

Universidade do Estado do Amazonas**Data:** 1 de outubro de 2020**Disciplina:** Fundamentos da Ciência de Dados**Professor(a):** Carlos Maurício Serodio Figueiredo**Alunos:** Lucas Pereira Reis**ATIVIDADE AVALIATIVA #5**

As empresas estão cada vez mais se preocupando em modernizar suas soluções e trazer novas implementações para sua realidade, e uma delas é a utilização de *data science* em seus produtos, focando em análise, engenharia de dados, inteligência artificial, etc. Assim, o mercado da área de *Data Science* está extremamente aquecido, tendo uma demanda muito alta. Para se inserir neste mercado, é necessário entender as *skills* necessárias para esta profissão.

Para se inserir neste mercado, é necessário entender como ele funciona no momento. Atualmente temos diversas funções presentes na área de *Data Science*, como *Data Analyst*, *Data Engineer*, *Machine Learning Engineer* e o próprio *Data Scientist*. Cada uma compartilha ferramentas e conhecimentos e estão interligadas dentro do processo de desenvolvimento de projeto de *Data Science*, porém devido à suas particularidades, é importante entendê-las para de fato focar o conhecimento em um e então entrar no mercado de trabalho.

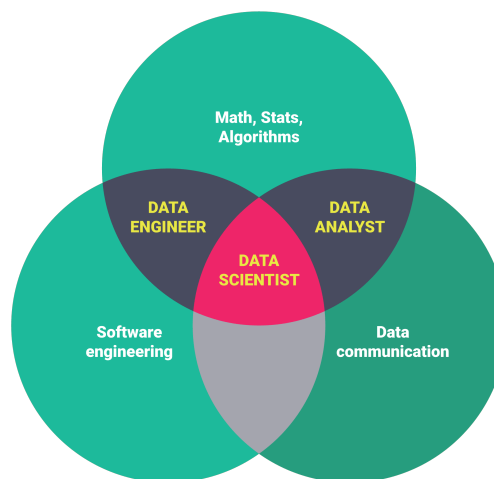


Figura 1: Relação entre as funções de trabalho.

A Figura 1 ilustra a relação entre três funções citadas anteriormente, onde o *Data Analyst* é responsável pela exibição e demonstração dos dados, além de realizar análises estatísticas. O *Data Engineer* cumpre o dever de preparar a estrutura de dados para serem consumidas pelo restante da equipe. O *Data Scientist* tem a possibilidade de atuar na interseção entre as funções, podendo atuar em diferentes etapas, normalmente acaba sendo responsável por entender e interpretar as regras de negócio do produto. Por fim, embora não esteja presente, o *Machine Learning Engineer* entra em uma posição semelhante ao *Data Engineer* focando-se em algoritmos de *Machine Learning* e *Deep Learning*.

Quando esta área começou a se popularizar, inicialmente as ofertas de emprego não estavam muito divididas conforme foi mostrado, era comum encontrarmos vagas onde solicitavam que o candidato dominasse todo o conjunto de ferramentas disponíveis. Aos poucos o mercado foi moldando e evoluindo até os dias de hoje onde as empresas entendem que precisamos de um time completo de *Data Science* para seu processo, mas dependendo do tamanho da empresa, ainda é comum termos *Data Scientists* que desempenham atividades de todas as etapas.

Como dito anteriormente, embora estejam interligadas, cada função possui seus pré-requisitos, a Figura 2 apresenta as ferramentas mais solicitadas nas vagas de *Data Analyst*. Nesta função é necessário

apresentar os dados para outras pessoas da empresa, assim é comum solicitarem conhecimentos em *Tableau*, *PowerPoint* e *PowerBI*, suas principais linguagens de programação são *Python* e *R* e conhecimentos em *SQL* são importantes pois normalmente os dados estão presentes em banco de dados, porém ainda é comum empresas terem seus dados catalogados em planilhas em *Excel*, explicando assim à alta demanda desta ferramenta.

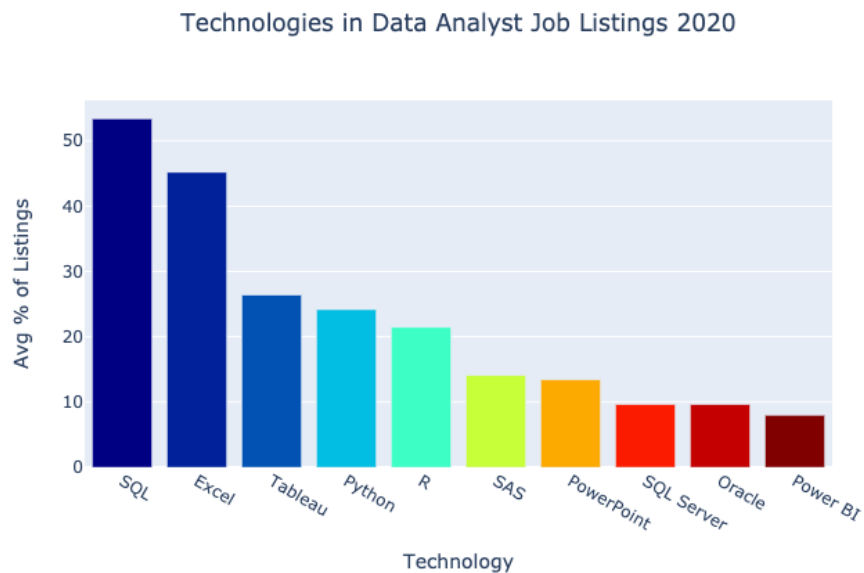


Figura 2: As tecnologias mais requisitadas para *Data Analyst* em 2020.

A Figura 3 apresenta as ferramentas solicitadas para *Data Engineer*. Como dito antes, esta função realiza o processamento de dados e o estrutura para então o restante da equipe poder consumi-los com mais facilidade, assim ferramentas para processamento de *Big Data* como *Spark*, *Hadoop* e *Hive* estão presentes. Novamente temos o *Python* como principal linguagem e *SQL* para realizar consultas.

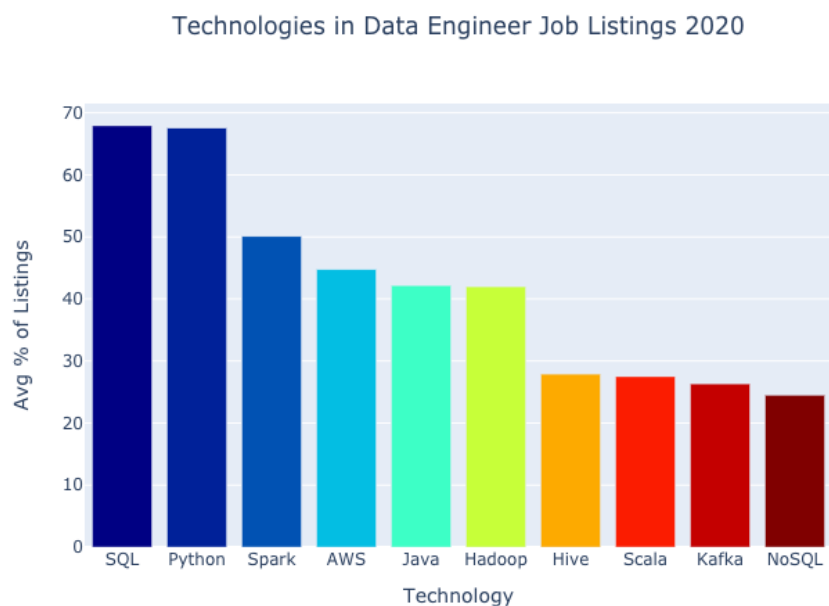


Figura 3: As tecnologias mais requisitadas para *Data Engineer* em 2020.

Para o *Machine Learning Engineer* temos novamente o foco em *Python* mas agora também vemos ferramentas como *Tensorflow*, *Keras* e *Scikit-learn* pois sua função é responsável por desenvolver os modelos de *machine learning* para a aplicação.

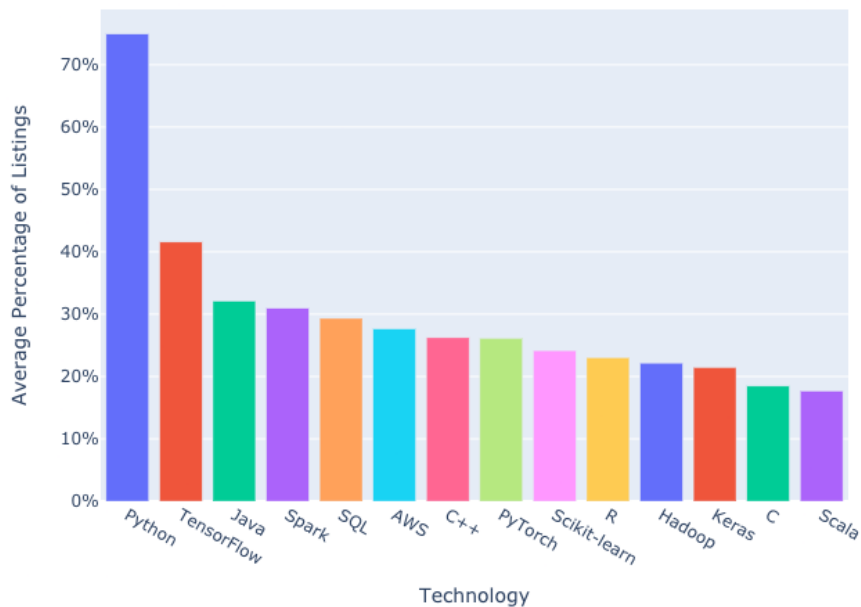


Figura 4: As tecnologias mais requisitadas para *Machine Learning Engineer* em 2020.

Por fim, temos o *Data Scientist* onde podemos claramente observar na Figura 5 que suas tecnologias são a junção das ferramentas apresentadas nas outras funções, justamente por este profissional ter a flexibilidade atuar em diferentes etapas. A linguagem *Python* continua extremamente requisitada para esta função, junto com o *R* e *SQL*.

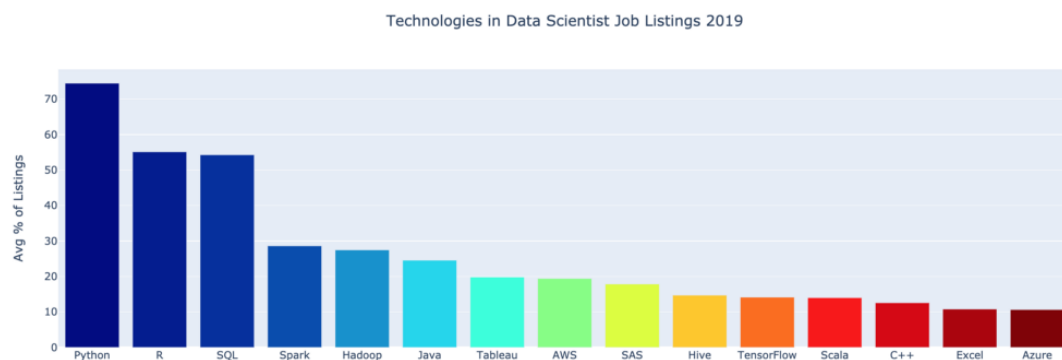


Figura 5: As tecnologias mais requisitadas para *Data Scientist* em 2019.

Podemos então concluir que a linguagem *Python* é a principal *skill* necessária para qualquer vaga na área de *data science*, existem empresas que utilizam também *R* e *Java* mas cada vez mais o *Python* está dominando este mercado, sendo requisito obrigatório para o mercado de trabalho. Em seguida observamos que é necessário ter conhecimento sobre banco de dados e como realizar consultas *SQL* para obter as informações armazenadas, após isso irá depender de qual função a pessoa deseja seguir.

Acredito que uma pessoa que está entrando no mercado de trabalho e quer ter um diferencial em seu currículo, é necessário estar sempre atento às novas tecnologias presentes e não se limitar a determinados tipos de ferramentas, estar sempre de mente aberta para aprender novas técnicas dentro do processo de desenvolvimento de *data science*.

A primeira etapa necessária é obviamente aprender a linguagem *Python* e como utilizá-la nesta

área, após isso é necessária é conhecer o fluxo completo de um projeto de *Data Science* e aos poucos pode ser adicionado novas ferramentas, onde você irá conhecer e entender as tecnologias presentes no mercado e implementá-las, de forma simples, em seus estudos.

Tendo o conhecimento em processamento de dados, análise de dados e criação de modelos, você poderá ser apresentando como um profissional que possa desempenhar diferentes papéis, além disso você também terá a possibilidade de testar diferentes novas soluções que estão sendo apresentadas constantemente em artigos, sendo um diferencial na sua carreira.