cade\_1.R

DELL

2024-09-29

# ANALISE DOS PROCESSOS DO CADE ENVOLVENDO CARTEIS INTERNACIONAIS #######  
# CONFIG ----  
# Instalar e carregar pacotes necessários  
if(!require(readxl)) install.packages("readxl", dependencies = TRUE)

## Carregando pacotes exigidos: readxl

if(!require(janitor)) install.packages("janitor", dependencies = TRUE)

## Carregando pacotes exigidos: janitor

##   
## Anexando pacote: 'janitor'

## Os seguintes objetos são mascarados por 'package:stats':  
##   
## chisq.test, fisher.test

if(!require(dplyr)) install.packages("dplyr", dependencies = TRUE)

## Carregando pacotes exigidos: dplyr

##   
## Anexando pacote: 'dplyr'

## Os seguintes objetos são mascarados por 'package:stats':  
##   
## filter, lag

## Os seguintes objetos são mascarados por 'package:base':  
##   
## intersect, setdiff, setequal, union

if(!require(lubridate)) install.packages("lubridate", dependencies = TRUE)

## Carregando pacotes exigidos: lubridate

##   
## Anexando pacote: 'lubridate'

## Os seguintes objetos são mascarados por 'package:base':  
##   
## date, intersect, setdiff, union

if(!require(ggplot2)) install.packages("ggplot2", dependencies = TRUE)

## Carregando pacotes exigidos: ggplot2

if(!require(tidyr)) install.packages("tidyr", dependencies = TRUE)

## Carregando pacotes exigidos: tidyr

if(!require(stringr)) install.packages("stringr", dependencies = TRUE)

## Carregando pacotes exigidos: stringr

if(!require(purrr)) install.packages("stringr", dependencies = TRUE)

## Carregando pacotes exigidos: purrr

if (!require(writexl)) install.packages("writexl", dependencies = TRUE)

## Carregando pacotes exigidos: writexl

library(readxl)  
library(janitor)  
library(dplyr)  
library(lubridate)  
library(ggplot2)  
library(tidyr)  
library(stringr)  
library(purrr)  
library(knitr)

## Warning: pacote 'knitr' foi compilado no R versão 4.4.1

library(writexl)  
  
# remove all ----  
rm(list = ls())  
  
# OBTENDO DADOS -----  
# Carregar a base de dados no dataframe df.cade  
df.cade <- read\_excel('./data/cade.xlsx', sheet = 'database\_adj')  
# Visualizar o resumo do dataframe  
summary(df.cade)

## Número do Processo Data de Autuação Conduta Geral   
## Length:186 Min. :2012-03-02 00:00:00.00 Length:186   
## Class :character 1st Qu.:2015-05-18 06:00:00.00 Class :character   
## Mode :character Median :2017-05-19 00:00:00.00 Mode :character   
## Mean :2017-05-27 02:42:34.83   
## 3rd Qu.:2019-04-30 00:00:00.00   
## Max. :2024-01-22 00:00:00.00   
##   
## Conduta Específica Coordenação da SG responsável Celebração do TCC   
## Length:186 Length:186 Length:186   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Suspensão do Processo por TCC Suspensão Judicial Trânsito em julgado  
## Length:186 Length:186 Length:186   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Data Decisão SG Teor da Decisão SG  
## Min. :2014-12-26 00:00:00.00 Length:186   
## 1st Qu.:2019-01-02 00:00:00.00 Class :character   
## Median :2021-04-13 00:00:00.00 Mode :character   
## Mean :2020-10-27 12:21:23.17   
## 3rd Qu.:2022-08-17 00:00:00.00   
## Max. :2024-05-10 00:00:00.00   
## NA's :85   
## Cooperação com Órgãos do Governo Cooperação Internacional  
## Length:186 Length:186   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##   
## Data da Decisão TR Teor da Decisão TR Houve Recurso?   
## Min. :2015-10-14 00:00:00.00 Length:186 Length:186   
## 1st Qu.:2019-10-16 00:00:00.00 Class :character Class :character   
## Median :2021-04-28 00:00:00.00 Mode :character Mode :character   
## Mean :2020-12-17 03:49:33.91   
## 3rd Qu.:2022-08-17 00:00:00.00   
## Max. :2024-05-22 00:00:00.00   
## NA's :117   
## Decisão Final Data Decisão Final Decisão do CADE   
## Length:186 Min. :2015-11-11 00:00:00.00 Length:186   
## Class :character 1st Qu.:2020-08-26 00:00:00.00 Class :character   
## Mode :character Median :2021-08-18 00:00:00.00 Mode :character   
## Mean :2021-08-07 09:06:12.41   
## 3rd Qu.:2022-11-09 00:00:00.00   
## Max. :2024-04-17 00:00:00.00   
## NA's :157   
## Origem Multas Aplicadas   
## Length:186 Min. : 0   
## Class :character 1st Qu.: 0   
## Mode :character Median : 0   
## Mean : 9512885   
## 3rd Qu.: 0   
## Max. :441882474   
##

# Verificar as classes das colunas de data  
str(df.cade)

## tibble [186 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ Número do Processo : chr [1:186] "08012.004280/2012-40" "08012.004422/2012-79" "08012.005324/2012-59" "08700.008576/2012-81" ...  
## $ Data de Autuação : POSIXct[1:186], format: "2012-03-02" "2012-03-07" ...  
## $ Conduta Geral : chr [1:186] "Cartel;" "Cartel;" "Cartel;" "Cartel;" ...  
## $ Conduta Específica : chr [1:186] "Abuso de Posição Dominante; Cartel em Licitações; Diferenciação de preços;" "Cartel em Licitações;" "Cartel Internacional;" "Cartel Internacional; Dificultar continuidade de relações comerciais; Fixar Preço de revenda; Troca de informações sensíveis;" ...  
## $ Coordenação da SG responsável : chr [1:186] "CGAA 8" "CGAA 8" "CGAA 6" "CGAA 7" ...  
## $ Celebração do TCC : chr [1:186] "Sim" "Sim" "Não" "Sim" ...  
## $ Suspensão do Processo por TCC : chr [1:186] "Não" "Não" "Não" "Não" ...  
## $ Suspensão Judicial : chr [1:186] "Não" "Não" "Não" "Não" ...  
## $ Trânsito em julgado : chr [1:186] "Sim" "Sim" "Sim" "Sim" ...  
## $ Data Decisão SG : POSIXct[1:186], format: "2018-01-18" "2016-12-16" ...  
## $ Teor da Decisão SG : chr [1:186] "Condenação" "Condenação" "Condenação" "Condenação" ...  
## $ Cooperação com Órgãos do Governo: chr [1:186] "-" "-" "-" "-" ...  
## $ Cooperação Internacional : chr [1:186] "-" "-" "-" "-" ...  
## $ Data da Decisão TR : POSIXct[1:186], format: "2019-10-30" "2018-08-08" ...  
## $ Teor da Decisão TR : chr [1:186] "Condenação" "Condenação" "Condenação" "Condenação" ...  
## $ Houve Recurso? : chr [1:186] "Sim" "Sim" "Sim" "Não" ...  
## $ Decisão Final : chr [1:186] "Condenação" "Condenação" "Condenação" "-" ...  
## $ Data Decisão Final : POSIXct[1:186], format: "2019-10-30" "2018-09-19" ...  
## $ Decisão do CADE : chr [1:186] "Unanimidade" "Maioria" "Maioria" "-" ...  
## $ Origem : chr [1:186] "Representação" "Ex-Officio" "Leniência" "Leniência" ...  
## $ Multas Aplicadas : num [1:186] 2122348 489461 88205622 732002 732002 ...

# LIMPANDO DADOS ----  
# Ajustar os nomes das variáveis para snake\_case e remover caracteres especiais  
df.cade <- df.cade %>% clean\_names()  
  
# Verificar duplicatas na variável numero\_do\_processo  
duplicados <- df.cade %>%  
 group\_by(numero\_do\_processo) %>%  
 filter(n() > 1)  
  
# Função para verificar se todas as colunas das linhas duplicadas são idênticas  
comparar\_duplicados <- function(grupo) {  
 # Identificar as colunas que têm valores diferentes  
 diferentes <- apply(grupo, 2, function(x) length(unique(x)) > 1)  
 if (all(!diferentes)) {  
 return("Idêntica")  
 } else {  
 # Retornar as colunas que têm valores diferentes  
 return(paste("Diferente nas colunas:", paste(names(grupo)[diferentes], collapse = ", ")))  
 }  
}  
  
# Adicionar uma coluna para identificar se os dados são idênticos ou diferentes  
if(nrow(duplicados) > 0) {  
 print("Processos Duplicados:")  
 print(unique(duplicados$numero\_do\_processo))  
   
 duplicados\_status <- duplicados %>%  
 group\_by(numero\_do\_processo) %>%  
 nest() %>%  
 mutate(status = map\_chr(data, comparar\_duplicados)) %>%  
 select(numero\_do\_processo, status)  
 # Exibir o status dos duplicados e onde estão as diferenças  
 print(duplicados\_status)  
 # Exibir as diferenças detalhadas para os processos marcados como "Diferente"  
 duplicados\_diferentes <- duplicados\_status %>%  
 filter(str\_detect(status, "Diferente"))  
   
 if (nrow(duplicados\_diferentes) > 0) {  
 for (processo in duplicados\_diferentes$numero\_do\_processo) {  
 print(paste("Diferenças encontradas no processo:", processo))  
 # Exibir as linhas do processo específico que possuem diferenças  
 print(df.cade %>% filter(numero\_do\_processo == processo))  
 }  
 } else {  
 print("Não foram encontradas diferenças nos processos duplicados.")  
 }  
   
} else {  
 print("Não foram encontrados processos duplicados.")  
}

## [1] "Processos Duplicados:"  
## [1] "08700.008576/2012-81" "08700.009509/2012-84" "08700.010979/2013-71"  
## [4] "08700.005789/2014-13" "08700.009161/2014-97" "08700.010056/2014-09"  
## [7] "08700.011474/2014-05" "08012.009732/2008-01" "08700.000448/2015-32"  
## [10] "08700.000949/2015-19" "08700.002247/2015-70" "08700.003826/2015-30"  
## [13] "08700.005146/2015-51" "08700.009029/2015-66" "08700.009167/2015-45"  
## [16] "08700.001043/2016-01" "08700.001094/2016-24" "08700.001836/2016-11"  
## [19] "08700.004455/2016-94" "08700.006377/2016-62" "08700.007777/2016-95"  
## [22] "08700.002443/2017-14" "08700.002938/2017-35" "08700.003237/2017-13"  
## [25] "08700.003244/2017-15" "08700.003247/2017-59" "08700.003249/2017-48"  
## [28] "08700.003344/2017-41" "08700.003699/2017-31" "08700.003709/2017-38"  
## [31] "08700.006006/2017-61" "08700.003855/2018-44" "08700.005375/2018-18"  
## [34] "08700.000556/2019-39" "08700.000881/2019-00" "08700.002290/2019-69"  
## [37] "08700.003910/2019-87" "08700.004287/2019-80" "08700.005020/2019-18"  
## [40] "08700.005876/2019-85" "08700.005714/2020-81" "08700.005726/2020-13"  
## [43] "08700.000413/2021-41" "08700.002012/2021-26" "08700.004914/2021-05"  
## [46] "08700.003153/2022-47"  
## # A tibble: 46 × 2  
## # Groups: numero\_do\_processo [46]  
## numero\_do\_processo status   
## <chr> <chr>   
## 1 08700.008576/2012-81 Idêntica  
## 2 08700.009509/2012-84 Idêntica  
## 3 08700.010979/2013-71 Idêntica  
## 4 08700.005789/2014-13 Idêntica  
## 5 08700.009161/2014-97 Idêntica  
## 6 08700.010056/2014-09 Idêntica  
## 7 08700.011474/2014-05 Idêntica  
## 8 08012.009732/2008-01 Idêntica  
## 9 08700.000448/2015-32 Idêntica  
## 10 08700.000949/2015-19 Idêntica  
## # ℹ 36 more rows  
## [1] "Não foram encontradas diferenças nos processos duplicados."

# Excluir duplicatas mantendo a primeira ocorrência  
df.cade <- df.cade %>%  
 distinct(numero\_do\_processo, .keep\_all = TRUE)  
# Remover duplicados  
rm(duplicados)  
  
write\_xlsx(df.cade, path = "./output/cartel\_internacional.xlsx")  
  
# Filtrar linhas onde conduta\_especifica contém 'internacional'  
df.cade <- df.cade %>%  
 filter(str\_detect(conduta\_especifica, regex('internacional', ignore\_case = TRUE)))  
  
# Ajustar o formato data com o lubridate  
df.cade <- df.cade %>%  
 mutate(  
 data\_de\_autuacao = as.Date(data\_de\_autuacao),  
 data\_decisao\_sg = as.Date(data\_decisao\_sg),  
 data\_da\_decisao\_tr = as.Date(data\_da\_decisao\_tr),  
 data\_decisao\_final = as.Date(data\_decisao\_final)  
 )  
  
# Visualizar o resumo do dataframe  
summary(df.cade)

## numero\_do\_processo data\_de\_autuacao conduta\_geral conduta\_especifica  
## Length:37 Min. :2012-05-07 Length:37 Length:37   
## Class :character 1st Qu.:2014-12-26 Class :character Class :character   
## Mode :character Median :2016-02-17 Mode :character Mode :character   
## Mean :2016-12-27   
## 3rd Qu.:2019-04-18   
## Max. :2024-01-22   
##   
## coordenacao\_da\_sg\_responsavel celebracao\_do\_tcc suspensao\_do\_processo\_por\_tcc  
## Length:37 Length:37 Length:37   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## suspensao\_judicial transito\_em\_julgado data\_decisao\_sg teor\_da\_decisao\_sg  
## Length:37 Length:37 Min. :2015-09-24 Length:37   
## Class :character Class :character 1st Qu.:2019-02-21 Class :character   
## Mode :character Mode :character Median :2020-08-13 Mode :character   
## Mean :2020-03-24   
## 3rd Qu.:2021-09-20   
## Max. :2023-12-22   
## NA's :14   
## cooperacao\_com\_orgaos\_do\_governo cooperacao\_internacional data\_da\_decisao\_tr   
## Length:37 Length:37 Min. :2015-10-14   
## Class :character Class :character 1st Qu.:2019-08-13   
## Mode :character Mode :character Median :2021-04-14   
## Mean :2020-06-08   
## 3rd Qu.:2021-09-18   
## Max. :2023-02-08   
## NA's :18   
## teor\_da\_decisao\_tr houve\_recurso decisao\_final data\_decisao\_final   
## Length:37 Length:37 Length:37 Min. :2020-03-18   
## Class :character Class :character Class :character 1st Qu.:2020-12-12   
## Mode :character Mode :character Mode :character Median :2021-06-16   
## Mean :2021-05-19   
## 3rd Qu.:2021-11-13   
## Max. :2022-05-11   
## NA's :30   
## decisao\_do\_cade origem multas\_aplicadas   
## Length:37 Length:37 Min. : 0   
## Class :character Class :character 1st Qu.: 0   
## Mode :character Mode :character Median : 0   
## Mean : 4714543   
## 3rd Qu.: 0   
## Max. :88205622   
##

# Verificar as classes das colunas de data  
str(df.cade)

## tibble [37 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ numero\_do\_processo : chr [1:37] "08012.005324/2012-59" "08700.008576/2012-81" "08700.009509/2012-84" "08700.010979/2013-71" ...  
## $ data\_de\_autuacao : Date[1:37], format: "2012-05-07" "2012-10-09" ...  
## $ conduta\_geral : chr [1:37] "Cartel;" "Cartel;" "Cartel;" "Cartel;" ...  
## $ conduta\_especifica : chr [1:37] "Cartel Internacional;" "Cartel Internacional; Dificultar continuidade de relações comerciais; Fixar Preço de revenda; Troca de informações sensíveis;" "Cartel Internacional; Preço Abusivo;" "Cartel Internacional;" ...  
## $ coordenacao\_da\_sg\_responsavel : chr [1:37] "CGAA 6" "CGAA 7" "CGAA 7" "CGAA 7" ...  
## $ celebracao\_do\_tcc : chr [1:37] "Não" "Sim" "Não" "Não" ...  
## $ suspensao\_do\_processo\_por\_tcc : chr [1:37] "Não" "Não" "Não" "Não" ...  
## $ suspensao\_judicial : chr [1:37] "Não" "Não" "Não" "Não" ...  
## $ transito\_em\_julgado : chr [1:37] "Sim" "Sim" "Sim" "Não" ...  
## $ data\_decisao\_sg : Date[1:37], format: "2019-05-28" "2022-07-05" ...  
## $ teor\_da\_decisao\_sg : chr [1:37] "Condenação" "Condenação" "Arquivamento" "Condenação" ...  
## $ cooperacao\_com\_orgaos\_do\_governo: chr [1:37] "-" "-" "-" "-" ...  
## $ cooperacao\_internacional : chr [1:37] "-" "-" "-" "-" ...  
## $ data\_da\_decisao\_tr : Date[1:37], format: "2021-03-31" "2023-02-08" ...  
## $ teor\_da\_decisao\_tr : chr [1:37] "Condenação" "Condenação" "Arquivamento" "-" ...  
## $ houve\_recurso : chr [1:37] "Sim" "Não" "Não" "-" ...  
## $ decisao\_final : chr [1:37] "Condenação" "-" "-" "-" ...  
## $ data\_decisao\_final : Date[1:37], format: "2021-03-31" NA ...  
## $ decisao\_do\_cade : chr [1:37] "Maioria" "-" "-" "-" ...  
## $ origem : chr [1:37] "Leniência" "Leniência" "Representação" "Leniência" ...  
## $ multas\_aplicadas : num [1:37] 88205622 732002 0 0 0 ...

# ANALISANDO DADOS ----  
## Calcular quantidade de "Sim", "Não" e Percentual para cada métrica  
  
## Quantidade e percentual de Leniência ----  
### Considera processos cuja 'origem' é 'Leniência'  
leniencia\_counts <- df.cade %>%  
 summarise(  
 Sim = sum(str\_detect(origem, regex('leniência', ignore\_case = TRUE))),  
 Nao = n() - Sim,  
 Percentual = Sim / n() \* 100  
 )  
  
## Quantidade e percentual de Trânsito em Julgado ----  
### Considera processos com 'transito\_em\_julgado' igual a 'Sim'  
transito\_counts <- df.cade %>%  
 summarise(  
 Sim = sum(str\_detect(transito\_em\_julgado, regex('sim', ignore\_case = TRUE))),  
 Nao = n() - Sim,  
 Percentual = Sim / n() \* 100  
 )  
  
## Quantidade e percentual de Arquivamento ----  
### Considera processos cuja 'teor\_da\_decisao\_sg' ou 'teor\_da\_decisao\_tr' é 'Arquivamento'  
arquivamento\_counts <- df.cade %>%  
 summarise(  
 Sim = sum(str\_detect(teor\_da\_decisao\_sg, regex('arquivamento', ignore\_case = TRUE)) |   
 str\_detect(teor\_da\_decisao\_tr, regex('arquivamento', ignore\_case = TRUE))),  
 Nao = n() - Sim,  
 Percentual = Sim / n() \* 100  
 )  
  
## Quantidade e percentual de Condenação ----  
### Considera processos cuja 'teor\_da\_decisao\_sg' ou 'teor\_da\_decisao\_tr' é 'Condenação'  
condenacao\_counts <- df.cade %>%  
 summarise(  
 Sim = sum(str\_detect(teor\_da\_decisao\_sg, regex('condenação', ignore\_case = TRUE)) |   
 str\_detect(teor\_da\_decisao\_tr, regex('condenação', ignore\_case = TRUE))),  
 Nao = n() - Sim,  
 Percentual = Sim / n() \* 100  
 )  
  
## Quantidade e percentual de Multas Aplicadas ----  
### Considera processos cuja 'multas\_aplicadas' é maior que 0  
multas\_counts <- df.cade %>%  
 summarise(  
 Sim = sum(multas\_aplicadas > 0),  
 Nao = n() - Sim,  
 Percentual = Sim / n() \* 100  
 )  
  
# OUTPUT ----  
## Criar um dataframe com os resultados ----  
df\_resultados <- data.frame(  
 "Variável" = c("Leniência", "Trânsito em julgado", "Arquivamento", "Condenação", "Multas aplicadas"),  
 "Não" = c(leniencia\_counts$Nao, transito\_counts$Nao, arquivamento\_counts$Nao, condenacao\_counts$Nao, multas\_counts$Nao),  
 "Sim" = c(leniencia\_counts$Sim, transito\_counts$Sim, arquivamento\_counts$Sim, condenacao\_counts$Sim, multas\_counts$Sim),  
 "Percentual (%)" = c(leniencia\_counts$Percentual, transito\_counts$Percentual, arquivamento\_counts$Percentual, condenacao\_counts$Percentual, multas\_counts$Percentual)  
)  
df\_resultados

## Variável Não Sim Percentual....  
## 1 Leniência 7 30 81.08108  
## 2 Trânsito em julgado 17 20 54.05405  
## 3 Arquivamento 24 13 35.13514  
## 4 Condenação 23 14 37.83784  
## 5 Multas aplicadas 30 7 18.91892

## Exportar para um arquivo .xlsx ----  
write\_xlsx(df\_resultados, path = "./output/percentuais\_cartel\_internacional.xlsx")