

Case 0

Camille Peixoto Almeida

26/03/23

1 Passos e descrições tomados no script do Case0

1. Importamos a biblioteca tidyverse já instalada no R-studio.
2. Lemos o arquivo imdb.rds (base de dados utilizada, no caso, é uma base de filmes com seus anos de lançamento, diretores, gênero, receita, orçamento, nota, atores), baixado na mesma pasta em que o projeto e script foram criados.
3. Retiramos os valores "not available" (N/A), ou seja, espaços em "branco" (faltantes) na base de dados para, por exemplo, facilitar a obtenção e manipulação de resultados.
4. Por indicação do documento oferecido no e-disciplinas, excluimos a coluna país.
5. Por indicação do documento oferecido no edisciplinas sobre o case 0, criamos a coluna idade (idade do filme).

6. Definimos um filme "velho" se possuir idade de 50 anos ou mais e "novo" se possuir menos que 50 anos, por meio do uso do "for (i in 1:2987)".

Observação: originalmente a base imdb.rds possuía um total de 3713 linhas (filmes), porém com a exclusão dos filmes com dados not available temos um total de 2987 filmes com dados totalmente preenchidos. Por esse motivo o uso do "for (i in 1:)" utilizamos o número 2987 de filmes em vez de 3713 para a classificação de filmes velhos e novos.

7. As classificações "velho" e "novo" estarão computadas em uma nova coluna chamada "status".
8. Para visualizarmos melhor, tabelamos ("contamos") o número de filmes classificados velhos e novos. Obtivemos como resultado, 2946 filmes novos e 41 filmes velhos.

2 Script-Case 0

```
# Camille Peixoto Almeida 12702259 - CASE 0

# importar a biblioteca
library(tidyverse)

# selecionar a base de dados
df <- readRDS("H:/Meu Drive/USP/1ºQuadri/reof_estat/
Estudo de Caso 0 Uso do R-20230322/imdb.rds")

# retirar os espaços nulos (NA)
df <- na.omit(df)

# excluir a coluna pais
df$pais = NULL

# definir a coluna idade
df$idade = 2023 - df$ano

# definir a coluna velho ou novo
df$status = "Novo/Velho"

# Caracterizar velho/novo idades >= a 50 (velhos) e < 50 (novos)
for (i in 1:2987) {
  if(df$idade[i]>=50){
    df$status[i] = "Velho"
  }
  else{
    df$status[i]="Novo"
  }
}

# contagem de velhos e novos
table(df$status)

# criar uma base de dados nova as novas modificações
write_rds(df,"banco_de_dados_case0_camille")
```