

PERGUNTAS REALIZADAS DURANTE A AULA

Curso: MBA em Data Science e Analytics

Disciplina: Data Wrangling I

Data: 09/11/2021

Guilherme Tavares Bassetto

Boa noite professor, Uma duvida. Como fazer uma realocação para deixar este exemplo como uma visão dimensional ?

Resposta: Boa tarde! Desculpe, não entendi a pergunta. O que seria uma visão dimensional?

Luis Filipe Souza de Almeida

caso eu queira extrair linhas ou colunas que não estão em ordem eu uso ";"?

Resposta: Boa tarde! Teria que ser feito com a função `c()`, ou seja, combinando em um vetor. Por exemplo, excluir as linhas 3, 5 e 7 da base de dados: `dataset[-c(3, 5, 7),]` ou excluir as colunas 1 e 5: `dataset[, -c(1, 5)]`. O argumento para linhas fica antes da vírgula e colunas após a vírgula.

Wilber Molina

é possível realizar um Select com [coluna:coluna]?

Resposta: Boa tarde! Sim, seria algo como `select(nova_base, observacoes:distancia)` na linha 291 do script 1.1. Neste caso, são selecionadas todas as colunas que estão da coluna observações até a coluna distância.

Igor Fillippe Goldstein

e o que é importante levar em conta exatamente pra definir qual type usar no cut?

Resposta: Boa tarde! O type é um argumento da função "quantile", isto é, alterando o type altera-se a forma de cálculo dos quantis da amostra. A documentação da função quantile apresenta 9 fórmulas de cálculo disponíveis. É uma escolha específica de acordo com a aplicação que deseja. Note que a função cut pode ser utilizada sem que exista o quantile como argumento dela.

Vinicius Fernandes Lourenzen de Oliveira

error: package or namespace load failed for 'tidyverse' in loadNamespace(j - i[[1L]], c(lib.loc, .libPaths()), versionCheck = vI[[j]]): there is no package called 'glue'

Resposta: Boa tarde! Parece algum erro na leitura do tidyverse. Poderia tentar instalar novamente o tidyverse - `install.packages("tidyverse")` - e verificar se carregando novamente ele funciona.

Cristiane Edna Cammboim

O vetor precisa conter os novos nomes na mesma sequencia das variaveis da base? Quando voce nao quer alterar um nome, por exemplo, qual o procedimento?

Resposta: Boa tarde! Sim, tem que estar na mesma sequência. Se não quiser mudar, é mais fácil só repetir o nome antigo dentro do vetor. Lembrando que o rename faz a tarefa mais facilmente.

Yuri Silva de Souza

Boa noite professor. Nesse seu exemplo da linha 188 script (1.1) qual é a diferença em inserir a base de dados antes de um pipe, ao invés de colocar esse objeto direto dentro da função mutate? Valeu!

Resposta: Boa tarde! A diferença é que o uso do pipe permite encadear diversas funções em sequência. Por exemplo, na linha 358 do script 1.1 foi usado o pipe e três funções encadeadas: group_by, summarise e arrange. Quando a base é inserida na função específica, só aquela função será executada.

Thaís Brenda Martins

Professor, porque não remover espaços, maiúsculas e acentos dos nomes das colunas?

Resposta: Boa tarde! Sim, é sempre importante fazer estes ajustes. Facilita a escrita e evita erros.

Murilo Brantegani Grizi

Por que algumas variáveis no comando rename estão com aspas, e outras não??

Resposta: Boa tarde! Utilizamos as aspas nas variáveis que contêm espaços (mais de uma palavra), assim, o comando “entende” que trata-se de um único termo. Também pode ser usada, sem problemas, nos casos em que é só uma palavra.

Walter Luiz Caram Saliba

No exemplo da linha 298, a função everything() não deveria repetir os campos observacoes e perfil?

Resposta: Boa tarde! Neste caso, o comando “entende” que foi solicitada uma mudança de posição das colunas, ou seja, as duas variáveis nomeadas antes e “todas as demais” na sequência. Por isto, não repete as variáveis já especificadas por nomes.

Igor Fillippe Goldstein

no nova_base %>% rename (...), a variável nova_base é alterada ou ele só faz print das colunas renomeadas sem mudar a original? não ficou claro pra mim

Resposta: Boa tarde! A “nova_base” é a base de dados, não a variável. Sim, se não for especificado um novo objeto com “novo objeto <-” então só será feito o print das colunas renomeadas e nada será alterado no objeto original “nova_base”. Quando especificar um novo objeto ou mesmo reescrever no antigo, então a mudança fica gravada.

Yuri Silva de Souza

Por que você usa o \$ para acessar a coluna da tabela sendo que já usou o pipe antes? Linhas 22:25 exemplo do jogo.

Resposta: Boa tarde! Naquele caso, é uma forma de escrita específica dentro da função “case_when” em que está sendo feita a subtração de duas variáveis em específico. Por isto, foi especificada a coluna com o \$.

Afonso Henrique Santos Silva

Se ao utilizar o pipe minhas colunas tivessem com espaço, por exemplo 'tempo percorrido' eu precisaria usar as aspas no rename?

Resposta: Boa tarde! Sim, se o nome da variável contém espaços, especifique-a entre aspas.

Paulo Kauai Amorim Pinto

Poderia exemplificar uma maneira de separar as strings das escalações dos times para contabilizar a quantidade de partidas disputadas por jogador?

Resposta: Boa tarde! É possível fazer com a mesma função utilizada para a extração do gol. Basta identificar o padrão existente no caso das escalações.

Seria este:

```
extraí_jogador_casa <- str_extract_all(jogos$lineup_home,
  "player_name": "\\w*(.*?)\\w*\\",
  simplify = TRUE)
```

Na sequência, seria fazer as mesmas “limpezas” feitas no caso do gol e depois fazer o mesmo procedimento para a variável jogadores visitantes e a tabela de frequência para as duas em conjunto (da mesma forma que foi feito no script).

Por acaso, fiz e há dois jogadores da base toda escritos fora deste padrão geral. Infelizmente, pode acontecer e como este comando trabalha buscando padrões, então o que está escrito de forma diferente acaba ficando de fora. Estes adicionais podem ser encontrados por:

```
extraí_jog_ad <- str_extract_all(jogos$lineup_home,
  "player_name": "\\w*(.*?)\\w*\\",
  simplify = TRUE)
```

E para limpar:

```
extraí_jog_ad <- gsub("player_name": ", ", extraí_jog_ad)
extraí_jog_ad <- gsub("\\", "", extraí_jog_ad)
```

Depois, é só fazer a tabela de frequências com as 3 informações geradas.

Tarcizio da Silva Barbosa

No Exercício 5, nas linhas 22 a 24, por que a subtração esta sendo concatenada com o c()?

Resposta: Boa tarde! Nenhum motivo em especial, foi simplesmente uma forma de identificar mais facilmente os três critérios que estavam presentes na base. É possível retirar os c(), o comando funciona normalmente.