

Importando a base de dados

Transcrição

[00:00] Vamos continuar o nosso projeto de análise utilizando o R, só lembrando aqui qual que é o nosso projeto de análise. Imaginemos que nós somos o cientista de dados de uma empresa de estratégia e essa empresa, ela está querendo expandir os seus negócios.

[00:18] Então, a área de planejamento passou para você uma base de dados de valores de aluguel e venda de imóveis, para você avaliar. Você, como cientista de dados, você vai avaliar essa base de dados e gerar informações para a empresa, para ela tomar decisão, justamente em que imóveis ela vai alugar ou comprar, em que bairro, que valor.

[00:45] Então, esse é o nosso desafio. A gente está aqui com o R aberto, o RStudio, que é a nossa plataforma que a gente vai avaliar, vai analisar os dados, só lembrando aqui, a gente vai usar aqui dentro do RStudio, eu vou abrir o R Notebook, que é aonde a gente vai estar fazendo o nosso projeto de dados.

[01:03] Então, eu abro um novo. Quando a gente cria, ele está... isso pode acontecer, se [XXIninteligívelXX], ele instalou automaticamente alguns pacotes, lembrando que o R, sempre ele... tem pacotes e bibliotecas que ele carrega, se ele não carregar, a gente tem que pedir para ele carregar.

[01:21] Quando ele abre a tela, ele dá algumas dicas do R Notebook. O R Notebook é uma linguagem R Markdown, tem mais informações. Algumas teclas de atalho aqui, como o “Ctrl + Alt + I”, para inserir o chunk de R, o código R. Então, eu vou... a gente pode limpar essa parte aqui, que ele dá só como dica, para ficar mais limpo o nosso projeto.

[01:43] Então, a primeira coisa que a gente faz para inserir aqui... Então, a gente vai colocar aqui um comentário que é aqui é o seguinte: a gente precisa fazer o quê? Carregar os pacotes necessários para a gente fazer a análise, antes de fazer a importação da base de dados.

[01:59] A gente vai importar essa base de dados para dentro do R, mas antes disso, a gente precisa de fato carregar as bibliotecas que a gente vai utilizar. Então, para carregar... vamos carregar... “#Carregando pacotes”. Então, o primeiro pacote que a gente vai carregar aqui, é o pacote de leitura.

[02:27] Eu vou ensinar para vocês como é que instala um pacote no R, como é que lê essa biblioteca e também, depois tem uma outra maneira aqui, que a gente pode fazer um click point, que fica até mais fácil de fazer essa importação. Lembrando, então, para tudo o que a gente escreve aqui fora, a gente escreve como se fosse um relatório mesmo.

[02:46] A gente precisa inserir esse código, a área de código, que a gente chama de chunk, então é um chunk R. Lembrando que, a gente pode... o notebook, ele propicia outras linguagens, até para a gente fazer esse projeto, mas a gente vai ficar aqui no R. Então, tudo o que a gente escrever aqui dentro, ele vai interpretar como código, dentro desse “{r}”, três pontinhos aqui.

[03:06] Então, se a gente quiser comentar uma expressão aqui dentro, como a gente não quer que rode aquela expressão, a gente coloca tracinho, tracinho aqui... tracinho não, pode colocar hashtag aqui dentro, só que ele vai comentar também. Então, eu vou colocar aqui.

[03:23] A primeira coisa, a gente precisa fazer a instalação dos pacotes, vamos instalar o primeiro pacote aqui, é... instalação do pacote chamado readr, o que que esse pacote faz? Justamente, ele propicia a leitura de alguns arquivos em txt., xls., alguns formatos.

[03:51] Para a gente instalar pacotes aqui no R, a gente tem três maneiras diferentes de fazer isso, eu vou explicar rapidamente as três para vocês. A primeira, a gente pode vim aqui pelo menu, Tools, Install Packages e aí, a gente... o que que ele vai fazer?

[04:08] Digita aqui o pacote, então o pacote que a gente vai querer é o readr, veja que ele preenche já, ele é "autopreenchível", se a gente esquecer o pacote e lembrar, aqui a gente consegue ver. Clicou aqui no nome, o que que ele vai fazer? Ele vai no repositório CRAN, que é o principal repositório, onde tem todas as library, todos os pacotes do R na nossa entrada, num servidor.

[04:30] E ele já vai fazer essa instalação automática, é só dar um install, ele já vai pegar e fica instalando. O console instalou rapidamente. Feita essa instalação, dessa maneira por click point, a gente pode também fazer ela pela linha de códigos.

[04:46] Se a gente fosse fazer pela linha de códigos, daí eu vou colocar "install.packages()", aqui dentro do parênteses, a gente coloca entre aspas o pacote que a gente quer carregar, é o readr e aí, executaria essa linha. Para executar o que está dentro do chunk, a gente pode... de novo, lembrando, pode vir aqui e executa todo o bloco no chunk, esse play ou então, eu dou o "Ctrl + Enter" na linha.

[05:11] Se eu der o "Ctrl + Enter", ele vai fazer a mesma coisa, vai instalar o pacote, a minha instalação. Instalou com sucesso. Depois da instalação, sempre quando a gente termina a instalação de um pacote, para usar, a gente precisa fazer o que? Carregar aquela biblioteca.

[05:29] Então: "library(readr)", está vendo que ela já está disponível aqui? Então, está vendo que aparecer o "readr" e o "readxl"? Que proízia carregar arquivos diretamente em Excel. Então, "readr", se eu executar o pacote aqui, eu vou comentar a linha, agora eu vou comentar essa linha, porque a gente não vai instalar de novo.

[05:52] Eu vou executar agora pelo play, ele vai executar toda essa linha aqui, todo esse grupo, todo esse grupo do setor chunk. Executou. Carregou. Não deu nenhum erro. Uma outra maneira... eu falei que tinha três, essa é a segunda maneira de instalar, tem uma outra maneira, que a gente pode ir pelo menu direito também, está vendo aqui?

[06:13] Nessa aba, é onde fica um diretório de arquivos, que a gente está utilizando, quando a gente fizer plot, ele vai aparecer aqui, tem o Help, que a gente pode também... o Help, sempre quando você quiser, por exemplo, acessar o Help de alguma palavra, o que que a gente faz?

[06:29] Clica em cima da palavra, da F1, que ele já vai para cá, veja que ele já abriu automaticamente um documento sobre aquela biblioteca, isso aqui é bastante útil também para saber mais sobre aquele pacote, aquela biblioteca. Então, essa parte de packages, poderia ser instalada por aqui também: Packages, Install, vai para o mesmo lugar também.

[06:55] Então, vamos na sequência, agora, já vou adicionar outro chunk, esse chunk, lembra que eu falei que a gente pode usar a tecla de atalho: "Ctrl + Alt + I"? Aqui, eu digitei "Ctrl + Alt + I", ela já apareceu automaticamente, eu não precisei ir por aqui, isso aqui facilita as vezes o código.

[07:14] Eu vou colocar mais um comentário aqui fora, vou colocar aqui: "#importando a base de dados", vamos importar agora para o R a base de dados. Para fazer essa importação, existem duas maneiras também, a gente pode fazer tudo pela linha de códigos ou então, também existe uma maneira bem intuitiva e rápida de fazer essa importação pelo click point, que eu vou mostrar para vocês até mais rápido.

[07:47] Para importar a base de dados, o que que acontece? A gente pode ir nesse menu aqui em cima, na parte Environment, nessa parte de cima tem Environment, History, ele vai aparecer tudo o que a gente já fez, ele vai aparecer (num log) do que a gente está fazendo.

[08:02] Se a gente quiser apagar esse (log), a gente clica na vassourinha aqui, ele apaga, ele vai falar, apagou, ficou limpo. A parte de Environment, é a parte onde aparecem as variáveis, quando a gente estiver criando as variáveis, carregando as bases, ela vai aparecer aqui.

[08:17] Para importar a base de dados, eu vou fazer o seguinte, eu vou pelo click point aqui, mais rápido. Então, (ele importa da) Dataset, eu vou clicar aqui no Import Dataset, o meu arquivo é um arquivo txt., que vai estar aí disponível para vocês também, é uma base de aluguel.csv.

[08:37] Então, ele já clica aqui, ele aponta para onde eu vou carregar a base, essa base, ela está no meu diretório chamado documentos, aqui onde ela vai estar disponível, eu vou carregar. Aponte... um momento, espera aí, voltando. Na verdade, se eu fizer isso... Eu vou voltar aqui de novo, Import, eu vou usar readr, estava base, não... eu vou usar exatamente a biblioteca que a gente usou: readr, aqui "From Text (readr)", isso, ok.

[09:06] Agora, aparece essa janela aqui, essa janela é bem intuitiva, ele dá o código que a gente está utilizando. Então, o que que a gente faz? Vai lá e aponta o caminho da base, vai em Browse, a minha base está aqui em documentos, aluguel.csv, vai estar disponível aí para vocês também.

[09:25] Carregou a base, quando carrega a base, o que que acontece? Ele aponta que o caminho... ele não coloca o caminho inteiro, está direto no documento, então ele tem um "~/" aqui, mas vai apontar automaticamente quando tiver a base.

[09:37] Veja que ele trouxe num bloco só isso aqui, o importante é ver que existe um delimitador aqui. Essa base, como eu comentei, ela tem o tipo do imóvel, o bairro onde está localizado aquele imóvel, a quantidade de quartos, vagas, se paga IPTU, se tem condomínio, valor do imóvel e valor de aluguel do imóvel.

[10:01] Existe um separador aqui, então a gente precisa definir esse Delimiter, eu vou colocar aqui como Other, eu sei que o delimitador aqui é ";", porque dá para ver aqui. Quando eu coloco ";", olha, ele arrumou a base, agora sim, então é: Tipo, Bairro, Quartos, Vagas, Suítes, Área, Valor, Condomínio, IPTU.

[10:22] Aqui do lado, essa funcionalidade é muito boa do RStudio, quando você está importando a base, porque ele já carrega exatamente o comando que você vai usar. Eu vou copiar esse comando e vamos jogar lá na minha área de trabalho. Então, eu vou dar o Import aqui, ele vai carregar a base, ele vai executar.

[10:38] Carregou a base. Quando ele carrega a minha base para o R, ele abriu a base automaticamente, por quê? Porque a gente já pediu para dar um view na base e automaticamente, o que que ele fez? Ele veio na parte Environment e colocou a base... algumas informações da base: aluguel, tipo, bairro, outras variáveis, então dá para dar uma visualização que carregou de fato a base.

[11:02] Eu vou fechar aqui, dar um clique, fechando aluguel, quando eu fiz isso, a base já está aqui dentro do R, eu não... Ela está exatamente aqui dentro, se eu chamar essa base, ela sempre vai estar. Lembra que eu coleí o código... copieí o código, eu vou copiar ele... eu vou colar ele aqui, para vocês verem que é exatamente como se a gente fosse fazer na mão.

[11:19] Então, o que que eu fiz? Acionou a library(readr), que a gente estava escrevendo aqui. Aí, a gente apontou aqui aluguel, que é uma variável, que é exatamente o aluguel, vai receber o quê? Ele vai usar o "read_delim", que é o delimitado, para a minha base aluguel.csv, com o delimitador que eu defini, que era o ";".

[11:40] E aí, ele deu alguns outros comandos aqui, que não necessariamente eu precisaria colocar, se eu colocasse só isso aqui: aluguel.csv e ";", que é principalmente o nome da base e o delimitador e vou... vou rodar de novo isso aqui, eu vou apagar a base, para apagar a base, clica nessa vassourinha, que ele vai limpar.

[12:04] “Tem certeza que quer limpar?”, sim, pronto. A minha base já não está mais aqui no R. Agora, eu vou executar pelo código aqui do chunk, vou executar todo o bloco. Executei. Carregou, está vendo? Por quê? Porque eu já dei o comando aqui.

[12:21] Ele já sabe para onde tem que carregar, por que que... Aqui, ele já... não precisa, não estou colocando nenhum tipo de diretório, porque ele já está indo e entendendo onde que está na minha raiz, a minha raiz está dentro de documentos e já vai e lê aqui.

[12:37] Quando eu dou esse View(aluguel) aqui, eu vou dar só o “Ctrl + Enter” no View(aluguel), é exatamente, ele vai a base. Então, abriu a base de dados que a gente vai analisar. Só lembrando, voltando, antes de explicar e comentar um pouco a nossa base de dados, então, o nosso objetivo, qual que vai ser?

[12:54] O nosso objetivo será avaliar essa base quanto ao tipo de imóvel, ao bairro que aquele imóvel se encontra, o número de quartos. Então, a gente vai receber... a gente recebeu algumas definições da área de planejamento, aonde a gente vai fazer algumas seleções, alguns filtros, para gerar justamente informação para a área de negócios, de planejamento e dar essas respostas da análise que a gente vai estar fazendo aqui no R.

[13:29] Então, esse aqui, a gente vai ver no próximo vídeo. Então, por enquanto é isso. Até mais.