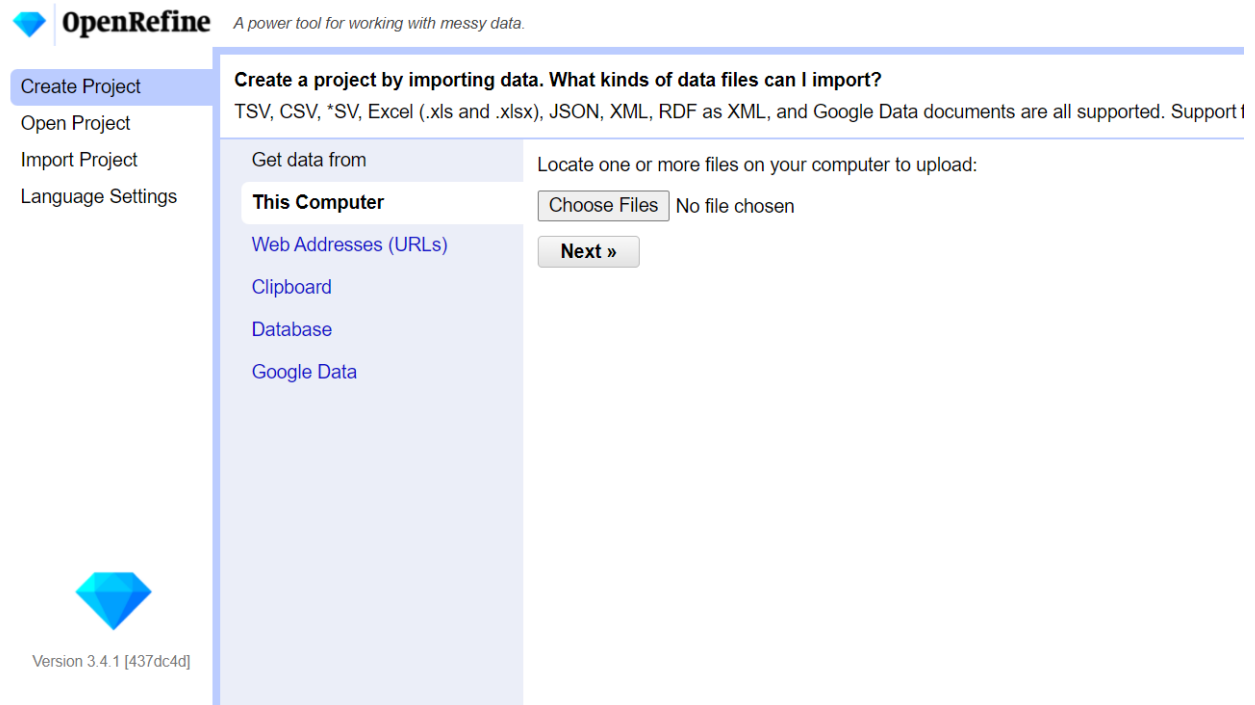


OPEN REFINE

Openrefine is an open-source desktop application for data cleanup and transformation to other formats. Installing it has a requirement of JAVA.

Running Openrefine let you choose the dataset to be uploaded or can select from any of the options like URL, Clipboard...



Choosing file and opening a project provides you an interface with blend of tools

OpenRefine CMPE-255_HW Vishnu Permalink Open... Export Help

Facet / Filter Undo / Redo 10 / 10

37352 rows Extensions: Wikidata

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

All	video_id	trending_date	title	channel_title	category_id	publish_time	views	like
1.	kzwfHumJyYc	17.14.11	Sharry Mann: Cute Munda (Song Teaser) Pamish Verma Releasing on 17 November	Lokdhun Punjabi	1	2017-11-12T12:20:30Z	1096327	3396
2.	zUZ1z7FwLc8	17.14.11	पौराणिक के समय, पेट पर पति करता ऐस, देखकर देन खे खे	HJ NEWS	25	2017-11-13T05:43:56.000Z	590101	73
3.	10L1hZ9qa58	17.14.11	Stylish Star Allu Arjun @ ChaySam Wedding Reception TFPC	TFPC	24	2017-11-12T15:48:08Z	473988	201
4.	N1vEBiEg64	17.14.11	Eruma Saani Tamil vs English	Eruma Saani	23	2017-11-12T07:08:48Z	1242680	7035
5.	kJzGH0PVQHQ	17.14.11	why Samantha became EMOTIONAL @ Samantha naga chaithanya marriage Reception Filmylooks	Filmylooks	24	2017-11-13T01:14:16Z	464015	49
6.	iLpSa5l98w	17.14.11	MCA (Middle Class Abbay) TEASER - Nani, Sai Pallavi, Dil Raju, Devi Sri Prasad, Srima Venu	Dil Raju	24	2017-11-10T04:29:50Z	6106669	9861
7.	7MxIQ4v0EnE	17.14.11	Daang (Full Video) Mankirt Aulakh Sukh Sanghera Latest Punjabi Song 2017 Speed Records	Speed Records	10	2017-11-11T16:41:15Z	5718766	12747
8.	c64l9HNpIOY	17.14.11	Padmavati - Ek Dil Ek Jaan Video Song Deepika Padukone Shahid Kapoor Sanjay Leela Bhansali	T-Series	10	2017-11-11T06:14:10Z	10588371	13273
9.	K0bFEYCaRx8	17.14.11	Chiranjeevi in Naga Chaitanya - Samantha Reception Chaitanya, Samantha Akkineni, Nag, Venkatesh	Top Telugu Media	24	2017-11-13T04:42:26Z	118223	52
10.	g8QstJhFpY	17.14.11	New bike vs Old bike - the reality	Jump Cuts	24	2017-11-12T04:30:01Z	969030	5979

Converting publish_time, likes and comment_count to date, numbers and numbers respectively

sets
et and
re top

37352 rows

Extensions: V

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 50 > next

trending_date	title	channel_title	category_id	publish_time	views	likes	dislikes	comment_count
7.14.11	Sharry Mann: Cute Munda (Song Teaser) Parmish Verma Releasing on 17 November	Lokdhun Punjabi	1	2017-11-12T12:20:39Z	1096327			
7.14.11	पौरियड्स के समय, पेट पर पति करता ऐसा, देखकर दंग रह जायेंगे	HJ NEWS	25	2017-11-13T04:42:26Z	590101	735		
7.14.11	Stylish Star Allu Arjun @ ChaySam Wedding Reception TFPC	Trim leading and trailing whitespace		Common transforms				
7.14.11	Eruma Saani Tamil vs English	Collapse consecutive whitespace		Fill down				
7.14.11	why Samantha became EMOTIONAL @ Samantha naga chaitanya marriage Reception Filmylooks	Unescape HTML entities		Blank down				
7.14.11	MCA (Middle Class Abbayi) TEASER - Nani, Sai Pallavi, Dil Raju, Devi Sri Prasad, Sriram Venu	Replace Smart quotes with ascii		Split multi-valued cells...				
7.14.11	Daang (Full Video) Mankirt Aulakh Sukh Sanghera Latest Punjabi Song 2017 Speed Records	To titlecase		Join multi-valued cells...				
7.14.11	Padmavati : Ek Dil Ek Jaan Video Song Deepika Padukone Shahid Kapoor Sanjay Leela Bhansali	To uppercase		Cluster and edit...				
7.14.11	Chiranjeevi in Naga Chaitanya - Samantha Reception Chaitanya, Samantha Akkineni, Nag, Venkatesh	To lowercase		Replace				
7.14.11	New bike vs Old bike - the reality	To number		2017-11-11T06:14:19Z	10588371	132738	8812	
7.14.11	Mehjabi Reveals Hina Khan Dirty Politics Exclusive Interview Bigg Boss	To date		2017-11-13T04:42:26Z	118223	520	53	
		To text		2017-11-12T04:30:01Z	969030	59798	1545	
		To null		2017-11-12T17:30:00Z	632747	4330	2183	
		To empty string						

Removing unnecessary columns

37352 rows

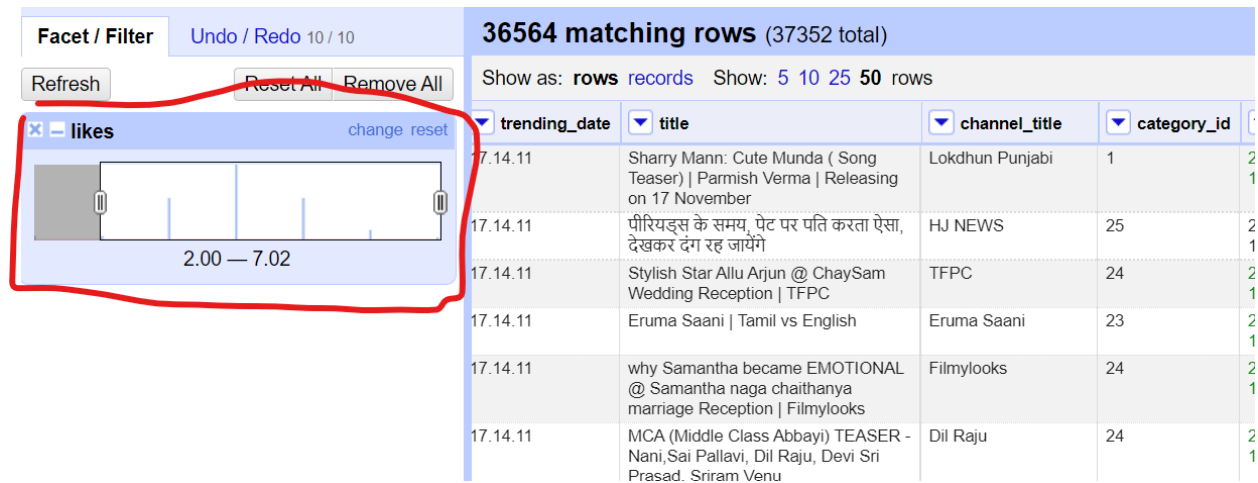
Extensions: Wiki

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 50 next

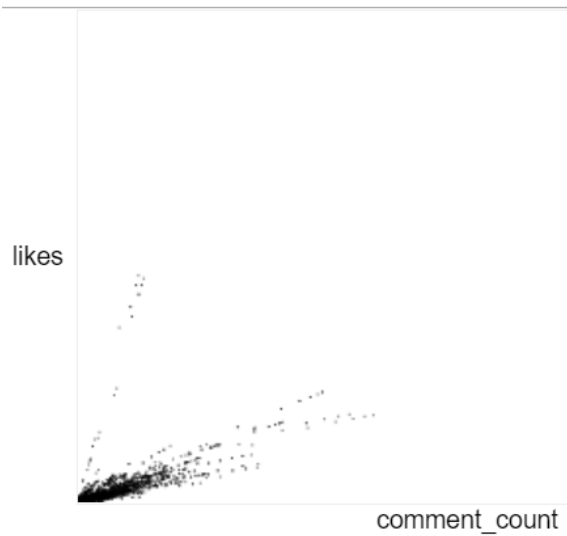
trending_date	title	channel_title	category_id	publish_time	views	likes	dislikes	comment
7.14.11	Sharry Mann: Cute Munda (Song Teaser) Parmish Verma Releasing on 17 November	Lokdhun Punjabi	1	2017-11-12T12:20:39Z	1096327	33966	Facet	
7.14.11	पौरियड्स के समय, पेट पर पति करता ऐसा, देखकर दंग रह जायेंगे	HJ NEWS	25	2017-11-13T05:43:56.000Z	590101	735	Text filter	
7.14.11	Stylish Star Allu Arjun @ ChaySam Wedding Reception TFPC	TFPC	24	Split into several columns... Join columns... Add column based on this column... Add column by fetching URLs... Add columns from reconciled values... Rename this column Remove this column Move column to beginning Move column to end Move column left Move column right			Edit cells	
7.14.11	Eruma Saani Tamil vs English	Eruma Saani	23				Edit column	
7.14.11	why Samantha became EMOTIONAL @ Samantha naga chaitanya marriage Reception Filmylooks	Filmylooks	24				Transpose	
7.14.11	MCA (Middle Class Abbayi) TEASER - Nani,Sai Pallavi, Dil Raju, Devi Sri Prasad, Sriram Venu	Dil Raju	24				Sort...	
7.14.11	Daang (Full Video) Mankirt Aulakh Sukh Sanghera Latest Punjabi Song 2017 Speed Records	Speed Records	10				View	
7.14.11	Padmavati : Ek Dil Ek Jaan Video Song Deepika Padukone Shahid Kapoor Sanjay Leela Bhansali	T-Series	10				Reconcile	
7.14.11	Chiranjeevi in Naga Chaitanya - Samantha Reception Chaitanya, Samantha Akkineni, Nag, Venkatesh	Top Telugu Media	24				7134	
7.14.11	New bike vs Old bike - the reality	Jump Cuts	24	2017-11-12T04:30:01Z	969030	59798	3812	
							53	
7.14.11	New bike vs Old bike - the reality	Jump Cuts	24	2017-11-12T04:30:01Z	969030	59798	1545	

Cleaning data using customized facets and timeline facet also using multiple filters to get required data



Scatterplot of the required data after cleaning

Scatterplot Matrix (focusing on 'publish_time')



Note: Please refer [Gitrepo](#) for colab and further details

References:

[\(32\) Reconciliation in OpenRefine part 2 - YouTube](#)

[\(32\) OpenRefine Data Cleaning Tutorial - YouTube](#)

[\(32\) Clean Your Data: Getting Started with OpenRefine \[workshop\] - YouTube](#)

[\(32\) Get Started with OpenRefine: Explore, Clean, and Transform your Data! - YouTube](#)

[Faceting - Getting Started with Data Cleaning and OpenRefine - LibGuides at University of Nevada, Las Vegas \(unlv.edu\)](#)