

2.3 - Entendendo o que são dados

00:00:00:10 - 00:00:20:24

Olá pessoal! Sejam muito bem-vindes ao nosso curso de Análise de Dados, essa jornada maravilhosa que a gente tem pela frente. Ah, gente, só antes de começar, eu preciso aqui falar urgente com a lana. Deixo eu abrir aqui minha lista telefônica, digitar lana. Ah, pronto, achei o número dela. Ainda bem que eu tinha o contato salvo. Na verdade, eu não preciso falar com a lana.

00:00:21:05 - 00:00:41:26

Mas você sabia que uma lista de contatos também é um dado? Pois é. Com a lista telefônica eu vou ter um conjunto de telefones e nomes associados, que são os meus contatos. Quando a gente organiza os números e nomes e eu consigo perguntar para o meu dado qual o número da lana ou quantos contatos eu tenho, meu dado passa a ter significado e vira informação.

00:00:42:09 - 00:01:07:05

Por exemplo, você quer falar com seu amigo... Então você procura o nome de fulane. E aí vem um número delu para você chamar. Ou também: preciso falar com a lana, mas ao invés do telefone eu quero o e-mail dela. Pode ser que na minha lista de contatos tenha outro atributo, que é o e-mail. Então também consigo essa informação. Esse exemplo mais pertinho da gente é pra gente poder entender o que são conjuntos de dados e que ele está no nosso dia a dia.

00:01:07:14 - 00:01:30:29

Os dados estão aí. Em resumo, então temos que dados são um conjunto de valores brutos atribuídos a algo ou alguma coisa. E é muito importante a gente entender o contexto nesse dado, para que a gente consiga elaborar perguntas pra ele e, principalmente, avaliar as respostas que ele nos dá. Um outro exemplo de dados em geral é a previsão de tempo.

00:01:31:10 - 00:01:53:02

Eu sou fissurada em ficar olhando a previsão do tempo. Às vezes eu olho umas três vezes no dia, mesmo sabendo que a previsão do tempo vai ser só dos próximos 7 dias. É através desse conjunto de dados que os meteorologistas coletaram que eu obtenho as minhas informações e consigo saber se vai chover amanhã e será que no final de semana vai estar chovendo?

00:01:53:03 - 00:02:12:19

Ou vou poder sair tranquilo ou é melhor ficar em casa e evitar me molhar? Se a gente tem um conjunto de dados e a gente não faz nenhuma pergunta para esse dado, ele não vai significar nada para a gente. Ou seja, esses dados não geram informação. Informação é o significado que a gente obtém a partir dos dados. Beleza? Então, vamos para outro exemplo de dado e de informação.

00:02:13:05 - 00:02:43:21

A gente costuma responder à pesquisa do IBGE e no final o governo tem ali todo um conjunto de dados, toda uma análise que ele vai nos passar. E às vezes em jornal apresenta assim: “Através dos dados coletados pelo Ministério Público, constatou-se que Minas Gerais possui cerca de 80% dos casos de dengue confirmados no Brasil.” Então, aqui o dado é o conjunto de casos confirmados no Brasil, enquanto a informação é de que 80% dos casos confirmados estão em Minas Gerais.

00:02:44:06 - 00:03:02:28

Nosso objetivo com os dados é gerar informações. Para que serve essa informação? Serve para dizer que Minas Gerais está com um número muito grande de dengue e que o governo do Estado precisa tomar alguma atitude para diminuir esse número. Inclusive, boa parte desses dados são públicos, a gente pode procurar na internet os dados dessa pesquisa, enfim, a gente pode dar uma olhadinha.

00:03:03:21 - 00:03:28:10

Deu para entender que os dados eles podem ser pequeninhos, como a nossa agenda telefônica aqui do nosso celular, mas eles podem ser grandiosos e através dessa grandiosidade a gente pode fazer questionamentos, ter respostas e dessas respostas a gente pode criar vários planos, várias soluções. Inclusive é assim que o nosso governo trabalha. O nosso governo trabalha fazendo essas análises de vários estados, da população no geral e tenta trabalhar por cima daqueles dados que estão piores.

00:03:29:11 - 00:03:55:23

Beleza, Carol, entendido. Eu posso ter um conjunto de dados, eu vou fazer perguntas para esse conjunto de dados e vou obter essas respostas. É isso, vou confiar nessas respostas e vou correr pro abraço. Calma, calma, não é assim. Nem sempre a gente vai poder confiar nas respostas desses conjuntos de dados. É preciso entender o contexto desses dados, de onde é que eles vêm, quem são as pessoas que responderam, quem são esses dados, de onde vieram?

00:03:56:00 - 00:04:16:02

Mais tarde, no Globo Repórter! Então, um exemplo: eu moro em BH e vamos supor que eu decido responder à pergunta de qual é o salário médio de BH. E aí, para responder essa pergunta eu vou no bairro mais chique de BH, colho ali qual é o salário da galera e aí, com esse salário dessa galera, eu chego à conclusão que BH inteira tem o mesmo salário aproximado.

00:04:16:17 - 00:04:36:27

Faz sentido eu ir no bairro mais chique e caro de BH e tirar conclusões para BH inteira? Imagina no jornal aparecendo lá: o salário médio do cidadão de BH é R\$50.000,00. Gente, não vai fazer o menor sentido. Porque de onde vieram esses dados? Esses dados vieram do bairro mais chique de BH, mais chique e caro. Eu não posso dizer que BH inteira é representada por esse bairro.

00:04:37:10 - 00:04:57:02

Esses dados eles estão enviesados, eles estão com um viés, com a perspectiva dessas pessoas mais ricas e chiques. Eu também não posso ir em um bairro extremamente pobre e carente de BH e falar que BH inteira é representada apenas por esse bairro. Existe toda uma BH complexa e gigante. Então a gente precisa tomar muito cuidado para saber de onde esses dados são, quem são as pessoas que coletam esses dados, entendeu?

00:04:57:03 - 00:05:15:21

Porque senão a gente vai acabar tirando conclusões que não são reais. E olha, isso acontece bastante. A gente vai falar um pouquinho mais pra frente, quando a gente estiver em outras aulas, mas acontece muito de a gente ver e tirar algumas conclusões de dados e acreditar nessas conclusões sem saber de onde vieram os dados.

00:05:15:26 - 00:05:33:28

Como vocês viram, eu dei vários exemplos e a gente está cercado de dados e a gente pode tirar dados até de uma imagem. Vamos supor essa imagem. Olha essa equipe maravilhosa da PrograMaria e vamos tirar alguns dados dessa imagem. Ah, Carol, duvido, eu estou vendo só pessoas ali, o que a gente vai tirar de informação? Pois nós vamos tirar informações dessa imagem e vamos criar uma tabela.

00:05:33:28 - 00:05:55:24

Observando essa imagem, a primeira coisa que a gente pode ver é qual a quantidade de pessoas que tem ali. A gente tem sete pessoas. Já é uma informação. Serve pra quê essa informação? Estou querendo responder alguma pergunta? Se alguém me perguntar quantas pessoas estavam na foto da PrograMaria, eu vou conseguir informar que eram sete pessoas porque eu olhei a foto e eu peguei esse dado de quantidade.

00:05:55:25 - 00:06:16:07

Ou seja, a quantidade de pessoas é um dado quantitativo, de quantidade, numérico. Agora, por exemplo, posso ver ali que quase todos estão com uma blusa azul, azul escuro. Tipo, seis pessoas estão com a blusa azul escuro e uma pessoa está com a blusa preta. Já é uma informação de qualidade, uma informação qualitativa.

00:06:16:07 - 00:06:37:21

Se eu precisar, eu posso falar “Nossa, aquela foto da PrograMaria tinha seis pessoas com a blusa azul escura e uma pessoa com a blusa preta”. Outra coisa: usando calça. Quantas pessoas estavam usando calça na imagem? Seis pessoas usando calça, uma usando saia. Quais as cores dessas calças? Tá vendo a quantidade de informação que a gente está tirando dessa imagem? A gente consegue olhar ali e falar qual que é a pessoa mais baixa da imagem, qual é a pessoa mais alta da imagem.

00:06:38:00 - 00:06:54:18

A gente consegue ver qual a pessoa que está usando óculos, qual que não está usando óculos. A gente consegue ver qual pessoa está com tênis, qual que não está. E essas informações podem virar uma tabela. E aqui a gente tem a nossa tabela. Eu coloquei aqui o nome das pessoas porque eu sou amiga de todes, da esquerda pra direita, Raíssa, até chegar na Stefani.

00:06:55:07 - 00:07:17:28

Então, aqui eu tenho os nomes das colunas na primeira linha e as linhas abaixo são as informações. Linha é horizontal, coluna é vertical. Beleza? De coluna a gente tem o nome das meninas, a cor de blusa, a vestimenta, a cor da vestimenta, o que tem nos pés e a altura delas. Por exemplo, contando a quantidade de nomes na coluna “nome”, temos que são sete pessoas, um dado quantitativo.

00:07:18:10 - 00:07:35:28

A coluna cor de blusa está aqui, azul pra maioria, só uma que está com a blusa preta. A maioria usando calça, só uma que está usando saia. A coluna cor de vestimenta a gente tem aqui quem tá de azul, quem está com a cor roxa, quem está de preto... A coluna nos pés responde o que tem nos pés, se é um tênis, se é uma sandália.

00:07:36:11 - 00:08:01:08

Esses valores aqui são valores qualitativos. A gente tem valores numéricos na coluna altura, altura com o “m” entre parênteses. Aqui, por exemplo, são valores numéricos. E tem esse “m” entre parênteses que indica a unidade de medida desses números, que no caso é em metros. Veja bem, eu não estou vendo a foto, nunca vi a foto, mas eu vejo aqui a tabela, a tabela que é esse conjunto de colunas e linhas.

00:08:01:15 - 00:08:25:16

Se alguém perguntar qual o nome da pessoa mais alta, eu posso vir aqui na coluna de altura, ver qual é a maior altura, que no caso é 1,74 e nessa linha de 1,74 eu vou pra coluna de nome e vejo que a pessoa mais alta é a Lana, ou seja, que o nome associado a 1,74 é o da Lana. Lana é a pessoa mais alta.

00:08:25:26 - 00:08:42:17

Se eu for na foto, eu vou conseguir ver que a Lana realmente é a pessoa mais alta. Beleza? Olha que legal, bacana. Outra pergunta: quem é que está usando uma blusa preta? A gente vai vir aqui na coluna cor de blusa, vai ver preto. Na mesma linha, vamos para a coluna nome e vemos que é a Stefani que está com a blusa preta.

00:08:43:19 - 00:09:02:02

Veja bem, a gente foi formulando várias perguntas e a gente conseguiu responder elas através da tabela, através dos nossos dados, sem olhar pra a imagem. Vendo a nossa tabela, a gente percebe que ela é bem estruturada e isso significa que a gente tem uma coluna que podemos

também chamar, nesse caso, de atributos ou variáveis. E cada coluna recebe a sua informação.

00:09:02:13 - 00:09:25:09

A altura é uma informação numérica, então ela vai aceitar só dados numéricos. E, nesse caso, temos aqui que a altura está representada em metros. A gente tem ali colunas categóricas, então a gente vai receber dados categóricos, de categoria, é um dado muito estruturado. Mas existem também os dados não estruturados, eles são imprevisíveis, digamos assim. Eles não estão tão fixos, igual a gente tem em uma tabela.

00:09:25:18 - 00:09:48:05

Eles estão se atualizando, se movimentando, se criando, mudando. Exemplos de dados não estruturados são vídeos, vídeo em movimento, vídeos como YouTube, TikTok, Instagram etc. Imagens também são dados não estruturados. Mensagens de texto, sei lá, do Twitter, são dados não estruturados. Agora que já sabemos o que são dados, que tabelas usam dados estruturados, eu deixo um desafio para vocês.

00:09:48:18 - 00:10:08:04

Quais outras informações vocês podem tirar da foto da PrograMaria? Vou dar um exemplo. A questão do óculos eu não coloquei, mas você pode criar uma coluna de óculos e colocar se a pessoa usa óculos ou não. Um outro desafio também que eu deixo para vocês é vocês pegarem uma foto de vocês com a família ou com os amigos e tentar tirar os dados e fazer uma tabelinha com esses dados.

00:10:08:15 - 00:10:19:05

Acho que vai ser super legal vocês olharem isso, além de mostrar para as pessoas também o que vocês estão aprendendo. Tipo “olha que legal, peguei a nossa foto, tirei várias informações e tal”. E é isso, gente!