

2.4 - Nossa primeira análise | Analisando distribuição de raça e gênero

00:00:00:07 - 00:00:22:12

Olá gente! Agora que já entendemos o que são dados, já podemos começar o nosso projeto. Lembra dele? Ah, vale um pequeno aviso que nessa e nas próximas três aulas nós vamos trabalhar alguns itens básicos de análise de dados com a planilha e no próximo módulo vamos reproduzir essas análises em código. Bom, algumas vezes no nosso dia a dia vamos receber dados como esse aqui na tela e vamos ouvir um:

00:00:22:13 - 00:00:42:25

“Tire insights desses dados”. Uma palavra chique para pedir informações úteis dos dados. Ou seja, nem sempre teremos o objetivo especificamente definido. Por isso, a primeira coisa que precisamos fazer é entender as informações que temos na mão e fazermos perguntas em cima dela. Então vamos lá. Vai estar disponível nessa aula um arquivo para que vocês façam o download para vocês usarem da mesma forma que eu vou usar aqui.

00:00:42:25 - 00:01:00:23

Vou aconselhar todos a entrar no Google Drive e criar uma pastinha lá para o curso da PrograMaria. Assim, vocês podem até concentrar todo o material do curso nessa pasta mesmo. E vocês fazem o upload desse arquivo lá nessa pasta. É só arrastar o arquivo para dentro da pasta, no drive e pronto.

00:01:01:09 - 00:01:27:26

Se vocês não quiserem levar a planilha para o Drive, vocês podem abrir no Excel mesmo. Mas é importante lembrar que as opções de barra de ferramentas pode mudar de uma ferramenta para outra. Por isso, eu aconselho que acompanhem a aula usando a mesma ferramenta que eu, que, no caso, é o Google Planilhas ou o Google Sheets, em inglês. Fechou? Bom, assim podemos abrir o nosso arquivo direto pelo Drive, pelo Google Planilhas. Vejam que aqui a gente tem a tabela que a gente vai trabalhar durante boa parte desse curso.

00:01:28:09 - 00:01:52:26

Essa tabela é da pesquisa do Data Hackers de 2022, que tem as respostas das pessoas que trabalham em dados. Ah, por enquanto vamos trabalhar com uma versão resumida da planilha para facilitar o nosso processo aqui. Depois nós vamos pegar a planilha completa. Temos aqui a idade, em termos de coluna, idade, faixa de idade, gênero, cor, raça, etnia.

00:01:53:22 - 00:02:14:08

Bom, temos várias possibilidades de análise com esses dados, dependendo de qual for o nosso objetivo. Veja bem, a gente tem aqui a coluna de gênero que está aqui ó, coluna D de gênero. Será que a gente consegue saber a quantidade de masculino e a quantidade de feminino? Já é uma coisa, já é uma pergunta. Indo além, já que a gente tem a coluna de cor, raça, etnia aqui do lado.

00:02:15:02 - 00:02:46:19

Imagina aí que queremos saber quantas pessoas que marcaram que são do sexo feminino e também marcaram que são brancas? Com isso a gente junta informações de duas colunas, pois queremos filtrar da coluna gênero e da coluna cor, raça e etnia. Uma forma de responder as nossas perguntas é filtrando essas colunas. Como assim filtrar? Quando a gente fala em filtrar é literalmente pensar em um filtro. Aqui na planilha Google é fácil realizar filtro, é só irmos nesse símbolo de funil aqui nessa barra de ferramentas que fica nas opções superiores.

00:02:47:01 - 00:03:08:29

E ao clicarmos nesse símbolo, a primeira linha da nossa tabela, que contém o nome das colunas, vai receber a opção de ser filtrada. Por exemplo, se a gente for aqui na coluna D de gênero e clicar aqui no cantinho, a gente vai ter algumas opções. Temos diferentes formas de realizar o filtro, mas o que é importante para nós nesse momento são as opções de seleção.

00:03:09:15 - 00:03:41:27

Vejam que as quatro opções de preenchimento de coluna estão selecionadas aqui abaixo. O espaço em branco, o feminino, o masculino, e o prefiro não informar. Se quisermos filtrar apenas as respostas das pessoas que responderam que são do gênero feminino, basta deixarmos apenas o gênero feminino selecionado. Então vamos tirar aqui a seleção dos outros. E assim que clicarmos em OK, a tabela vai ser atualizada com esse filtro e podemos observar que na coluna gênero está apenas o feminino.

00:03:42:14 - 00:04:09:27

Olha só! Tudo preenchido como feminino nessa coluna D de gênero. E na parte inferior direita da tabela, fica a informação de quantas linhas são mostradas durante esse filtro que realizamos. Olha que legal: a tabela que, no total, tem 4271 linhas, depois do filtro está exibindo apenas 1056 linhas. Ou seja, estamos deixando de visualizar a resposta das pessoas que responderam que são do sexo masculino ou prefiro não informar ou que deixaram o espaço em branco.

00:04:10:13 - 00:04:35:29

No total, tivemos 1056 respostas de mulheres. Muito legal, né? Lembrem de voltar o filtro para vermos tudo. Vamos lá tirar o filtro aqui em cima na barra de ferramentas e remover filtro. E se a gente quiser juntar todas as pessoas através da coluna de cor para sabermos a quantidade de pessoas por cada categoria? Aqui no Excel a gente tem uma forma muito legal de fazer isso, que é através da tabela dinâmica.

00:04:36:13 - 00:05:05:02

A tabela dinâmica é uma ferramenta em que a gente consegue juntar e/ou cruzar informações para resumir nossos números e facilitar a nossa análise. Vamos ver outras formas de fazer isso lá quando a gente chegar no código, mas aqui no Excel ela vai ser uma ajuda muito útil para gente. A gente pode selecionar a tabela toda aqui clicando no cantinho superior esquerdo da tabela, e depois ir aqui na barra de ferramentas em inserir e clicar em Tabela Dinâmica.

00:05:06:17 - 00:05:31:00

Aqui vai aparecer se a gente quer criar a tabela em uma nova página e a gente vai clicar em Criar e pronto. Vai ser criada em uma nova página uma tabela inicialmente em branco. E aqui a gente tem colunas e linhas. Os valores vão aparecer dentro dessa tabela, nessa parte de valores. E aqui do lado direito podemos ir colocando nas colunas e nas linhas o que a gente quer filtrar: linhas e colunas.

00:05:31:17 - 00:05:53:06

Lembra lá que a nossa primeira pergunta era qual a quantidade de pessoas que responderam que são do gênero feminino e quantas do gênero masculino? Vamos descobrir isso agora então. Queremos que as linhas da nossa tabela sejam os gêneros. Lembra lá da tabela que montamos com base na foto da PrograMaria? Nela, as linhas eram os nomes das pessoas, mas aqui, nesse caso, queremos que o gênero sejam as nossas linhas.

00:05:53:18 - 00:06:20:00

Para isso, vamos colocar aqui na opção de linhas a coluna Gênero. Então vamos aqui em adicionar e clicar em gênero, que é a nossa coluna. Ok. É só segurar mesmo a coluna gênero e arrastar para a opção de linhas e já temos aqui na tabela as linhas feminino e masculino. Essa linha vazia aqui são pessoas que não responderam nada, que deixaram o campo de resposta vazio.

00:06:20:25 - 00:06:45:09

Quando fizemos o filtro, eram os espaços em branco. Agora, como conseguimos saber a quantidade de pessoas? A gente pode usar a coluna de id. O id é um valor único para cada pessoa que respondeu o formulário, como se fosse um CPF, em que cada pessoa tem um número único de identificação. Então, podemos contar a quantidade de ids que respondeu que é do gênero masculino ou feminino e etc.

00:06:45:24 - 00:07:12:14

Podemos pegar aqui a coluna de id e arrastar para valores. Se a gente for aqui pro cantinho e arrastar para valores, pronto. Temos aí nossa tabela filtrada por gênero. O que fizemos aqui foi agrupar as informações por gênero e contar os únicos de cada gênero. Lembrem que isso aqui é um agrupamento. Vai ser útil pra gente lá na frente. Com essa tabela, podemos responder outras perguntas.

00:07:12:24 - 00:07:45:10

Veja bem, quantas pessoas deixaram o campo de gênero em branco? Nove pessoas. Quantas pessoas responderam que são do gênero feminino? 1056. Quantas pessoas responderam que são do gênero masculino? 3194. E quantas optaram por não responder, por não informar gênero? 12 pessoas. Total geral de 4271 pessoas que responderam um formulário. Podemos optar por não mostrar esse total geral. É só clicar aqui do lado direito.

00:07:45:11 - 00:08:07:21

(Des)selecionar, na verdade, aqui do lado direito e pronto. A pergunta que tínhamos pensado é a distribuição de raça. Quantas pessoas responderam que são brancas, amarelas, etc? Então vamos lá de novo. Dessa vez vamos agrupar por cor. Aqui na linha podemos tirar o gênero e colocar a coluna de cor. Aqui a gente pode tanto arrastar quanto clicar aqui em Adicionar. Cliquei em Adicionar.

00:08:08:03 - 00:08:35:03

Vou clicar agora na linha aqui de cor, raça e etnia e pronto. Com esse resultado, podemos facilmente responder quantas pessoas responderam que são amarelas. No caso, 128 pessoas. Podemos responder qual a cor da maioria das pessoas que responderam a pergunta, ou seja, a maior quantidade aqui são de pessoas brancas, com 2744. Dentre várias outras perguntas... Uma coisa legal é que podemos selecionar essa tabela e transformar num gráfico.

00:08:35:16 - 00:09:02:26

Fica ainda melhor de visualizar. Antes, nós precisamos tirar a parte do total geral, então vamos (des)selecionar aqui nesse canto direito. Agora sim, selecionamos a parte da tabela e vamos clicar aqui na barra de ferramentas em Inserir e agora vamos clicar nessa opção aqui de gráfico. Dá para observar que a cor predominante é a cor de pessoas brancas, com a maior parte em vermelho, é de pessoas brancas, 64,2%.

00:09:03:10 - 00:09:38:01

Em segundo lugar vem pessoas de cor parda, que marcaram que são pardas, com 24,7%, depois preta com 6,8% e depois amarela com 3%. Então na aula de hoje aprendemos a filtrar e agrupar informações para respondermos algumas perguntas sobre os nossos dados, transformando eles em informação. Eu gostaria que vocês explorassem a tabela, gerassem gráficos e compartilhassem entre vocês. Um exercício é responder à pergunta: Qual a quantidade de pessoas por nível de ensino?

00:09:38:17 - 00:09:43:19

Faça uma tabela como nós fizemos agora. Depois criem gráficos e explorem. E é isso, gente!