

# FGV - Preços

Lucas M. Oliveira

09/09/2020

## Pré-processamento dos dados

```
my_path<- "C:/Users/lucmo/Downloads/FGV_Iliada/FGV_PRECOS/" # caminho onde estao os seus arquivos
setwd(my_path) #set o diretorio
pacman::p_load('tidyverse',
               'stringr','stringi',
               'magrittr') #carregando os pacotes necessarios para manipulacao e Limpreza dos dados
amostra = 'amostra_pricecollector.csv' %>%
  data.table::fread()

Catalogo = 'Catalogo.xlsx' %>%
  readxl::read_excel(sheet = "1", skip = 3)
```

Selecionando dez alimentos com letra 'C':

```
Catalogo %>% select(`DESCRIÇÃO ITEM`) %>% rename(DI = `DESCRIÇÃO ITEM`) %>%
  mutate(DI = str_replace_all(DI, "-", " "),
  Letra = str_sub(DI, end = 1),
  Letra = str_to_upper(Letra)) %>% filter(Letra=='C') %>%
  slice(56:65)
```

```
## # A tibble: 10 x 2
##   DI                Letra
##   <chr>            <chr>
## 1 COENTRO          C
## 2 COLORIFICO       C
## 3 COMINHO EM PO    C
## 4 CONDIMENTO PREPARADO (CATCHUP), A BASE DE TOMATE C
## 5 CORAÇÃO BOVINO   C
## 6 COUVE            C
## 7 COUVE FLOR       C
## 8 CRAVO DA INDIA   C
## 9 CREME DE ARROZ   C
## 10 CREME DE LEITE  C
```

Processamentos adicionais

```

amostra %<>% mutate(product_name =str_to_upper(product_name),
product_name= stri_trans_general(product_name, "Latin-ASCII")) # retirando ç

```

```

amostra_filtrada = amostra %>%
select(product_id,product_name,price,website_id) %>%
mutate(Letra =str_sub(product_name,end=1)) %>%
filter(Letra%in%c('C','K')) # Por causa de Ketchup

```

Separando o nome do produto na amostra em um vetor:

```

amostra_filtrada %<>% mutate( pd =str_split(product_name," ") ) #

#
Catalogo_filtrado =
  Catalogo %>% select(`DESCRIÇÃO ITEM`,QTDE,U.M.) %>% rename(DI = `DESCRIÇÃO ITEM`) %>%
  mutate(DI = str_replace_all(DI, "-", " "),
    Letra = str_sub(DI, end = 1),
    Letra = str_to_upper(Letra),
    DI = stri_trans_general(DI, "Latin-ASCII")) %>% filter(Letra=='C') %>%
  slice(56:65) # selecionando os ultimos dez alimentos

```

Trocando o nome na amostra de Catchup

```

Catalogo_filtrado %<>%
  mutate(DI=if_else(DI=='CONDIMENTO PREPARADO (CATCHUP), A BASE DE TOMATE',
    'CATCHUP',DI))%>%
  add_row( DI= 'KETCHUP',
    QTDE = Catalogo_filtrado$QTDE[4],
    U.M.=Catalogo_filtrado$U.M.[4],
    Letra=Catalogo_filtrado$Letra[4]) %>%
  mutate(match =0)# Buscando exatamente por CATCHUP

```

Extraindo o comprimento de cada item no catálogo:

```

for (i in 1:dim(Catalogo_filtrado)[1]) {

  Catalogo_filtrado$match[i]<-lapply(str_split(Catalogo_filtrado$DI, " "), length)[[i]]
}
#-----#

Catalogo_filtrado %<>% mutate(DI_SEP =str_split(DI, " "))

amostra_filtrada %<>% mutate(DI=NA)

`%!in%` = Negate(`%in%`) #def operador de negacao

for(i in 1:length(amostra_filtrada$pd)){

  for (j in 1:dim(Catalogo_filtrado)[1]) {

```

```

# if(!is.na(table(unique(amostra_filtrada$pd[[i]])%in%Catalogo_filtrado$DI_SEP[[j]])["TRUE"])[1])
#                               Catalogo_filtrado$match[j])){

if(amostra_filtrada$pd[[i]][1] == Catalogo_filtrado$DI_SEP[[j]][1]){

  if(table(unique(amostra_filtrada$pd[[i]])%in%Catalogo_filtrado$DI_SEP[[j]])["TRUE"][[1]] ==
  Catalogo_filtrado$match[j]){
    amostra_filtrada$DI[i] = Catalogo_filtrado$DI[j]}
  # }else{
  #   amostra_filtrada$DI[i] = NA
  #   }
}

}
}

```

Obtendo unidades e Preços

```

amostra_filtrada %<>% drop_na() %>% mutate(pd_name2 = str_replace_all(product_name,pattern = " ", replace = ""))

#extraindo o peso/volume

grama =bind_cols(amostra_filtrada$product_id, str_extract(amostra_filtrada$pd_name2,"\\d+G")) %>% drop_na()
ml =bind_cols(amostra_filtrada$product_id, str_extract(amostra_filtrada$pd_name2,"\\d+ML")) %>% drop_na()
kilo =bind_cols(amostra_filtrada$product_id, str_extract(amostra_filtrada$pd_name2,"\\d+KG")) %>% drop_na()
unidade =bind_cols(amostra_filtrada$product_id, str_extract(amostra_filtrada$pd_name2,"UNID")) %>% drop_na()
litro = bind_cols(amostra_filtrada$product_id, str_extract(amostra_filtrada$pd_name2,"\\d+L")) %>% drop_na()

#to_find<- c("\\d+G","\\d+ML","\\d+KG","\\d+L","UNIDADE")

colnames(grama)<-c("product_id","qtd")
colnames(ml)<-c("product_id","qtd")
colnames(kilo)<-c("product_id","qtd")
colnames(unidade)<-c("product_id","qtd")
colnames(litro)<-c("product_id","qtd")

quant_amostra = rbind(grama,ml,kilo,unidade,litro)

amostra_filtrada= quant_amostra %>% inner_join(amostra_filtrada,by = 'product_id')

UM_amostra = str_extract(amostra_filtrada$qtd,"G|ML|L|UNID|KG") # UM da amostra
Qtd_amostra = str_extract(amostra_filtrada$qtd,'[0-9]+')

Cotacao =
  amostra_filtrada %>%
    inner_join(Catalogo_filtrado,by='DI') %>%
    mutate(QTDE = if_else(QTDE %in%c(NA),1,QTDE),
           price = as.numeric(str_replace(price,",","."))) %>%
    select(DI,qtd,QTDE,U.M.,price,website_id) %>% cbind(UM_amostra,Qtd_amostra)

Cotacao[,7] <- sapply(Cotacao[,7], as.character) # convertendo U.M. da amostra em caracter

```

```
Cotacao[,8] <- sapply(Cotacao[,8], as.character)
Cotacao[,8] <- sapply(Cotacao[,8], as.numeric)
```

OBS Se a variável UM no catálogo for igual à Molho prevalece o U.M. da amostra

```
Cotacao %<>% mutate(
  U.M. = if_else(U.M.=="MOLHO",UM_amostra,U.M.),
  Qtd_amostra = if_else( U.M=="UNID" & UM_amostra == "UNID",1,Qtd_amostra),
  Qtd_amostra = if_else( U.M=="KG" & UM_amostra == "UNID",1,Qtd_amostra),
  Preco_UM =(if_else(UM_amostra == U.M., (price/Qtd_amostra),0)),
  Preco_UM = if_else( U.M=="KG" & UM_amostra == "G",
    price/ (Qtd_amostra*0.001),Preco_UM),
  Preco_UM=round(Preco_UM,digits=2),
  DI = fct_recode(DI, CATCHUP = "KETCHUP"), # Colocando os Catchups em um só
  Cotacao = QTDE*price)

Cotacao %>% group_by(DI) %>% summarise(n=n()) %>% arrange(desc(n)) # QTD de Itens
```

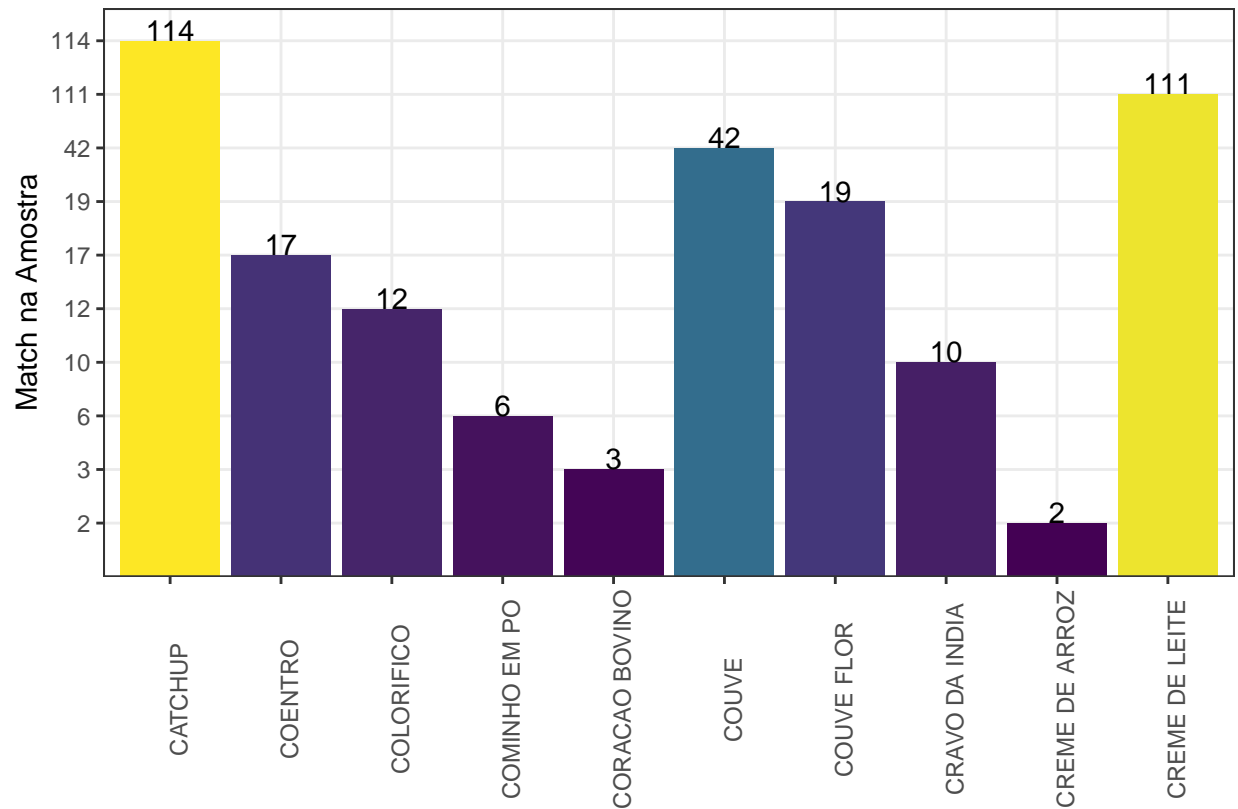
```
## # A tibble: 10 x 2
##   DI              n
##   <fct>          <int>
## 1 CATCHUP        114
## 2 CREME DE LEITE 111
## 3 COUVE          42
## 4 COUVE FLOR     19
## 5 COENTRO        17
## 6 COLORIFICO     12
## 7 CRAVO DA INDIA 10
## 8 COMINHO EM PO   6
## 9 CORACAO BOVINO  3
## 10 CREME DE ARROZ 2
```

```
Cotacao %>% group_by(DI,qtd,website_id) %>% arrange(DI) %>% filter(Cotacao==min(Cotacao)) %>% View()

#
```

Gráfico do número de itens encontrados na amostra

```
Cotacao %>% group_by(DI) %>% summarise(n=n()) %>% arrange(desc(n)) %>%
  ggplot(aes(DI,
    reorder(n,n,
      function(x)-length(x))
    ,fill=n))+geom_bar(stat = "identity")+
  scale_fill_continuous(type = "viridis")+theme_bw() +
  theme(axis.text.x = element_text(angle = 90),legend.position = "none") +geom_text(
    aes(label = n, n = n + 0.05),
    position = position_dodge(0.9),
    vjust = 0
  )+xlab('')+
  ylab('Match na Amostra')
```



Menores valores por item (desconsiderando a quantidade)

```
Cotacao %>% group_by(DI) %>% summarise(menor_preco=min(price)) %>% arrange(desc(menor_preco)) %>% head()
```

```
## # A tibble: 6 x 2
##   DI          menor_preco
##   <fct>          <dbl>
## 1 CORACAO BOVINO      5.99
## 2 CREME DE ARROZ      4.05
## 3 COUVE FLOR         3.85
## 4 CATCHUP            3.49
## 5 CRAVO DA INDIA      3.1
## 6 COLORIFICO         2.89
```

Tabela dos Preços dos itens agrupados por site e quantidade

```
Cotacao %>% group_by(DI,website_id,qtd) %>% summarise(menor_preco=min(price)) %>% arrange(desc(menor_preco)) %>% head()
```

```
## # A tibble: 6 x 4
## # Groups:   DI, website_id [3]
##   DI          website_id qtd  menor_preco
##   <fct>          <int> <chr>    <dbl>
## 1 CATCHUP            8 350G      23.5
```

## 2	CREME DE LEITE	1 490G	22.6
## 3	CREME DE LEITE	1 500G	22.0
## 4	CATCHUP	8 310G	20.0
## 5	CREME DE LEITE	5 500G	19.9
## 6	CATCHUP	8 300G	19.8