

INF100 – 09

filhåndtering, encoding, exceptions

Crystal Chang Din
David Grellscheid

File I/O

- `open(filename, mode)` mode: **r**ead, **w**rite, **a**ppend

```
with open("some.txt", "r") as f:  
    all_in_one = f.read()
```

```
with open("another.txt", "r") as f:  
    list_of_lines = f.readlines()
```

```
with open("third.txt", "r") as f:  
    for line in f:  
        one_line = line
```

File I/O

- `open(filename, mode)` mode: read, write, append

```
out = "Hello, how are you?"  
with open("hello.txt", "w") as f:  
    f.write(out)  
    f.write('\n')
```

```
in = 'inputdata.txt'  
out = 'reverse.txt'  
  
with open(in, 'r') as fi, open(out, 'w') as fo:  
    for line in fi:  
        outline = line[::-1] # reverse  
        fo.write(outline)
```

nyttige string-funksjoner

- `str.split(sep)` -> liste av ord, kan angi *sep*, f.eks. `'\n'` `','`

```
a = """Many  
different  
lines  
"""
```

```
'Many\ndifferent\nlines'
```

```
b = a.split() ['Many', 'different', 'lines']
```

- `'tekst'.join(liste)` -> bruker *tekst* som forbindelse

```
a = ['en', 'to', 'tre']
```

```
b = '=='.join(a)
```

```
'en==to==tre'
```

```
c = ''.join(a)
```

```
'entotre'
```

```
d = '\n'.join(a)
```

```
'en\nto\ntre'
```

```
"""en  
to  
tre"""
```


Koding / dekoding

Stud.kafp H-yte

07.03.2023



-47,50

Koding / dekoding

- Hvert bokstav kodert i et byte (8 bits):
bare $2^8 = 256$ muligheter
- ASCII bruker de første 128:

Hex	Value	Hex	Value	Hex	Value	Hex	Value	Hex	Value	Hex	Value	Hex	Value	Hex	Value
00	NUL	10	DLE	20	SP	30	0	40	@	50	P	60	`	70	p
01	SOH	11	DC1	21	!	31	1	41	A	51	Q	61	a	71	q
02	STX	12	DC2	22	"	32	2	42	B	52	R	62	b	72	r
03	ETX	13	DC3	23	#	33	3	43	C	53	S	63	c	73	s
04	EOT	14	DC4	24	\$	34	4	44	D	54	T	64	d	74	t
05	ENQ	15	NAK	25	%	35	5	45	E	55	U	65	e	75	u
06	ACK	16	SYN	26	&	36	6	46	F	56	V	66	f	76	v
07	BEL	17	ETB	27	'	37	7	47	G	57	W	67	g	77	w
08	BS	18	CAN	28	(	38	8	48	H	58	X	68	h	78	x
09	HT	19	EM	29	)	39	9	49	I	59	Y	69	i	79	y
0A	LF	1A	SUB	2A	*	3A	:	4A	J	5A	Z	6A	j	7A	z
0B	VT	1B	ESC	2B	+	3B	;	4B	K	5B	[	6B	k	7B	{
0C	FF	1C	FS	2C	,	3C	<	4C	L	5C	\	6C	l	7C	
0D	CR	1D	GS	2D	-	3D	=	4D	M	5D	]	6D	m	7D	}
0E	SO	1E	RS	2E	.	3E	>	4E	N	5E	^	6E	n	7E	~
0F	SI	1F	US	2F	/	3F	?	4F	O	5F	_	6F	o	7F	DEL

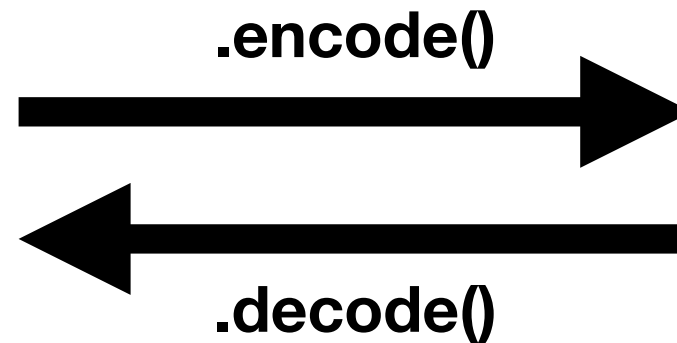
- Hva med Ø ä Ü Æ ? Bruk 128-255. Trenger mange encodings for ulike språk!

ISO/IEC 8859-1																
	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
0_0																
1_16																
2_32	SP 0020	! 0021	" 0022	# 0023	\$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
3_48	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
4_64	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
5_80	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
6_96	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
7_112	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	 007C	} 007D	~ 007E	
8_128																
9_144																
A_160	NBSP 00A0	¡ 00A1	¢ 00A2	£ 00A3	¤ 00A4	¥ 00A5	¦ 00A6	§ 00A7	¨ 00A8	© 00A9	ª 00AA	« 00AB	¬ 00AC	SHY 00AD	® 00AE	¯ 00AF
B_176	° 00B0	± 00B1	² 00B2	³ 00B3	´ 00B4	µ 00B5	¶ 00B6	· 00B7	¸ 00B8	¹ 00B9	º 00BA	» 00BB	¼ 00BC	½ 00BD	¾ 00BE	¿ 00BF
C_192	À 00C0	Á 00C1	Â 00C2	Ã 00C3	Ä 00C4	Å 00C5	Æ 00C6	Ç 00C7	È 00C8	É 00C9	Ê 00CA	Ë 00CB	Ì 00CC	Í 00CD	Î 00CE	Ï 00CF
D_208	Ð 00D0	Ñ 00D1	Ò 00D2	Ó 00D3	Ô 00D4	Õ 00D5	Ö 00D6	× 00D7	Ø 00D8	Ù 00D9	Ú 00DA	Û 00DB	Ü 00DC	Ý 00DD	Þ 00DE	ß 00DF
E_224	à 00E0	á 00E1	â 00E2	ã 00E3	ä 00E4	å 00E5	æ 00E6	ç 00E7	è 00E8	é 00E9	ê 00EA	ë 00EB	ì 00EC	í 00ED	î 00EE	ï 00EF
F_240	ð 00F0	ñ 00F1	ò 00F2	ó 00F3	ô 00F4	õ 00F5	ö 00F6	÷ 00F7	ø 00F8	ù 00F9	ú 00FA	û 00FB	ü 00FC	ý 00FD	þ 00FE	ÿ 00FF

ISO/IEC 8859-2 (Latin-2)																
	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
0_0																
1_16																
2_32	SP 0020	! 0021	" 0022	# 0023	\$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
3_48	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
4_64	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
5_80	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
6_96	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
7_112	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	 007C	} 007D	~ 007E	
8_128																
9_144																
A_160	NBSP 00A0	Ą 0104	Ć 02D8	Ł 0141	Ń 00A4	Ŕ 013D	Ś 015A	Ş 00A7	Š 00A8	Ž 0160	Ž 015E	Ţ 0164	Ž 0179	SHY 00AD	Ž 017D	Ž 017B
B_176	° 00B0	ą 0105	ć 02DB	ł 0142	ń 00B4	ŕ 013E	ś 015B	ş 02C7	š 00B8	ź 0161	ž 015F	ţ 0165	ž 017A	~ 02DD	ž 017E	ž 017C
C_192	Ř 0154	Á 00C1	Â 00C2	Ă 0102	Ä 00C4	Ĺ 0139	Ć 0106	Ç 00C7	Č 010C	É 00C9	Ę 0118	Ě 00CB	Ě 011A	Í 00CD	Î 00CE	Ď 010E
D_208	Đ 0110	Ń 0143	Ň 0147	Ó 00D3	Ô 00D4	Õ 0150	Ö 00D6	× 00D7	Ř 0158	Ů 016E	Ú 00DA	Ů 0170	Ü 00DC	Ý 00DD	Ť 0162	ß 00DF
E_224	ř 0155	á 00E1	â 00E2	ă 0103	ä 00E4	Í 013A	ć 0107	ç 00E7	č 010D	é 00E9	ę 0119	ě 00EB	ě 011B	í 00ED	î 00EE	ď 010F
F_240	đ 0111	ń 0144	ň 0148	ó 00F3	ô 00F4	õ 0151	ö 00F6	÷ 00F7	ř 0159	ů 016F	ú 00FA	ů 0171	ü 00FC	ý 00FD	ť 0163	· 02D9

File I/O: encoding

abstrakt representasjon
i Python-programmer



sekvens av bytes
for input/output
til filer, nett, osv.

str	encoding	bytes
'abc'	'ascii'	b'\x61\x62\x63'
'abc'	'iso8859-1'	b'\x61\x62\x63'
'abc'	'cp1252'	b'\x61\x62\x63'
'ø'	'ascii'	ERROR
'ø'	'iso8859-1'	b'\xf8'
'ø'	'cp1252'	b'\xf8'
'ř'	'iso8859-2'	b'\xf8'

File I/O: Unicode

- Unicodes mål:
representere ~alle symboler fra alle alfabeter
- Må bruke flere enn 1 byte i koding

str	encoding	bytes
'ø'	'iso8859-1'	b'\xf8'
'ř'	'iso8859-2'	b'\xf8'
'ø'	'utf8'	b'\xc3\xb8'
'ř'	'utf8'	b'\xc5\x99'
'Ã, '	'cp1252'	b'\xc3\xb8'
'Å™'	'cp1252'	b'\xc5\x99'
'酶'	'gb2312'	b'\xc3\xb8'
'好'	'utf8'	b'\xe5\xa5xbd'
'😊'	'utf8'	b'\xf0\x9f\x98\x8a'
'\u2615' == '☕'	'utf8'	b'\xe2\x98\x95'

abstrakt unicode
posisjon u2615

ord()
chr()

UTF8-koding
e2-98-95

I/O encoding

- Encoding er viktig å huske for input/output
- 'utf8' er *default* i de fleste Python-installasjonene (ikke på alle Windows ennå)
- kan være eksplisitt i **open()** ved bruk av **encoding=**'

```
msg = """\
سلام ، چطوری؟  
很好。  
"""
```

```
with open("hello.txt", "w", encoding="utf8") as f:  
    f.write(msg)
```

```
x = "سلام ، چطورید؟"
```

```
y = "很好"
```

```
五 = 5
```

```
اسم = "大卫"
```

```
print(x)
```

```
print(y)
```

```
print(五)
```

```
print(اسم)
```

```
文 = str
```

```
چاپ = print
```

```
چاپ(文(五)*五)
```

Unntak (Exceptions)

Alle feilmeldinger i Python er *Exceptions*

```
x = 2.0 / 0.0
```

```
ZeroDivisionError: float division by zero
```

```
a = [9,12]  
a[3]
```

```
IndexError: list index out of range
```

```
int('blob')
```

```
ValueError: invalid literal for int()  
with base 10: 'blob'
```

```
d = {'a':7}  
d['b']
```

```
KeyError: 'b'
```

```
open('foo.txt')
```

```
FileNotFoundError: [Errno 2] No such file or  
directory: 'foo.txt'
```

Exception handling

Vi kan behandle *exceptions* med `try: ... except: ...`

```
print('vi begynner')  
x = 2.0 / 0.0  
print('x er lik', x)
```

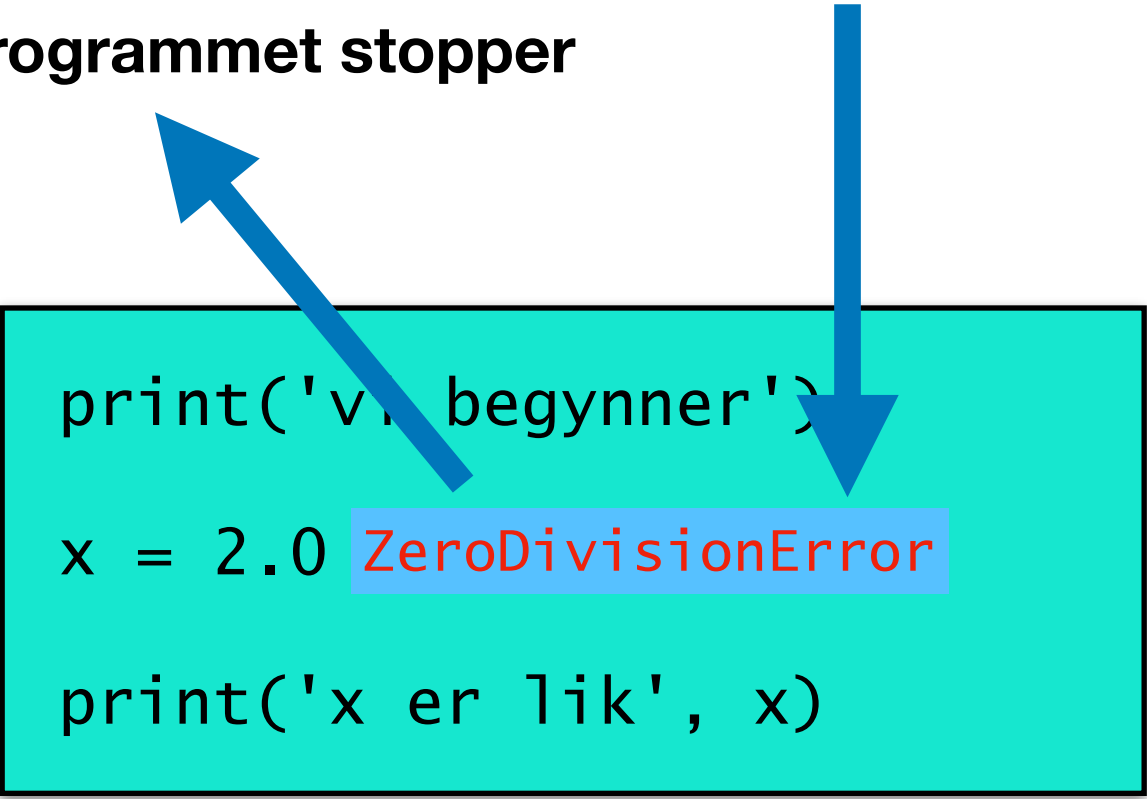
```
vi begynner  
...  
ZeroDivisionError: float division by zero
```

```
print('vi begynner')  
try:  
    x = 2.0 / 0.0  
    print('x er lik', x)  
except ZeroDivisionError:  
    print('Sorry!')  
  
print('vi går videre')
```

```
vi begynner  
Sorry!  
vi går videre
```

Exception handling

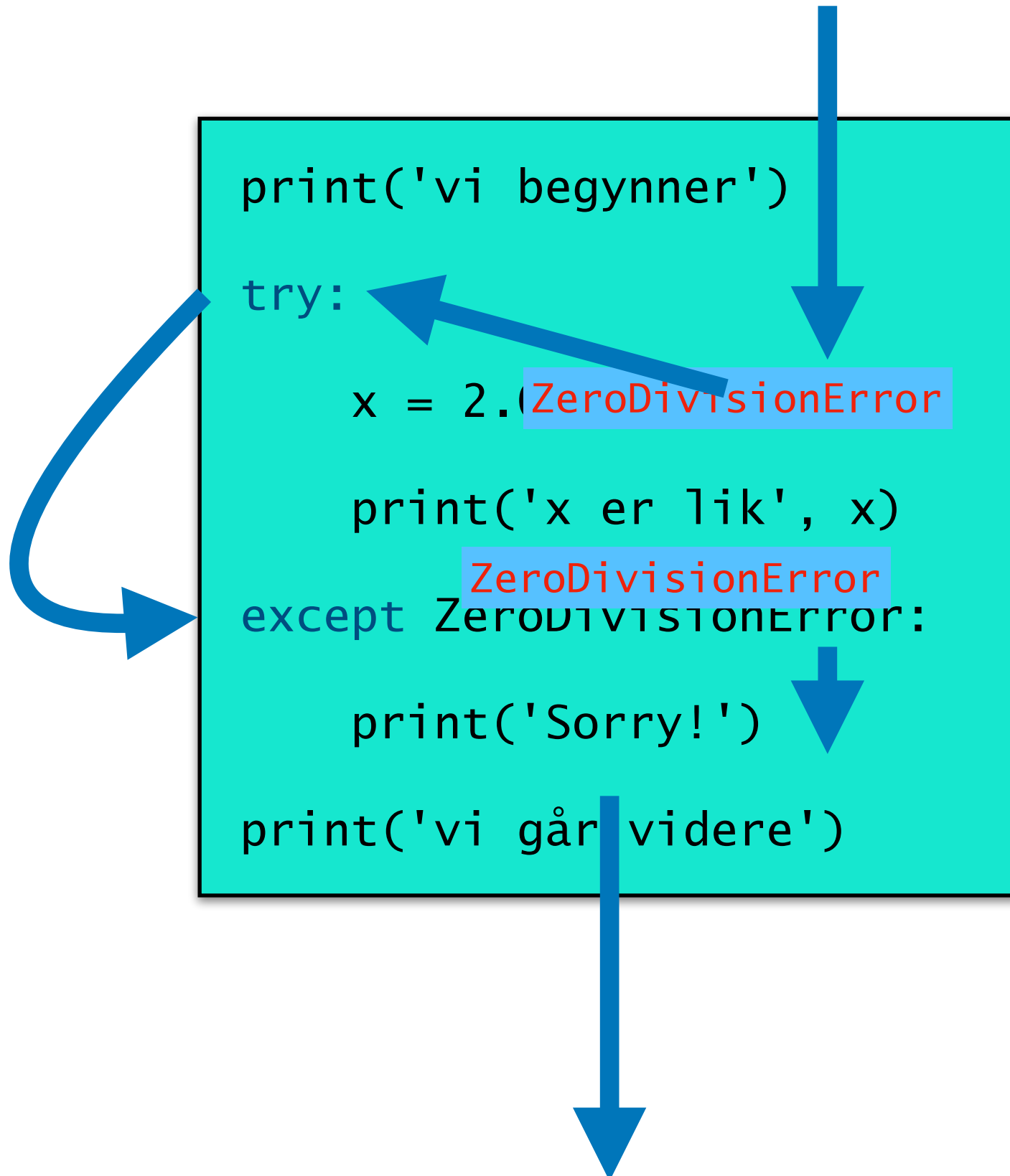
Feilmelding, programmet stopper



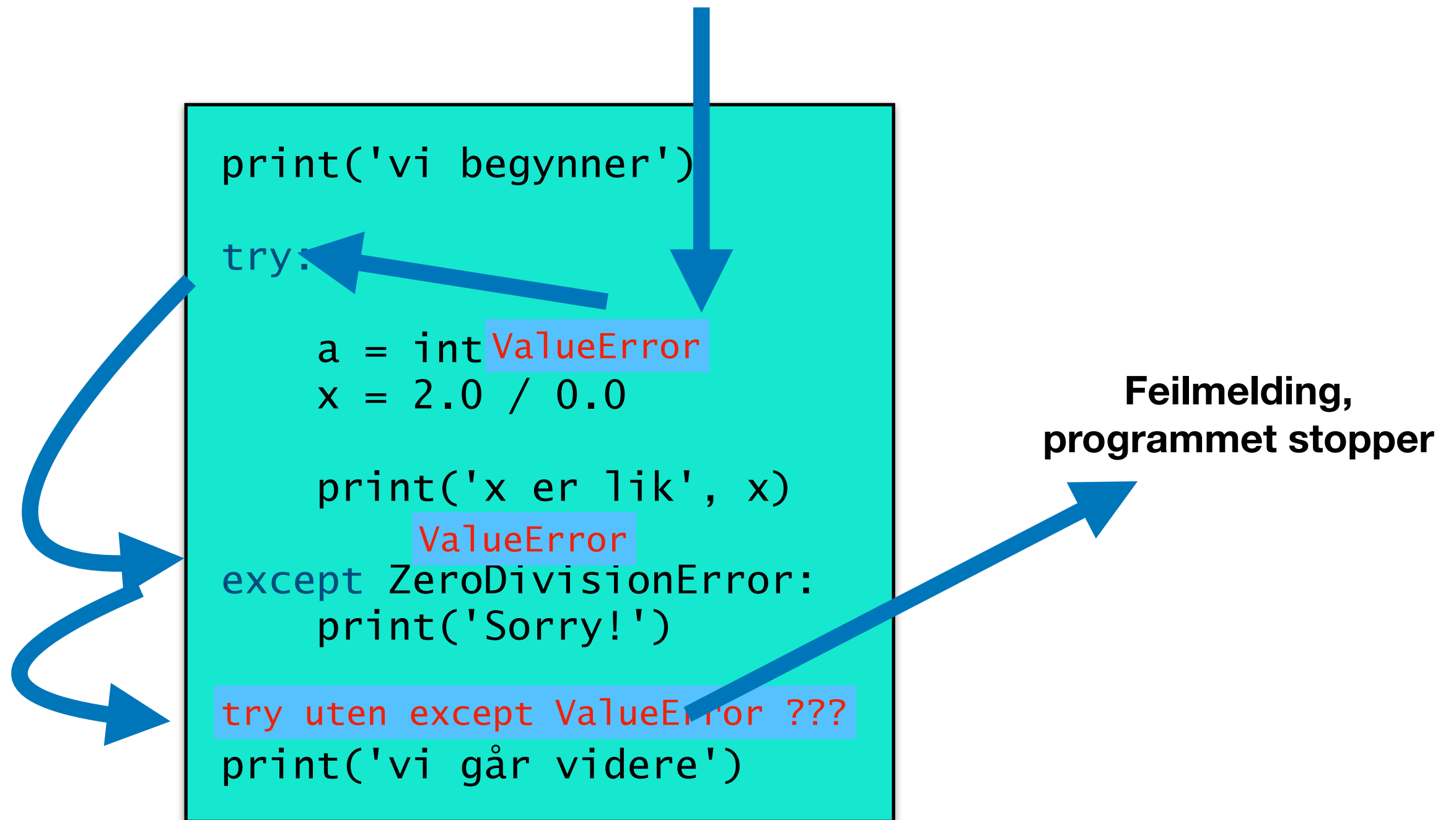
The diagram illustrates an exception handling process. A cyan rectangular box contains three lines of Python code. The second line, `x = 2.0 ZeroDivisionError`, is highlighted with a light blue background. A red text label `ZeroDivisionError` is positioned to the right of the `x = 2.0` assignment. A blue arrow points vertically downwards from above the box to the `ZeroDivisionError` label. Another blue arrow points diagonally upwards and to the left from the `ZeroDivisionError` label towards the text 'Feilmelding, programmet stopper' located above and to the left of the box.

```
print('vi begynner')  
x = 2.0 ZeroDivisionError  
print('x er lik', x)
```


Exception handling



Exception handling



Exception handling

Kan bruke flere *except*:

```
print('vi begynner')

try:
    a = int('hei')
    x = 2.0 / 0.0
    print('x er lik', x)

except ZeroDivisionError:
    print('Sorry!')

except ValueError:
    print('Also sorry')

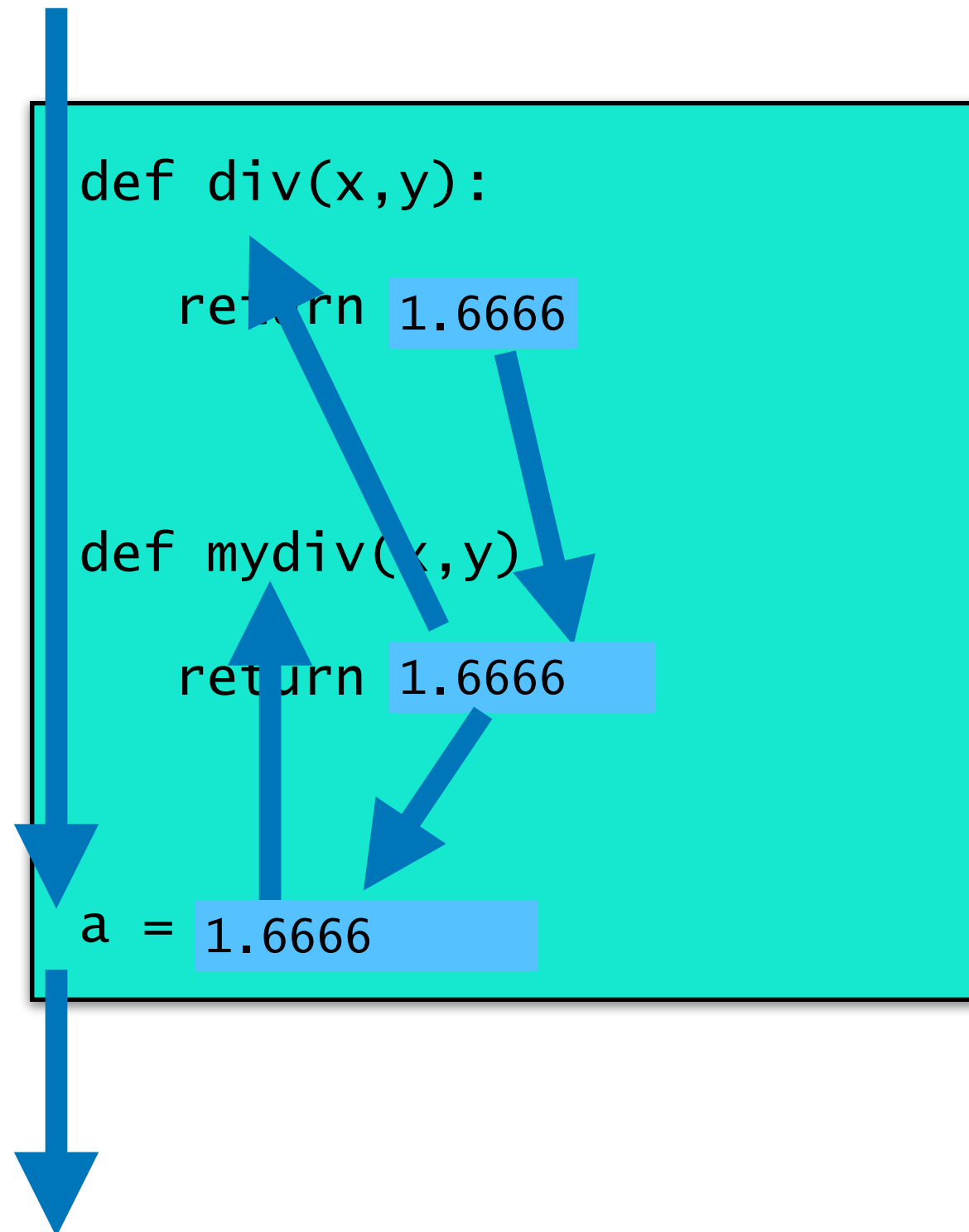
except: # alle andre error
    print('Huh?')

print('vi går videre')
```

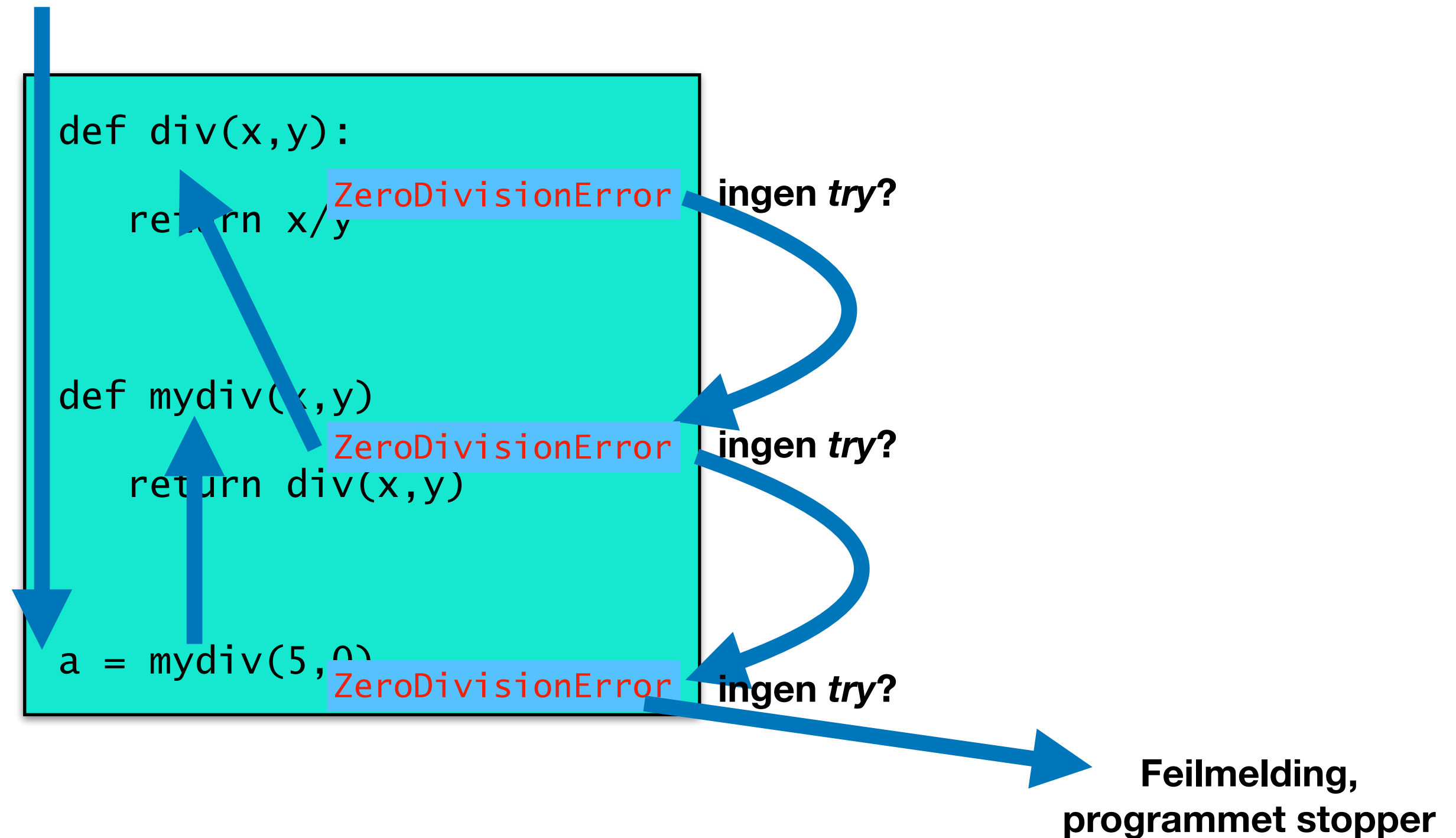
```
vi begynner
Also sorry
vi går videre
```

Exception i funksjoner

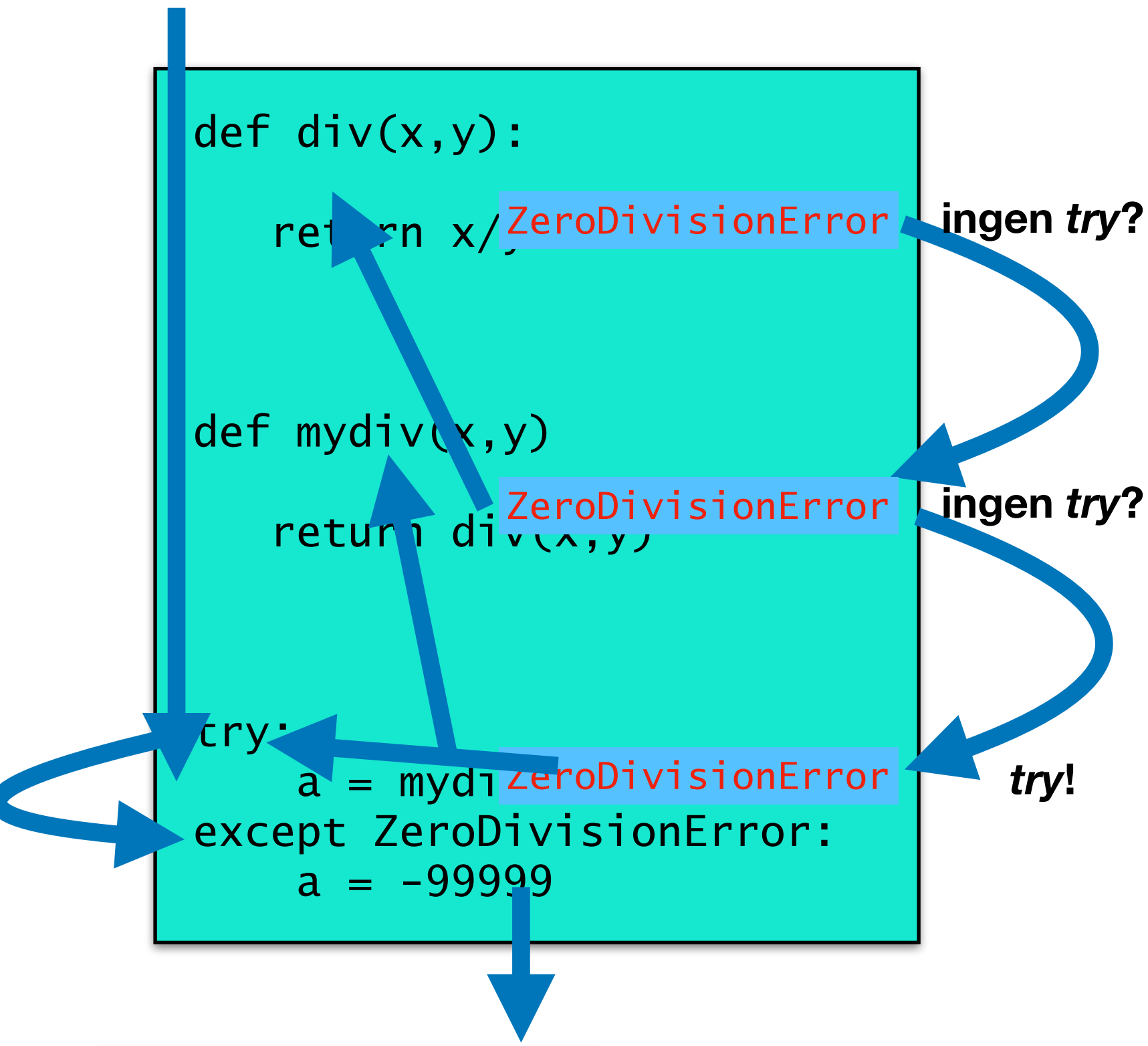
vanlig kjøring



Exception i funksjoner



Exception i funksjoner



Exception i funksjoner

```
def div(x,y):  
    return x/y
```

```
def mydiv(x,y)  
    return div(x,y)
```

```
try:  
    a = mydiv(5,0)  
except ZeroDivisionError:  
    a = -99999
```

problemet oppstår her
men kan ikke løses her
i en generell funksjon

Bruk *try-except* her hvor du har
en mulighet for å løse problemet

Hvis det ikke finnes en god løsning,
bedre å la *Exception* gå videre!