

```
In [28]: import json
import pandas as pd
```

```
In [67]: with open('new-fine.json') as f:
data = json.load(f)
```

```
In [30]: df = pd.DataFrame(data, columns=["Country", "Authority", "Date", "Fine",
", "Controller", "Category", "QuotedArt", "Type", "Summary", "Info"])
```

```
In [40]: df = df[df.Fine != 'Unknown']
df = df[df.Type != 'Unknown']
```

```
In [41]: df[['Fine']] = df[['Fine']].apply(lambda x: x.str.replace('Euro', '')
.replace(' ', ''))
df[['Fine']] = df[['Fine']].apply(lambda x: x.str.replace(' ', ''))
```

```
In [42]: df[['Fine']] = df[['Fine']].apply(lambda x: x.str.replace(',', ''))
df = df[df.Fine != 'ohnePunkte']

# df[['Fine']] = df[['Fine']].apply(lambda x: x.str.replace(',', '').a
stype(int))
```

```
In [46]: df[['Fine']] = df[['Fine']].apply(lambda x: x.str.replace(',', '').ast
pe(int))
# df.Fine.unique()
```

```
In [48]: df
```

```
Out[48]:
```

	Country	Authority	Date	Fine	Controller	Category	QuotedArt	T
0	SPAIN	Spanish Data Protection Authority (aepd)	2021-01-21	50000	Alterna Operador Integral S.L.	Transportation and Energy	Art. 6 (1) b) GDPR	Insuffic legal b: for c proces:
1	SPAIN	Spanish Data Protection Authority (aepd)	2021-01-20	1200	Individual	Individuals and Private Associations	Art. 5 (1) c) GDPR	N complia \ gen c proces:
2	NORWAY	Norwegian Supervisory Authority (Datatilsynet)	2021-01-19	9700	Aquateknikk AS	Industry and Commerce	Art. 5 GDPR, Art. 6 GDPR	Insuffic legal b: for c proces:
3	NORWAY	Norwegian Supervisory Authority (Datatilsynet)	2021-01-14	38600	Coop Finnmark SA	Industry and Commerce	Art. 5 (1) a) GDPR, Art. 6 GDPR	Insuffic legal b: for c proces:
4	SPAIN	Spanish Data Protection Authority (aepd)	2021-01-13	6000000	Caixabank S.A.	Finance, Insurance and Consulting	Art. 6 GDPR, Art. 13 GDPR, Art. 14 GDPR	Insuffic legal b: for c proces:
...
507	SPAIN	Spanish Data Protection Authority (aepd)	Unknown	0000	Restaurant (SANTI)	Employment	Art. 5 (1) a) GDPR,	Insuffic legal b:

527	SPAIN	Protection Authority (aepd)	Unknown	9000	(SAN III 3000, S.L.)	Employers	Art. 6 GDPR	- for c proces:
528	SPAIN	Spanish Data Protection Authority (aepd)	Unknown	60000	ENDESA (energy supplier)	Transportation and Energy	Art. 5 (1) f) GDPR	Insuffic legal b: for c proces:
529	SPAIN	Spanish Data Protection Authority (aepd)	Unknown	27000	Vodafone España, S.A.U.	Media, Telecoms and Broadcasting	Art. 5 (1) d) GDPR	Insuffic fulfillm of c subje rig
530	SPAIN	Spanish Data Protection Authority (aepd)	Unknown	60000	Debt collecting agency (GESTIÓN DE COBROS, YO ...	Finance, Insurance and Consulting	Art. 5 (1) f) GDPR	Insuffic legal b: for c proces:
531	SPAIN	Spanish Data Protection Authority (aepd)	Unknown	5000	Vodafone España, S.A.U.	Media, Telecoms and Broadcasting	Art. 5 (1) d) GDPR	N complia , gen c proces:

523 rows × 10 columns

```
In [51]: df = df.reset_index(drop=True)
df.to_csv("new-fine-db.csv", index=False)
```

```
In [65]: test = df.groupby("Country")["Fine"].count()
test = test.reset_index()
test.to_csv("new-fine-summary.csv", index=False)
test
```

Out[65]:

	Country	Fine
0	AUSTRIA	10
1	BELGIUM	21
2	BULGARIA	20
3	CYPRUS	11
4	CZECH REPUBLIC	12
5	DENMARK	8
6	ESTONIA	5
7	FINLAND	5
8	FRANCE	12
9	GERMANY	25
10	GREECE	12
11	HUNGARY	36
12	ICELAND	2
13	IRELAND	4

14	ISLE OF MAN	1
15	ITALY	41
16	LATVIA	4
17	LITHUANIA	2
18	MALTA	1
19	NORWAY	18
20	POLAND	18
21	PORTUGAL	4
22	ROMANIA	47
23	SLOVAKIA	2
24	SPAIN	175
25	SWEDEN	17
26	THE NETHERLANDS	6
27	UNITED KINGDOM	4

In []: