



GUIA DATA MASTERS



DATA MASTERS

CIENTISTA DE DADOS

Olá, seja muito bem-vindo ao Guia **DATA MASTERS** de Cientista de Dados, aqui você encontrará uma seleção de conteúdos completos para auxiliá-lo na sua certificação.

Sabemos que cada pessoa tem sua preferência em como estudar. Por isso, buscamos neste guia, disponibilizar conteúdos em formatos variados como vídeos, artigos, cursos, livros etc. Esta diversidade de materiais lhe proporcionará um aprendizado incrível.

Antes de começar, sugerimos que faça uma autoavaliação de seus conhecimentos, veja na próxima página a lista de temas que abordaremos ao longo deste guia.

Com estes conhecimentos, você será capaz de exercer a função de um Cientista de Dados! Mas você sabe qual a principal função deste cargo?

É analisar e solucionar problemas complexos, e atrelado a isso existe uma série de pré requisitos, dentre eles estão: a limpeza, a estruturação, a organização e a preparação de dados.

Vamos aos estudos então?

Esperamos que aproveite esta jornada ao máximo e conquiste sua Certificação!

VOCÊ ENCONTRARÁ...

07 Probabilidade

- Conceitos básicos;
- Probabilidade Condicional e Independência de Eventos;
- Variáveis Aleatórias;
- Variáveis Discretas;
- Variáveis Contínuas;
- Vetores Aleatórios;
- Principais Distribuições;
- Distribuições Discretas Mais Importante.

11 Estatística

- Análise exploratória;
- Medidas de Centro;
- Amostragem;
- Inferência.

14 Manipulação de Dados

- Conceitos de Modelagem de Dados;
- Comandos SQL e NoSQL;
- Conceitos de Big Data.

17 Programação

- Lógica;
- Boas Práticas de Programação;
- Bibliotecas de APIs de Machine Learning;

20 Técnicas de Regressão

- Linear;
- Múltipla;
- Ridge(E Lasso);
- Seleção de Variáveis;
- Avaliação de performance de modelos.

23 Técnicas de Classificação

- Regressão Logística;
- Árvores de Decisão;
- Naive Bayes.

25 Técnicas de Grupamento e Relacionadas

- K-means;
- Algoritmos hierárquicos aglomerativos e divisivos;
- Latent Dirichlet Allocation.

27 Cursos e Livros

SUMÁRIO

CIENTISTA DE DADOS

07 Probabilidade

11 Estatística

14 Manipulação de Dados

17 Programação

20 Técnicas de Regressão

23 Técnicas de Classificação

25 Técnicas de Grupamento
e Relacionadas

27 Cursos e Livros

ICONOGRAFIA

No conteúdo deste e-book, você encontrará diversos materiais, dentre eles opções pagas e abertas.

Abaixo você confere os significados de cada ícone e ao visualizá-los, basta clicar para acessar.



Artigo



PDF



Curso



Vídeo



Livro



Nível Básico



Nível Intermediário



Nível Avançado

PROBABILIDADE

Qual é a chance de um evento acontecer?
Essa pergunta parece vaga quando não há informações, certo?

Nessa seção do Guia **DATA MASTERS**, você encontrará uma série de conteúdos que auxiliarão no entendimento do conceito básico de Probabilidade até as principais distribuições, ou seja, você irá conferir como utilizar a teoria de Probabilidade para estimar a chance de um evento acontecer.

CONCEITOS BÁSICOS



Probabilidade:
Conceitos Básicos

🕒 13min



Cálculo de
Probabilidades

🕒 10min



Probabilidade da União
de Dois Eventos

🕒 13min

PROBABILIDADE CONDICIONAL E INDEPENDÊNCIA DE EVENTOS



● ○ ○
Probabilidade
Condicional

🕒 9min



● ○ ○
Probabilidade
Condicional

🕒 5min



● ○ ○
Probability Events
Conditional

🕒 15min 🇬🇧



● ○ ○
Independent Events (Basics
of Probability: Independence
of Two Events)

🕒 21min 🇬🇧

VARIÁVEIS ALEATÓRIAS



● ○ ○
O que são Variáveis Aleatórias
Discretas e Contínuas | Distribuição
de Probabilidades

🕒 14min



● ○ ○
Variáveis Aleatórias

🕒 6min



● ○ ○
Variáveis Discretas

🕒 7min



● ○ ○
Variáveis Contínua

🕒 15min

VETORES ALEATÓRIOS



● ○ ○

Vetores Aleatórios

🕒 30min



● ○ ○

Introdução à Estatística
- UFPE

Artigo Acadêmico



● ● ○

Variáveis Aleatórias
Independentes

🕒 11min

PRINCIPAIS DISTRIBUIÇÕES



● ● ○

O que são e como fazer
Distribuição de Probabilidades

🕒 9min



● ● ○

Montando uma distribuição de
probabilidades para variável discreta

🕒 8min

DISTRIBUIÇÕES DISCRETAS




● ● ○

Binomial Distribution

🕒 30min 



Binomial Distribution

🕒 12min 



Processo de Poisson 1

🕒 12min



Processo de Poisson 2

🕒 11min

DISTRIBUIÇÕES CONTÍNUAS



Distribuição Exponencial

🕒 15min



Distribuição Weibull

🕒 6min



Distribuição Gaussiana

🕒 7min



Distribuição Gamma

🕒 10min



Curso completo de
Probabilidade

🕒 35h

ESTATÍSTICA

Há quem diga que os primeiros estudos utilizando estatística foram iniciados em 1662, e suas principais aplicações estavam voltadas para a formulação de políticas públicas e fornecimento de dados demográficos e econômicos.

Apesar de ser uma ciência relativamente antiga, com o passar do tempo ela vem sendo aplicada de diversas maneiras. É comum encontramos estatística em Biostabiliometria, Controle de Qualidade, Pesquisa operacional e claro, em Ciência de Dados.

ANÁLISE EXPLORATÓRIA



● ○ ○


Análise Exploratória
de Dados - UEL

🕒 6h



● ○ ○

Exploratory Data Analysis

🕒 20min 

MEDIDAS DE CENTRO



● ○ ○

Medidas de Centro

🕒 10min



● ○ ○

Introdução à estatística:
média, mediana e moda

🕒 20min

AMOSTRAGEM



● ○ ○

Amostragem Estatística

🕒 25min



● ○ ○

Tipos de amostragens -
Introdução à Estatística

🕒 3min



● ○ ○

Revisão dos métodos
de amostragem

🕒 10min

INFERÊNCIA



Estimadores Pontuais

🕒 30min



Intervalos de Confiança

🕒 30min



Testes de Hipóteses

🕒 25min



Curso Completo
de Estatística

🕒 10h

MANIPULAÇÃO DE DADOS

Em Manipulação de Dados, vamos analisar requisitos, definir o modelo conceitual e o modelo lógico. Por fim, entenderemos como aplicar o modelo físico. Ficou confuso? Não se preocupe! Nessa seção, vamos abordar os conceitos e modelagem de dados e compreender como aplicar os comandos da Linguagem SQL.

Explicaremos o que é NoSQL e lhes apresentaremos de maneira simples o que é BIG DATA.

CONCEITOS DE MODELAGEM DE DADOS



● ○ ○

Entenda o que é a modelagem de banco de dados

🕒 8min



● ○ ○

Modelagem de Dados -
Conceitos de Bancos de Dados

🕒 21min

COMANDOS SQL E NoSQL



● ○ ○

Comandos básicos em SQL -
INSERT, UPDATE, DELETE e SELECT

🕒 15min



● ○ ○

Guia completo
de SQL

🕒 10h



● ○ ○

Introdução a SQL: Consulta e
gerenciamento de dados

🕒 8h



● ○ ○

SQL Tutorial

🕒 10h 



● ○ ○

O que é NoSQL?

🕒 2h



● ○ ○

Top 6 NoSQL
Databases

🕒 30min

CONCEITOS DE BIG DATA



Big Data

🕒 6min



Big Data - O que é e qual
sua importância?

🕒 30min



O que é Big Data -
Conceitos básicos

🕒 18min

PROGRAMAÇÃO

Muito bem! Agora que você chegou até aqui, vamos lhe apresentar os seguintes conteúdos: Conceitos básicos de Lógica, Algoritmos, Linguagem de Programação, Boas práticas e algumas bibliotecas e APIs de Machine Learning.

Se você não tiver familiaridade com a linguagem de programação, fique tranquilo! Pois, nessa seção, você encontrará uma introdução sobre Python e como ele está presente no cotidiano de um Cientista de Dados.

LÓGICA



● ○ ○

Lógica de Programação

🕒 1h30



● ○ ○

Lógica de Programação
com Python

🕒 2h



● ○ ○

Curso Python 01 - Introdução -
Aprenda Programar do ZERO

🕒 2h



● ○ ○

Introdução à programação
com Python por Nilo Ney Coutinho Menezes

Pago R\$ 73,00



● ○ ○

Learning Python
por Mark Lutz - Editora O'Reilly

Pago R\$ 270,01



● ○ ○

Lógica de Programação
e Algoritmos

🕒 3h Pago R\$ 29,99

BOAS PRÁTICAS DE PROGRAMAÇÃO



● ○ ○

DESIGN
PATTERNS

🕒 10min



● ○ ○

Design Patterns
Python

🕒 1h



● ○ ○

Design Patterns in Python
by Peter Ullrich

🕒 1h 

BIBLIOTECAS E APIS DE MACHINE LEARNING



● ○ ○

10 ferramentas e bibliotecas para trabalhar com data mining e Big Data

🕒 35min



● ● ○

Melhores bibliotecas de Machine e Deep Learning

🕒 40min



● ● ○

11 ferramentas open source para Machine Learning

🕒 20min



● ● ○

Machine learning APIs

🕒 5min 

TÉCNICAS DE REGRESSÃO

Aqui vão algumas perguntas para você pensar! Você sabe quais são os fatores mais importantes de um conjunto de dados? E quais podem ser ignorados?

Para enriquecer seu estudo, separamos as melhores técnicas utilizadas em Análise de Dados e realizamos uma completa seleção desde a Introdução aos Conceitos de Análise de Regressão até a Avaliações de Performance de Modelos.

ANÁLISE DE REGRESSÃO



● ● ●
Análise de Regressão

🕒 10min

LINEAR



● ● ●
Regressão Linear Simples

🕒 10min

MÚLTIPLA



Regressão Linear Múltipla

🕒 15min

RIDGE (E LASSO)



Métodos de Shrinkage
Pg. 16

🕒 30min

SELEÇÃO DE VARIÁVEIS



Seleção de Variáveis

🕒 25min



Comparing model selection
techniques for linear regression:
LASSO and Autometrics

🕒 1h 



Variable selection

🕒 1h 

AVALIAÇÃO DE PERFORMANCE DE MODELOS



Performance Evaluation of
Machine Learning Models

🕒 1h 🇬🇧



Model Evaluation -
Classification

🕒 30min 🇬🇧



Model Evaluation -
Regression

🕒 25min 🇬🇧

TÉCNICAS DE CLASSIFICAÇÃO

A criação de modelos que permitam a predição de valores é essencial para um Cientista de Dados. Neste tópico, você se aprofundará em mais algumas Técnicas Estatísticas de Regressão, irá entender o Conceito de Árvore de Decisão e o Naive Bayes (Algoritmo de Classificação).

REGRESSÃO LOGÍSTICA



Regressão Logística

🕒 15min



Regressão Linear Simples

🕒 1h



Regressão Linear Múltipla

🕒 15min



Regressão Logística

🕒 11min

ÁRVORES DE DECISÃO



Árvore Binária de Busca

🕒 15min



Árvores de Decisão

🕒 1h

NAIVE BAYES



Algoritmo de classificação
Naive Bayes

🕒 40min




Naive Bayes - Georgia Tech -
Machine Learning

🕒 8min 



Naive Bayes Theorem | Introduction
to Naive Bayes Theorem | Machine
Learning Classification

🕒 10min 

TÉCNICAS DE GRUPAMENTO E RELACIONADAS

Nesta seção, vamos abordar os temas que são extremamente importantes para o dia a dia de um Cientista de Dados. Você poderá conhecer o Algoritmo K-Means, os Métodos sobre Agrupamento de Dados, a Análise de Cluster, o Algoritmo LDA e a Introdução ao Agrupamento Hierárquico.

K-MEANS



Entenda o Algoritmo K-Means

🕒 41min



K-Means com Python Parte 1/2 Exemplo Visual

🕒 15min



K-Means com Python Parte 2/2 Exemplo Visual

🕒 13min



StatQuest: K-means clustering

🕒 9min 

Algoritmos hierárquicos aglomerativos e divisivos



Métodos de Agrupamento
de Dados

🕒 1h



O que é análise de Cluster

🕒 35min




Introdução ao
agrupamento hierárquico

🕒 13min

LATENT DIRICHLET ALLOCATION




LDA Algorithm Description

🕒 10min 



Latent Dirichlet Allocation

🕒 50min 



Explorando bases de texto
com LDA Latent Dirichlet

🕒 15min

CURSOS E LIVROS

Para você que aprecia um estudo de forma linear, como uma trilha de aprendizado, separamos alguns Cursos e Livros para complementar no seu processo de aprendizagem.

Os cursos abordam de forma transversal os assuntos do Guia **DATA MASTERS**, mas indicamos que você também estude todos os conteúdos, pois assim você vai garantir a total compreensão dos temas.

Bons estudos!



● ○ ○

Introdução à Ciência de Dados versão 2.0

🕒 8h

- Introdução;
- O Que é Ciência de Dados?;
- Ciência de Dados e Big Data;
- Ciência de Dados e Estatística;
- Aprendizado de Máquina (Machine Learning);
- Aplicações da Ciência de Dados;
- Ciclo de Vida de Projetos de Data Science;
- Carreiras em Data Science;
- Como Se Tornar um Cientista de Dados;
- Avaliação.



● ● ●

Big Data Fundamentos 2.0

🕒 8h

- Introdução;
- O Que é Big Data?;
- Introdução ao Hadoop;
- Arquitetura Hadoop;
- Ecossistema Hadoop;
- Soluções Comerciais com Hadoop;
- Introdução ao Apache Spark;
- Bancos de Dados NoSQL;
- Como Iniciar um Projeto de Big Data?;
- Avaliação e Certificado de Conclusão.



● ● ●

Formação Cientista de Dados

🕒 376h **Pago R\$ 2.720,00**



● ● ●

Data Scientist

🕒 6meses **Pago R\$ 1.439,00**



● ● ●

Python Fundamentos para Análise de Dados

🕒 54h

- Introdução;
- Variáveis, Tipos e Estruturas de Dados;
- Loops, Condicionais, Métodos e Funções;
- Tratamento de Arquivos, Módulos, Pacotes e Funções Built-in;
- Orientação a Objetos;
- Manipulando Banco de Dados em Python;
- Desafio DAS;
- Módulos Python para Análise de Dados;
- Introdução à Análise de Dados com Python;
- Introdução ao TensorFlow;
- Introdução a Machine Learning com Python;
- Bônus - Introdução a Deep Learning;
- Desenvolvimento Web;
- Introdução ao Web Scraping;
- Avaliação Final.



An Introduction to Statistical
Learning: With Applications
in R. Springer



An Introduction to Statistical
Learning: With Applications
in R. Springer

Pago R\$ 289,71



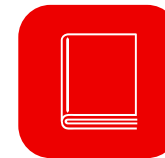
Python Para Análise de Dados:
Tratamento de Dados com
Pandas, NumPy e IPython

Pago R\$ 102,96



Probabilidade -
Aplicações à Estatística

Pago R\$ 200,76



Estatística Básica

Pago R\$ 104,90

