Spreadsheet Does What?

James Abel
www.abel.co
https://github.com/jamesabel/spreadsheetdoeswhat

An homage to https://www.pythondoeswhat.com

CSV/Spreadsheets are the Worlds Database

Me:

Please give me access the data in your database

Them (email):

Of course! Please see the attached CSV file!

Party on! \m/

Survey

I  Spreadsheets

?

Survey

I  Spreadsheets

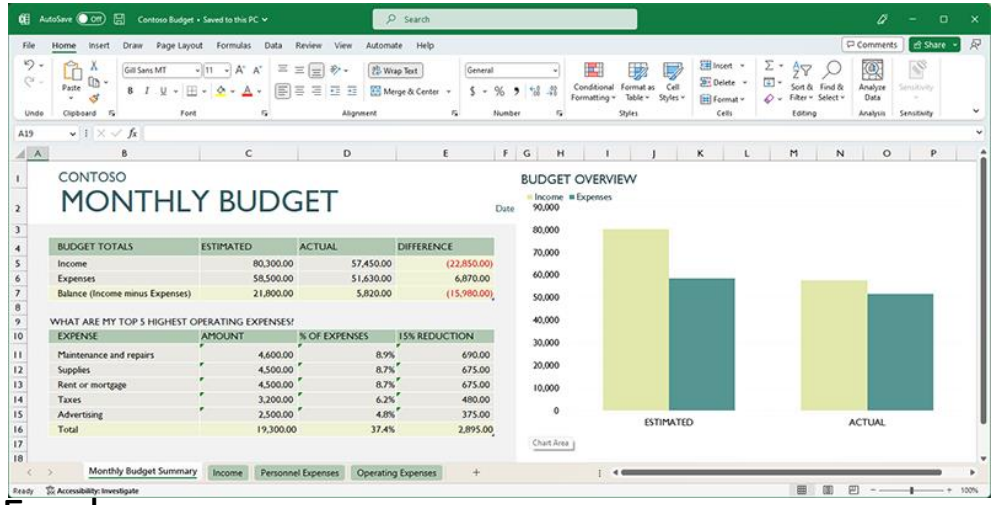
?


They're the same picture.

A Little History

- CSVs - 1972
- VisiCalc (Apple II) - 1979
- Excel - Mac 1985, Windows 1987
  - Python support - 2023
- CSV RFC 4180 - 2005

VisiCalc



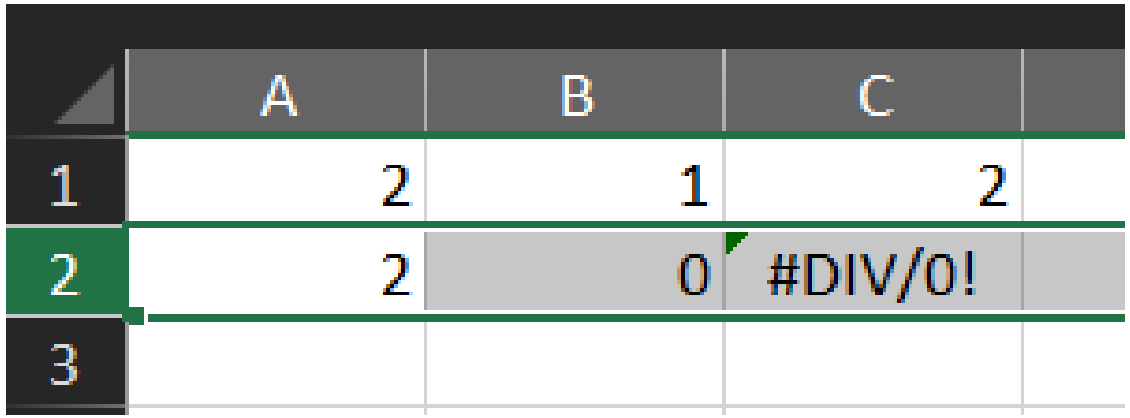
Excel

Python and Spreadsheet/Excel

- Python csv module (built-in)
- Packages include:
  - openpyxl
  - Pandas (openpyxl default engine)
  - xlrd

Tales From The Field
or
Do I Really Need to Validate My Data?

```
wb = open_workbook('totally_normal_spreadsheet.xls')
ws = wb.sheet_by_index(0)
print("xlrd:")
for column in (0, 1, 2):
    value = ws.cell(1, column).value
    print(value, type(value))
```



	A	B	C
1	2	1	2
2	2	0	#DIV/0!
3			

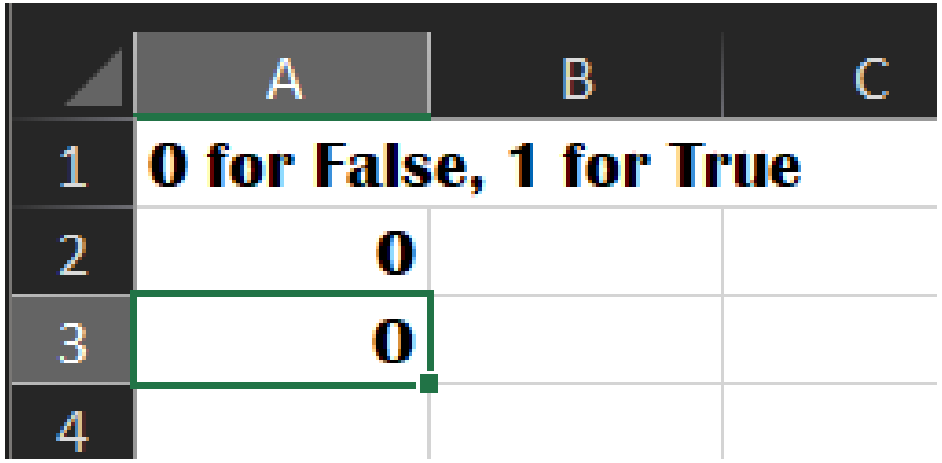
```
xlrd:
2.0 <class 'float'>
0.0 <class 'float'>
7 <class 'int'>
```

```
openpyxl:
2 <class 'int'>
0 <class 'int'>
#DIV/0! <class 'str'>
```

Tales From The Field

```
wb = load_workbook('o_no.xlsx', data_only=True)
ws = wb.active
a = ws["A2"].value
b = ws["A3"].value
a_bool = bool(a)
b_bool = bool(b)
print("openpyxl")
print("a", a, type(a), a_bool)
print("b", b, type(b), b_bool)
```

```
openpyxl
a 0 <class 'int'> False
b 0 <class 'str'> True
```



A2 is a number (proper)
A3 is a string (improper)

`tobool` to deal with Python's truthiness

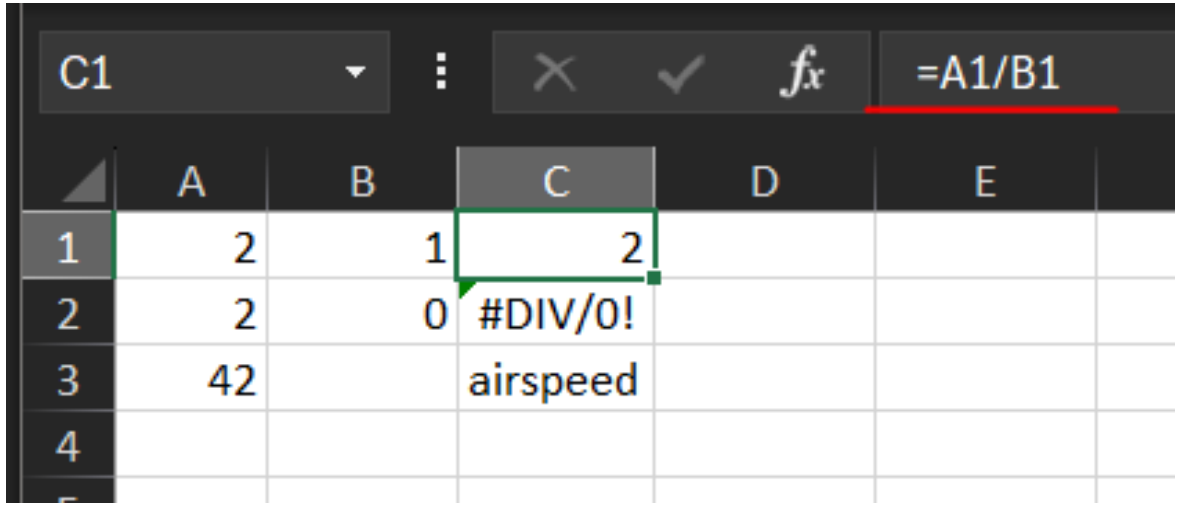
```
pip install tobool

from tobool import to_bool

bool(False)
False
bool("False")
True
to_bool("False")
False
```

(not only a CSV/Spreadsheet issue)

Another example

```
import pandas as pd
df = pd.read_excel('pandas_example.xlsx')
print(df)
```

```
    2    1          2.1
0   2  0.0          NaN
1  42  NaN  airspeed
```



(apparently a header issue)

Summary

Gorge of Eternal Peril

- Always check data formats and types
  - Example: pydantic
- Always sanity check answers
- Always check calculations done in spreadsheets
- Always check for irregular values such as NaNs
  - Be careful when choosing default ‘NA’s and associated filters

https://github.com/jamesabel/spreadsheetdoeswhat