post_date_gmt <dttm>, post_content <lgl>,
## #   post_title <chr>, post_excerpt <chr>, post_status <chr>,
## #   comment_status <chr>, ping_status <chr>, post_password <lgl>,
## #   post_name <chr>, post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

```
Cle(new_web, c('sku'))
```

```
## [1] "La clé n'est pas présente plusieurs fois dans le dataframe."
## [1] "Elle peut donc être utilisée comme clé primaire."
## [1] "Le dataframe est de la forme : (714,28) (lignes, colonnes)"
```

JOINTURES

Jointure entre erp et liaison

```
erp_liaison <- merge(erp, liaison, by="product_id", all=TRUE)
print(dim(erp_liaison))
```

```
## [1] 825   6
```

```
print(str(erp_liaison))
```

```
## 'data.frame': 825 obs. of 6 variables:  
## $ product_id : num 3847 3849 3850 4032 4039 ...  
## $ onsale_web : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ price      : num 24.2 34.3 20.8 14.1 46 34.3 32.7 31.2 60 42.6 ...  
## $ stock_quantity: num 0 0 0 0 0 15 34 12 66 ...  
## $ stock_status  : chr "outofstock" "outofstock" "outofstock" "outofstock" ...  
## $ id_web       : chr "15298" "15296" "15300" "19814" ...  
## NULL
```

```
head(erp_liaison)
```

```
##   product_id onsale_web price stock_quantity stock_status id_web  
## 1      3847          1   24.2            0  outofstock  15298  
## 2      3849          1   34.3            0  outofstock  15296  
## 3      3850          1   20.8            0  outofstock  15300  
## 4      4032          1   14.1            0  outofstock  19814  
## 5      4039          1   46.0            0  outofstock  19815  
## 6      4040          1   34.3            0  outofstock  15303
```

Il serait important de vérifier si chaque ligne a pu trouver une correspondance.

on vérifie si des id_product de liaison sont manquants de erp

```
anti_join(liaison, erp, by="product_id")
```

```
## # A tibble: 0 x 2  
## # ... with 2 variables: product_id <dbl>, id_web <chr>
```

inverse

```
anti_join(erp, liaison, by="product_id")  
  
## # A tibble: 0 x 5  
## # ... with 5 variables: product_id <dbl>, onsale_web <dbl>, price <dbl>,  
## #   stock_quantity <dbl>, stock_status <chr>
```

En somme, la jointure s'est bien passé. Chaque ligne a pu trouver une correspondance. De ce fait nous pouvons passer à la prochaine jointure.

jointure entre liaison et web

```
final <- merge(erp_liaison, new_web, by.x='id_web', by.y='sku', all=TRUE)  
print(dim(final))  
  
## [1] 825 33  
  
head(final)
```

```

##   id_web product_id onsale_web price stock_quantity stock_status virtual
## 1  10014      5913       1  36.0          9    instock     0
## 2  10459      4617       1  67.5          1    instock     0
## 3  10775      4709       1  44.0          0  outofstock     0
## 4  10814      4656       1  43.0          3    instock     0
## 5  11049      4619       1  59.0          5    instock     0
## 6  11225      4970       1  49.5          0  outofstock     0
##   downloadable rating_count average_rating total_sales tax_status tax_class
## 1              0            0             0            0 taxable      NA
## 2              0            0             0            0 taxable      NA
## 3              0            0             0            0 taxable      NA
## 4              0            0             0            0 taxable      NA
## 5              0            0             0            0 taxable      NA
## 6              0            0             0            0 taxable      NA
##   post_author      post_date    post_date_gmt post_content
## 1           2 2019-04-04 15:45:23 2019-04-04 13:45:23      NA
## 2           2 2018-04-13 15:58:19 2018-04-13 13:58:19      NA
## 3           2 2018-04-17 21:28:52 2018-04-17 19:28:52      NA
## 4           2 2018-04-17 10:12:40 2018-04-17 08:12:40      NA
## 5           2 2018-04-13 16:07:49 2018-04-13 14:07:49      NA
## 6           2 2018-06-06 16:47:38 2018-06-06 14:47:38      NA
##                           post_title
## 1          Darnley's London Dry Gin Original
## 2          Alphonse Mellot Sancerre Rouge Génération XIX 2011
## 3          Albert Mann Pinot Gris Vendanges Tardives Altenbourg 2011
## 4          Thierry Germain Saumur-Champigny Outre Terre 2013
## 5          Alphonse Mellot Sancerre Rouge En Grands Champs 2012
## 6 Champagne Larmandier-Bernier Terre de Vertus Premier Cru 2011
##
post_excerpt
## 1 <span title="">Un gin léger, floral et aromatisé aux agrumes dans le style London Dry. Au nez ressort de généreuses notes d'</span><span title="">aiguilles de pin, de genévrier, de fleur de sureau, ou encore de limonade. En bouche, les agrumes dominent, accompagnés de </span><span title="">genièvre, suivis de quelques épices. </span><span title="">La finale est toute en caractère, sur des notes fruitières et de fleur de sureau.</span>\r\n\r\n&ampnbsp
## 2 Robe d'un superbe rubis, dense et profond - Aspect mûr. Le nez est riche et complexe. Tout est séduction dans ce vin, la culture et la précision en plus. Il cueille pour nous la mûre, la myrtille et la griotte et nous offre sur un fond de fraîcheur les épices venues d'Orient : poivre rose, vanille et cannelle. La bouche est séduction et plénitude.
## 3

```

Belle bouteille dotée d'une grande complexité aromatique. Le vin est fruité, marqué par des notes de fruits jaunes, pêche et abricot. Et il dévoile des arômes de vanille et de marzipan. La bouche est suave avec une fine acidité. Ce vin ne demande qu'à vieillir quelques années.

4

La robe est rouge vif. Le nez est très floral, sur la pivoine. En bouche, l'attaque est ample et dense avec une belle sucrosité et la jolie amertume en fin de bouche ramène une belle fraîcheur.

5

Superbe robe carmin profond. Un très grand terroir situé sur les kimméridgiens au sommet de la Moussière où une poche de calcaire de Buzançais lui confère un caractère à grains denses et fins. La mûre, la griotte et le poivre rose se fondent dans des tanins soyeux et de grande garde.

6 C'est un vin à la robe pâle et aux reflets verts et argentés. Au nez, on découvre un champagne sur des notes de fruits blancs (pomme et poire), mais aussi d'épices et de fruits secs. Un nez élégant et net. En bouche, il est très équilibrée, avec une texture fine et crémeuse. Des notes élégantes d'agrumes, de fruits blancs naissent sur le palais, portées par une belle fraîcheur et une élégante minéralité. C'est un vin aérien, qui s'achève sur une belle finale iodée et poivrée.

post_status comment_status ping_status post_password

	post_name	guid	menu_order
## 1	darnleys-london-dry-gin-original		0
## 2	alphonse-mellot-sancerre-rouge-generation-xix-2011		0
## 3	albert-mann-pinot-gris-vendanges-tardives-altenbourg-2011		0
## 4	thierry-germain-saumur-champigny-outre-terre-2013		0
## 5	alphonse-mellot-sancerre-rouge-en-grands-champs-2012		0
## 6	champagne-larmandier-bernier-terre-de-vertus-premier-cru-2011		0
##			
## 1	2020-08-26 15:55:02 2020-08-26 13:55:02	NA	0
## 2	2020-08-24 14:05:02 2020-08-24 12:05:02	NA	0
## 3	2018-12-22 10:30:02 2018-12-22 09:30:02	NA	0
## 4	2020-08-05 14:35:02 2020-08-05 12:35:02	NA	0
## 5	2020-01-31 09:30:02 2020-01-31 08:30:02	NA	0
## 6	2019-05-17 10:10:03 2019-05-17 08:10:03	NA	0
##			
## 1	https://www.bottle-neck.fr/?post_type=product&p=5913	0	
## 2	https://www.bottle-neck.fr/?post_type=product&p=4617	0	
## 3	https://www.bottle-neck.fr/?post_type=product&p=4709	0	

```
## 4 https://www.bottle-neck.fr/?post_type=product&p=4656          0
## 5 https://www.bottle-neck.fr/?post_type=product&p=4619          0
## 6 https://www.bottle-neck.fr/?post_type=product&p=4970          0
##   post_type post_mime_type comment_count
## 1 product      <NA>        0
## 2 product      <NA>        0
## 3 product      <NA>        0
## 4 product      <NA>        0
## 5 product      <NA>        0
## 6 product      <NA>        0
```

on vérifie si les lignes web n'ont pas matchées

```
test <- anti_join(new_web, erp_liaison, by=c('sku' = 'id_web'))
print(dim(test))
```

```
## [1] 0 28
```

```
head(test)
```

```
## # A tibble: 0 x 28
## # ... with 28 variables: sku <chr>, virtual <dbl>, downloadable <dbl>,
## #   rating_count <dbl>, average_rating <dbl>, total_sales <dbl>,
## #   tax_status <chr>, tax_class <lgl>, post_author <dbl>, post_date <dttm>,
## #   post_date_gmt <dttm>, post_content <lgl>, post_title <chr>,
## #   post_excerpt <chr>, post_status <chr>, comment_status <chr>,
## #   ping_status <chr>, post_password <lgl>, post_name <chr>,
## #   post_modified <dttm>, post_modified_gmt <dttm>,
## #   post_content_filtered <lgl>, post_parent <dbl>, guid <chr>,
## #   menu_order <dbl>, post_type <chr>, post_mime_type <chr>,
## #   comment_count <dbl>
```

On regarde si des lignes de df_erp_liaison n'ont pas matchées

```
test2 <- anti_join(erp_liaison, new_web, by=c('id_web' = 'sku'))
print(dim(test2))
```

```
## [1] 111   6
```

```
head(test2)
```

```
##   product_id onsale_web price stock_quantity stock_status id_web
## 1      4055          0  86.1            0    outofstock    <NA>
## 2      4090          0  73.0            0    outofstock    <NA>
## 3      4092          0  47.0            0    outofstock    <NA>
## 4      4195          0  14.1            0    outofstock    <NA>
## 5      4209          0  73.5            0    outofstock    <NA>
## 6      4233          0  20.0            0    outofstock    <NA>
```

Lors de cette jointure, on remarque qu'il y a 111 lignes qui n'ont pas trouvé de correspondance. Ces lignes proviennent toutes de la jointure erp_liaison. Ce qui veut dire que les 'id_web' présents dans la table liaison n'ont pas trouvé de correspondance dans les 'sku' de la table web. On constate aussi que la colonne id_web comporte des NA. Nous devons vérifier si c'est le cas de tout le dataframe.

id_web non nul

```
id_web_non_nul <- test2[!is.na(test2['id_web'])]
print(dim(id_web_non_nul))
```

```
## [1] 20   6
```

```
head(id_web_non_nul)
```

```
##   product_id onsale_web price stock_quantity stock_status id_web
## 9      4289          0  22.8            0 outofstock 13771
## 11     4568          0  21.5            0 outofstock 15065
## 14     4584          0  32.3            0 outofstock 14785
## 25     4741          0  12.4            0 outofstock 12601
## 28     4864          0   8.3            0 outofstock 15154
## 29     4869          0  17.2            0 outofstock 14360
```

Par la suite, il serait important de conserver que les lignes qui ont matchées et supprimer les lignes qui n'ont pas trouver de correspondance avec new_web. Il seraient nécessaire de prendre en compte ce problème et d'y remédier.

conservation des lignes qui ont matchées

```
final_off <- merge(erp_liaison, new_web, by.x='id_web', by.y='sku')
print(dim(final_off))
```

```
## [1] 714 33
```

```
head(final_off)
```

```

##   id_web product_id onsale_web price stock_quantity stock_status virtual
## 1  10014      5913       1  36.0          9    instock     0
## 2  10459      4617       1  67.5          1    instock     0
## 3  10775      4709       1  44.0          0  outofstock     0
## 4  10814      4656       1  43.0          3    instock     0
## 5  11049      4619       1  59.0          5    instock     0
## 6  11225      4970       1  49.5          0  outofstock     0
##   downloadable rating_count average_rating total_sales tax_status tax_class
## 1              0            0             0            0 taxable      NA
## 2              0            0             0            0 taxable      NA
## 3              0            0             0            0 taxable      NA
## 4              0            0             0            0 taxable      NA
## 5              0            0             0            0 taxable      NA
## 6              0            0             0            0 taxable      NA
##   post_author      post_date    post_date_gmt post_content
## 1           2 2019-04-04 15:45:23 2019-04-04 13:45:23      NA
## 2           2 2018-04-13 15:58:19 2018-04-13 13:58:19      NA
## 3           2 2018-04-17 21:28:52 2018-04-17 19:28:52      NA
## 4           2 2018-04-17 10:12:40 2018-04-17 08:12:40      NA
## 5           2 2018-04-13 16:07:49 2018-04-13 14:07:49      NA
## 6           2 2018-06-06 16:47:38 2018-06-06 14:47:38      NA
##                           post_title
## 1 Darnley's London Dry Gin Original
## 2 Alphonse Mellot Sancerre Rouge Génération XIX 2011
## 3 Albert Mann Pinot Gris Vendanges Tardives Altenbourg 2011
## 4 Thierry Germain Saumur-Champigny Outre Terre 2013
## 5 Alphonse Mellot Sancerre Rouge En Grands Champs 2012
## 6 Champagne Larmandier-Bernier Terre de Vertus Premier Cru 2011
##
post_excerpt
## 1 <span title="">Un gin léger, floral et aromatisé aux agrumes dans le style London Dry. Au nez r
essort de généreuses notes d'</span><span title="">aiguilles de pin, de genévrier, de fleur de sureau, ou encore de limonade. E
n bouche, les agrumes dominent, accompagnés de </span><span title="">genièvre, suivis de quelques épices. </span><span title
="">La finale est toute en caractère, sur des notes fruitières et de fleur de sureau.</span>\r\n\r\n&ampnbsp
## 2
Robe d'un superbe rubis, dense et profond - Aspect mûr. Le nez est riche et complexe. Tout est séduction dans ce vin, la cultur
e et la précision en plus. Il cueille pour nous la mûre, la myrtille et la griotte et nous offre sur un fond de fraîcheur les é
pices venues d'Orient : poivre rose, vanille et cannelle. La bouche est séduction et plénitude.
## 3

```

Belle bouteille dotée d'une grande complexité aromatique. Le vin est fruité, marqué par des notes de fruits jaunes, pêche et abricot. Et il dévoile des arômes de vanille et de marzipan. La bouche est suave avec une fine acidité. Ce vin ne demande qu'à vieillir quelques années.

4

La robe est rouge vif. Le nez est très floral, sur la pivoine. En bouche, l'attaque est ample et dense avec une belle sucrosité et la jolie amertume en fin de bouche ramène une belle fraîcheur.

5

Superbe robe carmin profond. Un très grand terroir situé sur les kimméridgiens au sommet de la Moussière où une poche de calcaire de Buzançais lui confère un caractère à grains denses et fins. La mûre, la griotte et le poivre rose se fondent dans des tanins soyeux et de grande garde.

6 C'est un vin à la robe pâle et aux reflets verts et argentés. Au nez, on découvre un champagne sur des notes de fruits blancs (pomme et poire), mais aussi d'épices et de fruits secs. Un nez élégant et net. En bouche, il est très équilibrée, avec une texture fine et crémeuse. Des notes élégantes d'agrumes, de fruits blancs naissent sur le palais, portées par une belle fraîcheur et une élégante minéralité. C'est un vin aérien, qui s'achève sur une belle finale iodée et poivrée.

post_status comment_status ping_status post_password

	post_name	guid	menu_order
## 1	darnleys-london-dry-gin-original		0
## 2	alphonse-mellot-sancerre-rouge-generation-xix-2011		0
## 3	albert-mann-pinot-gris-vendanges-tardives-altenbourg-2011		0
## 4	thierry-germain-saumur-champigny-outre-terre-2013		0
## 5	alphonse-mellot-sancerre-rouge-en-grands-champs-2012		0
## 6	champagne-larmandier-bernier-terre-de-vertus-premier-cru-2011		0
##			
## 1	2020-08-26 15:55:02 2020-08-26 13:55:02	NA	0
## 2	2020-08-24 14:05:02 2020-08-24 12:05:02	NA	0
## 3	2018-12-22 10:30:02 2018-12-22 09:30:02	NA	0
## 4	2020-08-05 14:35:02 2020-08-05 12:35:02	NA	0
## 5	2020-01-31 09:30:02 2020-01-31 08:30:02	NA	0
## 6	2019-05-17 10:10:03 2019-05-17 08:10:03	NA	0
##			
## 1	https://www.bottle-neck.fr/?post_type=product&p=5913	0	
## 2	https://www.bottle-neck.fr/?post_type=product&p=4617	0	
## 3	https://www.bottle-neck.fr/?post_type=product&p=4709	0	

```
## 4 https://www.bottle-neck.fr/?post_type=product&p=4656          0
## 5 https://www.bottle-neck.fr/?post_type=product&p=4619          0
## 6 https://www.bottle-neck.fr/?post_type=product&p=4970          0
##   post_type post_mime_type comment_count
## 1 product      <NA>           0
## 2 product      <NA>           0
## 3 product      <NA>           0
## 4 product      <NA>           0
## 5 product      <NA>           0
## 6 product      <NA>           0
```

CALCUL DU CHIFFRE D'AFFAIRE PAR PRODUIT ET LE TOTAL

calcul du chiffre d'affaire par produit

```
final_off['CAproduit'] <- final_off['price'] * final_off['total_sales']
print(dim(final_off))
```

```
## [1] 714 34
```

```
head(final_off)
```

```

##   id_web product_id onsale_web price stock_quantity stock_status virtual
## 1  10014      5913       1  36.0          9    instock     0
## 2  10459      4617       1  67.5          1    instock     0
## 3  10775      4709       1  44.0          0  outofstock     0
## 4  10814      4656       1  43.0          3    instock     0
## 5  11049      4619       1  59.0          5    instock     0
## 6  11225      4970       1  49.5          0  outofstock     0
##   downloadable rating_count average_rating total_sales tax_status tax_class
## 1              0            0             0            0 taxable     NA
## 2              0            0             0            0 taxable     NA
## 3              0            0             0            0 taxable     NA
## 4              0            0             0            0 taxable     NA
## 5              0            0             0            0 taxable     NA
## 6              0            0             0            0 taxable     NA
##   post_author      post_date    post_date_gmt post_content
## 1           2 2019-04-04 15:45:23 2019-04-04 13:45:23     NA
## 2           2 2018-04-13 15:58:19 2018-04-13 13:58:19     NA
## 3           2 2018-04-17 21:28:52 2018-04-17 19:28:52     NA
## 4           2 2018-04-17 10:12:40 2018-04-17 08:12:40     NA
## 5           2 2018-04-13 16:07:49 2018-04-13 14:07:49     NA
## 6           2 2018-06-06 16:47:38 2018-06-06 14:47:38     NA
##                           post_title
## 1          Darnley's London Dry Gin Original
## 2          Alphonse Mellot Sancerre Rouge Génération XIX 2011
## 3          Albert Mann Pinot Gris Vendanges Tardives Altenbourg 2011
## 4          Thierry Germain Saumur-Champigny Outre Terre 2013
## 5          Alphonse Mellot Sancerre Rouge En Grands Champs 2012
## 6 Champagne Larmandier-Bernier Terre de Vertus Premier Cru 2011
##
post_excerpt
## 1 <span title="">Un gin léger, floral et aromatisé aux agrumes dans le style London Dry. Au nez r
essort de généreuses notes d'</span><span title="">aiguilles de pin, de genévrier, de fleur de sureau, ou encore de limonade. E
n bouche, les agrumes dominent, accompagnés de </span><span title="">genièvre, suivis de quelques épices. </span><span title
="">La finale est toute en caractère, sur des notes fruitières et de fleur de sureau.</span>\r\n\r\n&ampnbsp
## 2
Robe d'un superbe rubis, dense et profond - Aspect mûr. Le nez est riche et complexe. Tout est séduction dans ce vin, la cultur
e et la précision en plus. Il cueille pour nous la mûre, la myrtille et la griotte et nous offre sur un fond de fraîcheur les é
pices venues d'Orient : poivre rose, vanille et cannelle. La bouche est séduction et plénitude.
## 3

```

Belle bouteille dotée d'une grande complexité aromatique. Le vin est fruité, marqué par des notes de fruits jaunes, pêche et abricot. Et il dévoile des arômes de vanille et de marzipan. La bouche est suave avec une fine acidité. Ce vin ne demande qu'à vieillir quelques années.

4

La robe est rouge vif. Le nez est très floral, sur la pivoine. En bouche, l'attaque est ample et dense avec une belle sucrosité et la jolie amertume en fin de bouche ramène une belle fraîcheur.

5

Superbe robe carmin profond. Un très grand terroir situé sur les kimméridgiens au sommet de la Moussière où une poche de calcaire de Buzançais lui confère un caractère à grains denses et fins. La mûre, la griotte et le poivre rose se fondent dans des tanins soyeux et de grande garde.

6 C'est un vin à la robe pâle et aux reflets verts et argentés. Au nez, on découvre un champagne sur des notes de fruits blancs (pomme et poire), mais aussi d'épices et de fruits secs. Un nez élégant et net. En bouche, il est très équilibrée, avec une texture fine et crémeuse. Des notes élégantes d'agrumes, de fruits blancs naissent sur le palais, portées par une belle fraîcheur et une élégante minéralité. C'est un vin aérien, qui s'achève sur une belle finale iodée et poivrée.

post_status comment_status ping_status post_password

	post_name	guid	menu_order
## 1	darnleys-london-dry-gin-original		0
## 2	alphonse-mellot-sancerre-rouge-generation-xix-2011		0
## 3	albert-mann-pinot-gris-vendanges-tardives-altenbourg-2011		0
## 4	thierry-germain-saumur-champigny-outre-terre-2013		0
## 5	alphonse-mellot-sancerre-rouge-en-grands-champs-2012		0
## 6	champagne-larmandier-bernier-terre-de-vertus-premier-cru-2011		0
##			
## 1	2020-08-26 15:55:02 2020-08-26 13:55:02	NA	0
## 2	2020-08-24 14:05:02 2020-08-24 12:05:02	NA	0
## 3	2018-12-22 10:30:02 2018-12-22 09:30:02	NA	0
## 4	2020-08-05 14:35:02 2020-08-05 12:35:02	NA	0
## 5	2020-01-31 09:30:02 2020-01-31 08:30:02	NA	0
## 6	2019-05-17 10:10:03 2019-05-17 08:10:03	NA	0
##			
## 1	https://www.bottle-neck.fr/?post_type=product&p=5913	0	
## 2	https://www.bottle-neck.fr/?post_type=product&p=4617	0	
## 3	https://www.bottle-neck.fr/?post_type=product&p=4709	0	

```
## 4 https://www.bottle-neck.fr/?post_type=product&p=4656          0
## 5 https://www.bottle-neck.fr/?post_type=product&p=4619          0
## 6 https://www.bottle-neck.fr/?post_type=product&p=4970          0
##   post_type post_mime_type comment_count CAproduit
## 1 product      <NA>           0           0
## 2 product      <NA>           0           0
## 3 product      <NA>           0           0
## 4 product      <NA>           0           0
## 5 product      <NA>           0           0
## 6 product      <NA>           0           0
```

calcul du chiffre d'affaire total

```
CA_TOTAL<- sum(final_off$CAproduit,na.rm = TRUE)
print(paste("Le chiffre d'affaire total :" , CA_TOTAL , "euros."))
```

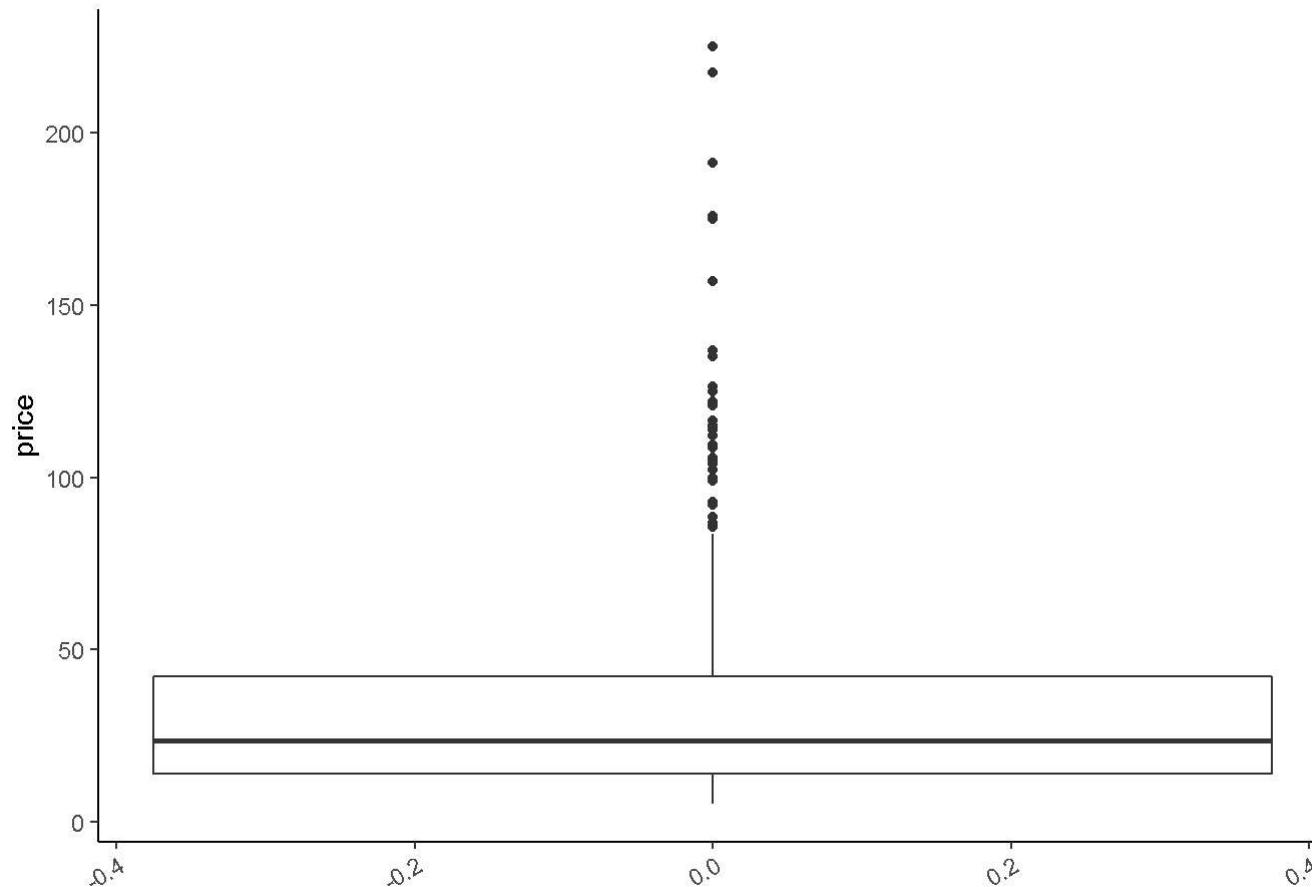
```
## [1] "Le chiffre d'affaire total : 70568.6 euros."
```

ANALYSE UNIVARIEE

Representation graphique des valeurs abérrantes

```
ggplot(final_off, aes(x=price)) +
  geom_boxplot() + coord_flip() + theme_classic() + theme(axis.text.x = element_text(angle=30, hjust=1, vjust=1))+ 
  theme(legend.position="none")+
  ggtitle("Représentation des points aberrants")
```

Représentation des points aberrants



LISTER LES VALEURS ABERRANTES SUR UN TABLEAU

```
outlier_val<-boxplot.stats(final_off$price)$out  
outlier_idx <-which(final_off$price %in% c(outlier_val))  
  
liste_outlier<-final_off[outlier_idx,]  
head(liste_outlier)
```

```

##      id_web product_id onsale_web price stock_quantity stock_status virtual
## 13    11601      6214          1  99.0           7 instock       0
## 14    11602      5008          1 105.0          10 instock       0
## 16    11668      4132          1  88.4           5 instock       0
## 39    12589      4355          1 126.5           2 instock       0
## 46    12790      6215          1 115.0           4 instock       0
## 47    12791      5007          1 105.0          17 instock       0
##      downloadable rating_count average_rating total_sales tax_status tax_class
## 13            0             0            0            0 taxable     NA
## 14            0             0            0            0 taxable     NA
## 16            0             0            0            0 taxable     NA
## 39            0             0            0           11 taxable     NA
## 46            0             0            0            0 taxable     NA
## 47            0             0            0            0 taxable     NA
##      post_author      post_date    post_date_gmt post_content
## 13          2 2019-07-25 09:15:41 2019-07-25 07:15:41      NA
## 14          2 2018-07-17 10:52:41 2018-07-17 08:52:41      NA
## 16          2 2018-02-13 11:43:55 2018-02-13 10:43:55      NA
## 39          2 2018-03-02 10:46:10 2018-03-02 09:46:10      NA
## 46          2 2019-07-25 09:30:16 2019-07-25 07:30:16      NA
## 47          2 2018-07-17 10:36:03 2018-07-17 08:36:03      NA
##
##                                         post_title
## 13          Domaine des Comtes Lafon Volnay 1er Cru Champans 2013
## 14          Domaine des Comtes Lafon Volnay 1er Cru Santenots du Milieu 2013
## 16 Zind-Humbrecht Pinot Gris Grand Cru Rangen De Thann Clos Saint-Urbain 2012
## 39          Champagne Egly-Ouriet Grand Cru Blanc de Noirs
## 46          Domaine des Comtes Lafon Volnay 1er Cru Champans 2014
## 47          Domaine des Comtes Lafon Volnay 1er Cru Santenots du Milieu 2014
##
## post_excerpt
## 13
La couleur rouge intense annonce un belle concentration, le nez évoque les fruits rouges, légèrement épicé, la bouche d'une très grande élégance livre une texture onctueuse et soyeuse malgré une structure bien présente.
## 14
"Il s'agit là de la meilleure partie de l'appellation Santenots, autrefois classée tête de cuvée par le docteur Lavalle". <span class="font5">Voilà qui nous plonge dans le bain de cette cuvée emblématique du domaine qui fut assemblée pièce par pièce, tel un puzzle, par le comte Jules Lafon. Ces 3 ha de vignes situées sur le village de Meursault et non de Volnay, sont enracinés sur une terre rouge très argileuse qui donne vie à l'un des plus grands vins rouges de la Côte de Beaune. </span>
## 16
Le nez dévoile déjà une belle intensité de fruit. Il est facile de ressentir

```

r le besoin de ce vin de continuer à évoluer, mais le verre déborde littéralement d'arômes propres au Rangen (fumée légère, fruits blancs, tourbe, pierre à fusil...). Ce vin dégage une sensation de sérénité et d'achèvement, confirmé par une bouche ample et délicate. La finale est très longue mais ce vin aura besoin de quelques années de vieillissement en bouteille et bénéficiera d'un passage à l'air avant le service.

39 Le Blanc de Noirs représente le meilleur du savoir-faire de la maison Egly-Ouriet. C'est une cuvée d'une grande finesse. Ce champagne est élevé en fût de chêne et la fermentation malolactique est volontairement bloquée afin de conserver une grande tension. Il possède une robe vieil or, et des bulles d'une très grande finesse. Au nez, c'est une cuvée qui s'ouvre sur des arômes gourmands de fruits compotés, de fruits jaunes, et une touche d'épices. La bouche est superbe, mêlant à la fois fraîcheur, rondeur, tension. Une immense cuvée !

46

La couleur rouge intense annonce un belle concentration, le nez évoque les fruits rouges, légèrement épicé, la bouche d'une très grande élégance livre une texture onctueuse et soyeuse malgré une structure bien présente.

47

"Il s'agit là de la meilleure partie de l'appellation Santenots, autrefois

classée tête de cuvée par le docteur Lavalle". Voilà qui nous plonge dans le bain de cette cuvée emblématique du domaine qui fut assemblée pièce par pièce, tel un puzzle, par le comte Jules Lafon. Ces 3 ha de vignes situées sur le village de Meursault et non de Volnay, sont enracinés sur une terre rouge très argileuse qui donne vie à l'un des plus grands vins rouges de la Côte de Beaune.

```
##   post_status comment_status ping_status post_password
## 13   publish      closed      closed       NA
## 14   publish      closed      closed       NA
## 16   publish      closed      closed       NA
## 39   publish      closed      closed       NA
## 46   publish      closed      closed       NA
## 47   publish      closed      closed       NA
##
##                                         post_name
## 13                 domaine-des-comtes-lafon-volnay-1er-cru-champans-2013
## 14                 domaine-des-comtes-lafon-volnay-1er-cru-santenots-du-milieu-2013
## 16 zind-humbrecht-pinot-gris-grand-cru-rangen-de-thann-clos-saint-urbain-2012
## 39                   champagne-egly-ouriet-grand-cru-brut-blanc-de-noirs
## 46                 domaine-des-comtes-lafon-volnay-1er-cru-champans-2014
## 47                 domaine-des-comtes-lafon-volnay-1er-cru-santenots-du-milieu-2014
##   post_modified  post_modified_gmt post_content_filtered post_parent
## 13 2020-07-04 11:35:02 2020-07-04 09:35:02       NA       0
## 14 2020-06-23 15:35:02 2020-06-23 13:35:02       NA       0
## 16 2020-02-20 09:55:02 2020-02-20 08:55:02       NA       0
## 39 2020-08-13 10:15:02 2020-08-13 08:15:02       NA       0
## 46 2019-11-04 09:30:25 2019-11-04 08:30:25       NA       0
## 47 2020-07-02 09:30:03 2020-07-02 07:30:03       NA       0
##
##                                         guid menu_order
```

```
## 13 https://www.bottle-neck.fr/?post_type=product&p=6214          0
## 14 https://www.bottle-neck.fr/?post_type=product&p=5008          0
## 16 https://www.bottle-neck.fr/?post_type=product&p=4132          0
## 39 https://www.bottle-neck.fr/?post_type=product&p=4355          0
## 46 https://www.bottle-neck.fr/?post_type=product&p=6215          0
## 47 https://www.bottle-neck.fr/?post_type=product&p=5007          0
##   post_type post_mime_type comment_count CAproduit
## 13  product      <NA>        0     0.0
## 14  product      <NA>        0     0.0
## 16  product      <NA>        0     0.0
## 39  product      <NA>        0  1391.5
## 46  product      <NA>        0     0.0
## 47  product      <NA>        0     0.0
```

les valeurs aberrantes que nous avons dans ce jeu de données correspondent à des produits très chers.

```
ABERANTE<- liste_outlier %>%
  select(id_web, product_id, price)
head(ABERANTE)
```

```
##   id_web product_id price
## 13  11601      6214  99.0
## 14  11602      5008 105.0
## 16  11668      4132  88.4
## 39  12589      4355 126.5
## 46  12790      6215 115.0
## 47  12791      5007 105.0
```