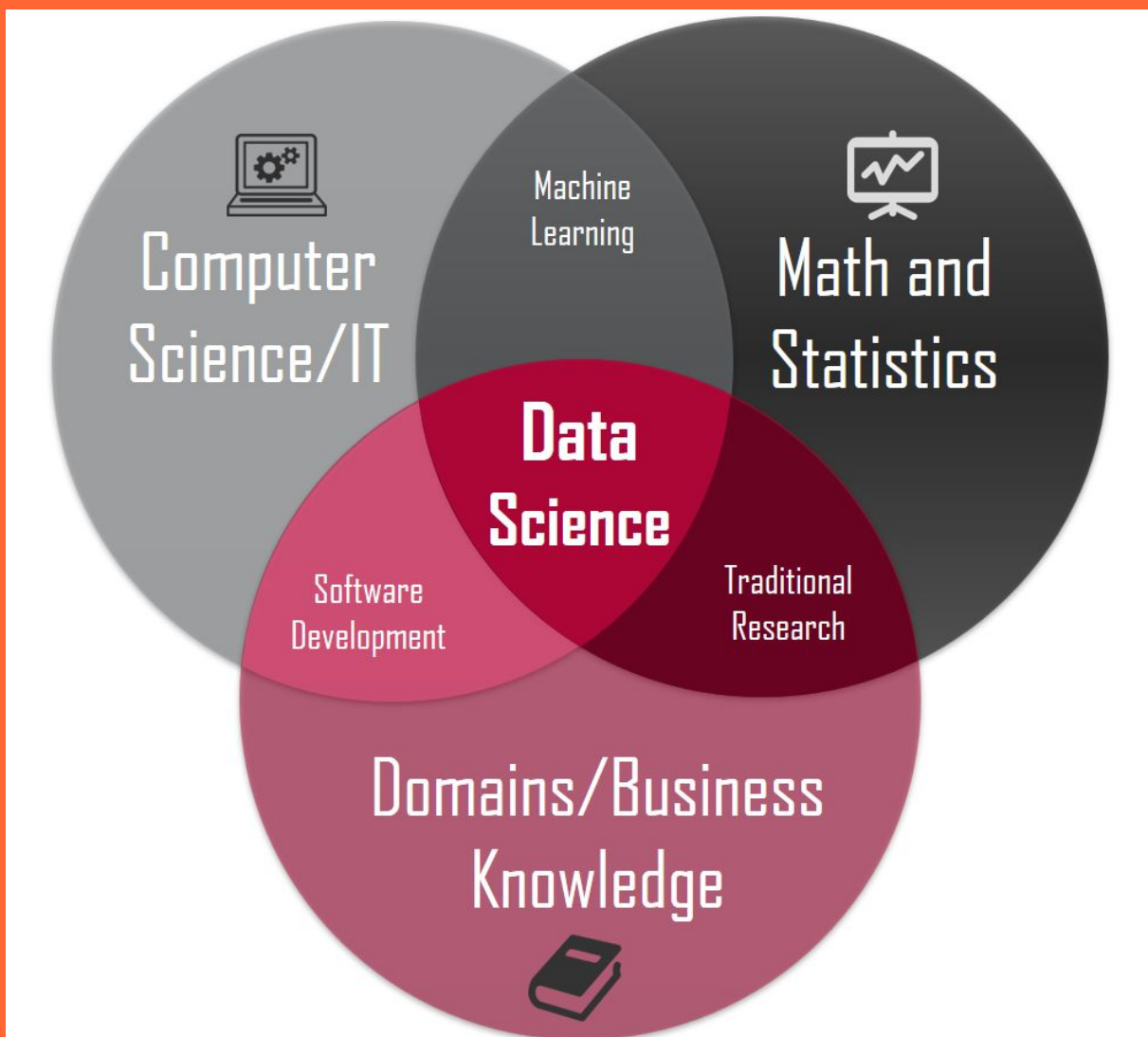


# O DESAFIO

---

Estágio para Ciência de Dados





## O DESAFIO

---

A InfoProp é uma startup de inteligência imobiliária que tem como pilar duas tecnologias: plataforma web para captura e consulta dos dados imobiliários e ciência de dados para enriquecimento da base e construção de inteligência para os usuários. Essa vaga tem como objetivo conhecimentos básicos de estatística, matemática e do RStudio (ou Python - Jupyter).

Por isso, queremos saber como você constrói o seu raciocínio e toma decisões para o tratamento dos dados levantados, assim como na construção de modelos de precificação e construção do algoritmo no RStudio.

O algoritmo que você irá construir irá avaliar o preço de imóveis baseado no valor da oferta (preço dos anúncios presentes na internet). Na InfoProp, nós avaliamos os imóveis com base nos valores transacionados da venda, mas também extraímos, analisamos e enriquecemos as nossas bases com bases de terceiros.

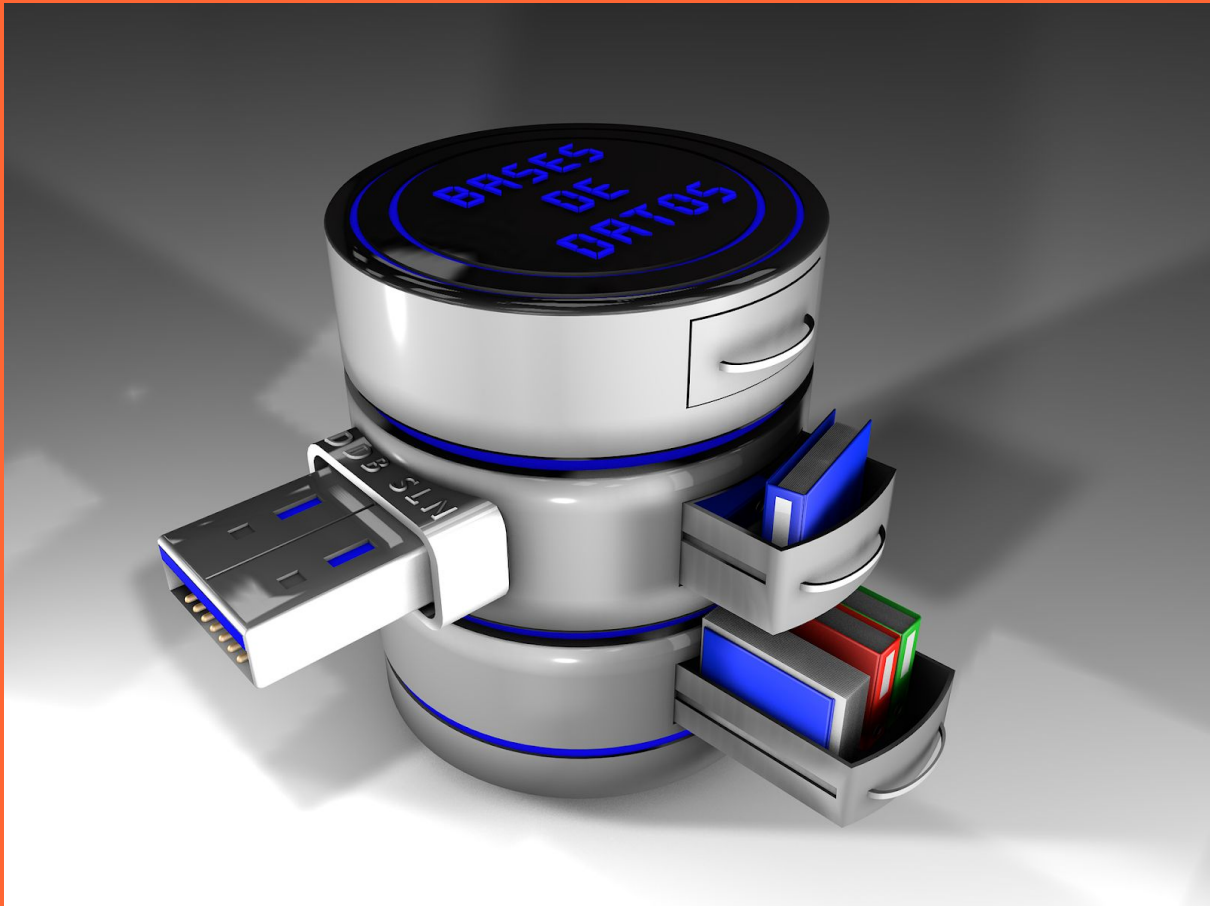
Se tiver qualquer dúvida, pode entrar em contato com qualquer um de nós:

[julio@infoprop.com.br](mailto:julio@infoprop.com.br)

[felipe@infoprop.com.br](mailto:felipe@infoprop.com.br)

[vagas@infoprop.com.br](mailto:vagas@infoprop.com.br)

Boa sorte =D



## Base de Dados de Imóveis Anunciados

---

A base de dados da InfoProp envolve diversas fontes de dados. São dados públicos, privados, proprietários adquiridos pela InfoProp e extraídos da internet.

Para facilitar a construção do seu modelo, nós iremos entregar para você uma base de dados de preços de imóveis extraída em 2017. [Só clicar aqui.](#)

Você irá limpar, classificar, analisar e preparar a base de dados para o modelo. Você deverá importar os dados para o RStudio ou Jupyter Notebook para realizar as ações acima.

Resultado: entregar os seus insights / análises sobre os dados.



## Modelo de Precificação

Na InfoProp, nós temos que desenvolver vários modelos e ir adequando para explicar melhor a realidade de precificação. Os nossos cientista de dados são responsáveis pela qualidade dos dados e dos resultados gerados para nossos clientes.

Nessa segunda etapa do desafio, você deve utilizar a base que foi enviada acima (após o seu tratamento) e criar um modelo linear descritivo dos imóveis disponibilizados no [dataset](#).

Resultado: entregar um documento em markdown ou jupyter explicando o seu modelo, quais hipóteses utilizadas e qual processo de verificação de aderência.

OBSERVAÇÃO: Esse desafio não é um artigo acadêmico e tem como objetivo entender como você construiu o seu raciocínio.

**Stack:** RStudio ou Python (Jupyter)

### DICAS:

- <https://www.coursera.org/learn/machine-learning/resources/JXWWS>
- <https://towardsdatascience.com/create-a-model-to-predict-house-prices-using-python-d34fe8fad88f>

Após finalizar a aplicação, enviar documentações por email para [felipe@infoprop.com.br](mailto:felipe@infoprop.com.br).