

## 6.1 - O que é uma boa visualização de dados

00:00:00:06 - 00:00:01:01

Olá, pessoal. E ufa também, né, gente? A gente acabou de passar por um módulo mais intenso, que foi o módulo de banco de dados. Eu peço para que vocês parem aí um pouco, revejam o conteúdo. Se ficou com alguma dúvida, procurem a nossa tutora. Não continuem se vocês estiverem com alguma dúvida ou até mesmo insegurança em alguma coisa do que a gente já viu até aqui. Deem uma olhada no material complementar também, sempre vai ter coisa interessante e converse entre vocês. Às vezes um amigão entendeu melhor um conteúdo que você está com dificuldade e vocês podem fazer essa troca bacana, tá? E com isso, bem-vindos a esse novo módulo, esse novo módulo aí que é o módulo de visualização de dados, que é um módulo que eu gosto bastante. A gente vai ver bastante gráfico, uma coisa mais interativa e colorida também. Bora lá. Lembram quando nós estávamos falando de outliers ou valores discrepantes?

00:00:01:01 - 00:00:02:33

Que nós fizemos um gráfico do tipo boxplot e ele nos ajudou a visualizar onde estavam os quartis, onde ficava a mediana, o que era um outlier, etc. Conseguir ver essas informações foi bem importante para entendermos como nossa amostra estava distribuída, né? E é aí que entra a visualização de dados. A visualização de dados é a representação e a apresentação gráfica dos dados. Ouso dizer que é a habilidade mais importante de um analista de dados, porque ela possibilita o analista contar a história dos dados e permite que todos, entendendo ou não dos dados, tirem conclusões de forma rápida e direta. No geral, nós temos duas regras de ouro para uma boa visualização de dados. A primeira é que nos primeiros cinco segundos que as pessoas olharem para o gráfico, elas precisam saber para o que elas estão olhando. E a segunda é que elas devem entender as conclusões que o gráfico está querendo dizer. Por exemplo, vou mostrar aqui para vocês esse gráfico. E façam aí o teste. 5 segundos olhem aí e os 5 segundos seguintes aí, rapidinho. Olhando rapidinho a gente consegue saber que o consumo de energia está aumentando. Pronto, de cara já temos nossa informação principal. Aí se pararmos para olhar mais um pouco, vemos que foi no período de janeiro a junho, né? Um gráfico simples, que passa tudo em poucos segundos. Prático, né? A gente consegue até ver que de abril para maio teve uma queda. E que depois voltou a aumentar, chegou até a dobrar de maio para junho. Vejam aqui nesse outro exemplo. Temos dois gráficos. Esses gráficos contêm as mesmas informações, a diferença é a visualização que eles têm.

00:00:02:33 - 00:00:04:07

No gráfico esquerdo temos barras. Pela primeira barra ser bem maior, a gente consegue concluir que ela provavelmente tem uma distância maior, já que é a informação que o gráfico quer passar, de distribuição total por quilômetro por região. As outras regiões aqui estão bem distribuídas em relação a distâncias. Tudo certo. Já o gráfico da direita é um gráfico de formato pizza. A gente levaria um tempo maior para entender. Vejam esse espaço azul aqui. É o maior, mas temos que olhar na legenda para saber do que é essa parte azul. Existem algumas ferramentas que auxiliam na organização dos pensamentos sobre a visualização de dados. Uma ferramenta interessante é o método McCandless. Nesse método, a junção de quatro elementos faz uma boa visualização.

São eles: a informação, ou seja, os dados, a história (a narrativa), o conceito, que são as conclusões da visualização, e a visualização em si, que são os gráficos. Então aqui, como eu disse, a gente tem história, objetivo, forma e informações que são os dados. E aí a junção de cada um deles dá algo em específico. Por exemplo, se a gente juntar esses quatro, a gente tem uma visualização de sucesso. Se temos só objetivo e uma visualização, uma forma visual, nós temos então só um template, uma coisa assim sem muita função. Não temos os dados ainda.

00:00:04:07 - 00:00:05:41

Se temos a história, aqui, e a gente tem os dados, informação dos dados, nós temos um artigo, uma pesquisa, mas nada de visualização ainda. A gente precisa desses quatro para ter essa visualização de sucesso. Então o método McCandless busca simplificar dados complexos para que as visualizações sejam facilmente compreendidas para qualquer pessoa e por qualquer pessoa. A visualização deve contar uma história. Precisamos organizar os dados de forma a criar uma narrativa visual que guie as pessoas através das informações apresentadas. Os gráficos precisam ser esteticamente bonitos, utilizando cores, fontes, layouts interessantes. Um outro método mais simples e conciso é a verificação trifecta de junk charts. Nome difícil, né, gente? Mas ela basicamente diz que a gente precisa fazer três perguntas para entender se o nosso gráfico faz sentido. São elas: qual é a questão prática? O que está tentando resolver? O que o nosso gráfico está tentando resolver? O que o gráfico diz? E o que os dados dizem? Em resumo, estamos perguntando se o gráfico está fazendo sentido. Caso o gráfico não esteja respondendo o que você gostaria de passar, ou o gráfico diz uma coisa e os dados outra, já é um sinal de alerta de que algo de errado não está certo. Vamos para alguns exemplos. Bom, vejam esse gráfico. O que está pior? Os dados, a visualização ou a narrativa do gráfico? Tá tudo ruim, né, gente? Mas esse gráfico está analisando a quantidade de gelo no Ártico em uma passagem de tempo, que é de 1979 a 2014.

00:00:05:41 - 00:00:07:09

A legenda aqui diz que tem os meses, né? Ó, coloridinho, vermelho e tal. Mas não conseguimos entender qual mês o volume do gelo aumentou ou diminuiu. Podemos dizer, então, que temos os dados, temos qual é a questão, mas não temos uma boa visualização. Poderíamos usar um gráfico único de linhas que a gente pudesse comparar ano a ano o que estaria acontecendo. Bom, nesse próximo exemplo a gente tem a média de bebidas consumidas por pessoas em cada uma das regiões. Temos algumas separações, parece que são minigráficos, né? Aqui, Rússia, Filipinas. Temos uma visualização que dá pra comparar por país e temos os dados pra isso. Mas a questão não ficou muito clara, né? Média de bebidas? Como assim? Não tá muito óbvio o que tá acontecendo. Só sabemos o que é pelo título, mas o gráfico mesmo, questão dos eixos, não tá legal. Bom, nesse próximo exemplo temos esse gráfico aqui, e estamos comparando o nível de felicidade do eixo Y, se é progressivo ou não, pelo eixo X. Então, o nível de felicidade aqui e se é progressivo ou não aqui. Isso tudo de acordo com a quantidade de taxas de imposto do lugar. Assim conseguimos ver que tem uns pontos verdes aqui em cima, que deve ser algo positivo, mas de resto está bem confuso de entender.

00:00:07:09 - 00:00:08:27

Podemos dizer que os dados estão confusos: como assim felicidade por progressão? E a questão também não está muito clara do que está sendo respondido nesse gráfico, sabe? A visualização em si talvez fosse algo positivo se a questão e os dados fossem mais claros. Um último exemplo aqui. Ao olhar esse gráfico, a gente pode até ficar feliz, porque a imagem é colorida. Eu gosto de amarelo, por exemplo. Uma bicicleta aqui e tal. Bom, aqui temos o número médio de bikes ativas no City Bike durante a semana comparados no último outono em dois dias. E aí, dá para entender o que esse gráfico quer responder? O que esses números ali no gráfico representam? Não tem escala, não tem nome, visualmente o gráfico é bonito, mas observando, não está fazendo sentido. Então, os dados não fazem sentido, a questão não está clara e a visualização também não está legal, já que não tem escala e tal. E é isso, gente. Nessa aula, nós vimos o que é visualização de dados, vimos métodos que auxiliam a elaborar uma boa visualização, vimos vários exemplos de visualizações que não tiveram sucesso no que queriam apresentar e representar. Na próxima aula, nós vamos aprender sobre elementos que podemos usar para deixar os nossos gráficos claros, bonitos e fáceis de entender. Até mais!