================================================================================

RQ1: CATEGORIAS PREVALENTES DE SATD EM NOTEBOOKS DO KAGGLE

================================================================================

📊 Distribuição de Tipos de SATD:

------------------------------------------------------------

Tipo de SATD Quantidade de Notebooks Percentual (%)

dataDependency 9 27.27

awareness 9 27.27

codeDependency 8 24.24

readability 6 18.18

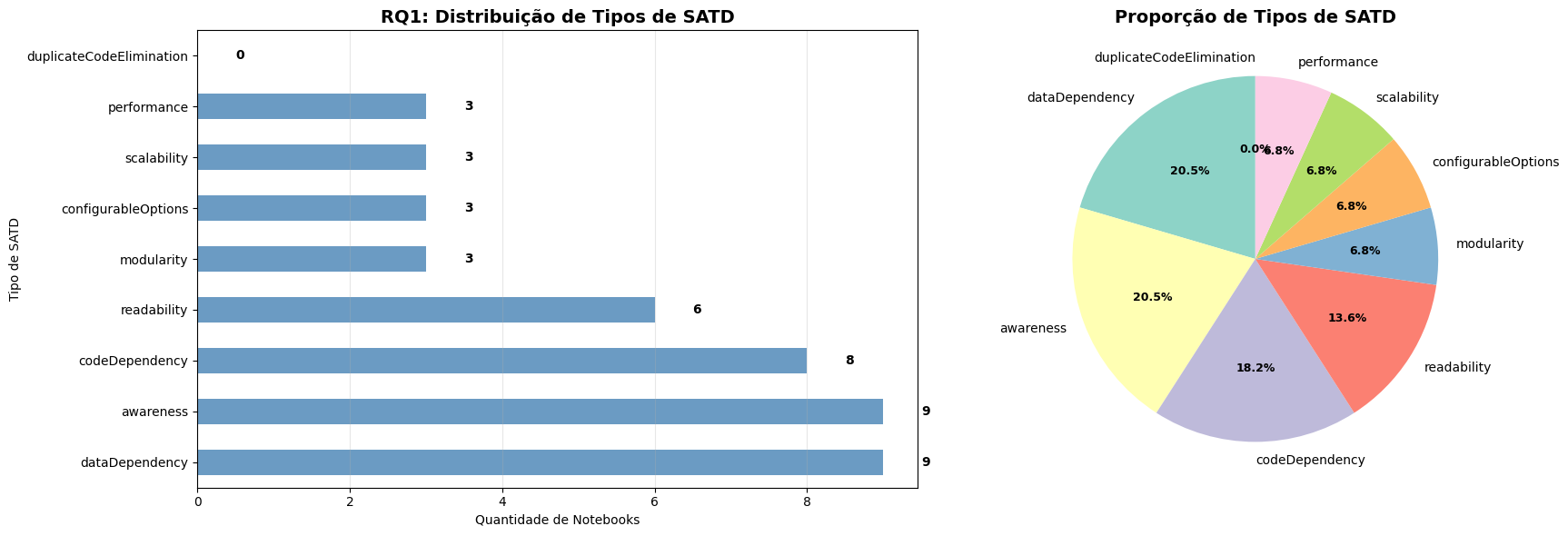
modularity 3 9.09

configurableOptions 3 9.09

scalability 3 9.09

performance 3 9.09

duplicateCodeElimination 0 0.00



================================================================================

RQ2: CORRELAÇÃO ENTRE SATD E MÉTRICAS DE SUCESSO (totalVotes)

================================================================================

📈 Estatísticas de Total Votes:

------------------------------------------------------------

count 33.000000

mean 14379.212121

std 74097.995245

min 8.000000

25% 25.000000

50% 101.000000

75% 345.000000

max 426522.000000

Name: totalVotes, dtype: float64

📊 Correlação entre Total SATDs e Total Votes: 0.2402

🔍 Correlação de cada tipo de SATD com Total Votes:

------------------------------------------------------------

dataDependency: 0.2775

codeDependency: -0.0951

awareness: -0.1176

modularity: -0.0588

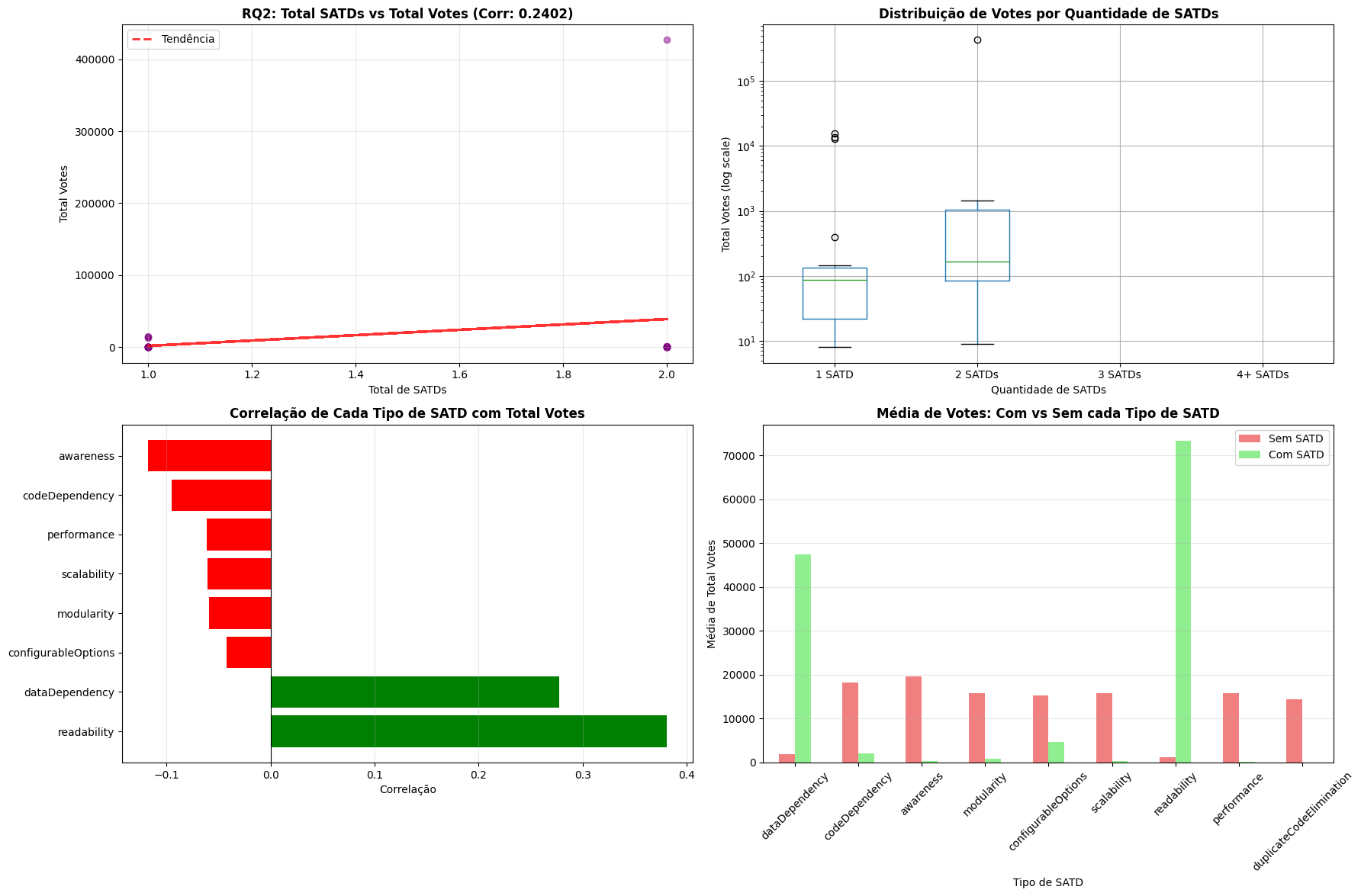
configurableOptions: -0.0422

scalability: -0.0605

readability: 0.3810

performance: -0.0616

duplicateCodeElimination: nan



📊 Análise de Total Votes por Presença de SATD:

------------------------------------------------------------

dataDependency:

Média COM SATD: 47450 votes

Média SEM SATD: 1978 votes

Diferença: +2299.5%

codeDependency:

Média COM SATD: 2109 votes

Média SEM SATD: 18306 votes

Diferença: -88.5%

awareness:

Média COM SATD: 363 votes

Média SEM SATD: 19635 votes

Diferença: -98.2%

modularity:

Média COM SATD: 809 votes

Média SEM SATD: 15736 votes

Diferença: -94.9%

configurableOptions:

Média COM SATD: 4631 votes

Média SEM SATD: 15354 votes

Diferença: -69.8%

scalability:

Média COM SATD: 418 votes

Média SEM SATD: 15775 votes

Diferença: -97.3%

readability:

Média COM SATD: 73360 votes

Média SEM SATD: 1272 votes

Diferença: +5665.6%

performance:

Média COM SATD: 155 votes

Média SEM SATD: 15802 votes

Diferença: -99.0%

================================================================================

RQ3: DISTRIBUIÇÃO DE CATEGORIAS DE SATD POR DOMÍNIO

================================================================================

🌍 Domínios encontrados: ['Dados Tabulares', 'NLP']

📊 Contagem de SATDs por Domínio:

--------------------------------------------------------------------------------

dataDependency codeDependency awareness modularity \

domain

Dados Tabulares 8 5 8 2

NLP 1 3 1 1

configurableOptions scalability readability performance \

domain

Dados Tabulares 2 2 6 2

NLP 1 1 0 1

duplicateCodeElimination

domain

Dados Tabulares 0

NLP 0

📊 Percentual de SATDs por Domínio (% dentro de cada domínio):

--------------------------------------------------------------------------------

dataDependency codeDependency awareness modularity \

domain

Dados Tabulares 22.86 14.29 22.86 5.71

NLP 11.11 33.33 11.11 11.11

configurableOptions scalability readability performance \

domain

Dados Tabulares 5.71 5.71 17.14 5.71

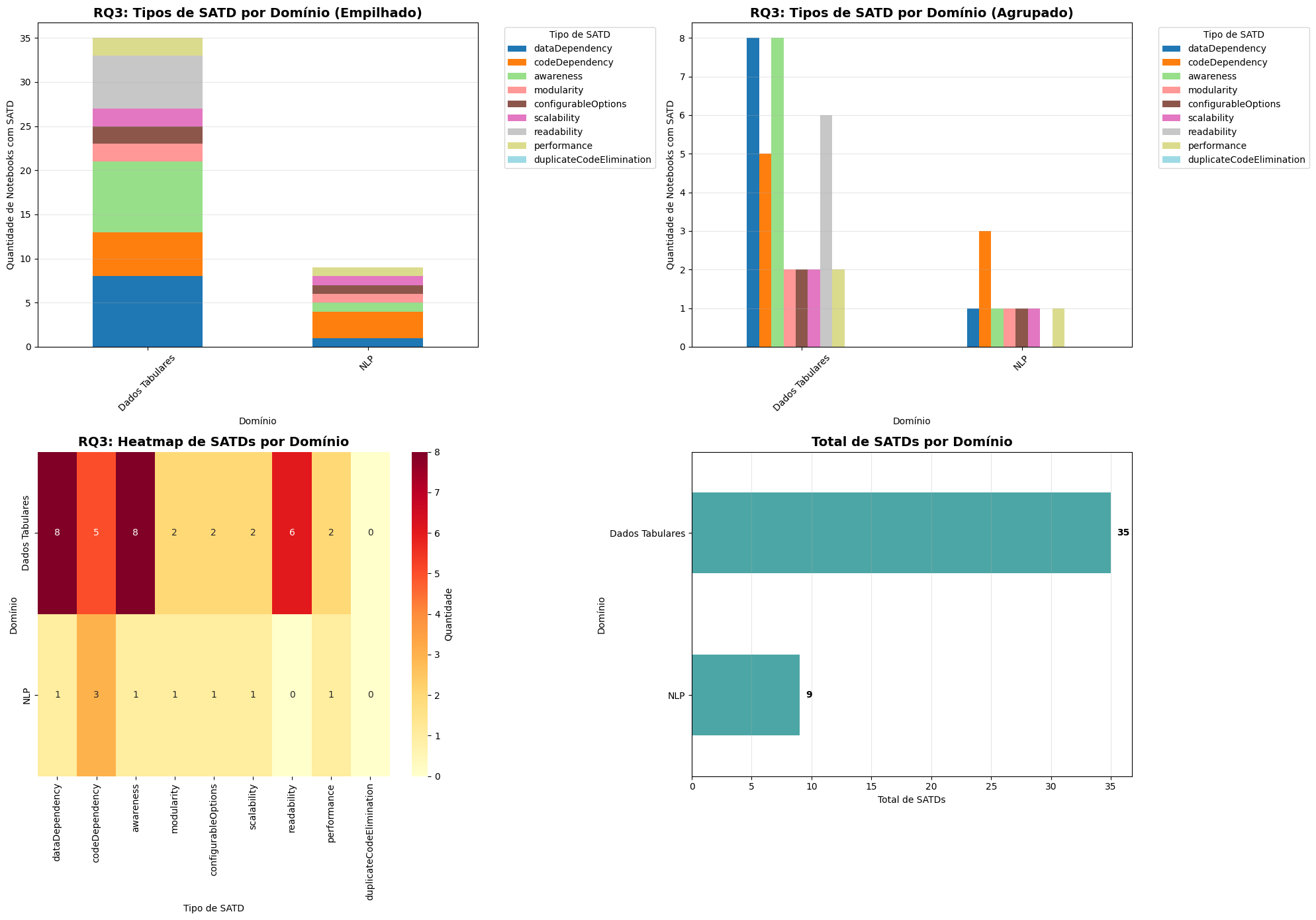
NLP 11.11 11.11 0.00 11.11

duplicateCodeElimination

domain

Dados Tabulares 0.0

NLP 0.0



📈 Estatísticas Detalhadas por Domínio:

--------------------------------------------------------------------------------

🔹 Dados Tabulares:

Notebooks com SATD: 26

Total de SATDs: 35

Média de SATDs por notebook: 1.35

Média de Votes: 18238

Tipo mais comum: dataDependency (8 notebooks)

🔹 NLP:

Notebooks com SATD: 7

Total de SATDs: 9

Média de SATDs por notebook: 1.29

Média de Votes: 46

Tipo mais comum: codeDependency (3 notebooks)

================================================================================

📋 SUMÁRIO EXECUTIVO

================================================================================

PRINCIPAIS ACHADOS:

1️⃣ RQ1 - Categorias Prevalentes:

• Total de notebooks COM SATD: 33

• Total de notebooks ANALISADOS: 3600

• Percentual com SATD: 0.9%

• Categoria mais prevalente: dataDependency

(9 notebooks, 27.27%)

• Total de SATDs identificados: 44

2️⃣ RQ2 - Relação com Sucesso (Total Votes):

• Correlação geral: 0.2402

• Tipo com maior correlação positiva: readability

(r=0.3810)

• Média de votes (notebooks COM SATD): 14379

• Média de votes (notebooks SEM SATD): 242

3️⃣ RQ3 - Distribuição por Domínio:

• Domínio com mais SATDs: Dados Tabulares (35 SATDs)

• Domínios analisados: 2

• Tipo de SATD dominante por domínio:

- Dados Tabulares: dataDependency (8 notebooks)

- NLP: codeDependency (3 notebooks)

================================================================================

✅ ANÁLISE CONCLUÍDA COM SUCESSO!

================================================================================

💾 Salvando resultados...

✓ resultados\_rq1\_final.csv

✓ resultados\_rq3\_dominio.csv

✓ resultados\_rq2\_correlacoes.csv

✓ notebooks\_com\_satd\_consolidado.csv

================================================================================

🏆 TOP 10 NOTEBOOKS COM MAIS VOTES E SUAS SATDs

================================================================================

title totalVotes total\_satds domain

Exercise: Arithmetic and Variables 426522 2 Dados Tabulares

Introduction to Ensembling/Stacking in Python 15380 1 Dados Tabulares

Exercise: Intro to AutoML 13769 1 Dados Tabulares

House Prices Prediction using TFDF 12946 1 Dados Tabulares

Pytanic 1458 2 Dados Tabulares

10 Simple hacks to speed up your Data Analysis 1221 2 Dados Tabulares

House Prices: EDA to ML (Beginner) 861 2 Dados Tabulares

Linear Regression from scratch (Gradient Descent) 399 1 Dados Tabulares

Titanic Survival: Seaborn and Ensembles 345 2 Dados Tabulares

HousePrice - end to end project 165 2 Dados Tabulares

================================================================================

📉 TOP 10 NOTEBOOKS COM MAIS SATDs

================================================================================

title totalVotes total\_satds domain

House Prices: EDA to ML (Beginner) 861 2 Dados Tabulares

HousePrice - end to end project 165 2 Dados Tabulares

Advanced Reg Techniques (Linear Models) Top 6% 101 2 Dados Tabulares

Housing prices prediction with scikit-learn 71 2 Dados Tabulares

Exercise: Arithmetic and Variables 426522 2 Dados Tabulares

Pytanic 1458 2 Dados Tabulares

10 Simple hacks to speed up your Data Analysis 1221 2 Dados Tabulares

Titanic Survival: Seaborn and Ensembles 345 2 Dados Tabulares

DNN Keras Santander 16 2 Dados Tabulares

📖 II.CommonLit: BERT vs RoBERTa + W&B testing 108 2 NLP