

Relatório 6 - O Que É Ciência De Dados (I)

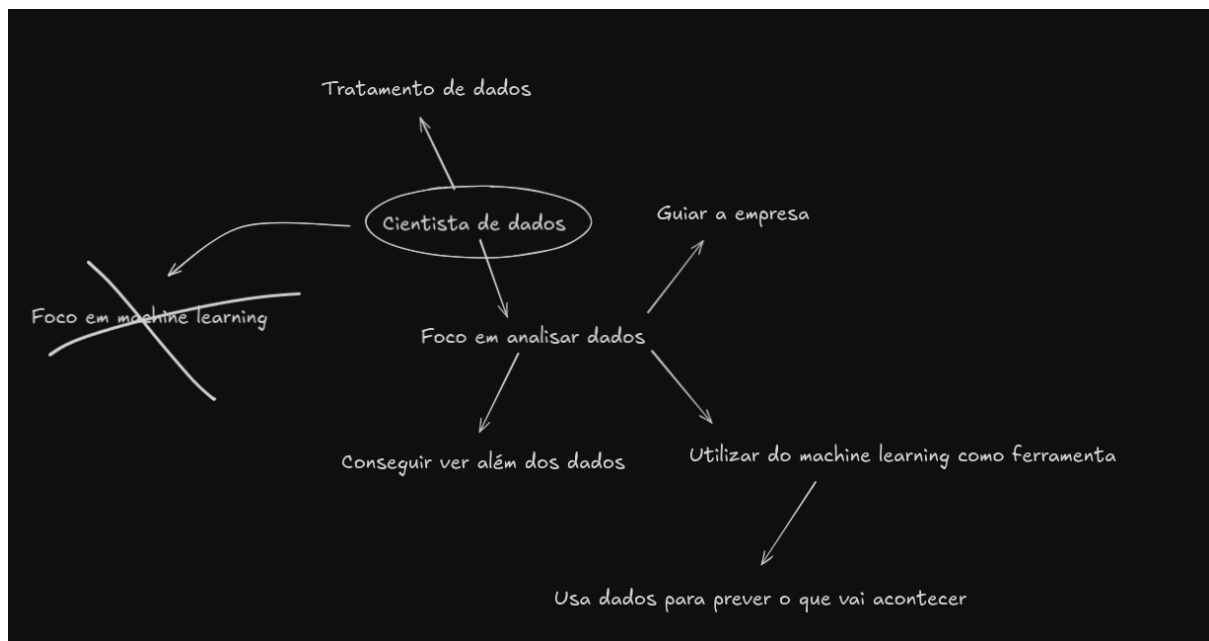
Felipe Fonseca

Descrição da atividade

<Nesta seção o participante deverá descrever com suas próprias palavras os temas abordados nas tarefas do card. É importante demonstrar que houve entendimento dos temas, isto mediante evidências com imagens, gráficos e textos originais.>

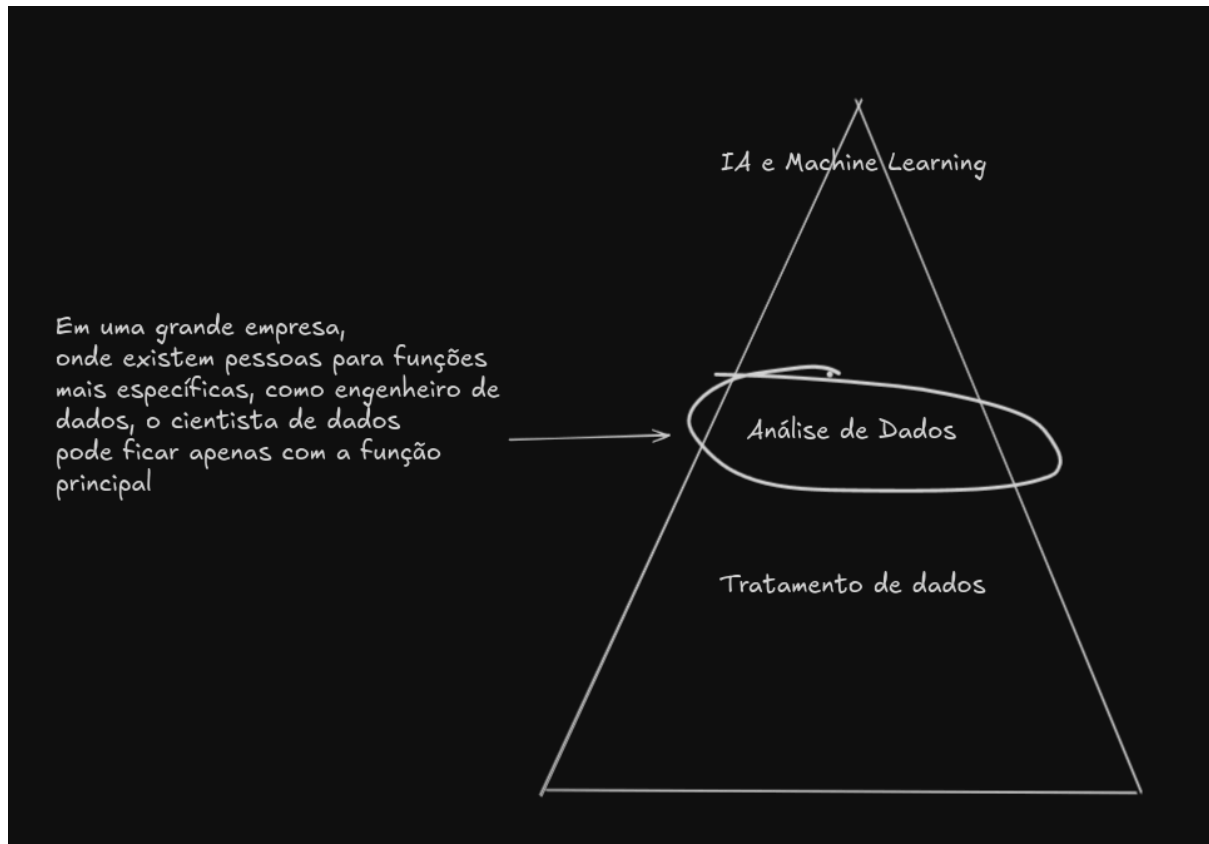
No primeiro vídeo do card tinha como tema a área de data science. Inicialmente o apresentador começa falando sobre a análise que um cientista fez sobre qual área do avião deveria ser blindada. Através dos dados de onde os aviões estavam baleados ele conseguiu chegar a conclusão de que as áreas que não estavam baleadas eram porque, quando eram atingidas, os aviões caíam e não voltavam, e por isso nunca chegavam até ele para a análise. Esse caso foi utilizado para exemplo de como um cientista de dados deve pensar, pois não adianta analisar apenas as informações trazidas pelos dados, invés disso é importante traçar uma relação entre os dados e o contexto geral, analisando as possibilidades e pensando além dos dados.

Após isso, o apresentador também fala sobre a quantidade de dados que se tem hoje em dia, e na possibilidade de utilizar modelos de algoritmos para, com base nos dados que temos, conseguir prever o que vai acontecer. E aí que entre o cientista de dados, responsável por garantir que os dados estão todos nos formatos corretos, o cientista de dados tem que analisar a estrutura dos dados que serão enviadas para treinar o modelo de machine learning, isso passa por avaliar a formatação, se os dados estão consistentes, se os dados vão ser o suficiente para que a IA entenda realmente o tema, etc. Isso porque os dados nem sempre estão em números, como visto no card anterior existem dados categóricos por exemplo, ou então gráficos em vídeos e em imagens, o cientista de dados é responsável por cuidar de todo esse tratamento desses dados não estruturados.



No segundo vídeo o apresentador já começa com uma frase impactante, dizendo que ser um cientista de dados não é sobre escrever códigos, e sim sobre criar impacto para a empresa utilizando dados. A tese do segundo vídeo é de que Data Science é a capacidade de descobrir coisas analisando dados utilizando da computação. Ele também diz sobre a confusão que há sobre o uso do termo, que muitos pensam em data science como algo puramente focado no machine learning, quando na verdade o mais importante para o cientista de dados é a análise dos dados em si.

Basicamente o foco principal do vídeo é sobre a importância do cientista de dados em analisar os dados para guiar a empresa, e que sua principal importância é essa.



Conclusões

As principais conclusões que eu consegui tirar desses dois vídeos foram que cientista de dados é muito mais do que um programador ou um treinador de IA, e sim alguém responsável por analisar os dados e conseguir guiar a empresa para o melhor rumo a partir dos resultados obtidos. Além disso é de suma importância que o cientista de dados consiga abstrair mais do que os dados mostram, e isso pode ser feito de várias formas diferentes, a ferramenta em si não é tão importante.

Referências

O que é ciência de dados (Nerdologia Tech):

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

What really is data science? (Told by a Data Scientist):

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