

# Web Scraping



# Sobre a Curso-R

# Nossos cursos

## Programação em R

---

Introdução à Programação em R

R para Ciência de Dados I

R para Ciência de Dados II

Pacotes

Introdução ao R com C++

## Visualização de dados

---

Relatórios e Visualização de Dados

Dashboards

Deploy

## Modelagem

---

Regressão linear

Machine Learning

XGBoost

Deep Learning

## Extração de dados

---

Faxina de Dados

Web scraping

# Além da Curso-R



Projetos de análise estatística

Modelagem preditiva

Aplicativos e dashboards em Shiny

Web scraping

Treinamentos in-company

Contribuímos ativamente com a comunidade!



# Sobre o curso

# Web scraping

## Introdução

- O que é e quando fazer web scraping
- O ciclo do web scraping
- Noções de protocolo HTTP
- Acessando dados de APIs

# Web scraping

## Web scraping de fato

- Como fazer requisições HTTP e baixar páginas web a partir do R
- Como estruturar dados de arquivos .xml, .html e .json
- Como iterar algoritmos no R

# Web scraping

## Tópicos

- Tratamento de erros
- Paralelização
- Páginas dinâmicas com Selenium



# Dinâmica do curso

- As aulas terão uma seção teórica, de exposição de conceitos, e prática, de aplicação de conceitos.
- Teremos diversos exercícios para serem feitos "em casa".
- O objetivo dos exercícios é gerar dúvidas. **Com exceção da entrega final, nenhum exercício precisará ser entregue.**
- O certificado será emitido mediante uma **entrega final**. O prazo da entrega final será de pelo menos 2 semanas após a última aula.
- Haverá monitoria para esclarecimento de dúvidas sempre 30 minutos antes do início das aulas.

# Dinâmica das aulas

- Mande dúvidas e comentários no chat em qualquer momento.
- Para falar, levante a mão.
- Algumas dúvidas serão respondidas na hora. Outras serão respondidas mais tarde na própria aula ou em aulas futuras.
- Só interrompa quem estiver falando em caso de emergência.

# Tirando dúvidas

- **Não existe dúvida idiota.**
- Nem sempre é trivial fazer a pergunta certa para que outra pessoa esclareça a sua dúvida. Neste curso, **vamos mostrar melhores práticas na hora de fazer perguntas sobre programação.**
- Fora do horário de aula ou monitoria:
  - perguntas gerais sobre o curso deverão ser feitas no Classroom.
  - perguntas sobre R ou Shiny, principalmente as que envolverem código, deverão ser enviadas no [nosso discourse](#).
- [Veja aqui dicas de como fazer uma boa pergunta.](#)

# Pratique

- Não se preocupe com estilos de estudo. Use o que te deixar mais confortável.
- Estude um pouco todo dia. Se você tem apenas uma hora para estudar na semana, faça 3 sessões de 20 minutos em dias diferentes.
- Se teste! Coloque em prática o que você aprendeu resolvendo exercícios ou inserindo o R no seu dia-a-dia.

## Referência

Nahurhodo 205 - Powerpoint é útil para a aprendizagem?

# Extras

- As **duas entregas** finais mais legais ganharão uma vaga em qualquer curso da Curso-R.
- A gravação das aulas ficará disponível no Google Classroom por 1 ano após o final do curso.