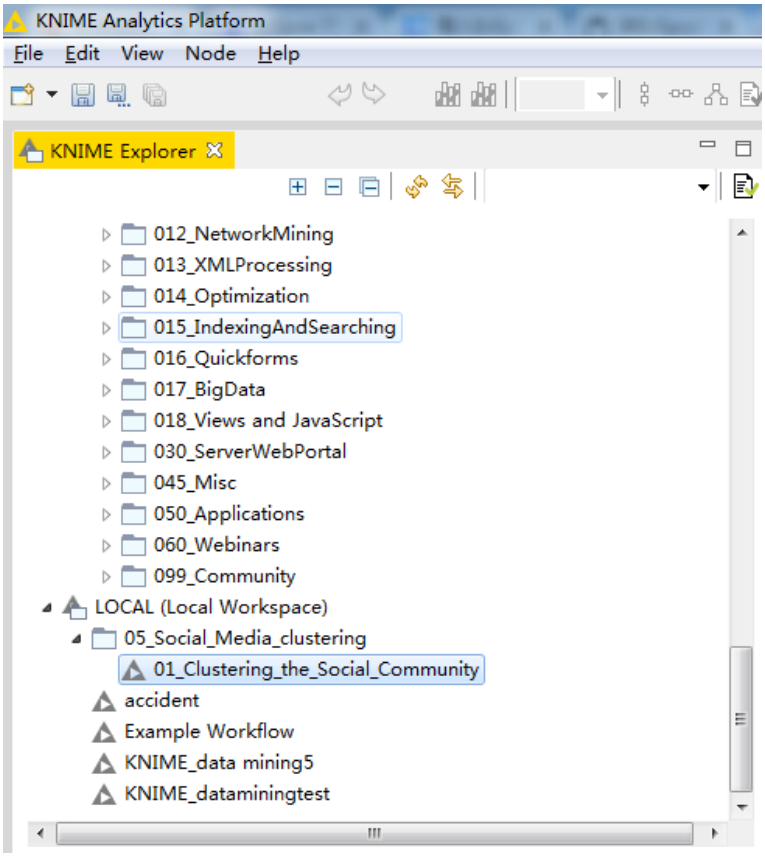
TP 4

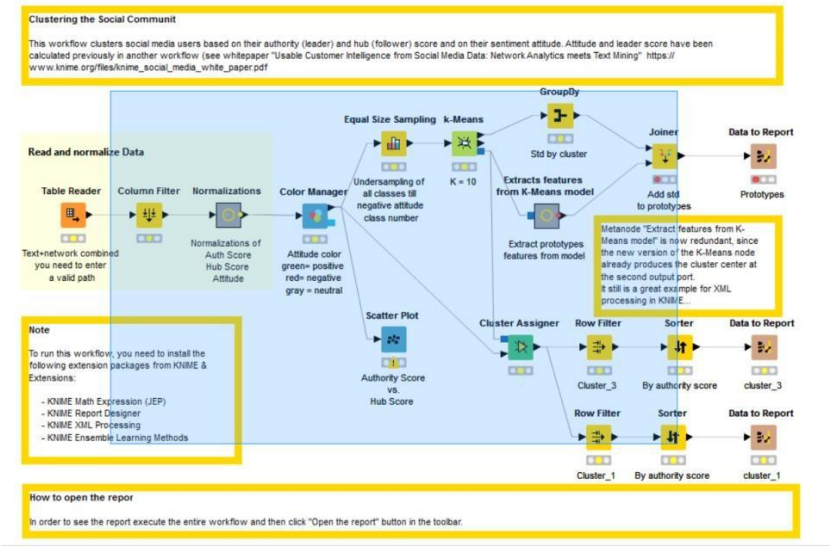
Applications in Knime

(BI 1) YAO Zeliang
(BI 2) ZHANG Meng

Exercise 1. Social Media Sentiment Analysis

1. Import the workflow Social_Media_Clustering





2. Analyze the data and the workflow by reading all the node's descriptions (some nodes are meta-nodes).

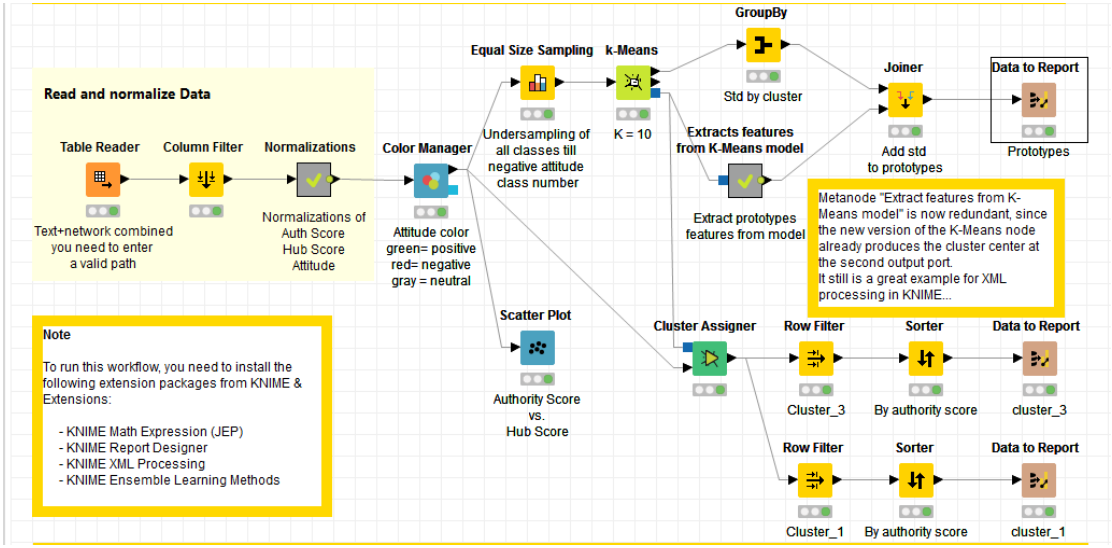
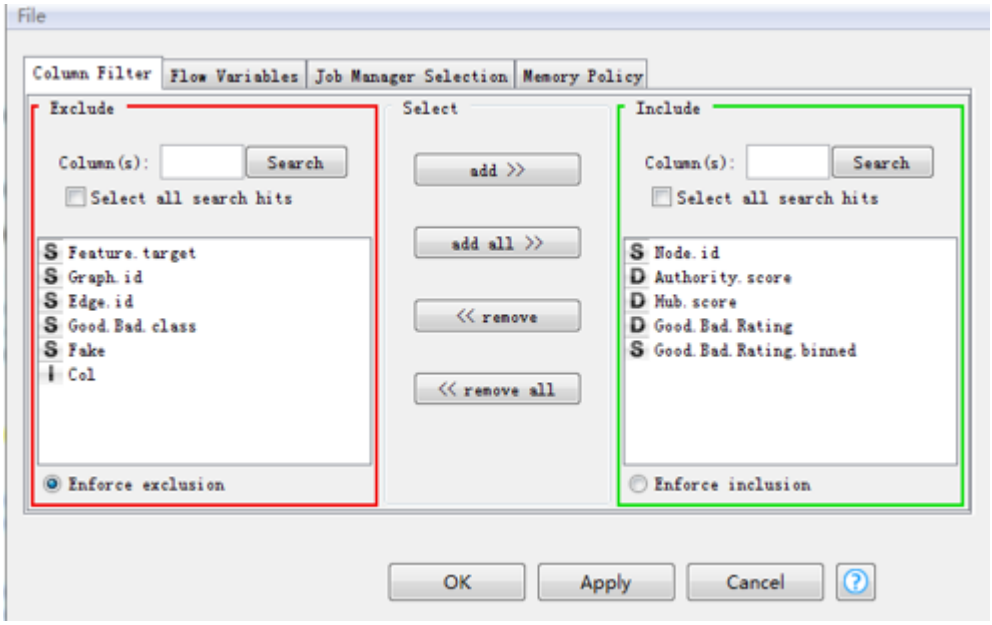
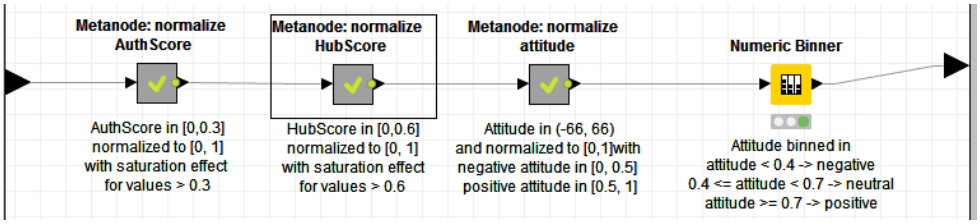


Table Reader: reads the original data. It retains all meta information such as domain, properties, colors, size.

Column Filter: This node allows columns to be filtered from the input table while only the remaining columns are passed to the output table. Within the dialog, columns can be moved between the Include and Exclude list.



Normalizations: normalize Auth Score, Hub Score, attitude and classifie the data into three categories (negative, neutral, positive) according to the value of Authority.score and Hub.score.



Color Manager: Colors can be assigned for either nominal (possible values have to be available) or numeric columns (with lower and upper bounds). If these bounds are not available, a '?' is provided as a minimum and maximum value. The values are then computed during execute. If a column attribute is selected, the color can be changed with the color chooser.

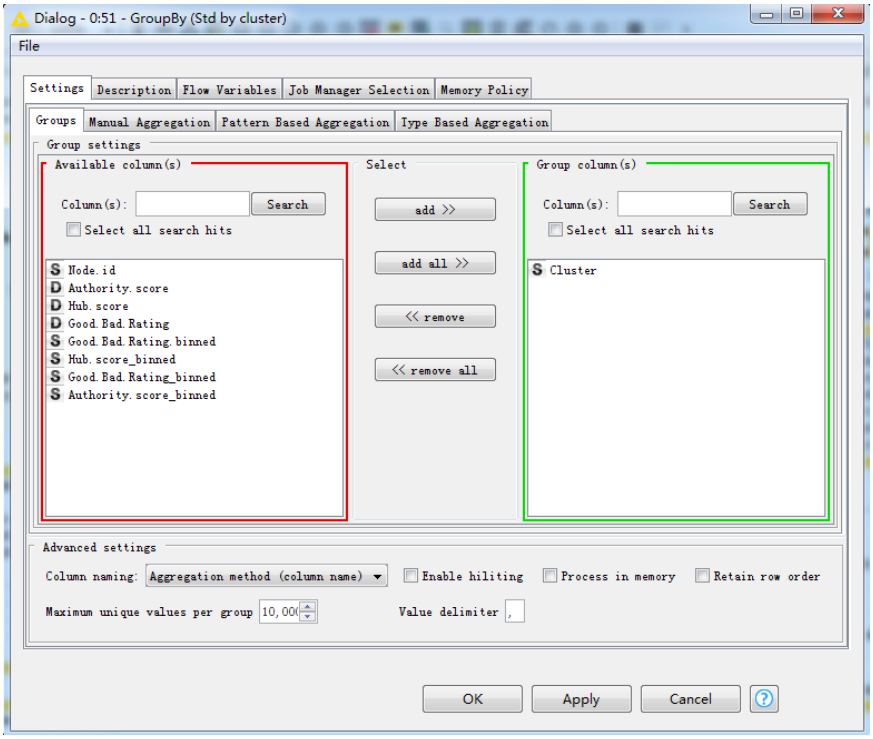
Equal Size Sampling: keeps all the rows of the minority class, and remove randomly the rows of the majority classes, making the majority classes be the same size as the minority class.

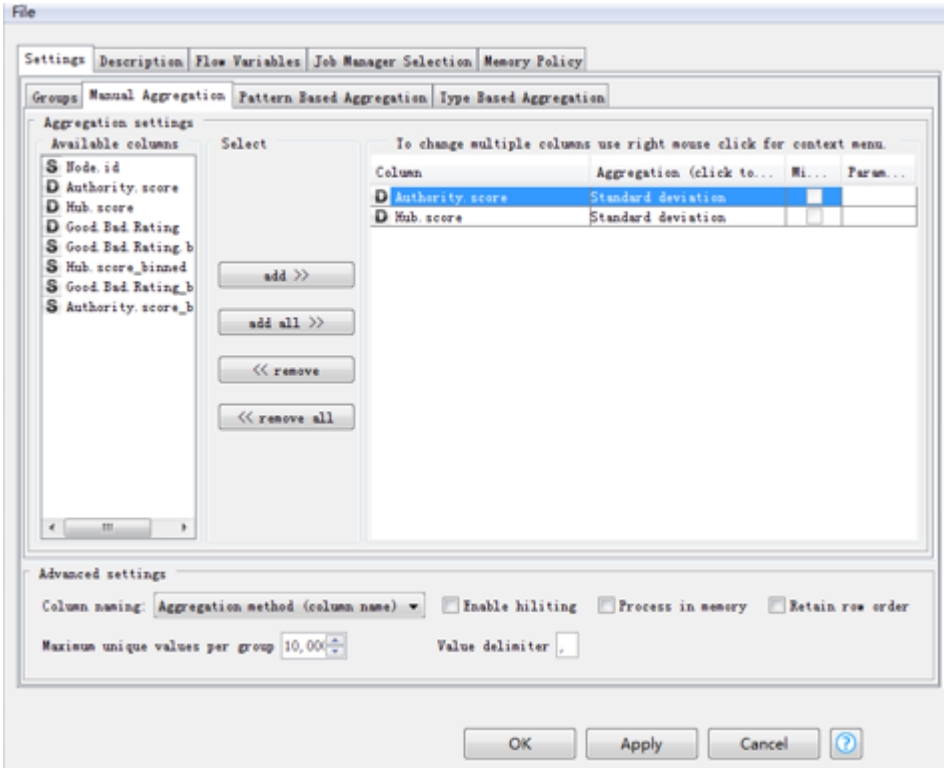
The node will remove random rows belonging to the majority classes. The rows returned by this node will contain all records from the minority class(es) and a random sample from each of the majority classes, whereby each sample contains as many objects as the minority class contains.

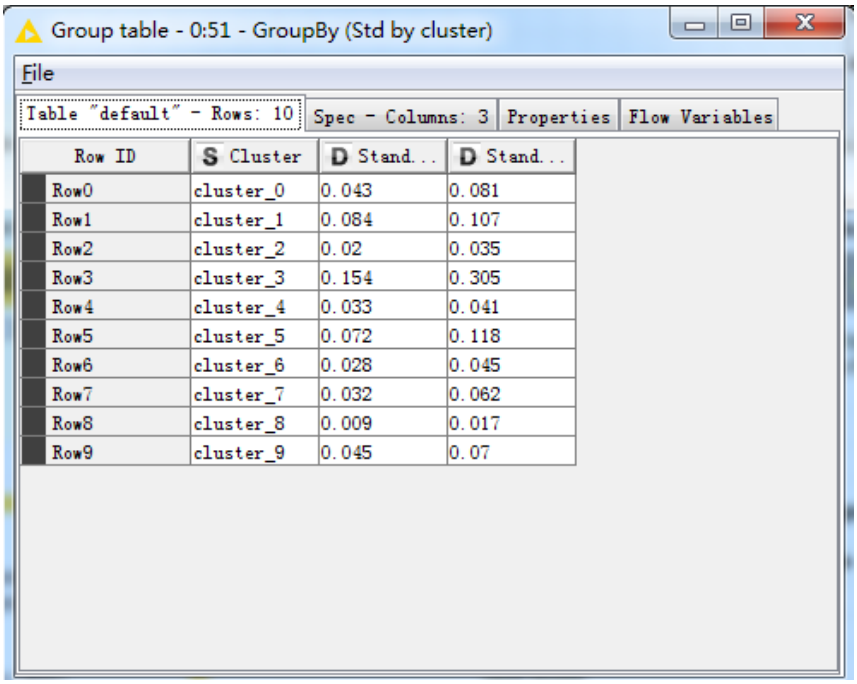
K-means: This node outputs the cluster centers for a predefined number of clusters (no dynamic number of clusters). K-means performs a crisp clustering that assigns a

data vector to exactly one cluster. The algorithm terminates when the cluster assignments do not change anymore. The clustering algorithm uses the Euclidean distance on the selected attributes. The data is not normalized by the node.
(K=10) clusters the sampling data into 10 clusters by K-means algorithm.

GroupBy: Groups the rows of a table by the unique values in the selected group columns. A row is created for each unique set of values of the selected group column. The remaining columns are aggregated based on the specified aggregation settings. The output table contains one row for each unique value combination of the selected group columns.
Authority.score and Hub.score was aggregated by Standard deviation method (in Manual Aggregation tab).



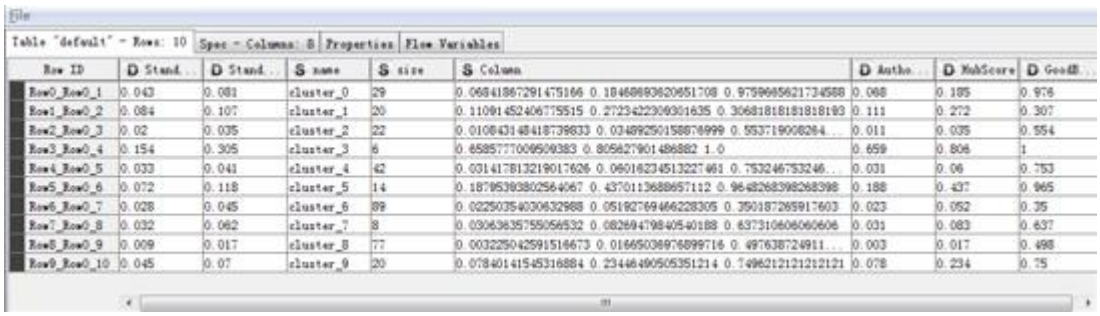




Row ID	Cluster	Stand...	Stand...
Row0	cluster_0	0.043	0.081
Row1	cluster_1	0.084	0.107
Row2	cluster_2	0.02	0.035
Row3	cluster_3	0.154	0.305
Row4	cluster_4	0.033	0.041
Row5	cluster_5	0.072	0.118
Row6	cluster_6	0.028	0.045
Row7	cluster_7	0.032	0.062
Row8	cluster_8	0.009	0.017
Row9	cluster_9	0.045	0.07

Extracts features from K-Means model: takes the range of each K-means cluster as a model.

Joiner: This node joins two tables in a database-like way. The join is based on the joining columns of both tables.



Row ID	Stand...	Stand...	name	size	Column	Autho...	HubScore	Goodl...
Row0_Row0_1	0.043	0.081	cluster_0	29	0.06841867291475166 0.18468693620651708 0.9759685621734588	0.068	0.185	0.976
Row1_Row0_2	0.084	0.107	cluster_1	20	0.11091452406775515 0.2723422306901635 0.3068181818181818193	0.111	0.272	0.307
Row2_Row0_3	0.02	0.035	cluster_2	22	0.010843148418739833 0.034892901588769999 0.553719008264...	0.011	0.035	0.554
Row3_Row0_4	0.154	0.305	cluster_3	6	0.6585777009509383 0.8056279014868821.0	0.659	0.806	1
Row4_Row0_5	0.033	0.041	cluster_4	42	0.03141781321901762 0.06016234513227461 0.753246753246...	0.031	0.06	0.753
Row5_Row0_6	0.072	0.118	cluster_5	14	0.18795393802564067 0.43701136886571120.9648268398268398	0.188	0.437	0.965
Row6_Row0_7	0.028	0.045	cluster_6	89	0.022503540306329880.05192769466228305 0.3501872659176030	0.023	0.052	0.35
Row7_Row0_8	0.032	0.062	cluster_7	8	0.030636357550568320.08269479840540188 0.6373106060606606	0.031	0.083	0.637
Row8_Row0_9	0.009	0.017	cluster_8	77	0.003225042591516673 0.016650389768999716 0.497638724911...	0.003	0.017	0.498
Row9_Row0_10	0.045	0.07	cluster_9	20	0.07840141545316884 0.23446490505351214 0.7496212121212121	0.078	0.234	0.75

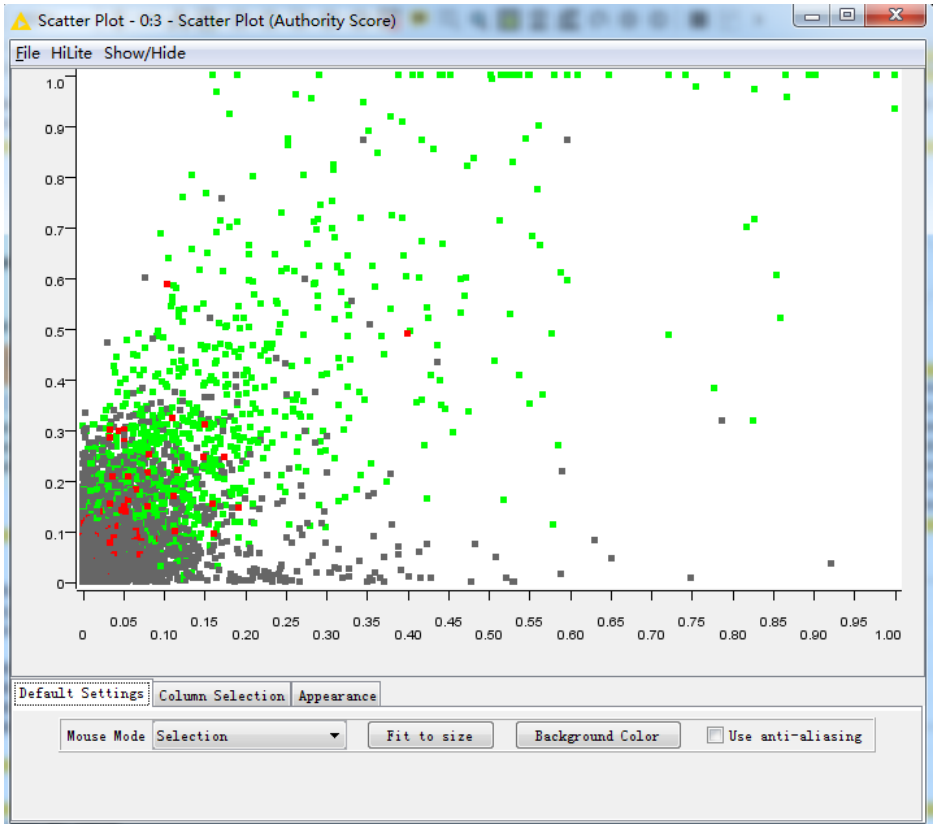
The second branch:

Cluster Assigner: assigns the original data to the set of prototypes, which are obtained by the k-means clustering. Each data point is assigned to its nearest prototype.

Two Row Filters: filter *Cluster_3* and *Cluster_1* respectively from the output of Cluster Assigner, and the two **Sorter**s sort the results by authority.score.

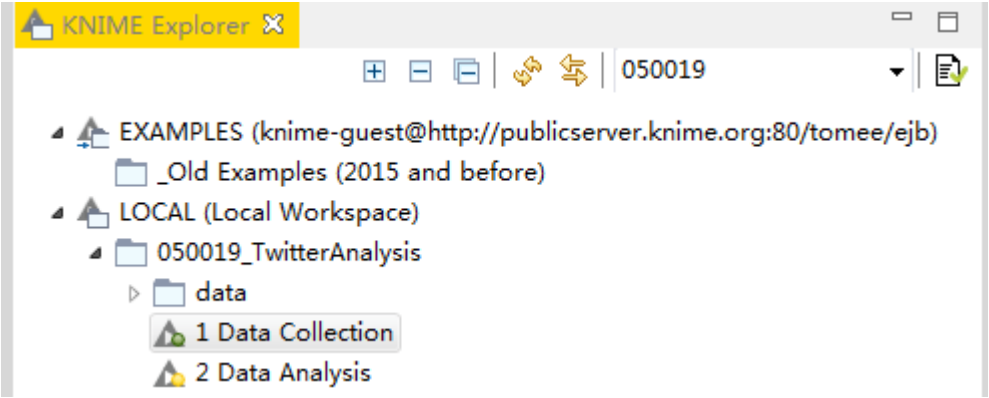
The third branch:

Scatter Plot: Creates a scatterplot of two selectable attributes. Then each datapoint is displayed as a dot at its corresponding place, dependent on its values of the selected attributes. The dots are displayed in the color defined by the Color Manager, the size defined by the Size Manager, and the shape defined by the Shape Manager.



Exercise 2. Twitter Analysis

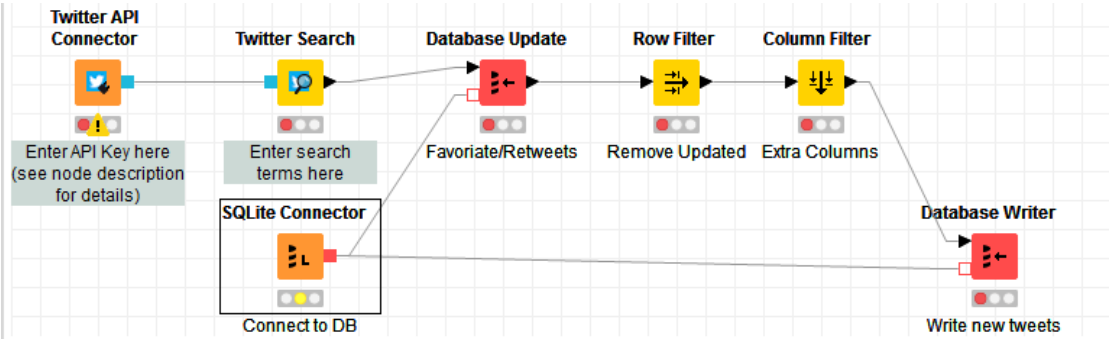
2.1 Import the Twitter Analysis (050019_TwitterAnalysis) workflow and the tweets data.



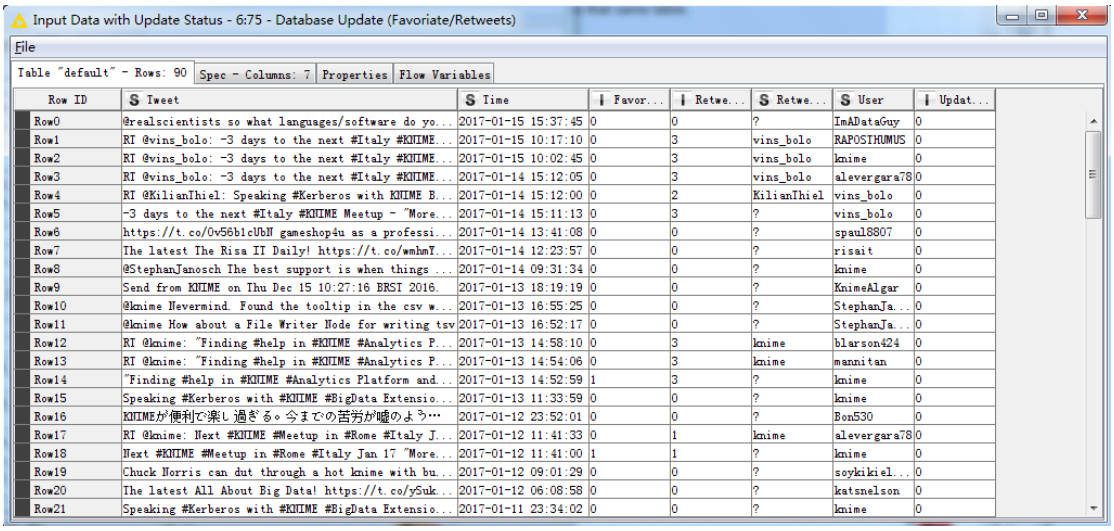
The twitterAnalysis includes two workflow one is for the Data Collection, and the another is for the Data Analysis.

2.2 Twitter Data Collection

The following workflow is designed to collect and store a sample of tweets on a particular search term. With each execution of the workflow, tweets are collected, favorite and retweet numbers are updated in existing records in a SQLite Database and new tweets are written to that same table.

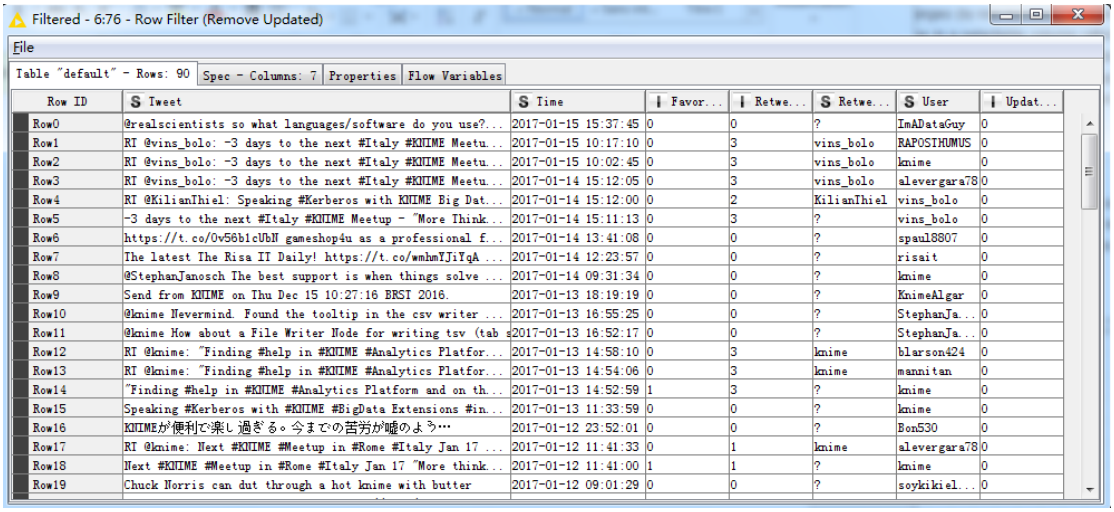


SQLite Connector: creates a connection to a SQLite database file via its JDBC driver.
Database Update: updates the data rows in the database with the data values from the input tables. The results is like below:



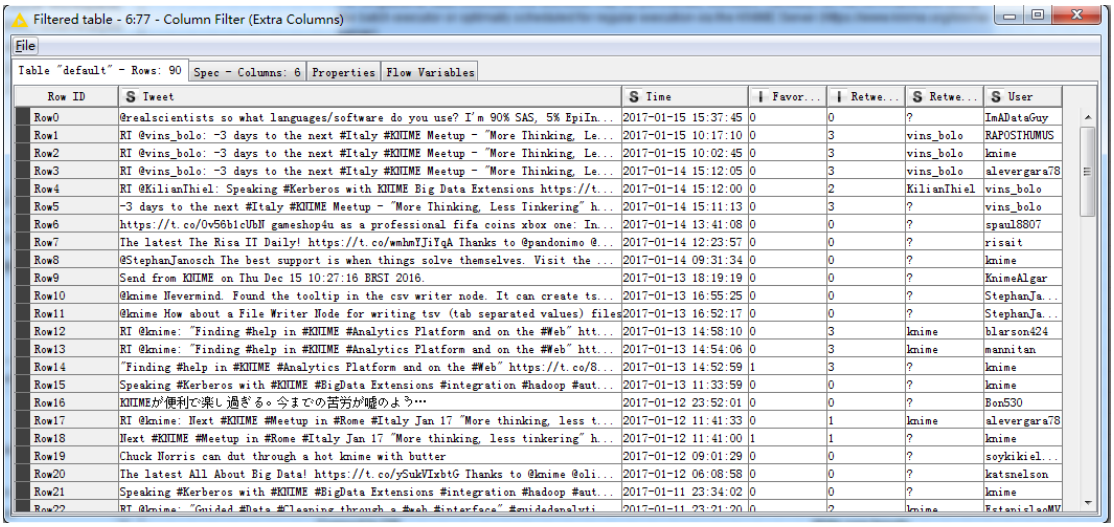
Row ID	Tweet	Time	Favor...	Retwe...	Retwe...	User	Updat...
Row0	@realscientists so what languages/software do yo...	2017-01-15 15:37:45	0	0	?	ImADataGuy	0
Row1	RT @vins_bolo: -3 days to the next #Italy #KNIME...	2017-01-15 10:17:10	0	3	vins_bolo	RAPOSTHUMUS	0
Row2	RT @vins_bolo: -3 days to the next #Italy #KNIME...	2017-01-15 10:02:45	0	3	vins_bolo	knime	0
Row3	RT @vins_bolo: -3 days to the next #Italy #KNIME...	2017-01-14 15:12:05	0	3	vins_bolo	alevergara78	0
Row4	RT @KilianThiel: Speaking #Kerberos with KNIME B...	2017-01-14 15:12:00	0	2	KilianThiel	vins_bolo	0
Row5	-3 days to the next #Italy #KNIME Meetup - "More...	2017-01-14 15:11:13	0	3	?	vins_bolo	0
Row6	https://t.co/0v56b1cUbN gameshop4u as a professi...	2017-01-14 13:41:08	0	0	?	spaul8807	0
Row7	The latest The Risa IT Daily! https://t.co/wmhmY...	2017-01-14 12:23:57	0	0	?	risait	0
Row8	@StephanJanosch The best support is when things ...	2017-01-14 09:31:34	0	0	?	knime	0
Row9	Send from KNIME on Thu Dec 15 10:27:16 BRST 2016.	2017-01-13 18:19:19	0	0	?	KnimeAlgar	0
Row10	@knime Nevermind. Found the tooltip in the csv w...	2017-01-13 16:55:25	0	0	?	StephanJa...	0
Row11	@knime How about a File Writer Node for writing tsv	2017-01-13 16:52:17	0	0	?	StephanJa...	0
Row12	RT @knime: "Finding #help in #KNIME #Analytics P...	2017-01-13 14:58:10	0	3	knime	blarson424	0
Row13	RT @knime: "Finding #help in #KNIME #Analytics P...	2017-01-13 14:54:06	0	3	knime	mannitan	0
Row14	"Finding #help in #KNIME #Analytics Platform and...	2017-01-13 14:52:59	1	3	?	knime	0
Row15	Speaking #Kerberos with #KNIME #BigData Extensio...	2017-01-13 11:33:59	0	0	?	knime	0
Row16	KNIMEが便利で楽し過ぎる。今までの苦労が嘘のよう…	2017-01-12 23:52:01	0	0	?	Bon530	0
Row17	RT @knime: Next #KNIME #Meetup in #Rome #Italy J...	2017-01-12 11:41:33	0	1	knime	alevergara78	0
Row18	Next #KNIME #Meetup in #Rome #Italy Jan 17 "More...	2017-01-12 11:41:00	1	1	?	knime	0
Row19	Chuck Norris can dut through a hot knime with bu...	2017-01-12 09:01:29	0	0	?	soykikiel...	0
Row20	The latest All About Big Data! https://t.co/ySuk...	2017-01-12 06:08:58	0	0	?	katsnelson	0
Row21	Speaking #Kerberos with #KNIME #BigData Extensio...	2017-01-11 23:34:02	0	0	?	knime	0

Row Filter: allows for row filtering according to certain criteria. It can include or exclude: certain ranges (by row number), rows with a certain row ID, and rows with a certain value in a selectable column (attribute). Here it is for removing each row' Update, the results like following:



Row ID	Tweet	Time	Favor...	Retwe...	Retwe...	User	Updat...
Row0	@realscientists so what languages/software do you use?...	2017-01-15 15:37:45	0	0	?	ImADataGuy	0
Row1	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetu...	2017-01-15 10:17:10	0	3	vins_bolo	RAPOSTHUMUS	0
Row2	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetu...	2017-01-15 10:02:45	0	3	vins_bolo	knime	0
Row3	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetu...	2017-01-14 15:12:05	0	3	vins_bolo	alevergara78	0
Row4	RT @KilianThiel: Speaking #Kerberos with KNIME Big Dat...	2017-01-14 15:12:00	0	2	KilianThiel	vins_bolo	0
Row5	-3 days to the next #Italy #KNIME Meetup - "More Think...	2017-01-14 15:11:13	0	3	?	vins_bolo	0
Row6	https://t.co/0v56b1cUbN gameshop4u as a professional f...	2017-01-14 13:41:08	0	0	?	spaul8807	0
Row7	The latest The Risa IT Daily! https://t.co/wmhmYJiYqA ...	2017-01-14 12:23:57	0	0	?	risait	0
Row8	@StephanJanosch The best support is when things solve ...	2017-01-14 09:31:34	0	0	?	knime	0
Row9	Send from KNIME on Thu Dec 15 10:27:16 BRST 2016.	2017-01-13 18:19:19	0	0	?	KnimeAlgar	0
Row10	@knime Nevermind. Found the tooltip in the csv writer ...	2017-01-13 16:55:25	0	0	?	StephanJa...	0
Row11	@knime How about a File Writer Node for writing tsv (tab s	2017-01-13 16:52:17	0	0	?	StephanJa...	0
Row12	RT @knime: "Finding #help in #KNIME #Analytics Platfor...	2017-01-13 14:58:10	0	3	knime	blarson424	0
Row13	RT @knime: "Finding #help in #KNIME #Analytics Platfor...	2017-01-13 14:54:06	0	3	knime	mannitan	0
Row14	"Finding #help in #KNIME #Analytics Platform and on th...	2017-01-13 14:52:59	1	3	?	knime	0
Row15	Speaking #Kerberos with #KNIME #BigData Extensions #in...	2017-01-13 11:33:59	0	0	?	knime	0
Row16	KNIMEが便利で楽し過ぎる。今までの苦労が嘘のよう…	2017-01-12 23:52:01	0	0	?	Bon530	0
Row17	RT @knime: Next #KNIME #Meetup in #Rome #Italy Jan 17 ...	2017-01-12 11:41:33	0	1	knime	alevergara78	0
Row18	Next #KNIME #Meetup in #Rome #Italy Jan 17 "More think...	2017-01-12 11:41:00	1	1	?	knime	0
Row19	Chuck Norris can dut through a hot knime with butter	2017-01-12 09:01:29	0	0	?	soykikiel...	0

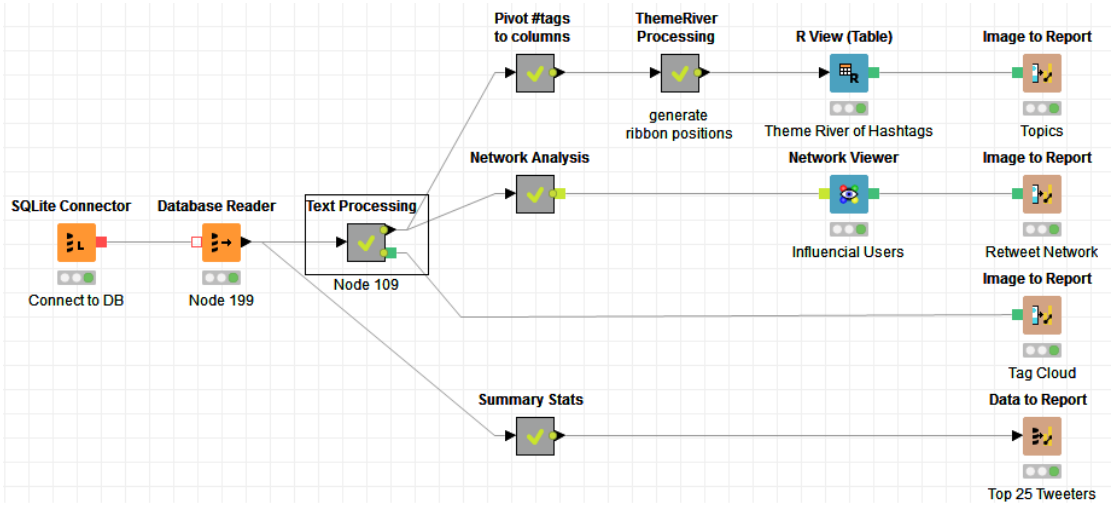
Column Filter: allows columns to be filtered from the input table while only the remaining columns are passed to the output table. Here is for getting the remaining columns, the results like following:



Row ID	Tweet	Time	Favor...	Retwe...	Retwe...	User
Row0	@realscientists so what languages/software do you use? I'm 90% SAS, 5% EpiIn...	2017-01-15 15:37:45	0	0	?	ImADataGuy
Row1	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetup - "More Thinking, Le...	2017-01-15 10:17:10	0	3	vins_bolo	RAPOSTHUMUS
Row2	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetup - "More Thinking, Le...	2017-01-15 10:02:45	0	3	vins_bolo	knime
Row3	RT @vins_bolo: -3 days to the next #Italy #KNIME Meetup - "More Thinking, Le...	2017-01-14 15:12:05	0	3	vins_bolo	alevergara78
Row4	RT @KilianThiel: Speaking #Kerberos with KNIME Big Data Extensions https://t...	2017-01-14 15:12:00	0	2	KilianThiel	vins_bolo
Row5	-3 days to the next #Italy #KNIME Meetup - "More Thinking, Less Tinkering" h...	2017-01-14 15:11:13	0	3	?	vins_bolo
Row6	https://t.co/Ov56b1cUbH gameshop4u as a professional fifa coins xbox one: In...	2017-01-14 13:41:08	0	0	?	spaul8807
Row7	The latest The Risa II Daily! https://t.co/wmhnYJiYqA Thanks to @pandonimo @...	2017-01-14 12:23:57	0	0	?	risait
Row8	@StephanJanosch The best support is when things solve themselves. Visit the ...	2017-01-14 09:31:34	0	0	?	knime
Row9	Send from KNIME on Thu Dec 15 10:27:16 BRST 2016.	2017-01-13 18:19:19	0	0	?	KnimeAlgar
Row10	@knime Nevermind. Found the tooltip in the csv writer node. It can create ts...	2017-01-13 16:55:25	0	0	?	StephanJa...
Row11	@knime How about a File Writer Node for writing tsv (tab separated values) files	2017-01-13 16:52:17	0	0	?	StephanJa...
Row12	RT @knime: "Finding #help in #KNIME #Analytics Platform and on the #Web" htt...	2017-01-13 14:58:10	0	3	knime	blarson424
Row13	RT @knime: "Finding #help in #KNIME #Analytics Platform and on the #Web" htt...	2017-01-13 14:54:06	0	3	knime	mannitan
Row14	"Finding #help in #KNIME #Analytics Platform and on the #Web" https://t.co/8...	2017-01-13 14:52:59	1	3	?	knime
Row15	Speaking #Kerberos with #KNIME #BigData Extensions #integration #hadoop #aut...	2017-01-13 11:33:59	0	0	?	knime
Row16	KNIMEが便利で楽し過ぎる。今までの苦労が嘘のよう…	2017-01-12 23:52:01	0	0	?	Bon530
Row17	RT @knime: Next #KNIME #Meetup in #Rome #Italy Jan 17 "More thinking, less t...	2017-01-12 11:41:33	0	1	knime	alevergara78
Row18	Next #KNIME #Meetup in #Rome #Italy Jan 17 "More thinking, less tinkering" h...	2017-01-12 11:41:00	1	1	?	knime
Row19	Chuck Norris can dut through a hot knime with butter	2017-01-12 09:01:29	0	0	?	soykikiel...
Row20	The latest All About Big Data! https://t.co/ySukVIxbtG Thanks to @knime @oli...	2017-01-12 06:08:58	0	0	?	katsnelson
Row21	Speaking #Kerberos with #KNIME #BigData Extensions #integration #hadoop #aut...	2017-01-11 23:34:02	0	0	?	knime
Row22	RT @knime: "Guided #Data #Cleaning through a #web #interface" #guidedanalyti...	2017-01-11 23:21:20	0	2	knime	EstanislaoMV

Database Writer: establishes and opens a database access connection to which the entire input table is written to.

2.3 Data Analysis Workflow



This workflow examines a sample of tweets from the days surrounding the Scottish referendum for independence in 2014 [1]. After reading the data from a local database, basic text processing is used to extract hashtags from the dataset and term frequencies calculated and used to build a tag cloud. Subsequently, hashtag trending is examined over time, with a notable post-election surge in the #the45 movement. Additionally, network analysis is performed in order to look at the most influential social graph surrounding this issue.

SQLite Connector: creates a connection to a SQLite database file via its JDBC driver.

Database Reader: establishes and opens a database access connection to read data from. The following is the data we read:

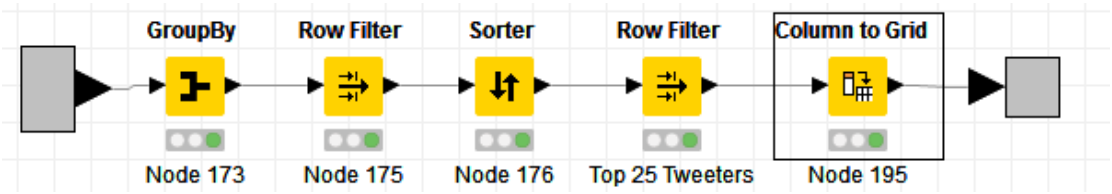
Data from Database - 7:199 - Database Reader

File

Table "database" - Rows: 25000 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	User	Tweet	Time	Retwe...	Favor...	Retwe...
Row0	Beirutiyat	Looks like #Russia and @RT_com are eager for Scottish,...	2014-09-12 00:25:49	?	0	0
Row1	JTWright71	RT @jsteve372: Nationalists will try to convince you L...	2014-09-12 00:24:49	jsteve372	0	4
Row2	Stockpile...	Meet The English People Voting For Scottish Independen...	2014-09-12 00:24:23	?	0	0
Row3	RealTimeHack	Scottish independence support slips 1 week before refe...	2014-09-12 00:24:16	?	0	0
Row4	BilalGohar7	RT @mehdifoundation: MFI UK' s Stance on Scottish Inde...	2014-09-12 00:23:59	mehdifoun...	0	483
Row5	meriemk	@crowd42 http://t.co/5yxVqPgiCp	2014-09-12 00:23:09	?	0	0
Row6	tomhbatty	Poll: support for Scottish independence cools as refer...	2014-09-12 00:22:23	?	0	0
Row7	muirkate	RT @iankatz1000: Kissinger: Support for Scottish indep...	2014-09-12 00:21:31	iankatz1000	0	23
Row8	DrudgeLive	ONE WEEK TO VOTE...: http://t.co/oRp1A72S1x	2014-09-12 00:21:28	?	0	0
Row9	15Seaton	RT @AlanRoden: Tomorrow's Scottish Daily Mail: Letter ...	2014-09-12 00:21:21	AlanRoden	0	26
Row10	ccmilne	RT @AlanRoden: Tomorrow's Scottish Daily Mail: Letter ...	2014-09-12 00:21:15	AlanRoden	0	26
Row11	MangoJelly91	RT @GdnScotland: Scottish independence: 97% register t...	2014-09-12 00:20:47	GdnScotland	0	68
Row12	WeKnowSFA	RT @Under_Radar_Mag: Scotland Week: James Yorkston on ...	2014-09-12 00:20:22	Under_Rad...	0	5
Row13	iAmJayPlatt	Poll: support for Scottish independence cools as refer...	2014-09-12 00:20:01	?	0	0
Row14	McAdeyemix	Poll: support for Scottish independence cools as refer...	2014-09-12 00:19:49	?	0	0
Row15	iamidunnu	Poll: support for Scottish independence cools as refer...	2014-09-12 00:19:43	?	0	0
Row16	Jcqdnld	RT @cad_11: BBC bias corruption petition. https://t.co...	2014-09-12 00:19:41	cad_11	0	8
Row17	newstalk_LL	Scottish independence: yes may be the result Alex Salm...	2014-09-12 00:18:05	?	0	0
Row18	smuji	True RT @iankatz1000: Kissinger: Support for Scottish...	2014-09-12 00:18:02	?	0	0
Row19	amangelone	RT @lionelbarber: Reader alert: We want to contribute ...	2014-09-12 00:17:50	lionelbarber	0	64
Row20	Raju2627	RT @MFI_Media: GoharShahi.us | News | MFI UK' s Stance...	2014-09-12 00:17:14	MFI_Media	0	62
Row21	BilalIm18	RT @MFI_Media: GoharShahi.us | News | MFI UK' s Stance...	2014-09-12 00:17:11	MFI_Media	0	62

Summary Stats: get the top 25 Tweeters



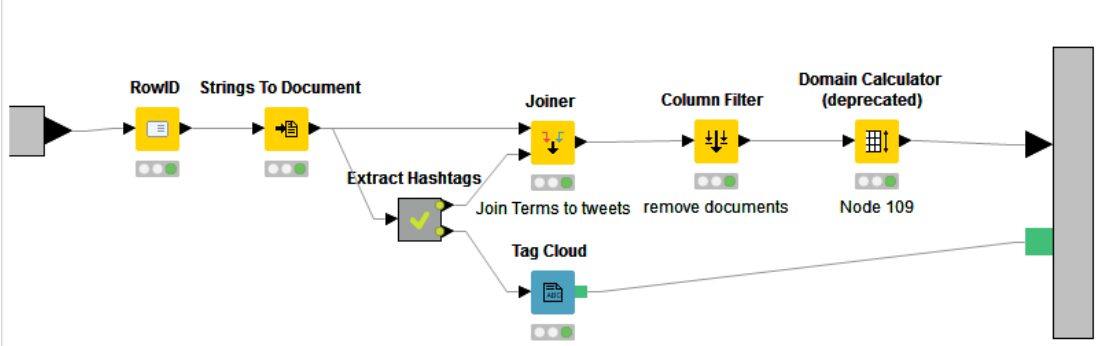
Grid table - 7:196:195 - Column to Grid

File

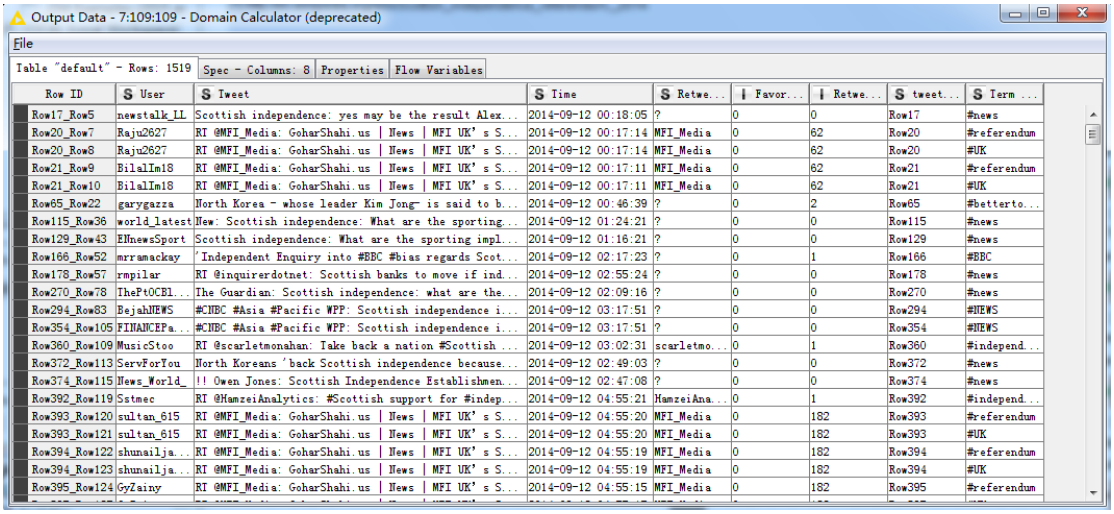
Table "default" - Rows: 5 | Spec - Columns: 10 | Properties | Flow Variables

Row ID	User (0)	Count...	User (1)	Count...	User (2)	Count...	User (3)	Count...	User (4)	Count...
Row0	2014_Scotland	66	barbjonn	39	austin89ca	31	debracolli	30	_dpaj	26
Row1	donfrancis...	26	kellrolle	26	christyacani	25	elene_rush	25	allenwaa	24
Row2	dalettoomo	24	janinenatt	24	aneucoz	23	husondo	22	kellyroosve	22
Row3	megannewstar	22	AndyMurra...	21	CollectedN	21	ideas4the...	20	maybee2day	20
Row4	millermarias	20	jesssparme	19	karendobr	19	kimbbyoung	19	madeli_ne_d	19

Text Processing: includes node RowID, Strings To Document, Extract Hashtags, Tag Cloud, Joiner, Column Filter, and Domain Calculator. Through this, it has two outputs, Data and Image like following:



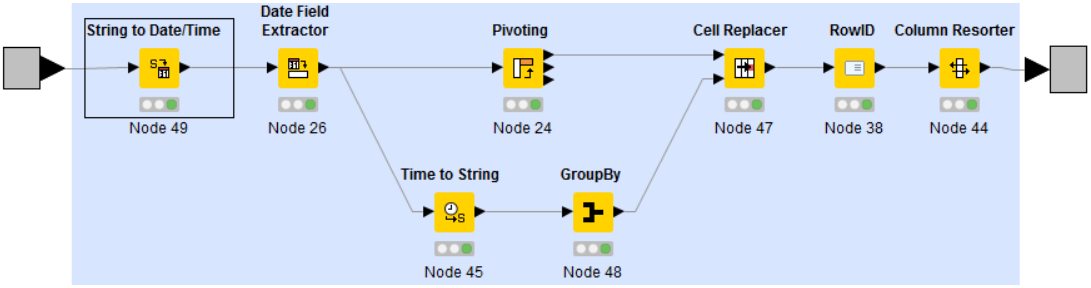
Through the process of calculating, it has two outputs, Data and Image like following:

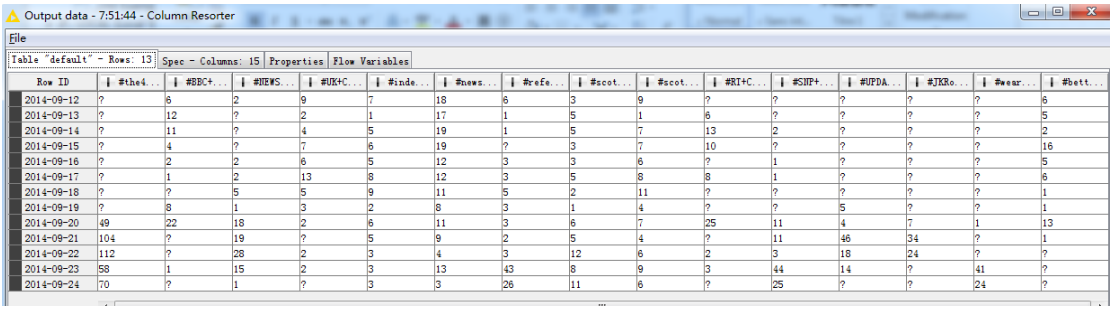


Row ID	User	Tweet	Time	Retwe...	Favor...	Retwe...	tweet...	Term ...
Row17_Row5	newstalk_LL	Scottish independence: yes may be the result Alex...	2014-09-12 00:18:05	?	0	0	Row17	#news
Row20_Row7	Raju2627	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 00:17:14	MFI_Media	0	62	Row20	#referendum
Row20_Row8	Raju2627	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 00:17:14	MFI_Media	0	62	Row20	#UK
Row21_Row9	BilalIm18	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 00:17:11	MFI_Media	0	62	Row21	#referendum
Row21_Row10	BilalIm18	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 00:17:11	MFI_Media	0	62	Row21	#UK
Row65_Row22	garygazza	North Korea - whose leader Kim Jong- is said to b...	2014-09-12 00:46:39	?	0	2	Row65	#betterto...
Row115_Row36	world_latest	New: Scottish independence: What are the sporting...	2014-09-12 01:24:21	?	0	0	Row115	#news
Row129_Row43	ENnewsSport	Scottish independence: What are the sporting impl...	2014-09-12 01:16:21	?	0	0	Row129	#news
Row166_Row52	mrramackay	'Independent Enquiry into #BBC #bias regards Scot...	2014-09-12 02:17:23	?	0	1	Row166	#BBC
Row178_Row57	rmpilar	RT @inquirerdotnet: Scottish banks to move if ind...	2014-09-12 02:55:24	?	0	0	Row178	#news
Row270_Row78	ThePtOCBl...	The Guardian: Scottish independence: what are the...	2014-09-12 02:09:16	?	0	0	Row270	#news
Row294_Row83	BejahNEWS	#CNBC #Asia #Pacific WPP: Scottish independence i...	2014-09-12 03:17:51	?	0	0	Row294	#NEWS
Row354_Row105	FINANCEPa...	#CNBC #Asia #Pacific WPP: Scottish independence i...	2014-09-12 03:17:51	?	0	0	Row354	#NEWS
Row360_Row109	MusicStoo	RT @scarletmonahan: Take back a nation #Scottish ...	2014-09-12 03:02:31	scarletmo...	0	1	Row360	#independ...
Row372_Row113	ServForYou	North Koreans 'back Scottish independence because...	2014-09-12 02:49:03	?	0	0	Row372	#news
Row374_Row115	News_World_	!! Owen Jones: Scottish Independence Establishmen...	2014-09-12 02:47:08	?	0	0	Row374	#news
Row392_Row119	Sstmec	RT @HamzeiAnalytics: #Scottish support for #indep...	2014-09-12 04:55:21	HamzeiAna...	0	1	Row392	#independ...
Row393_Row120	sultan_615	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 04:55:20	MFI_Media	0	182	Row393	#referendum
Row393_Row121	sultan_615	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 04:55:20	MFI_Media	0	182	Row393	#UK
Row394_Row122	shunailja...	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 04:55:19	MFI_Media	0	182	Row394	#referendum
Row394_Row123	shunailja...	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 04:55:19	MFI_Media	0	182	Row394	#UK
Row395_Row124	GyZsiny	RT @MFI_Media: GoharShahi.us | News | MFI UK' s S...	2014-09-12 04:55:15	MFI_Media	0	182	Row395	#referendum

The first Branch:

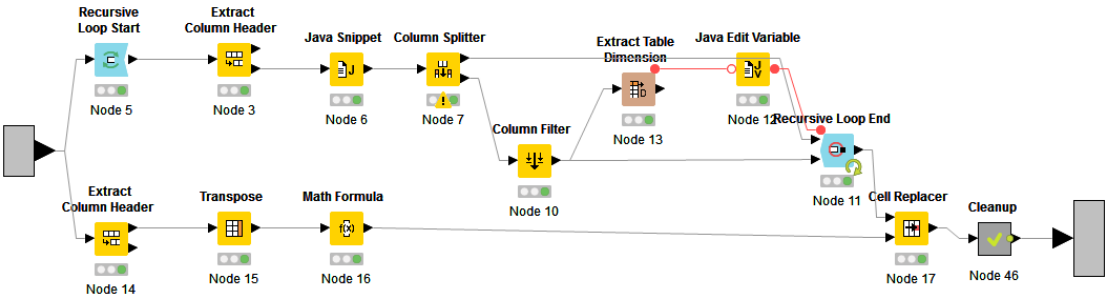
Pivot #tags to columns: reshapes data to show hashtag counts by day. The components and the result like following:





Row ID	#the4...	#BBC+...	#NEWS...	#UK+C...	#inde...	#news...	#refe...	#scot...	#scot...	#RI+C...	#SNP+...	#UPDA...	#JKRo...	#wear...	#bett...
2014-09-12	?	6	2	9	7	18	6	3	9	?	?	?	?	?	6
2014-09-13	?	12	?	2	1	17	1	5	1	6	?	?	?	?	5
2014-09-14	?	11	?	4	5	19	1	5	7	13	2	?	?	?	2
2014-09-15	?	4	?	7	6	19	?	3	7	10	?	?	?	?	16
2014-09-16	?	2	2	6	5	12	3	3	6	?	1	?	?	?	5
2014-09-17	?	1	2	13	8	12	3	5	8	8	1	?	?	?	6
2014-09-18	?	?	5	5	9	11	5	2	11	?	?	?	?	?	1
2014-09-19	?	8	1	3	2	8	3	1	4	?	?	5	?	?	1
2014-09-20	49	22	18	2	6	11	3	6	7	25	11	4	7	1	13
2014-09-21	104	?	19	?	5	9	2	5	4	?	11	46	34	?	1
2014-09-22	112	?	28	2	3	4	3	12	6	2	3	18	24	?	?
2014-09-23	58	1	15	2	3	13	43	8	9	3	44	14	?	41	?
2014-09-24	70	?	1	?	3	3	26	11	6	?	25	?	?	24	?

Theme River Processing: reshapes data to show hashtag minimum and maximum counts by day. The components and the result like following:



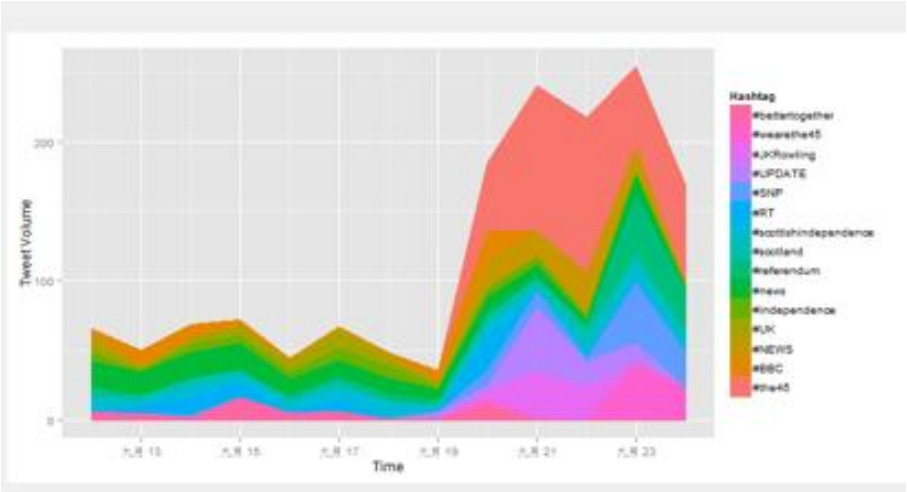
File

Table "default" - Rows: 195 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	D min	D max	S Itera...	S date	S Time	I date_...
2014-09-12#0	66	66	#the45	2014-09-12#0	2014-09-12	0
2014-09-13#0	50	50	#the45	2014-09-13#0	2014-09-13	0
2014-09-14#0	69	69	#the45	2014-09-14#0	2014-09-14	0
2014-09-15#0	72	72	#the45	2014-09-15#0	2014-09-15	0
2014-09-16#0	45	45	#the45	2014-09-16#0	2014-09-16	0
2014-09-17#0	67	67	#the45	2014-09-17#0	2014-09-17	0
2014-09-18#0	49	49	#the45	2014-09-18#0	2014-09-18	0
2014-09-19#0	36	36	#the45	2014-09-19#0	2014-09-19	0
2014-09-20#0	185	136	#the45	2014-09-20#0	2014-09-20	0
2014-09-21#0	240	136	#the45	2014-09-21#0	2014-09-21	0
2014-09-22#0	217	105	#the45	2014-09-22#0	2014-09-22	0
2014-09-23#0	254	196	#the45	2014-09-23#0	2014-09-23	0
2014-09-24#0	169	99	#the45	2014-09-24#0	2014-09-24	0
2014-09-12#1	66	60	#BBC	2014-09-12#1	2014-09-12	1
2014-09-13#1	50	38	#BBC	2014-09-13#1	2014-09-13	1
2014-09-14#1	69	58	#BBC	2014-09-14#1	2014-09-14	1
2014-09-15#1	72	68	#BBC	2014-09-15#1	2014-09-15	1
2014-09-16#1	45	43	#BBC	2014-09-16#1	2014-09-16	1
2014-09-17#1	67	66	#BBC	2014-09-17#1	2014-09-17	1
2014-09-18#1	49	49	#BBC	2014-09-18#1	2014-09-18	1
2014-09-19#1	36	28	#BBC	2014-09-19#1	2014-09-19	1
2014-09-20#1	136	114	#BBC	2014-09-20#1	2014-09-20	1
2014-09-21#1	136	136	#BBC	2014-09-21#1	2014-09-21	1
2014-09-22#1	105	105	#BBC	2014-09-22#1	2014-09-22	1

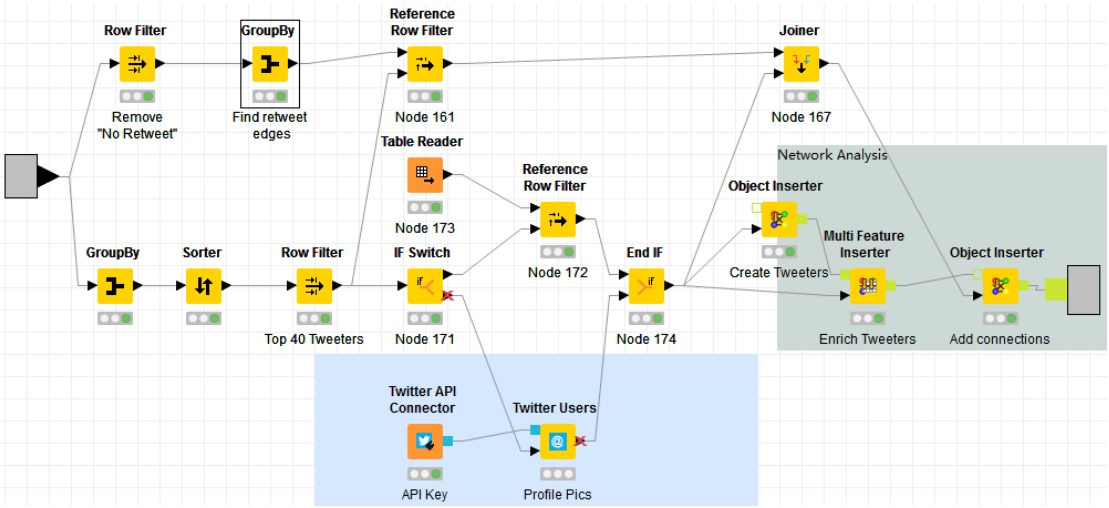
R View (Table): allows execution of an R script from within KNIME. The view resulting from this script is returned in the output image port of this node.

Data: 25,000 lines to analyze, so 25,000 tweets. The workflow will bring out the top 25 tweets. More specifically, this will highlight the people who posted the most tweets on the subject concerned. It will be able to give the following information, like the topics affected by the tweets:



The second branch:

Network Analysis: get the twitter user's information.



Network Viewer: visualizes the input network.

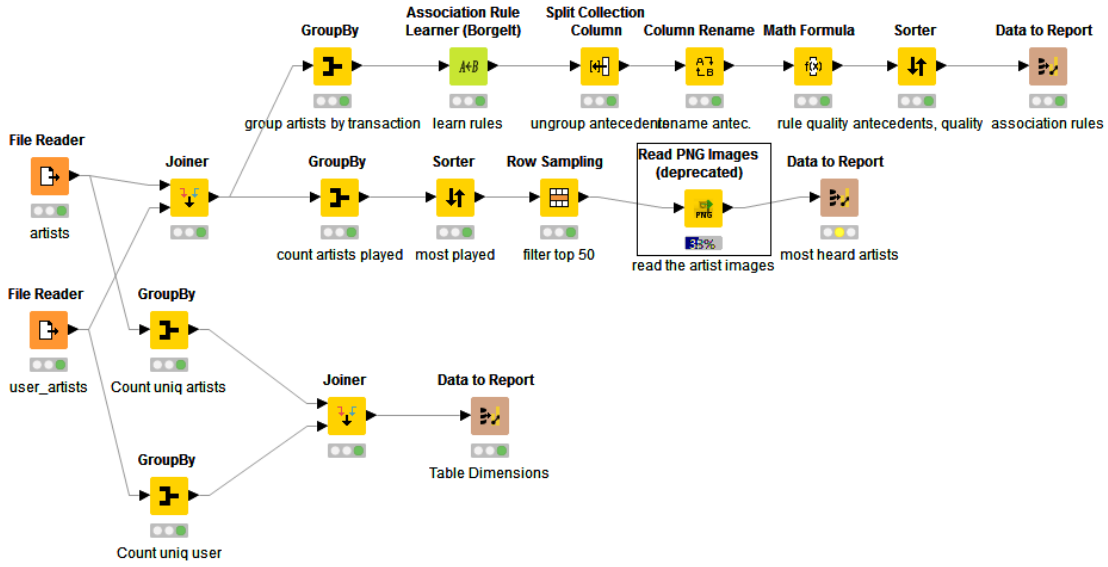
Conclusion:

This workflow examines a sample of tweets from the days surrounding the scottisch referendum for independence in 2014 [1]. After reading the data from a local database, basic text processing is used to extract hashtags from the dataset and term frequencies calculated and used to build a tag cloud. Subsequently, hashtag trending is examined over

time, with a notable post-election surge in the #the45 movement. Additionally, network analysis is performed in order to look at the most influential social graph surrounding this issue. And the workflow includes a report which can be edited and viewed from the KNIME Reporting perspective

Exercise 3. Recommender system

Import the Recommender System workflow (050004_lastfm_Recommendations) and the data. The workflow is like following:



From the workflow, we can see that there are two inputs (File reader) and three outputs (Data to Report).

3.1 The first input

The first input is the data of Artists. There are 17632 rows, 4 columns data. It includes their id, name, url and pictureURL.

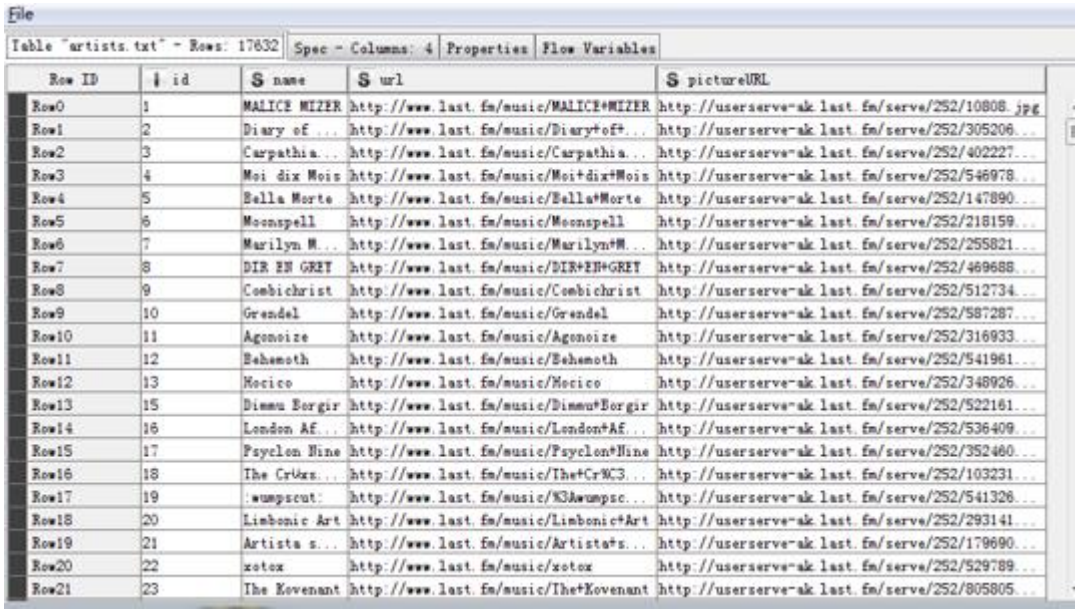
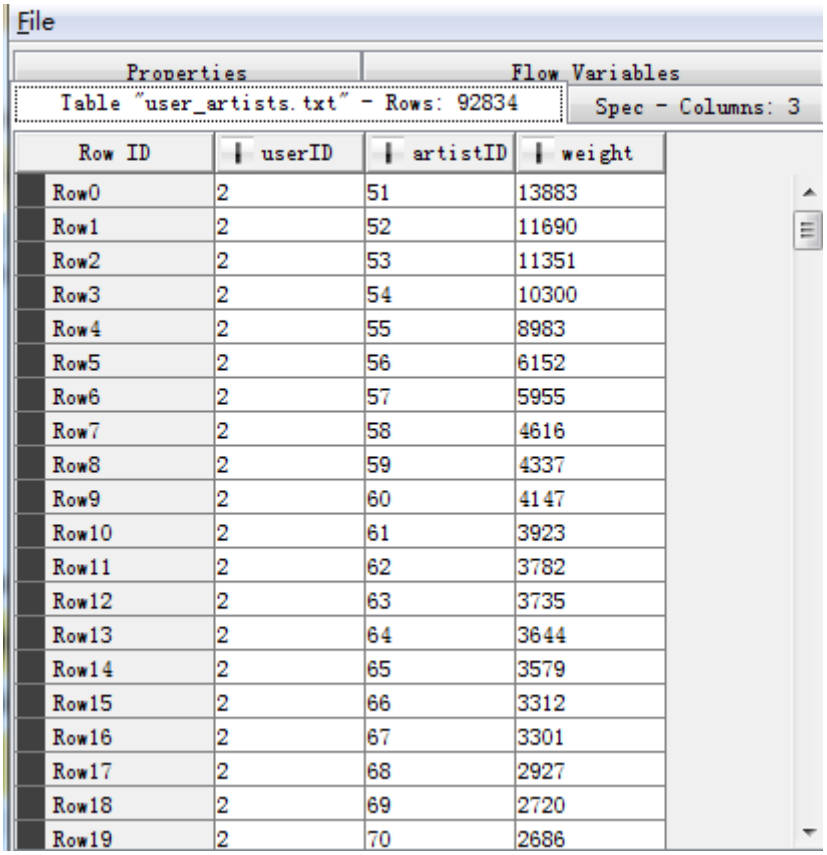

File

Table "artists.txt" - Rows: 17632 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	id	name	url	pictureURL
Row0	1	MALICE MIZER	http://www.last.fm/music/MALICE+MIZER	http://userserve-ak.last.fm/serve/252/10808.jpg
Row1	2	Diary of ...	http://www.last.fm/music/Diary+of+...	http://userserve-ak.last.fm/serve/252/305206...
Row2	3	Carpathia...	http://www.last.fm/music/Carpathia...	http://userserve-ak.last.fm/serve/252/402227...
Row3	4	Moi dix Mois	http://www.last.fm/music/Moi+dix+Mois	http://userserve-ak.last.fm/serve/252/546978...
Row4	5	Bella Morte	http://www.last.fm/music/Bella+Morte	http://userserve-ak.last.fm/serve/252/147890...
Row5	6	Moonspell	http://www.last.fm/music/Moonspell	http://userserve-ak.last.fm/serve/252/218159...
Row6	7	Marilyn M...	http://www.last.fm/music/Marilyn+M...	http://userserve-ak.last.fm/serve/252/255821...
Row7	8	DIR EN GREY	http://www.last.fm/music/DIR+EN+GREY	http://userserve-ak.last.fm/serve/252/469688...
Row8	9	Combichrist	http://www.last.fm/music/Combichrist	http://userserve-ak.last.fm/serve/252/512734...
Row9	10	Grendel	http://www.last.fm/music/Grendel	http://userserve-ak.last.fm/serve/252/587287...
Row10	11	Agonoize	http://www.last.fm/music/Agonoize	http://userserve-ak.last.fm/serve/252/316933...
Row11	12	Behemoth	http://www.last.fm/music/Behemoth	http://userserve-ak.last.fm/serve/252/541961...
Row12	13	Mocico	http://www.last.fm/music/Mocico	http://userserve-ak.last.fm/serve/252/348926...
Row13	15	Dimmu Borgir	http://www.last.fm/music/Dimmu+Borgir	http://userserve-ak.last.fm/serve/252/522161...
Row14	16	London Af...	http://www.last.fm/music/London+Af...	http://userserve-ak.last.fm/serve/252/536409...
Row15	17	Psyclon Nine	http://www.last.fm/music/Psyclon+Nine	http://userserve-ak.last.fm/serve/252/352460...
Row16	18	The Cr\0xs...	http://www.last.fm/music/The+Cr%C3...	http://userserve-ak.last.fm/serve/252/103231...
Row17	19	:wumpscut:	http://www.last.fm/music/%3Awumpsc...	http://userserve-ak.last.fm/serve/252/541326...
Row18	20	Limbonic Art	http://www.last.fm/music/Limbonic+Art	http://userserve-ak.last.fm/serve/252/293141...
Row19	21	Artista s...	http://www.last.fm/music/Artista+s...	http://userserve-ak.last.fm/serve/252/179690...
Row20	22	xotox	http://www.last.fm/music/xotox	http://userserve-ak.last.fm/serve/252/529789...
Row21	23	The Kovenant	http://www.last.fm/music/The+Kovenant	http://userserve-ak.last.fm/serve/252/805805...

3.2 The second input

The sacond input is the data of user_artists. There are 92834 rows, 3 columns data. It includes userID, artistID(the user like), weight.



File

Properties | Flow Variables

Table "user_artists.txt" - Rows: 92834 | Spec - Columns: 3

Row ID	userID	artistID	weight
Row0	2	51	13883
Row1	2	52	11690
Row2	2	53	11351
Row3	2	54	10300
Row4	2	55	8983
Row5	2	56	6152
Row6	2	57	5955
Row7	2	58	4616
Row8	2	59	4337
Row9	2	60	4147
Row10	2	61	3923
Row11	2	62	3782
Row12	2	63	3735
Row13	2	64	3644
Row14	2	65	3579
Row15	2	66	3312
Row16	2	67	3301
Row17	2	68	2927
Row18	2	69	2720
Row19	2	70	2686

3.3 The first output

The first output is a multimedia report that shows the top artists and the other musicians associated with each in the form “Others who like X also like….”
The result (data to report) includes the Antecedents of the artist, the consequent artist and this form’s support and confidence.

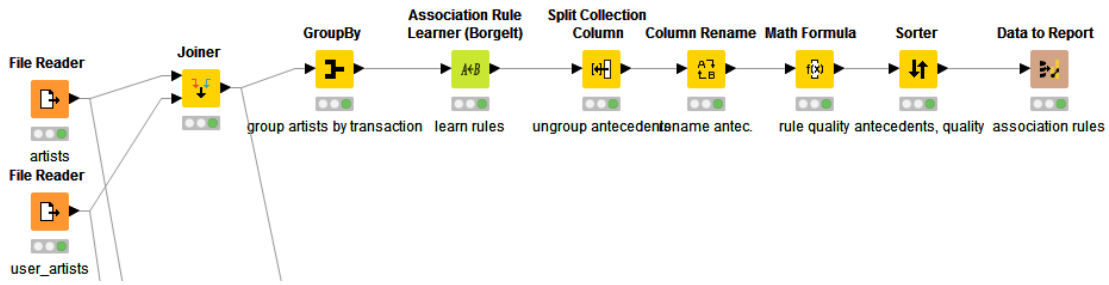


Table View - 8:18 - Data to Report (association rules)

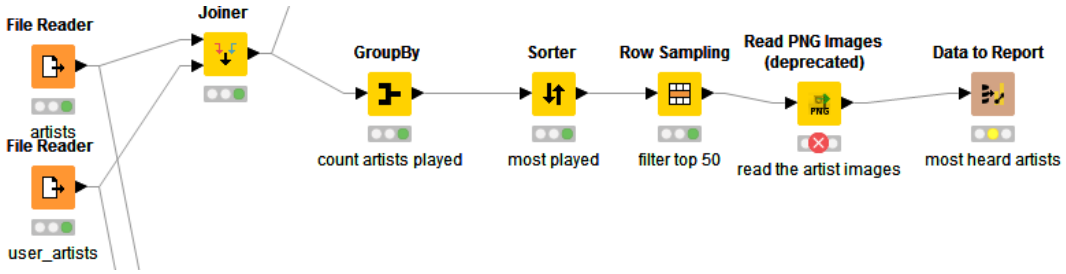
File Hilite Navigation View Output

Row ID	Conse...	Antecedants	ItemS...	Relat...	RuleC...	Absol...	Relat...	RuleLift	RuleL...	Absol...	Relat...	rule ...
Row111	Rihanna	The Pussycat Dolls	215	11.364	92.3	233	12.3	3.607	360.71	484	25.581	19,844.5
Row113	Britney S...	The Pussycat Dolls	213	11.258	91.4	233	12.3	3.313	331.34	522	27.59	19,468.2
Row115	Lady Gaga	The Pussycat Dolls	201	10.624	86.3	233	12.3	2.671	267.128	611	32.294	17,346.3
Row104	Beyoncé	The Pussycat Dolls	195	10.307	83.7	233	12.3	3.989	398.85	397	20.983	16,321.5
Row106	Christina...	The Pussycat Dolls	193	10.201	82.8	233	12.3	3.851	385.06	407	21.512	15,980.4
Row110	Katy Perry	The Pussycat Dolls	178	9.408	76.4	233	12.3	3.056	305.579	473	25	13,599.2
Row109	Madonna	The Pussycat Dolls	166	8.774	71.2	233	12.3	3.142	314.207	429	22.674	11,819.2
Row101	Shakira	The Pussycat Dolls	150	7.928	64.4	233	12.3	3.818	381.826	319	16.86	9,660
Row108	Avril Lav...	The Pussycat Dolls	149	7.875	63.9	233	12.3	2.901	290.145	417	22.04	9,521.1
Row103	Ke$ha	The Pussycat Dolls	143	7.558	61.4	233	12.3	3.208	320.769	362	19.133	8,780.2
Row97	Black Eye...	The Pussycat Dolls	141	7.452	60.5	233	12.3	3.766	376.626	304	16.068	8,530.5
Row89	Mariah Carey	The Pussycat Dolls	135	7.135	57.9	233	12.3	4.549	454.864	241	12.738	7,816.5
Row99	P!nk	The Pussycat Dolls	134	7.082	57.5	233	12.3	3.568	356.755	305	16.121	7,705
Row44	Jennifer ...	The Pussycat Dolls	128	6.765	54.9	233	12.3	4.973	497.312	209	11.047	7,027.2
Row95	Kylie Min...	The Pussycat Dolls	127	6.712	54.5	233	12.3	3.461	346.061	298	15.751	6,921.5
Row93	Miley Cyrus	The Pussycat Dolls	124	6.554	53.2	233	12.3	3.521	352.063	286	15.116	6,596.8
Row91	Kelly Cla...	The Pussycat Dolls	121	6.395	51.9	233	12.3	3.779	377.9	260	13.742	6,279.9
Row8	Ashley Ti...	The Pussycat Dolls	102	5.391	43.8	233	12.3	4.551	455.087	182	9.619	4,467.6
Row318	Muse	The Killers	156	8.245	51.3	304	16.1	2.427	242.724	400	21.142	8,002.8
Row315	Coldplay	The Killers	148	7.822	48.7	304	16.1	2.496	249.622	369	19.503	7,207.6
Row317	Radiohead	The Killers	125	6.607	41.1	304	16.1	1.98	197.954	393	20.772	5,137.5
Row319	The Beatles	The Killers	125	6.607	41.1	304	16.1	1.621	162.075	480	25.37	5,137.5
Row85	Depeche Mode	The Cure	117	6.184	51.1	229	12.1	3.428	342.785	282	14.905	5,978.7
Row88	The Beatles	The Cure	103	5.444	45	229	12.1	1.773	177.289	480	25.37	4,635

3.4 The second output

The second output is a dynamic multi-media report that shows the half top favorite artists and combines this list with overall facts about the sample and enhances the artist data with pictures.

But the pictureURL is invalid, so the Read PNG Images cannot execute. We just can see the most 50% popular artists results.



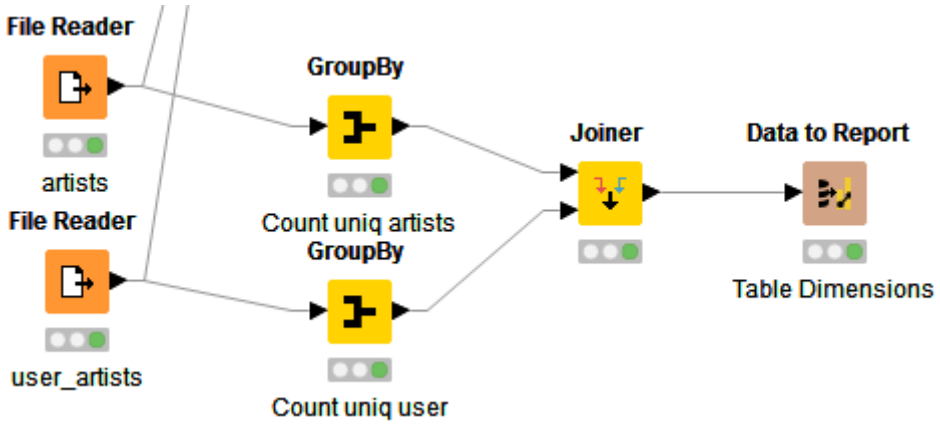
File

Table "default" - Rows: 50 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	name	First(pictureURL)	First...	Count...
Row8353	Lady Gaga	http://userserve-ak.last.fm/serve/252/47390093.png	89	611
Row2337	Britney S...	http://userserve-ak.last.fm/serve/252/60126439.png	289	522
Row12073	Rihanna	http://userserve-ak.last.fm/serve/252/53023109.png	288	484
Row14110	The Beatles	http://userserve-ak.last.fm/serve/252/2588646.jpg	227	480
Row7914	Katy Perry	http://userserve-ak.last.fm/serve/252/42128121.png	300	473
Row9131	Madonna	http://userserve-ak.last.fm/serve/252/340387.jpg	67	429
Row1412	Avril Lav...	http://userserve-ak.last.fm/serve/252/59708309.png	333	417
Row3024	Christina...	http://userserve-ak.last.fm/serve/252/47363849.png	292	407
Row10159	Muse	http://userserve-ak.last.fm/serve/252/416514.jpg	190	400
Row11070	Paramore	http://userserve-ak.last.fm/serve/252/35837991.png	498	399
Row1798	Beyoncé	http://userserve-ak.last.fm/serve/252/61958009.png	295	397
Row11763	Radiohead	http://userserve-ak.last.fm/serve/252/8461967.jpg	154	393
Row3221	Coldplay	http://userserve-ak.last.fm/serve/252/67770.jpg	65	369
Row7939	Ke$ha	http://userserve-ak.last.fm/serve/252/47829587.png	466	362
Row12819	Shakira	http://userserve-ak.last.fm/serve/252/52116105.png	701	319
Row10967	P!nk	http://userserve-ak.last.fm/serve/252/56011579.png	302	305
Row1949	Black Eye...	http://userserve-ak.last.fm/serve/252/53698409.png	306	304
Row14582	The Killers	http://userserve-ak.last.fm/serve/252/250099.jpg	229	304
Row8271	Kylie Min...	http://userserve-ak.last.fm/serve/252/12740835.jpg	55	298
Row9814	Miley Cyrus	http://userserve-ak.last.fm/serve/252/58452885.png	481	286

3.5 The third output

The third output is a report that shows there are how many artists and how many users in the database. Because one user can like several artists. From the table, we can see the number of artists is 17632, the number of users is 1892.



File Hilite Navigation View Output

Row ID	Unique count(name)	Unique count(userID)
Row0	17632	1892

Conclusion

This KNIME workflow takes Social Media data from a popular music site and uses a predictive analytics technique to make music preference recommendations for the top artists. In addition, the workflow creates a multimedia report that shows the top artists and the other musicians associated with each in the form “Others who like X

also like…."

This workflow transforms the Social Media data to make it suitable for association, performs and advanced association analysis, utilizes the resulting statistics to select lists of artists and recommendations, combines this list with overall facts about the sample and enhances the artist data with pictures to create a dynamic multi-media report.