

## **Relatório card 6 - O'Que é ciência de dados**

**Jefferson korte junior**

### **Descrição:**

Nesses dois vídeos aprendemos o'que é ciência de dados, e que não é simplesmente utilizar uma ferramenta para montar gráficos de dados, ciência de dados vai muito além de você apenas saber analisar dados, é entender os dados e interpretá-los de maneira correta para assim voce poder causar um impacto positivo no contexto ou no que você está analisando.

No primeiro vídeo do nerdologia ele cita um ótimo exemplo sobre Abraham Wald na segunda guerra mundial, que tinha como objetivo colocar blindagens em aviões para proteger dos tiros, Com dados sobre as regiões que mais receberam tiros nos aviões eles percebeu que na verdade eram bem essas regiões que não precisavam de blindagem pelo fato de que eram esses aviões que tinham voltado, então a região do motor estavam sem marca de projéteis, e os que tinham tiros de projéteis no motor, não tinham voltado para a base. Isso é um bom exemplo do que é ciência de dados.

É um fato que o aprendizado de máquina é uma excelente ferramenta para Compreender padrões em dados, mas é apenas uma ferramenta, diante da quantidade massiva de dados nos quais somos constantemente expostos diariamente surge a necessidade de utilizar novos métodos para manipular e processar esses dados e compreender outras questões de Ciência de Dados.

Já no segundo vídeo com uma visão mais aprofundada do funcionamento desse campo de atuação, o cientista de dados Joma explica a construção da ideia do que é a ciência de dados, assim como deve se comportar o cientista de dados, executando sua função de maneira que utilize os dados para gerar resultados significativos, tais como produtos ou até mesmo estratégias.

Joma traz também o histórico da criação do conceito de 'Data Science' onde inicialmente se chamava 'Data Mining', no qual consistia em extrair informações úteis de dados. Só que a partir de 2001, com William Cleveland foi dado o nome e o conceito de 'Data Science', isso devido a trazer o poder estatístico da ciência da computação aliada ao Data Mining.

Ele também fez uma reflexão sobre a mídia e como é tratado a ciência de dados, ele comenta que a ciência de dados é tratada somente de aplicações de técnicas como aprendizado de máquina, muito raramente ressaltando o papel fundamental e importante de quem atua nessa área, de ser um estrategista, ser um solucionador de problemas, independente de quais habilidades e técnicas sejam usadas.

Joma também mostra e explica através de uma pirâmide hierárquica que nos permite compreender as áreas de atuação e o processo de um cientista de dados.

# THE DATA SCIENCE HIERARCHY OF NEEDS

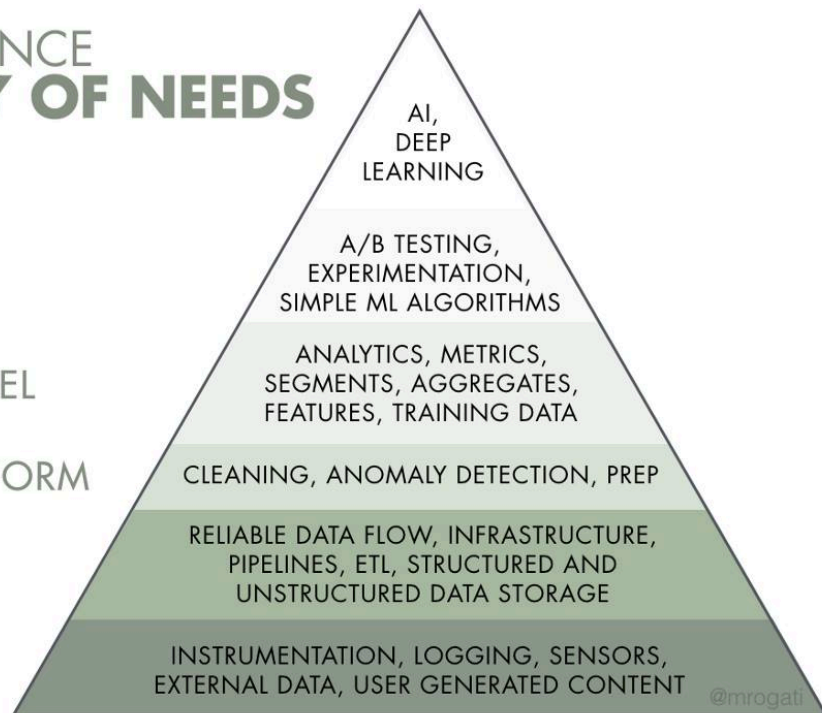
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



Como podemos perceber temos etapa de coleta, armazenamento, exploração, rotulamento e otimização. Dentro dessa pirâmide temos algumas etapas que muitas vezes é desconhecida ou ignorada quando vista por pessoas de fora da área, no caso são as etapas de análise, métrica e testes.

Onde para o cientista de dados, é essencial. E claro um bom cientista deve compreender essa necessidades principalmente as relacionadas a AGGREGATE/LABEL e LEARN/OPTIMIZE antes de utilizar o Aprendizado de máquina

## Conclusões:

Um cientista de dados busca resolver problemas que necessitam a análise aprofundada de dados, estabelecendo sentido para eles, dado o contexto, e auxiliando na tomada de decisão, para produzir resultados significativos. Para isso tudo deve ser utilizado ferramentas e fazer um bom uso da estatística.

## Referências:

<https://www.youtube.com/watch?v=ykSILAQQu6o>

<https://www.youtube.com/watch?v=xG-c7E5PK0Y>