

Cientista de Dados - Por Onde Começar em 8 Passos

Ansioso por desbravar o universo da Ciência de Dados e não sabe por onde começar? Nós ajudaremos você. Preparamos um guia que vai ajudá-lo a compreender o que faz um **Cientista de Dados** e como iniciar sua preparação! Confira.

Vamos começar definindo o que é um **Cientista de Dados**:

Cientistas de Dados são uma nova geração de especialistas analíticos que têm as habilidades técnicas para resolver problemas complexos – e a curiosidade de explorar quais são os problemas que precisam ser resolvidos.

Eles também são um sinal dos tempos modernos. Cientistas de dados não estavam no radar há uma década, mas sua popularidade repentina reflete como as empresas agora pensam sobre **Big Data**. Essa incrível massa de informações não estruturadas já não pode mais ser ignorada e esquecida. É uma mina de ouro virtual que ajuda a aumentar receitas – contanto que haja alguém que escave e desenterre insights empresariais que ninguém havia pensado em procurar. Entra em cena o **Cientista de Dados**.

Para a comunidade em geral, um Cientista de Dados é um desses “Magos de Dados”, que pode adquirir massas de dados de diversas fontes e então limpar, tratar, organizar e preparar os dados; e, em seguida, explorar as suas habilidades em Matemática, Estatística e Machine Learning para descobrir insights ocultos de negócios e gerar inteligência.

Os dados utilizados por um Cientista de Dados podem ser tanto estruturados (bancos de dados transacionais de sistemas ERP ou CRM, por exemplo) e não estruturados (e-mails, imagens, vídeos ou dados de redes sociais). O Cientista de Dados cria algoritmos para extrair insights desses dados. Em seguida, cabe ao Cientista de Dados, apresentar estes dados, de forma que os tomadores de decisão possam utilizar o resultado da análise ao definir as estratégias empresariais ou mesmo para criar novos produtos ou serviços baseados em dados.

De acordo com Anjul Bhambhri, ex Vice Presidente de Big Data da IBM e atual Vice Presidente da Adobe, o Cientista de Dados é o profissional capaz de trazer a mudança para uma organização através da análise de diversas fontes de dados. Anjul Bhambhri escreve:

“Um Cientista de Dados representa uma evolução do papel de Analista de Negócios ou Analista de Dados. Estes profissionais possuem uma base sólida normalmente em ciência da computação, aplicações, modelagem, estatísticas, análises e matemática. O que define o Cientista de Dados é a forte visão de negócios, juntamente com a capacidade de comunicar os resultados, tanto para os líderes de negócios quanto para seus pares, de uma forma que influencie como uma organização posiciona-se diante dos desafios do mercado”.

Não existe uma formação que prepare Cientistas de Dados, pois esta é uma função relativamente nova. Tem havido muito debate sobre isso no ambiente acadêmico (principalmente nos EUA), pois o mercado precisa de profissionais agora e o tempo de preparação de um profissional com estas habilidades, pode levar algum tempo. E por isso formações técnicas em determinadas áreas, podem ajudar a preparar estes profissionais.

Com tantas informações sobre a profissão de Cientista de Dados e seu crescimento exponencial nos últimos anos, é fácil se perder diante de tantos artigos e materiais com fórmulas mágicas sobre qual caminho seguir. Vou fazer um alerta: não existe caminho fácil para se tornar um Cientista de Dados! É preciso estudar, aprender diferentes técnicas e ter conhecimento interdisciplinar. Por esse motivo, os Cientistas de Dados são bem remunerados e difíceis de encontrar no mercado.

Abaixo, os 8 passos que consideramos fundamentais para a preparação de um Cientista de Dados:

Passo 1: Faça uma auto avaliação

Este é o primeiro passo e acredite, é fundamental. Você, como profissional, precisa avaliar o momento atual da sua carreira e como pretende estar em 5 ou 10 anos. Se pretende seguir uma carreira em Analytics, seja como Cientista de Dados, Engenheiro de Dados ou Analista, precisa compreender quais são suas habilidades atuais, onde pretende chegar, avaliar os gaps e traçar um plano de ação.

Como não existe uma formação acadêmica específica para se tornar um Cientista de Dados, este profissional pode vir de áreas como Estatística ou Ciência da Computação, sendo comum encontrar profissionais de outras áreas atuando como **Cientistas de Dados** (Marketing, Economia, Ciências Sociais,

etc..). Mas independente da área de formação, algumas características serão comuns a todos os profissionais que trabalham com Ciência de Dados:

Programação – Conhecimento de programação é necessário. Linguagens de programação como R, Python, Julia, Scala, Java são parte do arsenal de ferramentas utilizadas em Data Science. Mesmo outros pacotes de análise de dados, como SAS, Matlab, Octave, SPSS e até o IBM Watson Analytics, requerem conhecimento em programação, para se extrair o melhor de cada ferramenta. É a habilidade de programação, que permite ao Cientista de Dados colocar em prática sua criatividade e extrair dos dados respostas para perguntas que ainda não foram feitas. Se você já tiver conhecimento em programação, isso será uma vantagem. Caso não tenha experiência em programação, mas tenha uma boa noção dos conceitos envolvidos em programação de computadores, isso vai ajudar muito. Avalie de forma clara seu nível de conhecimento em programação. Nós oferecemos na DSA um curso gratuito para quem pretende começar em programação. Acesse e se inscreva hoje mesmo: **[Python Fundamentos Para Análise de Dados](#)**.

Pensamento Lógico – Cientistas de Dados usam o pensamento lógico para fazer análises. Programação requer lógica. Se você já possui esta habilidade, isso vai acelerar seu aprendizado em Data Science.

Habilidade com Números – Matemática é a base da Ciência de Dados. Programação de computadores, envolve habilidade com números. Os algoritmos de Machine Learning, são baseados em conceitos matemáticos. A Estatística, parte fundamental da Ciência de Dados, requer habilidade com números. Avalie suas características e na sua auto avaliação, verifique se esse item será um problema ou não.

Conhecimento em Banco de Dados – Em diversas fases do processo de análise de dados, interações com bancos de dados serão necessárias. Bancos de dados relacionais, Data Warehouses, bancos de dados NoSQL, Hadoop, linguagem SQL. Todas estas tecnologias estão diretamente ligadas ao trabalho do Cientista de Dados e pelo menos sua compreensão será um ponto que poderá fazer diferença. Avalie se você compreende o conceito de banco de dados, entende as diferenças entre bancos de dados relacionais e NoSQL e como utilizar linguagem SQL para consultas.

A esta altura, talvez você já esteja se perguntando: como você pretende que eu aprenda tudo isso? Aqui entra um dos conceitos mal interpretados sobre a

profissão de Cientista de Dados. Acredita-se que este profissional precisa conhecer todas as ferramentas. Isso não é verdade e nem mesmo necessário. Escolha suas ferramentas e se especialize nelas. Por exemplo: conhecimento em linguagem R e Hadoop, permitirá fazer análises de grandes volumes de dados (Big Data). Você não precisa conhecer todas as linguagens de programação, bem como não tem que conhecer todos os bancos de dados. O mais importante é o pensamento lógico, esse sim indispensável (e esta habilidade talvez você já tenha). A tecnologia oferece ferramentas e nenhuma delas resolve sozinha 100% dos problemas, pois todas possuem suas limitações.



Passo 2: Prepare seu computador

A Ciência de Dados é computacionalmente intensa (isso não deve ser uma novidade para você!). Portanto, você precisa de um computador que permita processar seus scripts e aprender sobre análise de dados. Além disso, você vai precisar instalar ferramentas, interpretadores, pacotes office, etc... Para trabalhar com Ciência de Dados, um computador com 8GB de memória RAM, com um processador intel i5/i7/i9 ou equivalente é a nossa recomendação. Naturalmente, quanto maior a capacidade do seu computador, melhor! É possível também utilizar serviços como o **Cloud9** ou **Amazon AWS** e montar um ambiente virtual de trabalho.

(Windows, MacOS e Linux) vai atender as suas necessidades. De qualquer forma, você poderá instalar máquinas virtuais com outro sistema operacional. Boa parte do framework de Data Science e Big Data, foi construída sobre plataforma Unix. Para um servidor Hadoop ou Spark, um servidor Linux é a melhor recomendação. Já para a parte de apresentação de dados, Microsoft Office e outras ferramentas de visualização podem depender de um sistema Windows. Não há uma regra aqui, mas para usuários mais avançados, um sistema Unix é recomendado. Para aqueles que se sentem mais confortáveis com o Windows, não há problema algum. Utilize o Windows como seu sistema operacional base e, se necessário, crie uma máquina virtual com Linux, se quiser processar arquivos com Hadoop e/ou Spark ou realizar outros testes. Os principais fornecedores do Hadoop fornecem gratuitamente máquinas virtuais com Linux e Hadoop, prontas para uso em poucos cliques. Já o Microsoft Azure Machine Learning pode ser utilizado online, por exemplo. É possível também fazer o download o SAS University Edition, uma máquina virtual com Linux e SAS, que em poucos segundos permite você utilizar o SAS (uma das principais soluções de Analytics atualmente) para seu aprendizado e totalmente gratuito.

Softwares – Independente da linguagem de programação que você escolher, você vai precisar instalar o interpretador e uma IDE (ambiente integrado de desenvolvimento). Se a sua escolha for pela linguagem R, por exemplo, além de instalar o interpretador, você poderá instalar o RStudio (IDE para linguagem R). O mesmo vale para outras linguagens de programação. É possível criar seus scripts de Data Science 100% online, via browser, usando o Jupyter Notebook. Mas nem sempre você pode estar online e ter suas ferramentas instaladas localmente vai trazer uma série de vantagens. Além disso, considere instalar:

- Editores de texto: Sublime, Atom, Notepad++
- Software para Máquinas Virtuais: VirtualBox, Parallels
- Git e Github: para criar seu portfólio de projetos em Data Science
- Suite Office: Microsoft Office, Libre Office

Com exceção do Microsoft Office, todas as demais ferramentas são gratuitas.



Passo 3: Estatística e Matemática

Conhecimentos de Estatística e Matemática fazem parte do pacote essencial para quem pretende trabalhar como Cientista de Dados. Modelos estatísticos e algoritmos de Machine Learning, dependem de conhecimentos em regressão linear, regressão múltipla, clustering, Álgebra Linear, etc... Você precisa ser especialista em Estatística ou Matemática ou mesmo ter feito uma graduação nestas áreas? A resposta é não. Apesar dessas áreas permitirem uma compreensão mais abrangente, é possível aprender estes conceitos e aplicá-los, ao longo da sua jornada de aprendizagem em Data Science. Você não precisa aprender todos os tópicos relacionados à Estatística ou Matemática.

Existem muitas formas de aprender os conceitos de Estatística e Matemática aplicada e isso leva tempo. Para qualquer aspirante a Cientista de Dados a recomendação é aprender Estatísticas codificando, de preferência em Python ou R, de forma que você possa aplicar imediatamente um conceito aprendido. Nada substitui uma graduação em Estatística ou Matemática claro, mas você pode aprender os conceitos que serão usados no seu dia a dia em Data Science, aplicando estes conceitos através de uma linguagem de programação. Data Science é uma área multi-disciplinar.

No fim deste artigo, você encontra alguns recursos indicados por nosso time de



especialistas!

Passo 4: Big Data

Big Data é a matéria prima da Ciência de Dados. A profissão de Cientista de Dados, surgiu da necessidade de criar novos métodos de análise do imenso volume de dados que vem crescendo exponencialmente. Técnicas analíticas já existem há muitas décadas (talvez há séculos), mas nunca na história da humanidade, gerou-se tantos dados como atualmente. Novas formas de coleta, armazenamento e análise de dados são necessárias e o Big Data está revolucionando o mundo atual, pois com tantos dados a nossa disposição, podemos tomar decisões em tempo real e isso gera impacto direto na vida de todos nós. Caso não saiba o que é Big Data, confira nosso curso gratuito: **Big Data Fundamentos 2.0**.

O Cientista de Dados vai consumir Big Data, ou seja, vai utilizar o Big Data como matéria prima, aplicar diversas técnicas e colher insights. Mas a responsabilidade por coletar e armazenar os dados normalmente é do **Engenheiro de Dados**. Criação de clusters Hadoop, streaming de dados com Spark, integração entre diferentes fontes de dados são todas atribuições novas e normalmente exercidas por **Engenheiros de Dados**. Mas é importante que o Cientista de Dados conheça bem como funciona a infraestrutura que armazena os dados que serão analisados, pois isso pode fazer a diferença na hora de analisar 1 trilhão de registros, por exemplo.

Hadoop – O Hadoop está se tornando o coração da infraestrutura de Big Data, o que vai revolucionar o sistema tradicional de armazenamento em bancos de dados como conhecemos hoje. Além de gratuito, o Hadoop foi criado para ser usado em hardware de baixo custo, uma combinação essencial para empresas

que buscam reduzir seus custos de infraestrutura de TI e ainda capitalizar os benefícios do Big Data.

Spark – Spark é um projeto open source, mantido por uma comunidade de desenvolvedores que foi criado em 2009 na Universidade da Califórnia, Berkeley. O Spark foi concebido com o principal objetivo de ser veloz, tanto no processamento de queries quanto de algoritmos, além de processamento em memória e eficiente recuperação de falha. É atualmente um dos assuntos mais quentes em Data Science e vem ganhando muita popularidade.

Bancos de Dados NoSQL – Bancos de Dados tradicionais RDBMS (Relational Database Management Systems) são foram projetados para tratar grandes quantidades de dados (Big Data). Bancos de Dados tradicionais foram projetados somente para tratar conjuntos de dados que possam ser armazenados em linhas e colunas e portanto, possam ser consultados através do uso de queries utilizando linguagem SQL (Structured Query Language). Bancos de Dados relacionais não são capazes de tratar dados não-estruturados ou semi-estruturados. Ou seja, Bancos de Dados relacionais simplesmente não possuem funcionalidades necessárias para atender os requisitos do Big Data, dados gerados em grande volume, alta velocidade e alta variedade. Esta é a lacuna preenchida por Bancos de Dados NoSQL, como o **MongoDB** por exemplo. Bancos de Dados NoSQL, são bancos de dados distribuídos e não-relacionais, que foram projetados para atender os requerimentos deste novo mundo de dados em que vivemos.

Bancos de Dados Relacionais e Data Warehouses – Nas últimas décadas, todos os dados corporativos tem sido armazenados em bancos de dados relacionais e soluções de Business Intelligence usaram DataWarehouses para criar soluções analíticas. Estes dados estruturados, serão fonte de dados para Data Science e daí a importância do conhecimento em linguagem SQL, a linguagem padrão para consultar estes tipos de dados.

Como Cientista de Dados, você precisa ser especialista em todas as tecnologias? Não. Mas parte do trabalho do Cientista de Dados, será coletar dados do HDFS (Hadoop File system), criar RDD's no Spark, aplicar algoritmos de Machine Learning em streaming de dados, cruzar dados não estruturados coletados de redes sociais, com bancos de dados de CRM, etc...portanto, o Cientista de Dados precisa estar confortável com a forma como os dados estão armazenados e extrair da tecnologia o melhor que ela pode oferecer.



Passo 5: Linguagem de Programação e Machine Learning

Existem diversas ferramentas de análise e o número de soluções não para de crescer. Mas a recomendação para quem está iniciando, é obter o conhecimento básico, antes de tentar usar ferramentas de análise ou pacotes comerciais, de forma a conseguir extrair o melhor destas ferramentas. Algumas linguagens de programação se tornaram ícones em Ciência de Dados, como Python e R, por diversas razões: são gratuitas, contam com uma comunidade ativa e crescente, já atravessaram o período de maturação, são amplamente utilizadas, tanto no meio acadêmico quanto no meio empresarial e se especializaram em Data Science.

Python – É uma linguagem de uso geral, que tem recebido nos últimos anos mais e mais módulos e pacotes para Data Science como Pandas, Matplotlib, Scikit-Learn, Stats Models e TensorFlow. Python é mais fácil de aprender em comparação a outras linguagens, tem uma comunidade ativa, muita documentação disponível (inclusive em português) e pode ser usada para outras atividades além de Data Science.

Linguagem R – Linguagem estatística, que existe há mais de 30 anos. Sua capacidade de processar estatísticas de grandes volumes de dados e criar gráficos sofisticados, fizeram com que gigantes do mercado de tecnologia, como Oracle e Microsoft, adotassem R como linguagem padrão para análises estatísticas. Um dos problemas mais comuns que as pessoas enfrentam em aprender R é a falta de um guia. As pessoas não sabem, por onde começar,

como proceder e nem que caminho seguir. Há uma sobrecarga de bons recursos gratuitos disponíveis na Internet e isso torna o caminho de aprendizado muito mais tortuoso.

Por exemplo: a plataforma de Aprendizado de Máquina da Microsoft (Microsoft Azure Machine Learning), possui uma série de módulos Python e R, prontos para uso. O conhecimento de pelo menos uma destas linguagens é fundamental.

Outras linguagens como Julia, Scala e Java também são muito utilizadas em Data Science, mas se estiver começando, opte pelas linguagens R ou Python. Estas linguagens vão permitir uma base sólida, fazendo com o que o profissional avance para soluções comerciais como SAS, Microsoft Azure Machine Learning, Oracle Advanced Analytics, Microstrategy, SAP Predictive Analytics, Tibco Analytics, entre outros.

Se você já possui conhecimento em Matlab, Octave, Stata ou Minitab, saiba que seu conhecimento já pode ser utilizado em Data Science.

Uma dica importante: não tente aprender tudo! Selecione 2 ou 3 ferramentas e se dedique ao aprendizado delas de forma detalhada.

E claro, não dá para falar em Data Science, sem falar em Machine Learning.

Machine Learning (ou Aprendizado de Máquina) é uma das tecnologias atuais mais fascinantes. Você provavelmente usa algoritmos de aprendizado várias vezes por dia sem saber. Sempre que você usa um site de busca como “**Google**” ou “**Bing**”, uma das razões para funcionarem tão bem é um algoritmo de aprendizado. Um algoritmo implementado pelo “**Google**” aprendeu a classificar páginas web. Toda vez que você usa o aplicativo para “marcar” pessoas nas fotos do “**Facebook**” e ele reconhece as fotos de seus amigos, isto também é Machine Learning. Toda vez que o filtro de spam do seu email filtra toneladas de mensagens indesejadas, isto também é um algoritmo de aprendizado.

Algumas razões para o crescimento Machine Learning são o crescimento da web e da automação. Isso significa que temos conjuntos de dados maiores do que nunca. Por exemplo, muitas empresas estão coletando dados de clicks na web, também chamados dados de “clickstream”, e estão criando algoritmos para minerar esses dados e gerar sistemas de recomendação, que “aprendem” sobre os usuários e oferecem produtos que muito provavelmente eles estão buscando. O Netflix é um dos exemplos mais bem sucedidos de aplicação de

Machine Learning. Cada vez que você assiste um filme ou faz uma avaliação, o sistema “aprende” seu gosto e passa a oferecer filmes de forma personalizada para cada usuário.

Existem diversos algoritmos de aprendizagem de máquina, dependendo se a aprendizagem é supervisionada ou não supervisionada, tais como: Linear Regression, Ordinary Least Squares Regression (OLSR), Logistic Regression, Classification and Regression Tree (CART), Naive Bayes, Gaussian Naive Bayes, k-Nearest Neighbour (kNN), k-Means, Bootstrapped Aggregation (Bagging), Natural Language Processing (NLP), Principal Component Analysis (PCA), Principal Component Regression (PCR), BackPropagation e muito mais.

Cada algoritmo será ideal para determinado tipo de dado e de acordo com a análise pretendida. Não há necessidade de aprender todos os algoritmos. Mas é importante compreender os conceitos e como implementá-los. Daí a importância da Matemática e da Estatística. Confira as dicas no fim do artigo.

Passo 6: Conhecimento de Negócios

Qual o objetivo da sua análise? Para que você vai coletar montanhas de dados e aplicar modelos de análise? Que problema você pretende resolver, analisando dados? O principal objetivo da Ciência de Dados é resolver problemas. As empresas não vão iniciar um projeto de Data Science, se isso não for relevante para o negócio. Portanto, o Cientista de Dados deve estar familiarizado com a área de negócio para a qual ele está iniciando um projeto, utilizando Data Science.

Normalmente o mercado interpreta de forma equivocada este requerimento para a profissão de Cientista de Dados, fazendo crer que o profissional precisa ser expert em determinado segmento de negócio. Mas aqui vale fazer algumas considerações importantes. Primeiro, os profissionais de Business Intelligence sempre tiveram este requerimento, conhecer bem uma área de negócio, a fim de coletar os KPI's (indicadores) e com isso prover soluções de BI que atendessem as necessidades do cliente. Isso não mudou, o que mudou foi a forma como a análise é feita, uma vez que o Big Data entrou na equação. Em segundo, dada a amplitude de projetos de Data Science e Big Data, dificilmente haverá um único profissional atuando e sim uma equipe de Data Science, normalmente liderada pelo Cientista de Dados. E no Data Science Team, diferentes perfis irão atuar, como por exemplo especialistas em segmentos de negócio.

A dica aqui é simples. Procure compreender a área de negócio na qual você pretende atuar como Cientista de Dados. Se vai trabalhar em uma mineradora por exemplo, quais são os principais indicadores? De onde vem os dados? Que problemas a empresa precisa resolver? Que tipos de dados devem ser analisados e correlacionados? Como técnicas de Machine Learning podem ser empregadas para melhorar o faturamento da empresa? Como a análise de dados permite oferecer um serviço melhor aos clientes? Cada área de negócio tem as suas particularidades e uma compreensão ampla disso, vai permitir um trabalho que realmente gere valor.

Passo 7: Técnicas de Apresentação e Visualização de Dados

Com a massiva quantidade de dados aumentando a cada dia, um grande desafio vem surgindo para aqueles responsáveis por analisar, sumarizar e apresentar os dados: fazer com que a informação gerada, possa ser facilmente compreendida.

É uma das tarefas mais importantes do trabalho do **Cientista de Dados**, é ser capaz de transmitir tudo aquilo que os dados querem dizer. E às vezes os dados querem dizer coisas diferentes, para públicos diferentes. Pode parecer fácil em princípio. Hoje temos à nossa disposição os mais variados recursos para apresentação e exatamente aí que está o desafio. Nunca foi tão fácil gerar tabelas e gráficos, com diferentes estruturas, formatos, tamanhos, cores e fontes. Os gráficos estão deixando de ser gráficos e se tornando infográficos. Ter um volume cada vez maior de dados à nossa disposição, não torna mais fácil a apresentação da informação gerada. Pelo contrário, torna a tarefa mais complicada. Quase uma arte.

Uma das tarefas do Cientista de Dados, é apresentar seus resultados. Ninguém melhor que o profissional que faz a análise, desde a coleta, limpeza e armazenamento dos dados, até a aplicação de modelos estatísticos, para explicar seus resultados. Uma visualização efetiva de dados, pode ser a diferença entre sucesso e falha nas decisões de negócio. Particularmente, eu acredito que em breve, a capacidade de comunicar e contar as histórias dos dados, será uma das características mais valorizadas e buscadas pelas empresas. Técnicas de apresentação é um dos skills que fará a diferença na hora de contratar um Cientista de Dados, à medida que o conhecimento técnico estiver difundido.

O Cientista de Dados deve ser um contador de histórias e deve ser capaz de contar a mesma história de maneiras diferentes. O profissional que for capaz de unir as habilidades técnicas necessárias para análise de dados, com a capacidade de contar histórias, será um profissional único. Este é o verdadeiro conceito de unicórnio, atribuído aos Cientistas de Dados.

Diversas ferramentas possuem funcionalidades avançadas para visualização de dados: Tableau, QlikView, Microsoft Excel, Microsoft Power BI, Microstrategy, Weka, NetworkX, Gephi, bibliotecas Java Script (D3.js, Chart.js, Dygraphs), além de visualizações alto nível que podem ser feitas em Python ou R.

É importante não apenas estar familiarizado com uma ou mais ferramentas para visualizar dados, mas também os princípios por trás da codificação visual de dados e comunicação de informações.



Passo 8: Pratique!!

Não há outra forma de aprender qualquer que seja o assunto. É preciso praticar, testar, experimentar, cometer erros, aprender com eles, testar novamente, interagir com a comunidade.



Ufa. Se você chegou até aqui, parabéns! Agora você compreende melhor porque o Cientista de Dados é uma profissão em ascensão e porque sua remuneração está entre as maiores em qualquer pesquisa que se faça? Mas ainda não terminamos, continue sua leitura!

Erros que devem ser evitados ao longo da caminhada:

1- Achar que o aprendizado é fácil e rápido – Em nenhuma profissão, o aprendizado é rápido e fácil. Adquirir conhecimento e experiência requer tempo, esforço, investimento e bastante dedicação. Não caia nesta armadilha de achar que é rápido e fácil. Comece sua preparação hoje mesmo, mas esteja ciente que adquirir conhecimento leva tempo.

2- Aprender muitos conceitos ao mesmo tempo – Este é o erro mais comum. Por se tratar de uma área interdisciplinar, há normalmente a tendência em querer aprender muita coisa ao mesmo tempo. Não foque em quantidade e sim qualidade. Aprenda um conceito, consolide, pratique e só então avance para outra área de estudo. O começo será mais difícil, mas à medida que aprende e adquire experiência, o aprendizado de novas ferramentas fica mais fácil.

3- Começar por problemas muito complexos – A solução de problemas mais complexos em Data Science, requer tempo e experiência. Não tente fazer isso no começo da sua jornada.

4- Focar apenas na programação – Data Science não é apenas programação. Outros conceitos são tão importantes quanto. Estatística é importante. Visualização de dados e apresentação são importantes. Tenha seu foco em outras áreas e não apenas programação.

E como estudar?

Ouvimos esta pergunta com frequência aqui na Data Science Academy. Antes de investir em formação analítica, tome essas medidas para ter certeza de obter o valor real do seu investimento.

1. Defina suas metas. Como diz o ditado: “Quem não sabe para onde vai, qualquer caminho serve!”. Você fez a auto avaliação sugerida lá no início do texto? Qual foi o resultado? Quais são as áreas que você precisa se dedicar? Como será sua disponibilidade? Qual seu objetivo? Sei que são muitas perguntas, mas elas precisam ser respondidas, antes de definir sua trilha de aprendizagem.

2. Participe de comunidades. Cientistas de Dados costumam se encontrar em Meetups e em alguns blogs (links abaixo). Estas comunidades vão permitir a troca de experiências e isso é muito valioso. Mas cuidado: cada pessoa tem sua opinião, que muitas vezes são baseadas em gostos pessoais e não avaliações técnicas. Ouça o que todos tem a dizer, filtre o que é relevante, experimente por si mesmo e então monte suas conclusões. Temos ainda a Comunidade DSA, o ponto de encontro dos alunos de todos os cursos pagos em nosso portal.

3. Experimente. Enquanto você não executar um algoritmo de aprendizado de máquina em um dataset com milhões de registros, não criar um algoritmo de limpeza e transformação de dados, não coletar streaming de dados de redes sociais, você não vai compreender como as coisas funcionam. Experimente! Aprenda, faça, erre, faça novamente e quando você menos esperar, você vai ser capaz de analisar dados e contribuir para a empresa onde trabalha ou para seu próprio negócio.

Use as informações que você recolheu para selecionar opções de treinamento que ajudam você a alcançar seus objetivos, sem perder tempo e dinheiro.

Conhecimento em Ciência de Dados requer tempo e dedicação. O treinamento que você escolher deve ser um mix de fundamentação teórica, com prática e experimentação.

Recursos

Existem muitos recursos disponíveis para aprender Data Science e tantos recursos acabam gerando sobrecarga nos iniciantes, que podem perder o foco. Separamos aqui uma lista com as que consideramos as melhores fontes de aprendizagem em Data Science.