



Preditiva.ai

Certificamos para todos os fins que

Giovana Perin Campbell Penna

Concluiu o curso Data Science com carga horária de 185 horas.

Emitido em: 16/12/2025

Fernando Itano
Co-fundador

Jefferson Magalhães
Coordenador Educacional

CONTEÚDO PROGRAMÁTICO

Módulo 1: Estatística Descritiva

- Tabelas de Frequência
- Medidas Resumo (Média, Mediana, Quartis e Desvio Padrão)
- Gráficos analíticos (Histograma, Box-Plot, Pareto entre outros)
- Correlação de Pearson
- Coeficiente de Determinação R2
- Information Value (IV)

Módulo 2: Fundamentos de ETL para Análise de Dados

- Extração de Dados
- Transformação de Dados
- Carga de Dados

Módulo 3: Probabilidades

- Teoria Clássica de Probabilidade
- Probabilidade Freqüentista
- Distribuições de Probabilidade
- Análise Preditiva

Módulo 4: Introdução à Inferência Estatística

- Conceitos de população, amostra, parâmetros e estimadores
- Estimação pontual e por intervalo
- Margem de erro e tamanho de amostra

Módulo 5: Metodologia de Projetos de Dados

- CRISP-DM
- Entendimento do negócio
- Coleta dos Dados e Análise Exploratória
- Data Preparation e estudo analítico
- Validação e implantação da solução
- Cálculo do valor de um Projeto de Dados
- Técnicas de apresentação de resultados
- Mensurando o impacto gerado após a implantação

Módulo 6: Testes de Hipóteses

- Teste de Hipóteses para Média de 1 população
- Testes de Hipóteses para Média e Variância de 2 populações
- Teste de Hipóteses para Média de 2 populações pareadas
- Método Científico

Módulo 7: Regressão Linear Múltipla

- Introdução à modelagem estatística
- Variáveis Dummy
- Métricas de Ajuste

Módulo 8: SQL para Análise de Dados

- Agregando Dados
- Junção de Tabelas
- Fundamentos de BI
- Subquery, CTE, Tabelas Temporárias
- Window Functions
- Tratamento de Dados com SQL
- Criação e Edição de Banco de Dados

Módulo 9: Power BI para Análise de Dados

- Preparação de Dados no Power BI
- Tipos de gráficos
- Modelagem de Dados
- Linguagem DAX
- Filtros, Parâmetros e Campos Calculados
- Criação de Dashboards

Módulo 10: Storytelling com dados

- Como contar histórias
- Transformando análises em histórias

Módulo 11: Fundamentos de Cloud

- Introdução à Computação em Nuvem
- Armazenamento e Processamento de Dados na AWS
- Integração de Dados e Visualização na AWS

Módulo 12: Apache Spark e Databricks

- Noções de Big Data
- Processamento de Dados Distribuído e Spark
- Databricks
- Spark SQL

Módulo 13: Python para Análise de Dados

- Variáveis e Tipos de Dados
- Controle de Fluxo
- Funções e Classes
- Numpy
- Pandas
- Matplotlib

Módulo 14: Versionamento de código com Git e Github

- Configurando o ambiente com VENV e Poetry
- Versionamento com o Git
- Github

Módulo 15: Matemática para Data Science

- Funções
- Limites e Derivadas
- Álgebra linear
- Técnicas de Otimização

Módulo 16: Preparação de dados com Python

- Valores faltantes (missing values)
- Dados duplicados
- Conversão de tipos de dados
- Transformações de Colunas e Linhas

Módulo 17: Aprendizado Supervisionado

- Regressão Linear Múltipla
- Regressão Logística
- Árvores de Decisão
- Redes Neurais Artificiais
- Ensemble de Modelos

Módulo 18: Diagnóstico de Modelos

- Métricas de Classificação
- Precision, Recall e F1 Score
- Curva RoC e AuRoC
- Cross Validation
- Técnicas de regularização
- Shap-Values e Interpretação de modelos

Módulo 19: Aprendizado Não Supervisionado

- Clusterização: Kmeans e DBSCAN
- Redução de dimensionamento: PCA e t-SNE e AutoEncoder
- Detecção de Anomalias: Isolation Forest e AutoEncoder

Módulo 20: Otimização de Modelos

- Tuning de Hiperparâmetros
- Random Search e Grid Search
- Otimização Bayesiana
- Oversampling e Undersampling

Módulo 21: Deep Learning

- Processamento de Linguagem Natural (NLP)
- Visão Computacional