In [49]:

import pandas as pd import numpy as np

In [16]:

main = pd.read\_csv('E:\爬虫\jjwx\作品库所有作品(数据少但是优质)\作品库.csv')

Out

ut[16]:		<b>a</b> _作者	a_作者_url	<b>b</b> _ 作 品	b_作品_url	<b>d</b> _ 风 格	<b>e</b> _ 进度	<b>f</b> _字数	<b>g</b> _作品积分	<b>h</b> _发表 时间	作品库类型	类 型 _ <b>1</b>	类 型 _ <b>2</b>	类 型 _ <b>3</b>
	0	竹已	oneauthor.php? authorid=1938546	偷偷 藏不 住	onebook.php? novelid=3676002	轻松	完结	372127	15400449024	2018- 07-24 23:08:19	完结作品	原创	言情	近代现代
	1	Zoody	oneauthor.php? authorid=3310540	云边 咖啡 馆	onebook.php? novelid=5095598	轻松	完结	87661	384190016	2020- 08-16 06:46:34	完结半价包月	原创	言情	近代现代
	2	竹已	oneauthor.php? authorid=1938546	难哄	onebook.php? novelid=4001734	轻松	完结	389894	14779252736	2019- 01-21 00:29:54	完结作品	原创	言情	近代现代
	3	栖见	oneauthor.php? authorid=585107	白日梦我	onebook.php? novelid=3525826	正剧	完结	427269	9028750336	2018- 03-16 13:21:32	完结作品	原创	言情	近代现代
	4	竹已	oneauthor.php? authorid=1938546	偷偷 藏不 住	onebook.php? novelid=3676002	轻松	完结	372127	15391267840	2018- 07-24 23:08:19	VIP 作 品	原创	言情	近代现代
	•••													
	61543		oneauthor.php? authorid=2417786	嫌人天会侦气半疑每都把探个死	onebook.php? novelid=6273634	轻松	连载	61162	8173361	2021- 10-18 12:58:04	免费作品	衍生	无 CP	近代现代
	61544	闲云向 晚	oneauthor.php? authorid=2415488	我妖迷的些	onebook.php? novelid=4011416	轻松	连载	163686	23800602	2019- 01-27 23:53:35	免费作品	衍生	纯爱	古色古香

	a_作者	a_作者_url	<b>b</b> _ 作 品	b_作品_url	<b>d</b> _ 风 格	e_ 进 度	<b>f</b> _字数	<b>g</b> _作品积分	<b>h</b> _发表 时间	品库类型	类 型 _ <b>1</b>	类 型 _ <b>2</b>	类 型 _ <b>3</b>	
61545	河鳝的 脑花	oneauthor.php? authorid=2213495	【职恋】修白起你风里	onebook.php? novelid=3500790	正剧	连载	317075	12327059	2018- 02-21 17:10:15	免费作品	衍生	纯爱	近代现代	
61546	聪明机 智菜菜 菜	oneauthor.php? authorid=1351301	西游 记之 九世 一生	onebook.php? novelid=2507295	正剧	完结	282775	39535896	2015- 07-16 11:56:54	免费作品	衍生	纯爱	古色古香	
61547	残阳落 暖	oneauthor.php? authorid=642847	我特的写 写	onebook.php? novelid=3554034	正剧	暂停	119208	19726530	2018- 04-10 22:56:55	免费作品	衍生	纯爱	近代现代	

作

61548 rows × 14 columns

In [17]:

author = pd.read\_csv('E:\爬虫\jjwx\作品库所有作品(数据少但是优质)\作者信息.csv')
author

Out[17]:		a_作者	<b>b</b> _被收	<b>c</b> _最近更新作品	d_作品	e_作品	f_最后更新	g_作者所发	<b>h</b> _小说	<b>i</b> _小说	<b>j</b> _小说	
			藏数		状态	字数	时间	送红包数	完本数	连载数	暂停数	
	0	Zoody	15070	《我的神他心比石 硬》	完结	143517	2022-02-15 17:34:52	27	7	10	0	
	1	栖见	253543	《桃枝气泡》	完结	394389	2021-04-12 05:01:00	23305	29	4	0	
	2	木苏里	721056	《不见上仙三百 年》	连载	309531	2022-03-12 00:55:29	193238	13	6	1	
	3	唧唧的 猫	140448	《是心跳说谎》	完结	330722	2021-08-10 20:03:51	11880	6	2	0	
	4	竹已	355330	《折月亮》	连载	318390	2022-03-12 21:59:59	23944	8	1	0	
	•••											
	19239	冬蝉鸣 夏	192	《cos旧剑后脚踩咒 术界高层》	连载	143640	2022-02-16 03:21:01	4	0	7	0	
	19240	上秋未 晴	313	《[综]诱捕三日月》	连载	203646	2021-08-06 20:20:20	47	1	7	0	
	19241	逆风or 随落	44	《斗罗之封宇终始》	完结	143531	2017-03-26 10:00:52	0	3	0	1	
	19242	番茄通 心粉	113	《npc在恋爱游戏花 式be后》	连载	498982	2022-03-13 18:00:00	5	3	3	0	

	a_作者	<b>b</b> _被收 藏数	c_最近更新作品	d_作品 状态	e_作品 字数	<b>f</b> _最后更新 时间	<b>g</b> _作者所发 送红包数	<b>h</b> _小说 完本数	i_小说 连载数	<b>j_</b> 小说 暂停数
19243	摩卡滋 味	3672	《直播异世界种田 的日子》	连载	329540	2021-12-26 23:40:00	29	10	5	1

## 19244 rows × 10 columns

In [18]: main\_author = pd.merge(main, author, on='a\_作者', how='left')
main\_author

Ou

out[18]:														
		a_作者	a_作者_url	<b>b</b> _ 作 品	b_作品_url	<b>d</b> _ 风 格	<b>e</b> _ 进 度	f_字数	g_作品积分	<b>h</b> _发表 时间	作品库类型	•••	类 型 _ <b>4</b>	<b>b</b> _被 薅
-	0	竹已	oneauthor.php? authorid=1938546	偷偷 藏不 住	onebook.php? novelid=3676002	轻松	完结	372127	15400449024	2018- 07-24 23:08:19	完结作品		爱情	35533
	1	Zoody	oneauthor.php? authorid=3310540	云边 咖啡 馆	onebook.php? novelid=5095598	轻松	完结	87661	384190016	2020- 08-16 06:46:34	完结半价包月		爱情	1507
	2	竹已	oneauthor.php? authorid=1938546	难哄	onebook.php? novelid=4001734	轻松	完结	389894	14779252736	2019- 01-21 00:29:54	完结作品		爱情	35533
	3	栖见	oneauthor.php? authorid=585107	白日梦我	onebook.php? novelid=3525826	正剧	完结	427269	9028750336	2018- 03-16 13:21:32	完结作品		爱情	25354
	4	竹已	oneauthor.php? authorid=1938546	偷偷 藏不 住	onebook.php? novelid=3676002	轻松	完结	372127	15391267840	2018- 07-24 23:08:19	VIP 作 品		爱情	35533
	•••													
	61543	千佑 Alleycat	oneauthor.php? authorid=2417786	嫌人天会侦气半疑每都把探个死	onebook.php? novelid=6273634		连载	61162	8173361	2021- 10-18 12:58:04	免费作品		轻小说	124
	61544	闲云向 晚	oneauthor.php? authorid=2415488	我妖迷的些	onebook.php? novelid=4011416	轻松	连载	163686	23800602	2019- 01-27 23:53:35	免费作品		东方衍生	3

		<b>a</b> _作者	í	a_作者_u	b_ url 作 品		b_作品_ur	d_ I 风 格	<b>e</b> _进度	<b>f</b> _字数	<b>g</b> _作品和	<sub>只分</sub>	_发表 时间	作品库类型		类 型 _ <b>4</b>	b_	被
	61545	河鳝的 脑花	oneau authorid	uthor.ph I=22134		one novelid	ebook.php? d=3500790		连载	317075	12327	059	2018- 02-21 :10:15	免费作品		东方衍生		1
	61546	聪明机智菜菜菜	oneau authorid	uthor.ph I=13513	•	one	ebook.php? d=2507295		完结	282775	39535	896	2015- 07-16 :56:54	免费作品		古典衍生	1(	64
	61547	残阳落 暖		uthor.ph id=6428		noveli	ebook.php? d=3554034		暂停	119208	19726	530	2018- 04-10 :56:55	免费作品	;	西方衍生	14	49
	61548 rd	ows × 23	3 column	ıS														
In [19]:	piece	['piece = piece	e_url']	<b>=</b> pie	ece['pi	.ece_ur		oly(1	amb	da x: x	质)\文章 [22:].st							
Out[19]:		b_	作品_url	作品 视角	作 品 全 风 格	文 字数 字数 雷地雷数	总书评数	总共 地雷 数量	法	是击点总	文章名 称	文章当 前被收 藏数	文	章积分	) }	章 类	进	签丝状态
	0	onebo novelid=	ook.php? 2595385	女主	正 698 剧	3356 89	88954	48892	2 14	4047236	有匪	409775	6455	11884	创 情想 多	原小言,架空历,武侠	完结	巨签丝

	b_作品_url	作品视角	作品风格	全文字数	前进一名所需地雷数	总书评数	总共 地雷 数量	总点击量	文章名 称	文章当 前被数 藏数	文章积分	文章类型	文章进度	签丝妆态
1	onebook.php? novelid=3429203	不明	正剧	187828	1	45820	6186	91041994	你是我的 荣耀	250945	906172928	原创言情近代现代爱情	完结	<b>未签约</b>
2	onebook.php? novelid=5095598	女主	轻松	87661	1	2393	234	862970	云边咖啡 馆	47165	382963168	原创言情近代现代爱情	完结	三签丝
3	onebook.php? novelid=205496	女主	轻松	55520	1	1091	0	251539	《死神 Bleach同 人》水月 纱夜	25153	156832048	衍生言情古色古香东方衍生	完结	正签丝
4	onebook.php? novelid=15488	女主	正剧	36201	1	959	2	170962	推理短文八篇	27309	143094304	原创言情近代现代悬疑	完结	三名

... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

	b_作品_url	作品视角	作品风格	全文字 数	前进一名所需地雷数	总书评数	总共 地雷 数量	总点击量	文章名 称	文章当 前被收 藏数	文章积分	文章类型	文章进度	签丝步态
41297	onebook.php? novelid=237304	NaN	正剧	1297466	1	9674	5	3445218	风 *******)	2340	159107120	衍生未知古色古香东方衍生	完结	巨签丝
41298	onebook.php? novelid=5619492	主攻	正剧	58139	1	513	0	382106	当一梦江 湖的奶爸 来到山河 令	1950	6340120	衍生纯爱古色古香其他衍生	连载	未签丝
41299	onebook.php? novelid=2966014	不明	正剧	542964	1	1707	73	3725932	[综]红楼 死亡不完 全手册	1950	39957068	衍生无P的想未来古典衍生	完结	三名公
41300	onebook.php? novelid=2520185	主受	正剧	143531	1	222	4	2057553	斗罗之封 宇终始	1934	7821952	衍生纯爱架空历史东方衍生	完结	未签丝

		<b>b</b> _f⁄	乍品 <b>_url</b>	作品视角	作品全	全文字 数		总书 评数	总共 地雷 数量	忘	金点击量	文章名 称	文章当 前被收 藏数	文i	章积分	文章类型	进	签丝妆态
	41301	oneboo novelid=	ok.php? 491678	主受	轻松	46363	1 1	1572	5		640879	[HP] 懒散度日	1949	80	03392	衍生纯爱近代现代西方衍生	连载	- 未签丝
	41302 rd	ows × 16	column	S														
In [37]:	data data	= pd.me:	rge(ma	in_au	thor,	piec	e, on=	:'b_作	ពី_u	rl'	, how=	'left')						
Out[37]:		a_作者	í	a_作者_ <b>ι</b>	url ∱	<b>)</b> _ 乍 品	<b>b</b> _作	묘_url	<b>d</b> _ 风	<b>e</b> _ 进度	<b>f</b> _字数	g_作品利	<sub>只分</sub> h	_发表 时间	作品库		共地数量	
	0	竹已	oneau authorid	uthor.ph =19385	46	<b></b>	onebool velid=36		轻松	完结	372127	154004490	024	2018- 07-24 :08:19	完结作品	. 168	378.0	2
	4	7	oneau	ıthor.ph	p? 云边	力	onebool	k.php?	轻	完	07664	20.44000	216	2020-	完结半		240	

onebook.php? 轻 完 novelid=5095598 松 结

87661

价/

包 月

234.0

08-16

06:46:34

384190016

oneauthor.php? authorid=3310540

Zoody

咖啡

馆

前

	a_作者	a_作者_url	<b>b</b> _ 作 品	b_作品_url	<b>d</b> _ 风 格	<b>e</b> _ 进 度	<b>f</b> _字数	<b>g</b> _作品积分	<b>h</b> _发表 时间	作品库类型	•••	总共地 雷数量	
2	竹已	oneauthor.php? authorid=1938546	难哄	onebook.php? novelid=4001734	轻松	完结	389894	14779252736	2019- 01-21 00:29:54	完结作品		33901.0	,
3	栖见	oneauthor.php? authorid=585107	白日梦我	onebook.php? novelid=3525826	正剧	完结	427269	9028750336	2018- 03-16 13:21:32	完结作品		17660.0	2
4	竹已	oneauthor.php? authorid=1938546	偷偷 藏不 住	onebook.php? novelid=3676002	轻松	完结	372127	15391267840	2018- 07-24 23:08:19	VIP 作 品		16878.0	2
•••													
61543	千佑 Alleycat	oneauthor.php? authorid=2417786	嫌人天会侦气半疑每都把探个死	onebook.php? novelid=6273634	轻松	连载	61162	8173361	2021- 10-18 12:58:04	免费作品		0.0	
61544	闲云向晚	oneauthor.php? authorid=2415488	我妖迷的些	onebook.php? novelid=4011416	轻松	连载	163686	23800602	2019- 01-27 23:53:35	免费作品		17.0	

	a_作者	a_作者_url	<b>b</b> _ 作 品	b_作品_url	<b>d</b> _ 风 格	<b>e</b> _ 进 度	f_字数	<b>g</b> _作品积分	<b>h</b> _发表 时间	作品库类型	•••	总共地 雷数量	
61545	河鳝的 脑花	oneauthor.php? authorid=2213495	【职 与叶 x 等在全儿恋】修白起你风里	onebook.php? novelid=3500790	正剧	连载	317075	12327059	2018- 02-21 17:10:15	免费作品		13.0	
61546	聪明机 智菜菜 菜	oneauthor.php? authorid=1351301	西游 记之 九世	onebook.php? novelid=2507295	正剧	完结	282775	39535896	2015- 07-16 11:56:54	免费作品		205.0	
61547	残阳落暖	oneauthor.php? authorid=642847	我特的写有殊侧技巧	onebook.php? novelid=3554034	正剧	暂停	119208	19726530	2018- 04-10 22:56:55	免费作品		20.0	

61548 rows × 38 columns

In [38]: data = data.drop(['类型\_1', '类型\_2', '类型\_3', '类型\_4', 'g\_作品积分', 'b\_作品', 'e\_进度', 'd\_ data 作 Out[38]:  $\mathbf{d}_{-}$ 作品 品 f\_最后 **h**\_发表 时间 **b**\_被收 **c**\_最近更 新作品 e\_作品 总雷 库 a\_作者 a\_作者\_url b\_作品\_url 更新时 藏数 字数 类 状 间 型 态

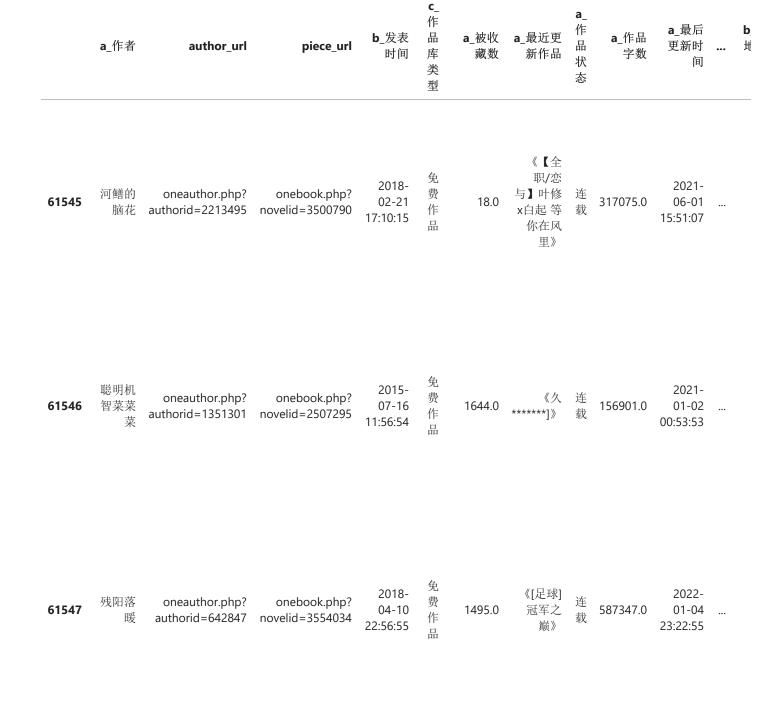
	<b>a</b> _作者	a_作者_url	b_作品_url	<b>h</b> _发表 时间	作品库类型	<b>b</b> _被收 藏数	<b>c</b> _最近更 新作品	d_作品状态	<b>e</b> _作品 字数	<b>f</b> _最后 更新时 间	•••	<b>总</b>
0	竹已	oneauthor.php? authorid=1938546	onebook.php? novelid=3676002	2018- 07-24 23:08:19	完结作品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		16
1	Zoody	oneauthor.php? authorid=3310540	onebook.php? novelid=5095598	2020- 08-16 06:46:34	完结半价包月	15070.0	《我的神 他心比石 硬》	完结	143517.0	2022- 02-15 17:34:52		
2	竹已	oneauthor.php? authorid=1938546	onebook.php? novelid=4001734	2019- 01-21 00:29:54	完结作品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		33
3	栖见	oneauthor.php? authorid=585107	onebook.php? novelid=3525826	2018- 03-16 13:21:32	完结作品	253543.0	《桃枝气 泡》	完结	394389.0	2021- 04-12 05:01:00		17
4	竹已	oneauthor.php? authorid=1938546	onebook.php? novelid=3676002	2018- 07-24 23:08:19	VIP 作 品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		16
•••												

	a_作者	a_作者_url	b_作品_url	<b>h</b> _发表 时间	作品库类型	<b>b</b> _被收 藏数	<b>c</b> _最近更 新作品	d_作品状态	<b>e</b> _作品 字数	<b>f</b> _最后 更新时 间	•••
61543	千佑 Alleycat	oneauthor.php? authorid=2417786	onebook.php? novelid=6273634	2021- 10-18 12:58:04	免费作品	1242.0	《嫌疑人 每天都会 把侦探气 个半死》	连载	61162.0	2022- 03-04 01:10:18	
61544	闲云向晚	oneauthor.php? authorid=2415488	onebook.php? novelid=4011416	2019- 01-27 23:53:35	免费作品	36.0	《我做妖王迷弟的那些年》	连载	163686.0	2019- 03-26 23:01:01	
61545	河鳝的 脑花	oneauthor.php? authorid=2213495	onebook.php? novelid=3500790	2018- 02-21 17:10:15	免费作品	18.0	《【全 职/恋 与】叶修 x白起等 你在风 里》	连载	317075.0	2021- 06-01 15:51:07	
61546	聪明机 智菜菜 菜	oneauthor.php? authorid=1351301	onebook.php? novelid=2507295	2015- 07-16 11:56:54	免费作品	1644.0	《久 ******]》	连载	156901.0	2021- 01-02 00:53:53	

总雷

		a_作者	a_作者_url	b_作品_url	<b>h</b> _发表 时间	作品库类型	<b>b</b> _被收 藏数	<b>c</b> _最近更 新作品	d_作品状态	<b>e</b> _作品 字数	<b>f</b> _最后 更新时 间	•••	<b>岩</b>
	61547	残阳落暖	oneauthor.php? authorid=642847	onebook.php? novelid=3554034	2018- 04-10 22:56:55	免费作品	1495.0	《[足球] 冠军之 巅》	连载	587347.0	2022- 01-04 23:22:55		
	61548 rd	ows × 29	columns										
In [44]:	data.	columns	'b_前进一名	'author_url',' 时间', 'a_作者所 所需地雷数', 'b <sub>_</sub> ', 'b_签约状态'	斤发送红包 总书评数	数', ', '	'a_小说' b_总共地	完本数', 雷数量',	'a_/ 'b_/	小说连载数 总点击量' <mark>,</mark>	', 'a 小	说暂	停
In [45]:	data												
Out[45]:		a_作者	author_url	piece_url	<b>b</b> _发表 时间	<b>c</b> _作品库类型	<b>a</b> _被收 藏数	<b>a</b> _最近更 新作品	a_作品状态	<b>a</b> _作品 字数	<b>a</b> _最后 更新时 间	•••	<b>b</b> . 地
	0	竹已	oneauthor.php? authorid=1938546	onebook.php? novelid=3676002	2018- 07-24 23:08:19	完结作品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		16
	1	Zoody	oneauthor.php? authorid=3310540	onebook.php? novelid=5095598	2020- 08-16 06:46:34	完结半价包月	15070.0	《我的神 他心比石 硬》	完结	143517.0	2022- 02-15 17:34:52		

	a_作者	author_url	piece_url	<b>b</b> _发表 时间	c_作品库类型	<b>a</b> _被收 藏数	<b>a</b> _最近更 新作品	a_作品状态	<b>a</b> _作品 字数	<b>a</b> _最后 更新时 间	•••	b. 划
2	竹己	oneauthor.php? authorid=1938546	onebook.php? novelid=4001734	2019- 01-21 00:29:54	完结作品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		33
3	栖见	oneauthor.php? authorid=585107	onebook.php? novelid=3525826	2018- 03-16 13:21:32	完结作品	253543.0	《桃枝气 泡》	完结	394389.0	2021- 04-12 05:01:00		17
4	竹已	oneauthor.php? authorid=1938546	onebook.php? novelid=3676002	2018- 07-24 23:08:19	VIP 作 品	355330.0	《折月 亮》	连载	318390.0	2022- 03-12 21:59:59		16
•••												
61543	千佑 Alleycat	oneauthor.php? authorid=2417786	onebook.php? novelid=6273634	2021- 10-18 12:58:04	免费作品	1242.0	《嫌疑人 每天都会 把侦探气 个半死》	连载	61162.0	2022- 03-04 01:10:18		
61544	闲云向 晚	oneauthor.php? authorid=2415488	onebook.php? novelid=4011416	2019- 01-27 23:53:35	免费作品	36.0	《我做妖 王迷弟的 那些年》	连载	163686.0	2019- 03-26 23:01:01		



61548 rows × 29 columns

[n [46]:	# 保存至文件 data.to_csv('E:\爬虫\jjwx\作品库所有作品(数据少但是优质)\作品库爬取.csv', index=False)
In [ ]:	
In [ ]:	