

Relatório EDA sobre salários para ciência de dados

Este projeto tem como foco a análise dos salários de cientistas de dados utilizando a linguagem de programação R e uma base de dados obtida na plataforma “Kaggle”. O objetivo é obter insights sobre as tendências salariais e os fatores que influenciam a remuneração na área de ciência de dados. Para a análise, utiliza-se um conjunto de dados contendo informações relevantes, como cargos, níveis de experiência, moedas correntes e respectivos salários.

A análise faz uso de diversas técnicas e funções disponíveis no R para explorar o conjunto de dados. Estatísticas descritivas, visualização de dados e modelagem estatística são empregadas para identificar padrões e relacionamentos nos dados. Os pacotes R utilizados incluem dplyr, ggplot2.

As principais descobertas da análise revelam os fatores que impactam os salários na área de ciência de dados. Isso inclui a influência dos níveis de experiência, cargos, formação educacional e possivelmente outras variáveis. Os resultados fornecem insights valiosos para profissionais que atuam na área de ciência de dados, empregadores em busca de referências salariais e pesquisadores interessados em compreender a dinâmica salarial nesse campo.

O projeto reconhece as limitações do conjunto de dados e sugere possíveis direções futuras. Isso pode envolver a incorporação de conjuntos de dados adicionais, consideração de fatores externos e realização de análises mais avançadas para aprimorar a compreensão das tendências salariais na indústria de ciência de dados.

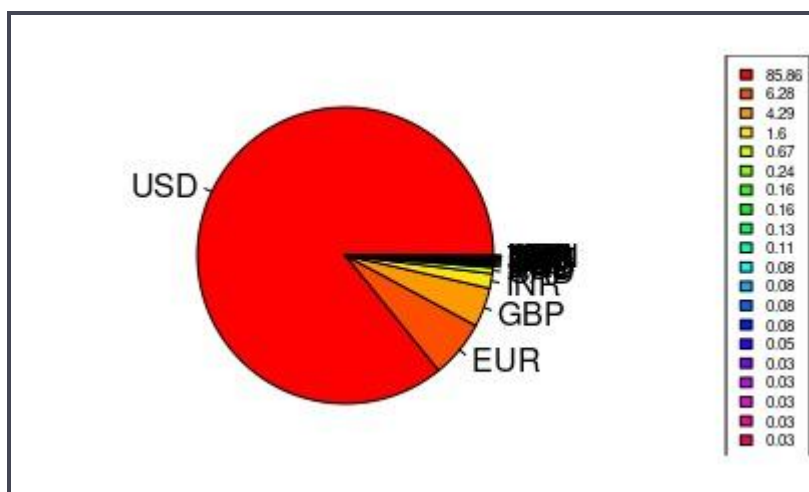
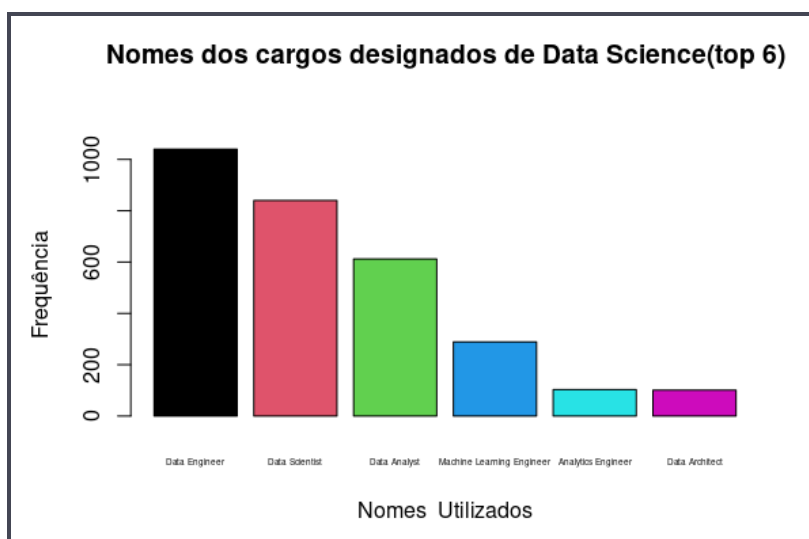
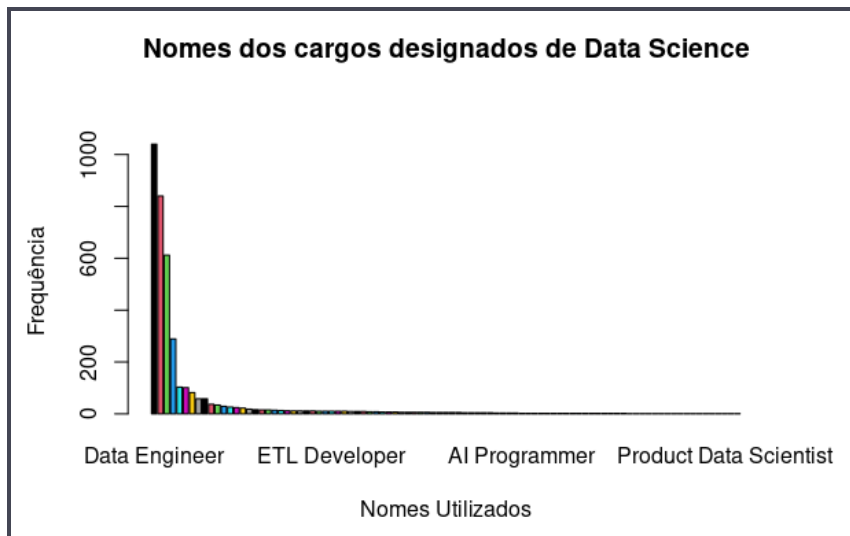
Ao conduzir essa análise em R, você adquire um entendimento mais profundo do panorama salarial na área de ciência de dados e contribui para a base de conhecimento nesse campo. Os insights obtidos a partir desse projeto podem ser utilizados para tomada de decisões, benchmarking e pesquisas futuras relacionadas aos salários na área de ciência de dados.

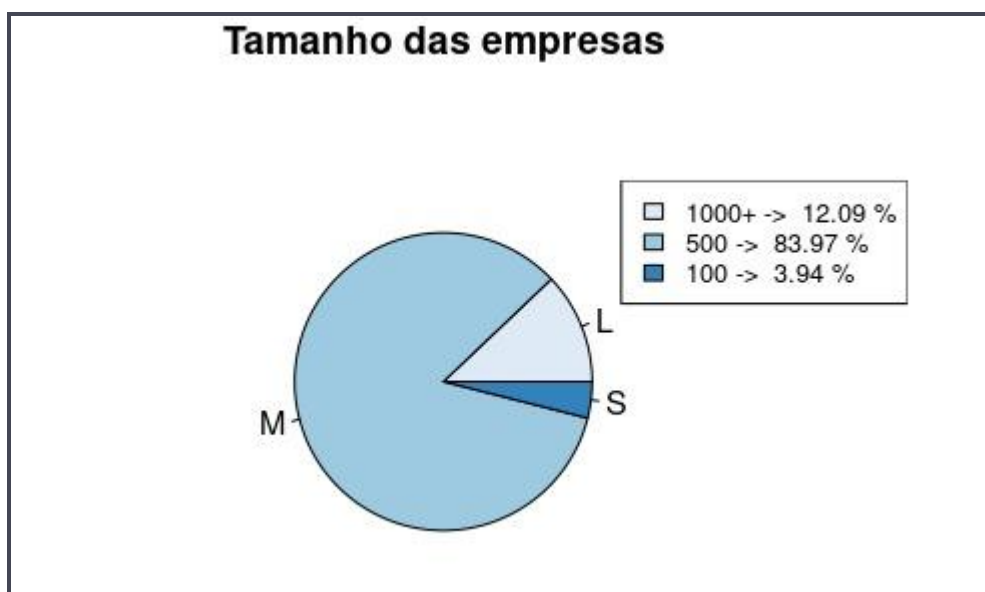
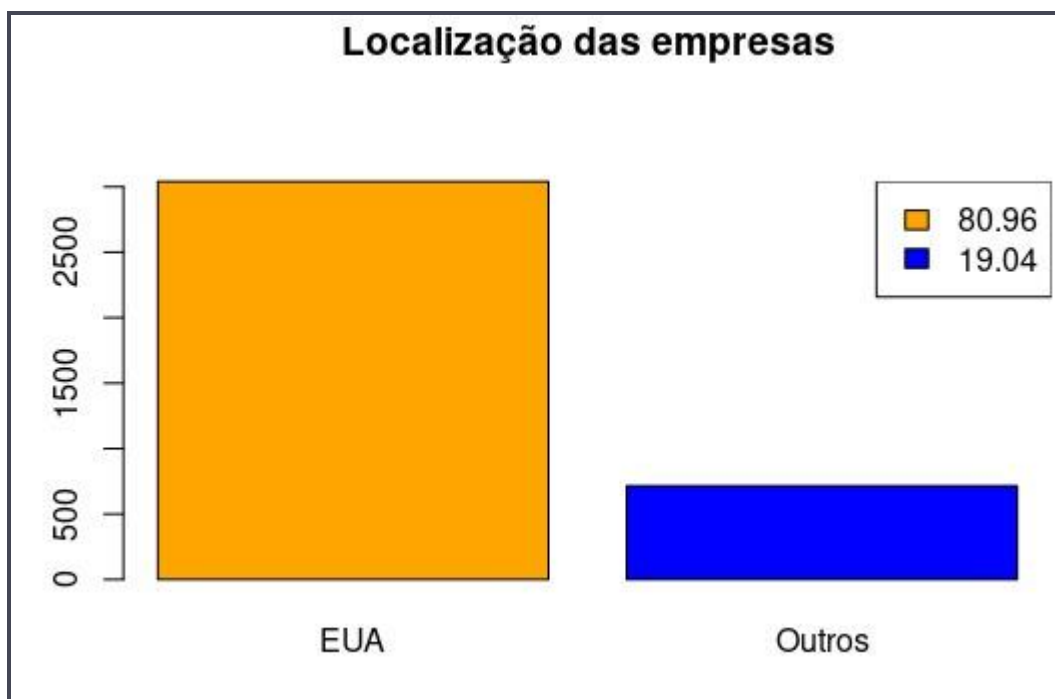
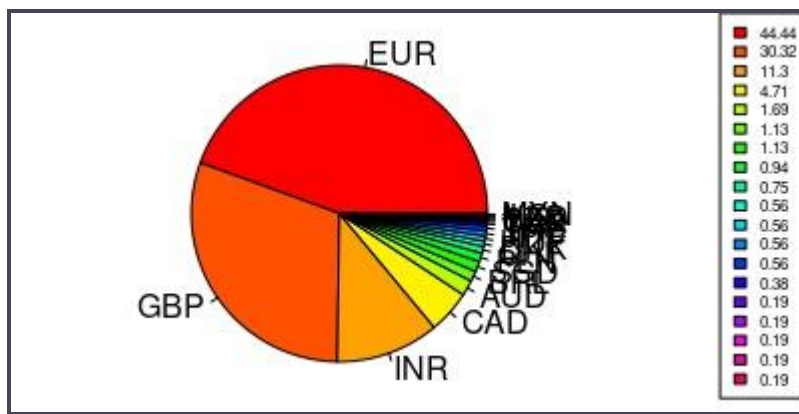
Com esse relatório podemos perceber, por exemplo que a maioria dos salarios sao pagos em dólar, em compensação a grande maioria das empresas são norte americanas tornando lógica a conclusão anterior, porém é visto que cerca de metade dos empregos

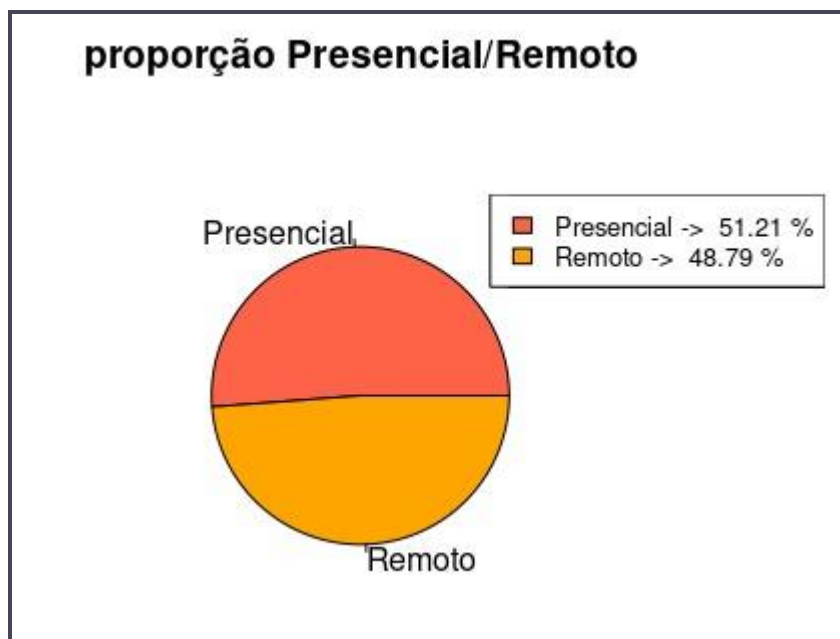
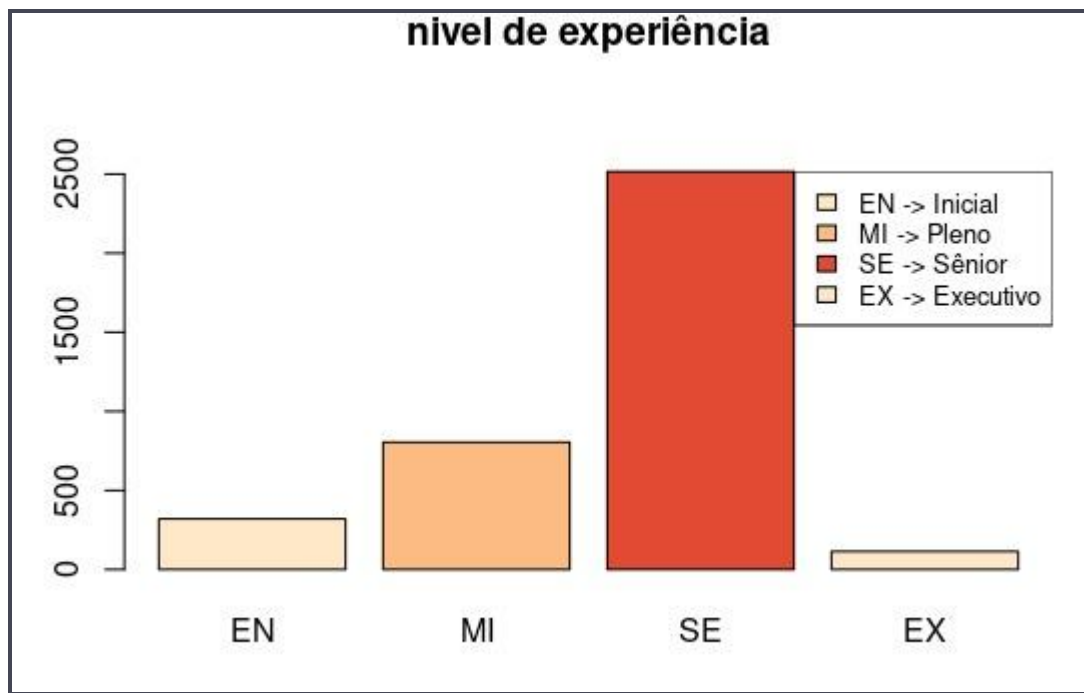
presentes na base de dados é de trabalho remoto, ou seja, dá para trabalhar nessas empresas de qualquer lugar do mundo.

aqui farei uma breve explicação de cada coluna do DataSet para se entender melhor, logo depois colocarei os gráficos tirados da própria base de dados para obter os insights:

1. **work_year:** Traz o ano em que os dados foram obtidos sendo que, a base tem dados de 2020 até 2023
2. **experience_level:** O nível de experiência do funcionário durante o ano, esse nível é um nível hierárquico que é feito com base na carreira, podendo ser Junior: quem acaba de entrar no mercado de trabalho, pleno : profissional que tem cerca de 5 anos de experiência, já possui certa experiência e pode até obter pós-graduação. Sênior: possui mais de 10 anos de experiência , trabalhador experiente e pronto para assumir lideranças e o executivo : veio para assumir cargos altos de liderança e tem alto nível de conhecimento do assunto
3. **employment_type:** Não obtive informações o suficiente para confirmar o que seria mas com base em meu conhecimento e nas discussões da plataforma Kaggle, acredito que seja tipos de funcionários divididos em “freelancers” e contratados.
4. **job_title:** É o cargo em que a pessoa se encontra, visto que ciência de dados é uma área de atuação, existem milhares .
5. **salary:** A quantidade de salário anual recebido.
6. **salary_currency:** A moeda corrente em que o salário foi pago.
7. **employee_residence:** O local da residência primária do contratado.
8. **remote_ratio:** Diz se o trabalho é feito de maneira remota ou presencial, a base traz dados como “0”, “50” e “100”, acredito que seja a porcentagem do trabalho realizado, visto que existem trabalhos que são feitos de maneira metade presencial e método remota, porém nao tive informações o suficiente para concretizar essa hipótese.
9. **company_location:** O local principal da empresa.
10. **company_size:** O tamanho da empresa baseado na quantidade de funcionários anuais que possui.







proporção Presencial/Remoto nos EUA

