

Segundo trabalho (27-11-2025)

Segue-se o segundo trabalho da segunda unidade. Para isso, serão proposto dois problemas e a **equipe deverá escolher um deles e responder ao problema proposto**. Use tudo o que você sabe até agora em CD. É possível construir um modelo que se ajuste ao problema. Seja criativo.

Observação: deve enviar um arquivo do Jupyter Notebook (envios em outros formatos não serão avaliados) com documentação completa do seu trabalho, bem como a interpretação de cada análise e gráfico, conclusões e referências utilizadas. o Arquivo jupyter notebook deve descrever o Grupo_xx e data de entregue. Exemplo: **Grupo_xx_27_11_2025.ipynb**

Deve apresentar seu trabalho. Em uma apresentação de 20 e 25 minutos.

ordem de apresentação dos grupos será feito em Sala,

PROBLEMA Nº 1

Uma empresa que oferece empréstimos a pessoas físicas precisa saber, com base nos dados de seus clientes, quais novos clientes poderão ou não acessar o crédito (empréstimo) solicitado.

Faça uma pergunta, por exemplo: Qual é a probabilidade de aprovação/risco de cada novo cliente, considerando suas informações cadastrais e financeiras? A partir daí, defino os objetivos para resolver o problema em CD.

O conjunto de dados para este problema é: empretamo.csv.

PROBLEMA Nº 2

Um executivo do setor imobiliário em Boston, EUA, solicitou à sua equipe de Cientistas de Dados a criação de um modelo de previsão de preços de imóveis com base em algumas características dos imóveis e sua localização, como: número de cômodos, distância entre grandes centros, patrimônio, idade, taxa de criminalidade, etc.

Faça uma pergunta, por exemplo: Quanto vale um imóvel residencial em Boston (preço de mercado), considerando seu conjunto de características e localização? A partir daí, defina os objetivos para resolver o problema em CD.

O conjunto de dados para este problema é: data_imovel.csv.

Anexo do Dicionário de Dados

Dicionário de Dados — emprestamo.csv

Nome da coluna	Tipo de dado	Descrição	Valores possíveis / Observações	Exemplo de valor	% Faltantes
Loan_ID	Texto	Identificador único do pedido de empréstimo.		LP001002	0.0%
Gender	Texto	Gênero do(a) requerente.	Female, Male Categorias usuais: Male, Female.	Male	2.12%
Married	Texto	Estado civil do(a) requerente.	No, Yes Categorias usuais: Yes, No.	No	0.49%
Dependents	Texto	Número de dependentes do(a) requerente.	0, 1, 2, 3+ Categorias usuais: 0, 1, 2, 3+.	0	2.44%
Education	Texto	Escolaridade do(a) requerente.	Graduate, Not Graduate Categorias usuais: Graduate, Not Graduate.	Graduate	0.0%
Self_Employed	Texto	Indica se o(a) requerente é trabalhador(a) autônomo(a).	No, Yes Categorias usuais: Yes, No.	No	5.21%
ApplicantIncome	Inteiro	Renda do(a) requerente (geralmente mensal).		5849	0.0%
CoapplicantIncome	Decimal	Renda do(a) co-requerente (se houver), geralmente mensal.		0.000	0.0%
LoanAmount	Decimal	Valor do empréstimo solicitado/aprovado (usualmente em milhares de unidades monetárias).	Atenção: em versões deste dataset, o valor pode estar em milhares (ex.: 128 ≈ 128 mil).	128.000	3.58%
Loan_Amount_Term	Decimal	Prazo do empréstimo (em meses).	12.0, 120.0, 180.0, 240.0, 300.0, 36.0, 360.0, 480.0, 60.0, 84.0 Valores comuns: 360, 180, 120, 84, 60, 36, 240, 300, 480 (meses).	360.000	2.28%

Credit_History	Decimal	Indicador de histórico de crédito conforme regra do provedor (1=atende, 0=não atende).	0.0, 1.0 Forte preditor de aprovação; 1.0=histórico adequado, 0.0=histórico inadequado.	1.000	8.14%
Property_Area	Texto	Zona/área de residência do(a) requerente.	Rural, Semiurban, Urban Categorias usuais: Urban, Semiurban, Rural.	Urban	0.0%
Loan_Status	Texto	Status final do pedido de empréstimo (Y=aprovado, N=não aprovado).	N, Y Variável alvo típica em modelos de aprovação; Y=aprovado, N=não aprovado.	Y	0.0%

Dicionário de Dados — data_Imoveis.csv

Variável	Tipo	Descrição (PT)	Unidade / Observações
CRIM	Numérica (contínua)	Taxa de criminalidade per capita por cidade.	Taxa per capita (escala original do conjunto Boston).
ZN	Numérica (contínua)	Proporção de terrenos residenciais zoneados para lotes acima de 25.000 pés ² .	Proporção (0–100 ou 0–1 conforme versão do dataset).
INDUS	Numérica (contínua)	Proporção de acres destinados a atividades não varejistas por cidade.	Proporção (percentual).
CHAS	Binária (0/1)	Variável indicadora do Rio Charles (=1 se o setor é lindeiro ao rio; 0 caso contrário).	Indicador dummy.
NOX	Numérica (contínua)	Concentração de óxidos de nitrogênio.	Partes por 10 milhões (ppm $\times 10^{-5}$).
RM	Numérica (contínua)	Número médio de cômodos por residência.	Média de cômodos.
AGE	Numérica (contínua)	Proporção de unidades ocupadas pelos proprietários construídas antes de 1940.	Proporção (percentual).
DIS	Numérica (contínua)	Distâncias ponderadas para cinco centros de emprego de Boston.	Distância ponderada (unidade adimensional própria do dataset).
RAD	Numérica (discreta – índice)	Índice de acessibilidade a rodovias radiais.	Índice (valores inteiros).
TAX	Numérica (contínua)	Alíquota do imposto predial para valor total por US\$10.000.	Taxa por US\$10.000.
PTRATIO	Numérica (contínua)	Relação aluno/professor por cidade.	Razão (alunos por professor).
B	Numérica (contínua – transformada)	Métrica transformada: $1000 \times (B_k - 0,63)^2$, em que B_k é a proporção de população negra por cidade.	Transformação específica do dataset (adimensional).
LSTAT	Numérica (contínua)	Percentual da população de menor status socioeconômico.	Percentual (%).
MEDV	Numérica (contínua – alvo em muitos estudos)	Valor mediano das residências ocupadas pelos proprietários (em milhares de US\$).	US\$ milhares (ex.: 21.2 = US\$21.200).