

02

Entendimentos Iniciais

Transcrição

[00:00] Vamos voltar na nossa análise de projeto de dados utilizando o R, só relembrando, a gente já carregou a nossa base de dados para dentro do R, a base de dados de aluguel.

[00:15] Lembrando que o nosso objetivo é justamente fazer uma análise de modo exploratório e para gerar insumos para outra área da empresa, área de planejamento, que ela está pretendendo fazer uma expansão da empresa e precisa alugar novos imóveis.

[00:31] Então, ela precisa entender quais são imóveis disponíveis, o tipo, casa, apartamento, que bairro que está e os respectivos valores. Então, a ideia desse projeto é justamente gerar insumo analítico para a tomada de decisão por essa área de planejamento.

[00:49] A base está dentro do R, a gente já carregou nessa base com 32.960 observações, o próximo passo que a gente vai fazer é o seguinte: a gente vai fazer o entendimento melhor da base. Então, a gente vai... vou colocar aqui: “#entendendo a base de dados”.

[01:15] Esse entendimento da base de dados, que a gente vai estar fazendo, eu vou abrir um novo chunk, então o primeiro comando no R que a gente pode estar utilizando, que a gente vai utilizar para entender a estrutura da base... se a gente olhar no Environment, ele já dá algumas dicas de [XXIninteligívelXX], a gente consegue ver o número de observações, nove colunas, a gente conseguiria olhar por aqui, mas a gente... vamos olhar pelo comando.

[01:50] Então, se a gente der... olhar o comando “str”, o nome da base que é aluguel, que a gente salvou essa variável aluguel, quando a gente carregou. Então, eu vou executar o chunk, o que que ele traz? Ele vai trazer a classe das variáveis, o nome da... Aqui é um data.frame, esse é o data.frame, é o nome de... quando é um tipo de tabela que ele carrega.

[02:17] Tem 32.960 observações, nove variáveis, que são os nove campos, o tipo, quais são essas nove variáveis? Esses nove campos. Então, tem o tipo, o que que é o tipo? É “chr”, então, se é uma quitinete, casa, conjunto comercial, sala, apartamento, etc.

[02:39] O bairro que está localizado aquele aluguel, Copacabana, Jardim... lembrando que essa base, é uma base de aluguel do Rio de Janeiro, então vão ter bairros aqui disponíveis no Rio de Janeiro. Número de quartos disponíveis, se tem vaga, se não tem vaga, quantas vagas, suíte, se tem suíte, quantas suítes.

[02:57] Área em metros quadrados, o valor do respectivo aluguel, se tem condomínio, valor do condomínio e se tem IPTU. Então, essa é a estrutura da nossa base de dados. Tem uma outra forma de visualização da nossa base de dados, que ela é até um pouco mais amigável, um pouco melhor.

[03:19] Então, se a gente está trabalhando com uma base de dados muito grande, nesse caso é uma base pequena, até 32.960, mas se a gente está trabalhando com uma base maior, então a gente pode estar visualizando as primeiras linhas.

[03:32] Quando a gente abre pelo comando view, eu até vou abrir de novo, vou rodar de novo aqui, coloquei em cima da... na linha aqui, eu vou executar só essa linha do view, ele vai abrir a base toda, como se tivesse a planilha mesmo, ele abre, até o formato é diferente.

[03:47] Então, aqui eu consigo rodar a base. Então, essa base não é uma base de dados muito grande, o meu exemplo, mas a gente poderia estar trabalhando com base de milhões de linhas. Então, quando a gente fosse abrir esse view aqui, ia demorar bastante.

[03:59] Um outro jeito, é usar o comando head, se a gente usar esse comando head, a minha base aluguel, eu vou executar só essa linha, veja o que que ele faz, ele até dá uma tabela com as primeiras seis linhas, então até a forma de visualização é melhor.

[04:21] Aqui em baixo, ele dá o tipo da variável, então “chr”, se a gente consegue visualizar o tipo. As variáveis, a gente consegue visualizar aqui, ir para o lado. Então, tem uma setinha que a gente consegue andar aqui.

[04:36] Para baixo, a gente não vai rodar, porque de novo, como ele está pegando só as primeiras linhas da tabela, então a gente vai visualizar, mas já tem uma ideia dos campos, que aqui é campos numéricos, tem algumas variáveis aqui que não tem, que tem ausência de valor.

[04:51] No R, quando aparecer esse NA, é que não tem valor disponível, depois a gente vai fazer um tratamento, eu vou explicar como é que a gente vai fazer esse tratamento desses NAs. Então, é essa parte de entendimento de base de dados aqui, que a gente já consegue visualizar.

[05:07] Agora, uma outra coisa que... a gente começou o nosso projeto de dados, a gente já está escrevendo aqui no relatório as etapas do projeto, como eu falei, a primeira etapa, era a parte de importação, depois entendimento. Uma coisa muito importante que a gente não fez ainda é o seguinte, vamos salvar esse nosso relatório, que esse relatório, a gente vai levar até o final da nossa análise.

[05:32] Para salvar, para ele abrir o documento aqui... o primeiro nome, é simplesmente, a gente pode ir aqui pelo menu, salvar ou pela... aparece a tecla de atalho aqui, que é o “Ctrl + S” também. Eu vou salvar esse documento na minha pasta aqui, ok, eu vou salvar isso aqui como... eu vou salvar em cima do meu projeto: “projeto de análise – base aluguel”, salvando.

[05:59] Eu já tinha um nome, vou salvar em cima. Então, ele salvou aqui “projeto de análise – base aluguel”, é onde a gente vai fazer toda a nossa análise. Vou mostrar um outro recurso bastante interessante também, antes de prosseguir aqui, que quando a gente salva o R Notebook, a gente já consegue dar um Preview do nosso projeto.

[06:20] A gente clica no Preview ou a tecla de atalho, ele dá a tecla de atalho que é o “Ctrl + Shift + H”, eu vou clicar em cima do Preview, então ele vai executar com tudo o que a gente já fez até o momento, veja que os comandos que a gente deu... Ah, antes disso, espera aí, eu só vou... como te uma parte de... Eu vou executar esses dois chunks aqui.

[06:44] Então ele vai... quando você executa... o melhor é separar até... quando ele fizer isso, ele vai executar um em cima do outro, então eu vou separar esses dois aqui. Eu vou criar um outro chunk aqui em baixo. Não precisaria exatamente fazer isso, mas só para ficar um pouquinho mais estruturado.

[07:07] Então, aqui é “Visualizando as primeiras linhas”, aqui é o head, eu vou executar aqui. Aí, eu vou dar o Preview de novo aqui. Ele começa a aparecer exatamente tudo o que você está fazendo, ele vai guardando esse código e vai guardando os outputs também, os respectivos outputs.

[07:38] Está salvo, depois a gente melhora essa visualização aqui também ao longo da estrutura. A gente já fez o entendimento da base de dados, então o próximo passo que a gente vai fazer é exatamente começar a análise, a gente vai receber algumas demandas lá da área de planejamento.

[07:57] E aí, a gente continua fazendo exatamente... resolvendo essas [XXInteligívelXX], filtrando os imóveis de interesse, se a área quer que a área quer a gente verifique, análise só apartamento ou quer que a gente verifique só casa,

o número de quartos, então a gente vai receber alguns inputs da área lá de planejamento, para a gente fazer essa seleção e tratamento da base dados.

[08:23] A gente continua no próximo vídeo, ok? É isso aí.