WWW.SAFE.COM





DATA PREP TIPS FOR BI PLATFORMS by riley greene

OVERVIEW

Business intelligence (BI) platforms such as Tableau, Power BI, and Qlik (and many others) provide tools that enable organizations to more easily identify insightful patterns in their data. Although they all read common formats like Excel and CSV, to take full advantage of their power, the data should be formatted in a certain way.

Tips:

- 1. Merge and/or append data that is spread across multiple sources into a single dataset.
- 2. Consolidate columns as much as possible. For instance, if a dataset has a column for each individual month, the data will work better in a BI application if these columns are condensed into two columns that represent Month and Value.
- 3. Eliminate rows and columns representing totals.
- 4. Create column headers that are unique and descriptive.
- 5. Eliminate duplicate headers, merged cells and nested tables.

These challenges can be resolved using Excel or data preparation tools within the BI platform. However, manual alterations to data can risk introducing errors, and for situations where many changes are required, repetitive steps can become tedious. FME® can execute the same tasks for data preparation in an automated, repeatable way. This guide will outline a few tips for preparing tabular data for business intelligence platforms with FME®.

4	LANG
6	FME [®] TOO
8	APPENDING
14	CONSOL
20	DUPLICATE
22	RENAN
24	CC

CONTENTS

UAGE IN FME® DLS FOR DATA PREP MULTIPLE DATASETS IDATE COLUMNS **HEADERS OR TOTALS** MING COLUMNS ONCLUSION

A Note on Language in FME

When working in FME and reading its documentation, keep these equivalencies in mind:

```
Attributes = columns
Features = rows
Feature Types = sheets (tables)
Readers = inputs (data connections)
Writers = outputs
Transformers = data transformations
Non-spatial = tabular
Spatial = lat/longs (mapping data)
```



FME Tools for Data Prep

Several tools in FME, called transformers, are used frequently for data remodelling tasks.

AttributeManager

Attributes in FME are equivalent to columns in a tabular format like Excel spreadsheets. This transformer enables users to rename, remove or add columns all in one place. Related transformers are the AttributeRenamer, AttributeRemover, and AttributeCreator which individually execute the same tasks as the AttributeManager.

TestFilter

Enables users to filter rows out of a dataset based on conditional formatting rules.

AttributeExploder

When used in conjunction with the TestFilter, this transformer is great for consolidating multiple columns that contain values for the same metric - month or location for instance - into one. It "explodes" a dataset into attribute (column) name and value pairs, listing them in two new columns, with the option to keep all other columns in the output. This will be explained further in the "Consolidate Columns" section.

Sorter

The sorter allows users to sort data based on different criteria.

FME Transformer Gallery

Manipulate your data exactly as needed by using any combination of FME's 450 transformers.



SAFE.COM/TRANSFORMERS

SCENARIO #1 - APPENDING MULTIPLE DATASETS

Oftentimes the data requiring analysis is spread across multiple worksheets because of a unique factor like time or place of collection.

•	• •							Passing St	ats by Year.xl	SX							
0			🖹 🔏 👔		· A	🔽 • 🕢 🏝	1	00% -	2				(Q- (Search i	n Sheet		
					~ .									Cocaron	II ONCOL	1.022	
	A Home Layo	out Table	S Charts	SmartArt	Formula	ns Data	Review									×	- Q 7
	A1 🛟	80(-	fx Player														
	A	В	C	D	E	F	G	Н	1	J	К	L	M	N	0	Р	(
1	Player	Sum of Comp	Sum of Att Ave	rage of CompPct S	Sum of Yds	Average of Avg	Sum of TD	Sum of Int	Average of Qbr	Sum of FstDn	Average of FstDnPct	Max of Lng	Sum of Sck	Sum of SckYds	Sum of Fum	Sum of FumL	
2	Anderson, Derek	36	53	59.55	453	5.10	2	5	58.53	22	21.20	48	0	0	0	0	
3	Baldwin, Doug	1	1	100.00	15	15.00	1	0	158.30	1	100.00	15	0	0	0	0	
4	Barkley, Matt	129	216	59.53	1611	7.71	8	14	69.21	89	41.19	3/	6	43	2	2	
5	Beasley, Cole	0	1	0.00	0	#DIV/01	0	0	#DIV/01	0	0.00	0	1		0	0	
7	Bell Lo'Voon	0	1	0.00	0	#010/01	0	0	#010/01	0	0.00	0	1	0	0	0	
8	Benjamin Travis	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	0	0	
9	Bortles Blake	368	625	58.30	3005	6.20	23	16	78 31	206	32.50	51	34	197	6	5	
10	Boykin Trevone	13	18	70.40	145	8.97	1	10	96.90	200	42 57	31	1	5	1	0	_
11	Bradford Sam	395	552	71.56	3877	7.07	20	5	99.65	197	35.75	71	37	276	9	5	_
12	Brady, Tom	291	432	67.58	3554	8.35	28	2	113.03	163	38.13	79	15	87	4	0	_
13	Brees, Drew	471	673	70.16	5208	7.81	37	15	101.93	264	39.73	98	27	184	5	4	_
14	Brissett, Jacoby	34	55	62.53	400	7.73	0	0	86.43	18	34.17	58	6	46	1	0	_
15	Bryant, Dez	1	1	100.00	10	10.00	1	0	147.90	1	100.00	10	0	0	0	0	
16	Carr, Derek	357	560	63.65	3937	7.00	28	6	96.09	189	33.97	75	16	79	2	2	
17	Cassel, Matt	30	51	71.90	284	7.00	2	2	82.60	16	21.27	50	5	23	1	0	
18	Clemens, Kellen	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	0	0	
19	Cook, Connor	14	21	66.70	150	7.10	1	1	83.40	6	28.60	32	2	7	2	1	
20	Cooper, Pharoh	0	1	0.00	0	0.00	0	1	0.00	0	0.00	0	0	0	0	0	
21	Cousins, Kirk	406	606	66.73	4917	8.27	25	12	98.89	225	37.66	80	23	190	6	2	
22	Cutler, Jay	81	137	60.06	1059	7.84	4	5	77.64	45	32.52	54	17	104	5	2	
23	Dalton, Andy	364	563	65.38	4206	7.68	18	8	93.69	204	36.83	86	41	264	4	1	
24	Daniel, Chase	1	1	100.00	16	16.00	0	0	118.80	1	100.00	16	0	0	0	0	
25	Dixon, Riley	1	1	100.00	16	16.00	0	0	118.80	1	100.00	16	0	0	0	0	_
26	Fales, David	2	5	40.00	22	4.40	0	0	53.80	2	40.00	12	1	5	0	0	_
27	Fitzgerald, Larry	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	0	0	_
28	Fitzpatrick, Ryan	228	403	55.71	2710	6.69	12	17	68.79	128	31.06	57	19	81	6	1	_
29	Flacco, Joe	436	672	64.71	4317	6.51	20	15	83.85	204	30.62	95	33	243	4	2	_
30	Foles, Nick	36	55	66.65	410	7.90	3	0	110.75	15	29.55	49	4	34	0	0	
	Cobhort Blains	2000	2001 2002	2003 200	4 2005	2006 2 2	007 20	08 200	2010	2011 2012	2 2013 2014	2015	2016 +		•		TI
	Normal View	Ready	A	A		A			Sum	=0	-						

For analysis in a BI platform these sheets need to be appended while adding a column that enables each individual dataset to still be uniquely

identified. For example, in the case of this NFL data, by season.

To do this in FME. first add a reader tool to the canvas, find the file and click "Parameters".

Add Reader	
Reader	
Format: Microsoft Excel	
Dataset: siness Intelligence/NFL Players/Passing Stats by Year.xlsx"	
Please verify that the reader parameters are correct.	
Workflow Options	
• Individual Feature Types	
Help Cancel OK	

Select "fme feature type".

		Micro
heeto	to Read	
meets	to Read	
		Sheet/Named Ranges
	2000	
	2001	
	2002	
	2003	
	2004	
	2005	
	2006	
	2007	
	2008	
Preview	N	
_		😑 🔘 🕒 🔂 Select Add
1	Player	
2	Aguiar, L	fme_basename
3	Aikman,	fme_color
4	Alstott, M	fme_dataset
-	Banks, T	fme_fill_color
5		
5 6	Batch, C	fme primary ax
5 6 7	Batch, C Berger, I	fme_primary_ax
5 6 7 8	Batch, C Berger, I Bettis, Ju	fme_primary_axi fme_rotation fme_secondary_
5 6 7 8	Batch, C Berger, I Bettis, Ju	fme_primary_axi fme_rotation fme_secondary_ fme_start_angle
5 6 7 8 9	Batch, C Berger, I Bettis, Ju Beuerlein	fme_primary_axi fme_rotation fme_secondary_ fme_start_angle
5 6 7 8 9 10	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_awaaa and Q. Filter
5 6 7 8 9 10	Batch, C Berger, I Bettis, J Beuerlei Bishop, I	fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_start_angle
5 6 7 8 9 10	Batch, C Berger, I Bettis, J Beuerlein Bishop, I tes	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_sweep_app Q Filter
5 6 7 8 9 10 Xttribu	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I tes	fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_succedent
5 6 7 8 9 10 	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I tes	fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_awage_awage fme_awage_awage fme_awage_awage Filter Player Sum of Comp
5 6 7 8 9 10 ttribu	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I tes Exposed C	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_start_angle fme_start_angle fme_start_angle Sum of Comp Sum of Att
5 6 7 8 9 10 xttribu 8 8 C D	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I tes Exposed C C	 ∫ fme_primary_axi ∫ fme_rotation ∫ fme_secondary_ ∫ fme_start_angle ∫ fme_start_angle
5 6 7 8 9 10 10 8 8 8 6 8 6 7 8 9 10 10 8 8 6 7 10 8 8 7 10 8 8 7 10 8 10 10 10 10 10 10 10 10 10 10 10 10 10	Batch, C Berger, t Bettis, Ju Beuerlein Bishop, t tes Exposed C C C C C	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_st
5 6 7 8 9 10 10 8 8 9 10 10 8 8 0 0 2 5 7 8 9 10 10 8 8 7 10 8 8 9 10 10 8 8 8 9 10 10 8 8 8 9 10 10 10 10 10 10 10 10 10 10 10 10 10	Batch, C Berger, t Bettis, Ju Beuerlein Bishop, t tes Exposed C C C C C C C C C C C C C C C C C C C	 fme_primary_axi fme_rotation fme_rotation fme_secondary_ fme_start_angle fme_start
5 6 7 8 9 10 Attribu E A D E F G	Batch, C Berger, I Bettis, Ju Beuerlein Bishop, I tes Exposed C C C C C C C C C C C C C C C C C C C	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_sta
5 6 7 8 9 10 ttribu R B C D E F G H	Batch, C Berger, I Bettis, Ju Beuerlei Bishop, I tes Exposed C C C C C C C C C C C C C C C C C C C	 fme_primary_axi fme_rotation fme_secondary_ fme_start_angle fme_start_angle<

Advanced

Schema Attributes

Additional Attributes to Expose: No ite

Help

To create a column that identifies which individual sheet the data point originated from, head down to Schema Attributes in the Excel Parameters dialog and click the ellipsis square next to "Additional Attributes to Expose".

~
~
~
~
~
~
~
~
~
Yds
on
ct Al

Back on the canvas, an FME "feature type" has been created for each individual worksheet from Excel. Browse a worksheet's columns by clicking the arrow next to its name. Observe that a new column has been created that represents the season of the data called "fme_feature_type".



+ - * * * *

Help

Defaults 🔽

Filter:

using the AttributeRenamer tool.



A GUIDE TO EXCEL & FME

TABLE OF CONTENTS

Vhy Process Excel Data with EME

Getting Started with Excel in FME

- Language Add Reader tool Creating Excel file

- Integrating Data from Multiple Worksheets
- Worksheets with the
- same structure Running a Vlookup

Manipulating Excel Data & Structures

- Columns and values
- iltering

Analytics & Reporting

EXCEL & FME





Import ...

OK

Cancel

The appended datasets look like this:

_	B112	: 80	(fx Bra	ady, Tom														-
		B	C	D	E	F	G	Н		J	K	L	М	N	0	Р	Q	
- 1	Season I	layer	Sum of Comp	Sum of Att	Average of CompPct	Sum of Yds	Average of Avg	Sum of TD	Sum of Int	Average of Qbr	Sum of FstDn A	verage of FstDnPct	Max of Lng	Sum of Sck	Sum of SckYds	Sum of Fum	Sum of FumL	
9)	2000	tewart, Kordell	151	289	50.59	1860	6.02	11	8	72.61	91	26.15	45	30	150	2	1	
9 L	2000	toerner, Clint	3	5	60.00	53	10.60	1	0	135.80	2	40.00	29	2	21	0	0	
92	2000	estaverde, Vinny	328	590	52.34	3732	5.86	21	25	66.11	179	28.03	63	13	71	2	2	
<mark>9</mark> 8	2000	homas, Thurman	0	0		0		0	0		0		0	1	2	0	0	
9	2000	hompson, Kevin	1	1	100.00	8	8.00	0	0	100.00	0	0.00	8	0	0	0	0	
95	2000	'an Pelt, Alex	4	8	50.00	67	8.40	0	0	78.60	2	25.00	36	0	0	0	0	
95	2000	Valter, Ken	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	1	0	
97	2000	Varner, Kurt	235	347	67.65	3429	10.19	21	18	95.75	153	44.33	85	20	115	2	0	
98	2000	Villiams, Ricky	1	1	100.00	34	34.00	0	0	118.80	1	100.00	34	0	0	0	0	
99	2000	Vright, Anthony	22	53	42.58	237	6.13	0	3	49.15	10	22.50	46	12	92	2	2	
100	2000	Vycheck, Frank	2	2	100.00	53	26.50	1	0	138.55	2	100.00	30	0	0	0	0	
101	2000	Vynn, Spergon	22	54	40.67	167	2.90	0	1	46.67	5	6.02	32	13	89	2	2	
102	2000	eier, Eric	3	3	100.00	19	6.30	0	0	93.10	1	33.30	14	0	0	0	0	
103	2000	irand Total	9497	16322	57.09	110131	6.92	634	531	77.07	5317	31.72	85	1232	7525	175	89	
104	2001	anks, Tony	198	370	53.63	2386	6.31	10	10	72.45	106	28.96	85	29	173	2	2	
105	2001	atch, Charlie	198	341	57.06	2392	6.80	12	12	72.75	107	29.20	76	33	176	5	3	
106	2001	erger, Mitch	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	0	0	
107	2001	ettis, Jerome	1	2	50.00	32	16.00	1	0	98.95	1	50.00	32	0	0	0	0	
108	2001	lake, Jeff	0	1	0.00	0	0.00	0	0	39.60	0	0.00	0	0	0	0	0	
109	2001	ledsoe, Drew	40	66	61.10	400	6.00	2	2	71.95	18	26.05	58	5	21	0	0	
110	2001	ooker, Marty	1	2	50.00	34	17.00	1	0	98.95	1	50.00	34	0	0	0	0	
111	2001	ouman, Todd	51	89	47.68	795	6.58	8	4	83.66	32	22.76	80	4	27	0	0	
112	2001	rady, Tom	264	413	63.74	2843	6.82	18	12	85.99	143	34.50	91	41	216	9	2	
113	2001	rees, Drew	15	27	55.60	221	8.20	1	0	94.80	11	40.70	40	2	12	2	0	
114	2001	rooks, Aaron	312	558	56.90	3832	7.01	26	22	78.86	183	33.45	63	50	330	4	0	
115	2001	rown, Travis	15	33	45.50	201	6.10	1	2	50.20	11	33.30	34	1	2	1	0	
116	2001	runell, Mark	289	473	61.28	3309	6.96	19	13	82.87	164	33.77	44	57	387	7	3	
117	2001	arter, Quincy	90	176	48.08	1072	5.69	5	7	62.06	46	24.26	64	12	56	2	0	
118	2001	handler, Chris	223	365	62.06	2847	7.86	16	14	81.90	127	34.66	94	41	261	4	1	
119	2001	ollins, Kerry	327	568	56.84	3764	6.59	19	16	76.48	189	33.21	76	36	206	10	4	
120	2001	ollins, Todd	3	4	75.00	40	10.00	0	0	106.30	1	25.00	26	0	0	0	0	
121	2001	ouch, Tim	272	454	60.04	3040	6.81	17	21	75.75	134					4	1	
122	2001	iraig, Dameyune	4	8	50.00	34	4.30	0	0	61.50	2					0	0	
123	2001	ulpepper Daunte	235	366	64 63	2612	7 27	14	13	83 88	126					9	5	
1		Player	Stats by Seaso	on +														
		Normal View Read	ły							Sum=	0	Note: Th	is will o	nly wor	k in FME i	f		1

Note: This will only work in FME if worksheets are named uniquely.

Scenario #2 Consolidate Columns

Business intelligence platforms do not work well with cross-table formatting, and columns need to be consolidated as much as possible.

•							Passing S	Stats by St	at Type.xlsx						
•			X h 🕻	§ 🔮 🗵	• 🕥 •	∑ • A	• 搔 • [fx 🔛 🛓	100% -	2		Q	•		⊗ ◄
1	Home	Layout	Tables	Charts	Smart	Art Forr	nulas D	ata Re	view						~
Ec	dit ;	Fo	nt		Alig	nment		Nu	mber	Fc	ormat		Cells		Themes
	Calik	ori (Body)	▼ 12 ▼		ab	wra	p Text 🔻 🛛 G	eneral			-	⇐ – –		Aa	•
Pa	ste B	ΙU					Merge 🔻 🖪	§ • %	0. 0. \$	Condition Formattir	nal Styles	Insert	Delete Fo	rmat Then	nes Aa•
	E12	÷ 6	3 🔿 (= fs	c											
		Α	В	С	D	E	F	G	H	I	J	К	L	М	N
1	Player		2000	2001	200	2 2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
2	Addai, Josep	h							0			0			
3	Aguiar, Louie	e	0					8		2		1			
4	Aikman, Tro	у	2												
5	Akers, David								0					0	
6	Alexander, D	Danario													0
7	Allen, James	1	8 2			0				8 2					
8	Alstott, Mike	9	0												
9	Amendola, D	Danny													
10	Anderson, D	erek		-					1	3	4	3	5		0
11	Anderson, R	ichie	5 2			0 0	0	6		2					
12	Andrews, An	ntonio													
13	Anger, Bryar	ſ													
14	Arrington, J.	J.									0				
15	Austin, Tavo	n	Completion D		Manda T	a v als al a v mar	lute un entire entire	Casha							
			Completion Pe	ercentage	Tards 1	ouchdowns	interceptions	Sacks							
	Norn	nal View	Ready								Sum=0		•		

Visit the

FME KNOWLEDGE CENTER

Q&A Forum Knowledge Base Ideas Exchange FME Hub

knowledge.safe.com

In the case of this table, a player's passing stat is displayed in a column for each season. To consolidate these columns - also sometimes referred to as "unpivoting" - into one "Season" column in FME, connect an AttributeExploder to the Reader. The AttributeExploder splits your data into two columns: attribute names and attribute values, and writes a row for each name-value pair.

For this particular dataset we are interested in the Season-Yards data for each player. In the AttributeExploder Parameters dialog, set the Attribute Name Label to "Seasons" and the Attribute Value Label to "Yards". Then set "Keep Attributes" to Yes. This keeps the column with the players' names in the table.

Here's what comes out of the AttributeExploder:

80

Table View

Table: no-conn-ffs-1508275334897_25716 [FFS] - AttributeExploder_OUTPUT

	Season	Player	Yards	2003	2000	2016	2005	2011	2007	2002	2004	2008	2012	2010
1	Player	Addai, Joseph	Addai, Joseph	<missing></missing>	<missii< th=""></missii<>									
2	xlsx_row_id	Addai, Joseph	2	<missing></missing>	<missii< th=""></missii<>									
3	2006	Addai, Joseph	0	<missing></missing>	<missii< th=""></missii<>									
4	fme_type	Addai, Joseph	fme_no_geom	<missing></missing>	<missi< th=""></missi<>									
5	2009	Addai, Joseph	22	<missing></missing>	<missii< th=""></missii<>									
6	Player	Aguiar, Louie	Aguiar, Louie	<missing></missing>	13	<missing></missing>	<missii< th=""></missii<>							
7	xlsx_row_id	Aguiar, Louie	3	<missing></missing>	13	<missing></missing>	<missii< th=""></missii<>							
8	2000	Aguiar, Louie	13	<missing></missing>	13	<missing></missing>	<missi< th=""></missi<>							
9	fme_type	Aguiar, Louie	fme_no_geom	<missing></missing>	13	<missing></missing>	<missii< th=""></missii<>							
10	Player	Aikman, Troy	Aikman, Troy	<missing></missing>	1632	<missing></missing>	<missi< th=""></missi<>							
11	xlsx_row_id	Aikman, Troy	4	<missing></missing>	1632	<missing></missing>	<missii< th=""></missii<>							
12	2000	Aikman, Troy	1632	<missing></missing>	1632	<missing></missing>	<missii< th=""></missii<>							
13	fme_type	Aikman, Troy	fme_no_geom	<missing></missing>	1632	<missing></missing>	<missii< th=""></missii<>							
٩			in any	column									349	94 row(s)

Columns...

3494 row(s)

17

From here, use a TestFilter to eliminate any rows that have a value in the "Season" column that don't represent a year.

Here is what the data looks like coming out of the TestFilter:

B112	🛟 😣 🛇 (* f	Brady, Tom	
4	Α	B	С
1 Season		Player	Yards
87	2000	Small, Torrance	0
88	2000	Smith, Akili	1253
89	2000	Smith, Lamar	0
90	2000	Stewart, Kordell	1860
91	2000	Stoerner, Clint	53
92	2000	Testaverde, Vinny	3732
93	2000	Thomas, Thurman	0
94	2000	Thompson, Kevin	8
95	2000	Van Pelt, Alex	67
96	2000	Walter, Ken	0
97	2000	Warner, Kurt	3429
98	2000	Williams, Ricky	34
99	2000	Wright, Anthony	237
100	2000	Wycheck, Frank	53
101	2000	Wynn, Spergon	167
102	2000	Zeier, Eric	19
103	2000	Grand Total	110131
104	2001	Banks, Tony	2386
105	2001	Batch, Charlie	2392
106	2001	Berger, Mitch	0
107	2001	Bettis, Jerome	32
108	2001	Blake, Jeff	0
109	2001	Bledsoe, Drew	400
110	2001	Booker, Marty	34
111	2001	Bouman, Todd	795
112	2001	Brady, Tom	2843
113	2001	Brees, Drew	221
114	2001	Brooks, Aaron	3832
115	2001	Brown, Travis	201
116	2001	Brunell, Mark	3309
117	2001	Carter, Quincy	1072
118	2001	Chandler, Chris	2847
119	2001	Collins. Kerry	3764
120	2001	Collins. Todd	40
121	2001	Couch. Tim	3040
122	2001	Croig Domoutino	24

	😪 TestFilter P	arameters	Pass	s Criteria				
sforme	er		P d d d	Griteria				
ansfor	rmer Name: TestFilter			Pass Criteria:	One Test (OR)			<
			Co	omposite Expression:				
Definit	itions							
	Test Condition	Output	Test	Clauses				
14	@Volue(Seesee) RECING WITH 2	Cutput		Left Value	Operator	Right Value	Negate	Mode
"	@value(Season) BEGINS_WITH Z	@valu				-	-	
1			1	Season	Begins With	2		Automatio
Else If Else	<all conditions="" other=""></all>	<unfi< td=""><td>1</td><td>Season</td><td>Begins With</td><td>2</td><td></td><td>Automati</td></unfi<>	1	Season	Begins With	2		Automati
Else If Else	<all conditions="" other=""></all>	<unfi< th=""><th>1</th><th>4 Season</th><th>Begins With</th><th>2</th><th></th><th>Automat</th></unfi<>	1	4 Season	Begins With	2		Automat
Else If Else	<all conditions="" other=""></all>	<unfi< th=""><th>1</th><th>Season</th><th>Begins With</th><th>2</th><th></th><th>Automati</th></unfi<>	1	Season	Begins With	2		Automati
Else If Else elp	<pre><all conditions="" other=""></all></pre>	<unfi< td=""><td>1</td><td>♠ Season</td><td>Begins With</td><td>2</td><td></td><td>Automatio</td></unfi<>	1	♠ Season	Begins With	2		Automatio
Else If Else elp	<all conditions="" other=""></all>	<unfi< td=""><td>1</td><td> </td><td>Begins With</td><td>2</td><td></td><td>Automatio</td></unfi<>	1	 	Begins With	2		Automatio
elp	<all conditions="" other=""></all>	<unfi< td=""><td>1 + Outp</td><td>◆ Season</td><td>Begins With</td><td>2</td><td></td><td>Automatio</td></unfi<>	1 + Outp	◆ Season	Begins With	2		Automatio

19

SCENARIO #3 FILTERING OUT DUPLICATE HEADERS OR TOTALS

Business intelligence platforms don't work properly if the data has two rows of headers or contains rows for totals, so they should be removed. Going back to Scenario #1, the data isn't 100% ready for a business intelligence platform, as scrolling through the data coming out of the AttributeRenamer reveals that each appended worksheet has a row at the bottom for the grand totals.

80

Table: Passing Stats by Year - Prepared for BI [XLSXR] - 2005

	Season	Player	Sum of Comp	Sum of Att
96	2000	Warner, Kurt	23	347
97	2000	Williams, Ric		1 1
98	2000	Wright, Anth	2	2 53
99	2000	Wycheck, Fr		2 2
100	2000	Wynn, Sperg	2	2 54
101	2000	Zeier, Fric		3 3
102	2000	Grand Total	949	16322
103	2001	Banke, Teny	19	8 370
104	2001	Batch, Charlie	19	8 341
105	2001	Berger, Mitch		0 1
106	2001	Bettis loromo		1 2
107	2001	Bla		
108	2001	Ble Hint: to	do this double	click the little ov

To eliminate these rows, use the TestFilter.

nsforme	er							
ransfor	rmer Name: TestFilter			Pass Criteria:	One Test (OR)			\$
t Defiei	tions		Co	mposite Expression:				
t Defini	tions		Test	Clauses				
	Test Condition	Output		Left Malue	Onerster	Diebt Melve	Manada	Marda
lf	@Value(Season) BEGINS_WITH 2	@Valu		Left Value	Operator	Right Value	Negate	Mode
Else If								
				V Season	begins with			Automatic
Else	<all conditions="" other=""></all>	<unfi< th=""><th></th><th>Season</th><th></th><th></th><th></th><th></th></unfi<>		Season				
Else	<all conditions="" other=""></all>	<unfi< th=""><th></th><th>3225011</th><th></th><th></th><th></th><th></th></unfi<>		3 225011				
Else + – Help	<all conditions="" other=""></all>	<unfi< td=""><td></td><td></td><td></td><td></td><td></td><td>Automatic</td></unfi<>						Automatic
Else + -	<all conditions="" other=""></all>	<unfi< td=""><td>+</td><td>· - A V I I</td><td></td><td></td><td></td><td>Duplicate</td></unfi<>	+	· - A V I I				Duplicate
Eise + – Help	<all conditions="" other=""></all>	<unfi< td=""><td>+ Outp</td><td>v - A v I X</td><td></td><td></td><td></td><td>Duplicate</td></unfi<>	+ Outp	v - A v I X				Duplicate

these values to a rejected output port.

To filter out columns that contain totals, use the AttributeRemover.

In the Test Conditions dialog, set the Test Clause to find rows where the value for Player equals "Grand Total". Check the Negate box to send

Scenario #4 - Renaming Columns

Business intelligence platforms require that columns be unique and descriptive. There are scenarios where it is necessary to rename the columns of datasets. For example, when data is read from an HTML table, often the column names will be Col1, Col2, Col3 and so on.

80

able: defense [HTMLTABLE] - Table1												
	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10		
1	RK	TEAM	YDS	YDS/G	PASS	P YDS/G	RUSH	R YDS/G	PTS	PTS/G		
2	1	Denver	1309	261.8	958	191.6	351	70.2	97	19.4		
3	2	Cincinnati	1314	262.8	798	159.6	516	103.2	83	16.6		
4	3	Atlanta	1562	312.4	1052	210.4	510	102.0	109	21.8		
5	4	Miami	1577	315.4	1175	235.0	402	80.4	84	16.8		
6	5	Washington	1580	316.0	1140	228.0	440	00 0	AttributeRenamer F	Parameter	rs	
7	6	Buffalo	1612	322.4	1174	234.8	438	Transformer				
8	7	Pittsburgh	1632	272.0	921	153.5	711 Trans					
9	8	Seattle	1650	330.0	1014	202.8	636 Tra	Transformer Name: AttributeRenamer_4				
10	9	Carolina	1680	280.0	1180	196.7	500 Attrib					
11	10	Dallas	1699	339.8	1109	221.8	590					
12	11	Minnesota	1773	295.5	1301	216.8	472	1 Attribute	Rank		Default Value	
13	12	Chicago	1816	302.7	1190	198.3	626 Co	2	Team	~		
٩	1		any column									
		U: ar	se the Attribund descriptive	+ He	 ▲ ▼ 革 ェ Defaults ▼ 	Filter:		Import Cancel OK				

Conclusion: Use FME for Automated, Hands-Free Data Prep

Hopefully this guide helps you prepare datasets to fit the rigid model requirements of business intelligence platforms. Using these tips in FME Workbench you can design automated, hands-free workflows that save time and errors and help you get the insights you need from your data.



FOLLOW US ON TWITTER

@SAFESOFTWARE



WWW.SAFE.COM

