

6 - Vídeo: O Que É Ciência De Dados (I)

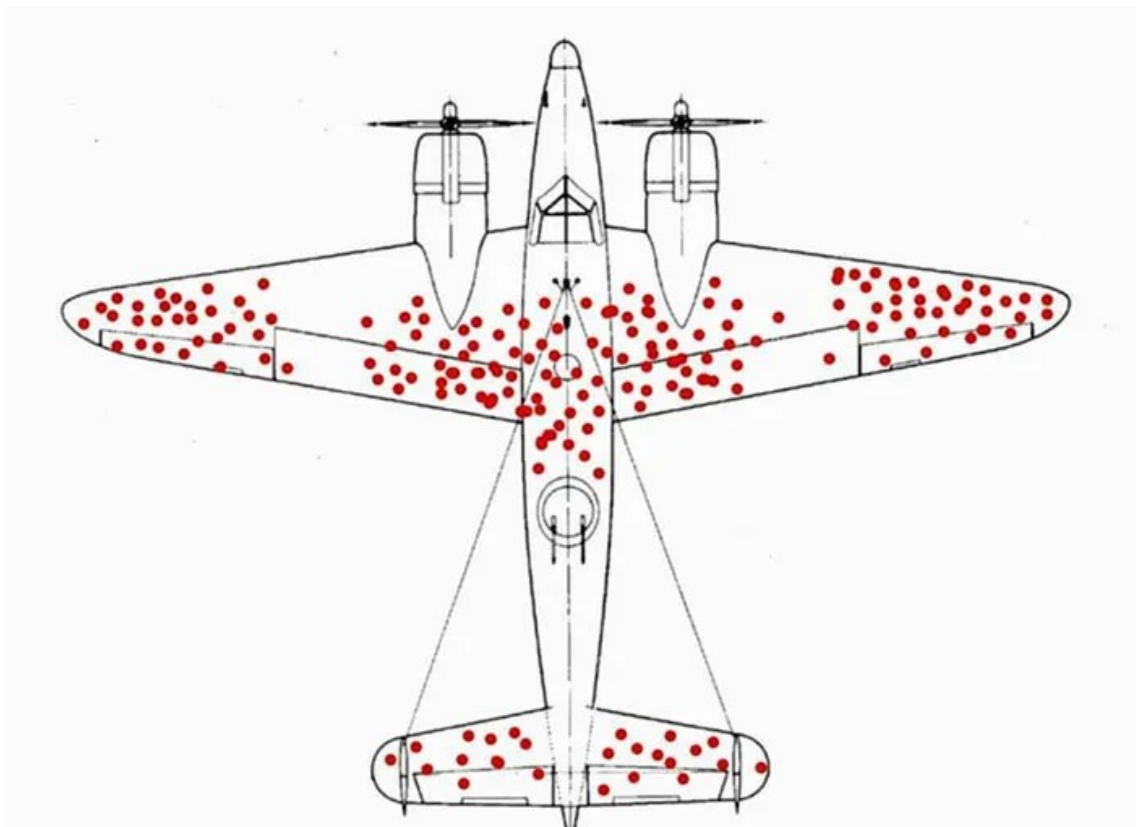
Ronny Gabryel Colatino de Souza

Descrição da atividade:

Abraham Wald:

O primeiro vídeo do Nerdologia começou com uma história que eu nunca imaginei que teria a ver com ciência de dados. Durante a Segunda Guerra Mundial, um matemático chamado **Abraham Wald** recebeu uma missão que parecia simples: descobrir onde colocar blindagem nos aviões ingleses. O problema é que blindagem é pesada, então não dava pra blindar o avião inteiro senão ele nem decolava.

A primeira ideia que qualquer um teria (incluindo eu) seria: vamos ver onde os aviões que voltaram levaram mais tiros e blindar essas partes, né Faz todo sentido Mas aí que tá o pulo do gato.



Durante a guerra, ele analisaram os aviões que voltavam das missões. Notaram que muitos deles tinham furos nas asas e na fuselagem, mas o motor e o sistema balístico estavam quase sempre intactos. A princípio, parecia lógico reforçar as

áreas com mais furos. Mas aí veio o verdadeiro insight esses aviões conseguiram voltar **apesar** dos danos. Já os que **não voltaram** provavelmente foram atingidos em pontos críticos como o motor ou o sistema de armas e por isso caíram.

A conclusão foi surpreendente, mas genial: **era preciso reforçar justamente as áreas onde os aviões que voltaram não tinham furos**, porque esses eram os pontos onde, se atingidos, não havia chance de sobrevivência.

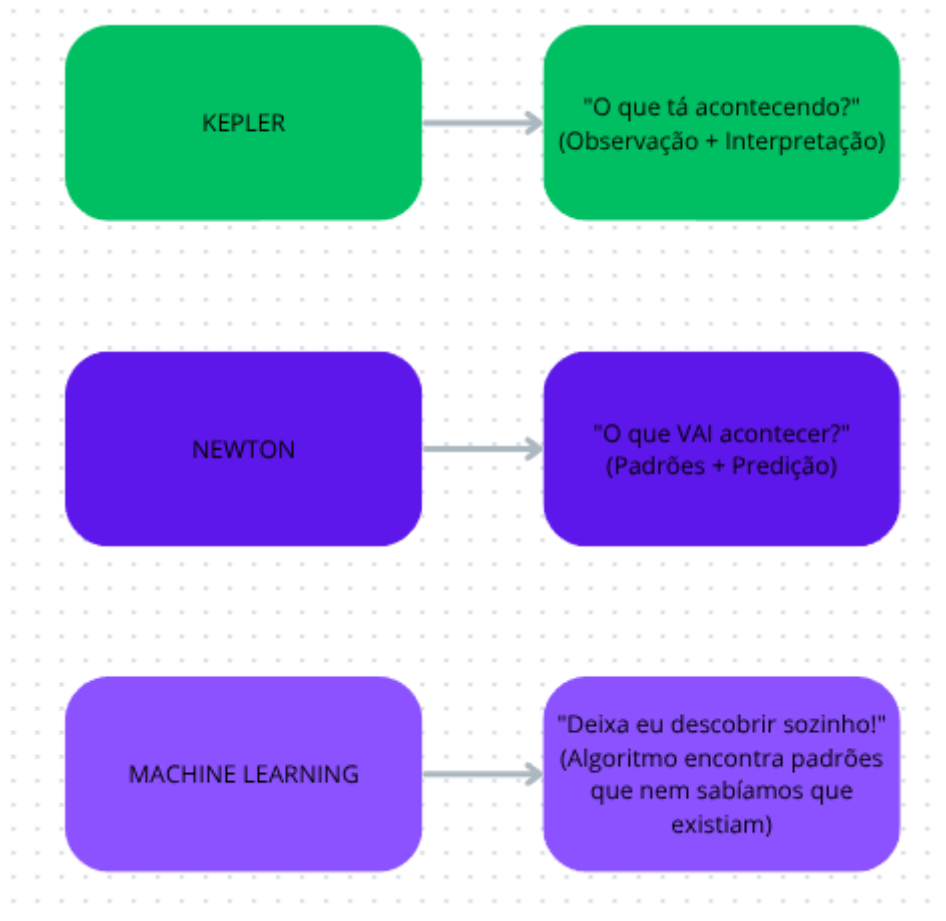
Cara, quando entendi isso, foi tipo um "click" na minha cabeça. Wald percebeu que os aviões só conseguiam voltar justamente porque **não** foram atingidos nas partes vitais. Os furos que eles viam eram nas partes que o avião conseguia "aguentar" levar tiro e ainda assim voar.

Como o vídeo explicou, seria tipo criar um colete à prova de balas nas pernas só porque no hospital você vê mais gente baleada nas pernas do que nos órgãos vitais. Óbvio que quem leva tiro nos órgãos vitais não chega vivo no hospital pra ser contado

De Kepler a Newton: A Evolução do Raciocínio

O vídeo fez uma analogia muito boa com a astronomia que me ajudou a entender melhor. Kepler ficou lá observando os planetas pelo telescópio e conseguiu calcular que a gente gira em órbitas elípticas em torno do sol. Ele interpretou os dados pra entender o que tava rolando.

Já Newton foi mais longe. Pegou essas informações do Kepler, botou mais observações na conta e não só entendeu o que tava acontecendo, como criou uma fórmula que consegue **prever** onde a gente vai estar em qualquer ponto da órbita. Tipo, os foguetes e satélites que a gente usa hoje ainda seguem essas regras que ele descobriu.



Isso me fez pensar na evolução do conhecimento primeiro a gente observa, depois interpreta, e só então consegue prever. É exatamente isso que a ciência de dados faz hoje, mas de forma muito mais rápida e com muito mais dados.

Machine Learning: A Ferramenta que Faz as Próprias Perguntas

Uma coisa que o vídeo falou que me marcou foi: “Agora a gente tem uma ferramenta que consegue achar resposta mesmo quando não sabemos que pergunta fazer. Isso é o machine learning.”

Antes, com estatística tradicional, a gente tinha que saber o que perguntar pros dados. Agora o algoritmo pode vasculhar os dados e descobrir padrões que nem

passaram pela nossa cabeça. E não só números organizadinhos textos, vídeos, qualquer coisa.

O exemplo que me impressionou foi o da Target que descobriu que uma consumidora tava grávida antes do pai dela saber, só porque ela comprou hidratante. Tipo, que loucura é essa mas ok né interessante isso

Mas aqui vem um problema sério que o vídeo destacou “**Dados ruins geram resultados ruins**”. No machine learning isso fica ainda pior Se você treina um programa pra reconhecer motos usando só fotos de pessoas montadas em motos, ele pode entender que "moto" significa as duas coisas juntas. Aí quando ele vê uma moto vazia, não reconhece.

Para finalizar esse primeiro, o vídeo foi bem honesto numa parte como essas informações vêm de todo lado (textos, vídeos, fotos de satélite), o grosso do trabalho de um cientista de dados pode ser gasto garantindo que os dados estão limpos. Não é a parte glamourosa, mas é a mais importante

O segundo vídeo foi um balde de água fria. O cara foi direto ao ponto: **"Data science não é sobre fazer modelos complicados, não é sobre fazer visualizações incríveis, não é sobre escrever códigos. Data science é sobre usar dados para criar o maior impacto possível pra sua empresa"**.

Ele falou uma coisa que me fez pensar a mídia só mostra uma parte da ciência de dados (IA, deep learning), e isso cria expectativas irreais. Mesmo em empresas gigantes como Google, Apple, Facebook, a realidade é bem diferente do que a gente vê por aí.

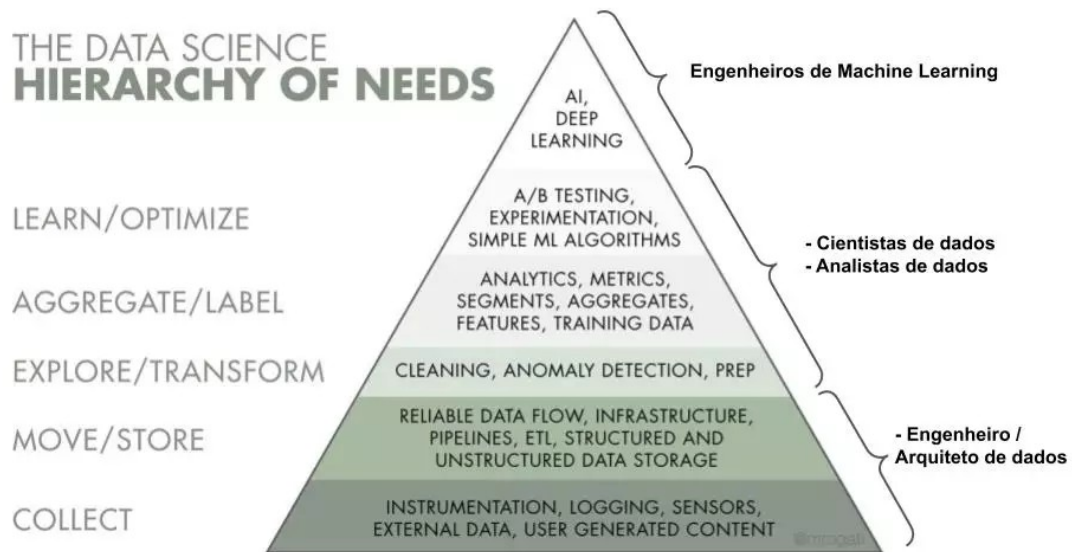
A evolução foi interessante de acompanhar:

1996: Já tinha o termo "Data Mining" - descobrir informações úteis em bancos de dados.

2001: William S. Cleveland expandiu esse conceito. Ele combinou ciência da computação com data mining e fez a estatística ficar muito mais técnica. Basicamente, conseguiu usar o poder do computador pra estatística, e chamou isso de "Data Science".

2005-2010: Nasceu a Web 2.0. MySpace, Facebook, YouTube - todo mundo curtindo, comentando, compartilhando. E adivinha? Isso gera MUITO dado. Tanto que as tecnologias tradicionais não conseguiam processar direito.

2010+: Era do Big Data. Dados demais pras ferramentas antigas, então tiveram que inventar MapReduce, Hadoop, Spark só pra dar conta.



O segundo vídeo mostrou uma pirâmide que meio que destruiu minhas ilusões. A parte de IA e deep learning que todo mundo acha que é ciência de dados Tá lá no topo da pirâmide. A base Coletar dados, limpar dados, organizar dados.

E dependendo do tamanho da empresa, você vai fazer coisas completamente diferentes:

Empresa pequena: Você faz tudo, desde coletar até limpar dados. IA Esquece, você vai estar ocupado explicando pra empresa o que fazer com os produtos deles.

Empresa média: Agora tem engenheiro de software pra coletar dados, engenheiro de dados pra organizar, e você fica com a parte de análise. Aqui te pedem pra ser um super-herói porque os problemas são mais complexos.

Empresa grande: Cada um faz uma coisinha específica. Você pode passar a vida toda só fazendo analytics e nunca tocar em IA. Tipo uma linha de produção, mas de dados.

Cara, isso me pegou desprevenido. Eu achava que cientista de dados passava o dia criando algoritmos super complexos e fazendo gráficos incríveis. Na real, a maior parte do tempo é gasta garantindo que os dados são limpos e organizados. É a parte "chata", mas é a mais importante.

O que mais me surpreendeu foi descobrir que **a única coisa igual em todas as empresas é que você precisa ser um resolvidor de problemas.** Não importa

se você tá numa startup ou no Google, no final das contas te contratam pra resolver as paradas mais difíceis e guiar na direção certa.

O que mais me impressionou foi descobrir que existe uma linha tênue entre o que é popular na mídia e o que a indústria realmente usa. Todo mundo fala de IA e deep learning, mas na prática a maioria das empresas te contrata pra fazer coisas bem mais básicas e que impactam muito mais o negócio.

Confesso que fiquei meio frustrado quando descobri que 70% do trabalho pode ser limpeza de dados. Eu tinha uma visão muito romântica da profissão imaginava que era só insight genial atrás de insight genial.

Mas pensando melhor, isso é até bom. Significa que não preciso ser um gênio da matemática pra começar a contribuir. Posso começar aprendendo a organizar dados direito, fazer as perguntas certas e ir evoluindo aos poucos.

Conclusões

Esses vídeos me abriram os olhos pra algumas coisas importantes. Primeiro, que fazer as perguntas certas é mais importante que ter as ferramentas mais avançadas. O Wald resolveu um problema complexo com estatística básica, mas com raciocínio brilhante.

Segundo, que ciência de dados é muito mais sobre resolver problemas reais do que sobre mostrar o quanto você é inteligente com algoritmos complexos. No final das contas, se os dados não gerarem impacto real pra empresa, não servem pra nada.

E terceiro, que a área não é tão glamourosa quanto eu pensava, mas isso não é necessariamente ruim. Significa que posso começar do básico aprender a limpar e organizar dados direito e ir crescendo aos poucos.

O que mais me motiva agora é saber que tô vivendo um momento histórico. Nunca tivemos tantos dados disponíveis, e ao mesmo tempo nunca precisamos tanto de pessoas que saibam fazer as perguntas certas. É uma oportunidade única de fazer diferença de verdade.

Como o vídeo disse, o papel do cientista de dados só tem aumentado ao longo do tempo, e quanto mais entramos no mundo digital, mais importante fica extrair insights pras empresas melhorarem seus produtos. A profissão se tornou uma peça-chave no sucesso de qualquer negócio, seja pequeno ou grande.

Referencias

NERDOLOGIA. O que é ciência de dados | Nerdologia Tech. 2019. Disponível em: [https://www.youtube.com/watch?v=\[URL\]](https://www.youtube.com/watch?v=[URL]). Acesso em: 30 de julho de 2025.

WHAT REALLY IS Data Science: Told by a Data Scientist. Disponível em: [https://www.youtube.com/watch?v=\[URL\]](https://www.youtube.com/watch?v=[URL]). Acesso em: 30 de julho de 2025.

E a Imagem do avião tirada de: <https://www.boredpanda.com/world-war-2-aircraft-survivorship-bias-abraham-wald/>

E a Imagem da Hierarquia foi tirada de: <https://datauniverse.com.br/a-hierarquia-das-necessidades-em-data-science-e-seus-grandes-papeis/>