

Why are the dates and titles missing?

Question: Why are the dates and titles missing in some entries/keyword matches in the final output?

Answer: It's because these entries belong to firms not matched with a gvkey, and dates and titles are extracted only from the dataset of gvkey-matched firms. This is also why the gvkeys are missing.

Details

The relevant stages of the processing pipeline are Stages 3 (Firm Identification) and 4 (Keyword Identification).

Stage 3: Firm Identification (gvkey Matching)

- Output: *CC_List.csv*, a table of gvkey-matched firms with identifying information (e.g. dates, titles, report ID).

AutoSave On

CC List2020

Search

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

CutCopyPasteFormat Painter

ClipboardFontAlignmentNumberStylesCellsEditingIdeasSensitivity

Calibri11A A

B I U

Font

Align Center

Number

Styles

Cells

Editing

Ideas

Sensitivity

General

\$ % & #

Conditional FormattingTable Styles

InsertDelete Format

AutoSumFillClear

Find & Select

ShareComments

112X V English

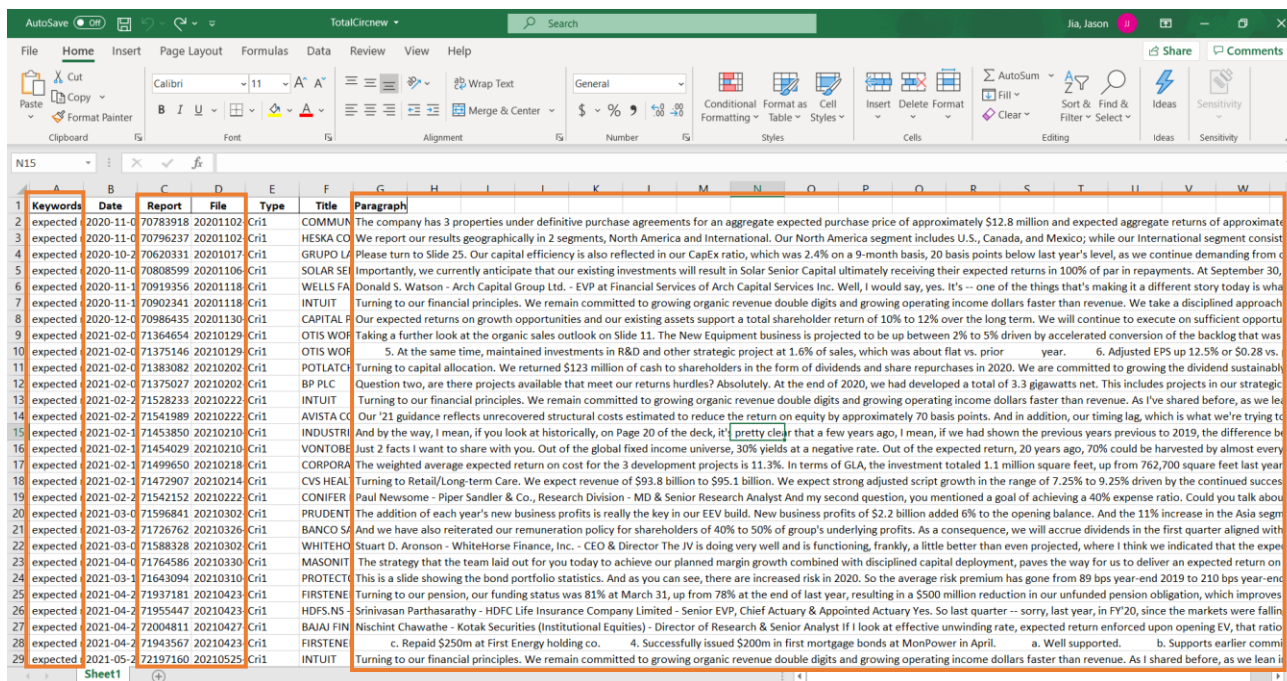
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
	PPV	TOC	Title	Subtitle	Date	Pages	Price	Contribut	Analyst	Language	Report	Collection	gkey_h	gkey_c	gkey_v	prob	gues_by_d	gues_nam	countryid	Filename			
59	N	Y	CHARLES F.CRL.N - Ev	#####		8	Subscription	THOMSON ANON	English	70894501	INV	0	209416	0.834758	0	charloer1	21	20201118-20201121	10				
60	N	Y	CHARLES F.CRL.N - Ev	#####		9	Subscription	THOMSON ANON	English	70975880	INV	0	209416	0.834758	0	charloer1	21	20201118-20201121	3,7				
61	N	Y	WISE TALE 6100.HK -	#####		12	Subscription	REFINITIV ANON	English	72885874	INV	0	203391	0.836421	0	island info	162	20201110-20201113	1,8				
62	N	Y	COMMERJ CBKG.DE -	#####		13	Subscription	THOMSON ANON	English	7090861	INV	0	213023	0.837103	0	bertrand t	76	20201102-20201105	1,8				
63	N	Y	- EVENT TI - Event Trg	#####		5	Subscription	THOMSON ANON	English	71201041	INV	0	247686	0.837427	0	a n	13	20201013-20201016	4,4				
64	N	Y	WALLENST WALLB.ST	#####		5	Subscription	THOMSON ANON	English	70636819	INV	0	213058	0.837995	0	wallenstar	194	20201021-20201024	9,4				
65	N	Y	FIRST BAN FB.P.N - Ev	#####		11	Subscription	THOMSON ANON	English	70872332	INV	0	16821	0.84	0	first banc	213	20201029-2020101	2,2				
66	N	Y	BRIDGEPO ZVO.OQ -	#####		10	Subscription	THOMSON ANON	English	70711587	INV	0	349716	0.84127	0	bridgepoint	212	20201025-20201028	4,4				
67	N	Y	DRAEGER DRWG.DE -	#####		13	Subscription	THOMSON ANON	English	70725001	INV	0	102218	0.84127	0	draegerwe	76	20201029-2020101	10,1				
68	N	Y	EXTENDIC EXE.TO - E	#####		14	Subscription	THOMSON ANON	English	70864995	INV	0	263690	0.841385	0	extendic	213	20201110-20201113	2,2				
69	N	Y	ROCKWOC ROCKb.CO	#####		18	Subscription	THOMSON ANON	English	70937787	INV	0	288571	0.842424	0	roctool sa	71	20201126-20201129	2,2				
70	N	Y	ROCKWOC ROCKb.CO	#####		19	Subscription	THOMSON ANON	English	71045193	INV	0	288571	0.842424	0	roctool sa	71	20201122-20201215	3,3				
71	N	Y	SAFETY IN SAFE.N - E	#####		15	Subscription	THOMSON ANON	English	703839763	INV	0	204654	0.842597	0	troy inc	212	20201021-20201024	6,6				
72	N	Y	SAFETY IN SAFE.N - E	#####		15	Subscription	THOMSON ANON	English	70902160	INV	0	204654	0.842597	0	troy inc	212	20201118-20201121	5,5				
73	N	Y	KCELL JC KCEL.KZ -	#####		8	Subscription	THOMSON ANON	English	70862028	INV	0	314492	0.843137	0	kcell joint	104	20201110-20201113	1,1				
74	N	Y	QGEF PAR ENAT3.SA	#####		14	Subscription	THOMSON ANON	English	70860300	INV	0	245349	0.843293	0	cma partic	30	20201110-20201113	9,9				
75	N	Y	BONAVA P BONAV.S	#####		13	Subscription	THOMSON ANON	English	70649121	INV	0	321757	0.843915	0	bonava ab	194	20201021-20201024	2,2				
76	N	Y	NETENT P NETB.ST -	#####		10	Subscription	THOMSON ANON	English	70639941	INV	0	291480	0.843915	0	netent ab	194	20201021-20201024	9,9				
77	N	Y	SOFTBANK 9434.T - E	#####		16	Subscription	THOMSON ANON	English	70785291	INV	0	223150	0.844444	0	softbank	102	20201102-20201015	21,1				
78	N	Y	SOFTBANK 9434.T - E	#####		17	Subscription	THOMSON ANON	English	70781032	INV	0	223150	0.844444	0	softbank	102	20201102-20201015	24,4				
79	N	Y	INNOQAT.INOD.OQ -	#####		8	Subscription	THOMSON ANON	English	70860316	INV	0	28777	0.844444	0	innodata	213	20201110-20201113	6,6				
80	N	Y	ITT INCOR ITT.N - Ev	#####		18	Subscription	THOMSON ANON	English	70755550	INV	0	15024	0.846164	0	united cor	37	20201029-2020101	2,2				
81	N	Y	ITT INCOR ITT.N - Ev	#####		21	Subscription	THOMSON ANON	English	70755549	INV	0	15024	0.846164	0	united cor	37	20201029-2020101	2,2				
82	N	Y	BILFINGER GBFG.DE -	#####		13	Subscription	THOMSON ANON	English	70853310	INV	0	100102	0.849415	0	bilfinger se	76	20201110-20201113	9,9				
83	N	Y	PURAVANI PURA.NS -	#####		14	Subscription	THOMSON ANON	English	70866611	INV	0	285996	0.85	0	puravanka	99	20201110-20201113	2,2				
84	N	Y	DONACO I DNA.AX -	#####		13	Subscription	THOMSON ANON	English	71076647	INV	0	217731	0.85	0	donaco int	13	20201220-20201223	2,2				
85	N	Y	WESTERN WES.N - E	#####		15	Subscription	THOMSON ANON	English	70835336	INV	0	16225	0.851801	0	western m	213	20201110-20201113	20,20				
86	N	Y	BLUEGREE BXG.N - Ev	#####		5	Subscription	THOMSON ANON	English	70894495	INV	0	285884	0.851852	0	sclegreen	98	20201118-20201121	9,9				

CC List2020

Stage 4: Keyword Identification (find paragraphs containing keywords)

Stages 4.1 and 4.2: Run *keyword_ident_1.py* and *keyword_ident_2.py*.

- Output: *TotalCircnew.xlsx*, a table of keyword matches with identifying information (e.g. report ID)



Keywords	Date	Report	File	Type	Title	Paragraph
expected	2020-11-0	70783918	20201102	Cri1	COMMUN	The company has 3 properties under definitive purchase agreements for an aggregate expected purchase price of approximately \$12.8 million and expected aggregate returns of approximate
expected	2020-11-0	70796237	20201102	Cri1	HESKA CO	We report our results geographically in 2 segments, North America and International. Our North America segment includes U.S., Canada, and Mexico; while our International segment consist
expected	2020-11-0	70620331	20201017	Cri1	GRUPO U	Please turn to Slide 25. Our capital efficiency is also reflected in our CapEx ratio, which was 2.4% on a 9-month basis, 20 basis points below last year's level, as we continue demanding from c
expected	2020-11-0	70808599	20201106	Cri1	SOLAR SE	Importantly, we currently anticipate that our existing investments will result in Solar Senior Capital ultimately receiving their expected returns in 100% of par in repayments. At September 30,
expected	2020-11-1	70919356	20201118	Cri1	WELLS FA	Donald S. Watson - Arch Capital Group Ltd. - EVP at Financial Services of Arch Capital Services Inc. Well, I would say, yes. It's -- one of the things that's making it a different story today is wha
expected	2020-11-1	70902341	20201118	Cri1	INTUIT	Turning to our financial principles. We remain committed to growing organic revenue double digits and growing operating income dollars faster than revenue. We take a disciplined approach
expected	2020-12-0	70986435	20201130	Cri1	CAPITAL P	Our expected returns on growth opportunities and our existing assets support a total shareholder return of 10% to 12% over the long term. We will continue to execute on sufficient opportu
expected	2021-02-0	71364654	20210129	Cri1	OTIS WOR	Taking a further look at the organic sales outlook on Slide 11. The New Equipment business is projected to be up between 2% to 5% driven by accelerated conversion of the backlog that was
expected	2021-02-0	71375146	20210129	Cri1	POTLATCO	5. At the same time, maintained investments in R&D and other strategic project at 1.6% of sales, which was about flat vs. prior year. 6. Adjusted EPS up 12.5% or \$0.28 vs.
expected	2021-02-0	71383082	20210202	Cri1	BP PLC	Turning to capital allocation. We returned \$123 million of cash to shareholders in the form of dividends and share repurchases in 2020. We are committed to growing the dividend sustainabl
expected	2021-02-2	71375027	20210202	Cri1	INTUIT	Question two, are there projects available that meet our returns hurdles? Absolutely. At the end of 2020, we had developed a total of 3.3 gigawatts net. This includes projects in our strategic
expected	2021-02-2	71528233	20210222	Cri1	AVISTA CO	Turning to our financial principles. We remain committed to growing organic revenue double digits and growing operating income dollars faster than revenue. As I've shared before, as we lea
expected	2021-02-2	71541989	20210222	Cri1	INDUSTRI	Our '21 guidance reflects unrecovered structural costs estimated to reduce the return on equity by approximately 70 basis points. And in addition, our timing lag, which is what we're trying to
expected	2021-02-1	71453850	20210210	Cri1	VONTONE	And by the way, I mean, if you look at historically, on Page 20 of the deck, it's pretty clear that a few years ago, I mean, if we had shown the previous years previous to 2019, the difference b
expected	2021-02-1	71454029	20210210	Cri1	CORPORA	Just 2 facts I want to share with you. Out of the global fixed income universe, 30% yields at a negative rate. Out of the expected return, 20 years ago, 70% could be harvested by almost every
expected	2021-02-1	71472907	20210214	Cri1	CVS HEAL	The weighted average expected return on cost for the 3 development projects is 11.3%. In terms of GLA, the investment totaled 1.1 million square feet, up from 762,700 square feet last year
expected	2021-02-2	71542152	20210222	Cri1	CONIFER	Turning to Retail/Long-term Care. We expect revenue of \$93.8 billion to \$95.1 billion. We expect strong adjusted script growth in the range of 7.25% to 9.25% driven by the continued succes
expected	2021-03-2	71596841	20210302	Cri1	PRUDENT	Paul Newsome - Piper Sandler & Co., Research Division - MD & Senior Research Analyst And my second question, you mentioned a goal of achieving a 40% expense ratio. Could you talk about
expected	2021-03-2	71726762	20210326	Cri1	BANCO SA	The addition of each year's new business profits is really the key in our EEV build. New business profits of \$2.2 billion added 6% to the opening balance. And the 11% increase in the Asia segm
expected	2021-03-0	71588328	20210302	Cri1	WHITEHO	And we have also reiterated our remuneration policy for shareholders of 40% to 50% of group's underlying profits. As a consequence, we will accrue dividends in the first quarter aligned with
expected	2021-04-0	71764586	20210330	Cri1	MASONIT	Stuart D. Aronson - WhiteHorse Finance, Inc. - CEO & Director The JV is doing very well and is functioning, frankly, a little better than even projected, where I think we indicated that the expe
expected	2021-04-2	71937181	20210423	Cri1	FIRSTENE	The strategy that the team laid out for you today to achieve our planned margin growth combined with disciplined capital deployment, paves the way for us to deliver an expected return on
expected	2021-04-2	71955447	20210423	Cri1	PROTECTO	This is a slide showing the bond portfolio statistics. And as you can see, there are increased risk in 2020. So the average risk premium has gone from 89 bps year-end 2019 to 210 bps year-end
expected	2021-04-2	72004811	20210427	Cri1	HDFC.NS	Turning to our pension, our funding status was 81% at March 31, up from 78% at the end of last year, resulting in a \$500 million reduction in our unfunded pension obligation, which improves
expected	2021-04-2	71943567	20210423	Cri1	BAJAJ FIN	Srinivasan Parthasarathy - HDFC Life Insurance Company Limited - Senior EVP, Chief Actuary & Appointed Actuary Yes. So last quarter -- sorry, last year, in FY'20, since the markets were fallin
expected	2021-05-2	72197160	20210525	Cri1	FIRSTENE	Nischint Chawathe - Kotak Securities (Institutional Equities) - Director of Research & Senior Analyst If I look at effective unwinding rate, expected return enforced upon opening EV, that ratio
expected	2021-05-2	72197160	20210525	Cri1	INTUIT	c. Repaid \$250m at First Energy holding co. 4. Successfully issued \$200m in first mortgage bonds at MonPower in April. a. Well supported. b. Supports earlier commi
expected	2021-05-2	72197160	20210525	Cri1	INTUIT	Turning to our financial principles. We remain committed to growing organic revenue double digits and growing operating income dollars faster than revenue. As I shared before, as we lean i

Stage 4.3: Run *mergclean.do*.

- Inputs: *CC_List.csv*, *TotalCircnew.xlsx*
- Purpose: Merge the 2 types of datasets using report ID to get a consolidated dataset with keyword matches, gvkeys and identifying information. Keyword matches in *TotalCircnew.xlsx* with report IDs that don't appear in *CC_List.csv* are still retained.
- Output: *cric1_newtotal.xlsx*
 - Columns in blue come from *CC_List.csv* and columns in orange come from *TotalCircnew.xlsx*.
 - Dates and titles are missing in about 2.5%-3% of the entries. This happens in both the old (2001-2020) and new (2020-2021) compilations.
 - After comparing the report IDs, it is found that most (>99%) of the missing dates in ... \Jason-Kilian\Conference Calls with missing dates\paragraph_datemissing.xlsx belong to this pool of incomplete entries.

	Keywords	Paragraph	Date	gvkey	Title	Subtitle	gvkey_h	gvkey_c	prob	gues_by	gues_nam	countryid	Report	File
5036	cost of de	Our loan-	2021-02-1	332533	VASTNED	VASN.AS -	0	0	1	0	fastned bv	143	71454090	20210210-20210213_9.csv
5037	cost of de	And the th	2020-10-2	332739	AIRTEL AF	AAF.L - Ev	332739	332739	1	0	airtel afric	212	70648749	20201021-20210204_1.csv
5038	interest r	Jaldeep K	2020-10-2	332739	AIRTEL AF	AAF.L - Ev	332739	332739	1	0	airtel afric	212	70648749	20201021-20210204_1.csv
5039	cost of de	Pier Falcic	2021-01-2	332739	AIRTEL AF	AAF.L - Ev	332739	332739	1	0	airtel afric	212	71347762	20210129-20210201_2.csv
5040	hurdle rat	Kim Henr	2021-01-2	333866	EQT AB	EQTAB.ST	333866	333866	1	0	eqt ab	194	71314316	20210125-20210128_11.csv
5041	cost of de	In Februa	2021-05-0	333885	TEAMVIEW	TMV.DE - I	333885	333885	1	0	teamview	76	72069851	20210501-20210504_7.csv
5042	cost of de	We contin	2021-03-0	338557	JDE PEETS	JDEP.AS -	0	0	1	0	jde peets	143	71719133	20210306-20210309_6.csv
5043	cost of de	And these	2021-08-0	338557	JDE PEETS	JDEP.AS -	0	0	1	0	jde peets	143	72688887	20210801-20210804_2.csv
5044	interest r	In Septen	2020-11-1	339015	METSO OL	MOCORP.	339015	339015	1	0	metso out	69	70866443	20201110-20201113_17.csv
5045	interest r	Moving fo	2021-04-2	339015	METSO OL	MOCORP.	339015	339015	1	0	metso out	69	71937152	20210423-20210426_3.csv
5046	cost of de	As regard	2021-05-1	339361	MINDSPA	MINS.NS -	339361	339361	1	0	mindspaci	93	72137468	20210517-20210520_17.csv
5047	cost of de	On the de	2021-08-1	339361	MINDSPA	MINS.NS -	339361	339361	1	0	mindspaci	93	72778784	20210813-20210816_3.csv
5048	cost of ca	While Por	2021-05-1	344052	CLEVER LE	CLVR.OQ -	0	0	1	0	everfuel a	55	72766261	20210517-20210520_16.csv
5049	interest r	The financ	2021-08-1	344052	CLEVER LE	CLVR.OQ -	0	0	1	0	everfuel a	55	72766431	20210809-20210812_4.csv
5050	roic	Moving or	2020-11-1	345256	AERIS IND	AERIS.SA -	0	0	1	0	aeris indu	30	71607515	20201110-20201113_12.csv
5051	roic	Througho	2021-02-1	345256	AERIS IND	AERIS.SA -	0	0	1	0	aeris indu	30	71607261	20210210-20210213_18.csv
5052	return on	Bruno Lol	2021-05-1	345256	AERIS IND	AERIS.SA -	0	0	1	0	aeris indu	30	72125032	20210513-20210516_13.csv
5053	roic	So the se	2021-05-1	345256	AERIS IND	AERIS.SA -	0	0	1	0	aeris indu	30	72125032	20210513-20210516_13.csv
5054	roic	Moving or	2021-08-1	345256	AERIS IND	AERIS.SA -	0	0	1	0	aeris indu	30	72975318	20210809-20210812_9.csv
5055	cost of de	Unidentif	2021-08-1	346069	ANTONY V	ANTO.NS	346069	346069	1	0	antony we	93	72754079	20210809-20210812_3.csv
5056	interest r	Alexandre											72160309	20210517-20210520_4.csv
5057	interest r	I won't sp											70697904	20201025-20201028_8.csv
5058	interest r	Interst ir											70630902	20201021-20201024_11.csv
5059	discount r	The weigh											70851128	20201110-20201113_15.csv
5060	irr	On Slide 1											71518754	20210222-20210225_28.csv
5061	interest r	So in term											70694227	20201025-20201028_8.csv
5062	interest r	We endec											70756326	20201102-20201105_32.csv
5063	interest r	In the nea											70630196	20201021-20201024_11.csv
5064	interest r	At the enc											70834794	20201110-20201113_23.csv
5065	interest r	On Page 4											71205011	20201029-20201101_3.csv
5066	interest r	Taken tog											70860640	20201110-20201113_16.csv
5067	interest r	Net intere											72669530	20210801-20210804_20.csv
5068	interest r	1, Se											72669558	20210801-20210804_21.csv
5069	expected	Donald S.											70919356	20201118-20201121_7.csv
5070	irr	The third											71065345	20201216-20201219_2.csv
5071	interest r	Michael A											70968864	20201130-20201203_8.csv
5072	interest r	Unidentif											71070700	20201216-20201219_2.csv