



## PLANO DE ENSINO

2º/2021

CURSO DE MBA EM BIG DATA, BUSINESS ANALYTICS E GESTÃO DE NEGÓCIOS /  
MBA EM BIG DATA E DATA SCIENCE APLICADO AO SETOR PÚBLICO

**DISCIPLINA:** Técnicas Avançadas de Captura e Tratamento de Dados

**PROFESSOR:** Bernardo Alves Furtado

**E-MAIL:** furtadobb@gmail.com

**CARGA HORÁRIA:** 21 h/a

**EMENTA:** Introdução a técnicas avançadas de captura e tratamento de dados. Conceito dados estruturados e não-estruturados. Captura básica. Captura por meio de API. Armazenamento. Análise e tratamento dados. Relatórios básicos.

### OBJETIVOS:

**A) Objetivo Geral:** Identificar dados estruturados e não-estruturados. Acessar e avaliar dados de várias fontes. Realizar persistência de dados e análise básica. Produzir relatórios.

### B) Objetivos Específicos:

- Compreender principais problemas e armadilhas de bases de dados.
- Realizar leitura de dados em arquivo e por captura estruturada e não-estruturada.
- Salvar e acessar arquivos. Noções básicas de formatos.
- Realizar tratamento e análise iniciais e produzir relatórios. Uso biblioteca pandas em Python.

### CONTEÚDO PROGRAMÁTICO E CRONOGRAMA:

**AULA 1 (3/8):** Introdução. Armadilhas. Aperitivo. Dados estruturados e não-estruturados. Exercícios.

**AULA 2 (5/8):** Formatos: tabulares, hierárquicos. Persistência. TXT, JSON, pickle, SQL

**AULA 3 (10/8):** Captura. Web. Api. Twitter. Exercícios. Exercícios

**AULA 4 (12/8):** Captura massiva. DOU. Datalake.

**AULA 5 (17/8):** Tratamento. Texto. Filtrando texto. NOÇÕES de REGEX. Regular expressions. `import re`.

**AULA 6 (19/8):** Tratamento. Pandas. Pandas. Exercícios.

**AULA 7 (21/8):** Captura, tratamento, análise e resultado. Exercício.

### BIBLIOGRAFIA BÁSICA:

**1. MERTZ, David. Cleaning Data for Effective Data Science.** Birmingham, UK: Packt Publishing Ltd., 2021.



2. MOLIN, Stephanie. **Hands-On Data Analysis with Pandas: Efficiently perform data collection, wrangling, analysis, and visualization using Python**. Packt Publishing Ltd, 2019.

3. JARMUL, K.; LAWSON, R. **Python Web Scraping**. 2. ed. Birmingham, UK: Packt Publishing Ltd, 2017.

**BIBLIOGRAFIA COMPLEMENTAR:**

1. SHIKIDA, C. D.; MONASTERIO, L.; NERY, P. F. **Guia brasileiro de análise de dados: armadilhas & soluções**. Brasília, DF: ENAP, 2021.

2. CIELEN, D.; MEYSMAN, A. **Introducing data science: big data, machine learning, and more, using Python tools**. [s.l.] Simon and Schuster, 2016.

3. MCKINNEY, W. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython**. 1 edition ed. Beijing: O'Reilly Media, 2012.

4. MITCHELL, R. **Web scraping with Python: Collecting more data from the modern web**. O'Reilly Media, Inc., 2019. Edição em português.