

INTRODUÇÃO À CIÊNCIA DE DADOS

**Mercado Profissional para Ciência
de Dados**

QUEM É O CIENTISTA DE DADOS?

Importante ressaltar que há um **conjunto de nomenclaturas** e principalmente de funções que podem ser atribuídas a quem estuda Ciência de Dados, mas é importante saber separar as formações e atividades a serem desenvolvidas.

Importante saber também que por ser uma área recente, até o mercado (empregadores e clientes) **não faz ideia de como contratar, ou ainda, como chamar o profissional** que estuda Ciência de Dados.

Dentro do próprio segmento de Ciência de Dados, há uma variedade de funções que podem ser exercidas dentro de um projeto ou ainda de um time de Ciência de Dados.

CIENTISTA OU ENGENHEIRO DE DADOS?

Duas das atividades principais no contexto de DS, que às vezes são executadas pela mesma pessoa ou por grupos distintos, dependendo do projeto, da empresa, do time...e de muitas outras variáveis.

Essa disciplina tem como objetivo principal formar o que chamamos de **Cientista de Dados**, isso não impede que com estudos adicionais o mesmo profissional se qualifique para ser um **Engenheiro de Dados**.

CIENTISTA DE DADOS

Profissional capaz de fazer análises aprofundadas, dominar linguagens de programação, pensar em modelos a serem aplicados, dominar fluentemente probabilidade e álgebra linear, experimentar novas técnicas matemáticas e estatísticas, pensar em modelos apropriados de visualização dos dados e principalmente ter a capacidade de compreender o contexto dos dados que está trabalhando.

Em geral uma pessoa com capacidade técnica para operar um Jupyter Notebook com Python ou RStudio com R.

PAPEL DO CIENTISTA DE DADOS

- Planejar a coleta dos dados;
- Analisar os dados;
- Criar hipóteses;
- Escolher, definir e/ou criar um modelo estatístico;
- Realizar muitos testes;
- Demonstrar resultados (visualização), contar histórias (story telling);
- Estruturar um produto (definir como será a cadeia de alimentação e de consumo dos dados);
 - Serializar o seu código para alimentar um outro sistema;
 - Usar container;
 - Criar API;

ENGENHEIRO DE DADOS

Profissional capaz de operar tecnicamente ferramentas e sistemas que possibilitem coletar, manipular, transformar e limpar dados brutos.

Inclui aqui todo o conhecimento de ferramentas, bancos de dados, frameworks e sistemas de infraestrutura de grande escala.

É papel do Engenheiro de Dados construir um pipeline para que o modelo preditivo criado pelo Cientista de Dados possa ser executado com boa performance e em larga escala.

TECNOLOGIA - ENGENHEIRO DE DADOS

- Em geral os engenheiros de dados estão mais aptos para lidar com algumas tecnologias como:
 - Bancos de Dados NoSQL: ElasticSearch, Hbase, Cassandra, MongoDB, Neo4j, Apache Phoenix (engine) entre outros...
 - Frameworks e Ferramentas:
 - Apache Spark (linguagem Scala), Apache Hadoop, Apache Storm, Apache Mahout, OxData h2o – Computação distribuída e paralelismo
 - Apache Hive - Datawarehouse

TECNOLOGIA - ENGENHEIRO DE DADOS

- Apache Flume, Kafka, Sqoop – Coleta, agregação e transferência de grande quantidade de dados
- Solr, Lucene – Indexação e Pesquisa
- Docker, Swarm e Kubernetes – virtualização e containers
- TensorFlow – Machine Learning
- Gestão de Servidores em Nuvem (entre outros)
- Um cientista de dados também pode estudar tudo isso, mas não é o foco!

ANALISTA OU CIENTISTA DE DADOS

- Como já abordamos anteriormente, geralmente o papel de analista de dados é mais atribuído a profissionais que atuam com BI, o profissional de DS é nomeado como Cientista de Dados (data scientist).
- Há uma grande confusão a respeito de nomes de profissão, o que se faz efetivamente e principalmente uma dificuldade entre os gestores em saber que tipo de profissional precisam.
- Fique sempre atento a oportunidades do mercado de trabalho.

OFERTAS DE EMPREGO

- Meios tradicionais:
 - LinkedIn (mantenha perfil atualizado), o profissional de dados é alguém que sabe o valor de um perfil nesse tipo de rede.
 - Sites de emprego (alguns internacionais, inclusive com vagas para atuação de brasileiros fora do Brasil)
 - Indeed
 - Simply Hired

OUTROS MEIOS

- Além dos meios tradicionais, o mundo está requisitando profissionais de DS o tempo todo, é talvez a profissão que mais precise de profissionais.
- Há muitas maneiras de se destacar com esse tipo de trabalho e receber ofertas de trabalho.
 - Participação em listas de discussão.
 - Construção de portfólio rico
 - Enriquecer currículo com competições
 - Atuação como freelancer

LISTAS DE DISCUSSÃO

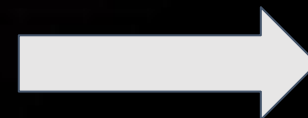
- É sabido que para trabalhar na área de DS é preciso estudar muito, estar sempre se atualizando e principalmente abarcar a maior quantidade de conhecimento possível sobre tecnologias e ferramentas que envolvem DS.
- Isso dá condições de se destacar entre os demais.
- Uma das maneiras de obter destaque é participar ativamente da comunidade de developers, lista de discussão (como stack overflow por exemplo), ou criar conteúdo com blog (medium por exemplo), instagram entre outros.

PORTFÓLIO

- Durante os estudos de DS é muito provável que você desenvolverá muitos projetos, reproduzindo algo que alguém fez, testando com dados públicos ou privados (há muitos repositórios com conjuntos de dados que podem ser utilizados).
- Tudo que se produz precisa estar disponível, precisa virar conhecimento para outras pessoas também, usar o github é uma boa alternativa.
- Tudo que se produz vira portfólio, no final é isso que interessa a quem quer pagar você por um projeto.

COMPETIÇÕES

- O Kaggle é um dos principais sites de competição de DS do mundo, ser protagonista num ambiente desses abre muitas portas de trabalhos.
- Mas além da própria competição o Kaggle é uma oportunidade de aprender muito sobre DS, ver o que outros competidores estão fazendo, entender o que se espera de resultado em um projeto de DS.
- Além do Kaggle, é possível participar de Hackathons e também há muitos outros sites de competição ou com propostas parecidas, veja:



KAGGLE E OUTROS

- Kaggle (<https://www.kaggle.com/>)
- TopCoder (<https://www.topcoder.com/>)
- DrivenData (<https://www.drivendata.org/>)
- Codalab (<https://competitions.codalab.org/competitions/>)
- HackerRank (<https://www.hackerrank.com/>)
- Hash Code
(<https://codingcompetitions.withgoogle.com/hashcode>)
- Machine Hack (<https://machinehack.com/hackathons>)

ATUAÇÃO COMO FREELANCER

- Freelancer é aquele profissional que faz trabalhos pontuais, que podem durar um dia, uma semana ou até alguns meses, dependendo do projeto.
- Há muito trabalho de freelancer para quem estuda DS, inclusive com pagamento em US\$.
- Alguns sites (não exclusivos para DS):
 - Workana (<https://www.workana.com/pt/>)
 - Upwork (<https://www.upwork.com/>)
 - Freelancer (<https://freelancer.com.br/>)
 - Codementor (<https://www.codementor.io/>)
 - Toptal (<https://www.toptal.com/>)

FINALIZANDO

- Mercado para DS está com uma demanda muito grande.
- Saiba exatamente que profissional você é e como pode atuar.
- Crie conteúdo (portfólio) e autoridade (listas, grupos)
- Treine e estude muito
- Experimente competições
- Comece com pequenos trabalhos de freelancer
- Mergulhe no mundo dos devs
- Dica final:
 - <https://free-for.dev/#/>

INTRODUÇÃO À CIÊNCIA DE DADOS

**Mercado Profissional para Ciência
de Dados**