

Descrição do projeto

Transcrição

[00:00] Então, vou aproveitar aqui o desafio desse projeto. O desafio desse projeto, a gente recebe uma base com valores de imóveis, valores de aluguel e venda, você está trabalhando numa empresa como cientista de dados.

[00:17] Você precisa fazer a análise dessa base de dados e encontrar oportunidades e ajudar a empresa a dar respostas quanto à estratégia de expansão dela, verificando nessa base, se existem oportunidades e imóveis de interesse para a empresa.

[00:38] Nós vamos estar realizando essa análise no R e para isso, a gente precisa preparar o ambiente antes. Para preparar o ambiente, nós precisaremos fazer download de alguns arquivos. O primeiro arquivo que a gente vai precisar fazer é a instalação do próprio R.

[00:58] Para fazer a instalação do R, a gente baixa aqui, só digitar “R download”, o primeiro link que aparecer, a gente vai baixar esse arquivo. A versão do R mais recente é a 3.61, o R é um software livre, free, bastante divulgado e bastante utilizado por cientista de dados.

[01:17] A gente primeiro vai fazer a instalação do R, depois a gente vai fazer a instalação do RStudio, o RStudio é uma versão onde tem uma interface mais agradável, que eu vou mostrar para vocês, facilita um pouco mais a nossa... a execução da análise de dados.

[01:37] Então, a gente vai fazer primeiramente o download do R, feito o download do R, a próxima parte é só instalar, ele vai baixar esse “R-3.6.1-win”, vai utilizar como padrão a instalação do Windows, é bastante simples para fazer essa instalação, é só dar próximo que aceita os termos de uso.

[01:59] Não existe nenhum tipo de cobrança, porque acaba sendo um software livre, eu recomendo que na instalação, você deixe os diretórios padrões que estiverem apontados no momento da instalação, não mude. Instale a versão 32 a 64 bits, também não tem problema, avança.

[02:23] Aceita a instalação padrão, acaba sendo um pouco mais fácil de fazer essa instalação e também não muda o diretório que for apontado. Eu não vou continuar a instalação aqui, porque já está instalado na minha máquina, vou dar um cancelar, mas vocês façam essa instalação até o final e depois vai ser gerado um atalho, depois eu mostro para vocês.

[02:47] O próximo arquivo que a gente... eu vou sair daqui então, o próximo download que a gente precisa fazer a instalação, é do próprio RStudio. Para fazer a instalação do RStudio, também a gente pode buscar no Google o “RStudio download”, provavelmente o primeiro ícone que aparece aqui, ele vai entrar na página do RStudio.

[03:10] O RStudio, existem outras versões que são pagas, tem aqui... existem suportes e tal, mas a gente vai utilizar essa versão do “RStudio Desktop”, que ela é free, não existe nenhum tipo de cobrança. Rolando para baixo, clica em download, ele vai aparecer aqui em baixo as versões para Windows, Mac, Linux.

[03:33] Como padrão, a gente vai utilizar o Windows, mas se alguém quiser instalar alguma outra versão aqui, também pode instalar e não tem problema. A gente vai baixar esse arquivo da última versão disponível, 64 bits. Ele, terminando de fazer o download, ele vai baixar esse arquivo aqui: “RStudio – 1.2.1335”.

[03:57] Para executar esse arquivo, também é bastante simples, é só dar próximo aqui também, não existe nenhuma configuração adicional, é só dando próximo, próximo. A gente também mantém o diretório que o próprio software está sugerindo. Avança, ele vai criar essa versão de RStudio, vai dando [XXIninteligívelXX], depois é só dar o instalar.

[04:20] Também não vou instalar aqui, porque senão ele vai subscrever o software que está lá, mas vocês instalem e após a instalação, a gente... vai aparecer esse seguinte ícone, vai aparecer tanto o R, quando o RStudio. Vou mostrar para vocês rapidamente a cara do R aqui.

[04:38] A gente vai... como eu falei, a gente vai usar o RStudio para fazer as nossas análises e ver, mostrar para vocês qual que é a vantagem de utilizar o RStudio sobre o R. A princípio, a gente poderia até usar o próprio R, mas o jeito de fazer a análise no R, seria bem mais complicado por causa dessas funcionalidades que não existem nessa versão do R.

[05:02] O RStudio, ele roda em cima do R, como eu falei, é só uma interface de melhoria, mas eu vou mostrar a cara do R rapidinho, abrindo ele aqui, então ele abre essa cara aqui, bem seco. No fundo, ele abre só o Console, R Console, esse R Console é um ambiente de programação mesmo.

[05:19] Então, só um exemplo rápido aqui, por exemplo, se a gente fizer uma conta que “A” recebe 1, enter; “B” recebe 3, enter, se eu fizer a operação “A + B”, ele vai fazer a soma, vai dar 4, 3 + 1, 4. Então, ele é um console de programação mesmo, então teria que fazer toda a programação aqui em cima.

[05:45] A vantagem de se instalar o RStudio é a seguinte, vou abrir o RStudio aqui. Aqui, quando abre o RStudio, ele abre também o mesmo console que a gente abriu ali no R puro, mas aqui, ele já dá algumas outras opções. Basicamente, o RStudio, ele é dividido como?

[06:01] Ele tem a parte de console, que essa aqui, que é exatamente igual ao do R puro, mas aqui ele já dá umas outras funcionalidades, já dá um gerenciador de arquivos que dá para a gente navegar, ele aparece uma pasta de Plots, depois aqui, eu vou mostrar para vocês.

[06:17] Os Packages, para ser instalados aqui ao longo do (template), estão instalados e depois a gente vai instalar outros. Tem uma parte de Environment aqui, que é a parte das variáveis... ao longo, quando a gente for fazendo a programação, vão aparecendo.

[06:33] Salvar, importar a base de dados, aquilo que posteriormente, também vou mostrar como é que possibilita essa importação via click point de diferentes tipos de arquivos, via texto, Excel, SAS e outros bancos e a grande funcionalidade... Então, essa visualização desse framework aqui já é bem melhor.

[06:55] Outra vantagem muito boa que eu vejo de utilizar o RStudio, é que o seguinte, a nossa análise, a gente vai usar o R Notebook, se a gente estivesse utilizando o R, a gente não conseguiria fazer a análise, utilizando essa análise de... essa interface de R Notebook ou R Markdown também, que é uma linguagem marcada por marcadores, que facilita bastante o modo de você fazer o projeto, de você escrever o projeto.

[07:22] O R Notebook, quando a gente clica, ele vai abrir uma janela nova aqui, abriu, já vou... Então, como é que funciona o R Notebook? Ele abre aqui rapidamente uma instrução para depois, quem quiser dar uma olhada no “rmarkdown.rstudio”, ver como é que funciona.

[07:42] E eu vou apagar essas recomendações, só para ficar mais limpa a nossa área aqui de desenvolvimento. Então, como é que funciona essa página de R Notebook? É como se fosse um notebook com anotação de notas mesmo, então, a gente pode escrever.

[07:59] A gente escreve, por exemplo, aqui, como se a gente estivesse escrevendo um projeto de análise de imóveis, que vai ser a nossa base, que a gente vai avaliar. Aqui, então, ele escreve como se a gente estivesse escrevendo um texto normal.

[08:24] Existe... como é que a gente roda um... como é que a gente executa o código aqui dentro do notebook? Existe essa parte de Insert, o que que acontece? A gente consegue executar diferentes códigos aqui, a gente vai utilizar o R, mas a gente até... existe a possibilidade, fazendo algumas configurações de executar outros códigos aqui também, é... Python, até mesmo o Python, SQL e algumas outros códigos.

[08:51] Mas como a gente vai trabalhar com o R, então a gente vai inserir esse novo... é... sempre ele insere esse chunk, que é chamado de chunk, o que que é o chunk? O chunk é onde exatamente você vai abrir a biblioteca e fazer a programação. Tudo o que você escrever dentro desse chunk, ele é o interpretador de programa.

[09:15] Então aqui é o programa, se a gente fizer como a gente fez... como a gente estava fazendo, por exemplo, “A” recebe 1; “B” recebe 3, se eu fizer... “C” recebe “A + B”, ele vai interpretar isso aqui como programação. Como é que a gente executa esse programa?

[09:41] A gente pode executar dando “Ctrl + Enter”, apertando o “Ctrl + Enter”, ele vai executar ou então nessa teclinha aqui do lado, nesse play, que ele dá o Run Current Chunk, ele vai executar somente esse bloco que estiver entre esse comando interpretador. Eu vou executar aqui, ó, ele executou, está vendo?

[10:07] E aonde que ele joga essas variáveis? Aonde que ele vai aparecer? Ele vai jogar aqui, ó, nas variáveis globais. Então, “A” é igual a 1, “B” é igual a 3, “C” é igual a 4, ele executou, interpretou exatamente o “C”, fazendo a soma de “A + B”. Então, sempre quando a gente for executar o nosso projeto, a nossa interpretação do R, o que que ele vai fazer?

[10:26] Ele vai estar jogando aqui do lado direito. A gente vai conseguir, vendo essa execução das variáveis que a gente está guardando e quando a gente for fazer algum gráfico, alguma tabela, ele vai aparecer aqui no próprio relatório, que eu vou mostrar para vocês exatamente na sequência [XXIninteligívelXX] no nosso projeto, na base de dados, que é o próximo assunto, certo? Ok?