

# Recommending Alpha

## Group #16

### Data Collection

**Group Members:**

Vedanta Dhobley: [vjd41@scarletmail.rutgers.edu](mailto:vjd41@scarletmail.rutgers.edu)

Avani Bhardwaj: [ab1572@scarletmail.rutgers.edu](mailto:ab1572@scarletmail.rutgers.edu)

Shazidul Islam: [si194@scarletmail.rutgers.edu](mailto:si194@scarletmail.rutgers.edu)

Akshat Shah: [avs91@scarletmail.rutgers.edu](mailto:avs91@scarletmail.rutgers.edu)

Kutay Kerimoglu: [kk851@scarletmail.rutgers.edu](mailto:kk851@scarletmail.rutgers.edu)

Alan Patel: [akp122@scarletmail.rutgers.edu](mailto:akp122@scarletmail.rutgers.edu)

Anthony Matos: [amm720@scarletmail.rutgers.edu](mailto:amm720@scarletmail.rutgers.edu)

Joel Cruz: [jc2125@scarletmail.rutgers.edu](mailto:jc2125@scarletmail.rutgers.edu)

**AutoSave** | File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1				best_book_work_id	books.co isbn	isbn13	authors	original_p	j	t_title	language	average_r	atings_ccwork_rati	work_textratings_1	ratings_2	ratings_3	ratings_4	ratings_5	image_url	small			
2	2767052	2792775	272	4.39E+08	9.78E+12	Suzanne C	2008	The Hung	The Hung	eng	4.34	4780653	4942365	155254	66715	127936	560092	1481305	2706317	https://ir	https:		
3	3	4640799	491	4.4E+08	9.78E+12	J.K.R. Rowli	1997	Harry Pott	Harry Pott	eng	4.44	4602479	4800065	75867	75504	101676	455024	1156318	3011543	https://ir	https:		
4	41865	3212258	226	3.16E+08	9.78E+12	Stephenie	2005	Twilight	Twilight i	-en-US	3.57	3866839	3916824	95009	456191	436802	793319	875073	1355439	https://ir	https:		
5	2657	3275794	487	6.1120081	9.78E+12	Harper Lei	1960	To Kill a M	To Kill a M	eng	4.25	2198671	3340896	72586	60427	117415	446835	1001952	1714267	https://ir	https:		
6	4671	245494	1356	7.43E+08	9.78E+12	F. Scott Fi	1925	The Great	The Great	eng	3.89	2683664	2773745	51992	86236	197621	606158	936012	947718	https://ir	https:		
7	11870085	16827462	226	5.25E+08	9.78E+12	John Gree	2012	The Fault	The Fault	eng	4.26	2346404	2478609	140739	47994	92723	327550	698471	1311871	https://ir	https:		
8	5907	1540236	969	6.18E+08	9.78E+12	J.K.R. Tolk	1937	The Hobbi	The Hobbi	-en-US	4.25	2071616	2196809	37653	46023	76784	288649	665365	1119718	https://ir	https:		
9	5107	3036731	360	3.17E+08	9.78E+12	D.J. Saling	1951	The Catch	The Catch	eng	3.79	2042421	2120637	44920	109383	185520	455042	661516	709176	https://ir	https:		
10	960	3338963	311	1.42E+09	9.78E+12	Dan Brown	2000	Angels &	Angels &	i-en-CA	3.85	2001311	2078754	25112	77481	145479	458429	716596	680175	https://ir	https:		
11	1885	3060926	3455	6.8E+08	9.78E+12	Jane Austi	1813	Pride and	Pride and	eng	4.24	2035490	2191465	49152	54700	86485	284852	609755	1155673	https://ir	https:		
12	77203	3295919	283	1.59E+09	9.78E+12	Khaled Hc	2003	The Kite R	The Kite R	eng	4.26	1813044	1878095	59730	34288	59980	226062	628174	929591	https://ir	https:		
13	13335037	13155899	210	6.2024035	9.78E+12	Veronica I	2011	Divergent	Divergent	eng	4.24	1903563	2216814	101023	36315	82870	310297	737028	1114304	https://ir	https:		
14	5470	153313	995	4.52E+08	9.78E+12	George Or	1949	Nineteen	1984	eng	4.15	1956832	2053394	45518	41845	86425	324874	692021	908229	https://ir	https:		
15	7613	2207778	896	4.52E+08	9.78E+12	George Or	1945	Animal Fa	Animal Fa	eng	3.87	1881700	1982397	35472	68584	135147	433432	698642	648912	https://ir	https:		

FileHomeInsertPage LayoutFormulasDataReviewViewHelpTell me what you want to do

AutoSave | books.csv - Excel | Shazidul Islam

CutCopyFormat PainterClipboard

Paste

Clipboard

Calibri11

B

I

U

</







AutoSave book\_tags.csv - Excel Shazidul Islam

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

goodreads\_book\_id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
5117	415	5207	101																		
5118	415	31505	84																		
5119	415	7457	79																		
5120	415	18045	75																		
5121	415	236	75																		
5122	415	18326	74																		
5123	415	21152	74																		
5124	415	30672	66																		
5125	415	923	66																		
5126	415	2313	65																		
5127	415	18367	61																		
5128	415	21989	60																		
5129	415	32586	57																		
5130	415	22753	52																		
5131	415	8055	52																		
5132	415	24694	52																		
5133	415	24083	51																		
5134	415	9221	47																		
5135	415	30445	46																		
5136	415	24060	43																		
5137	415	2312	40																		
5138	415	14467	40																		
5139	415	14552	40																		

book\_tags

Ready

All the data that we have in the CSV file is from goodreads them self. Our teammate Alan is the one that actually maintained this database and brought it to the table.