

MÓDULO 1 - Acelera DEV Data Science Codenation

Professor: Túlio Vieira de Souza

<https://www.linkedin.com/in/tuliovieira/>

3. O que é Ciência?

Começa com a observação do nosso “mundo” observado. As observações são neutras, não interferimos no mundo ao observá-lo.

Trabalha em cima do raciocínio **lógico Indutivo** (Proposições, conclusões e regras)

René Descartes: Obra o discurso do método “Penso, logo existo”

Edgard Morin: Pensamento Complexo
O viés da variável omitida

Método Científico:

- Fazer uma observação
- Formular uma hipótese
- Realizar um experimento
- Analisar os Dados
- Reportar as descobertas
- Convidar terceiros para replicar os resultados.

4. Big data - um grande problema

Explosão no volume de dados no mundo, principalmente com o advento do IoT (Internet of Things)

5 V's of Big Data:

- Variedade (Dados estruturados ou não estruturados)
- Velocidade
- Volume
- Veracidade (Confiança, verificar e validar os dados)
- Valor

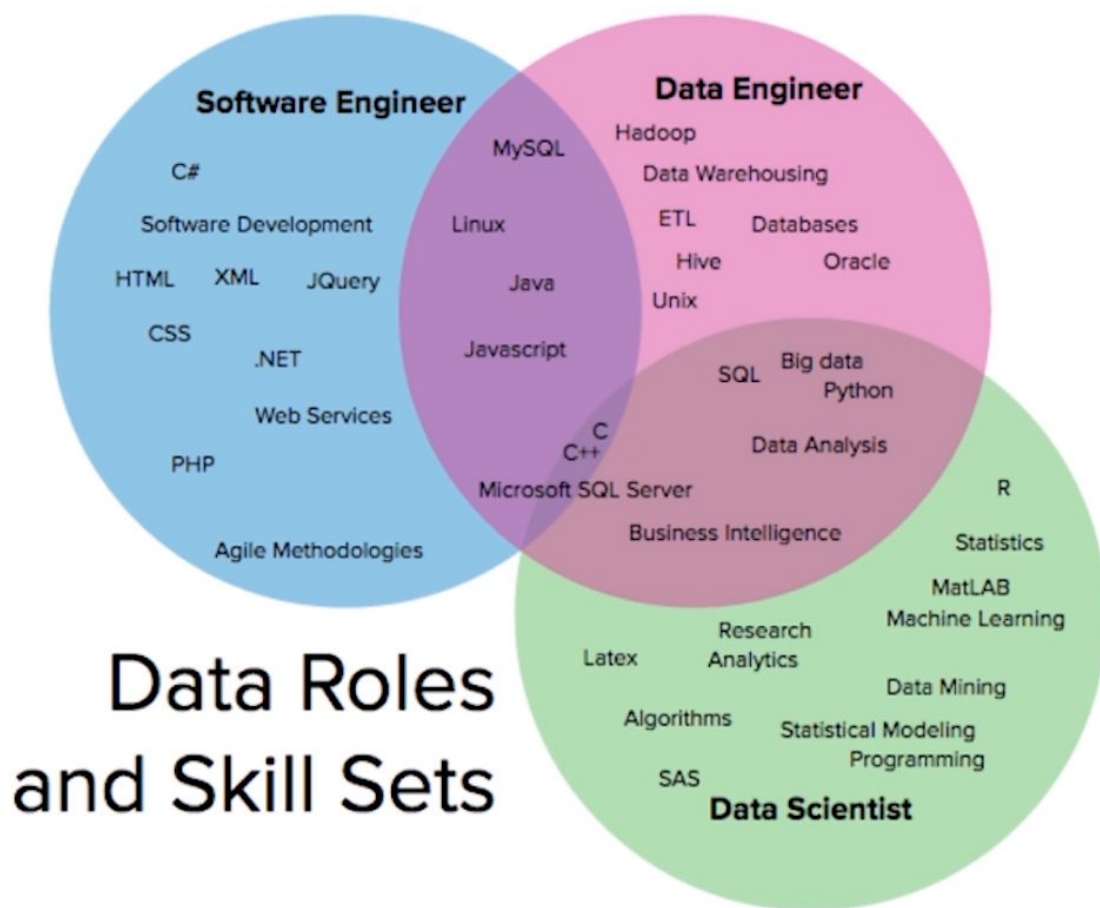
Governança de Dados:

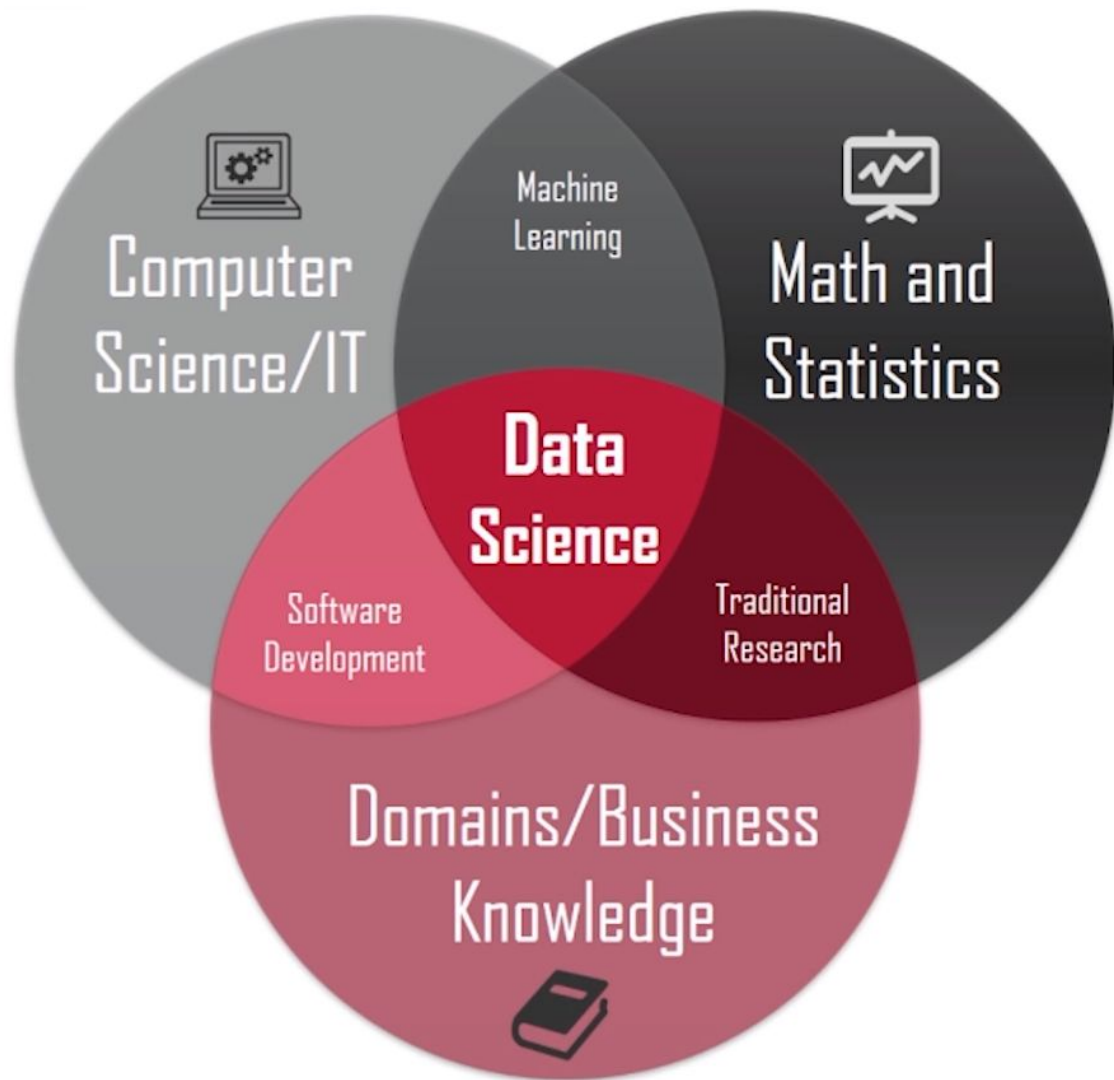
- Para onde vou (em termos de negócio), como vou, quando vou?
- Que dados serão necessários nesse caminho?
- Como obtê-los, como mantê-los?
- Que áreas serão prioritárias no tratamento dos dados, baseado no negócio?
- Leitura recomendada: [DMBOK](#)

5. Papéis dentro de um projeto de dados

Top Data Science Jobs:

- **Data Science Manager**
- **Data Scientist**
- **Data Architect:** Conceber o projeto de dados por inteiro, inicia os projetos de pesquisa. Entender qual a melhor forma “tecnologicamente” de entregar a solução ao negócio.
- **Data Engineer:** Focado em processos de manipulação dos dados “ETL”
- **Estatístico:**
- **Decision Scientists:** Perfil Sênior, como entregar o ML ao negócio, conhecimento do negócio.
- **Engenheiro de Machine Learning:** Interface Ciência de Dados e Desenvolvimento de Software.





6. Tipos de Soluções de Análises

Solução descritiva:

Descrever o que existe, olhar para o passado/atual. “**O que aconteceu / está acontecendo nos meus negócios?**”

Solução diagnóstica:

Descreve os porquês, olhando apenas para o passado/atual. “**Por que isso aconteceu / está acontecendo?**”

É muito importante validar com pessoal de negócio para entender através da experiência das pessoas.

Solução Preditiva:

Descreve o que vai existir;

Usa o passado para prever/predizer o futuro. “**O que vai acontecer?**”

Previsão: subconjunto da predição vinculado ao tempo.

Predição: Dizer de uma forma genérica o que vai acontecer no futuro.

Solução Prescritiva:

Recomenda uma tomada de decisões, recomenda uma ação. “O que vou fazer?”

7. Tipos de Problemas

Classificação Binária: Preveem um resultado binário.

- Este e-mail é SPAM ou não é?
- Esta revisão foi escrita por um cliente ou por um robô?
- Regressão logística.

Classificação multiclases: Permite gerar previsões para várias classes.

- Este produto é um livro, um filme ou vestuário?
- Este filme é uma comédia, documentário, suspense, etc...
- Regressão logística multinomial.

Regressão: Prever um valor numérico

- Qual será a temperatura amanhã?
- Quantas unidades deste produto serão vendidas?
- Regressão linear.

Agrupamento/Clusterização: Modelo aprende grupos / divisões nos dados.

- Quantos perfis de cliente minha empresa tem?
- Quantos grupos de alunos eu tenho na minha classe?
- KMeans

Sistemas de Recomendação: Retorna uma sugestão / recomenda algo à um usuário.

- Qual filme recomendar à um usuário baseado em seu gosto?
- Qual filme recomendar à um usuário baseado no gosto de pessoas.
- Collaborative Filtering / Content Based

8. Tipos de Aprendizado

Aprendizado Supervisionado: Onde eu ensino ao meu algoritmo através de **exemplos** o que ele deve aprender.

O objetivo é aprender uma regra geral que mapeia as entradas para entender as saídas.

Aprendizado não Supervisionado: O objetivo é que o algoritmo entenda os padrões e descubra as saídas sem a necessidade de exemplos prévios.

9. Maturidade com Analytics

Raw Data (Dados Crus): Pouca maturidade em relação aos dados. Os dados trazem pouco valor à companhia.

Information: Os dados trazem algum valor para a companhia. Pouca centralização e gestão das informações.

Business Intelligence: Tem uma boa visão sobre o passado e uma gestão de dados bem definida. Já possui rotinas de ETL e Data Warehouse por exemplo.

Advanced Analytics: Conhecimento profundo, já possui uma boa manipulação dos dados. Já consegue extrair previsões dos dados.

Em todo início de um projeto de Data Science é importante avaliar o nível de maturidade analítica atual da empresa, para assim implementar os processos de evolução.