

Análise de dados em



Projeto

Agenda

- **Projeto**
- **Datasets**
- **Input/output CSV**

Projeto

Grupos	2 a 3 pessoas
Como vai ser o projeto?	Cada semana, vão aplicar os conhecimentos que aprenderam nas aulas anteriores
Apresentação	Cada sexta, 2 em 2 semanas, começamos a aula com apresentação de 10 min. do trabalho realizado

Plano

29 Maio - 2 Junho	Profiling EDA <ul style="list-style-type: none">• Descriptive Statistics
5 Junho - 9 Junho	EDA <ul style="list-style-type: none">• Visualization• Test hypothesis• Correlation
9 Junho	Apresentações
12 Junho - 16 Junho	Clustering
19 Junho - 30 Junho	Classificação/Regressão
30 Junho	Apresentações

Datasets

IMDB - <https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>

Diabetes - <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>

COVID - <https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.csv>

Monkeypox - <https://github.com/owid/monkeypox/blob/main/owid-monkeypox-data.csv>

Energy - <https://github.com/owid/energy-data/blob/master/owid-energy-data.csv>

Cancer - <https://www.kaggle.com/datasets/erdemtaha/cancer-data>

Breast cancer - <https://www.kaggle.com/datasets/utkarshx27/breast-cancer-wisconsin-diagnostic-dataset>

Input/Output CSV

In Comma Separated Values File (CSV) the information is plain text with tabular format where each line is a row and the columns are separated by comma (can be other symbol like ; or tab)

	Description	Input	Output
read.csv("./myfile.csv", sep = ",")	Read CSV file data from your computer	First parameter - path to file name to read sep - Symbol that separates the columns	data.frame
write.csv(dataframe, "./myfile.csv", sep = ",")	Writes data from data.frame to a CSV file in your computer	First parameter - data.frame with data we want to write in file Second parameter - path to file name to write sep - Symbol that separates the columns	Boolean

Input/Output

Save and load R object in files

	Description	Input	Output
load("./myfile.RData")	Load R object previous saved	First parameter - path of the file with R object saved	Boolean The R object is added to global environment
save(<R object>, "./myfile.RData"))	Saves R objects like lists, data.frame, etc into a file	First parameter - R object we want to save in file Second parameter - path to file name to save	Boolean