



+ Code

+ Text

Connect ▾

Editing

⤴



▼ Import

```
[ ] import numpy as np
import pandas as pd
import string
```

```
[ ] df=pd.read_json('reports.json',lines=True)
```

```
[ ] df.head()
```

	count	url	name	text
0	0	https://www.cdc.gov/mmwr/volumes/70/wr/mm7015...	Update: COVID-19 Pandemic–Associated Changes i...	During March 29–April 25, 2020, emergency depa...
1	1	https://www.cdc.gov/mmwr/volumes/70/wr/mm7015...	Factors Associated with Participation in Eleme...	During December 3, 2020–January 31, 2021, CDC,...
2	2	https://www.cdc.gov/mmwr/volumes/70/wr/mm7015...	Trends in Racial and Ethnic Disparities in COV...	Persons from racial and ethnic minority groups...
3	3	https://www.cdc.gov/mmwr/volumes/70/wr/mm7015...	Emergency Department Visits for COVID-19 by Ra...	Hispanic or Latino (Hispanic), non-Hispanic Bl...
4	4	https://www.cdc.gov/mmwr/volumes/70/wr/mm7015...	Notes from the Field: Update on Excess Deaths ...	Estimates of excess deaths, defined as the num...

▼ Preprocesing/Cleaning

```
[ ]
```

```
[ ] def removePunct(txt):
    txt_nopunct="".join([c for c in txt if c not in string.punctuation])
    return txt_nopunct
```

