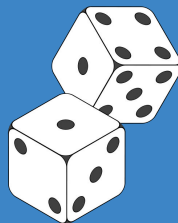


O cenário da Ciência de Dados no Brasil na atualidade

por Renata Ribeiro








25 de outubro de 2021

Sobre mim

- Bacharela em C&T e graduanda em Ciência da Computação (UFBA)
- Pós-graduanda no MBA Data Science and Analytics (USP)
- Estagiária em Análise de Dados na RD Station
- Seis anos em TI e desde 2019 estudando com foco em dados
- Criadora de Conteúdo no Instagram **@renadeveloper**



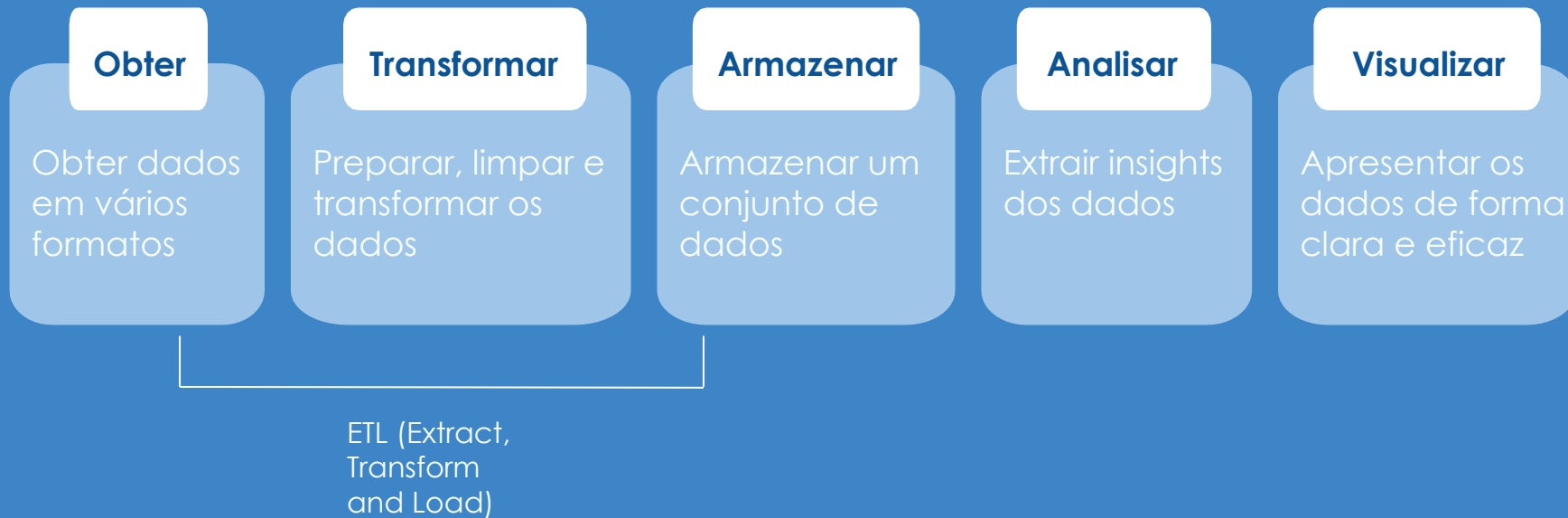
O Brasil na Internet em 2020*

-  47% das pessoas entrevistadas indicam **comprar mais pela internet** após Covid-19
-  Foram realizadas cerca de 14,2 milhões de **compras online** de livros, volume 44% maior do que em 2019
-  O Brasil possui cerca de 1,3 milhão de **lojas online**, crescimento de 40,7% em um ano
-  Houve alta de 110,52% no **comércio online**: o dobro, comparando junho de 2019 e 2020
-  Brasil é o **quarto país** do mundo com **mais usuários diários de internet**

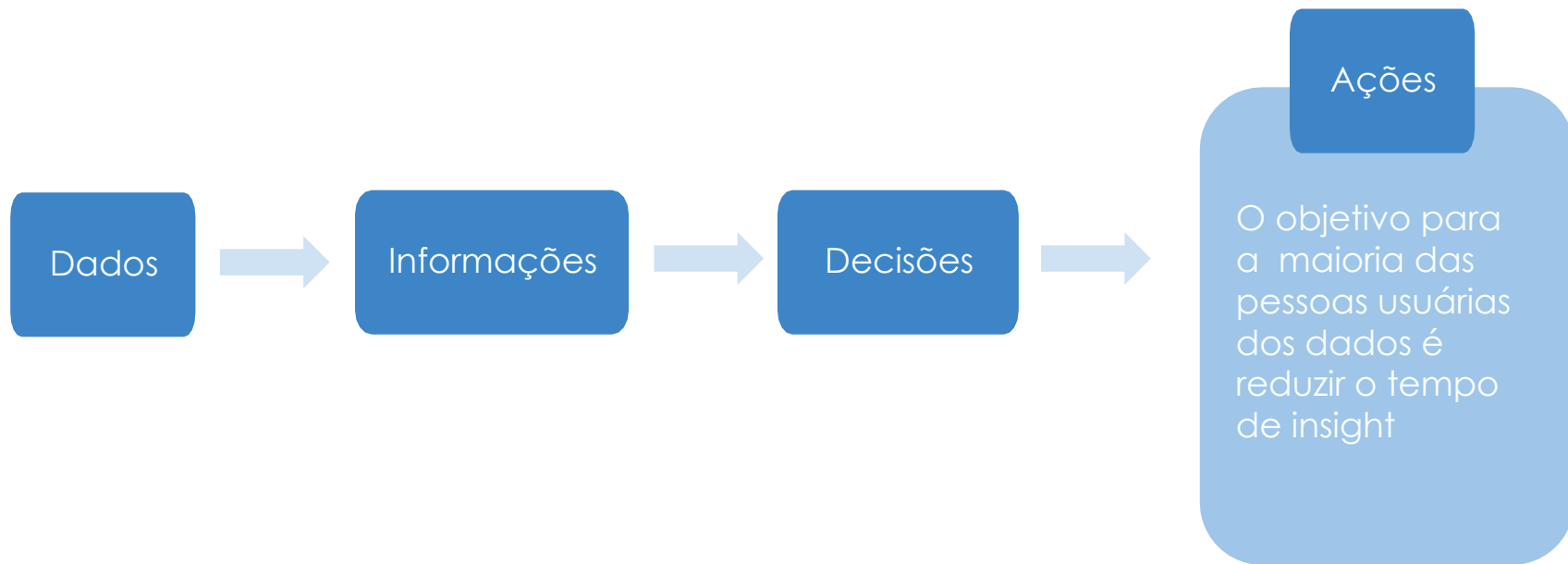
Dado *versus* Informação

Dados	Informação
<p>Dados são uma coleção de fatos brutos e desestruturados.</p> <p>Dados são baseados em observações, documentos, registros...</p>	<p>Informação são os dados após seu processamento, onde podem ser vistos de forma organizada e estruturada.</p> <p>Além disso, são acompanhadas de um contexto significativo que auxilia sua utilização para a tomada de decisão.</p>
<p>Ex: CADA registro de venda de um e-commerce é uma unidade de dado</p>	<p>Ex: Total de vendas de cada categoria de produtos</p>

Workflow de Dados



Percurso de Análise de Dados



Por que analisar Dados?



Aumentar a receita



Melhorar a experiência do usuário



Diminuir custos



Provar Retorno sobre Investimento (ROI)



Desenvolver produtos



Customizar entrega de conteúdo e venda de produto



Corrigir desvios



Predizer metas, comportamentos, etc.

Principais atores de Dados



Engenheiro de Dados

Constrói e otimiza a infraestrutura de dados



Cientista de Dados

Responde perguntas de negócio a partir dos dados



Analista de Dados

Consolida e elabora relatórios a partir dos dados

Análise do atual Cenário Brasileiro de Dados

Algoritmo de Web
Scraping e posterior
Análise de Dados*

+

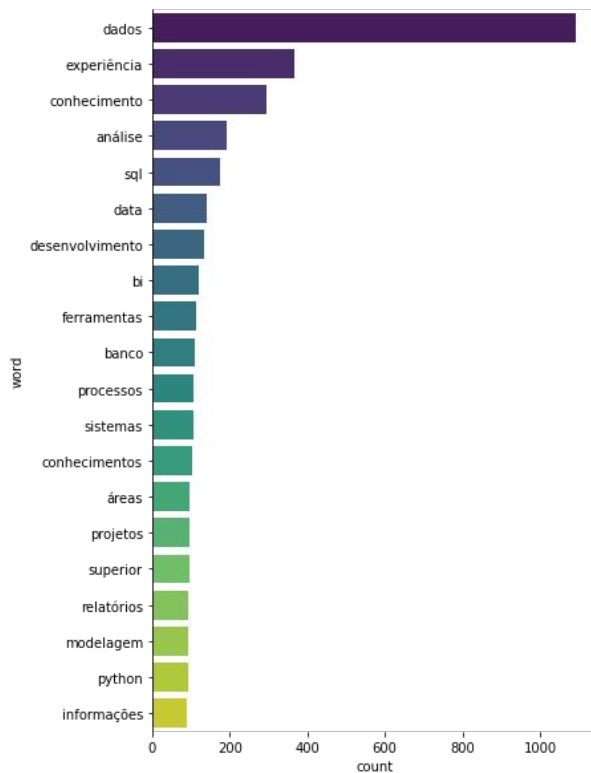
Percepções
enquanto
profissional atuante
em dados

*Disponível no repositório Job Positions in a Data Field com autoria de Daiane Klein (<https://github.com/daianeklein/jobs-stacks>)

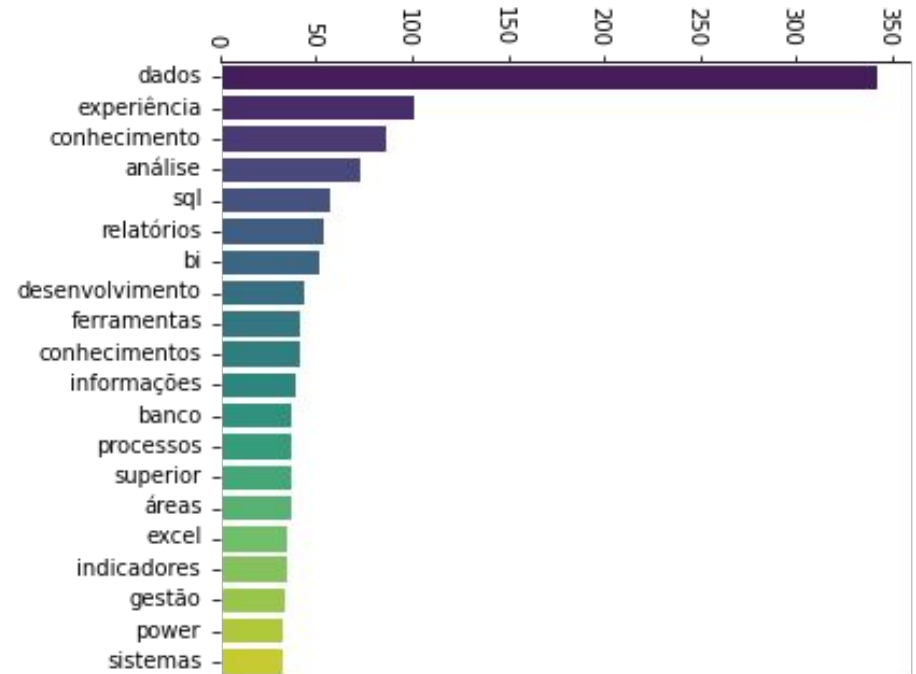
Analizando o Dataset

- Foram utilizadas as primeiras 50 páginas com vagas em PT-BR adquiridas com algoritmo de coleta de dados em um site de vagas de emprego de empresa privada
- Das 244 vagas encontradas, 83 são voltadas para Análise de Dados, 59 para Engenharia de Dados e 25 para Ciência de Dados
- Somente 100 vagas possuem descrição de senioridade sendo 35 para Sênior, 25 para Pleno, 14 para Júnior e 9 para Estágio ainda possuindo vagas para Especialista, Consultor e Assistente
- Serão mostradas aqui as palavras mais comuns na descrição de todas as vagas no geral e das vagas para Análise de Dados, Engenharia de Dados e Ciência de Dados
- Serão mostradas aqui as palavras mais comuns quanto as senioridade Estágio, Júnior, Pleno e Sênior (para vagas de dados em geral)

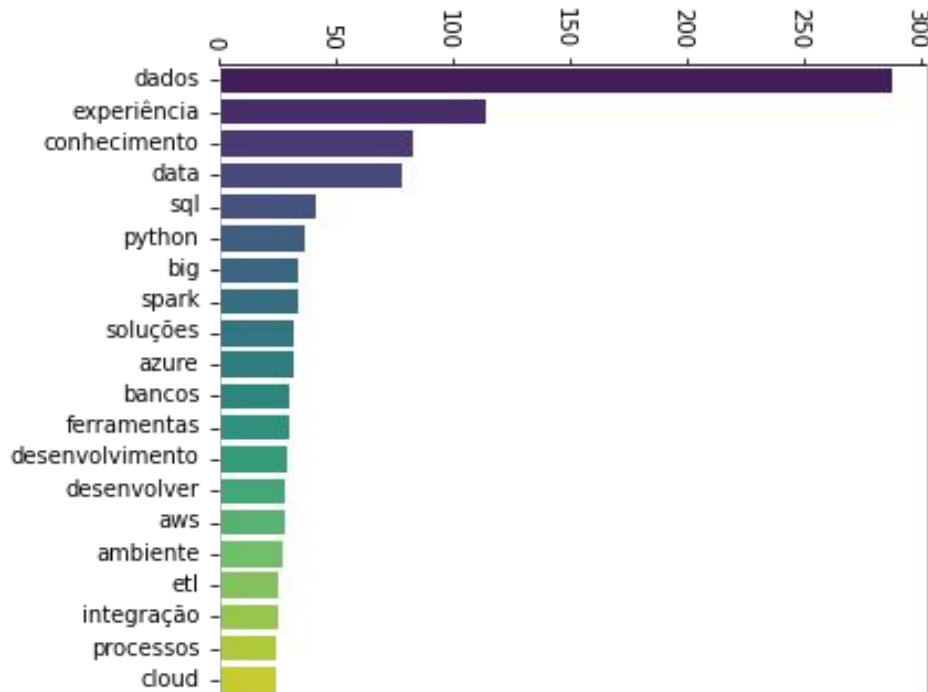
Vagas - Contexto Geral



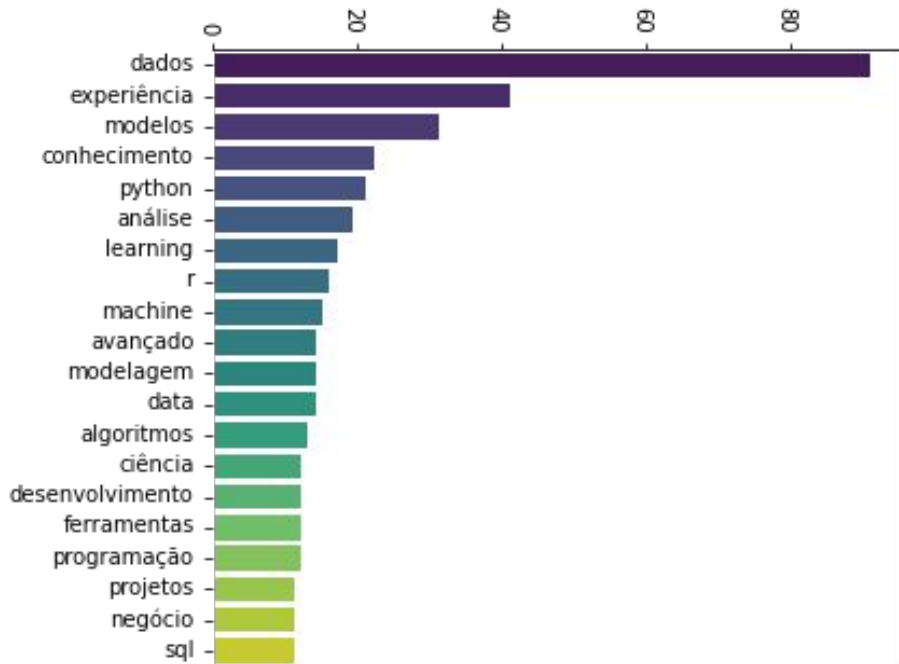
Vagas - Análise de Dados



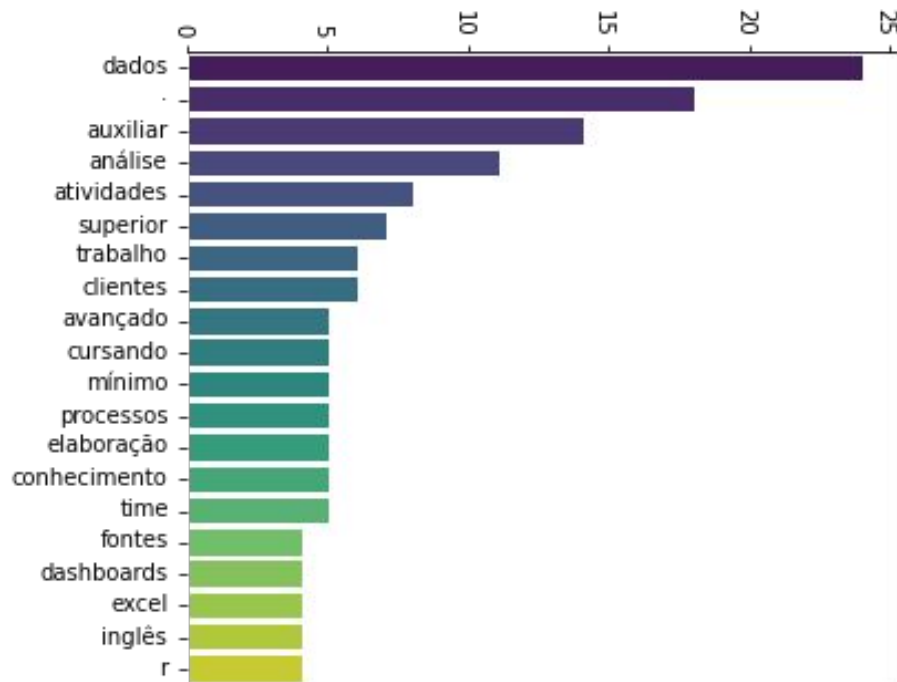
Vagas - Engenharia de Dados



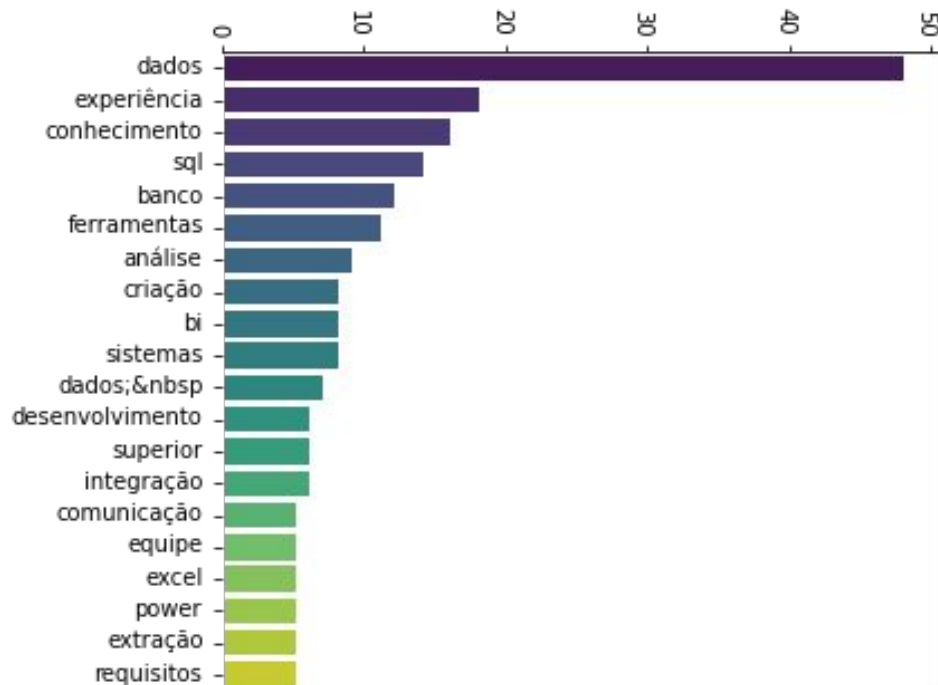
Vagas - Ciência de Dados



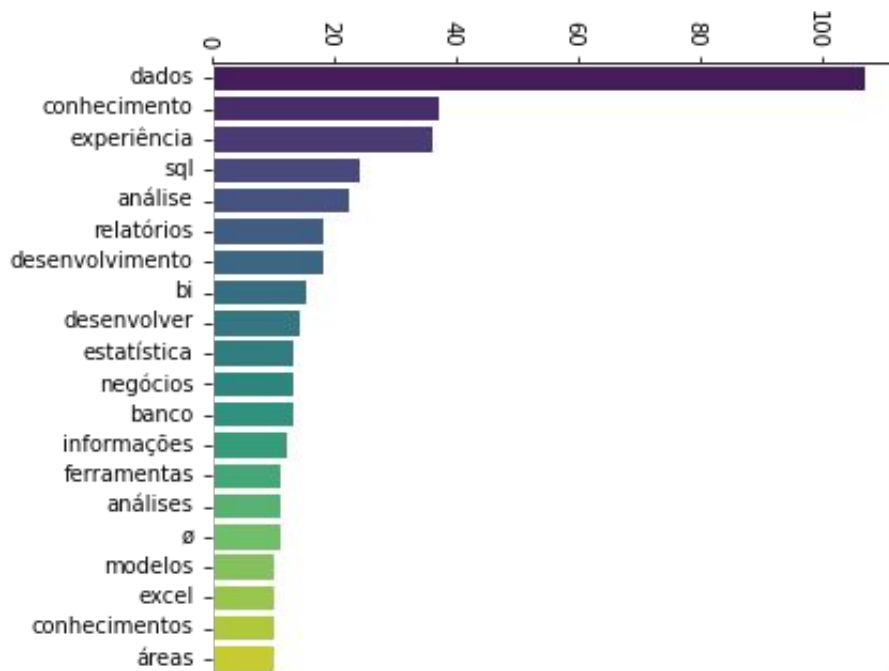
Senioridade - Estágio



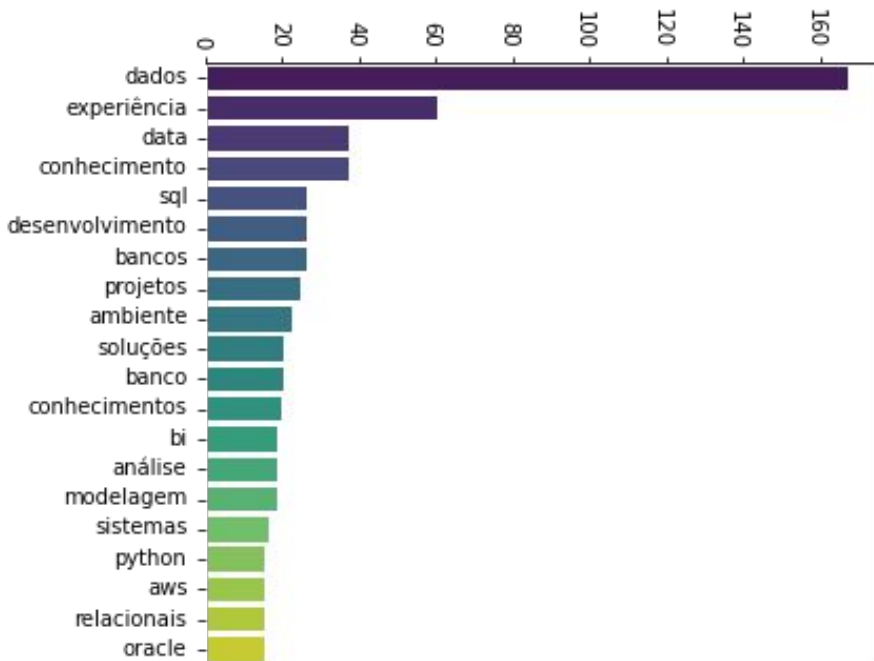
Senioridade - Júnior



Senioridade - Pleno



Senioridade - Sênior



Percepções e Pesquisas Futuras

- Apesar das definições diferentes para os três principais atores de Dados, na prática as habilidades necessárias são muito parecidas
- Experiência e vasto conhecimento contam muito na área! Um movimento de pessoas de Dados Estagiárias e Júniores começa a se formar mas ainda é muito embrionário
- Conhecimento básico em SQL e funcionamento de banco de dados (CRUD) é indispensável até para a senioridade estágio
- Vagas de senioridade pleno e sênior pedem uma visão de negócio maior do que as vagas de Júnior, que focam em desenvolvimento
- Inserir como feature no dataset a possibilidade de trabalho remoto e/ou presencial e modificar o algoritmo de coleta de dados para possibilitar seu uso no LinkedIn

MUITO OBRIGADA!!
Perguntas?!

