



THE DEVELOPER'S CONFERENCE

Trilha – Data Science

André Sionek

Engenheiro de Dados na EmCasa



THE
DEVELOPER'S
CONFERENCE

**O que faz um cientista de dados
ganhar mais de U\$100k/ano?**



THE
DEVELOPER'S
CONFERENCE

Quanto é o bastante?



THE
DEVELOPER'S
CONFERENCE



Previsão Inicial de Carreira



2009/12 - Bolsista de Iniciação Científica:

$$\text{R\$}360,00_{\text{mes}} \times 12_{\text{meses/ano}} \times \text{U\$}0,26/\text{R\$} = \text{U\$}1.123/\text{ano}$$

2013/14 - Bolsista de Mestrado:

$$\text{R\$}1.200,00_{\text{mes}} \times 12_{\text{meses/ano}} \times \text{U\$}0,26/\text{R\$} = \text{U\$}3.744/\text{ano}$$

2015/18 - Bolsista de Mestrado:

$$\text{R\$}2.500,00_{\text{mes}} \times 12_{\text{meses/ano}} \times \text{U\$}0,26/\text{R\$} = \text{U\$}7.800/\text{ano}$$

Expectativa Data Science



THE
DEVELOPER'S
CONFERENCE

≡ EXAME

↗ Sergio Moro Greves Previdência Compliance Revista Newsletter

Investimentos BTG

Cargo: Cientista de dados

O que faz: “garimpa”, analisa e percebe tendências em dados para trazer soluções para problemas da empresa e insights para o negócio.

Perfil: formação em grande parte ligada à área de exatas: matemática, ciências da computação, análise de sistemas, estatística, física.

Salário: de 12 mil reais a 15 mil reais

Percentual de aumento na demanda no 1º semestre: a posição é nova na Michael Page

Motivo em alta 2017: empresas estão desenvolvendo suas áreas de inteligência de mercado, CRM (customer relationship marketing) / DBM (Database Marketing), análise de dados estruturados e não estruturados. Área digital está se consolidando e tecnologia já está diretamente presente no negócio das empresas.

Realidade Data Science



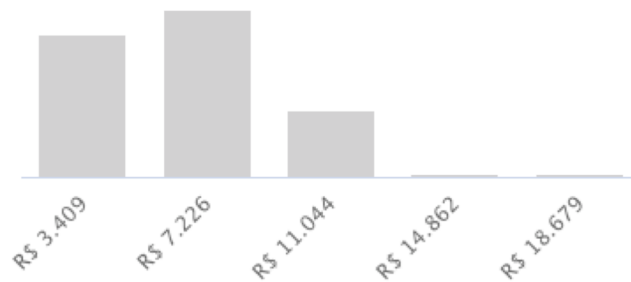
Salário médio bruto

96 Postados

R\$ 8.056/mensal



Gráfico de variação salarial



Realidade Data Science



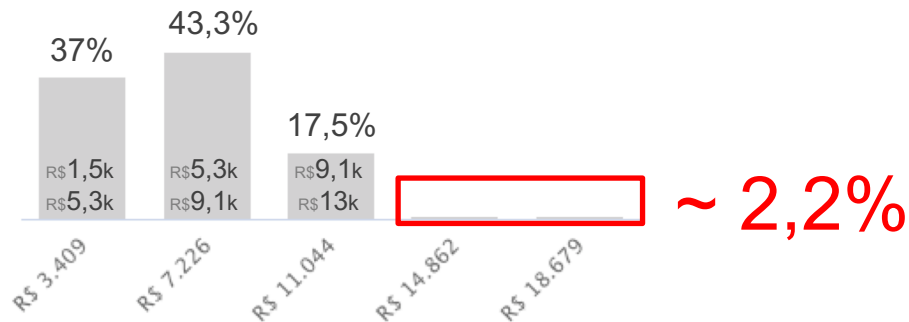
Salário médio bruto

96 Postados

R\$ 8.056/mensal



Gráfico de variação salarial



Realidade Data Science



37% dos cientistas de dados

$$\text{R\$}3.400_{\text{mes}} \times 13,3_{\text{meses/ano}} \times \text{U\$}0,26_{\text{R\$}} = \text{U\$}11.757_{\text{ano}}$$

43% dos cientistas de dados

$$\text{R\$}7.200_{\text{mes}} \times 13,3_{\text{meses/ano}} \times \text{U\$}0,26_{\text{R\$}} = \text{U\$}24.897_{\text{ano}}$$

17% dos cientistas de dados:

$$\text{R\$}11.000_{\text{mes}} \times 13,3_{\text{meses/ano}} \times \text{U\$}0,26_{\text{R\$}} = \text{U\$}38.038_{\text{ano}}$$



glassdoor®

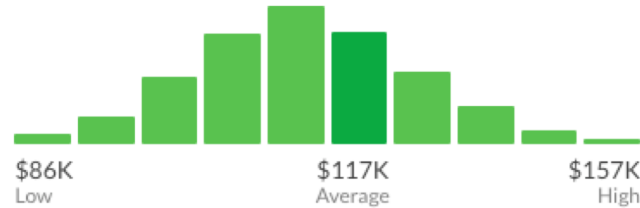
Data Scientist Salaries

4,354 Salaries Updated Jun 11, 2019

Industries Company Sizes Years of Experience

Average Base Pay

\$117,345 /yr



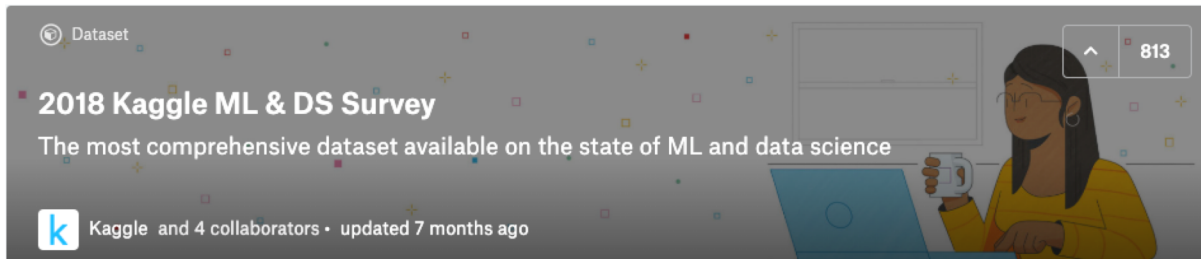
Additional Cash Compensation

Average **\$11,530**

Range **\$3,933 - \$26,784**

How much does a Data Scientist make?

The national average salary for a Data Scientist is \$117,345 in United States. Filter by location to see Data Scientist salaries in your area. Salary estimates are based on 4,354 salaries submitted anonymously to Glassdoor by Data Scientist employees. [Less](#)



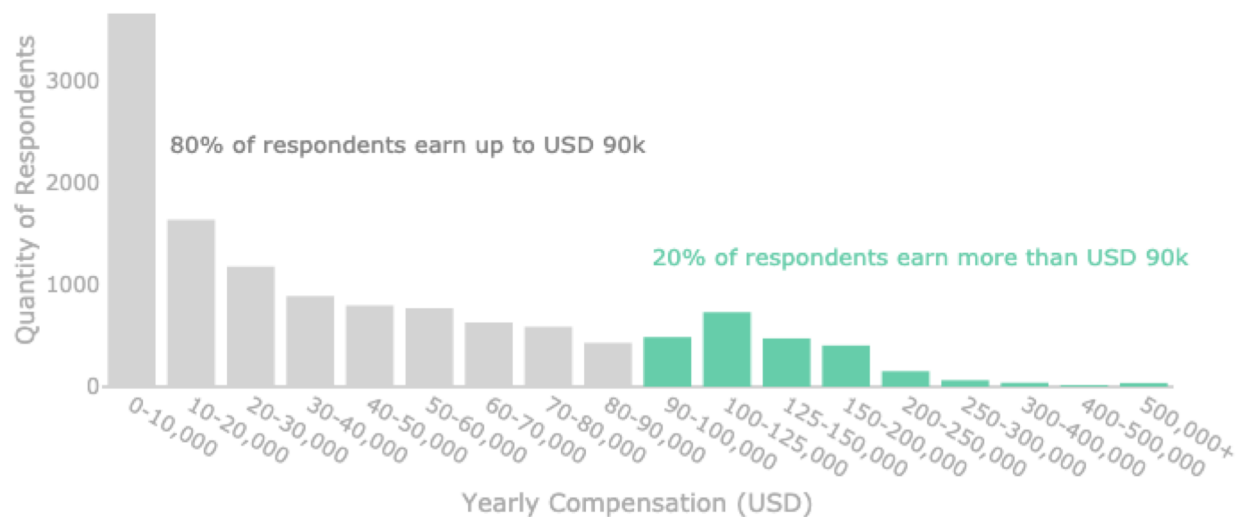
- Pesquisa online com membros da comunidade
- 23.859 respostas
- Não representa todo o universo de profissionais de dados, mas é uma boa base de comparação.

Quanto um Kaggler recebe?



THE
DEVELOPER'S
CONFERENCE

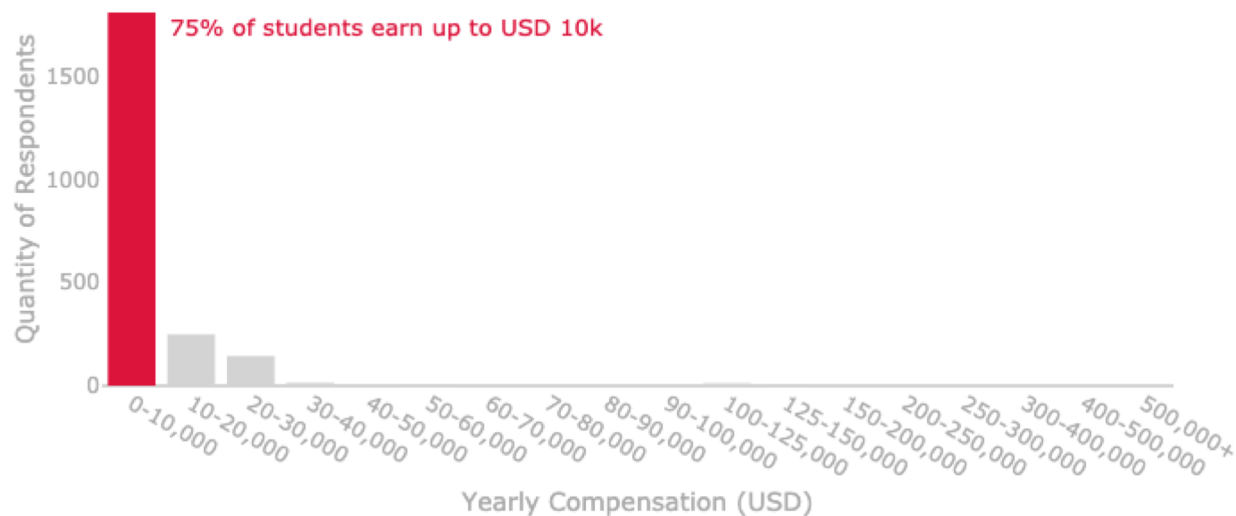
How Much Does Kagglers Get Paid?



Quanto um estudante recebe?



Do Students Get Paid at All? *only students*

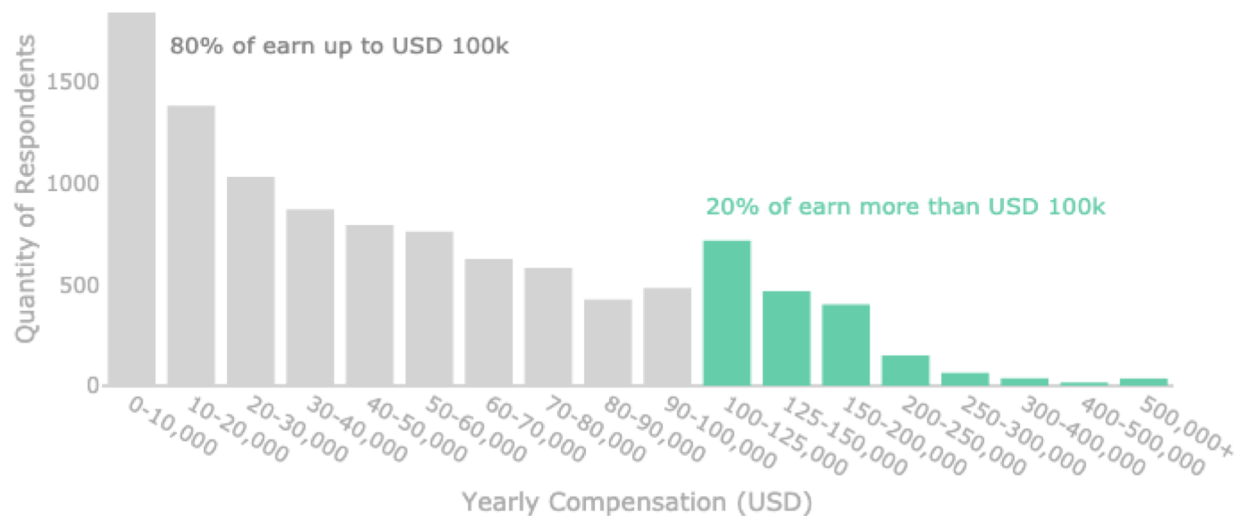


Tirando os estudantes



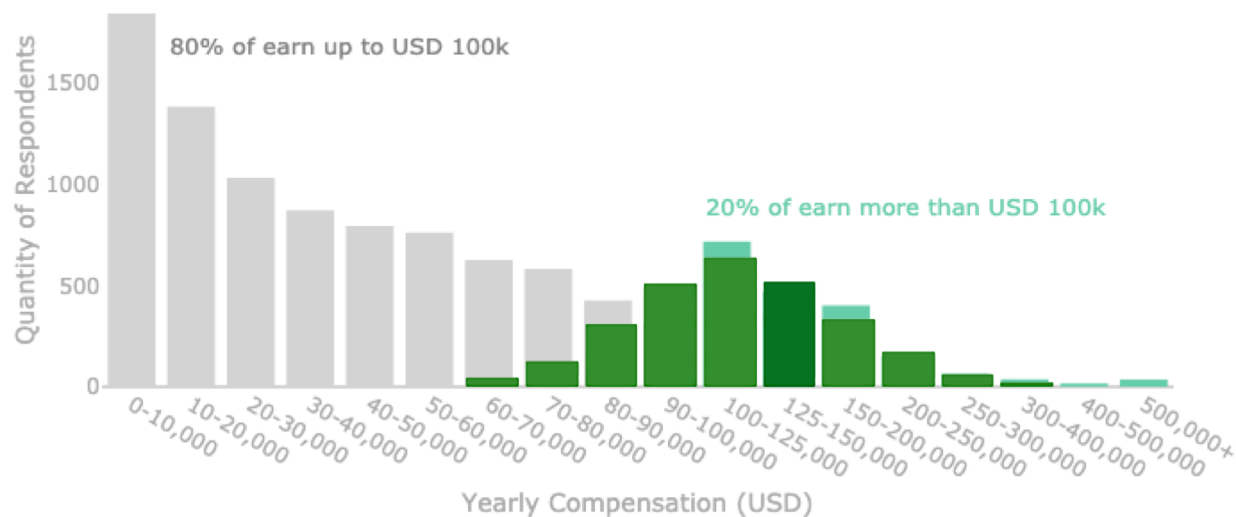
THE
DEVELOPER'S
CONFERENCE

How Much Does Kagglers Get Paid? *without students*



Comparando com Glassdoor

How Much Does Kagglers Get Paid? *without students*





THE
DEVELOPER'S
CONFERENCE

O que faz um cientista de dados estar no top 20%?

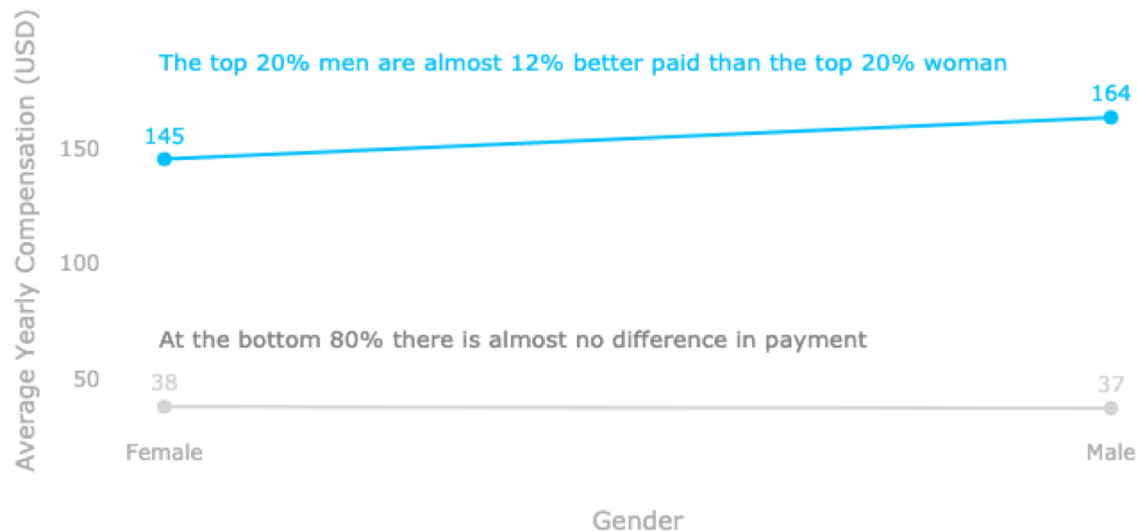
ganhar mais de U\$100k/ano

Impacto do gênero



THE
DEVELOPER'S
CONFERENCE

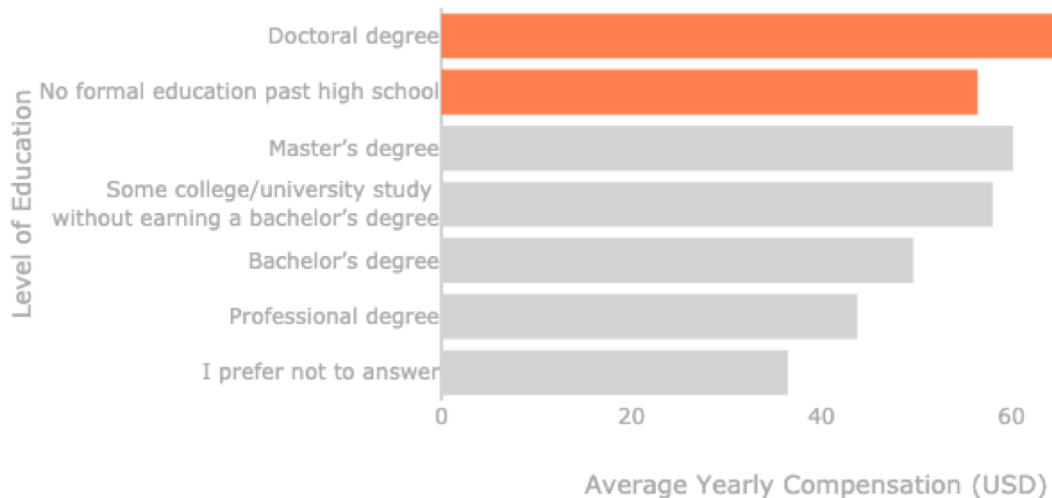
What is the gender difference in compensation at the top 20%?
without students



Impacto da educação formal



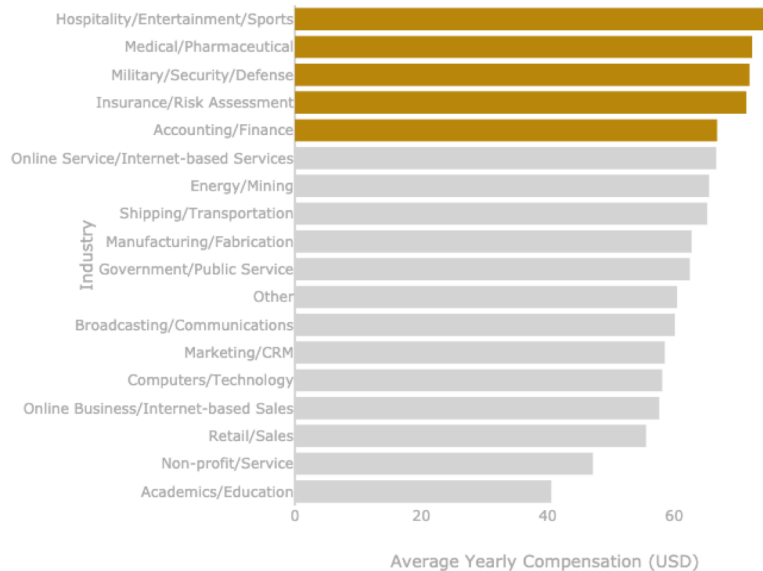
Impact of Formal Education on Compensation *without students*



Impacto da indústria



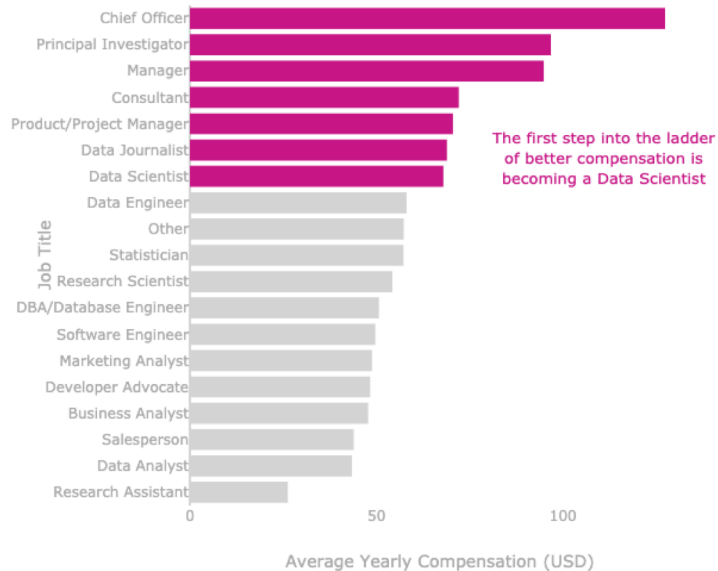
Average Compensation per Industry | Top 5 in Color
without students



Impacto do cargo



Average Compensation per Role | Top 7 in Color
without students

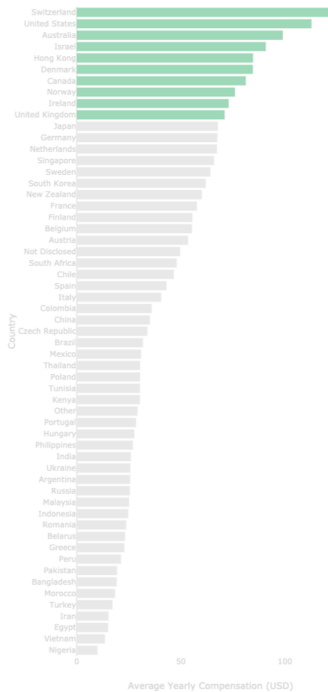


Impacto do país

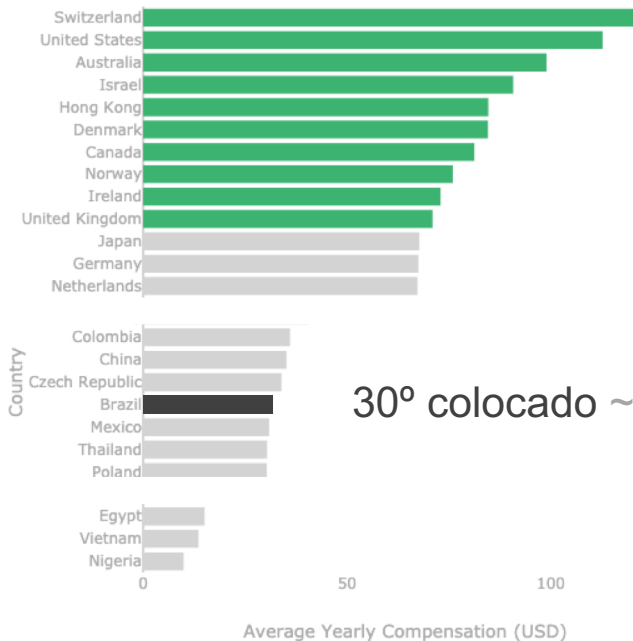


THE
DEVELOPER'S
CONFERENCE

Average Compensation per Country - Top 10 in Color
without students



Average Compensation per Country - Top 10 in Color
without students



30º colocado ~US\$31k/ano

Average Yearly Compensation (USD)

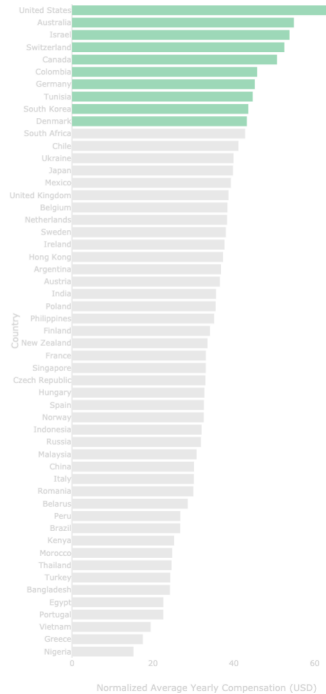
Impacto do país

(normalizado pelo custo de vida expatistan.com)

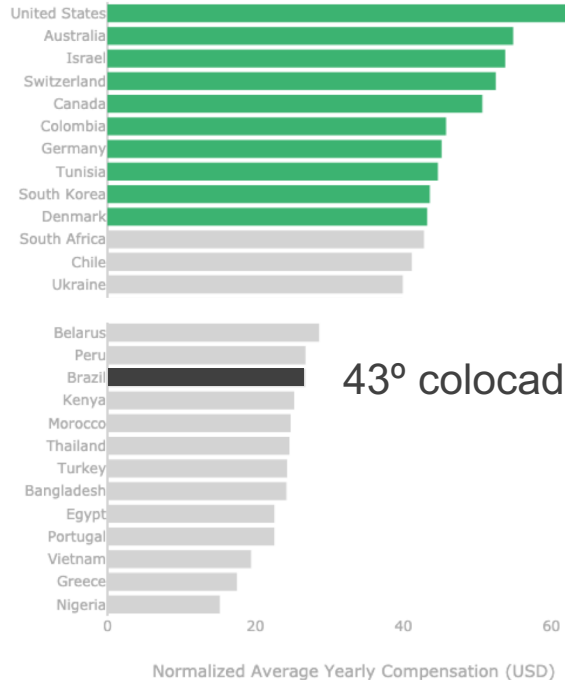


THE
DEVELOPER'S
CONFERENCE

Normalized Average Compensation per Country - Top 10 in Color
without students



Normalized Average Compensation per Country - Top 10 in Color
without students





THE
DEVELOPER'S
CONFERENCE

Criando um modelo para entender as características dos top 20%

Metodologia



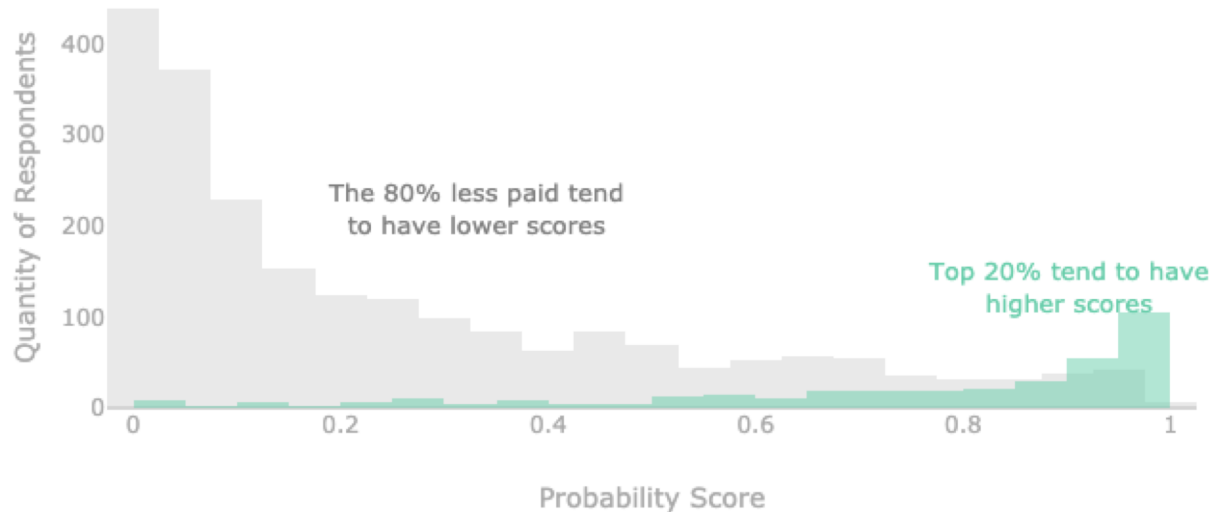
THE
DEVELOPER'S
CONFERENCE

- Seleção de questões relevantes 28 questões
- Desconsiderar respostas que tivessem menos de 5% de respondentes
- Desconsiderar questões altamente correlacionadas $> \text{abs}(\pm 0.5)$
- Undersampling
- Regressão Logística

Distribuição de Scores



**Distribution of Scores From the Top 20% and Bottom 80%
*test data***

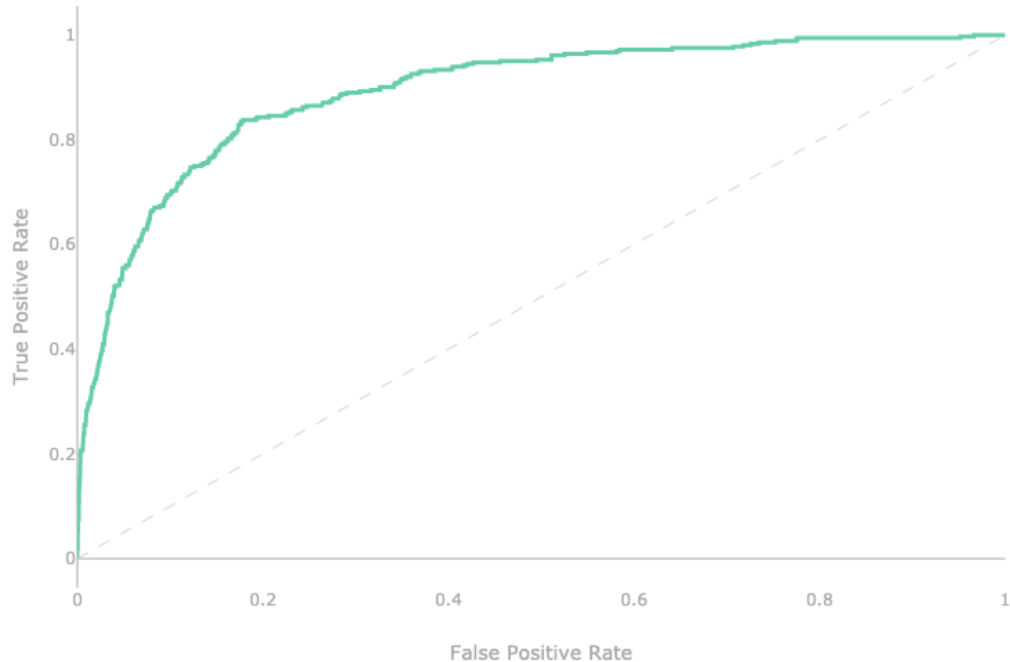


Curva ROC



THE
DEVELOPER'S
CONFERENCE

Receiver Operating Characteristic Curve
test data





THE
DEVELOPER'S
CONFERENCE

Como aumentar as suas chances de estar nos Top 20%



THE
DEVELOPER'S
CONFERENCE

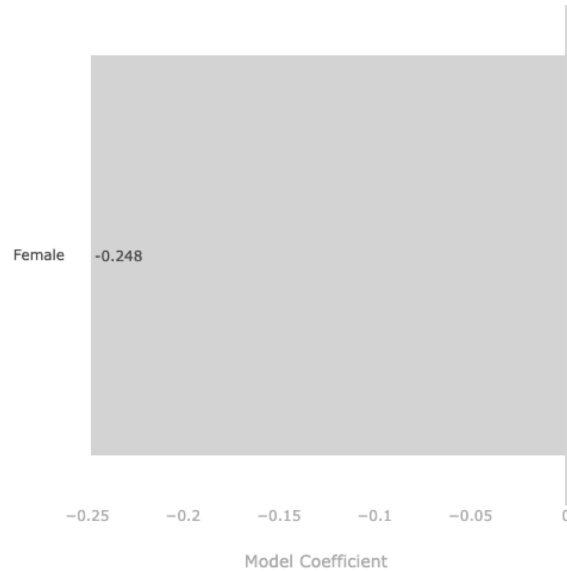
Coef. Linear (intercept) = 0.91

Gênero

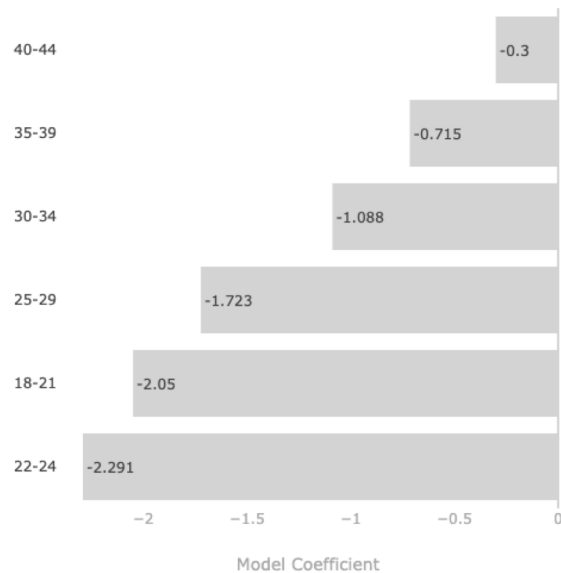


THE
DEVELOPER'S
CONFERENCE

Q1. What is your gender?



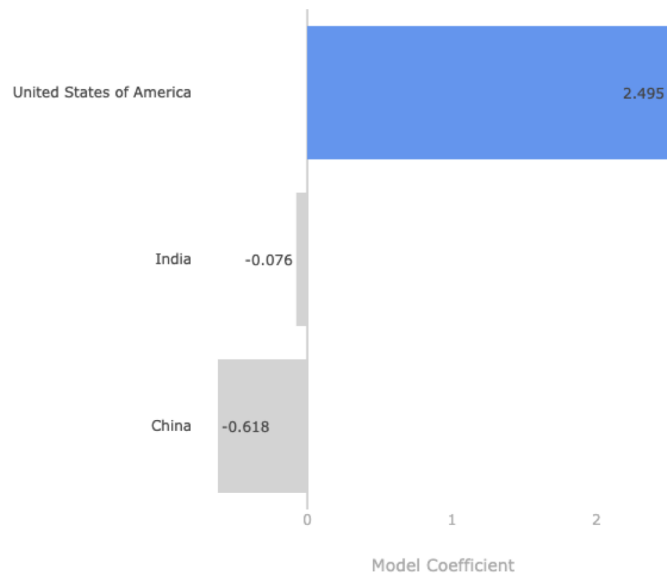
Q2. What is your age (# years)?



País

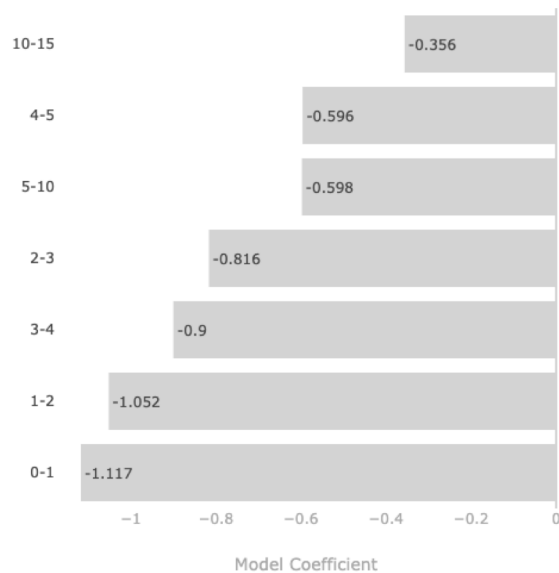


Q3. In which country do you currently reside?



Experiência

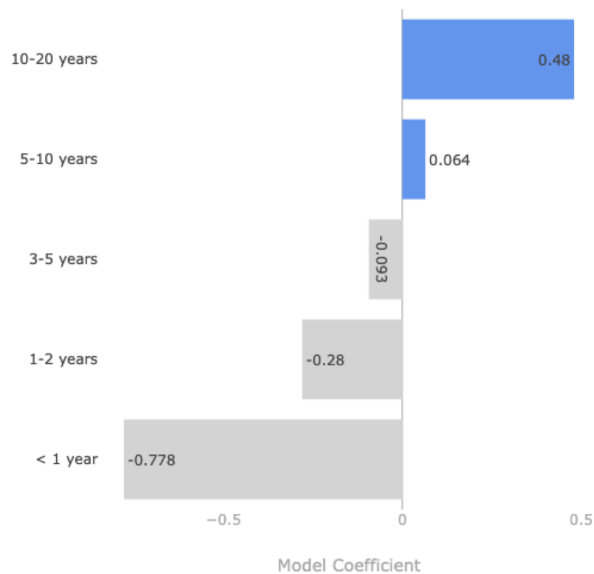
Q8. How many years of experience do you have in your current role?



Experiência analisando dados



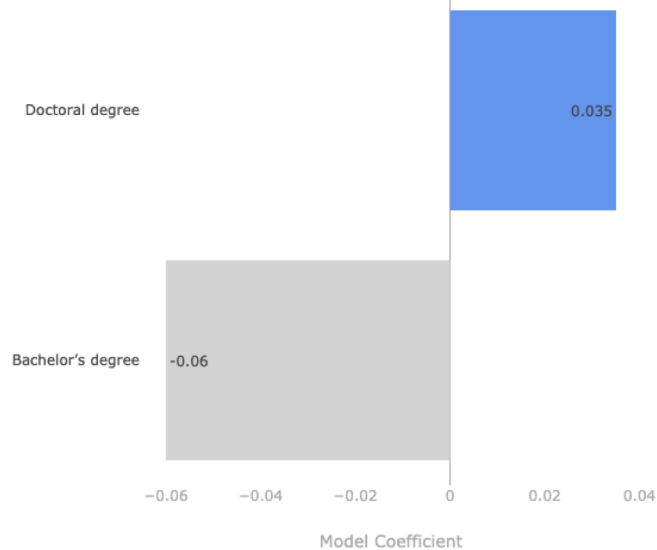
Q24. How long have you been writing code to analyze data?



Educação Formal

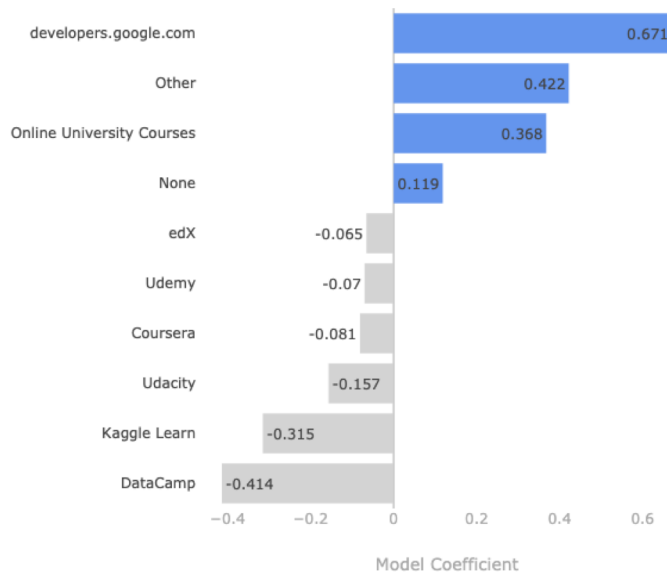


Q4. What is the highest level of formal education that you have attained or plan to attain within the next 2 years?



Cursos online

Q36. On which online platforms have you begun or completed data science courses?

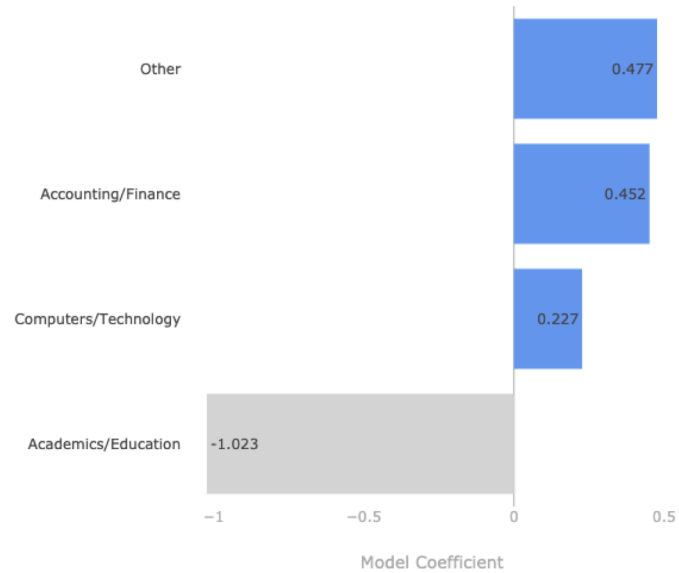


Indústria



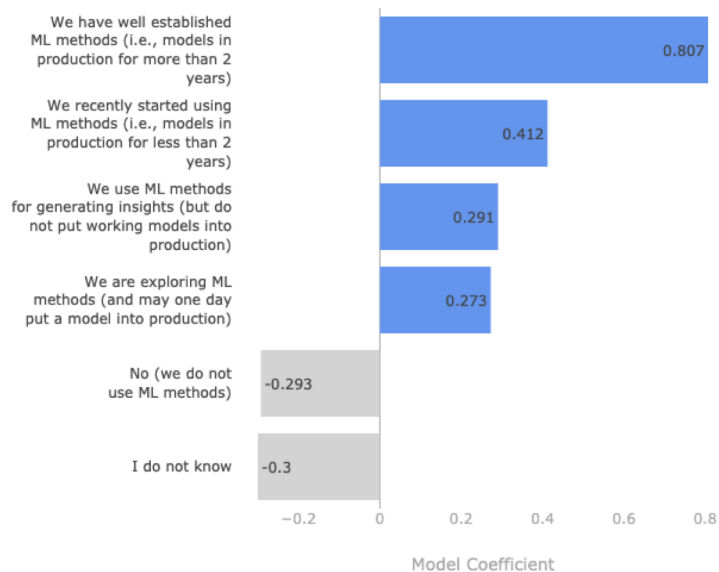
THE
DEVELOPER'S
CONFERENCE

Q7. In what industry is your current employer/contract (or your most recent employer if retired)?



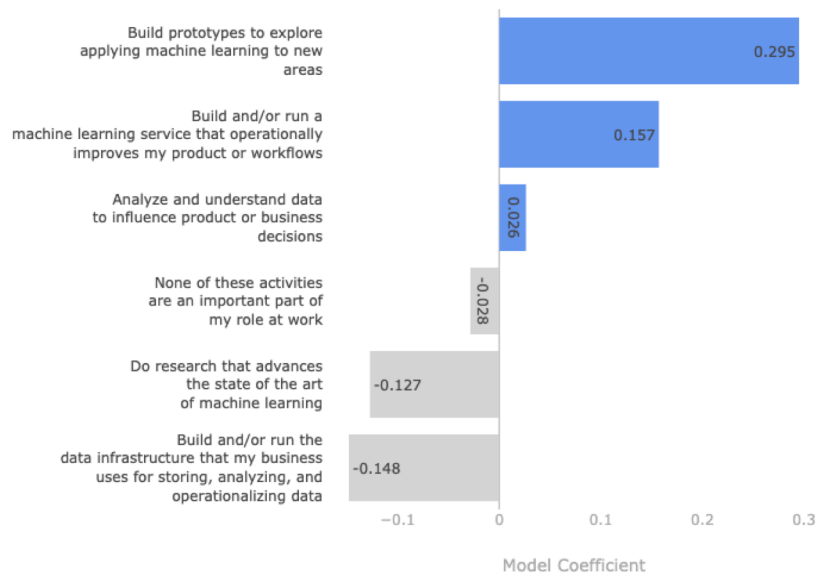
Uso de ML no trabalho

Q10. Does your current employer incorporate machine learning methods into their business?



Atividades no trabalho

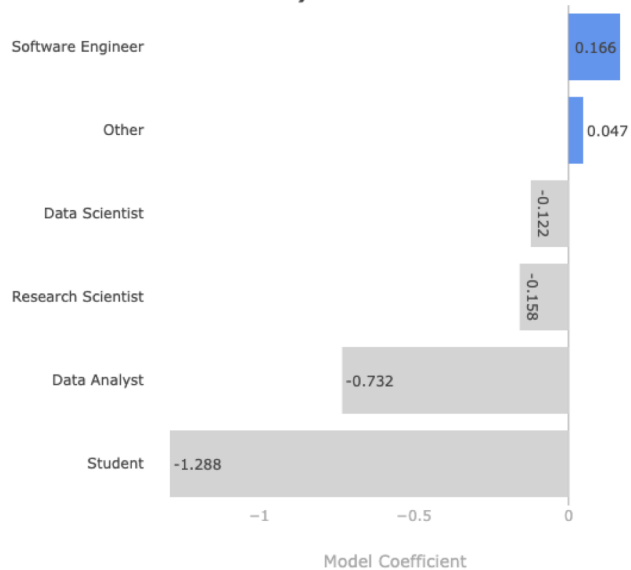
Q11. Select any activities that make up an important part of your role at work



Cargo

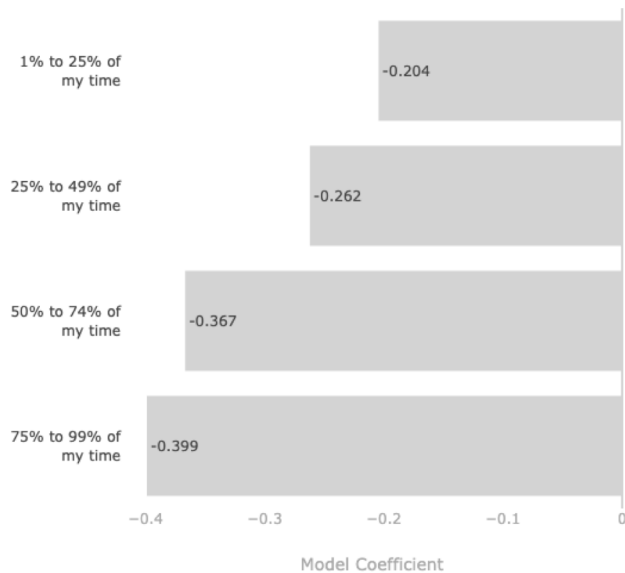


Q6. Select the title most similar to your current role (or most recent title if retired)



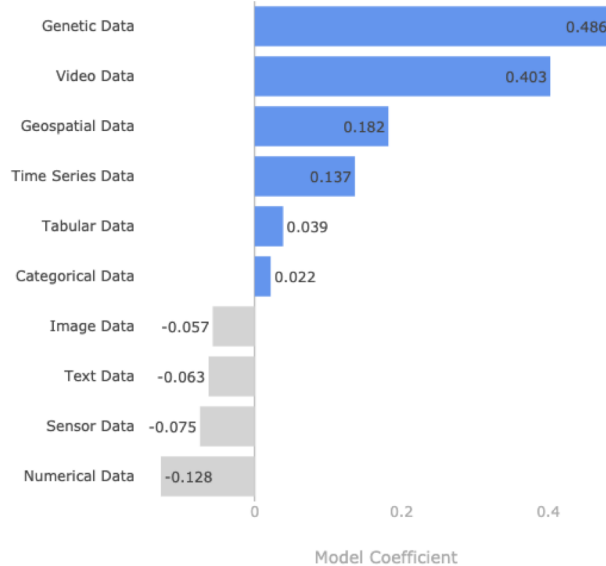
Porção de tempo codando

Q23. Approximately what percent of your time at work or school is spent actively coding?



Tipos de datos

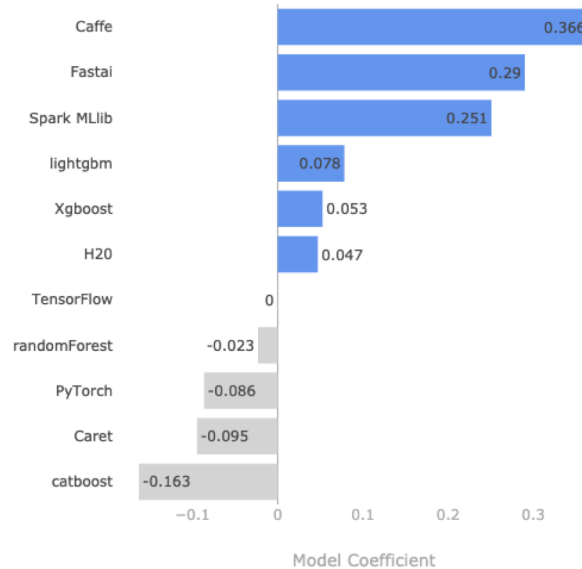
Q31. Which types of data do you currently interact with most often at work or school?



Frameworks de ML

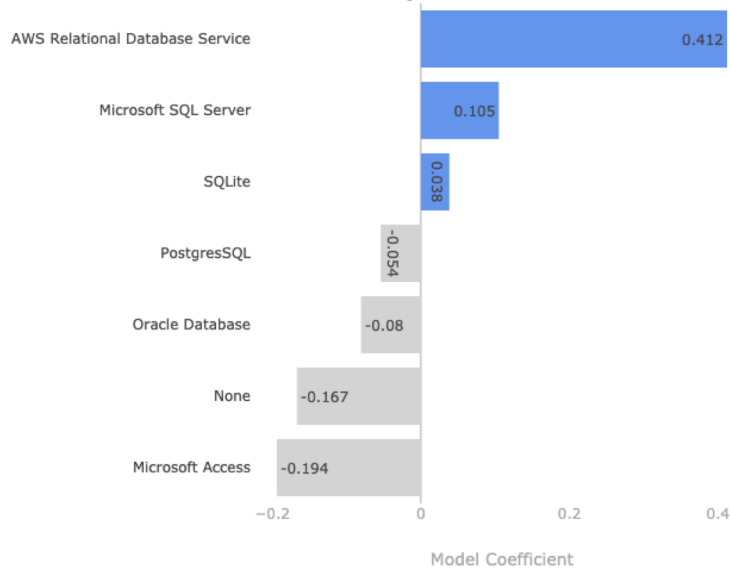


Q19. What machine learning frameworks have you used in the past 5 years?



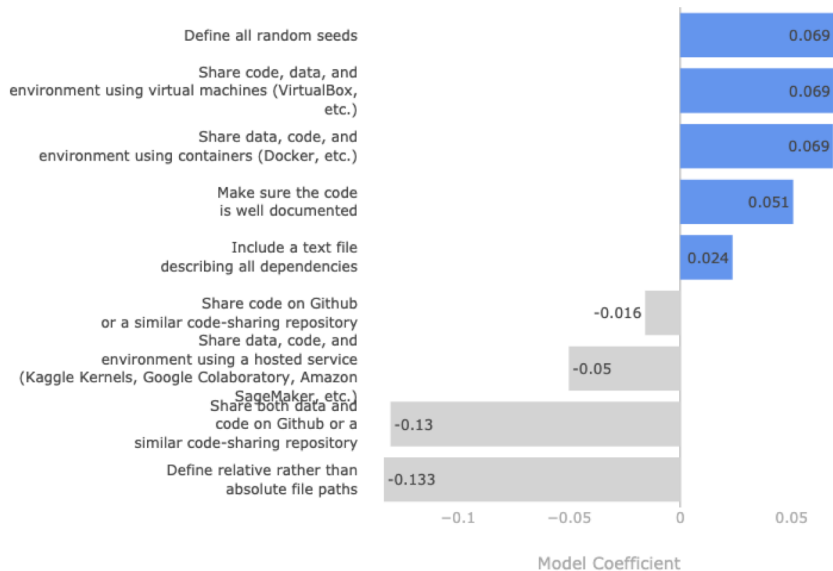
Banco de dados

Q29. Which of the following relational database products have you used at work or school in the last 5 years?



Boas práticas

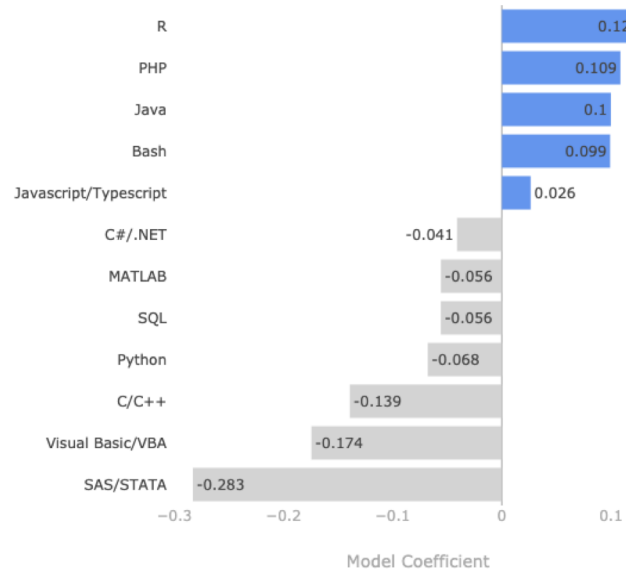
Q49. What tools and methods do you use to make your work easy to reproduce?



Linguagens



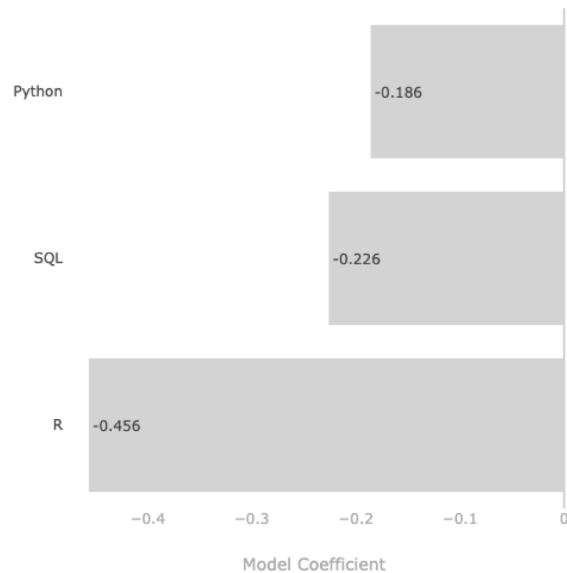
Q16. What programming languages do you use on a regular basis?



Linguagem de uso diário



Q17. What specific programming language do you use most often?

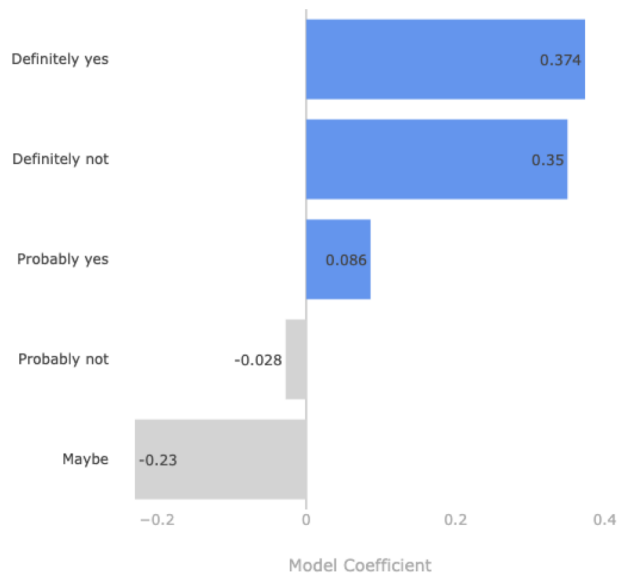


Se considera um cientista de dados?



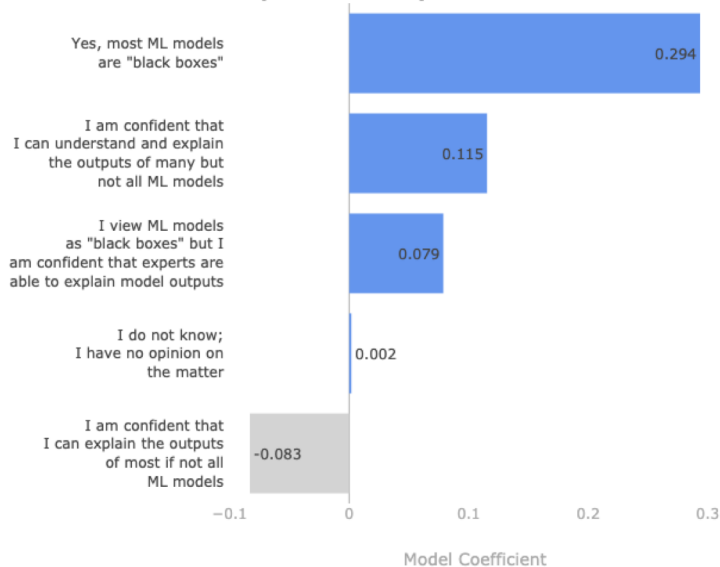
THE
DEVELOPER'S
CONFERENCE

Q26. Do you consider yourself to be a data scientist?



ML é “caixa preta”?

Q48. Do you consider ML models to be "black boxes" with outputs that are difficult or impossible to explain?





THE
DEVELOPER'S
CONFERENCE

Calcule as suas chances!
data-scientist-value.com

Análise completa disponível em:
<https://www.kaggle.com/andresionek/what-makes-a-kaggler-valuable>



THE DEVELOPER'S CONFERENCE