

DigitalHouse >

DATA SCIENCE

Programa do curso

3.32 Pages/Visit

Overview

■ Direct Traffic
3,097.00 (40.0%)

■ Search Engines
2,910.00 (38.04%)

■ Referring Sites
1,642.00 (21.47%)



Quem somos

Somos um **centro de formação de habilidades digitais** que desafia a repensar tudo, **aprendendo a passar da ideia para a ação**. Incorporamos o conceito de *Coding School*, oferecendo cursos intensivos em desenvolvimento de aplicativos *Web, mobile, marketing digital, UX Design, data science* e *analytics*, entre outros.

Sabemos que a melhor maneira de ensinar essas disciplinas é com uma metodologia de “aprender fazendo” que **integre, permanentemente, a teoria aos casos práticos**.

Queremos que os alunos tragam ideias e possam desenvolvê-las, recebendo um *feedback* constante dos professores e instrutores, que, por serem especialistas nesse setor atual, poderão orientá-los no caminho da aprendizagem.

Queremos passar do uso da tecnologia para sermos protagonistas da sua criação. Por isso, na Digital House formamos as novas gerações de programadores e profissionais do mundo digital, para que eles sonhem e inovem com sucesso.

Buscamos treinar e estimular nossos alunos para que rapidamente possam fazer parte de equipes de empresas globais, criar suas *startups* ou trabalhar como *freelancers* em qualquer lugar do mundo.

Para quem é o curso?



ESTUDANTES, PROFISSIONAIS DE DIVERSAS ÁREAS

Economistas, administradores de empresas, cientistas sociais, físicos, químicos, médicos, entre outros que desejam aprender Data Science (Ciência de Dados) para conquistar novas habilidades, criar novos projetos e melhorar sua comunicação com engenheiros de Big Data ou especialistas em Machine Learning.

PROFISSIONAIS DE MARKETING

que queiram enriquecer sua análise de dados para inovar nas suas estratégias de segmentação, personalização da oferta e previsões de todos os tipos sobre seus clientes.

ENGENHEIROS DE SISTEMAS E ANALISTAS DE SISTEMAS

que queiram se especializar, incorporar os conhecimentos de estatística e *Machine Learning* e adquirir habilidades práticas com as ferramentas específicas para fazer análises e organização de dados.

EMPREENDEDORES

que queiram criar seu próprio negócio com base em dados e/ou técnicas de inteligência artificial.

PROGRAMADORES

que precisam se atualizar e conhecer as principais técnicas da ciência de dados.


CIENTISTAS

de qualquer área em busca de ferramentas mais sofisticadas para realizar suas análises de dados

Programa do curso

Ao final do curso, os participantes serão capazes de:

- » Extrair, consultar, limpar e agregar dados para sua análise.
- » Realizar análises visuais e estatísticas de dados, usando o Python e suas bibliotecas.

- 
- » Construir, implementar e avaliar problemas de Data Science usando os algoritmos de Machine Learning.
 - » Usar as ferramentas de visualização adequadas para comunicar suas conclusões.
 - » Criar relatórios claros para as diversas partes envolvidas.
 - » Investigar, modelar e validar processos de resolução de problemas aplicados a *datasets* provenientes de diversos setores para oferecer experiências em vários tipos de problemas e soluções do mundo real.

Proposta do curso



- » A necessidade de contar com analistas de dados qualificados cresce rapidamente em todos os setores empresariais e industriais, assim como nos âmbitos onde se desenvolvem novos paradigmas educacionais, sociais e científicos.
- » O curso de Data Science da DH oferece conhecimentos sólidos no processo de extração, processamento e análise de grandes quantidades de dados, com o objetivo de apoiar a tomada de decisões estratégicas em organizações de todos os tipos.
- » O programa está estruturado com base no ciclo de vida completo dos dados, da sua coleta e preparação até a sua análise e visualização efetiva.
- » Você aprenderá as habilidades e os conhecimentos fundamentais para ter o seu primeiro trabalho como analista de dados.
- » E descobrirá os caminhos possíveis para continuar a sua aprendizagem nos temas de especialização que sejam do seu interesse.

Programa do curso

INTRODUÇÃO À DATA SCIENCE e ANÁLISE EXPLORATÓRIA

MÓDULO 0:

Fundamentos: NumPy, Stats, SQL e DataViz

- **Python e Numpy:** demonstrar conceitos de programação usando as ferramentas Python e NumPy para tratar bases de dados e DataFrames.
- **Estatística descritiva:** rever e aplicar os fundamentos de estatística descritiva.
- **SQL e Bancos de Dados:** conhecer diferentes tipos de bancos de dados, rever comandos em SQL e realizar atividades práticas, obtendo dados de bancos de dados.
- **Introdução a Gráficos e Visualização:** Realizar atividades práticas usando notebooks Jupyter com biblioteca plotly e matplotlib para visualização de dados.

DESAFIO 0

Com um dataset com dados demográficos do IBGE, os participantes realizarão uma análise exploratória utilizando NumPy e Matplotlib, aplicando técnicas básicas de estatística descritiva.*

MÓDULO 1:

Exploratory Data Analysis (EDA), Pandas e SciPy

- **Pandas e Tabelas Pivot:** introdução a Pandas (biblioteca para ler, limpar, realizar o parsing e representar por gráficos os dados usando funções booleanas, indexação, séries, joins e outras funcionalidades).
- **Limpeza de dados:** conceitos de “*tidy data*”, estruturação de dados, introdução ao problema de dados faltantes e noções de expressões regulares.
- **EDA:** Análise Exploratória de Dados em Pandas, com visualização de insights sobre recursos de dados
- **Estatística Inferencial:** Noções de probabilidade, distribuições amostrais, intervalos de confiança, testes de hipótese. Uso de biblioteca SciPy.

DESAFIO 1

Usando dados do INEP, os participantes utilizarão Pandas para limpar os dados. Depois apresentarão formalmente um problema para uma empresa no setor da educação.

*Este e os outros desafios podem mudar.

INTRODUÇÃO A REGRESSÃO, WEB SCRAPING, MACHINE LEARNING, CLASSIFICAÇÃO e CLUSTERIZAÇÃO DE DADOS.

MÓDULO 2:

Introdução a Regressão Linear, Modelagem Estatística, Regressão Logística, Statsmodels e Scikit-Learn

- **Regressão Linear Simples e Múltipla:** estimativa e interpretação dos coeficientes, suposições, medidas de ajuste. Introdução a não linearidades no modelo.
- **Regressão Logística:** princípio de transformação matemática sobre a regressão linear, interpretação dos estimadores, introdução a métricas para modelos de classificação
- **Modelagem Estatística:** análise Exploratória de dados e *feature engineering* com foco no aumento da explicabilidade dos dados.
- **Statsmodels e Scikit-Learn:** introdução ao uso das duas bibliotecas para estimar modelos estatísticos e algoritmos de *Machine Learning* em Python. API, tipos e objetos na *Scikit-Learn*.

DESAFIO 2

A partir do acesso a determinados dados de vendas, os participantes poderão escolher entre realizar uma investigação de marketing ou uma auditoria tributária, usando Pandas, Statsmodels e scikit-learn para transformar os dados, fazer uma regressão linear e visualizar os resultados.

MÓDULO 3:

SQL, Web Scrapping, Consumindo APIs, Introdução ao Machine Learning, Compressão e Clusterização de dados

- **Machine Learning:** introdução a conceitos nodais: desvio, variância, overfitting, underfitting. Identificação de diferentes tipos de algoritmos de *Machine Learning* (supervisionados e não supervisionados). Formas de estimar o erro de generalização (*train-test split*, *cross-validation*). Apresentação geral das técnicas mais utilizadas.
- **APIs e Web Scraping:** Selenium e *Beautiful Soup* para raspar sites, e estrutura de dados JSON para uso de APIs
- **Redução de Dimensionalidade:** introdução ao PCA. Apresentação de Análise de Correspondência Múltipla. Manifold Learning: Técnicas para redução de dimensionalidade não lineares: Escala Multidimensional (MDS), Locally Linear Embeddings e *Isomaps*. Diferenças em relação ao PCA. Métricas de distância.

DESAFIO 3

Os alunos do curso aplicarão conceitos de ETL e agrupamento de dados para identificar segmentação de clientes em dados públicos de *e-commerce* do Brasil.

INTRODUÇÃO a ALGORITMOS DE CLASSIFICAÇÃO, ÁRVORES DE DECISÃO, REGULARIZAÇÃO, ENSEMBLES e SISTEMAS DE RECOMENDAÇÃO

MÓDULO 4:

Classificação, árvores de Decisão e regularização

- **Algoritmos usados para problemas de classificação:** Regressão Logística, Naïve Bayes, *Support Vector Machines* (SVM). Avaliação de algoritmos de classificação: métricas de erro, acurácia e limitações da acurácia. Precisão, *recall*, F1 score, curva ROC, área sob a curva (AUC).
- **Árvores de Decisão:** construção de árvores por meio do algoritmo CART, evitando *overfitting* e *underfitting*, passando pelo mecanismo de separação de dados utilizando as estratégias de entropia de gini.
- **Regularização:** noções de regularização. Regressão LASSO e Ridge. Feature scaling.

DESAFIO 4

Os alunos deverão trabalhar com um modelos de classificação preços de celular em faixas de valores, baseado nas catacterísticas oferecidas por diferentes aparelhos de diferentes fabricantes.

MÓDULO 5:

Ensembles e sistemas de recomendação

- **Modelos de Ensemble:** Noção de Ensemble Learning. Vantagens e desvantagens. Apresentação do meta-algoritmo Bagging e de duas aplicações para árvores de decisão: Random Forest e ExtraTrees. Apresentação do meta-algoritmo de Boosting e de AdaBoost e Gradient Boosting. Entrando na caixa preta: feature importance e partial dependence plot em modelos de ensemble.
- **Sistemas de Recomendação:** técnicas de filtragem de dados e operações com matrizes para relativizar a proximidade de pontos que sejam clientes ou produtos, como base de um sistema de recomendação.

DESAFIO 5

Neste desafio, os aluno deverão trabalhar uma grande base de dados de fraude, totalmente besbalanceado, para previsão de fraude em pagamentos de cartões de crédito em um estabelecimento comercial, e fazerem análise dos resultados sob a perspectiva dos diferentes tipos de erros (falsos positivos e falsos negativos)

CRIAÇÃO DE APIs, DEPLOY, SÉRIES TEMPORAIS, INTERPRETAÇÃO de MODELOS, AGILE, ÉTICA e PRIVACIDADE.

MÓDULO 6:

Criação de APIs, Deploy de Modelos, Séries Temporais e interpretação de Modelos

- **Criação de APIs:** Utilização de ferramentas de Deploy e configuração em URL pública..
- **Deploy de Modelos:** Pipelines: automatização e encapsulamento de etapas do workflow. Deploy de modelos de machine learning em ambiente Web
- **Séries Temporais:** Pré-processamento de séries. Enfoque clássico (tendência, ciclo e resíduo). Sazonalidade. Modelos básicos: média constante, tendência determinística, média móvel, suavização exponencial simples. Modelos avançados: processos AR, MA, ARMA, ARIMA e ARIMAX.
- **Interpretabilidade de Modelos:** Utilização de pacotes especiais do Python como Shap e Lime para interpretação de modelos caixa preta, que não oferecem uma possibilidade de entendimento sobre explicabilidade do modelo.

DESAFIO 6

Utilizando dados de tráfego em páginas Web, os alunos farão uma análise de como os acessos variam com o tempo e qual tendência de comportamento é possível extrair dos dados para os diferentes grupos e levando em conta os eventos do cenário mundial.

MÓDULO 7:

Agile, Ética e Privacidade e Apresentação do Projeto Integrador

- **Agile:** metodologia de gerenciamento de projeto de programação com curtos e testes frequentes para agilidade na execução de tasks, com trabalho em equipe.
- **Ética e Privacidade:** Apresentação dos desafios da Ciência de Dados na sociedade atual, e análise crítica da LGPD (Lei Geral de Proteção de Dados) com as permissões de utilização de dados, preservação de direitos e o que fazer para se adequar à legislação evitando tomada de risco empresarial
- **Projeto Integrador:** projeto de Data Science a ser executado pelos alunos, divididos em grupos de 5 a 7 participantes, contendo todas as etapas de um trabalho semelhante ao encontrado na vida real das empresas. Os professores caminham junto com os alunos durante o processo de construção da solução até a apresentação final

PROJETO INTEGRADOR

Ao longo do curso os alunos desenvolverão em grupo um projeto de ciência de dados, de ponta a ponta, sobre um problema ou desafio decidido pelo próprio grupo. O trabalho, então, será apresentado no último dia de aula e será avaliado e comentado pelos professores e por uma banca avaliadora convidada.



Requisitos Mínimos

Os candidatos a alunos devem ter conhecimentos sólidos em **estatística descritiva** (eg., medidas de tendência central e dispersão, histograma etc.), lógica e aritmética básicas e noções de **programação de computadores**, como estruturas de dados, controle de fluxo, definição de funções etc. É recomendável, mas não obrigatório, ter noções de estatística inferencial (amostragem, intervalos de confiança, testes de hipótese e distribuições de probabilidade) e programação com a linguagem Python (versão 2 ou 3).

1. Candidate-se
2. Estude (nivelamento)
3. Faça o desafio
4. Entrevista

Processo de Admissão

A disponibilidade de vagas nos cursos é limitada a você deverá passar pelo seguinte processo de admissão

01

PARA SE CANDIDATAR

Solicitar e Preencher o formulário de Inscrição no site

[FORMULÁRIO DE INSCRIÇÃO](#)

02

DESAFIO ON-LINE

Faça um exercício real que enviaremos para podermos entender a sua forma de pensar e de resolver situações

03

ENTREVISTA DE ADMISSÃO

Agende uma entrevista para conhecermos o seu perfil, seus interesses e expectativas com relação ao curso

FAQs

POSSO FAZER O CURSO SEM TER EXPERIÊNCIA E/OU CONHECIMENTOS PRÉVIOS COM DADOS?

Sim. Porém, nesse caso recomendamos o curso **Data Analytics**, que em geral é mais adequado para quem está começando nessa área.

AS VAGAS SÃO LIMITADAS?

Sim. Todos os cursos têm no máximo 50 alunos.

COMO É A ENTREVISTA DO PROCESSO DE ADMISSÃO?

É uma conversa, online com um professor. Ela visa (1) verificar se seu interesse pela ciência dos dados, seus motivos e objetivos estão alinhados com a proposta do curso, bem como (2) orientá-lo(a) sobre os desafios do curso e, finalmente, (3) verificar seus conhecimentos sobre os pré-requisitos do curso.

PRECISO TER CONHECIMENTOS MATEMÁTICOS?

Sim. A ciência de dados em geral utiliza análise quantitativa de dados. Por isso é importante que você domine alguns conceitos de lógica, álgebra e estatística do ensino médio.

SE JÁ TENHO CONHECIMENTOS BÁSICOS DE PROGRAMAÇÃO E ESTATÍSTICA VALE A PENA FAZER O CURSO?

Sim. No nosso curso você integrará metodologias e ferramentas para os seus projetos, orientado pelos melhores especialistas do setor e do universo acadêmico. Além disso, terá como colegas pessoas de setores variados, o que agrega um enorme valor à experiência e permite estabelecer relacionamentos de trabalho para o futuro.

PRECISO COMPRAR OU LEVAR UM NOTEBOOK PARA FAZER O CURSO?

Na modalidade **presencial** você contará com os equipamentos necessários em sala de aula (você também pode levar seu notebook, se preferir); na modalidade **online**, você precisa providenciar um computador. Qualquer que seja o caso, é imprescindível que você tenha um equipamento para praticar fora da sala de aula.

QUE TIPO DE COMPROMISSO IMPLICA ESTUDAR NA DIGITAL HOUSE?

Na Digital House você terá aproximadamente 7 horas de aula por semana durante 7 meses, o que representa um total de quase 200 horas. Mas não termina aí. Para aproveitar o curso ao máximo, você terá que (e vai querer!) praticar fora da aula, fazer exercícios, resolver desafios e trabalhar nos projetos que forem pedidos.

HÁ ALGUMA PROVA PARA ENTRAR? EM QUE CONSISTE? É OBRIGATÓRIA?

Alguns conhecimentos básicos você pode aprender online e por conta própria, com a ajuda do nosso material de nivelamento e com apoio dos professores, via colearnings. Depois você fará um teste online que chamamos de "desafio online" para avaliar seus conhecimentos quanto aos pré-requisitos do curso.

QUEM SERÃO OS MEUS PROFESSORES?

Um dos nossos principais ativos são os professores. Eles são profissionais da área, reconhecidos e bem-sucedidos, motivados e com vontade de ensinar, usando esses novos modelos de Educação. Além disso, são capacitados internamente pelo nosso departamento acadêmico e pedagógico para que os conteúdos estejam sempre alinhados. Algumas aulas contam ainda com a participação de especialistas que são referência no setor.

POSSO LEVAR CONVIDADOS?

Você poderá trazer amigos e colegas à sede da Digital House, mas eles só poderão ficar com você nos espaços comuns e de colearning, não nas salas de aula.

Se você tiver alguma dúvida, entre em contato conosco.
Estamos aqui para te ajudar!

DigitalHouse >

CONTATO@BR.DIGITALHOUSE.COM

BR.DIGITALHOUSE.COM

TEL.: (11) 4858-4000