

# Spreadsheet Does What?

James Abel

[www.abel.co](http://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*



**Spreadsheets**

**?**



**They're the same picture.**

# A Little History

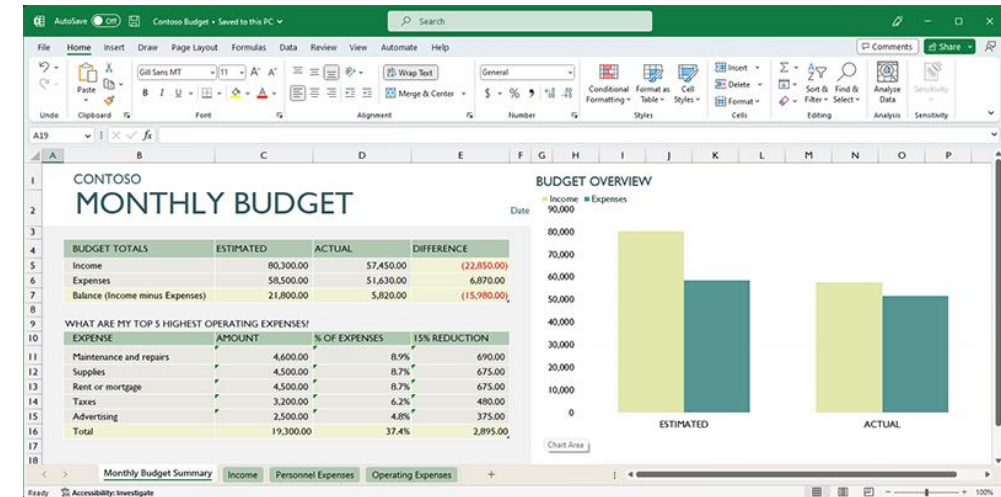
- CSVs - 1972
- VisiCalc (Apple II) - 1979
- Excel - Mac 1985, Windows 1987
  - Python support - 2023
- CSV RFC 4180 - 2005



A screenshot of the VisiCalc spreadsheet application on an Apple II. The spreadsheet is titled 'C11 (L) TOTAL' and shows a list of items with their respective quantities, units, and costs. The data is as follows:

ITEM	NO.	UNIT	COST
MUCK RAKE	42	12.95	556.85
BUZZ CUT	15	6.75	101.25
TOFF TONER	250	49.95	12487.50
EYE SNUFF	2	4.95	9.90
SUBTOTAL			13155.50
9.75% TAX			1282.66
TOTAL			14438.16

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))
```

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

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

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



# 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
```

	A	B	C
1	<b>0 for False, 1 for True</b>		
2	0		
3	0		
4			

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")
```

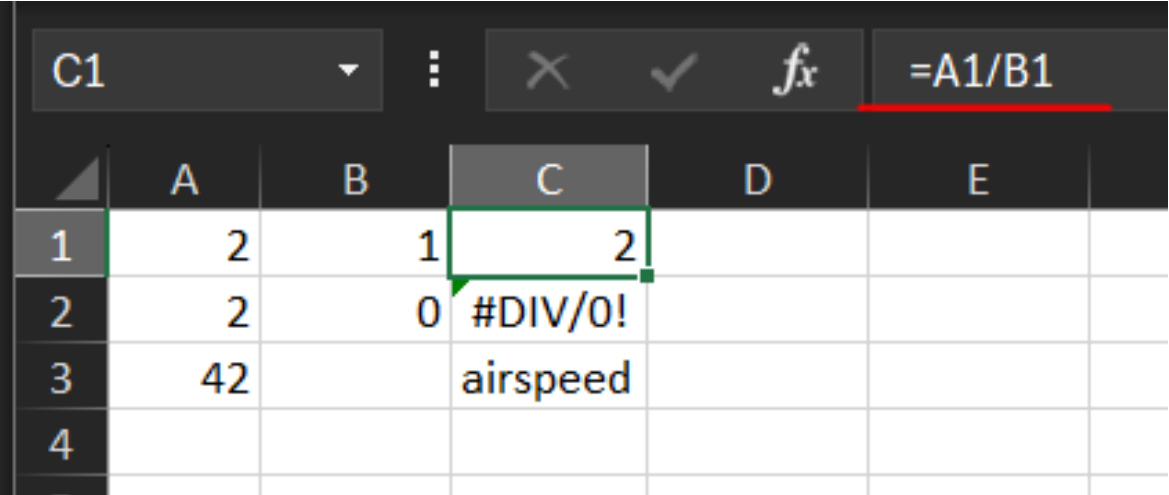
```
False
```

(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



The screenshot shows an Excel spreadsheet with columns A through E. The formula bar at the top displays '=A1/B1'. The spreadsheet data is as follows:

	A	B	C	D	E
1	2	1	2		
2	2	0	#DIV/0!		
3	42		airspeed		
4					

(apparently a header issue)

# Summary

- 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



Gorge of Eternal Peril

