

Correlação x Causalidade -> MUITO CUIDADO COM ISSO

Correlação NÃO NECESSARIAMENTE implica em causalidade

Ex: Foi notado que em países do hemisfério norte que as pessoas tendem a gastar mais em compras durante a época de frio.

Correlação x Causalidade -> MUITO CUIDADO COM ISSO

O gasto em compras parece estar correlacionado negativamente com a temperatura
(- temperatura, + compras)

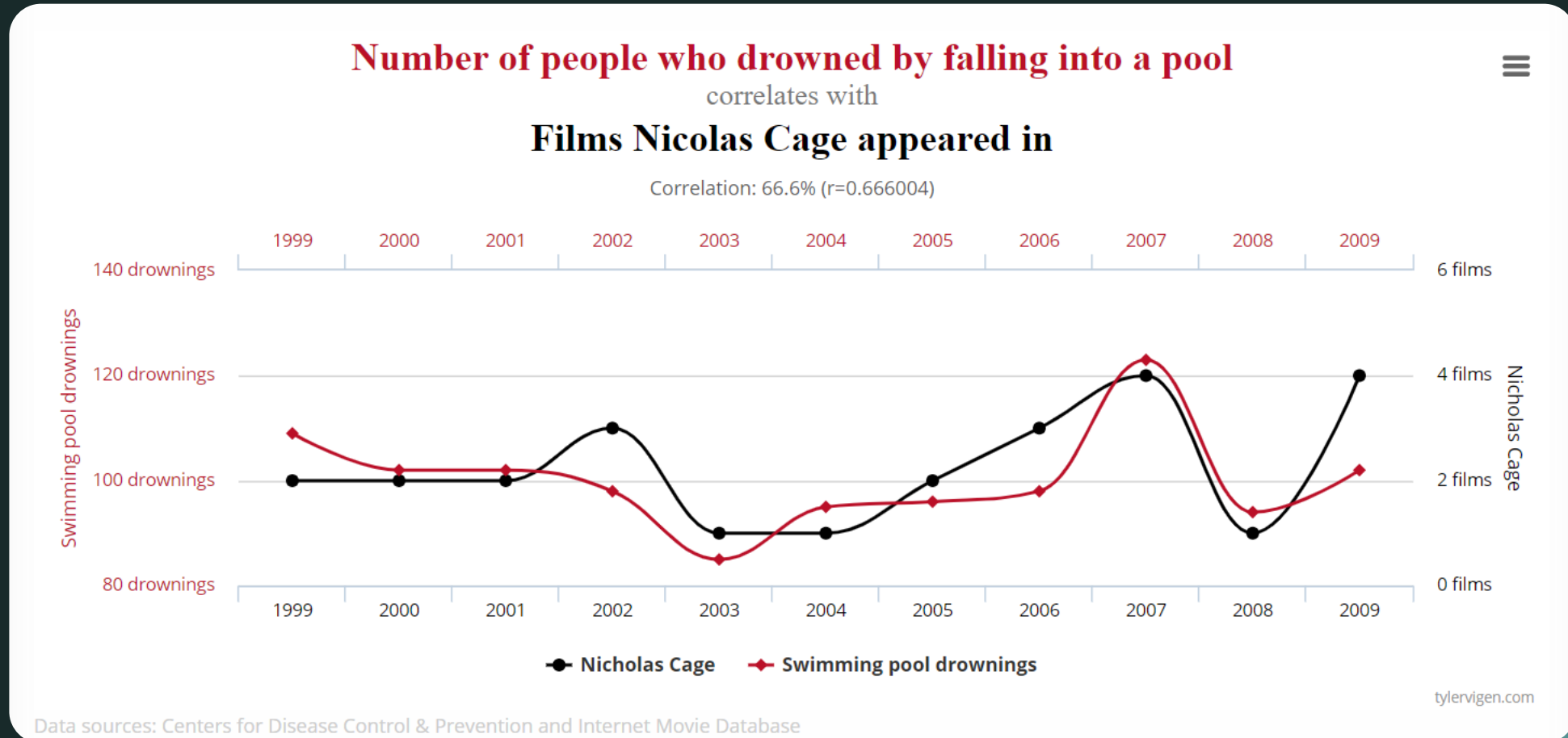
Mas isso não quer dizer que o frio cause um aumento nas vendas.

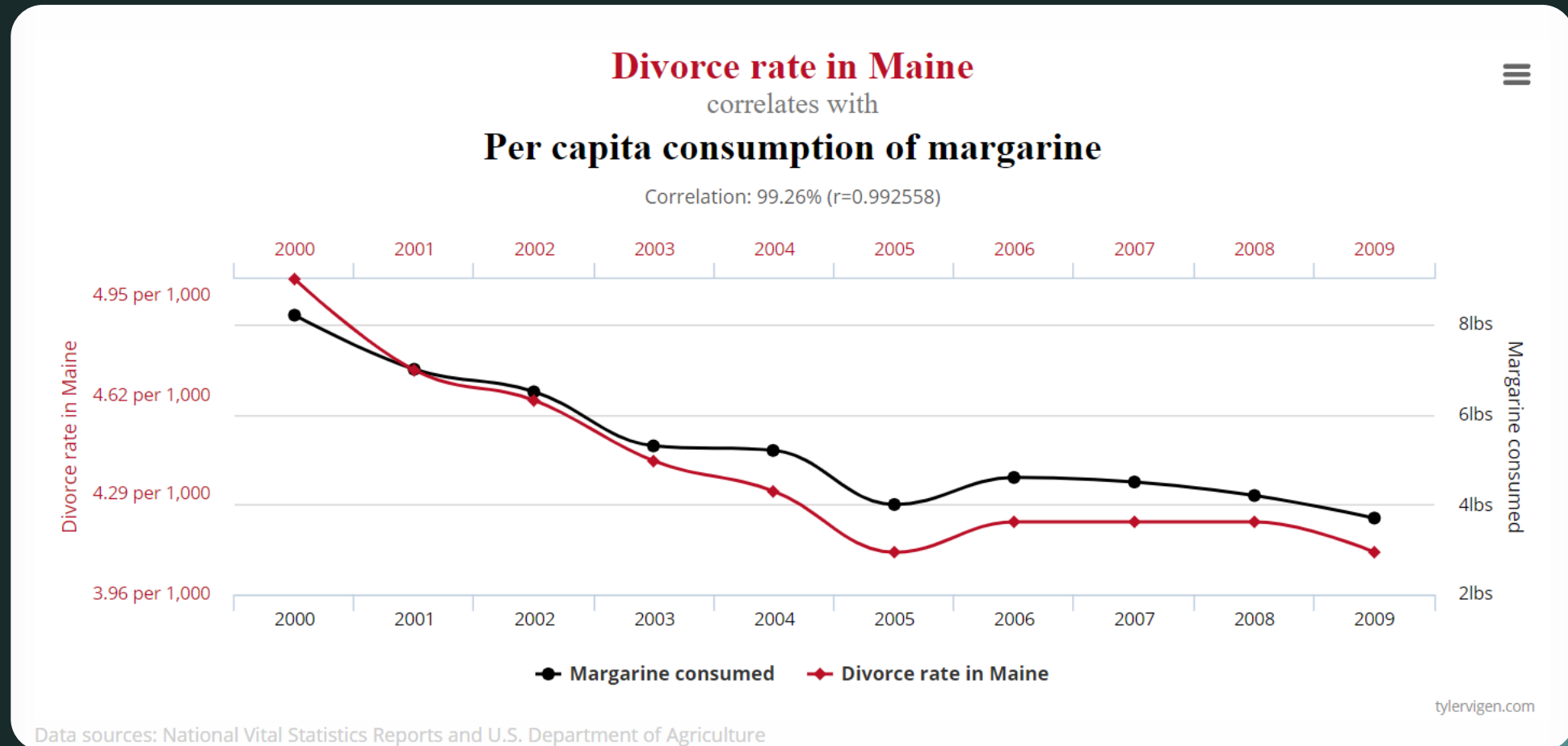
Uma explicação mais plausível que foi encontrada pra isso é que datas festivas — como o Natal, por exemplo — coincidem com épocas de frio nos países situados ao norte da Linha do Equador.

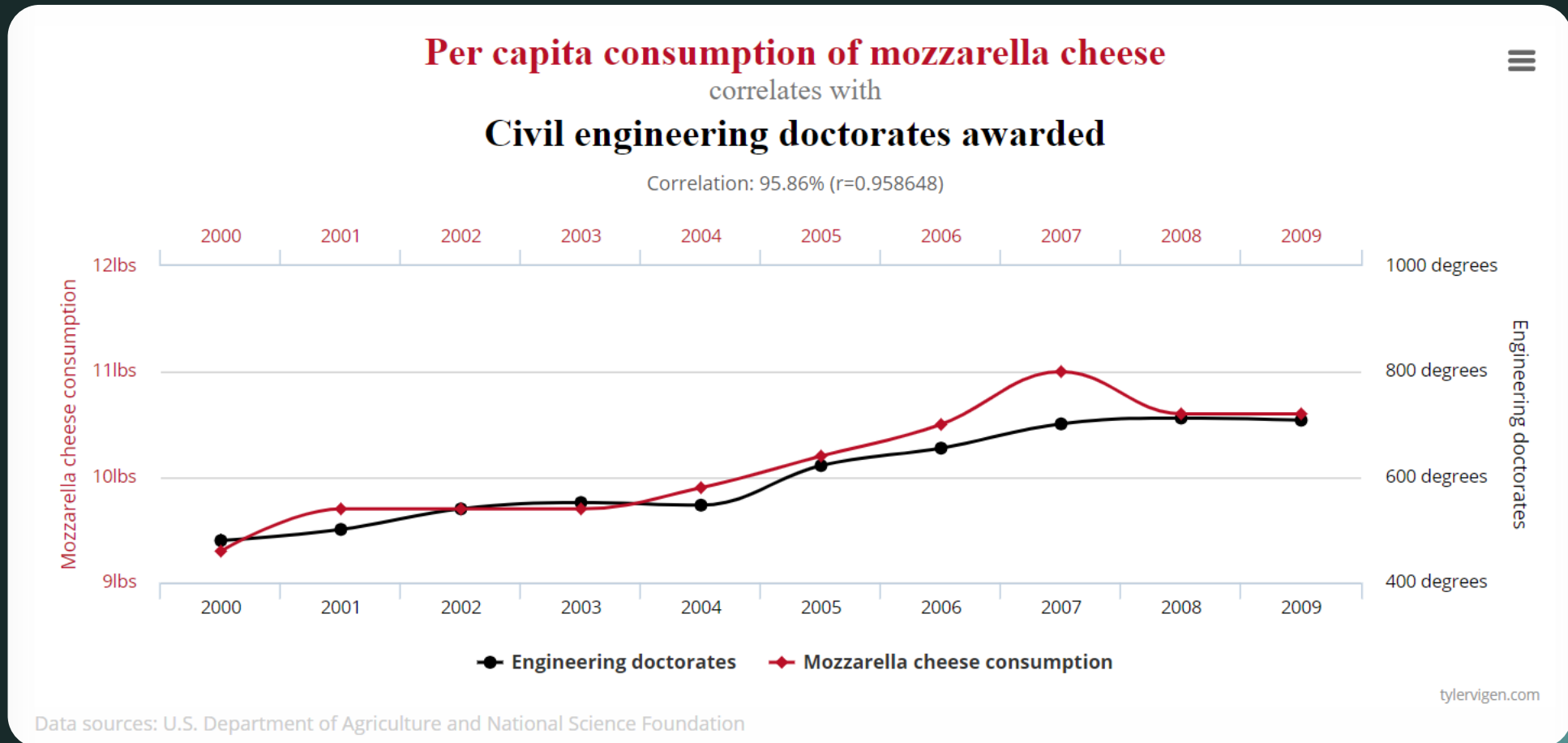
Processo de Análise de Dados

Ou seja, questione SEMPRE

- a) Levante as suas hipóteses
- b) Traga dados para comprovar essa hipótese
- c) Tome cuidado com a Correlação x Causalidade







Segundo exemplo:

No início do ano, estávamos analisando a receita da Hashtag, considerando o acumulado dos meses de janeiro e fevereiro. E percebemos que estava menor do que o acumulado de janeiro e fevereiro do ano passado.

O que não parecia fazer muito sentido, porque estávamos projetando um crescimento para 2023.

Objetivo: entender a queda de faturamento

Dados que impactam:

- > Vendas em eventos
- > Vendas fora de eventos
- > Vendas para empresas



Objetivo: entender a queda de faturamento

Dados que impactam:
80%-20%:

- > **Vendas em eventos**
- > Vendas fora de eventos
- > Vendas para empresas

Como esses 80-20 impactam no resultado?

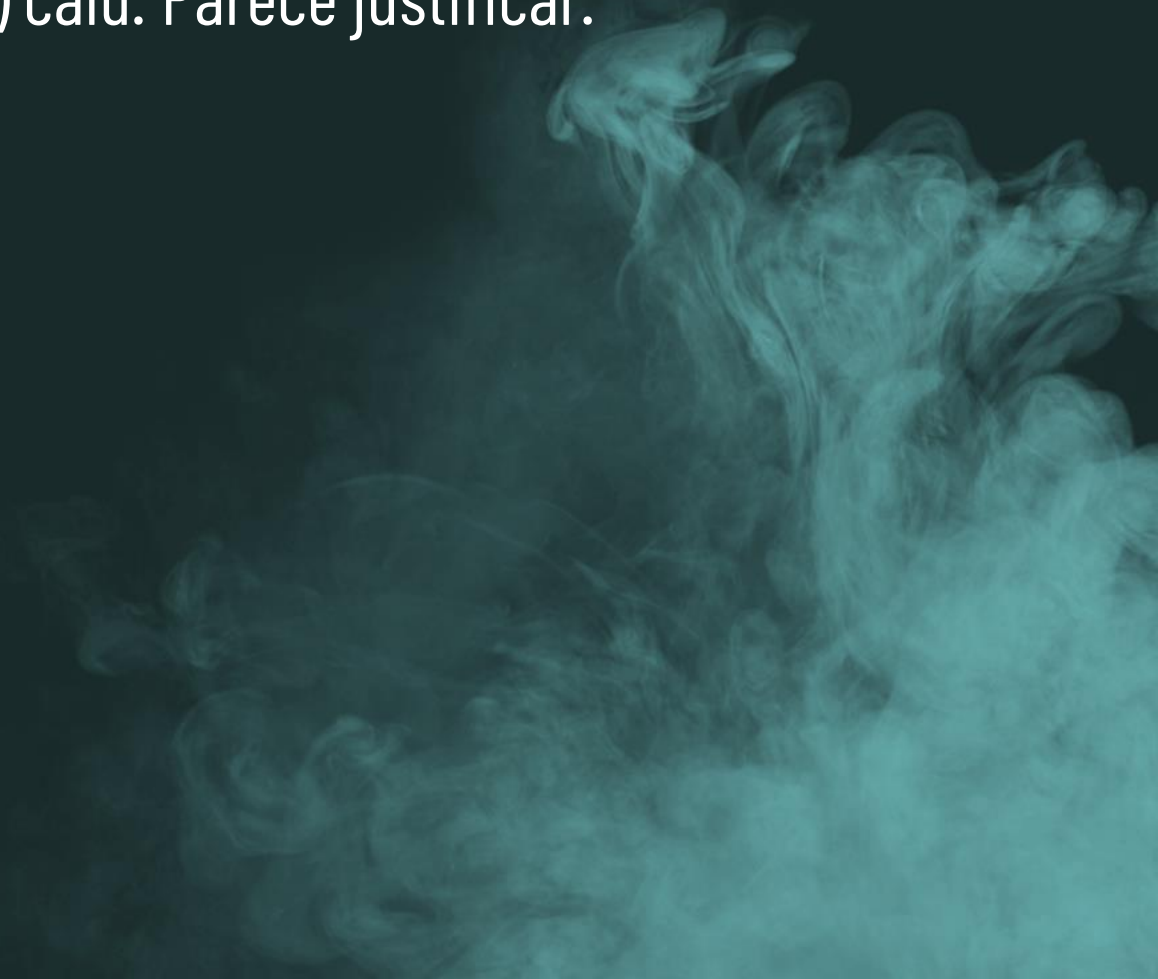
- > Comparecimento
- > Retenção entre as aulas
- > Vendas nos eventos



O comparecimento nos eventos (percentualmente) caiu. Parece justificar.

...

Mas lembra que tem que questiona sempre?



Mesmo assim sentimos que o faturamento estava muito descolado

E já que o grande 80-20 são os eventos, fomos analisar evento por evento.

E percebemos que, por conta da época do carnaval, fizemos menos eventos em 2023 do que em 2022.

Produto	Lançamento	Professor	Data
Python	LPY6	Lira	10/jan
Excel	L26	João	17/jan
Power BI	LPBI9	Alon	14/fev
Python	LPY7	Lira	07/mar
Excel	L27	João	21/mar

2022

Produto	Lançamento	Hashtag	Data
Python	LPY11	Programação	09/jan/23
Excel	L31	Treinamentos	23/jan/23
Power BI	LPBI13	Treinamentos	27/fev/23
Python	LPY12	Programação	13/mar/23

2023

Processo de Análise de Dados

Depois que a gente viu, parecia óbvio... E a gente pensou que tivesse acabado por aí.



Imagina que você trabalha na Netflix e você precisa analisar comparativamente o desempenho de duas séries:

Wandinha e Stranger Things



Processo de Análise de Dados

Se você tiver ainda um senso básico de análise de dados, o que você pode tender a olhar?

Quantos % da base de clientes da Netflix assistiu cada série



E faz sentido, porque você quer saber se a série fez sucesso ou não

E aí imagina que você identificou que tanto Wandinha quanto Stranger Things foram assistidos por 14% da base de assinantes da Netflix.

Parece que as duas tiveram o mesmo desempenho, certo?

Mas lembra... Questionar SEMPRE. "Po, mas senti que Stranger Things teve muito mais tração... Foi muito mais falado pela mídia, bateu vários recordes"

O que pode ter acontecido?

14% da base de 130 milhões de assinantes assistiram Stranger Things = 18,2 Milhões

14% da base de 220 milhões de assinantes assistiram Wandinha = 30,8 Milhões

Até aí tudo bem... A gente pode até pensar que Wandinha teve desempenho melhor porque teve mais assinantes brutos

Processo de Análise de Dados

Mas lembra que a gente tem a visão de negócios...

Por que a Netflix produz filmes e séries? Investe milhões e milhões de dólares todo ano, ao invés de usar só séries e filmes de terceiros pra fazer o streaming

Processo de Análise de Dados

- 1) Reter assinantes antigos
- 2) Atrair novos assinantes
- 3) Diferenciação no mercado
- 4) Expansão global
- 5) E muito mais...



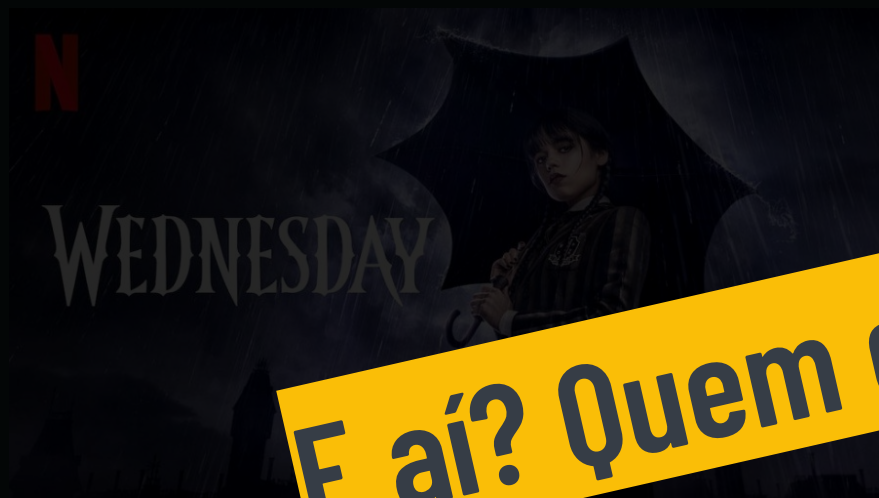
E se a gente analisasse o desempenho das duas séries já com essa ótica?



% Assinantes que assistiram = 14% (30,8 Milhões)
Novos assinantes = 1,5 Milhão (4,9% de crescimento de 29,3MM)
Churn médio = 2,2%
Churn no mês de lançamento = 1,3%

% Assinantes que assistiram = 14% (18,2 Milhões)
Novos assinantes = 2,4 Milhão (15% de crescimento de 15,8MM)
Churn médio = 1,9%
Churn no mês de lançamento = 1,1%

E se a gente analisasse o desempenho das duas séries já com essa ótica?



E aí? Quem ganha essa disputa?

% Assinantes que assistiram = 14% (30,8 Milhões)
Novos assinantes = 1,5 Milhão (4,9% de crescimento de 29,3MM)
Churn médio = 2,2%
Churn no mês de lançamento = 1,3%

% Assinantes que assistiram = 14% (18,2 Milhões)
Novos assinantes = 2,4 Milhão (15% de crescimento de 15,8MM)
Churn médio = 1,9%
Churn no mês de lançamento = 1,1%