

M2 Practical Challenge:

Creating Your Own Alteryx Workflow

The objective is to get a list of books base on “Goodreads Books”, for that I’m going to use two datasets, and combine them to obtain a bigger one.

Once that’s done, I’m going to make use of a third dataset in order to get a Bestseller Ranking for over 1 million books.

The goal is to obtain a list of the top 10 best “Goodreads Books” based on the Bestseller Ranking. As well as make some Scatter plots render, comparing the BestSeller Ranking to the Average Rating, Rating Count, and Number of pages. That way we can see the influence in the ranking of these variables.

Table of Contents

1. Importing the data	2
2. Create the WorkFlow:	4
2.1. Creating the Books dataset	4
2.2. Creating the BestSeller Ranking map	7
2.3. Transforming the data	8
3. Outputting the data	10
4. Review	14

1. Importing the data

First, we need to get 3 data sources. For that I have use Kaggle to obtain 3 .csv files related with books:

<https://www.kaggle.com/jealousleopard/goodreadsbooks/version/2>

<https://www.kaggle.com/dylanjcastillo/7k-books-with-metadata>

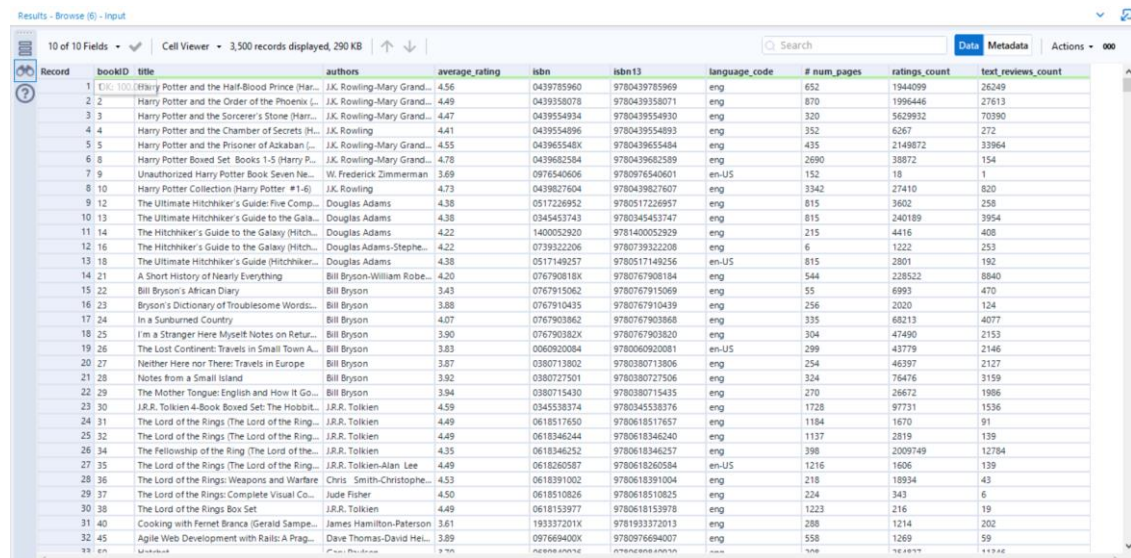
<https://www.kaggle.com/sp1thas/book-depository-dataset?select=dataset.csv>

The first .csv file include 3500 books obtain from the app “[goodreads books](#)”.

The second .csv file include 6810 books, made by “DylanCastillo”, based on the “Soumik's Goodreads-books dataset” and the “Google Books API”

The third .csv file contains 1,109,383 books, and its bestseller rank.

The input from the first dataset is:



The screenshot shows a Kaggle dataset viewer interface. At the top, it says "Results - Browse (6) - input". Below that, there's a search bar and tabs for "Data", "Metadata", and "Actions". The main area displays a table with 10 columns: Record, bookID, title, authors, average_rating, isbn, isbn13, language_code, # num_pages, ratings_count, and text_reviews_count. The table contains 35 records, with the first few rows showing books from the Harry Potter series by J.K. Rowling. The interface includes a "Cell Viewer" showing 3,500 records displayed out of 290 KB. There are also icons for zooming in and out, and a "Data" tab is selected.

Record	bookID	title	authors	average_rating	isbn	isbn13	language_code	# num_pages	ratings_count	text_reviews_count
1	100	Harry Potter and the Half-Blood Prince (Har...	J.K. Rowling-Mary Grand...	4.56	0439785960	9780439785969	eng	652	1944099	26249
2	2	Harry Potter and the Order of the Phoenix (...)	J.K. Rowling-Mary Grand...	4.49	0439358078	9780439358071	eng	870	1996446	27613
3	3	Harry Potter and the Sorcerer's Stone (Har...	J.K. Rowling-Mary Grand...	4.47	0439554934	9780439554930	eng	320	5629932	70390
4	4	Harry Potter and the Chamber of Secrets (H...	J.K. Rowling	4.41	0439554896	9780439554893	eng	352	6267	272
5	5	Harry Potter and the Prisoner of Azkaban (...)	J.K. Rowling-Mary Grand...	4.55	043965548X	9780439655484	eng	435	2149872	33964
6	8	Harry Potter Boxed Set: Books 1-5 (Harry P...	J.K. Rowling-Mary Grand...	4.78	0439682584	9780439682589	eng	2690	38872	154
7	9	Unauthorized Harry Potter Book Seven Ne...	W. Frederick Zimmerman	3.69	0976540606	9780976540601	en-US	152	18	1
8	10	Harry Potter Collection (Harry Potter #1-6)	J.K. Rowling	4.73	0439827604	9780439827607	eng	3342	27410	820
9	12	The Ultimate Hitchhiker's Guide: Five Comp...	Douglas Adams	4.38	0517226952	9780517226957	eng	815	3602	258
10	13	The Ultimate Hitchhiker's Guide to the Gala...	Douglas Adams	4.38	0345453743	9780345453747	eng	815	240189	3954
11	14	The Hitchhiker's Guide to the Galaxy (Hitch...	Douglas Adams	4.22	1400052920	9781400052929	eng	215	4416	408
12	16	The Hitchhiker's Guide to the Galaxy (Hitch...	Douglas Adams-Stephe...	4.22	0739322206	9780739322208	eng	6	1222	253
13	18	The Ultimate Hitchhiker's Guide (Hitchhiker...	Douglas Adams	4.38	0517149257	9780517149256	en-US	815	2801	182
14	21	A Short History of Nearly Everything	Bill Bryson-William Robe...	4.20	076790818X	9780767908184	eng	544	228532	8840
15	22	Bill Bryson's African Diary	Bill Bryson	3.43	0767915062	9780767915069	eng	55	6993	470
16	23	Bryson's Dictionary of Troublesome Words...	Bill Bryson	3.88	0767910435	9780767910439	eng	256	2020	124
17	24	In a Sunburned Country	Bill Bryson	4.07	0767903862	9780767903868	eng	335	68213	4077
18	25	I'm a Stranger Here Myself Notes on Retur...	Bill Bryson	3.90	076790382X	9780767903820	eng	304	47490	2153
19	26	The Lost Continent: Travels in Small Town A...	Bill Bryson	3.83	0060920084	9780060920081	en-US	299	43779	2146
20	27	Neither Here nor There: Travels in Europe	Bill Bryson	3.87	0380713802	9780380713806	eng	254	46397	2127
21	28	Notes from a Small Island	Bill Bryson	3.92	0380727501	9780380727506	eng	324	76476	3159
22	29	The Mother Tongue: English and How It Go...	Bill Bryson	3.94	0380715430	9780380715435	eng	270	26672	1986
23	30	J.R.R. Tolkien 4-Book Boxed Set: The Hobbit...	J.R.R. Tolkien	4.59	0345538374	9780345538376	eng	1728	97731	1536
24	31	The Lord of the Rings (The Lord of the Ring...	J.R.R. Tolkien	4.49	0618517650	9780618517657	eng	1184	1670	91
25	32	The Lord of the Rings (The Lord of the Ring...	J.R.R. Tolkien	4.49	0618346244	9780618346240	eng	1137	2819	139
26	34	The Fellowship of the Ring (The Lord of th...	J.R.R. Tolkien	4.35	0618346252	9780618346257	eng	398	2009749	12784
27	35	The Lord of the Rings (The Lord of the Ring...	J.R.R. Tolkien-Allan Lee	4.49	0618260587	9780618260584	en-US	1216	1606	139
28	36	The Lord of the Rings: Weapons and Warfare	Chris Smith-Christophe...	4.53	0618391002	9780618391004	eng	218	18934	43
29	37	The Lord of the Rings: Complete Visual Co...	Jude Fisher	4.50	0618510826	9780618510825	eng	234	343	6
30	38	The Lord of the Rings Box Set	J.R.R. Tolkien	4.49	0618153977	9780618153978	eng	1223	216	19
31	40	Cooking with Fernet Branca (Gerald Sampe...	James Hamilton-Paterson	3.61	193337201X	9781933372013	eng	288	1214	202
32	45	Agile Web Development with Rails: A Prag...	Dave Thomas-David Hel...	3.89	097669400X	9780976694007	eng	558	1269	59

The input from the second dataset is:

Results - Browse (10) - Input

12 of 12 Fields | Cell Viewer | 6,810 records displayed, 2.4 MB

Search

Rec...	isbn13	isbn10	title	subtitle	authors	categories	thumbnail	description	published_year	average_rating	num_pages	ratings_count
1	9780002005883	0002005883	Gilead	[Null]	Marilynne Robinson	Fiction	http://books.g...	A NOVEL THAT READ...	2004	3.85	247	361
2	9780002261982	0002261982	Spider's Web	A Novel	Charles Osborne/Agatha Christie	Detective and mystery...	http://books.g...	A new 'Christie for C...	2000	3.83	241	5164
3	9780006163831	0006163831	The One Tree	[Null]	Stephen R. Donaldson	American fiction	http://books.g...	Volume Two of Steph...	1982	3.87	479	172
4	9780006178736	0006178736	Page of Angels	[Null]	Sidney Sheldon	Fiction	http://books.g...	A memorable read...	1993	3.93	512	29532
5	9780006280897	0006280897	The Four Loves	[Null]	Clive Staples Lewis	Christian life	http://books.g...	Lewis' work on the n...	2002	4.15	170	33684
6	9780006280934	0006280934	The Problem of Pain	[Null]	Clive Staples Lewis	Christian life	http://books.g...	In The Problem of P...	2002	4.09	176	37569
7	9780006353287	0006353287	An Autobiography	[Null]	Agatha Christie	Authors, English	http://books.g...	Donation.	1977	4.27	560	3975
8	9780006380832	0006380832	Empires of the Moonson	A History of t...	Richard Hall	Africa, East	http://books.g...	Umbi Vasco da Gama...	1998	4.41	608	65
9	9780006470229	000647022X	The Gap Into Madness	Chaos and Or...	Stephen R. Donaldson	Hyland, Morn (Fictiti...	http://books.g...	A new-cover reissue...	1994	4.15	743	103
10	9780006472612	0006472612	Master of the Game	[Null]	Sidney Sheldon	Adventure stories	http://books.g...	Kate Blackwell is an e...	1982	4.11	489	43540
11	9780006479673	0006479673	If Tomorrow Comes	[Null]	Sidney Sheldon	Adventure stories	http://books.g...	One of Sidney Sheld...	1994	4.04	501	49170
12	9780006480099	0006480099	Assassin's Apprentice	[Null]	Robin Hobbs	American fiction	http://books.g...	Fantasy-roman.	1996	4.15	460	13392
13	9780006482079	0006482074	Warhost of Vastmark	[Null]	Janny Wurts	Fiction	http://books.g...	Tricked once more by...	1995	4.03	522	2966
14	9780006483014	0006483011	The Once and Future King	[Null]	Terence Hanbury White	Arthurian romances	http://books.g...	An omnibus volume...	1996	4.04	823	2805
15	9780006483892	0006483895	Murder in LaMuit	[Null]	Raymond E. Feist/Col Rosenberg	Adventure stories	http://books.g...	Available in the U.S. f...	2003	3.7	337	5083
16	9780006483908	0006483909	Jimmy the Hand	[Null]	Raymond E. Feist/S. M. Stirling	Fantasy fiction	http://books.g...	Jimmy the Hand, boy...	2003	3.95	368	5579
17	9780006486145	0006486142	Well of Darkness	[Null]	Margaret Weis/Tracy Hickman	[Null]	http://books.g...	Gareth is just a fright...	2001	3.66	599	68
18	9780006490456	000649045X	Witness for the Prosecution...	[Null]	Agatha Christie	English drama	http://books.g...	Newly Jacketed Editi...	1995	4.05	352	3908
19	9780006496434	0006496431	The Little House	[Null]	Philippa Gregory	Counrty life	http://books.g...	It was easy for Elizab...	1998	4.2	359	1832
20	9780006496878	0006496873	Mystical Paths	[Null]	Susan Howatch	English fiction	http://books.g...	1968 Andy Nicholas...	1996	4.23	576	1023
21	9780006496892	000649689X	Glimmering Images	[Null]	Susan Howatch	English fiction	http://books.g...	It is 1937, and Charle...	1996	4.07	512	2045
22	9780006496922	000649692X	Glamorous Powers	[Null]	Susan Howatch	Clergy	http://books.g...	Reissue of the author...	1996	4.2	512	1441
23	9780006498865	000649886X	The Mad Ship	[Null]	Robin Hobbs	Fantasy fiction	http://books.g...	Fantasy-master Robin...	2000	4.24	906	36101
24	9780006499169	0006499163	Post Captain	[Null]	Patrick O'Brian	Aubrey, Jack (Fictiti...	http://books.g...	This tale begins with...	1996	4.29	474	7815
25	9780006499268	0006499260	The Reverse of the Medal	[Null]	Patrick O'Brian	Adventure stories	http://books.g...	In this book, Jack Au...	1997	4.41	261	177
26	9780006499626	0006499627	Miss Marple	The Complete...	Agatha Christie	Detective and mystery...	http://books.g...	Miss Marple Features...	1997	4.2	359	8238
27	9780006511489	0006511481	The Years of Rice and Salt	[Null]	Kim Stanley Robinson	Black Death	http://books.g...	Hailed by 'The New Y...	2003	3.73	772	194
28	9780006512677	000651267X	Spares	[Null]	Michael Marshall Smith	Human cloning	http://books.g...	Spares - human clon...	1998	4.13	317	2481
29	9780006513087	0006513085	Gravity	[Null]	Tess Gerritsen	Science fiction	http://books.g...	Emma Watson a rese...	2004	4.04	342	8024
30	9780006514640	0006514642	The Wise Woman	[Null]	Philippa Gregory	Great Britain	http://books.g...	Alone and vulnerable...	2002	3.31	640	6583
31	9780006514855	0006514855	Girls' Night in	[Null]	Jessica Adams/Chris Marby/Fiona...	American fiction	http://books.g...	'Girls' Night in' featu...	2000	3.26	586	140
32	9780006545866	0006545866	The White Album	[Null]	John Didion	American essays	http://books.g...	This collection of ess...	1993	4.17	224	91

And the input from the third dataset is:

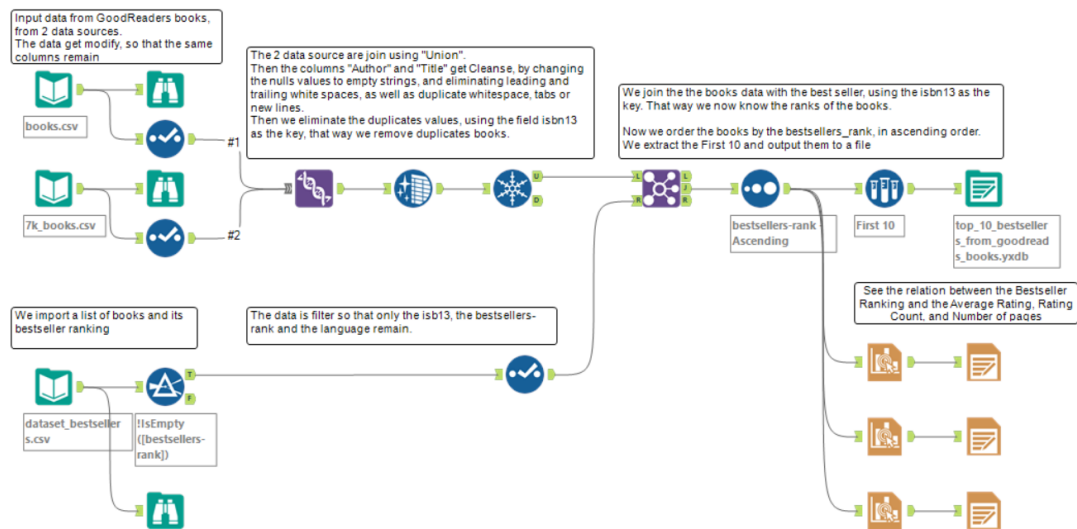
Results - Browse (8) - Input

28 of 28 Fields | Cell Viewer | 1,109,383 records displayed, 545 MB

Search

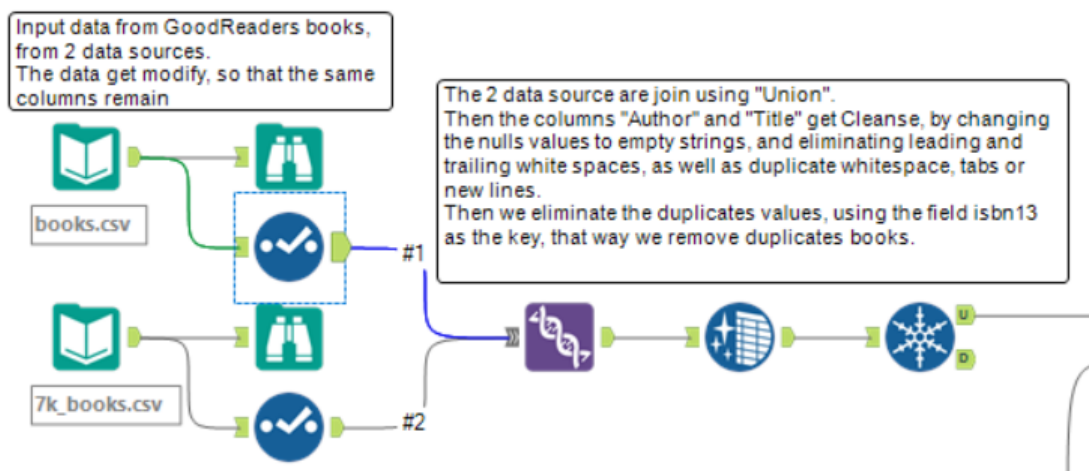
Rec...	authors	bestsellers_rank	catrg...	descrip...	dim...	di...	edition	edition...	for...	fo...	id	illustrations-note	ima...	ima...	im...	impr...	isbn10	isbn13	lang	pub...	p...	rating_avg	rating_count	title	url	weight
1	[1]	49948	[214, 2...]	SOLDIE...	129.0	198.0	20.0				1	97		97c...	full/c...	net...	Main...	184018907X	9781840189070	en	200...	4.03	292	Soldier Five...	/s...	2240
2	[2, 3]	115215	[235, 3...]	John M...	127.0	203.2	25.4				1	97		1dc...	full/c...	net...		184454737X	9781844547371	en	200...	3.6	335	Underbelly...	/u...	285.76
3	[4]	11732	[358, 2...]	Sir Phill...	150.0	234.0	28.0	New e...			1	97		3be...	full/c...	net...		8416327866	9788416327867	es	202...	3.88	37211	A Sir Phillip...	/s...	386.0
4	[5, 6, 7...]	114379	[377, 2...]	The Thir...	153.0	234.0	24.0		Export...		1	97		bcs...	full/c...	net...		571308996	9780571308996	en	201...	4.17	384	Q: The Thir...	/q...	436.0
5	[9]	98413	[2813, ...]	The Try...	191.0	240.0	29.0				2	97		39c...	full/c...	net...	Harp...	8352518	9780008352516	en	201...	3.9	5095	The Hidden...	/t...	980.0
6	[10, 11]	90674	[1529, ...]	When a...	130.0	200.0	13.0				1	97	35 colour illustra...	bed...	full/c...	net...	Banta...	533819224	978053819229	en	201...	4.04	64647	The Grand...	/g...	329.0
7	[6, 7, 8]	209150	[377, 2...]	Did you...	105.0	170.0	20.0			Main	1	97		d04...	full/c...	net...		571287955	9780571287951	en	201...	3.84	3923	1,227 Q: Fac...	/f...	164.0
8	[12, 13]	223191	[2980, ...]	We've a...	132.0	198.0	22.0				1	97	Integrated black...	4be...	full/c...	net...	Com...	1444762087	9781444762082	en	201...	3.56	268	The 100 Mo...	/t...	248.0
9	[14]	7004	[3223, ...]	For the...	140.0	213.0	15.0			Reprint	1	97		824...	full/c...	net...		1451636024	9781451636024	en	201...	3.82	33806	Waking Up...	/w...	181.0
1	[15]	66137	[37, 46...]	Part off...	193.0	272.0	15.24				1	97	1 Illustrations, u...	962...	full/c...	net...		1484788893	9781484788899	en	201...	4.2	5	Art Of Color...	/a...	476.27
1	[16]	318003	[2859, ...]	When a...	148.0	198.0	16.0				2	97		870...	full/c...	net...	Trape...	1409173356	9781409173359	en	201...	3.12	24	Stranger Fil...	/s...	384.0
1	[17]	76405	[1532, ...]		151.0	230.0	23.0				1	97		448...	full/c...	net...		1401936898	9781401936897	en	201...	4.14	12083	Biology of B...	/b...	405.0
1	[18]	509443	[2980, ...]	Olivia b...	89.0	91.0	61.0		6-8		1	97	colour illustra...	e6ef...	full/c...	net...	Runn...	762451017	9780762451012	en	201...			Olivia Dress...	/o...	84.0
1	[19]	662246	[146, 3...]	Based o...	133.0	196.0	27.0			UK ed.	1	97		588...	full/c...	net...	Corsar...	147211258X	9781472112583	en	201...	3.75	255	Bleak Spec...	/b...	242.0
1	[20, 21]	12741	[767, 1...]	What is...	129.0	198.0	18.0			Main	1	97		0e2...	full/c...	net...		1782395504	9781782395508	en	201...	4.01	3426	Reaching D...	/r...	254.0
1	[7, 8]	791174	[2980, ...]	Imagine...	196.0	285.0	8.0			Main	1	97		723...	full/c...	net...		057130883X	9780571308835	en	201...	3.8	61	Q: Annual 2...	/q...	367.0
1	[22]	868418	[2978, ...]	Griff Ri...	140.0	216.0	18.0			UK ed.	1	97	Illustrations (col...	c26...	full/c...	net...		1909844993	9781909844995	en	201...	3.46	68	Insufficient...	/i...	466.0
1	[23]	49141	[1532, ...]	Few thi...	129.0	198.0	19.0			Main	1	97	III	k7f...	full/c...	net...		1847672361	9781847672360	en	200...	3.84	52318	Book: The...	/b...	215.0
1	[24, 2...]	1303328	[146, 3...]	Series S...	138.0	140.0	23.0	Unabrid...	Unabrid...		3	97		636...	full/c...	net...	BBC...	1785296140	9781785296147	en	201...	5.0	2	Steptoe & S...	/s...	200.0
2	[29]	1214577	[142, 3...]	Buttling...	137.0	198.0	22.86				1	97		974...	full/c...	net...		1784183725	9781784183721	en	201...	3.29	56	A Travel Gu...	/t...	220.0
2	[30]	116088	[1532, ...]	In 1994...	127.0	196.0	23.0			Repr.	1	97	map	04...	full/c...	net...	Corgi...	555152188	9780555152181	en	200...	3.99	6413	The Seven...	/s...	251.0
2	[31, 32]	73234	[142, 3...]	The Sun...	162.0	240.0	30.0				2	97		250...	full/c...	net...	Blnk...	178870195X	9781788701952	en	201...	4.08	670	Mortimer B...	/m...	326.0
2	[33]	2417805	[334, 2...]	First pu...	129.0	198.0	22.0				1	97		7ae...	full/c...	net...		1526608359	9781526608352	en	201...	3.31	646	Otherhood	/o...	214.0
2	[34]	4804	[663, 8...]	In this...	153.0	234.0	14.0				1	97		820...	full/c...	net...	Portf...	241300754	9780241300756	en	201...	3.68	2935	The Fourth L...	/f...	256.0
2	[35]		[165, 3...]	Mans h...	126.0		198.0				1	97		371...	full/c...	net...	Ebury...	1529106303	9781529106305	en	202...			How to Be...	/h...	368.32
2	[36]	2717966	[2980, ...]	Harry Hi...	136.0	141.0	14.0	Unabrid...	Unabrid...		3	97		749...	full/c...	net...		571273386	9780571273386	en	201...	3.29	123	Live! the Dr...	/l...	149.0
2	[37]	61950	[2663, ...]	A fasci...	137.0	208.0	25.0			Reprint	1	97	Illustrations, uns...	4ca...	full/c...	net...	River...	1594485380	9781594485381	en	201...	4.03	12379	Where Goo...	/w...	340.0
2	[38, 3...]	3190426	[143, 3...]	Sixteen...	137.0	141.0	34.0	Unabrid...	Unabrid...		3	97		57c...	full/c...	net...	BBC...	1785299662	9781785299669	en	201...	4.5	4	Only Foots...	/o...	376.0
2	[39]		[165, 3...]	This bo...	144.0	222.0	35.0				2	97		7e9...	full/c...	net...	Ebury...	152910629X	9781529106299	en	202...			How to Be...	/h...	395.0
3	[41]	33687	[663, 1...]	A SUPNO...	129.0	198.0	21.0				1	97		47c...	full/c...	net...		141986069	9780141986067	en	201...	4.17	7151	The Future...	/t...	269.0
3	[42]	252306	[2978, ...]	TV Go...	190.0	245.0	7.0			Main	1	97	Illustrations (chi...	b1a...	full/c...	net...		571272193	9780571272198	en	201...	4.23	315	TV Go Home...	/t...	361.0
3	[43]	241278	[2980, ...]	In ALA...	138.0	138.0	16.0	Unabrid...	Unabrid...		3	97		439...	full/c...	net...	Trape...	1409160238	9781409160236	en	201...	4.13	4708	Alan Partid...	/a...	157.0

2. Create the WorkFlow:



2.1. Creating the Books dataset

On the upper part, the two Goodreads Books are imported. Before joining them, we need to make sure the have the same name and number of columns, as well as the same type of data.



I make use of the select tool to change the types of columns, change the names, and select only the needed columns.

Select (13) - Configuration

Options | TIP: To reorder multiple rows: select, right-click and drag

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	bookID	V_String	545		
<input checked="" type="checkbox"/>	title	V_String	545		
<input checked="" type="checkbox"/>	authors	V_String	545		
<input checked="" type="checkbox"/>	isbn	V_String	545	isbn10	
<input checked="" type="checkbox"/>	isbn13	Int64	8		
<input checked="" type="checkbox"/>	# num_pages	Int32	4	num_pages	
<input checked="" type="checkbox"/>	average_rating	Double	8		
<input checked="" type="checkbox"/>	ratings_count	Int32	4		
<input type="checkbox"/>	language_code	V_String	545		
<input type="checkbox"/>	text_reviews_count	Int32	4		
<input checked="" type="checkbox"/>	*Unknown	Unknown	0		Dynamic or...

Figure 2 - books.csv select

Select (14) - Configuration

Options | TIP: To reorder multiple rows: select, right-click and drag

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	title	V_String	6000		
<input checked="" type="checkbox"/>	authors	V_String	6000		
<input checked="" type="checkbox"/>	isbn10	V_String	6000		
<input checked="" type="checkbox"/>	isbn13	Int64	8		
<input type="checkbox"/>	published_year	Int32	4		
<input checked="" type="checkbox"/>	num_pages	Int32	4		
<input checked="" type="checkbox"/>	average_rating	Double	8		
<input checked="" type="checkbox"/>	ratings_count	Int32	4		
<input type="checkbox"/>	subtitle	V_String	6000		
<input type="checkbox"/>	description	V_String	6000		
<input type="checkbox"/>	categories	V_String	6000		
<input type="checkbox"/>	thumbnail	V_String	6000		
<input checked="" type="checkbox"/>	*Unknown	Unknown	0		Dynamic or Un...

Figure 2 – 7k_books.csv select

Once the two data source are modified, we can combine them. For that I made use of the Union tool. The main problem is that there are some books that have come from both datasets, for that we need to delete the duplicates books.

But first I clean the Title and Authors columns, for that I used the Data Cleansing tool, I decided to replace with blanks the nulls values, as the main identifier of the book is the ISBN, so a missing title or authors is not a big problem. Also deleted some Whitespaces and Tabs.

Data Cleansing (17) - Configuration

Options

Remove Null Data

☐ Remove null rows

☐ Remove null columns

Select Fields to Cleanse [All](#) [None](#)

☒ title

☒ authors

☐ isbn10

☐ isbn13

☐ num_pages

☐ average_rating

☐ ratings_count

Replace Nulls

☒ Replace with Blanks (String Fields)

☐ Replace with 0 (Numeric Fields)

Remove Unwanted Characters

☒ Leading and Trailing Whitespace

☒ Tabs, Line Breaks, and Duplicate Whitespace

☐ All Whitespace

☐ Letters

☐ Numbers

☐ Punctuation

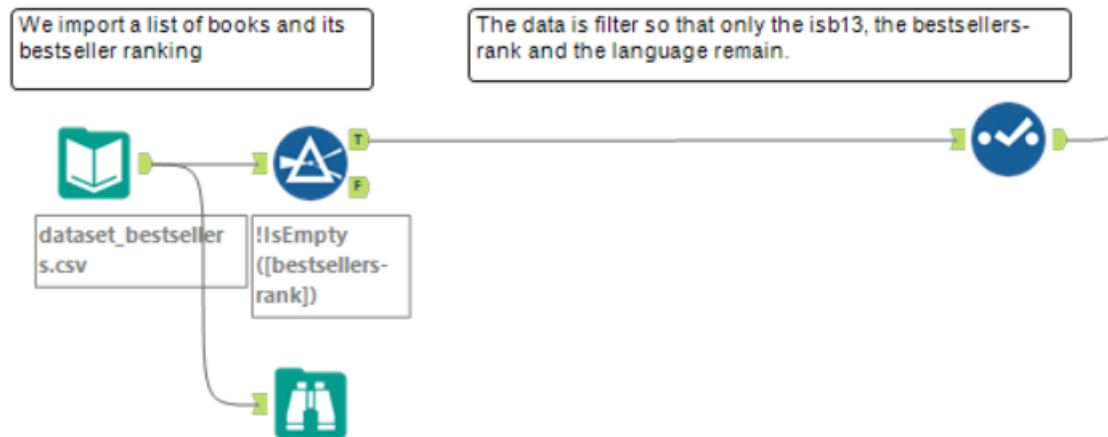
☐ Modify Case

Upper Case

Lastly, I made use of the Unique tool to remove duplicated books base on the ISBN 13 as the key.

2.2. Creating the BestSeller Ranking map

On the lower part, we can see the creation of the map between the books ISBN and the BestSeller Ranking.



The first thing I did was eliminate the rows that contains an empty value on the bestseller rank column. For that I made use of the Filter tool, to get the non-empty rows.

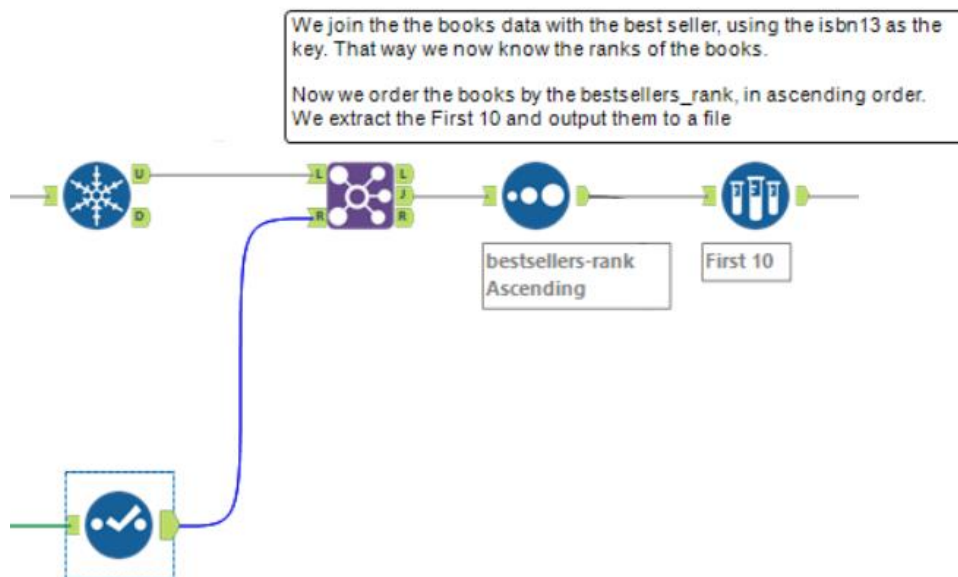
After that with the use of the Select tool, I selected the needed columns, such as title, bestsellers-rank, isbn10, isbn13 and lang.

The image shows two screenshots from the Apache NiFi interface. The left screenshot is the 'Filter (37) - Configuration' window. It shows the 'Basic filter' selected, with the filter expression 'bestsellers-rank' and the condition 'Is not empty'. The right screenshot is the 'Select (11) - Configuration' window. It shows a list of fields with checkboxes for selection. The selected fields are: title, bestsellers-rank, isbn10, isbn13, and lang. The table below shows the configuration for the Select tool.

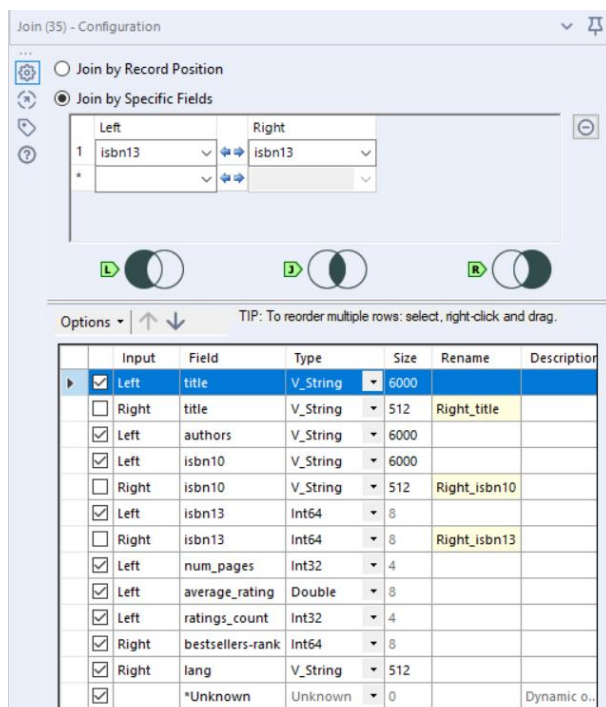
Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/> title	V_String	512		
<input checked="" type="checkbox"/> bestsellers-rank	Int64	8		
<input checked="" type="checkbox"/> isbn10	V_String	512		
<input checked="" type="checkbox"/> isbn13	Int64	8		
<input checked="" type="checkbox"/> lang	V_String	512		
<input type="checkbox"/> authors	V_String	512		
<input type="checkbox"/> categories	V_String	512		
<input type="checkbox"/> description	V_String	512		
<input type="checkbox"/> dimension-x	V_String	512		
<input type="checkbox"/> dimension-y	V_String	512		
<input type="checkbox"/> dimension-z	V_String	512		
<input type="checkbox"/> edition	V_String	512		
<input type="checkbox"/> edition-statement	V_String	512		
<input type="checkbox"/> for-ages	V_String	512		
<input type="checkbox"/> format	V_String	512		
<input type="checkbox"/> id	V_String	512		
<input type="checkbox"/> illustrations-note	V_String	512		
<input type="checkbox"/> image-checksum	V_String	512		
<input type="checkbox"/> image-path	V_String	512		
<input type="checkbox"/> image-url	V_String	512		
<input type="checkbox"/> imprint	V_String	512		
<input type="checkbox"/> index-date	V_String	512		
<input type="checkbox"/> publication-date	V_String	512		
<input type="checkbox"/> publication-place	V_String	512		

2.3. Transforming the data

For this part, the next part of the workflow is used.



First the output of the previous sections, are join, using the Join tool, and using the isbn13 field as the key. That way the output will contain the books from the merge dataset from “Goodreads Books”, and the associated bestseller ranking, from the other input.



After doing this step, I sample order the books by ascending order of the bestseller rank, by using the Sort tool. After that I used the Sample tool to get the first 10 books.

Sort (38) - Configuration

Use Dictionary Order

Inglés (Estados Unidos) (Predete

Fields

Name	Order
bestsellers-rank	Ascending

Sample (40) - Configuration

Select Sample Type

☒ First N rows

☐ Last N rows

☐ Skip 1st N rows

☐ 1 of every N rows

☐ 1 in N chance to include each row

☐ First N% of rows

N = 10

This way the output data from the sample is the following:

Results - Sample (40) - Output

9 of 9 Fields | Cell Viewer | 10 records displayed | Search

Re...	title	authors	isbn10	isbn13	num_pages	average_rating	ratings_count	bestsellers-rank	lang
1	Cloud Atlas	David Mitchell	0340822783	9780340822784	529	4.02	6763	1048	en
2	Into the Wild	Jon Krakauer	0330351699	9780330351690	206	3.97	1268	1099	en
3	A Moveable Feast	Ernest Hemingway	0099285045	9780099285045	192	4.04	78896	1254	en
4	The Art of Howl's Moving Castle	Hayao Miyazaki-Yuji Oniki-Joe Hisaishi	1421500493	9781421500492	256	4.51	2762	1255	en
5	Head, Shoulders, Knees and Toes ...		0859537285	9780859537285	12	4.06	1083	1480	en
6	Ulysses	James Joyce-Declan Kiberd	0141182806	9780141182803	933	3.74	1447	1640	en
7	Beyond Good and Evil	Friedrich Nietzsche	014044923X	9780140449235	240	4.02	43587	1660	en
8	Pride and Prejudice	Jane Austen-Tony Tanner-Vivien Jones	0141439513	9780141439518	480	4.25	11552	1853	en
9	J.R.R. Tolkien 4-Book Boxed Set: The Hobbit and...	J.R.R. Tolkien	0345538374	9780345538376	1728	4.59	97731	1981	en
10	Twelfth Night	William Shakespeare	0743482778	9780743482776	220	3.98	133832	2010	en

3. Outputting the data

For that I just connected the output from the Sample tool directly into an Output Data tool, and save it as “top_10_bestseller_s_from_goodreads_books.yxdb”



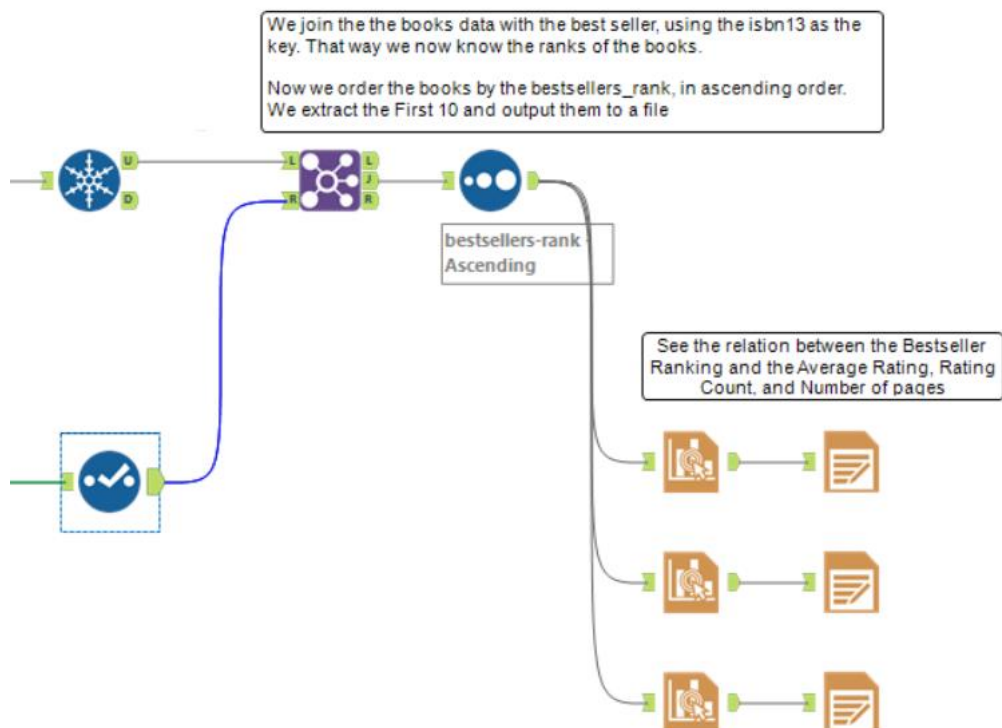
The output data will be the following:

Results - Sample (40) - Output

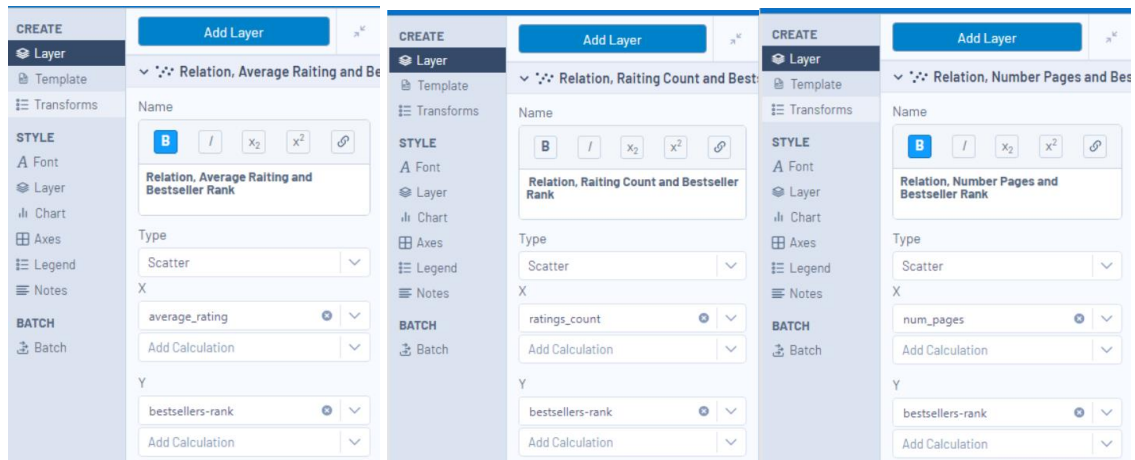
9 of 9 Fields | Cell Viewer | 10 records displayed | Search

Re...	title	authors	isbn10	isbn13	num_pages	average_rating	ratings_count	bestsellers_rank	lang
1	Cloud Atlas	David Mitchell	0340822783	9780340822784	529	4.02	6763	1048	en
2	Into the Wild	Jon Krakauer	0330351699	9780330351690	206	3.97	1268	1099	en
3	A Moveable Feast	Ernest Hemingway	0099285045	9780099285045	192	4.04	78896	1254	en
4	The Art of Howl's Moving Castle	Hayao Miyazaki-Yuji Oniki-Joe Hisaishi	1421500493	9781421500492	256	4.51	2762	1255	en
5	Head, Shoulders, Knees and Toes ...		0859537285	9780859537285	12	4.06	1083	1480	en
6	Ulysses	James Joyce-Decian Kiberd	0141182806	9780141182803	933	3.74	1447	1640	en
7	Beyond Good and Evil	Friedrich Nietzsche	014044923X	9780140449235	240	4.02	43587	1660	en
8	Pride and Prejudice	Jane Austen-Tony Tanner-Vivien Jones	0141439513	9780141439518	480	4.25	11552	1853	en
9	J.R.R. Tolkien 4-Book Boxed Set: The Hobbit and...	J.R.R. Tolkien	0345538374	9780345538376	1728	4.59	97731	1981	en
10	Twelfth Night	William Shakespeare	0743482778	9780743482776	220	3.98	133832	2010	en

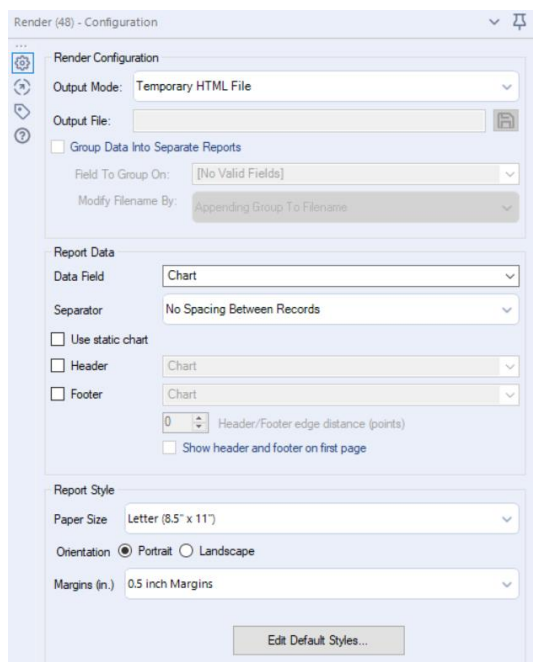
Another Output data that is presented is some Scatter plots renders.



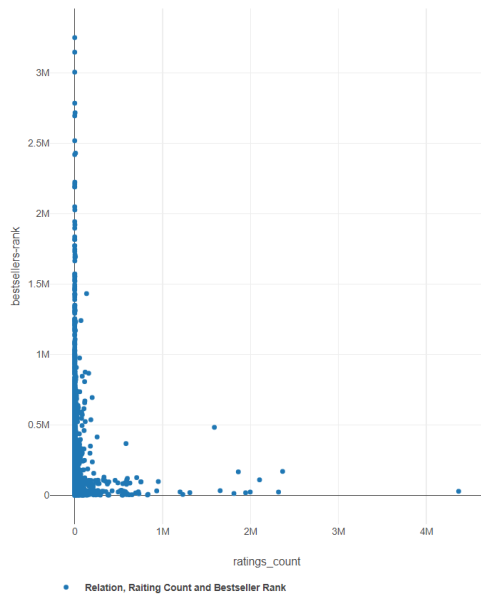
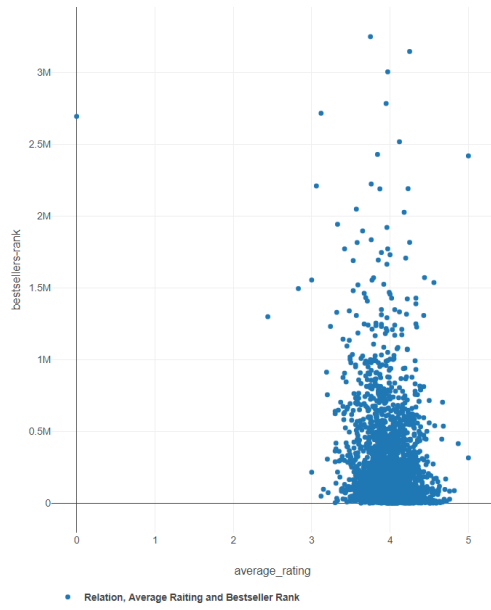
For that I connected the output of the Sort tool into 3 Interactive Chart tool, in order to make 3 graphs. Each one compare 3 thing to the Bestseller Rank. 1. Average Rating, 2. Rating Count, 3. Number of Pages.

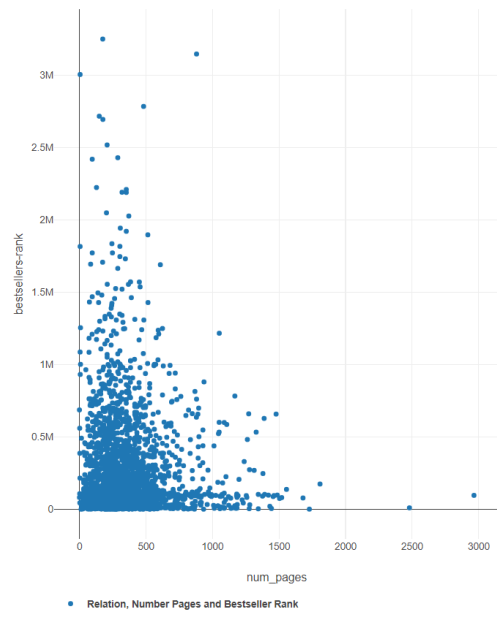


From than each graph is converted into a render an outputted as a temporal HTML file to visualize, using the Render tool.

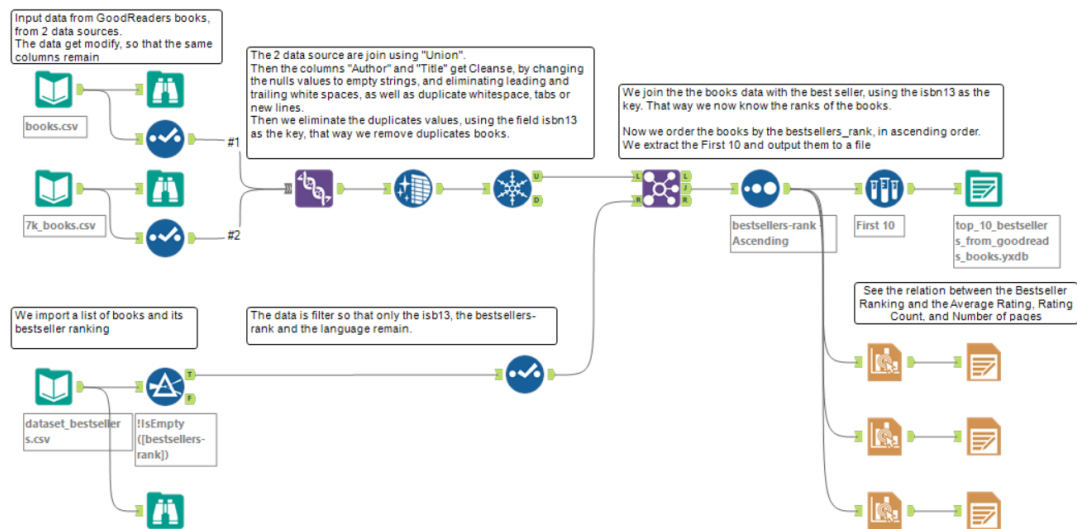


The 3 Scatter plots render are the following:





4. Review



As we can see from the workflow, there are 10 different tools used:

1. In/Out Category
 - a. Input Data (1)
 - b. Output Data (2)
2. Preparation Category
 - a. Select (3)
 - b. Data Cleansing (4) (NEW)
 - c. Unique (5) (NEW)
 - d. Filter (6)
 - e. Sort (7)
 - f. Sample (8)
3. Join Category
 - a. Union (9)
 - b. Join (10)

Also 2 new tools have been used, “Data Cleansing” and “Unique”, from the Preparation Category.

Apart from these tools, other tools have been used:

1. In/Out Category
 - a. Browse
2. Reporting Category
 - a. Interactive Chart
 - b. Render
3. Documentation Category
 - a. Comment