**1) O que é Ciência de Dados? O que você entende quando escuta o termo Ciência de Dados? Mostre um exemplo prático de um projeto de Ciência de Dados que você já trabalhou ou você conhece?**

Por definição, podemos citar ciência de dados como: “um campo multidisciplinar que usa métodos, processos, algoritmos e sistemas para extrair conhecimento e insights a partir de dados estruturados e não estruturados.

Para mim, ciência de dados é uma área de conhecimento destinada à geração de informações que embasem e facilitem o processo de tomada de decisão. O trabalho do cientista de dados pode ir desde a obtenção desses dados (com técnicas de Data Mining, por exemplo), até a projeção visual destes.

Já trabalhei com bases de dados históricos de uma empresa do ramo madeireiro, que ia de 2016 até 2021. O objetivo deste trabalho foi a partir destes dados, conceber um rearranjo para os armazéns da empresa.

Durante o período de pesquisa na UFPE, trabalhei com uma base de dados policial, que continha boletins de ocorrência das delegacias de estádios de futebol em Recife. O objetivo era ajudar o poder público a traçar estratégias de mitigação de atos violentos nos estádios.

**2) Ciência de Dados é um campo multidisciplinar que envolve várias disciplinas de outras áreas. Explique qual a relação de Ciência de Dados com as principais disciplinas envolvidas?**

Estatística & Ciência de dados: A estatística é um fundamento indispensável para área, uma vez o primeiro passo para elaboração de um planejamento do que virá a ser feito com os dados é compreender esses dados. A estatística nos dá um panorama geral dos dados, mostrando como a base se comporta, como os dados se correlacionam uns com os outros. A partir desse entendimento é que se tem um ponto de partida para demais estudos.

Séries Temporais: As séries temporais tratam dados que se apresentam ao longo de um período de tempo. Diferente da Análise de Regressão, em séries é possível incorporar fatores de sazonalidade e tendência, que enriquecem as predições e análises a serem feitas pela ciência de dados.

**3) Explique a relação entre o processo Knowledge Discovery in Databases (KDD), Mineração de Dados e Ciência de Dados? Ciencia de Dados é a mesma coisa de Mineração de Dados?**

KDD é um processo interativo e iterativo usado na (e para a) mineração de dados com o objetivo de reconhecer padrões nos dados. Através de ferramentas de KDD, pode-se, por exemplo, seccionar a base dados de forma a trabalhar-se apenas com aquilo que é pertinente para receber tratamento e depois tornar-se o produto final, informações.

**4) CRISP DM (Cross Industry Standard Process for Data Mining) é um processo padrão que pode ser utilizado na indústria para mineração de dados. Explique as etapas de metodologia do CRIPS DM e discuta suas vantagens e desvantagens no desenvolvimento de um projeto de Ciência de Dados.**

Business Understanding: É o passo inicial do CRISP. Trata-se do conhecimento que se deve obter do modelo de negócio em que se planeja aplicar Data Science. É impossível criar uma boa solução para um negócio, baseada em dados, sem antes ter-se o conhecimento do negócio.

Data Unerstanding & Data preparation: Engloba todo o processo de obtenção e tratamento inicial das bases de dados.

Modeling: Começa-se nesta fase a identificar quais técnicas de modelagem devem ser usadas para alcançar o objetivo traçado no inicio do planejamento, podendo, dependendo do problema, aplicar técnicas puras ou ainda a mesclagem entre diversas.

**5) O cientista de dados se tornou um profissional muito requisitados nas empresas devido à capacidade de gerar informações relevantes para auxiliar o processo de tomada de decisão em negócios. Ao longo do tempo, muitas habilidades são requisitadas para o cientista de dados, consequentemente, outros nomes de profissionais sugiram na literatura. Explique a diferença entre as habilidades necessárias para se tornar um Cientista de Dados, Analista de Dados, Engenheiro de Dados e Engenheiro de Machine Learning?**

Cientista de Dados: É um profissional mais completo, que deve possuir conhecimentos de programação, do negócio, de banco de dados, estatística e matemática e além disso, deve ser capaz de comunicar bem, aos diversos públicos as estratégias que traça para os negócios.

Analista de Dados: É um profissional mais voltado a elaboração de relatórios. Ele deve ser capaz de condensar as informações que recebe de forma a torna-las de fácil compreensão para os gestores. É comumente exigido deste profissional domínio em PowerBI e demais ferramentas do pacote Office.

Engenheiro de Dados: É um profissional que trabalha menos com o cliente final, sendo geralmente responsável por desenvolver soluções e depois expandi-las, tendo uma atuação mais generalista

Engenheiro de Machine Learning: É um profissional que trabalha menos com o cliente final, sendo geralmente responsável por desenvolver soluções e depois expandi-las, porém, diferente do Engenheiro de Dados, é mais direcionado a área de Machine Learning.

**6) A área de Ciência de Dados acumulou ao longo do tempo diversas aplicações em diferentes domínios. Cite pelo menos cinco aplicações de Ciência de Dados que você tem interesse de tabalhar ou já trabalhou?**

Na realidade a área que me chama mais atenção é a consultoria. Assim é possível lidar com os mais diversos tipos de bancos de dados, trabalhando com criação de modelos preditivos complexos para redes de compras e logística; criação de soluções de melhorias em processos produtivos.

Também acredito ser possível usar DataScience em conjunto com **técnicas de modelagem multicritério**.

**7) Você precisa desenvolver várias habilidades se quiser se tornar um Cientista de Dados. Mostre quais são as habilidades necessárias para se tornar um Cientista de Dado? Quais dessas habilidades você precisa aprender? Quais dessas habilidades você já domina?**

Um cientista de dados precisa desenvolver uma gama de soft e hard skills. Dentre as habilidades técnicas, deve se especializar em conhecer linguagens de programação e ferramentas de gestão de banco de dados. É preciso não se prender a apenas uma metodologia e/ou linguagem.

No quesito técnico, já programo em R e estou aprendendo Python, no entanto, ainda preciso aprender a trabalhar com dados em larga escala e conhecimentos específicos de DataScience, como machine learning.

Nas habilidades não técnicas, o profissional precisa ter boa comunicação e expressão, sabendo lidar com o trabalho em grupo, tendo em vista que muitas equipes formadas para resolução de problemas são multidisciplinares. Então, cabe a um bom cientista de dados saber lidar com especialistas dos mais diversos ramos de forma amistosa e harmoniosa, ainda mais quando toma papel de líder dessas equipes. Além disso, deve ser uma pessoa com bom raciocínio lógico e objetividade.

Nestas habilidades, tenho me dado bem devido às experiências com liderança e público na vida pessoal, mais precisamente, na igreja, onde lido com jovens, idosos, adultos e crianças. No entanto, ainda sinto que preciso me desenvolver mais nesse aspecto.