

# EXCELSioR

## SatRday Johannesburg

Cesaire Tobias | [cesaire@fractalva.com](mailto:cesaire@fractalva.com)

Fractal Value Advisors

06 April 2019

# Bio

- Studied Business Science Economics.
- Started working as a quantitative analyst for Investec Asset Management (2015).
  - Focused predominantly on model migration (Excel to R) and econometric and fixed income analysis.
- Currently work as a consultant at Fractal Value Advisors.
  - Focus is on value based models and data munging.

# Overview

1. Why spreadsheets are important
2. Wrangling untidy data from spreadsheets using tidyxl
3. Challenges
4. Hardknocks
5. Conclusion
6. Useful links for further research

## Desired Outcomes

1. For those with the ability to code to take more time to appreciate the role of spreadsheets rather than loathe them.
2. For those without the ability to code to be aware of the challenges that poor spreadsheet design pose.
3. For everybody to not dismiss the value that spreadsheets can contribute.

# Why Spreadsheets are Important

*Spreadsheets are like the English of the data world. - Jenny Bryan*

- Most organisations create, store and disseminate data through spreadsheets.
- Enable knowledge experts without the ability to code to contribute their expertise.
- Great medium for prototyping.
- Provide a veil of trust/confidence/familiarity to non-technical users.
  - Important as sometimes this is the only means to communicate findings.

# Wrangling Data from Spreadsheets

## Why tidyxl?

- Great for unstructured, non tabular data
- `tidyxl::xlsx_cells()` imports data from spreadsheets without coercing it into a rectangle. Each cell is represented by a row in a data frame, giving the cell's address, contents, formula, height, width, and keys to look up the cell's formatting in the return value of `tidyxl::xlsx_formats()`.

*When your spreadsheet is too fat for readxl, tidyxl + unpivotr helps you tackle charming features like "data as formatting" and "data in the layout". - **Jenny Bryan***

# Turning Messy Spreadsheets to Tidy Tibbles

Statement of financial position						
	Actual	Actual	Actual	Actual	Actual	Actual
	30 September	30 September	30 September	30 September	30 September	30 September
Rm	2008	2009	2010	2011	2012	2013
<b>ASSETS</b>						
<b>Non-current assets</b>						
Property, plant and equipment	4,565,054	12,254,076	12,520,455	11,659,927	13,961,248	16,822,392
Loans and receivables	-	-	-	-	-	-
<b>Financial assets</b>	-	-	-	-	-	-
Deferred lease asset	-	-	-	-	-	-
Deferred taxation	-	-	-	-	-	-
<b>Total non-current assets</b>	<b>4,565,054</b>	<b>12,254,076</b>	<b>12,520,455</b>	<b>11,659,927</b>	<b>13,961,248</b>	<b>16,822,392</b>
<b>Current assets</b>						
Loans to group companies	-	-	-	-	-	-
Loans and receivables	-	-	-	-	-	-
<b>Financial assets</b>	-	-	-	-	-	-
<b>Property, plant and equipment</b>						
Cost	22,227,738	28,043,176	30,192,970	31,216,777	32,461,969	34,039,353
Accumulated depreciation and impairment	(17,662,684)	(15,789,100)	(17,672,514)	(19,556,851)	(18,500,721)	(17,216,961)
	4,565,054	12,254,076	12,520,455	11,659,927	13,961,248	16,822,392
<b>Square Meters</b>						
Land and buildings - Insurance value at 31 August 2016						
Plant and equipment - Insurance value at 30 September 2017						
<b>Cost</b>	<b>22,227,738</b>	<b>28,043,176</b>	<b>30,192,970</b>	<b>31,216,777</b>	<b>32,461,969</b>	<b>34,039,353</b>

Figure 1: Example client data

## Turning Messy Spreadsheets to Tidy Tibbles (cont.)

```
## # A tibble: 13,467 x 21
##   sheet address  row  col is_blank data_type error logical numeric
##   <chr> <chr>   <int> <int> <lgl>   <chr>   <chr> <lgl>   <dbl>
## 1 IS    B2         2    2 TRUE    blank   <NA>  NA      NA
## 2 IS    C2         2    3 TRUE    blank   <NA>  NA      NA
## 3 IS    D2         2    4 TRUE    blank   <NA>  NA      NA
## 4 IS    E2         2    5 TRUE    blank   <NA>  NA      NA
## 5 IS    F2         2    6 TRUE    blank   <NA>  NA      NA
## 6 IS    G2         2    7 TRUE    blank   <NA>  NA      NA
## 7 IS    H2         2    8 TRUE    blank   <NA>  NA      NA
## 8 IS    I2         2    9 TRUE    blank   <NA>  NA      NA
## 9 IS    J2         2   10 TRUE    blank   <NA>  NA      NA
## 10 IS   K2         2   11 TRUE    blank   <NA>  NA      NA
## # ... with 13,457 more rows, and 12 more variables: date <dtm>,
## #   character <chr>, character_formatted <list>, formula <chr>,
## #   is_array <lgl>, formula_ref <chr>, formula_group <int>, comment <chr>,
## #   height <dbl>, width <dbl>, style_format <chr>, local_format_id <int>
```



# Clean Data

## Produced in R, Written to Excel

lookup_id	file_name	fractal_variable	2015-09-30	2016-09-30	2017-09-30	2018-09-30	2019-09-30
Unit1Cash	Unit1	Cash	-5985.34	-59878.32	-58380.38	-60603.98	0
Unit1Cash and equivalents	Unit1	Cash and equivalents	-5985.34	-59878.32	-58380.38	-60603.98	0
Unit1Cash settled compensation liabilities	Unit1	Cash settled compensation liabilities	0	0	0	0	0
Unit1Comprehensive profit attributable to non controlling interests	Unit1	Comprehensive profit attributable to non controlling interests	0	0	0	0	0
Unit1Comprehensive profit attributable to ordinary shareholders	Unit1	Comprehensive profit attributable to ordinary shareholders	0	0	0	0	0
Unit1Comprehensive profit attributable to preference shareholders	Unit1	Comprehensive profit attributable to preference shareholders	0	0	0	0	0
Unit1Construction in progress (net)	Unit1	Construction in progress (net)	0	0	0	0	0
Unit1Corporate bond yield (average)	Unit1	Corporate bond yield (average)	0.093	0.093	0.093	0.093	0.093
Unit1Cost of goods sold	Unit1	Cost of goods sold	-856176.8	-1105943.95	0	1391.49	-42067.19
Unit1Cost of sales	Unit1	Cost of sales	-856176.8	-1105943.95	0	1391.49	-42067.19
Unit1Cost of sales depreciation	Unit1	Cost of sales depreciation	-17172883.44	-19351115.41	-21758129.64	-25222571.21	-25465373.69
Unit1Current assets	Unit1	Current assets	47950112.11	54185200	55340135.54	47396897.66	67888374.42
Unit1Current financial assets	Unit1	Current financial assets	0	0	0	0	0
Unit1Current financial liabilities	Unit1	Current financial liabilities	0	0	0	0	0
Unit1Current intercompany balances	Unit1	Current intercompany balances	340621047.8	360078055.3	395514181	432988840.7	417898991
Unit1Current liabilities	Unit1	Current liabilities	401057767.8	425431919.7	462638822.9	502955932.4	502759459
Unit1Current loans and receivables	Unit1	Current loans and receivables	0	0	0	0	0
Unit1Deferred lease asset	Unit1	Deferred lease asset	0	0	0	0	0
Unit1Deferred lease liability	Unit1	Deferred lease liability	0	0	0	0	0
Unit1Deferred tax asset	Unit1	Deferred tax asset	0	0	0	0	0
Unit1Deferred tax liability	Unit1	Deferred tax liability	0	0	0	0	0
Unit1Depreciation	Unit1	Depreciation	-17172883.44	-19351115.41	-21758129.64	-25222571.21	-25465373.69
Unit1Depreciation and amortisation	Unit1	Depreciation and amortisation	0	0	0	0	0
Unit1Diluted shares in issue	Unit1	Diluted shares in issue	0	0	0	0	0
Unit1Direct catering costs	Unit1	Direct catering costs	0	0	0	0	0
Unit1Direct payroll	Unit1	Direct payroll	-204841107.8	-220649949	-226395626.4	-237559771.2	-247463514.8
Unit1Dividends	Unit1	Dividends	0	0	0	0	0
Unit1Dividends paid	Unit1	Dividends paid	0	0	0	0	0

Figure 2: FVA Model Inputs

# The Model

<i>CFROI Calculation</i>	2010	2011	2012	2013	2014	2015	2016
EBITDA	74	56	56	45	62	60	59
+ Operating Lease Expense	0	0	0	0	0	0	0
+ R&D Expense	0	0	0	0	0	0	0
+ Investment Income	0	0	0	0	0	0	0
- Income Taxes	(17)	(12)	(10)	(7)	(11)	(11)	(11)
+ Income Tax on Non-recurring Gains (Losses)	0	0	0	0	0	0	0
- Monetary Holding Charge (Gain)	2	3	2	5	6	7	10
- Charge on FIFO Inventory	(1)	(2)	(2)	(4)	(3)	(3)	(4)
<b>Gross Cash Flow</b>	<b>57</b>	<b>45</b>	<b>45</b>	<b>40</b>	<b>54</b>	<b>53</b>	<b>53</b>
<b>Net Monetary Assets</b>	<b>(50)</b>	<b>(33)</b>	<b>(86)</b>	<b>(99)</b>	<b>(134)</b>	<b>(146)</b>	<b>(188)</b>
Inventories	32	37	58	56	52	60	55
Land (net)	0	0	0	0	0	4	3
Inflation adjustment	0	0	0	0	0	2	2
Other Long-term Assets	12	12	12	12	12	12	12
<b>Non-Depreciating Assets</b>	<b>-6</b>	<b>17</b>	<b>-15</b>	<b>-31</b>	<b>-70</b>	<b>-68</b>	<b>-116</b>
<b>Gross Plant (excluding Land &amp; CIP)</b>	<b>187</b>	<b>199</b>	<b>238</b>	<b>245</b>	<b>253</b>	<b>241</b>	<b>262</b>
Inflation-adjustment	101	118	126	157	203	160	169
Construction in Progress (net)	0	0	0	0	0	0	0
Capitalized Operating Leases	0	0	0	0	0	0	0
Capitalized R&D	0	0	0	0	0	0	0
Intangible Assets	0	(0)	27	22	22	23	23
<b>Depreciating Assets</b>	<b>288</b>	<b>317</b>	<b>391</b>	<b>424</b>	<b>478</b>	<b>424</b>	<b>454</b>
<b>Asset life calculated</b>	<b>14.9</b>	<b>15.0</b>	<b>12.9</b>	<b>12.1</b>	<b>11.2</b>	<b>12.9</b>	<b>14.8</b>
<b>Asset life used</b>	<b>15.0</b>	<b>15.0</b>	<b>15.0</b>	<b>15.0</b>	<b>15.0</b>	<b>15.0</b>	<b>15.0</b>
<b>CFROI</b>	<b>18.5%</b>	<b>10.8%</b>	<b>8.4%</b>	<b>5.3%</b>	<b>9.3%</b>	<b>11.5%</b>	<b>12.2%</b>
CFROI incl goodwill	18.5%	10.8%	8.4%	5.3%	9.3%	11.5%	12.2%
Real Asset Growth	-3.5%	11.3%	6.6%	-1.4%	-1.7%	-17.0%	-11.2%
Real Revenue Growth	-5.9%	-14.0%	4.8%	5.4%	11.5%	-5.7%	-2.8%
Inflation-adjusted Gross Investment Turns	1.05	0.81	0.80	0.85	0.97	1.10	1.20

Figure 3: FVA Model Example

# Summary Output in Excel

Business Unit1	Selected LFY		Reported LFY		Reporting unit		Scaling Factor	
	2018/09/30	2018/09/30	2018/09/30	2018/09/30	ZAR	ZAR	0.001	0.001
Summary	2008	2009	2010	2011	2012	2013	2014	2014
Revenue	281	296	296	271	299	335	395	
Revenue Growth		5.3%	0.1%	-8.4%	10.3%	11.9%	18.0%	
EBITDA	93	81	74	56	56	45	62	
EBITDA %	33%	28%	25%	21%	19%	13%	16%	
EBIT	82	69	61	43	37	24	39	
EBIT %	29%	23%	21%	16%	12%	7%	10%	
Invested Capital	109	110	108	130	146	113	53	
Invested Capital Turns (average)		2.7	2.7	2.3	2.2	2.6	4.8	
ROIC (average)		45%	40%	26%	19%	13%	34%	
Economic Profit (average)		34	28	13	6	(2)	15	
FCFF		49	47	9	10	50	89	
Interest-bearing Debt	0	0	0	0	5	6	(0)	
Net Debt (Cash)	(27)	(22)	(18)	(12)	(11)	(15)	(40)	
<b>CFROI</b>		<b>21.3%</b>	<b>18.5%</b>	<b>10.8%</b>	<b>8.4%</b>	<b>5.3%</b>	<b>9.3%</b>	
Gross Cash Flow		62	57	45	45	40	54	
Gross Investment	250	275	282	334	375	393	408	
Non Depr Assets	14	9	(6)	17	(15)	(31)	(70)	
Depr Assets	235	265	288	317	391	424	478	
Life	15	15	15	15	15	15	15	
Infl Adj Asset Turns	1.13	1.08	1.05	0.81	0.80	0.85	0.97	

Figure 4: FVA Model Summary

# Design (Excel)

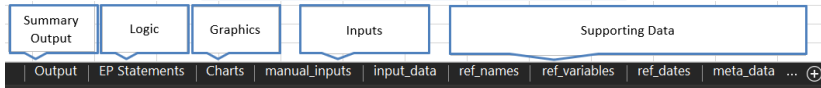


Figure 5: FVA Model Design

# Challenges

- Getting clients to supply consistent data.
  - Even consistently bad data is better than inconsistent data.
- Mixed object types in rows/columns.
- Missing data as blanks, NA's, dashes, and even 0's.
- Excel specific errors like #VALUE, #NAME?, #DIV/0.
- Writing functions that are generic enough to handle a broad range of cases without consuming too much time.

# Hardknocks

```
resolve_dupes <- function(raw_data, fs_vars) {  
  
  fractalAssert::assert_allpresent(names(raw_data), c("row", "col", "is_blank", "character"))  
  fractalAssert::assert_allpresent(names(fs_vars), c("received_var"))  
  
  dupe_vars <- raw_data %>%  
    dplyr::filter(character %in% fs_vars$received_var) %>%  
    dplyr::arrange(character) %>%  
    dplyr::mutate(dupe = duplicated(character)) %>%  
    dplyr::mutate(dupe2 = duplicated(character, fromLast = TRUE)) %>%  
    dplyr::mutate(dupes = ifelse(dupe | dupe2, TRUE, NA)) %>%  
    tidyr::drop_na(dupes) %>%  
    dplyr::select(row, col, character)  
  
  dupe_col <- unique(dupe_vars$col)  
  fractalAssert::assert_true(length(dupe_col) == 1, "snap - you've found duplicate variables")  
  
  resolved_dupe_df <- raw_data %>%  
    dplyr::left_join(dupe_vars, by = c("row", "col")) %>%  
    tidyr::fill(character.x, .direction = "down") %>%  
    dplyr::mutate(character = ifelse(is_blank & col == dupe_col + 1, character.x, character.y)) %>%  
    dplyr::mutate(char_type = ifelse(character %in% character.y, "sub", "head")) %>%  
    tidyr::drop_na(character) %>%  
    ...  
}
```

# Hardknocks (cont.)

Statement of financial position							
	Forecast 4	Forecast 4	Forecast 4	Forecast 4	Forecast 4		
	30 September	30 September	30 September	30 September	30 September		
Rm	2008	2009	2010	2011	2012		
<b>ASSETS</b>							
<b>Non-current assets</b>							
Property, plant and equipment	35,441,220	37,112,351	41,537,855	46,068,766	53,691,656	BLS000100	
Loans and receivables	-	-	-	-	-	BLS000200	
Financial assets	-	-	-	-	-	BLS010100	
Deferred lease asset	356,387	471,000	456,000	274,000	-	BLS000530	
Deferred taxation	-	-	-	-	-	BLS030800	
<b>Total non-current assets</b>	<b>35,797,607</b>	<b>37,583,351</b>	<b>41,993,855</b>	<b>46,342,766</b>	<b>53,691,656</b>		
<b>Current assets</b>							
Loans to group companies	-	-	-	-	-	BLS010710	
Loans and receivables	-	-	-	-	-	BLS010800	
Financial assets	-	-	-	-	-	BLS017100	
<b>Property, plant and equipment</b>							
Cost	71,400,049	79,305,430	90,526,223	102,448,540	118,253,061		
Accumulated depreciation and impairment	(35,958,829)	(42,193,079)	(48,988,368)	(56,379,774)	(64,561,405)		
	35,441,220	37,112,351	41,537,855	46,068,766	53,691,656		
Square Meters							
Land and buildings - Insurance value at 31 August 2016							
Plant and equipment - Insurance value at 30 September 2017							
<b>Cost</b>	<b>71,400,049</b>	<b>79,305,430</b>	<b>90,526,223</b>	<b>102,448,540</b>	<b>118,253,061</b>		

Figure 6: Don't outright dismiss spreadsheets as inferior

# Recap

- Why spreadsheets are important
- My workflow for wrangling data from spreadsheets using tidyxl
- Challenges
- Hardknocks



## Conclusion

Excel is here to stay. It is ubiquitous across many industries and a primary means of exploratory analysis. While the frustrations of wrangling data from spreadsheets are evident and often painful, the data community is privileged to:

1. Have the data to begin with
2. Be in a position to add real value to those data

*Excelsior! Ever upward. - Stan Lee*

SAT **R** DAY

## Useful Links

- [Spreadsheets in Data Science](#)
- [Tidying Messy Excel Data](#)
- [Data Organisation in Spreadsheets](#)
- [tidyxl](#)
- [readxl](#)
- [Beyond Earnings](#)