Projeto 1

Guia para a "Análise de excesso de mortes por COVID-19"

short line

# Introdução

## Contexto

Depois de ler esta interessante situação no [brief](https://docs.google.com/document/d/1nN47G0r3lcGTxa_KqLwZuM0xt2qgzqHjr1JRX_dRxLA/edit?usp=sharing) do projeto, sua tarefa será analisar e calcular o excesso de mortes em 5 países da América Latina com o objetivo de fazer comparações com as mortes causadas por COVID-19.

O objetivo do projeto é que você tenha um melhor entendimento das mortes causadas por COVID-19 e tire conclusões com base nos dados.

## Objetivo

Depois de revisar o contexto, voltemos ao estudo de caso: você é apaixonada por dados e quer calcular o excesso de mortes em tempos de pandemia, para poder compará-los com os números oficiais informados pelos 5 países da América Latina com o objetivo de descobrir se o número de mortes por COVID é semelhante aos seus cálculos.

Neste projeto iremos ajudá-la em algumas etapas da sua análise e, finalmente, você irá criar um relatório mostrando suas descobertas e conclusões.

# Desenvolvimento

## Etapas

1. Revise o dataset e a estrutura dos dados

Os dados que você usará contêm informações sobre o total de mortes e sobre as mortes relatadas por COVID-19 nos países do México, Peru, Chile, Colômbia e Brasil. Esse conjunto de dados foi processado e armazenado nesta [planilha do Google](https://docs.google.com/spreadsheets/d/1qs64yg3caddAEvkwbZDeEmZNDwba-IIvR-KuUQhuC1k/edit?usp=sharing):

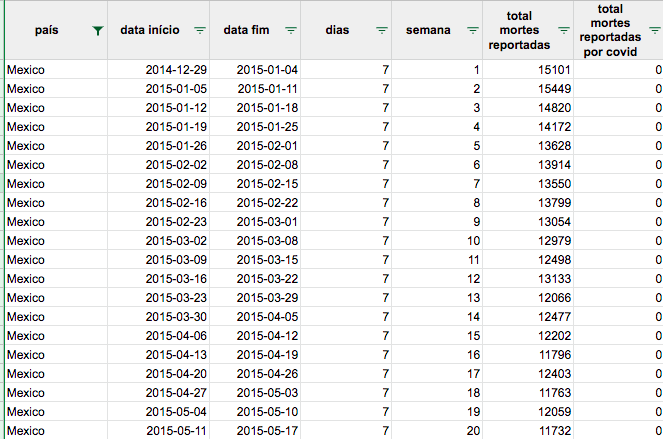
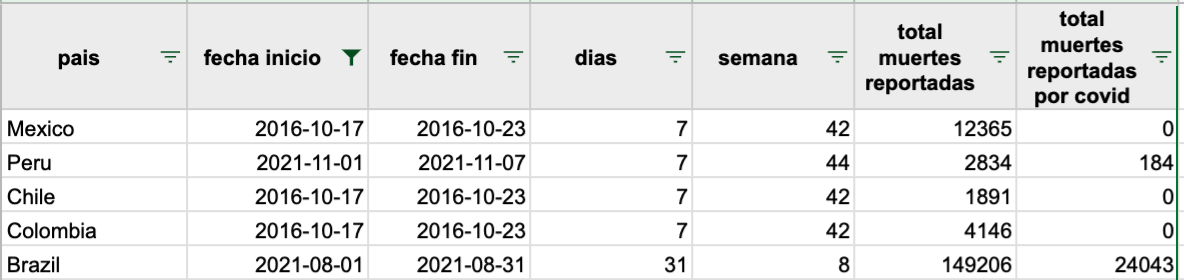


Imagem do dataset

O dataset tem:

* 1.332 linhas, cada linha representa o acumulado por semana ou mês do total de mortes e das mortes causadas por COVID-19. Abaixo detalhamos o significado.
* As colunas são:
  + **país**: nome do país (podem ser: México, Peru, Chile, Brasil e Colômbia)
  + **data de início**: data de início do intervalo de contagem de mortes
  + **data de término**: data de término do intervalo de contagem de mortes
  + **dias**: número de dias da contagem de óbitos. No caso do Chile, Peru, Colômbia e México é uma semana (7 dias). Para o Brasil, os dados são acumulados mensalmente (28, 30 e 31 dias).
  + **semana**: número da semana ou mês do ano.
  + **total de mortes relatadas**: é o total de óbitos notificados por qualquer causa (não apenas COVID-19) no intervalo de datas estabelecido (óbitos em uma semana ou mês).
  + **total de mortes relatadas por COVID-19**: é o total de óbitos notificados causados ​​por COVID-19 no intervalo de datas estabelecido (óbitos em uma semana ou mês)

1. Compreensão do dataset

A imagem acima mostra 5 linhas do dataset, cujo conteúdo está explicado a seguir:

● A primeira linha mostra o número de mortes no México de 2016-10-17 a 2016-10-23 (7 dias no total). Esse intervalo de datas representa a semana 42 de 2016 (lembre-se de que um ano tem 52 ou 53 semanas no total), e nessas datas foram relatadas 12.365 mortes no total e zero mortes causadas por COVID-19. Isso faz sentido, já que a COVID-19 só apareceu em 2020 e essa linha contém dados de 2016.

● A segunda linha mostra o número de mortes no Peru no intervalo de 2021-11-01 a 2021-11-07 (7 dias, semana 44 do ano) com um total de 2.834 mortes totais e 184 mortes causadas por COVID-19.

● A última linha mostra o número de mortes no Brasil no intervalo de 2021-08-01 a 2021-08-31 (31 dias, 8º mês do ano) com um total de 149.206 mortes e 24.043 mortes causadas por COVID.

Como você deve ter notado, cada linha contém o acumulado por semana ou mês do número total de mortes e o número de mortes causadas por COVID-19 nos 5 países.

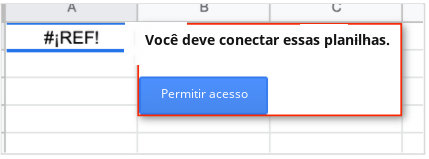
Depois de revisar o significado das linhas, vamos começar com a análise!

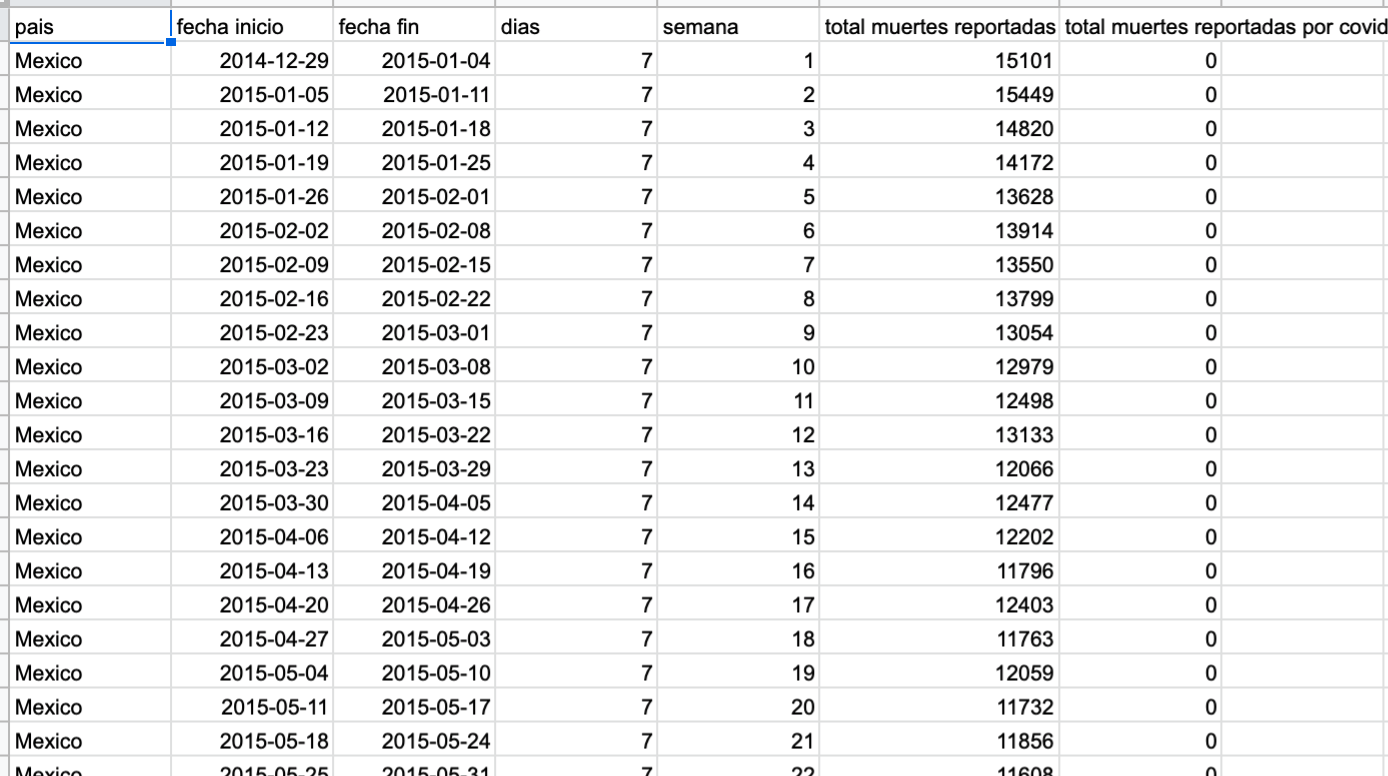
 Neste projeto iremos guiá-la no desenvolvimento da análise para o **Peru**, para que posteriormente você possa completar a análise dos 4 países restantes. Portanto, preste atenção nas etapas!

Para desenvolver este projeto usaremos o Planilhas Google. Trata-se de um aplicativo de planilhas totalmente gratuito, e tudo que você precisa é de uma conta do gmail para começar. Se você estiver usando o Planilhas Google pela primeira vez, recomendamos que você assista a estes vídeos introdutórios ([vídeo1 (📹)](https://vimeo.com/676035218/c9a03c0c94) , [vídeo 2](https://vimeo.com/676039554/2a8956ae53)[📹](https://youtu.be/Q_3g7sXga1g))

1. Criação de uma cópia do banco de dados

O banco de dados está nesta [planilha do Google](https://docs.google.com/spreadsheets/d/1qs64yg3caddAEvkwbZDeEmZNDwba-IIvR-KuUQhuC1k/edit?usp=sharing). Para iniciar o projeto, crie uma nova planilha com sua conta do Google e importe o banco de dados usando a função [IMPORTRANGE (📹)](https://vimeo.com/676071594/529e3bfaca).

Por ser a primeira vez, é muito provável que você veja um erro de permissão. Clique no botão “Permitir acesso” para importar todo o banco de dados. 

A imagem a seguir mostra o resultado da aplicação da função IMPORTRANGE.

1. Análise das mortes no Peru

Depois de carregar o dataset completo, comece revisando os dados. Esta é uma etapa muito importante, mesmo antes de fazer qualquer cálculo, então, para se familiarizar com seus dados, experimente o seguinte:

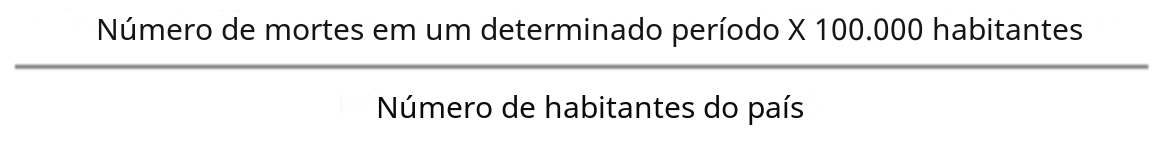
* [Aplicar filtros e classificar dados (📹)](https://vimeo.com/677262853/c4d1be2dd0)
* [Criar tabelas dinâmicas (📹)](https://vimeo.com/676078013/8fb6dd98f0)

Depois de entender os dados, revisaremos os entregáveis do projeto. O [primeiro](https://docs.google.com/document/d/1nN47G0r3lcGTxa_KqLwZuM0xt2qgzqHjr1JRX_dRxLA/edit?usp=sharing) é “Uma tabela comparativa com o **total de mortes relatadas por COVID-19 por 100.000 habitantes** e o ‘**excesso de mortes**’ **por 100.000 habitantes**”. Isso te confunde um pouco? Não se preocupe, a seguir explicaremos porque pedimos esses cálculos.

* 1. Cálculo do número total de mortes por COVID por 100.000 habitantes

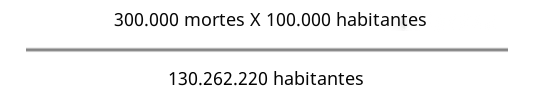
Para cumprir com o entregável, devemos realizar alguns cálculos preliminares. Você pode se perguntar por que pedimos o número total de mortes por 100.000 habitantes e por que não apenas o total.

O número de mortes é normalmente expresso como "o número de mortes por 1.000, 10.000, 100.000 ou um milhão de habitantes", com o propósito de fazer comparações objetivas com cidades ou países que têm diferentes números de habitantes. Em nosso caso, multiplicaremos as mortes por 100.000 e dividiremos o resultado pela população total.



Vamos ver um exemplo:

* Suponha que em 2020 houve 300.000 mortes no México e sabemos que a população total desse país é 130.262.220, então substituímos os valores na fórmula:



Depois de fazer o cálculo, o resultado é 230,30 e é interpretado como: **houve 230 mortes por 100.000 habitantes no México em 2020**.

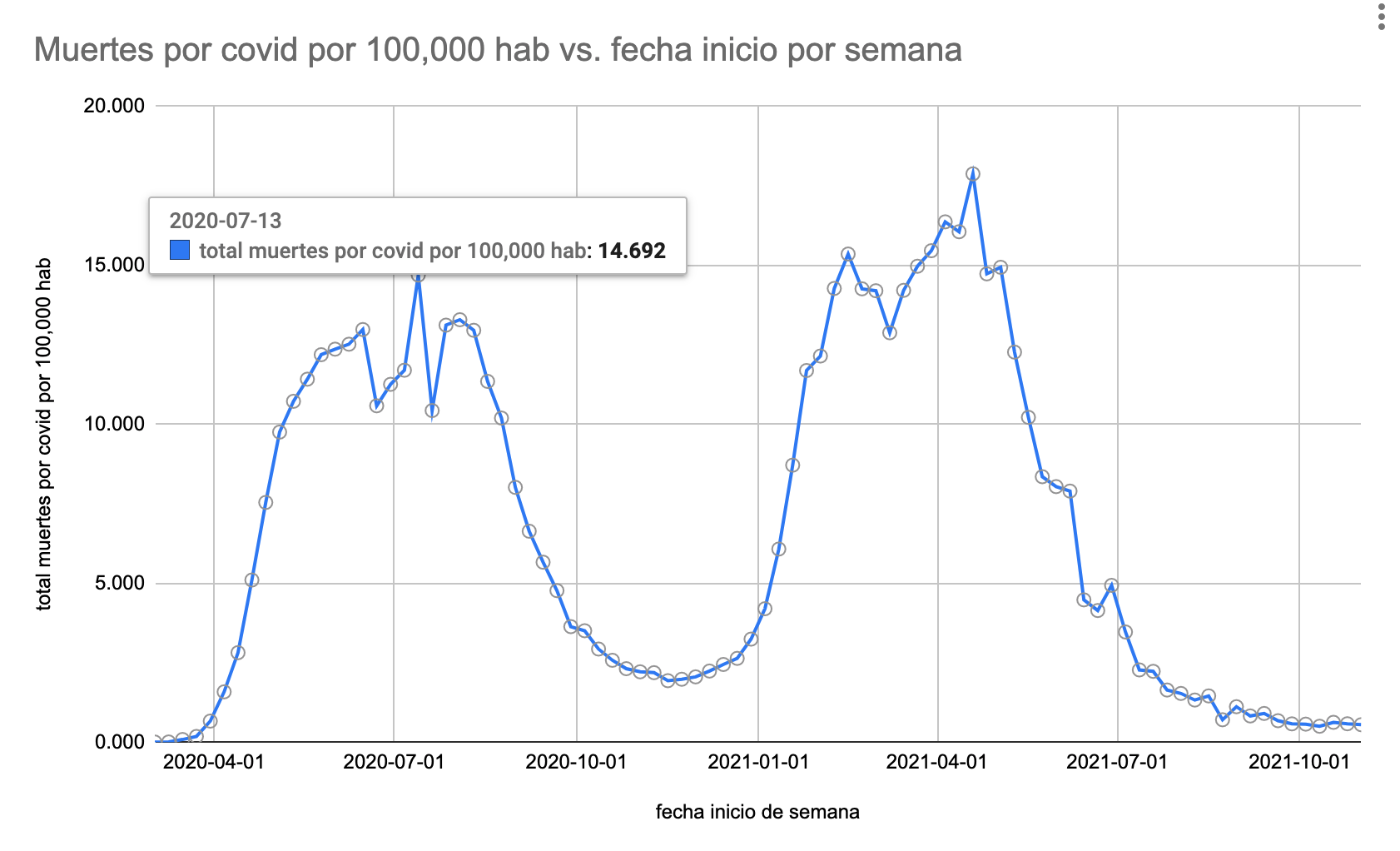
Conceito entendido, vamos realizar o cálculo na planilha do Google.

* 1. Cálculo do número total de mortes por COVID por 100.000 habitantes no Peru

Primeiro filtre o conteúdo do dataset para obter apenas os dados do Peru no momento em que houve mortes por COVID-19, usando a função [QUERY (📹)](https://vimeo.com/676082923/b75b89e83c)

Depois de carregar os dados, **crie uma nova coluna para o cálculo das mortes por COVID por 100.000 habitantes** e aplique a fórmula explicada acima. Para isso, utilize a função PROCV / VLOOKUP do Planilhas Google. Você pode revisar a aplicação desta função no projeto neste [vídeo (📹)](https://vimeo.com/676088587/55ed3332d5).

Uma boa maneira de verificar a tendência das mortes por COVID é por meio de gráficos. Crie um gráfico de linha para mostrar seus dados. Você pode assistir a este [vídeo (📹)](https://vimeo.com/676093043/6952f3588d) para saber quando escolher um gráfico de linha.



A imagem acima mostra a tendência de mortes por COVID por 100.000 habitantes no Peru. Cada número ou valor no gráfico representa o número de mortes por semana por 100.000 habitantes. Por exemplo, você pode ver que na semana que começa em 2020-07-13, 14,6 mortes foram relatadas por 100.000 habitantes. Continue analisando o gráfico e pergunte-se:

**\* Percebemos duas ondas ou picos no gráfico. Por que você acha que isso poderia acontecer?**

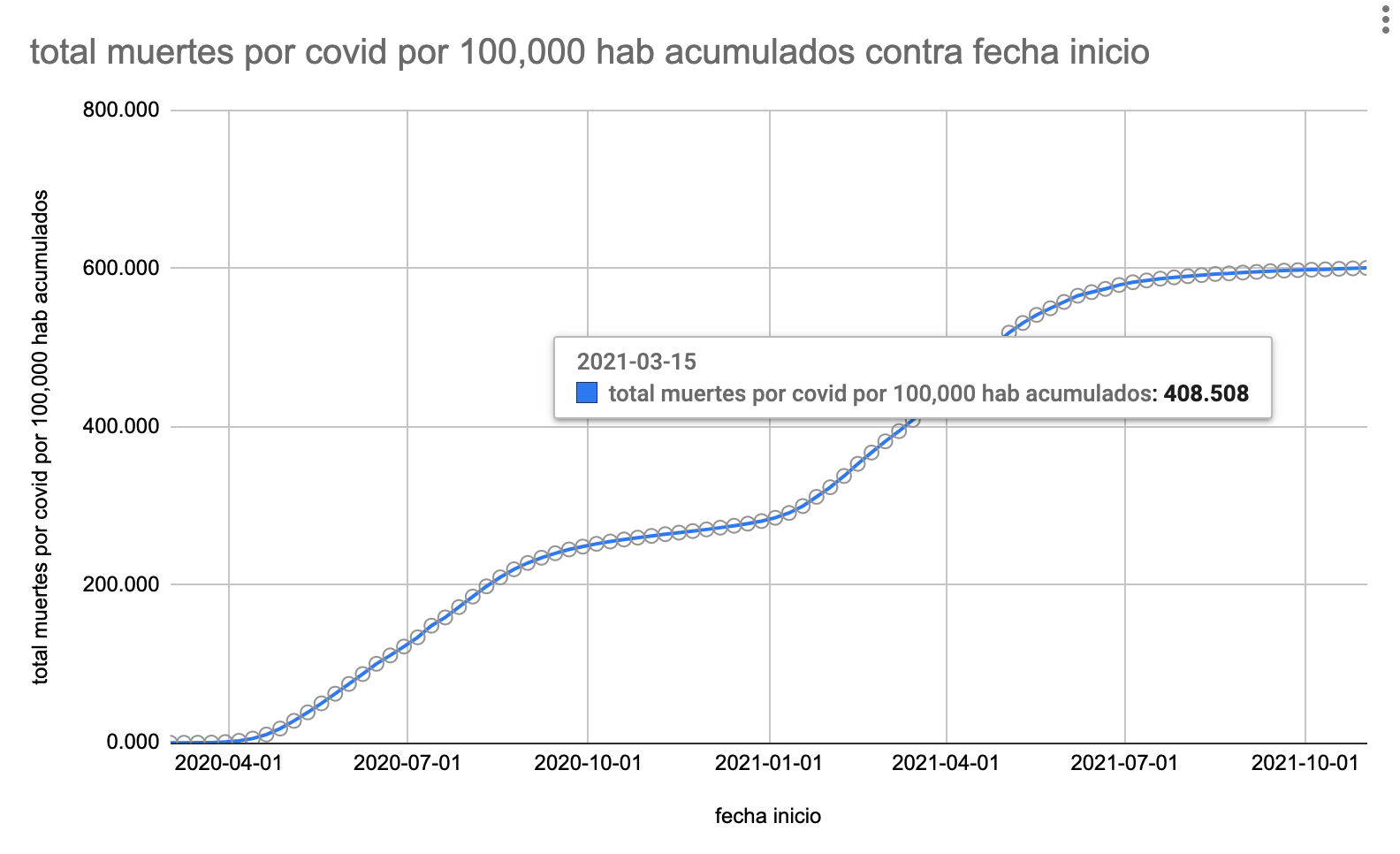
\* **As mortes estão aumentando ou diminuindo?**

Escreva e discuta suas ideias com suas colegas.

* 1. Cálculo do total acumulado de mortes por COVID por 100.000 habitantes no Peru

O próximo passo é saber o número acumulado de óbitos por data, muito semelhante aos resultados anteriores, com a diferença de que adicionaremos o número de óbitos acumulados do dia anterior com o dia atual. Se você tiver dúvidas, verifique esta [explicação (📹)](https://vimeo.com/676095385/c9dfc83a5f) para mais detalhes. Armazene os resultados em uma nova coluna.

Não se esqueça de representar graficamente seus dados para observar melhor seus resultados.



A imagem acima mostra a tendência de mortes acumuladas causadas pela COVID por 100.000 habitantes no Peru. O ponto destacado significa que houve aproximadamente 408 mortes acumuladas até 2021-03-15 por 100.000 residentes.

Este tipo de gráfico, por ser cumulativo, sempre aumentará, mas, neste caso de mortes por COVID, a linha de morte deveria permanecer plana.

**\* Qual é o intervalo de datas em que houve maior aumento de mortes?**

\* **O que você pode dizer sobre o último mês do gráfico?**

Escreva e discuta suas ideias com suas colegas.

* 1. Cálculo do prognóstico de mortes usando médias simples

Como comentamos no início do projeto, seu objetivo é fazer uma comparação das mortes notificadas por COVID-19 (cálculos já realizados acima) versus o excesso de mortes. Este último termo é o que explicaremos a seguir:

O excesso de mortes será calculado como a **diferença entre o total de mortes notificadas e o prognóstico de mortes.**

Você deve estar se perguntando: como eu vou prever as mortes? Não se preocupe, você não usará fórmulas complicadas. Você usará a medida da **MÉDIA**. Aconselhamos que você assista a este [vídeo](https://vimeo.com/676098521/62abc86716)[📹](https://youtu.be/Q_3g7sXga1g), onde encontrará mais detalhes.

* Prognóstico de mortes: para calcular a previsão para um determinado período de tempo, utilizaremos o cálculo de médias simples.

Por exemplo, supomos que a tabela a seguir representa o número de mortes por ano em uma cidade. Se quisermos calcular a previsão de mortes para 2020 usando médias simples, devemos somar os números de mortes e dividi-los pelo número total de anos:

| Ano | Número de mortes |
| --- | --- |
| 2016 | 250 |
| 2017 | 389 |
| 2018 | 450 |
| 2019 | 240 |
| 2020 | ? |

Previsão para 2020 = (250 + 389 + 450 + 240) / 4 = 332,25, então a previsão de morte para 2020 usando a média simples será de aproximadamente 332 mortes.

Use a função MÉDIA.SE.S./AVERAGEIFS para calcular a previsão de mortes nas Planilhas do Google. Se você tiver alguma dificuldade em prever as mortes usando as Planilhas do Google, recomendamos que você assista o seguinte [video](https://vimeo.com/676101618/2a7aeb0f24)[📹](https://youtu.be/Q_3g7sXga1g).

Se você quiser ver mais técnicas de prognóstico, leia o seguinte [artigo](https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/basics/methods-for-analyzing-time-series/).

Às vezes, ao realizar cálculos, encontramos diferentes tipos de erros. Para controlá-los você pode usar a função [IFERROR📹](https://vimeo.com/676104559/73bbb21165) das Planilhas do Google.

* 1. Cálculo do excesso de mortes por COVID

A próxima etapa é calcular o excesso de mortes. Basta subtrairmos **o número total de mortes notificadas menos o prognóstico de mortes.**

Seguindo o exemplo anterior, teríamos:

\* Com o cálculo da etapa anterior obtivemos uma média histórica de 332 óbitos aprox. para 2020, mas também sabemos que o número real de mortes é 450. Então, se quisermos saber o excesso de mortes, subtraímos apenas 450 (número real) - 332 (previsão) = 118

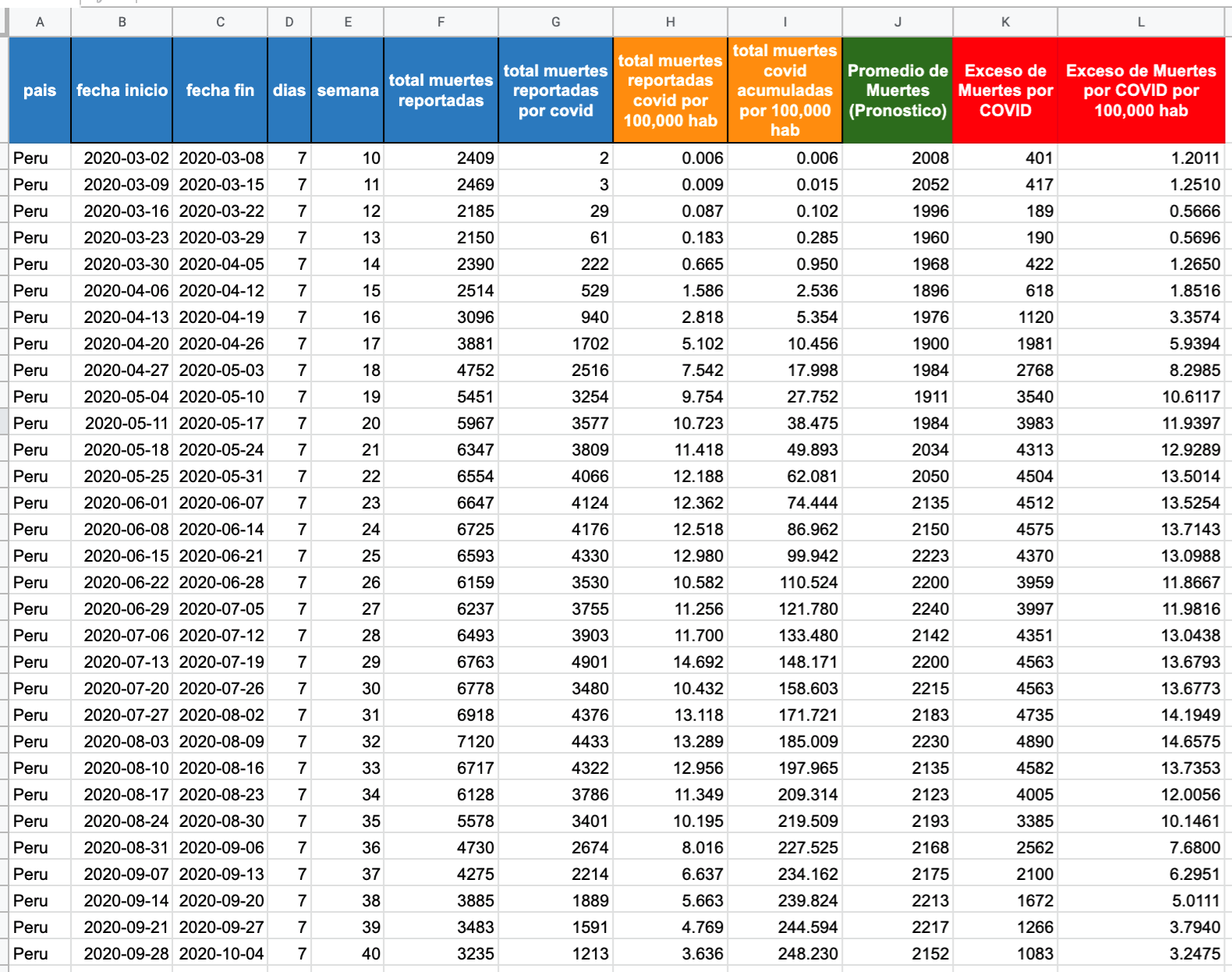
| Ano | Número de mortes reais | Prognóstico de mortes | Excesso de mortes |
| --- | --- | --- | --- |
| 2020 | 450 | 332 | 450 - 332 = 118 |

Portanto, podemos dizer que temos um excesso de 118 mortes.

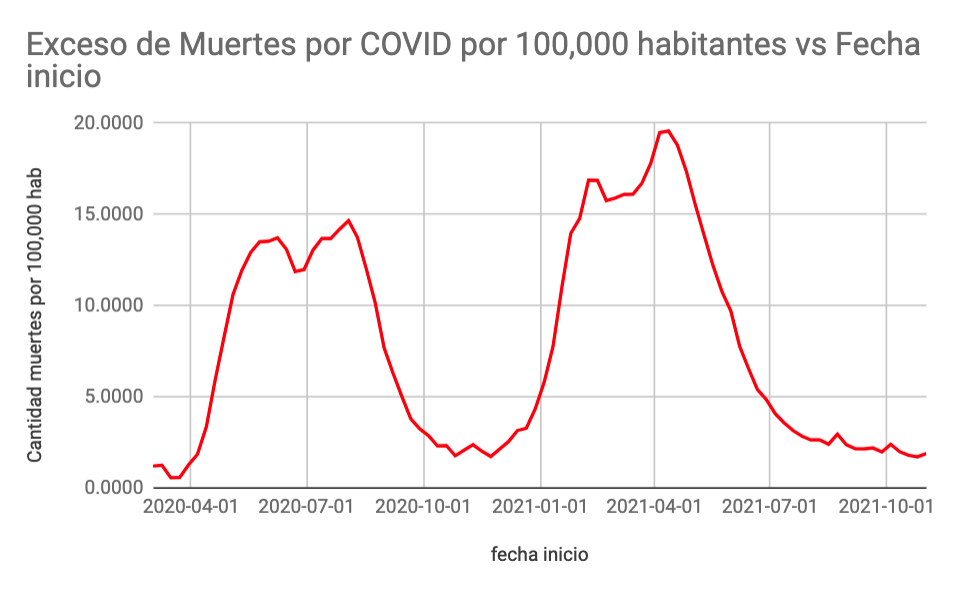
Se você tiver dúvidas para calcular o excesso de mortes nas Planilhas do Google, obtenha ajuda neste [vídeo📹](https://vimeo.com/676106169/5e12315832).

* 1. Cálculo do excesso de mortes por COVID por 100.000 habitantes

Para fazer comparações, vamos converter os números de mortes para valores por 100.000 habitantes, muito semelhante ao que desenvolvemos na Etapa 4.1. Crie uma coluna "Excesso de mortes por COVID por 100.000 habitantes" e aplique a fórmula aprendida nas etapas anteriores. A imagem a seguir mostra os resultados da nova coluna.

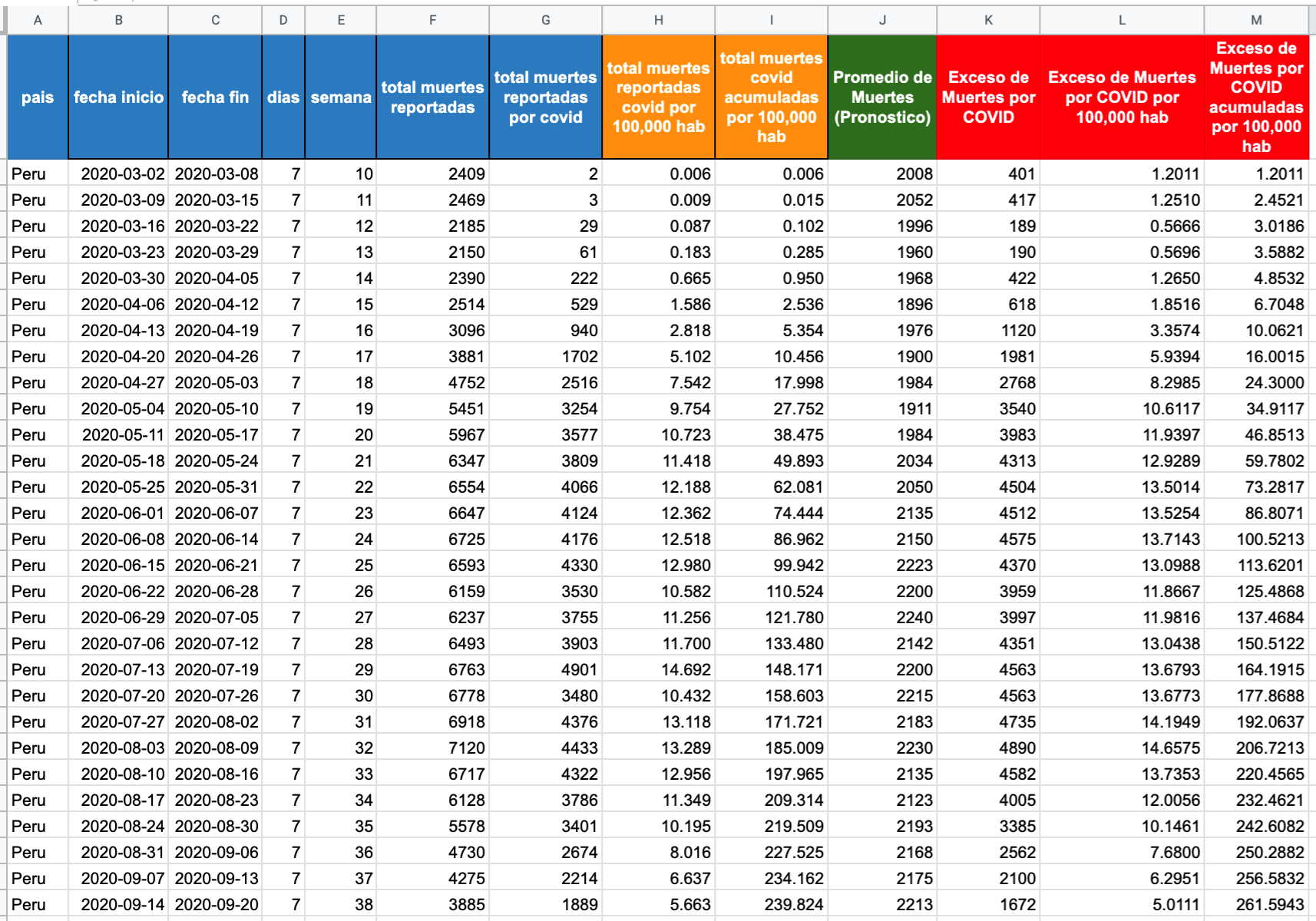


Lembre-se que os gráficos ajudam a melhor observar os insights. Use a coluna que acabou de criar e gere um gráfico de linha.

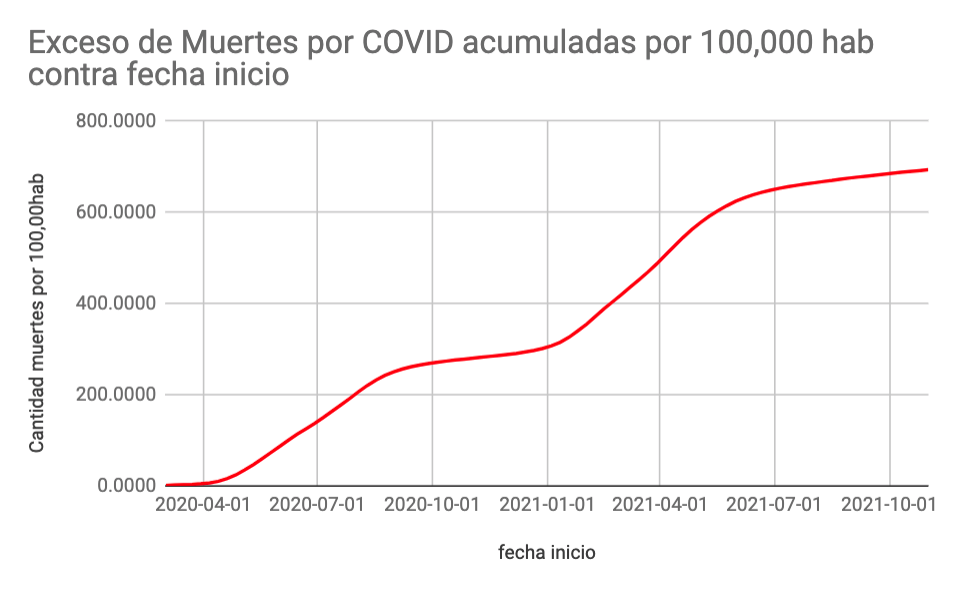


* 1. Cálculo do excesso de mortes por COVID por 100.000 habitantes acumuladas

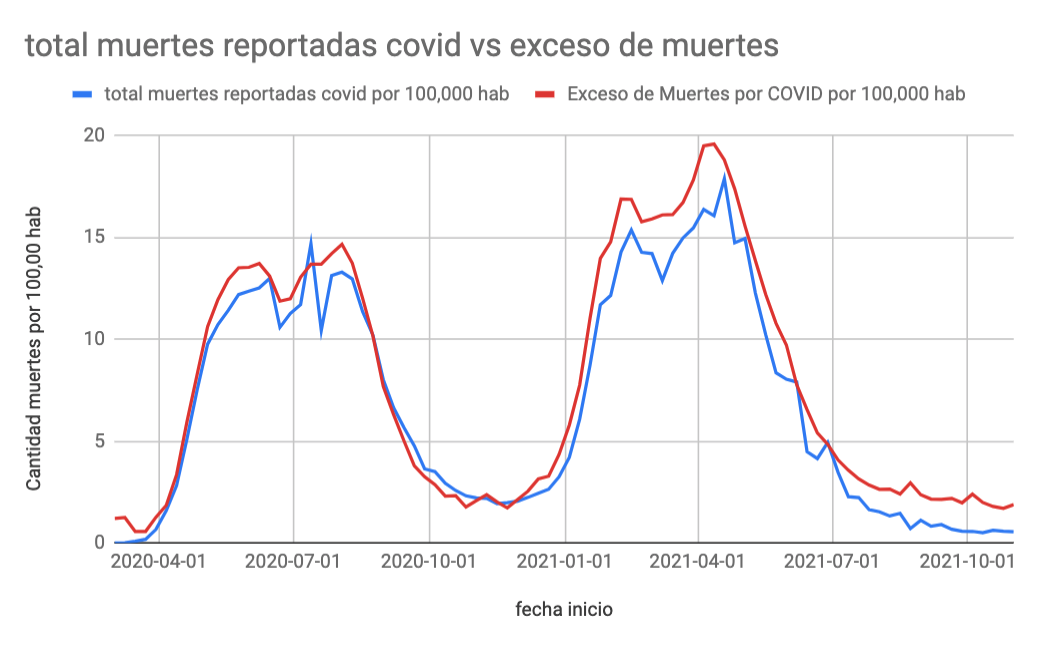
Muito semelhante ao que desenvolvemos na Etapa 4.3, vamos calcular o excesso de mortes acumuladas por 100.000 habitantes para fazer uma comparação com os números oficiais de mortes acumuladas. Crie uma coluna "Excesso de mortes acumuladas por COVID por 100.000 habitantes" e aplique a fórmula aprendida nas etapas anteriores. A imagem a seguir mostra os resultados da nova coluna.



Não se esqueça de representar graficamente seus resultados a partir da coluna calculada, bem como a comparação entre mortes em excesso e o número de mortes oficiais por COVID-19.



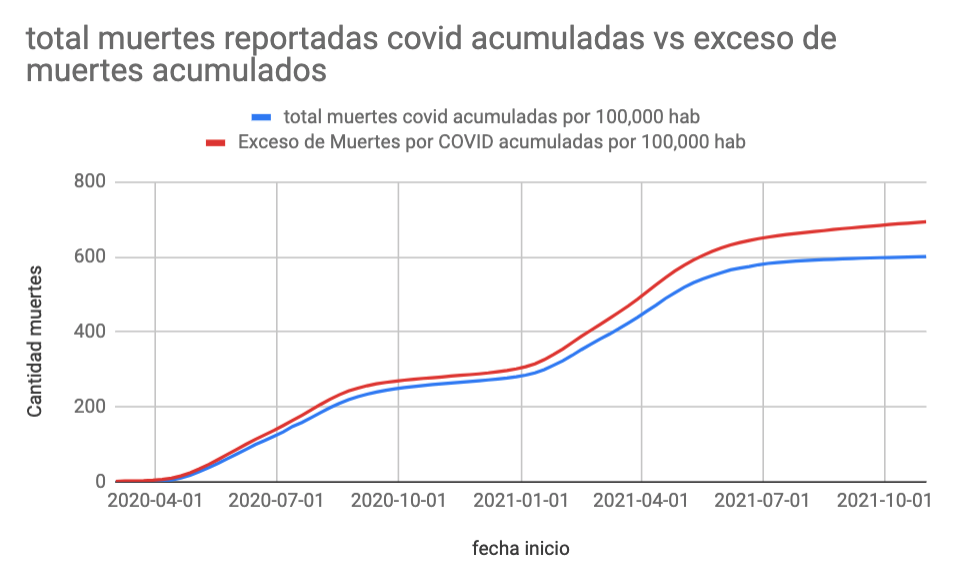
 Você vai ver que as formas dos dois gráficos anteriores são muito semelhantes às geradas nas etapas 4.1 e 4.3. portanto, para fazer uma melhor comparação, você vai juntar esses dois gráficos. Você pode ver este [vídeo📹](https://vimeo.com/676107695/213fa9641b) para aprender como exibir mais de duas variáveis ​​em um único gráfico.



A imagem acima mostra o número de mortes oficiais por COVID-19 versus o excesso de mortes calculado por 100.000 habitantes.

Podemos ver que o excesso de mortes (linha vermelha) superou os números oficiais notificados por COVID-19 (linha azul): que resultado interessante. Você deve se perguntar: por que eles estariam mostrando esses resultados?

Faça uma comparação semelhante, mas agora com o excesso de mortes e os números oficiais de mortes acumuladas.



Este gráfico mostra a comparação do excesso de mortes e das mortes oficiais por COVID-19 acumulados por 100.000 habitantes. Podemos ver claramente uma diferença entre essas duas linhas: o excesso de mortes **excedeu** os números oficiais de mortes por COVID-19. Por que se dá essa diferença? Você pode notar outro insight a partir dos seus cálculos?

Com isso, conseguimos entregar tudo o que foi solicitado para o caso do Peru. Muito bem!

Agora, sua **próxima etapa** será aplicar a mesma análise para os 4 países restantes.

 Se você precisar de um guia para montar a tabela comparativa por país, recomendamos que você assista este [vídeo📹](https://vimeo.com/676111352/65e60b1274).

# Entrega do projeto

É hora de montar seu entregável! A seguir, faremos uma lista de verificação das etapas que recomendamos que você siga para atender aos requisitos.

* **Analisar e estruturar o relatório**

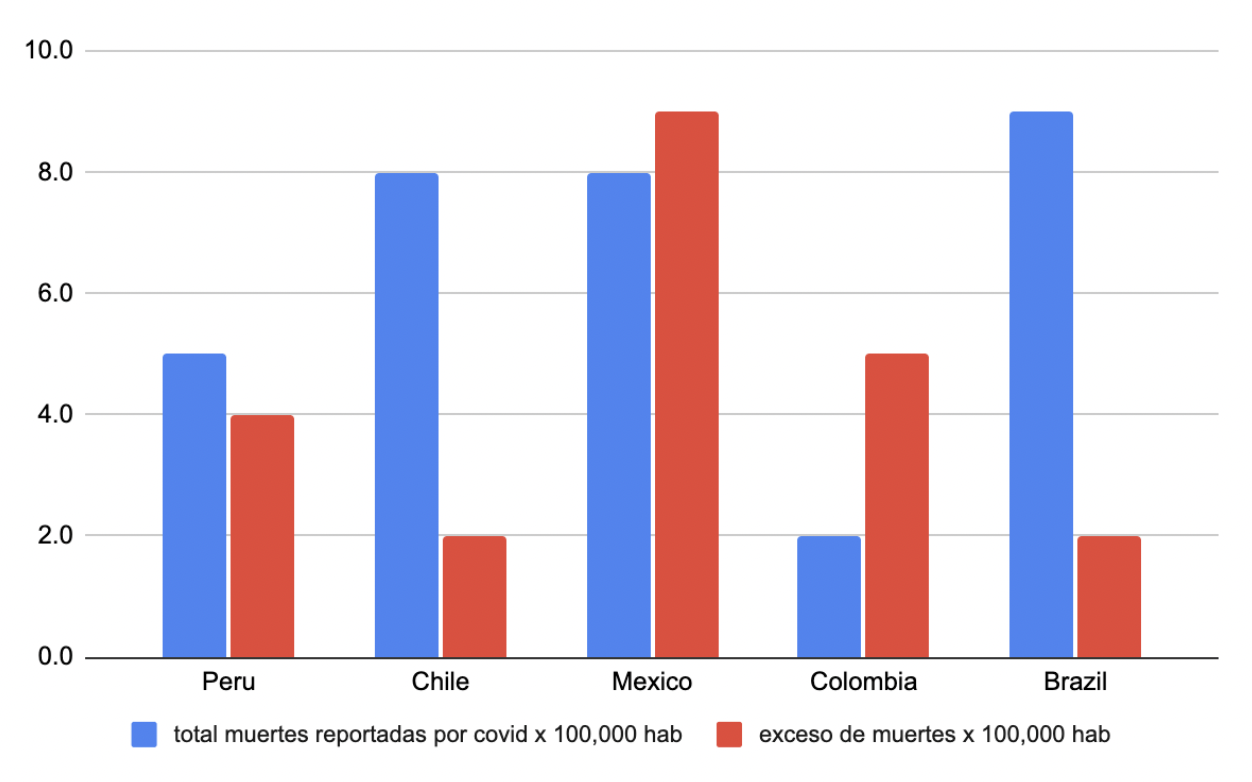
A análise desenvolvida gerará impacto desde que seja feita uma apresentação adequada. Se os dados não forem apresentados corretamente, é possível que eles passem despercebidos. Crie um relatório com os indicadores solicitados no [enunciado do projeto](https://docs.google.com/document/d/1nN47G0r3lcGTxa_KqLwZuM0xt2qgzqHjr1JRX_dRxLA/edit?usp=sharing). Deixamos um exemplo de como deveria ser apresentado cada um dos pontos solicitados no relatório:

* *Uma tabela comparativa com o total de mortes por COVID-19 notificadas por 100 mil habitantes, o total de “mortes em excesso” por 100 mil habitantes e a diferença entre os dois valores, por país:*

**

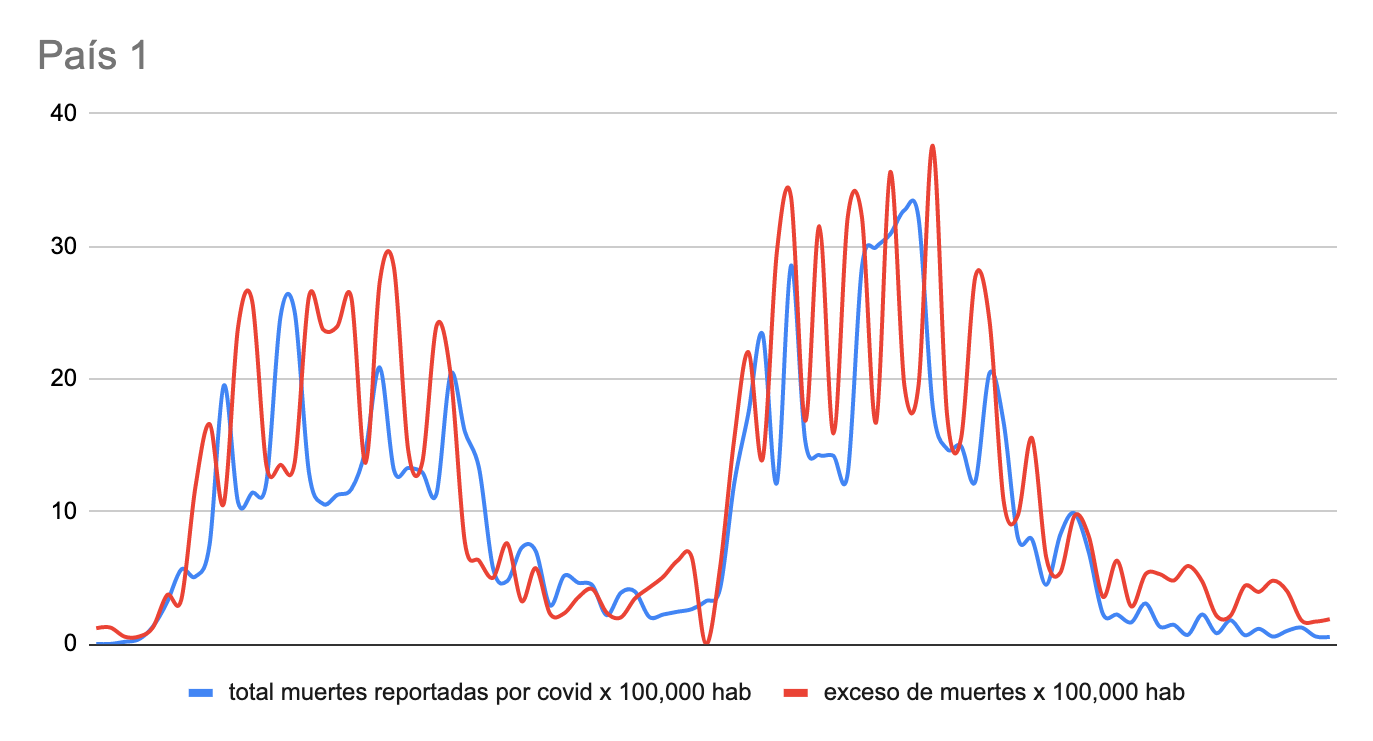
*Exemplo de tabela solicitada no ponto 1 do enunciado*

* *Um gráfico mostrando os dados da tabela acima como um gráfico de barras:*

**

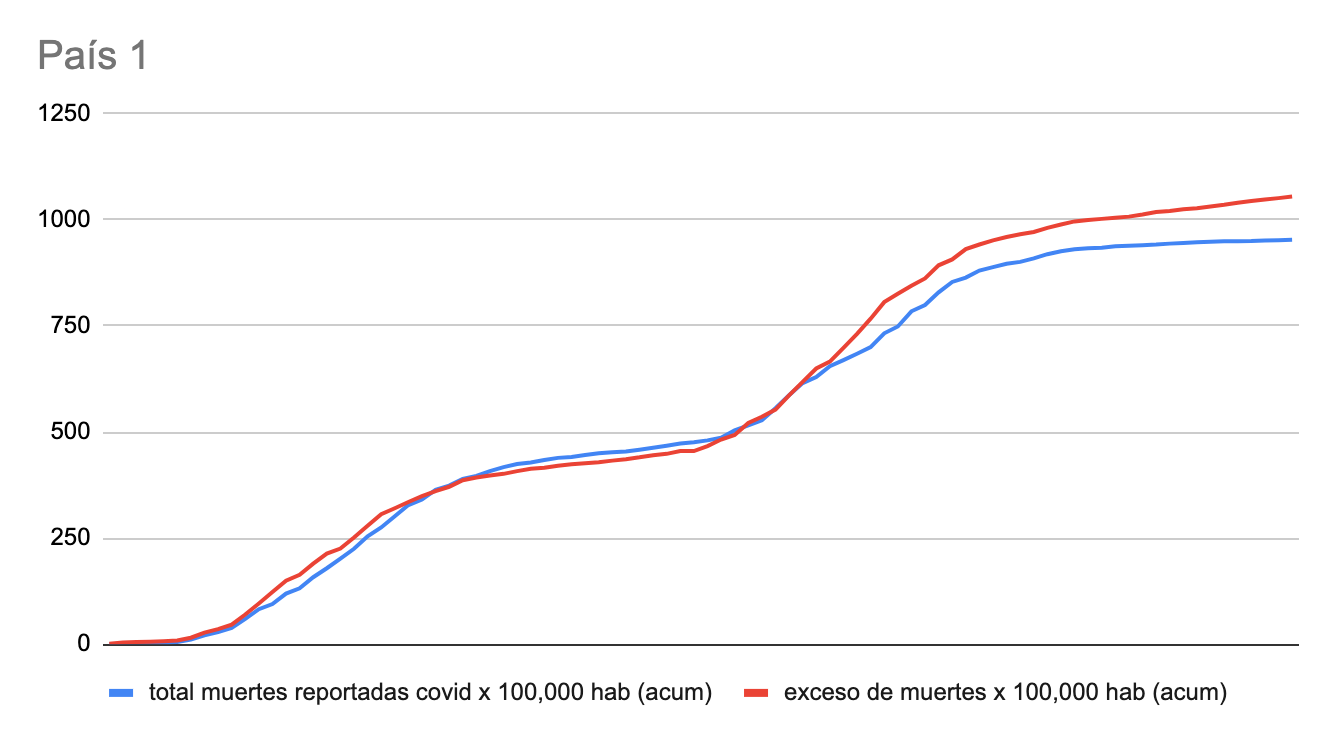
*Exemplo do gráfico solicitado no ponto 2 do enunciado (não é o gráfico real com os dados do projeto)*

* *Um gráfico para cada país que mostra a comparação da evolução ao longo do tempo das mortes notificados por COVID-19 versus o cálculo do "excesso de mortes" semana a semana, ambos dados por 100 mil habitantes:*



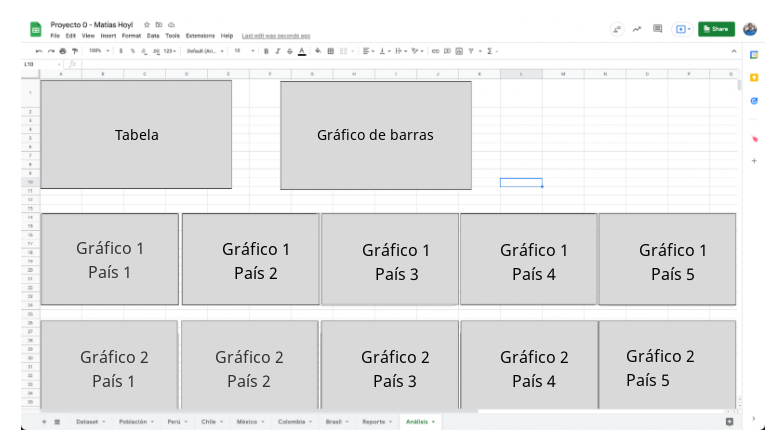
*Exemplo do gráfico solicitado no ponto 3 do enunciado para um país (não é o gráfico real com os dados do projeto). Lembre-se que você deve fazer este gráfico para cada um dos cinco países*

* *Um gráfico para cada país que mostra as mesmas variáveis ​​do ponto anterior, mas acumuladas ao longo do tempo:*

**

*Exemplo do gráfico solicitado no ponto 4 do enunciado para um país (não é o gráfico real com os dados do projeto). Lembre-se que você deve fazer este gráfico para cada um dos cinco países*

* *Com todas as informações acima, procure estruturar um relatório que resuma os gráficos e tabelas de forma ordenada em uma guia do seu relatório:*

**

* **Grave um vídeo falando sobre suas conclusões**

Uma boa analista deve ter a habilidade de explicar suas descobertas de forma clara e concisa. Para este projeto, você deve fazer uma apresentação de no máximo 3 minutos compartilhando suas conclusões e achados mais notáveis. Para fazer o vídeo recomendamos o uso do [Loom](https://www.youtube.com/watch?v=toj6_pxRqHE). Para saber como funciona a plataforma Loom, recomendamos que você assista a este [vídeo](https://www.youtube.com/watch?v=-8mwLqvNOPY).

* **Responda às perguntas de avaliação na plataforma de aprendizagem**

Depois de terminar seu relatório, é hora de validá-lo. Na seção "Entrega" da [plataforma de aprendizagem](http://app.laboratoria.la), você encontrará 3 questões onde deve escolher a resposta correta. Caso alguma resposta não esteja correta, a plataforma irá alertá-la com uma mensagem quando você marcar a opção “Entregar”. Por favor, **leia as perguntas com atenção.**

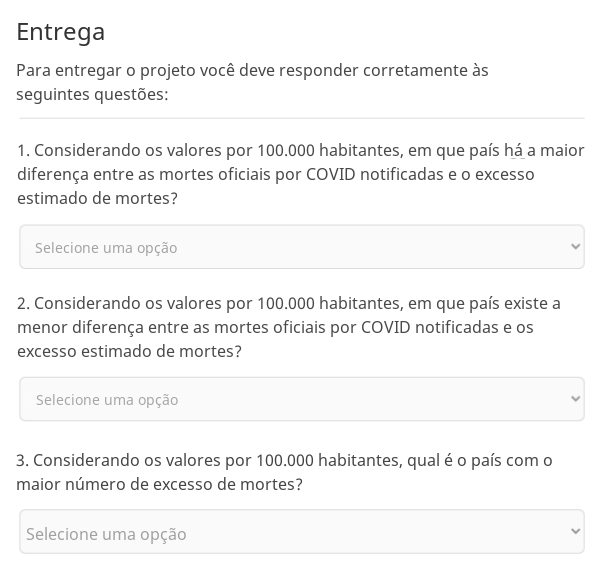
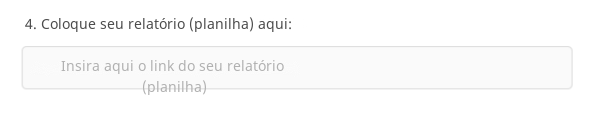
****

Figura “Perguntas de validação do projeto COVID”

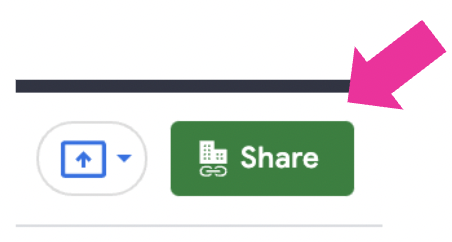
* **Obtenha o url da sua planilha**

Você já tem o relatório e as respostas ao questionário da plataforma de aprendizagem prontos, é hora de compartilhar seu relatório conosco na questão 4 do formulário de entrega:

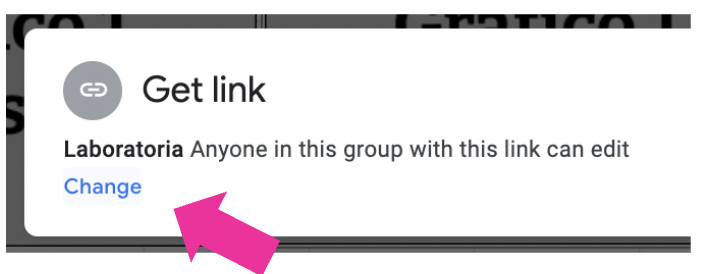
****

Para que o seu relatório fique publicamente visível (para que possamos revisá-lo), você deve seguir as seguintes etapas:

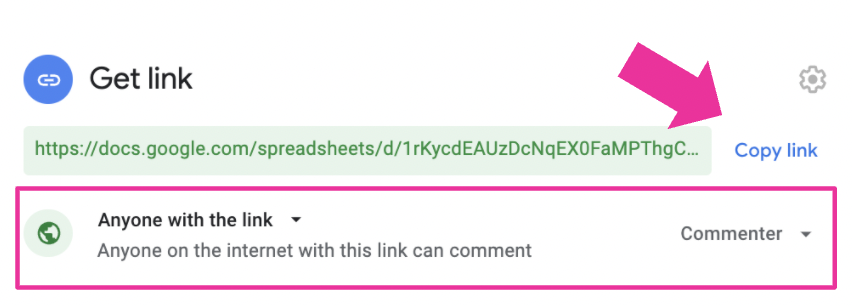
* No canto superior direito, você encontrará um botão verde que diz “Share”/ “Compartilhar”



* Dentro da caixa, você deve pressionar o botão Change/ Alterar

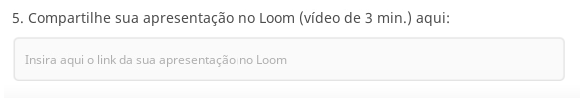


* Lá, você deve alterar a visibilidade do link para que qualquer pessoa com o link possa comentar (“qualquer pessoa com o link pode comentar”) e, em seguida, copiar o link para colá-lo na plataforma de aprendizagem.



* **Cole o link do seu vídeo na plataforma**

Na última pergunta da plataforma de aprendizagem, você deve colar o link para sua apresentação de vídeo no Loom.

****

O link será algo como:

**https://www.loom.com/share/5c4fce76d53e4347afa86fa6e5b0b9e2**

Depois disso, você está pronta! Parabéns por enviar seu projeto. Fique ligada na comunidade online do [Slack](https://join.slack.com/t/certificado-datos-lab/shared_invite/zt-zw3s89lh-lAJTLMEK3f3LHZyBLiMp9g) para apoiar suas colegas. A gente se encontra na sessão de onboarding (a data aparece no portal da [plataforma de aprendizagem](http://app.laboratoria.la)).