

# Pageviews-edits-data-viz-Clean

May 31, 2020

```
"\nLots of plots for the Wiki-HUEs/Jammu and Kashmir paper. Not all to be
included in the paper.\n\nTo export this to PDF without the plots, run the
following in the command line. You'll need xelatex or xetex, and
nbconvert.\n\njupyter nbconvert --to pdf Pageviews-edits-data-viz-Clean.ipynb
--TemplateExporter.exclude_input_prompt=True
--TemplateExporter.exclude_output_prompt=True
--TemplateExporter.exclude_input=True\n"
```

## 1 Page views time series.

Log scale makes local peaks more visible. Linear scale plots are in plots/pageviews if you want to see them. Note that these are all from article creation through March 2020. Later in the notebook there are plots of only 2019, with both page views and edits time series.

Unnamed: 0	project	article	granularity	access	\	
0	0	hi.wikipedia	Kashmir	conflict	daily	all-access
1	1	hi.wikipedia	Kashmir	conflict	daily	all-access
2	2	hi.wikipedia	Kashmir	conflict	daily	all-access
3	3	hi.wikipedia	Kashmir	conflict	daily	all-access
4	4	hi.wikipedia	Kashmir	conflict	daily	all-access

	agent	views	date
0	all-agents	104	2019-01-01
1	all-agents	144	2019-01-02
2	all-agents	159	2019-01-03
3	all-agents	174	2019-01-04
4	all-agents	145	2019-01-05

/home/molly/anaconda3/lib/python3.7/site-packages/ipykernel\_launcher.py:23:

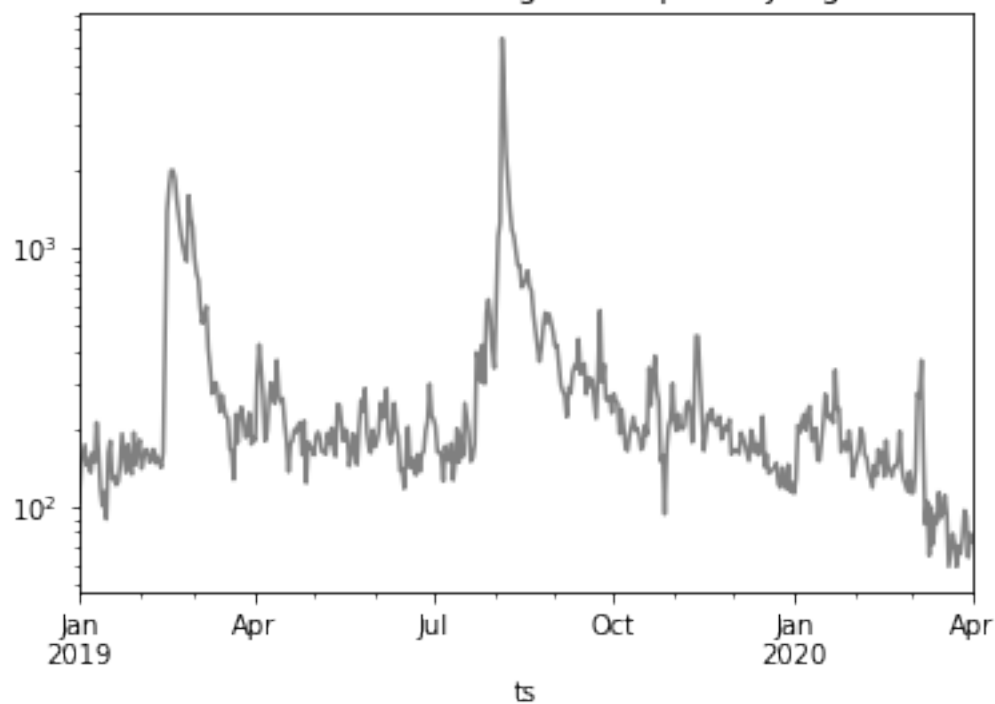
SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

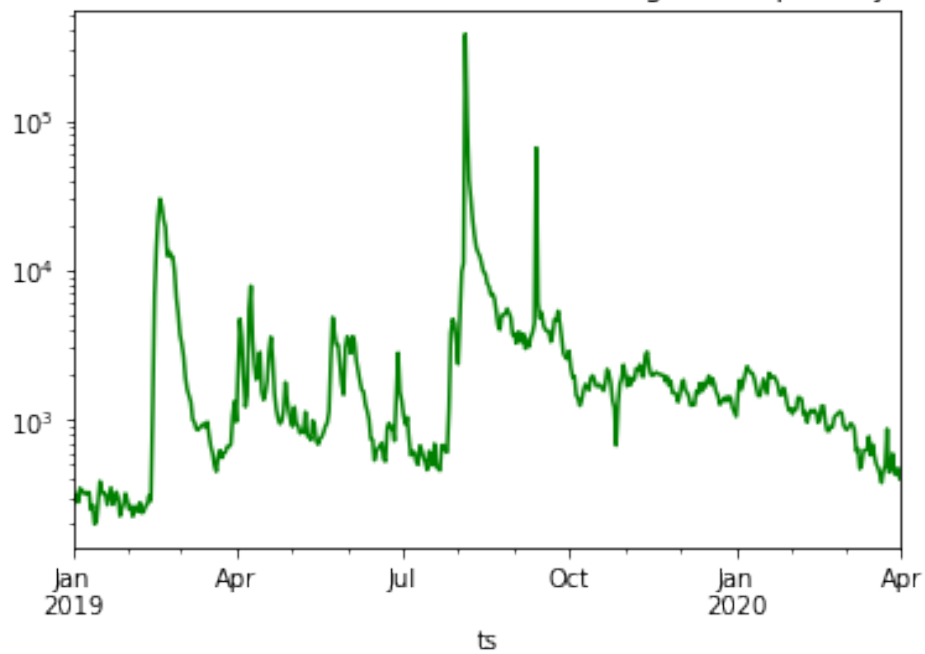
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

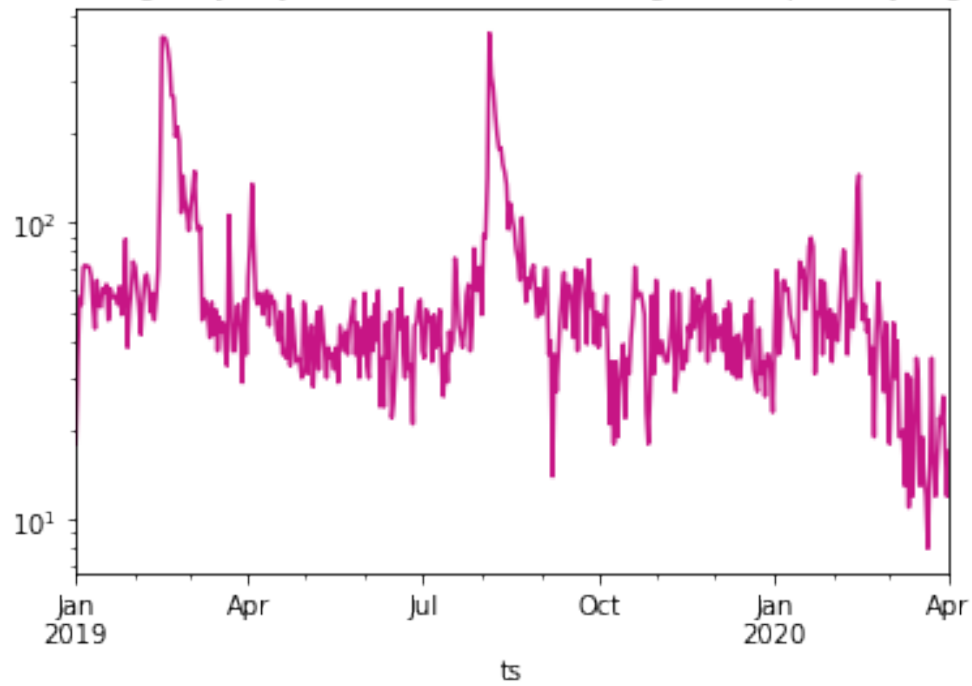
Hindi Kashmir conflict Pageviews per day log scale



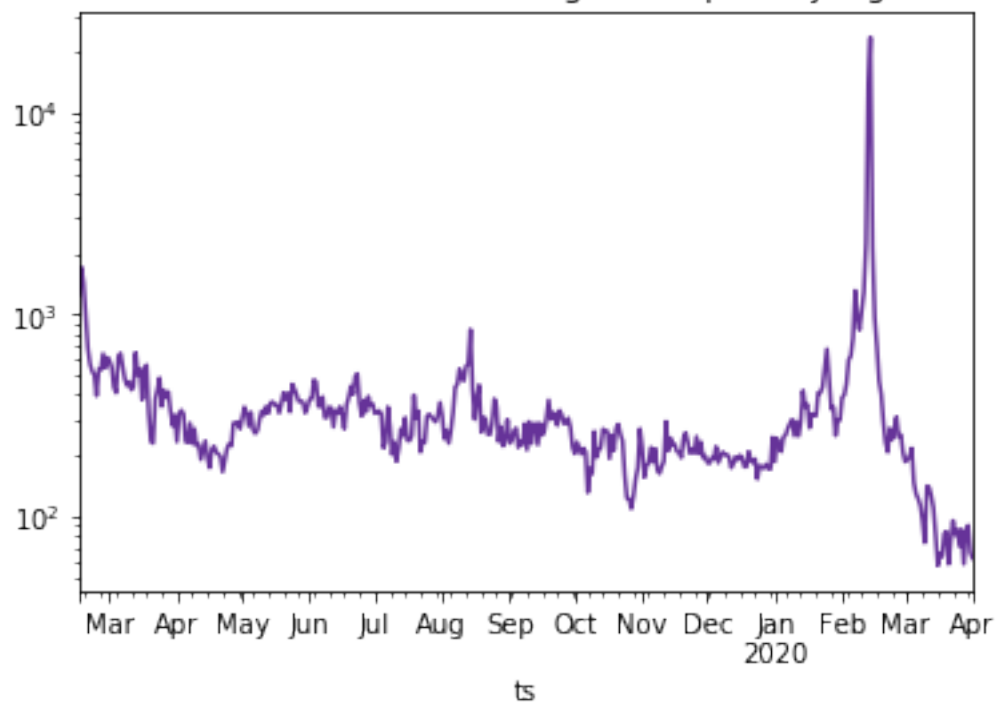
Hindi Article 370 of the Constitution of India Pageviews per day log scale



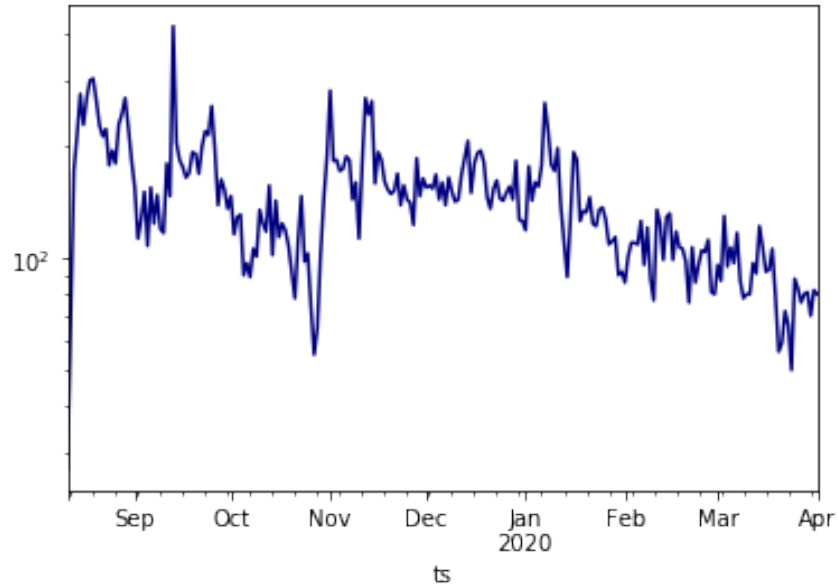
Hindi Insurgency in Jammu and Kashmir Pageviews per day log scale



Hindi 2019 Pulwama attack Pageviews per day log scale



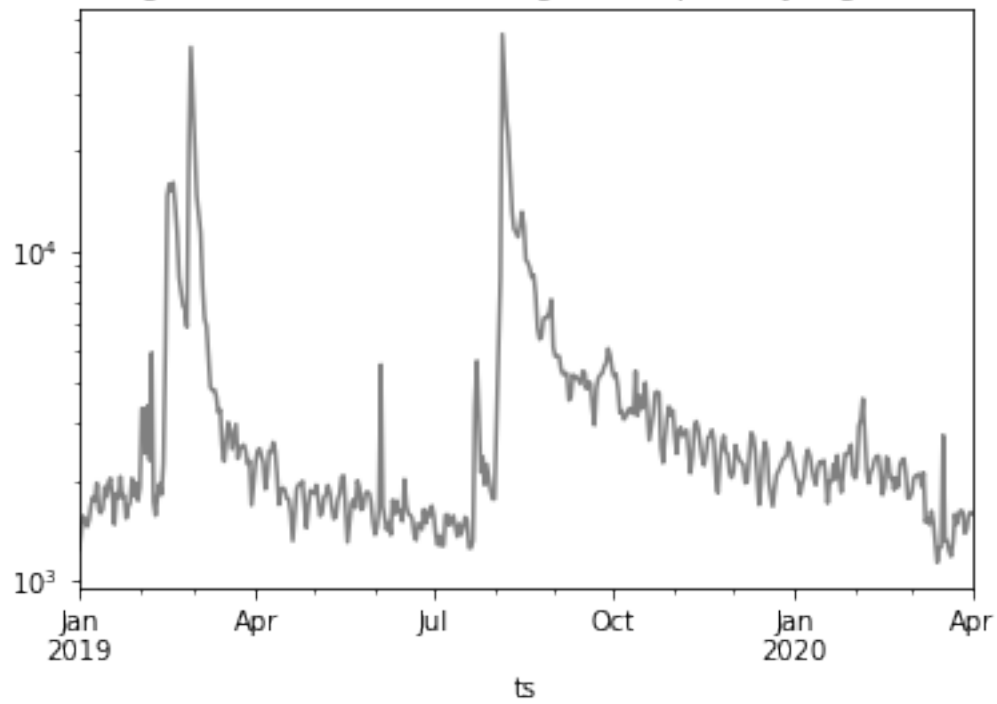
Hindi Jammu and Kashmir Reorganisation Act, 2019 Pageviews per day log scale



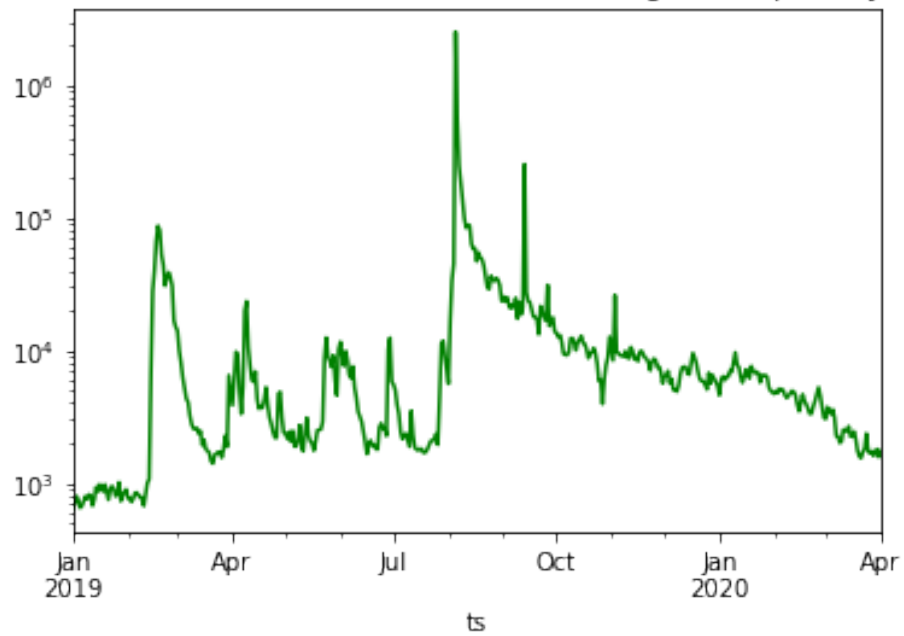
	Unnamed: 0	project	article	granularity	access	\
0	0	en.wikipedia	Kashmir conflict	daily	all-access	
1	1	en.wikipedia	Kashmir conflict	daily	all-access	
2	2	en.wikipedia	Kashmir conflict	daily	all-access	
3	3	en.wikipedia	Kashmir conflict	daily	all-access	
4	4	en.wikipedia	Kashmir conflict	daily	all-access	

	agent	views	date
0	all-agents	1233	2019-01-01
1	all-agents	1425	2019-01-02
2	all-agents	1558	2019-01-03
3	all-agents	1522	2019-01-04
4	all-agents	1447	2019-01-05

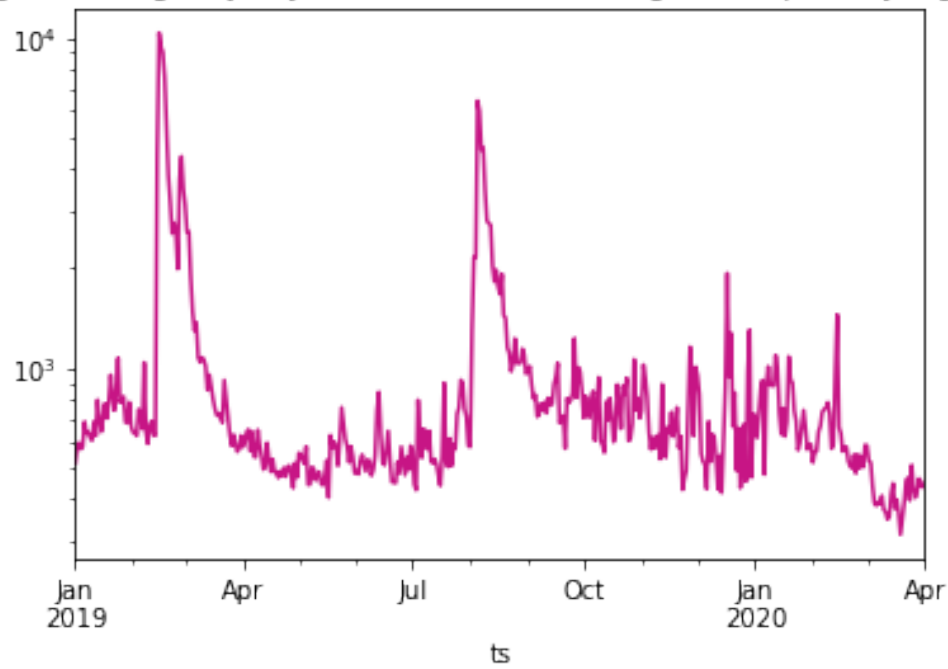
English Kashmir conflict Pageviews per day log scale



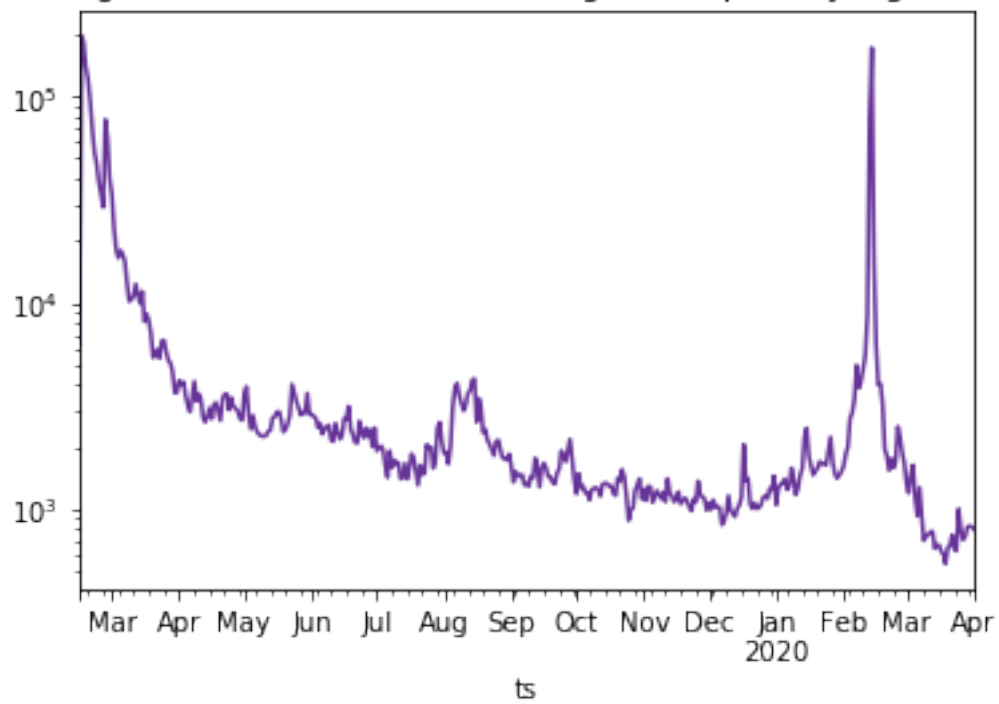
English Article 370 of the Constitution of India Pageviews per day log scale



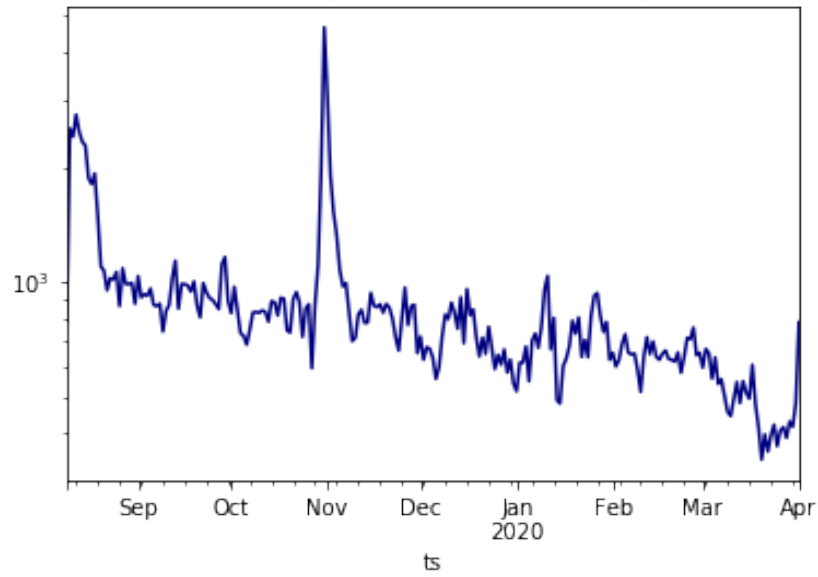
English Insurgency in Jammu and Kashmir Pageviews per day log scale



English 2019 Pulwama attack Pageviews per day log scale



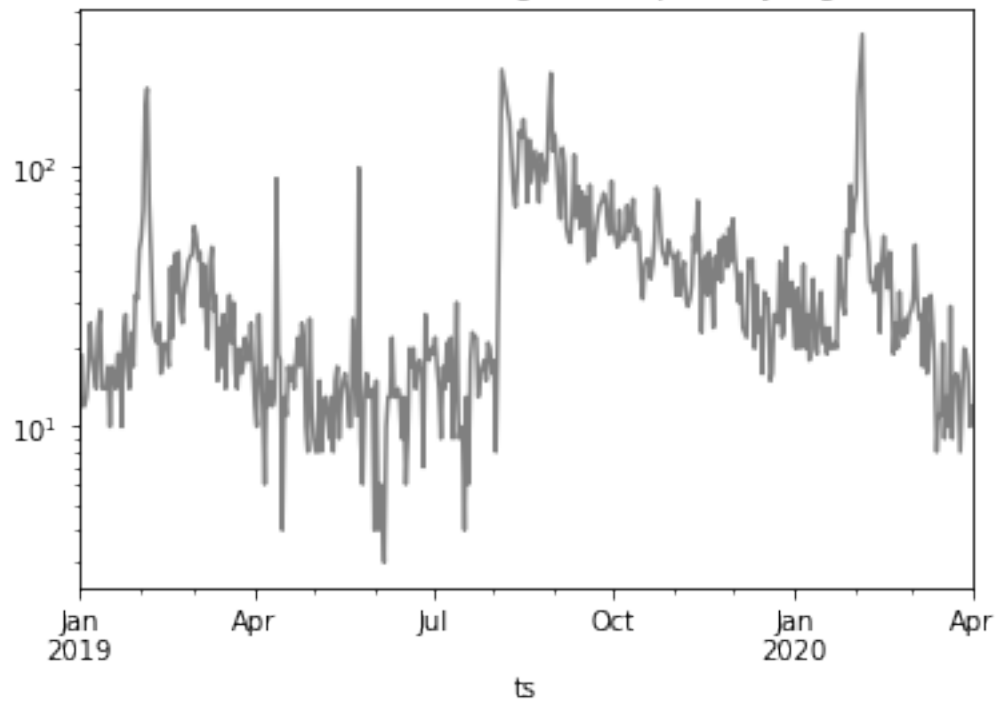
English Jammu and Kashmir Reorganisation Act, 2019 Pageviews per day log scale



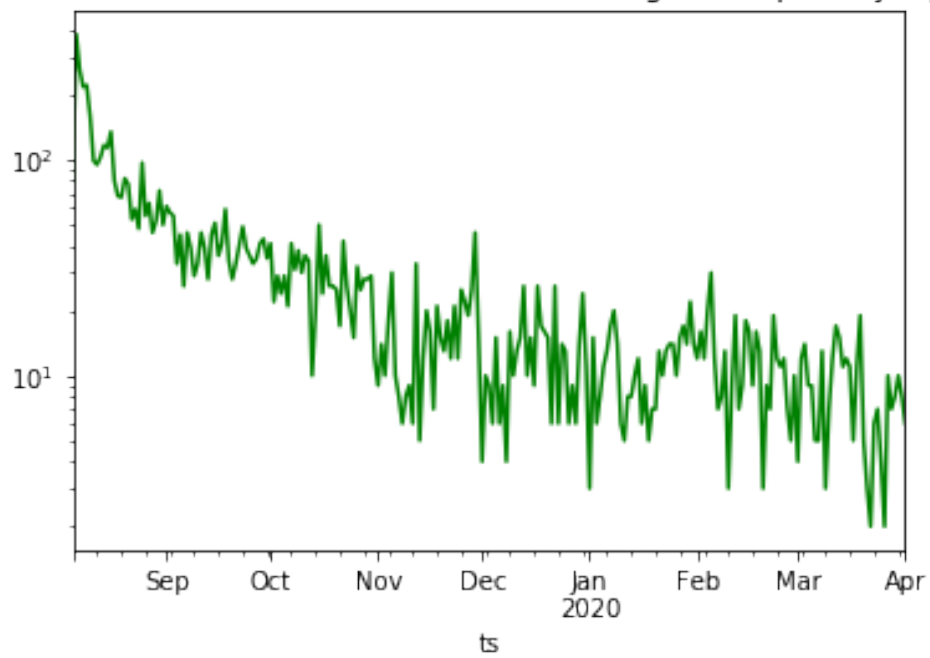
	Unnamed: 0	project	article	granularity	access	\
0	0	ur.wikipedia	Kashmir conflict	daily	all-access	
1	1	ur.wikipedia	Kashmir conflict	daily	all-access	
2	2	ur.wikipedia	Kashmir conflict	daily	all-access	
3	3	ur.wikipedia	Kashmir conflict	daily	all-access	
4	4	ur.wikipedia	Kashmir conflict	daily	all-access	

	agent	views	date
0	all-agents	9	2019-01-01
1	all-agents	19	2019-01-02
2	all-agents	12	2019-01-03
3	all-agents	12	2019-01-04
4	all-agents	13	2019-01-05

Urdu Kashmir conflict Pageviews per day log scale

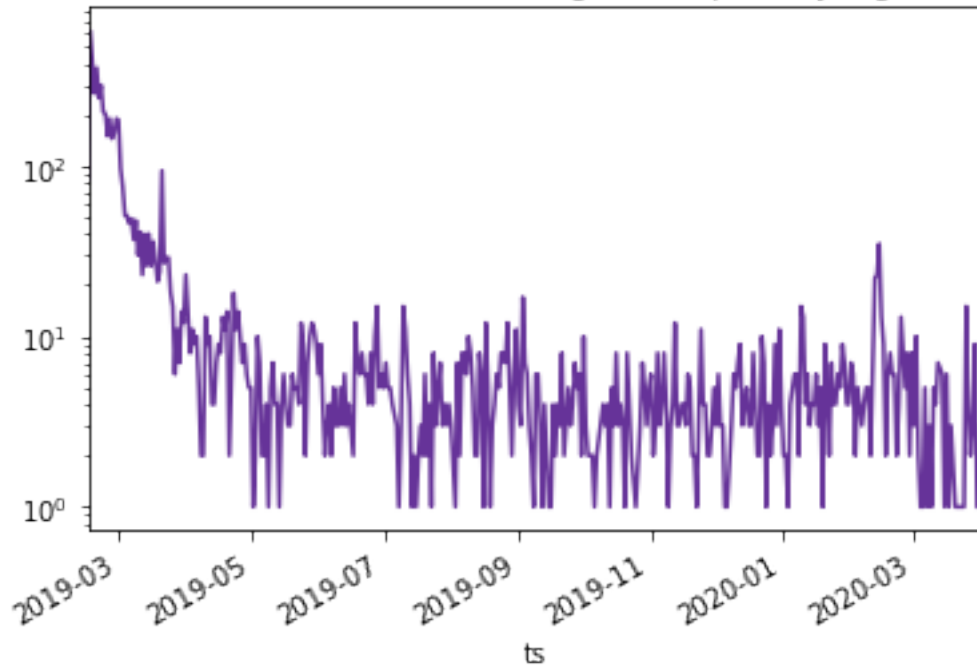


Urdu Article 370 of the Constitution of India Pageviews per day log scale

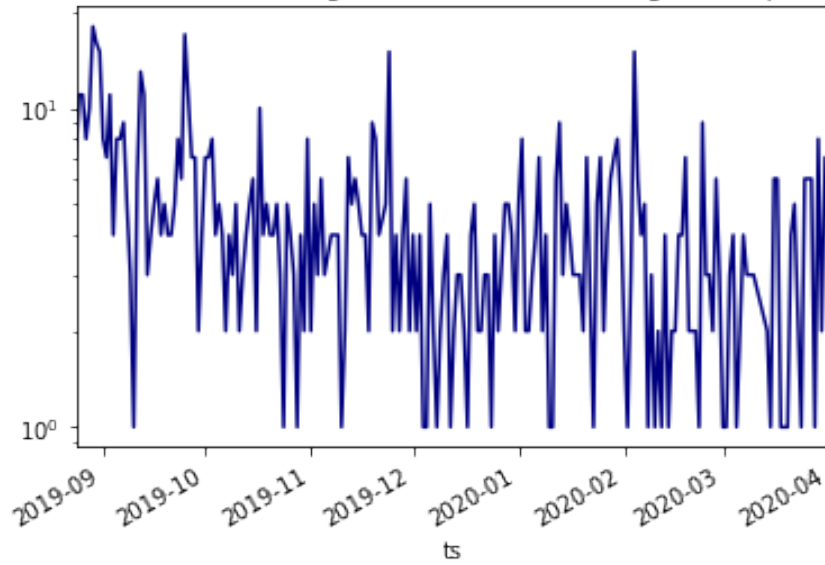




Urdu 2019 Pulwama attack Pageviews per day log scale

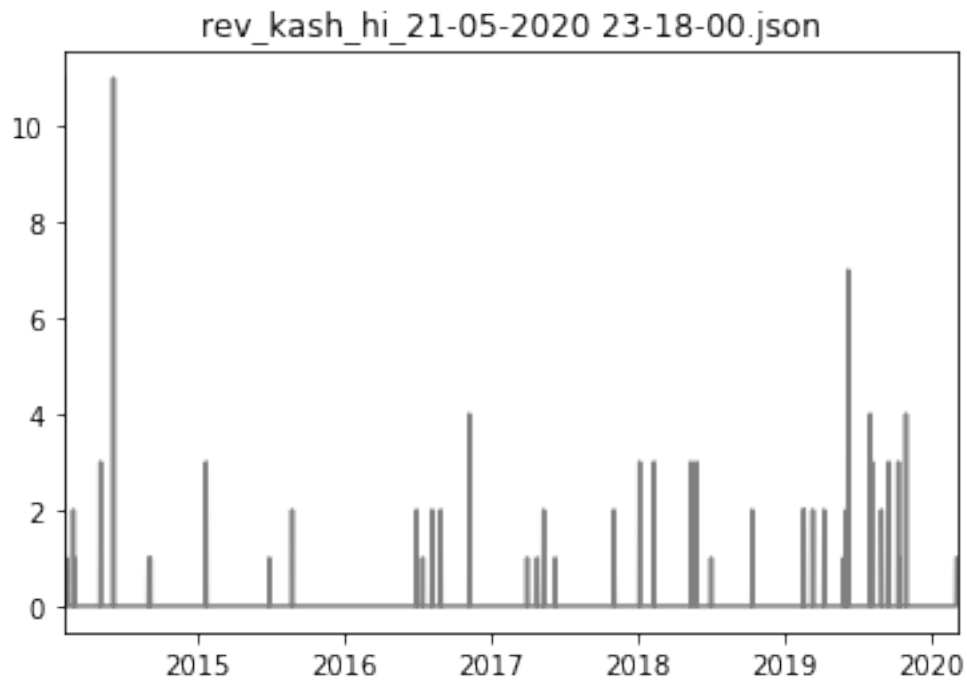
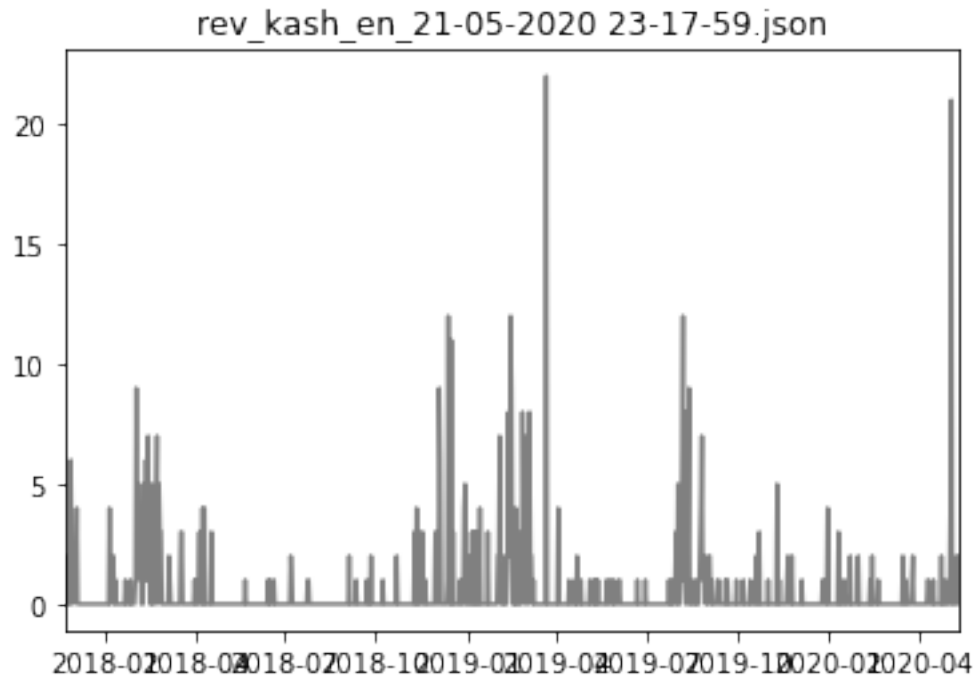


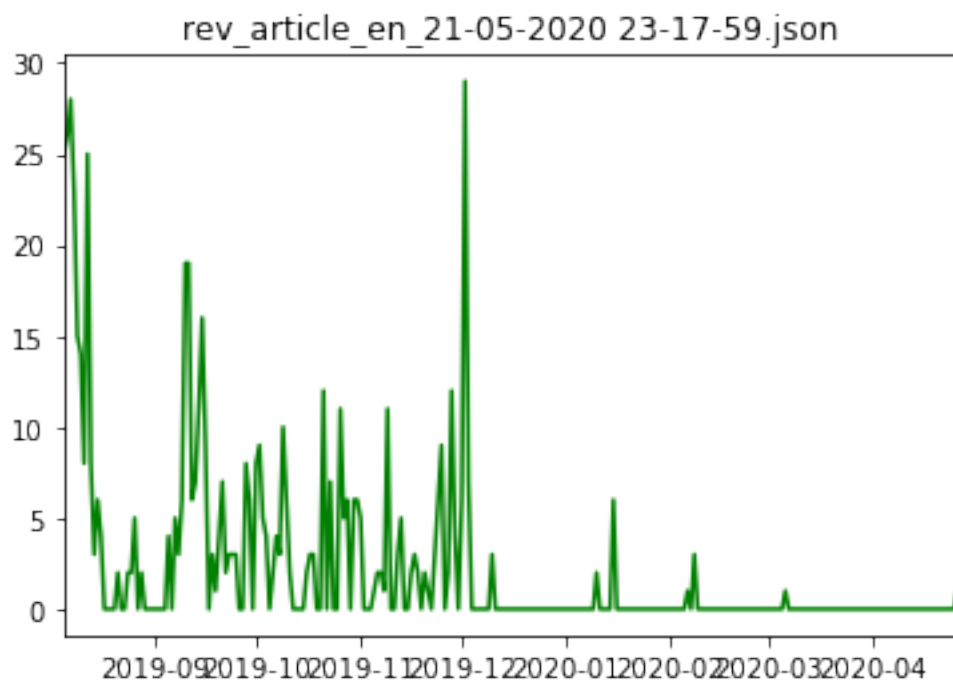
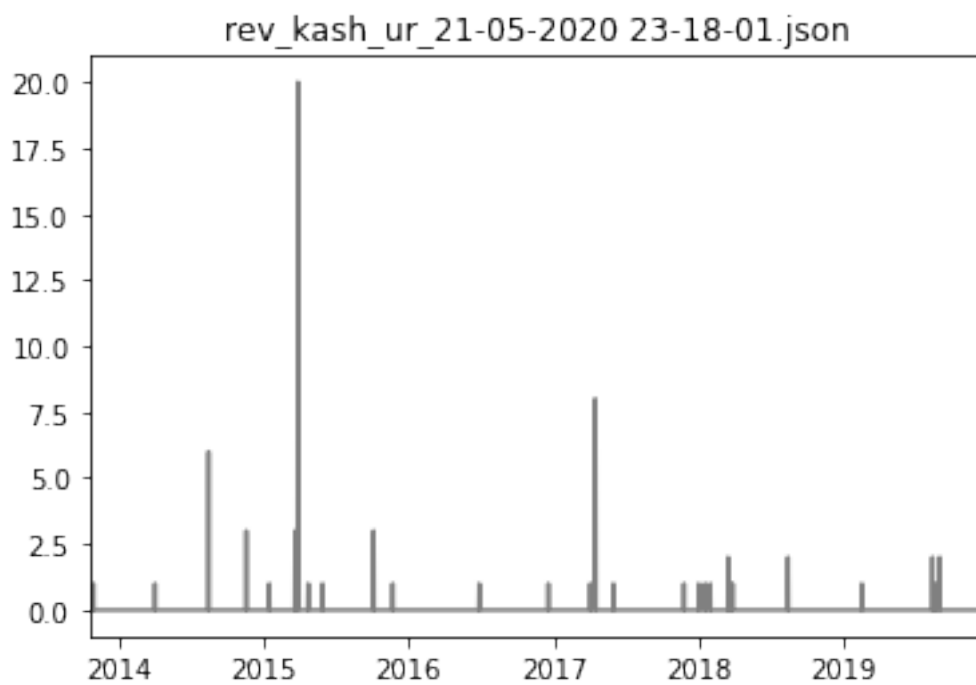
Urdu Jammu and Kashmir Reorganisation Act, 2019 Pageviews per day log scale

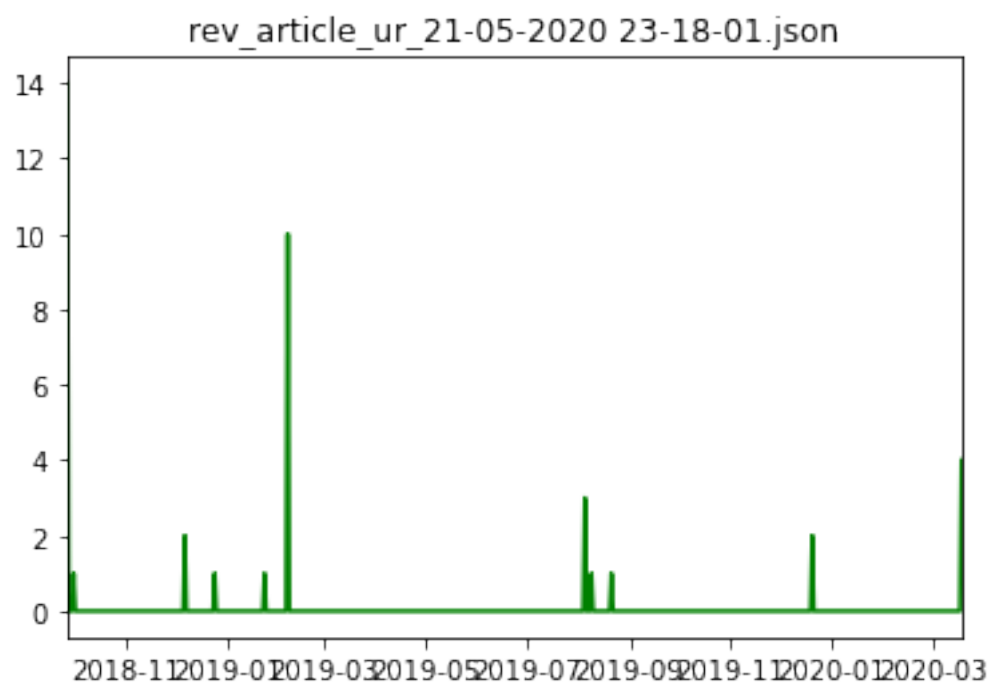
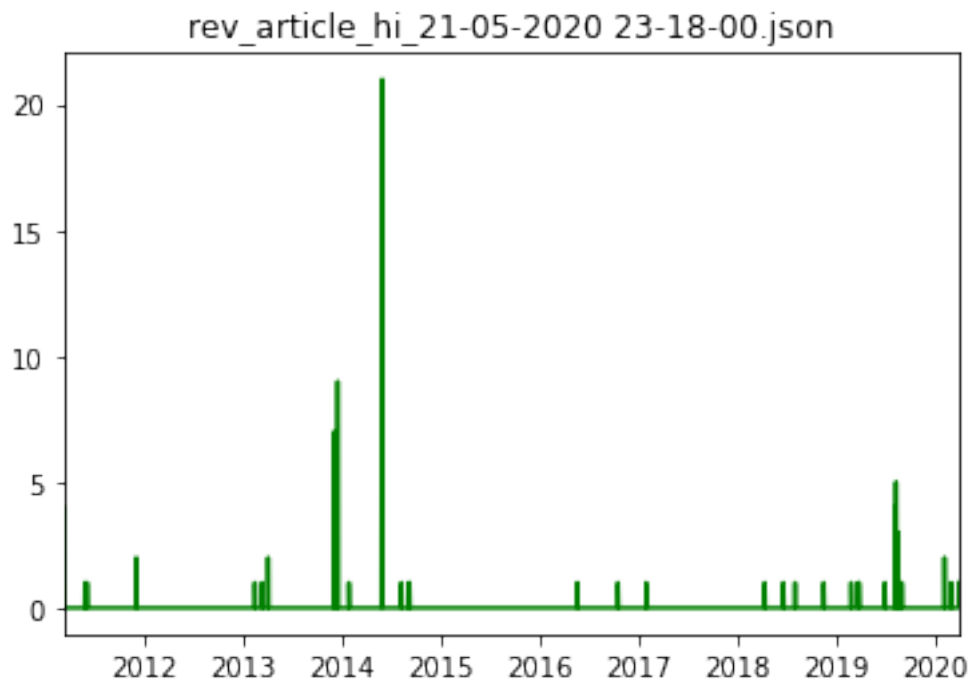


## 2 Edits per day timeseries

Also from article creation through March 2020. The original three articles: \* Kashmir Conflict \* Article 370 of the Constitution of India \* Insurgency in Jammu and Kashmir

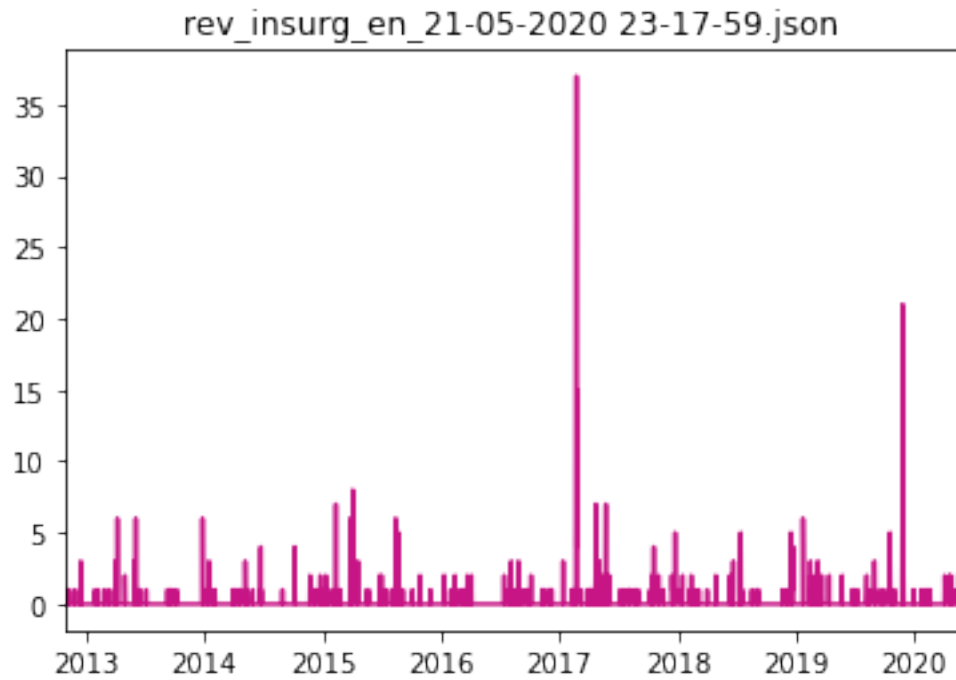




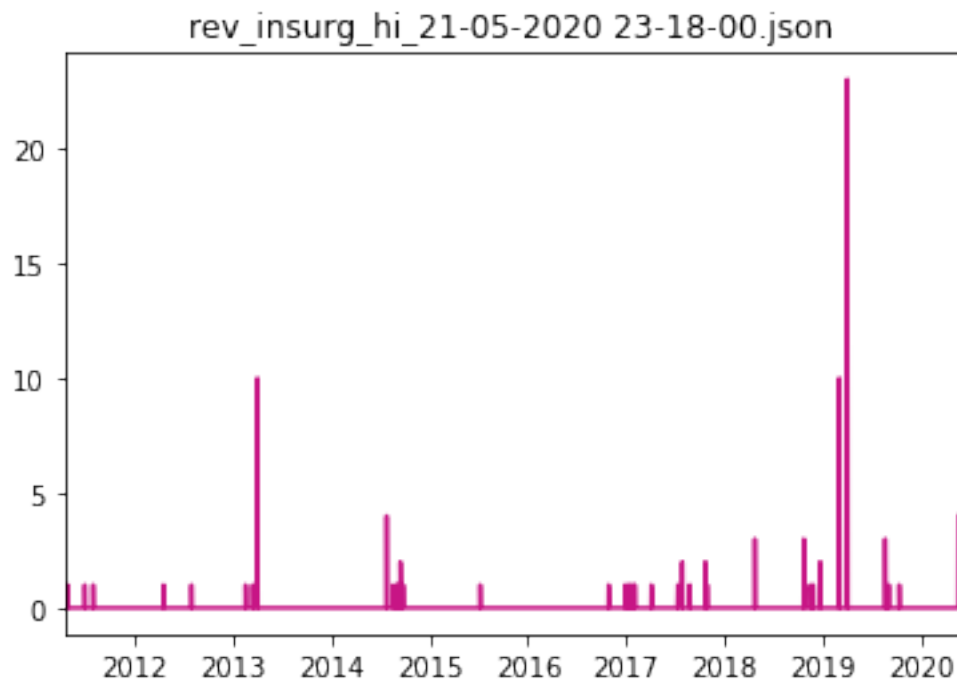


	revid	parentid	user	timestamp	size	comment	total
2012-10-26	1	1	1	1	1	1	1

2012-10-27	0	0	0	0	0	0	0
2012-10-28	0	0	0	0	0	0	0
2012-10-29	0	0	0	0	0	0	0
2012-10-30	0	0	0	0	0	0	0

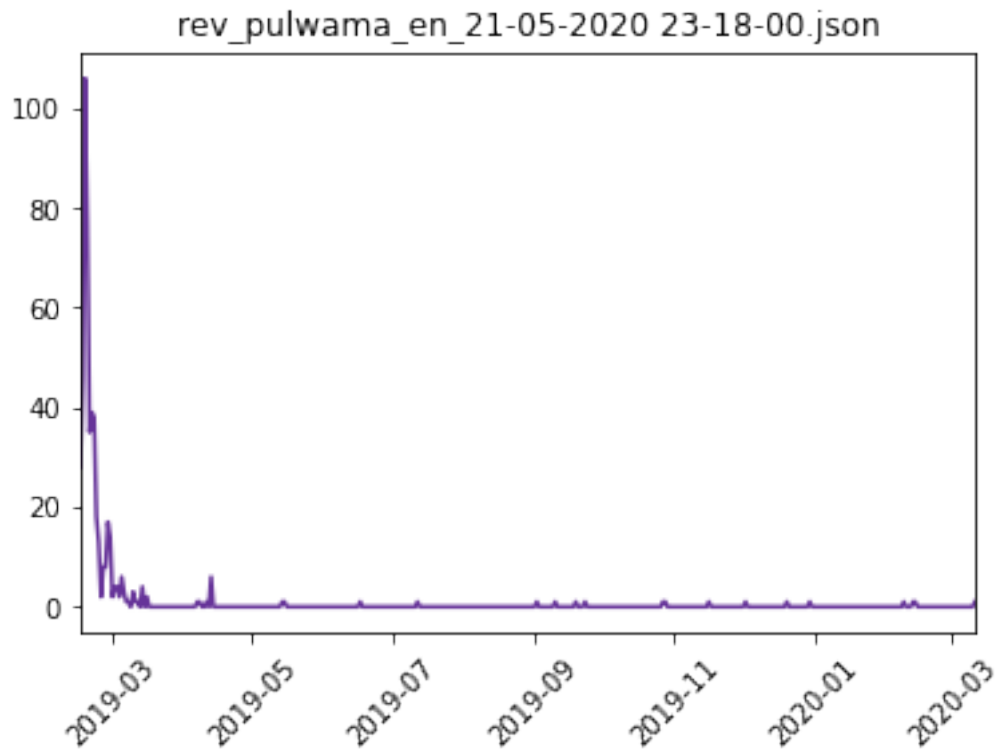


	revid	parentid	user	anon	timestamp	size	comment	total
2011-04-14	1	1	1	0	1	1	1	1
2011-04-15	0	0	0	0	0	0	0	0
2011-04-16	0	0	0	0	0	0	0	0
2011-04-17	0	0	0	0	0	0	0	0
2011-04-18	0	0	0	0	0	0	0	0



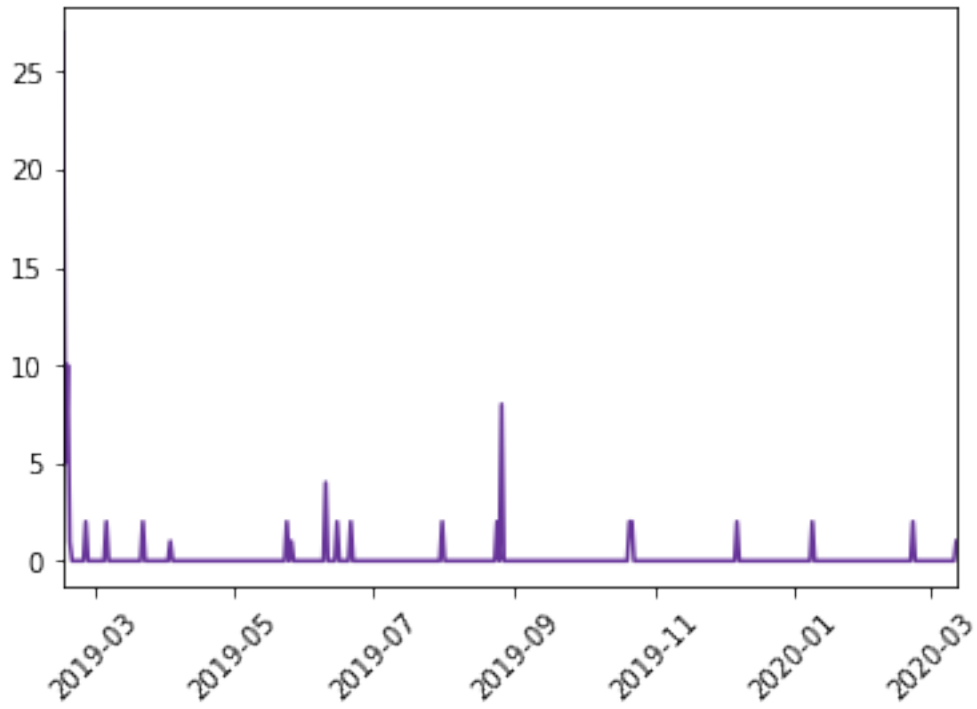
**“Pulwama Attack 2019”** pages

	revid	parentid	user	timestamp	size	comment	anon	total
2019-02-15	28	28	28	28	28	28	0	28
2019-02-16	46	46	46	46	46	46	0	46
2019-02-17	106	106	106	106	106	106	0	106
2019-02-18	67	67	67	67	67	67	10	77
2019-02-19	35	35	35	35	35	35	0	35



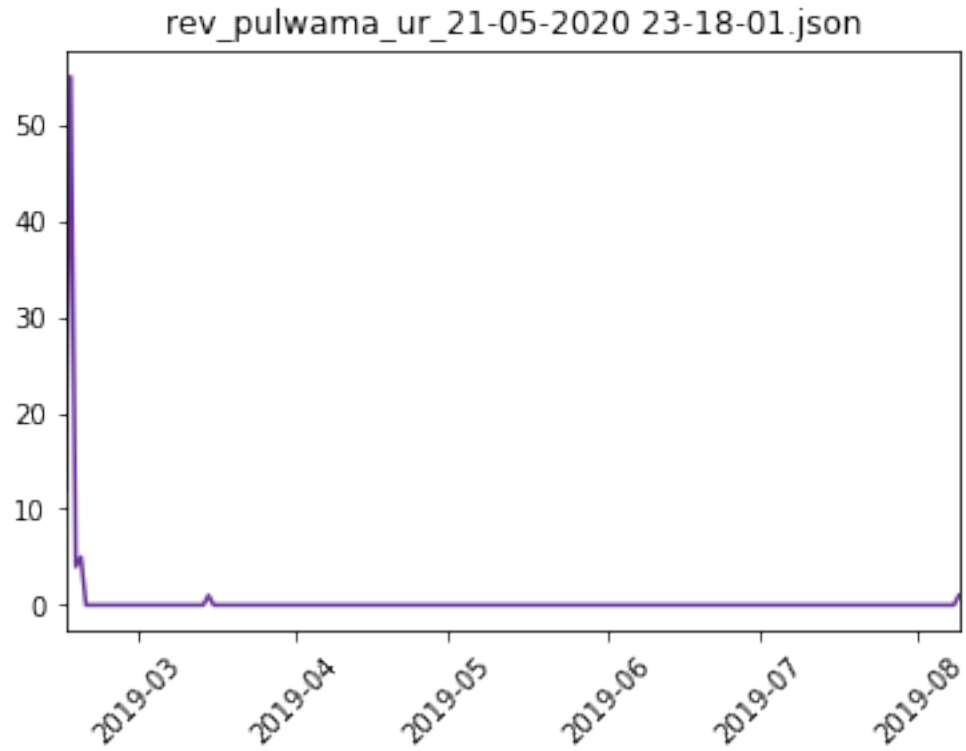
	revid	parentid	user	timestamp	size	comment	anon	total
2019-02-15	27	27	27	27	27	27	0	27
2019-02-16	5	5	5	5	5	5	0	5
2019-02-17	7	7	7	7	7	7	3	10
2019-02-18	1	1	1	1	1	1	0	1
2019-02-19	0	0	0	0	0	0	0	0

rev\_pulwama\_hi\_21-05-2020 23-18-00.json

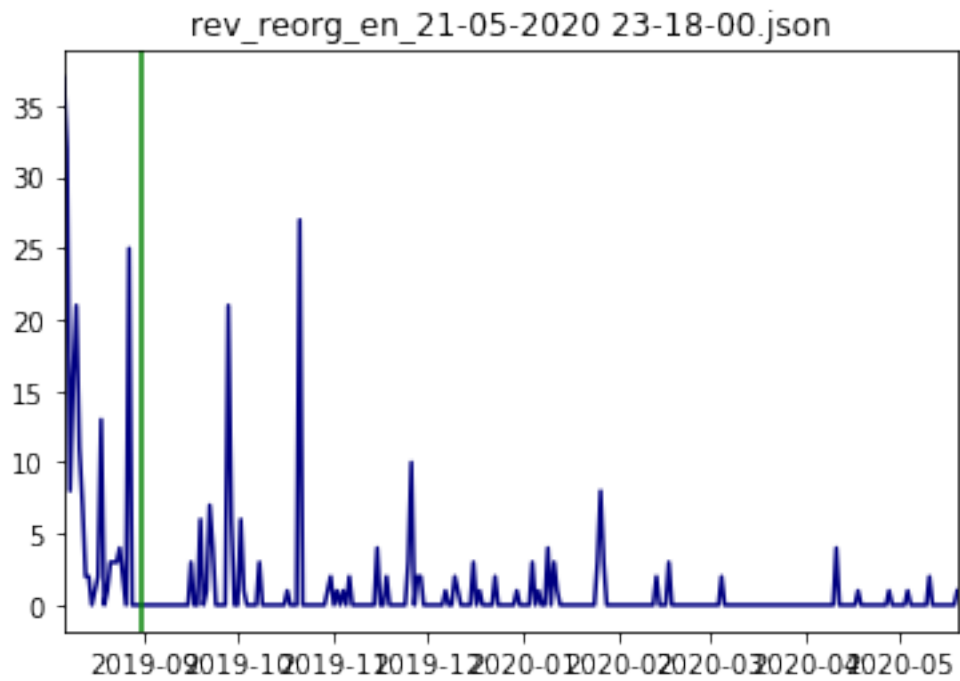


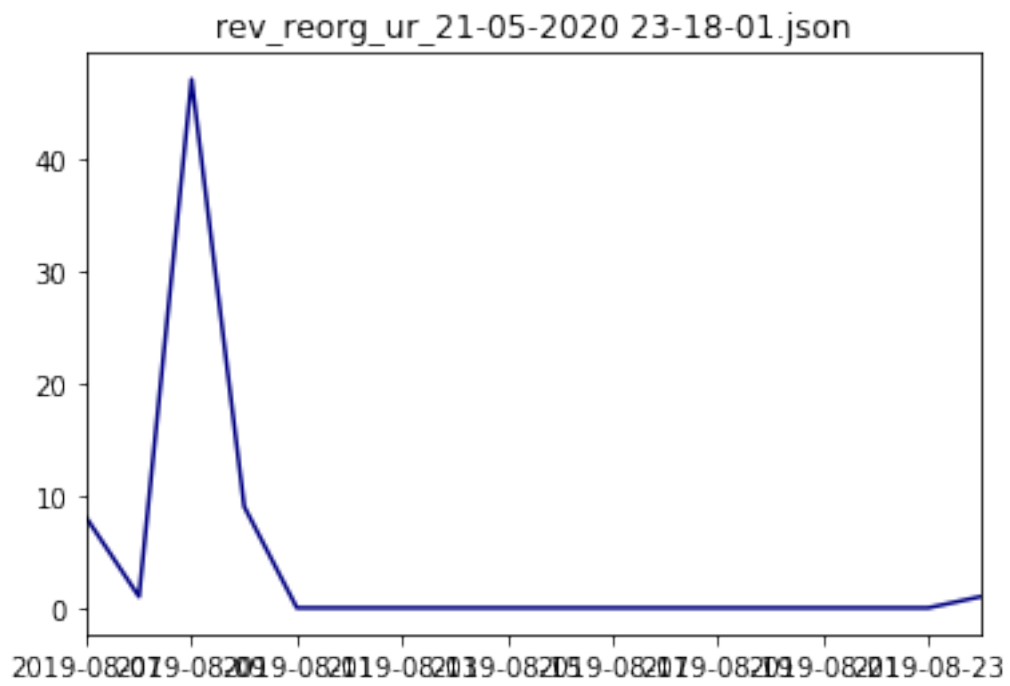
	revid	parentid	user	timestamp	size	comment	total
2019-02-15	12	12	12	12	12	12	12
2019-02-16	55	55	55	55	55	55	55
2019-02-17	4	4	4	4	4	4	4
2019-02-18	5	5	5	5	5	5	5
2019-02-19	0	0	0	0	0	0	0





“Reorganization of Jammu & Kashmir” pages





```
'for f in os.listdir(mypath):\n    if fnmatch.fnmatch(f,\n\n\\'rev_article_en_*.json\\'):\n        file = f\n        \\narticleEnPath =\n\n\"../data/revisions/\" + file\n        \\ndfEdits =\n\ngetEdits(articleEnPath)\n\\ndfEdits.tail()'
```

### 3 TLCC plots overlaid

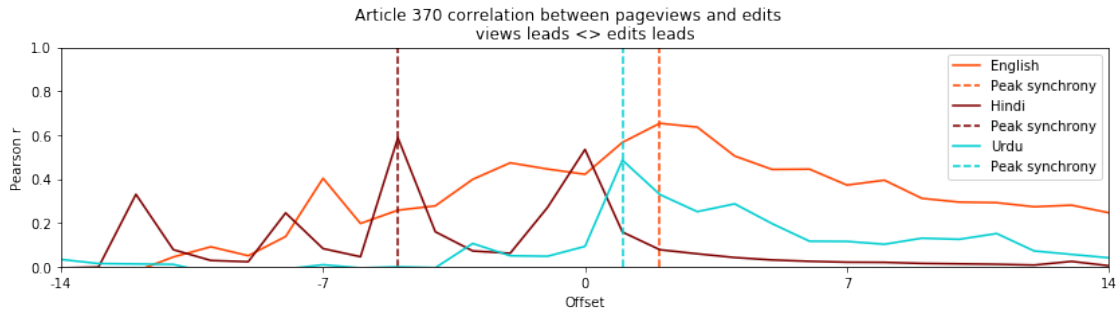
...and plots of views vs edits, with significant events. For each article, there is one time-lagged cross-correlation plot (these take into account whole lifespan of article) and a time series plot for each language of just 2019 (edits and page-views).

- Article 370=====

**NOTE:** There's something stupid going on that prevents me from getting the whole page-views series for the Urdu version of Article 370. I do have all the data somewhere. Will fix this.

The PostScript backend does not support transparency; partially transparent artists will be rendered opaque.

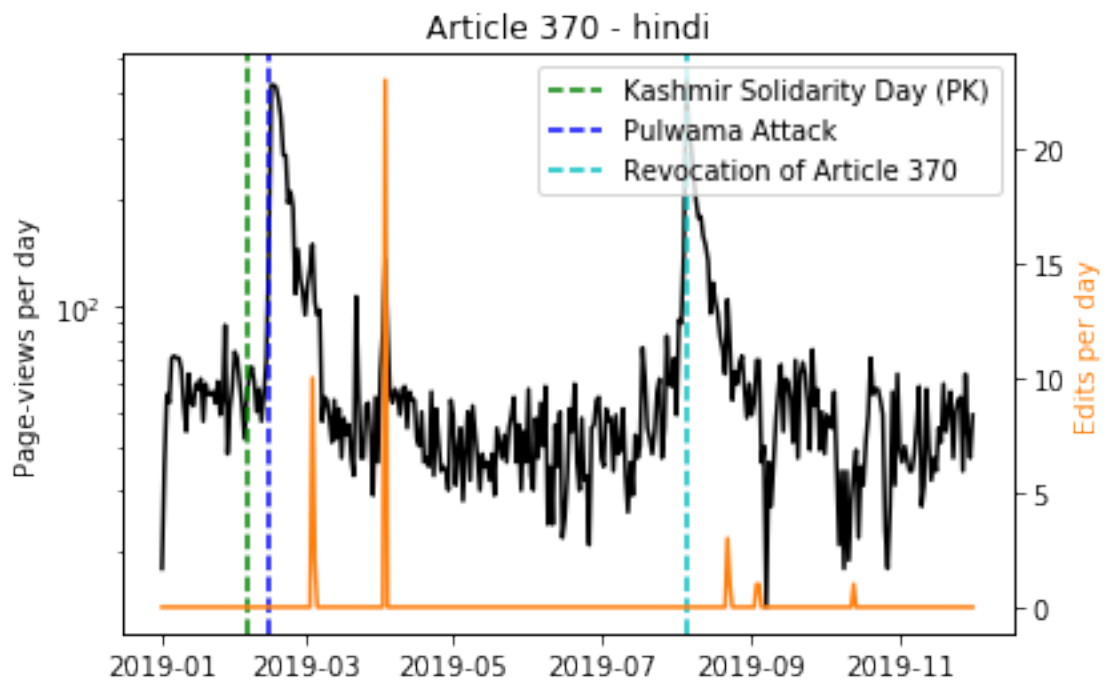
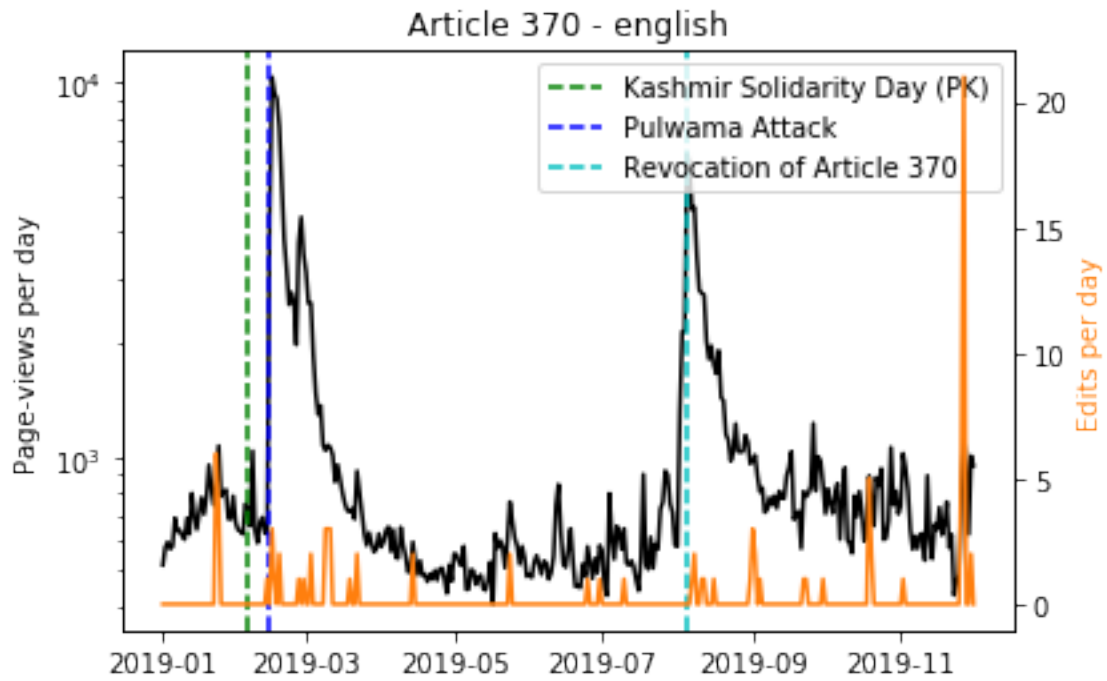
The PostScript backend does not support transparency; partially transparent artists will be rendered opaque.

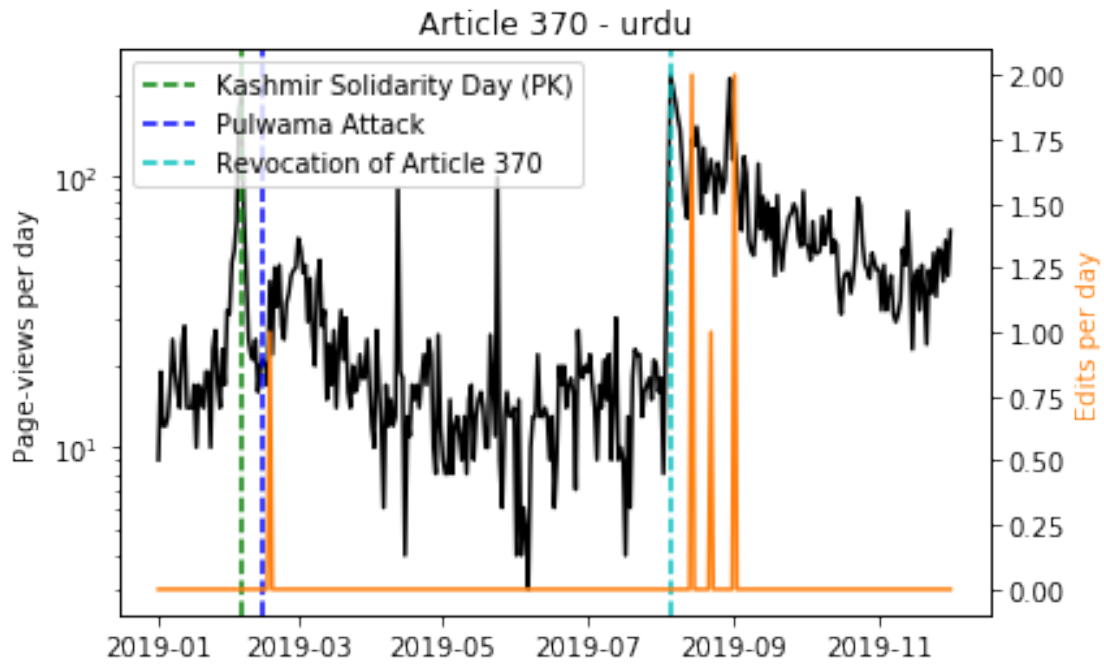
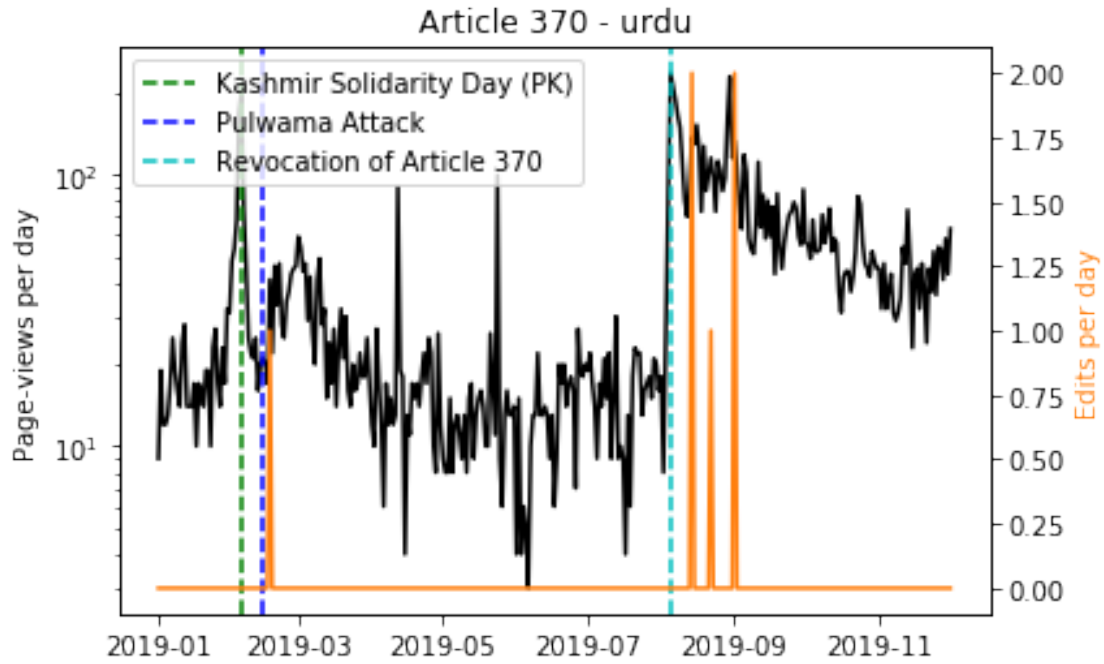


	revid	parentid	user	timestamp	size	comment	anon	total
2019-08-13	0	0	0	0	0	0	0	0
2019-08-14	1	1	1	1	1	1	1	2
2019-08-15	0	0	0	0	0	0	0	0
2019-08-16	0	0	0	0	0	0	0	0
2019-08-17	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
2019-12-16	0	0	0	0	0	0	0	0
2019-12-17	0	0	0	0	0	0	0	0
2019-12-18	0	0	0	0	0	0	0	0
2019-12-19	0	0	0	0	0	0	0	0
2019-12-20	1	1	1	1	1	1	0	1

[130 rows x 8 columns]

2013-10-18  
True





- Reorganisation Act=====

/home/molly/anaconda3/lib/python3.7/site-packages/numpy/lib/function\_base.py:2522: RuntimeWarning: Degrees of freedom <=

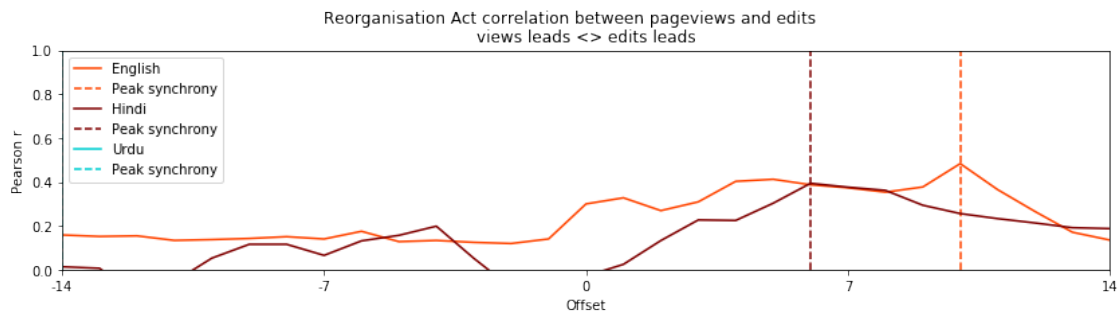
```

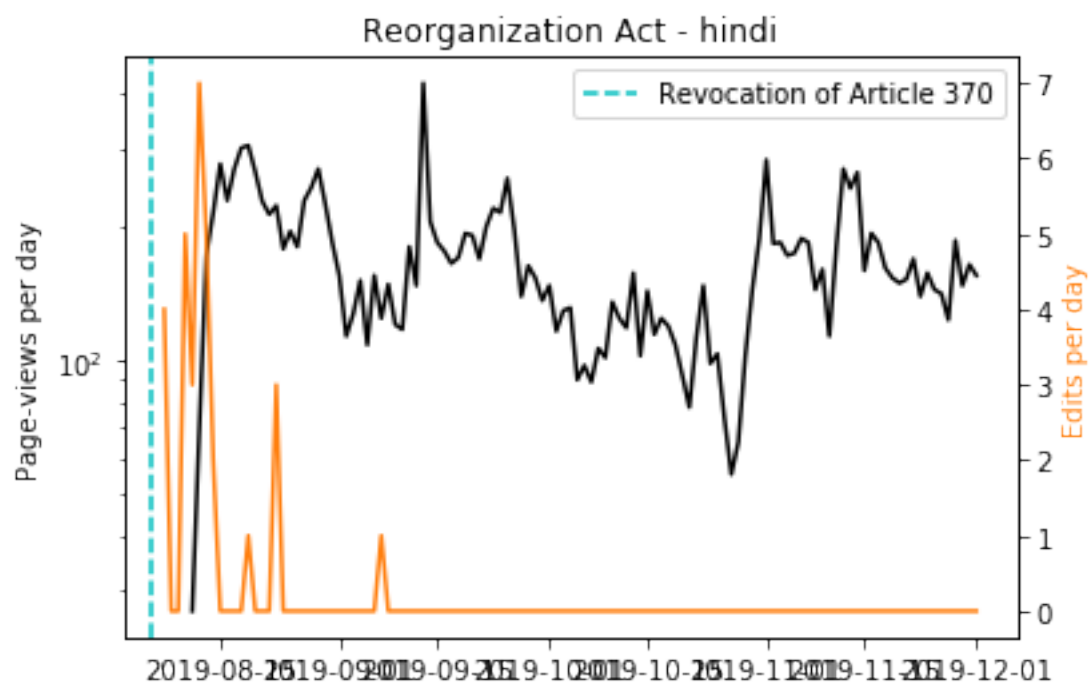
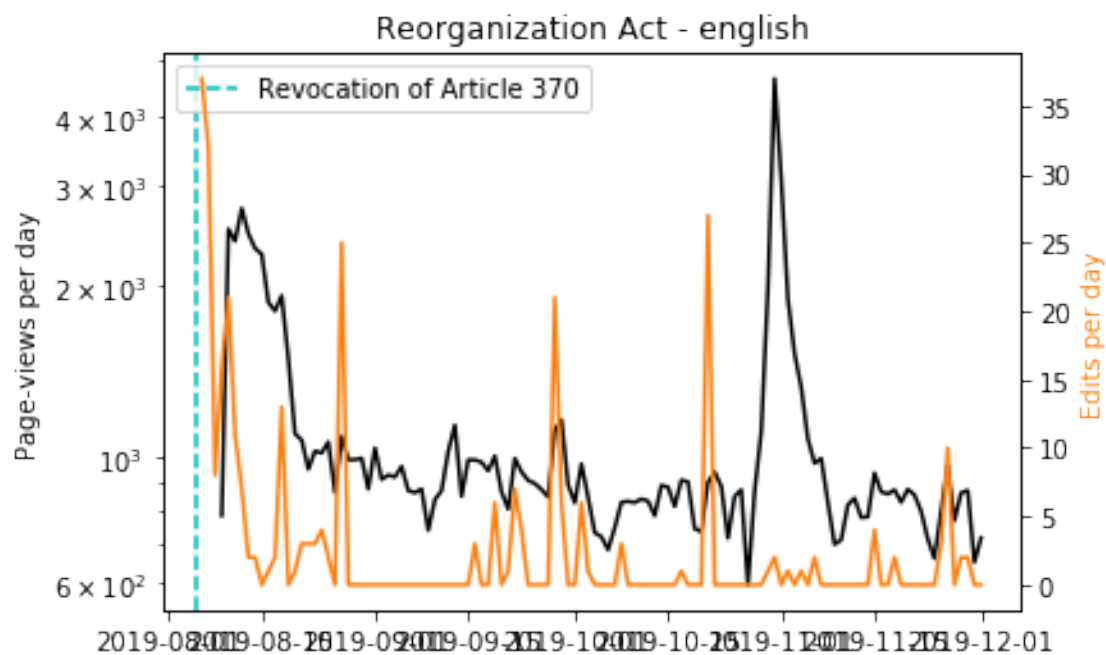
0 for slice
  c = cov(x, y, rowvar)
/home/molly/anaconda3/lib/python3.7/site-
packages/numpy/lib/function_base.py:2451: RuntimeWarning: divide by zero
encountered in true_divide
  c *= np.true_divide(1, fact)

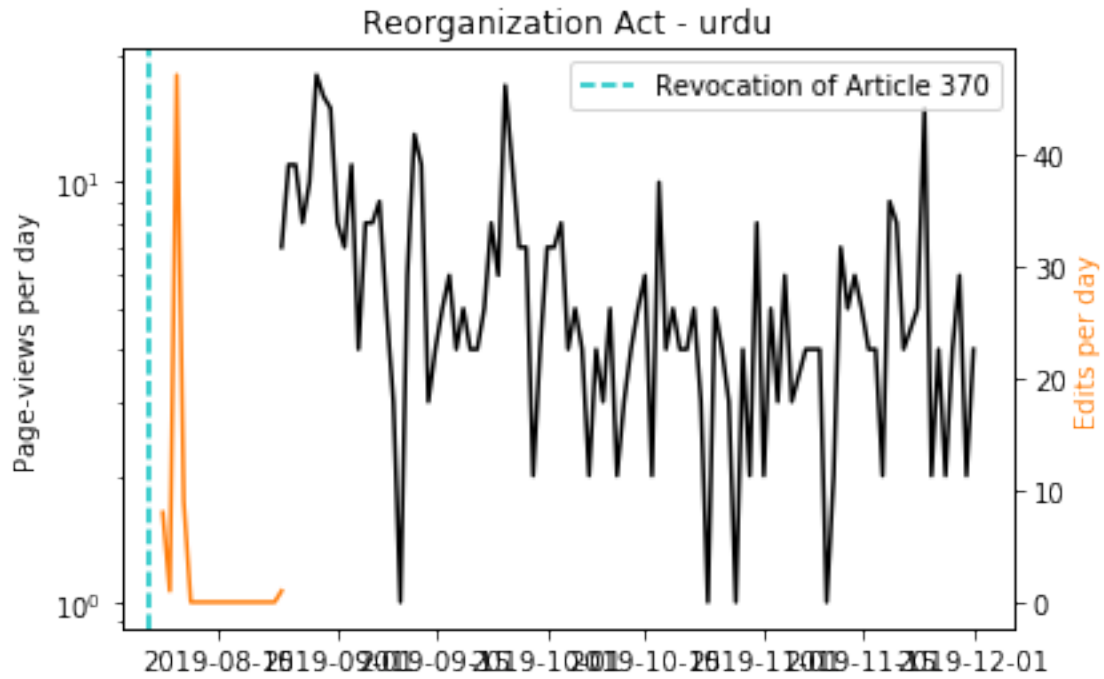
```

	lang	pearson	offset
0	en	0.158154	-14
1	en	0.151425	-13
2	en	0.154047	-12
3	en	0.133342	-11
4	en	0.137080	-10
..	...	...	...
82	ur	NaN	10
83	ur	NaN	11
84	ur	NaN	12
85	ur	NaN	13
86	ur	NaN	14

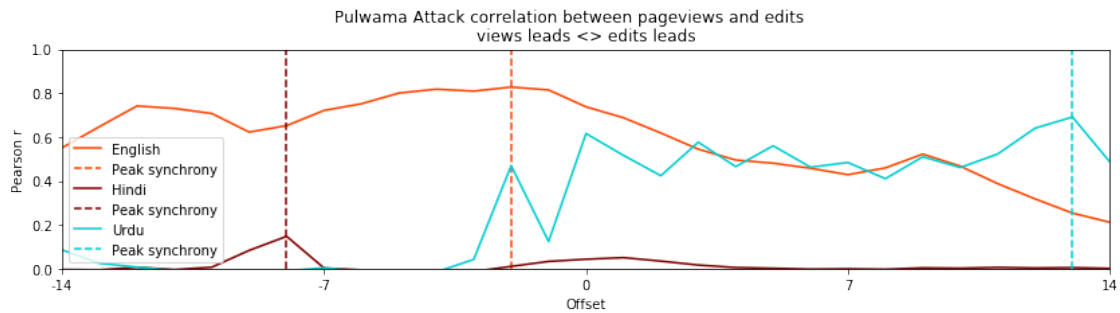
[87 rows x 3 columns]



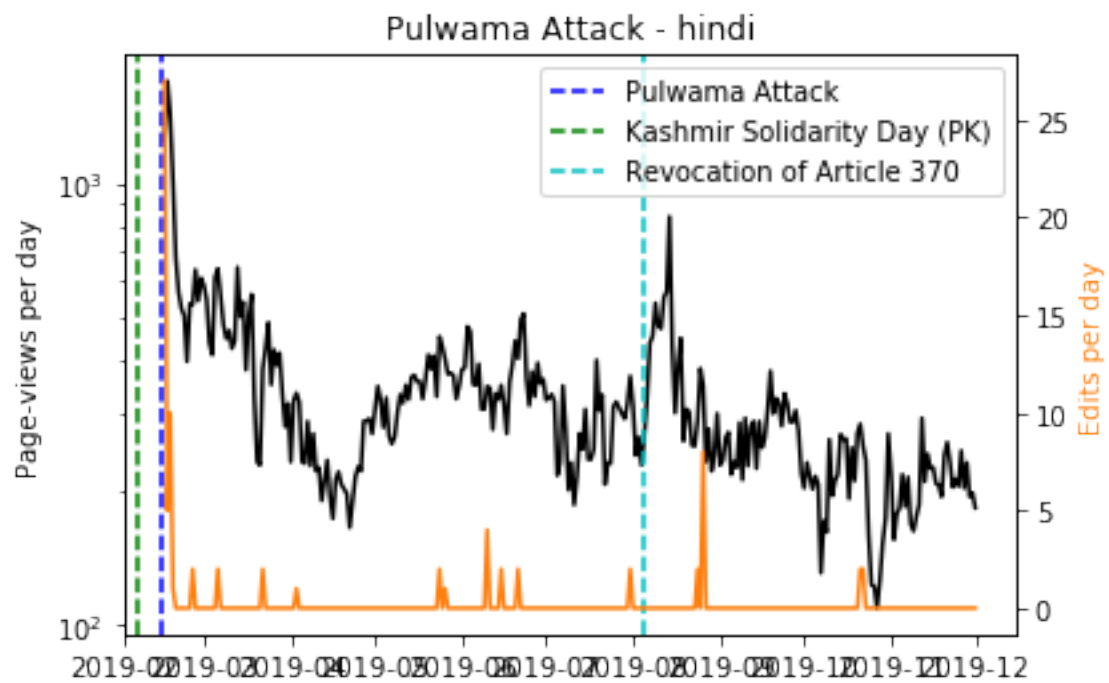
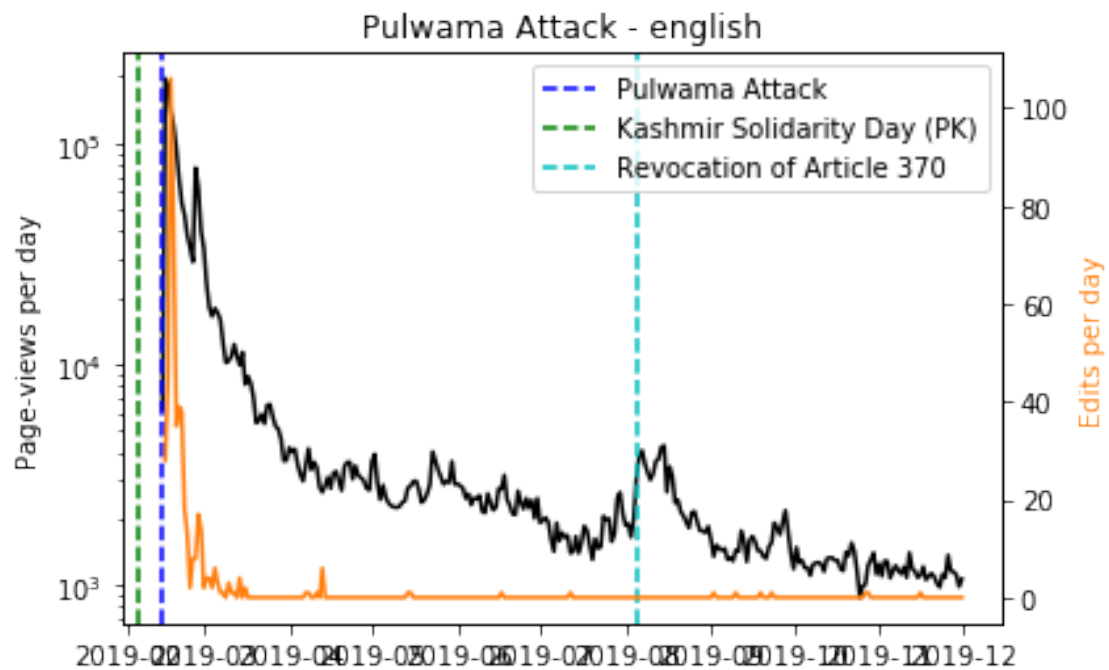


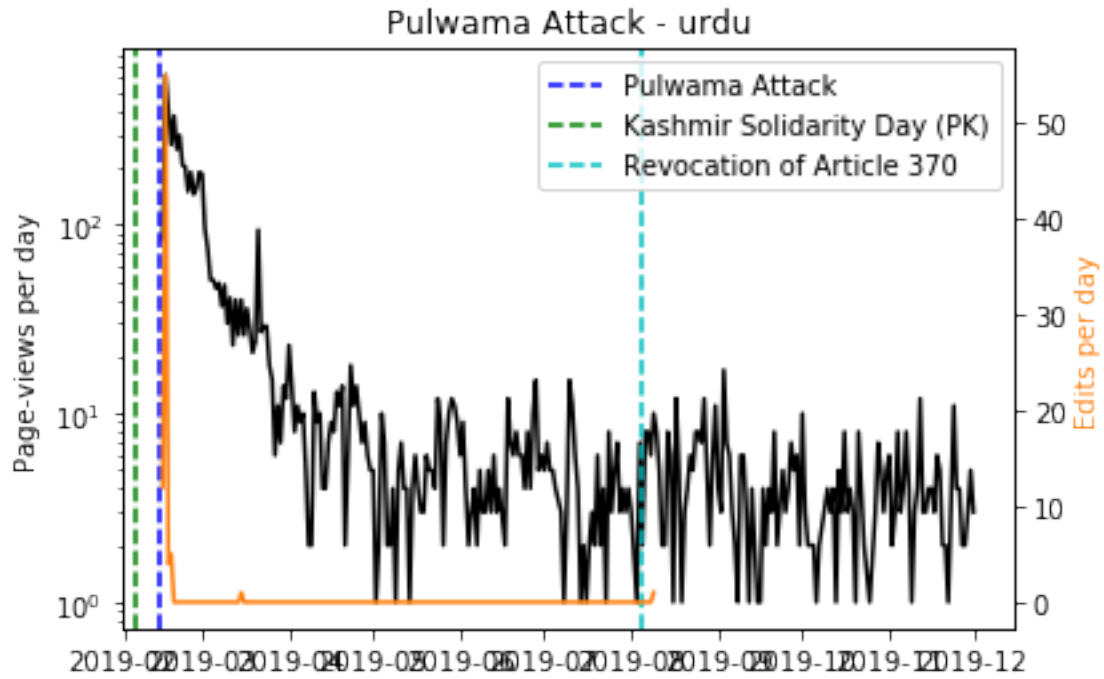


- Pulwama attack=====









- Kashmir conflict=====

