TEXT TO NUMERIC COLUMNS

WHY CHANGE TEXT VALUES?

Machine learning models need numbers as input.

Text is ambiguous, while numbers are precise.

Text will break ML models.

TEXT COLUMN

Here is an example of a text column.

Book

ebook

print

audio

ORDINAL COLUMN

A text column may be converted to an ordinal column as follows.

Book	Value
ebook	0
print	1
audio	2

There is an issue with this approach. Can you spot it?

ORDINAL COLUMN

Book	Value
ebook	0
print	1
audio	2

2 is closer to 1 than 0. But is an audio book more close to a print book than an ebook?

Is this any better? →

Book	Value
ebook	1
print	0
audio	2

ONE-HOT ENCODING

Called one-hot encoding. 1 indicates presence, 0 indicates absence.

Book	ebook	print	audio
ebook	1	0	0
print	0	1	0
audio	0	0	1

FINAL RESULT

One-hot encoded DataFrame.

ebook	print	audio
1	0	0
0	1	0
0	0	1

HAPPY CODING!