

```
In [2]: import pandas as pd
```

```
In [4]: import numpy as np
```

```
In [8]: import matplotlib.pyplot as plt
```

```
In [11]: %matplotlib inline
```

```
In [14]: from sklearn import metrics
```

```
In [17]: from sklearn.neighbors import KNeighborsClassifier
```

```
In [19]: from sklearn.linear_model import LogisticRegression
```

```
In [21]: from sklearn.model_selection import train_test_split
```

```
In [23]: df = pd.read_csv('https://raw.githubusercontent.com/okfn/dataportals.org/master/data/portals.csv')
```

```
In [25]: print(df)
```

```
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
..      ...
599      CKAN API
600      NaN
601 Custom/in-house implementation
602 Custom/in-house implementation
603      NaN

full_metadata_download
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
..      ...
599      NaN
600      NaN
```

In [27]: `df.isnull()`

Out[27]:

	name	title	url	author	publisher	issued	publisher_classification	description	tags	license_id	...	place	lo
0	False	False	False	False	False	True	True	False	False	True	...	False	
1	False	False	False	False	False	False	False	False	False	False	...	False	
2	False	False	False	False	False	True	True	False	False	True	...	False	
3	False	False	False	False	False	True	False	False	False	True	...	False	
4	False	False	False	False	False	True	False	False	False	True	...	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	
599	False	False	False	False	False	False	False	False	False	False	...	False	
600	False	False	False	False	False	True	False	False	False	False	...	False	
601	False	False	False	False	False	False	False	False	False	False	...	False	
602	False	False	False	False	False	True	False	False	False	True	...	False	
603	False	False	False	False	False	True	False	False	False	True	...	False	

604 rows × 22 columns

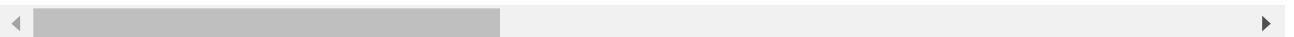


In [29]: `df.describe()`

Out[29]:

	name	title	url	author	publisher	issued	publisher_class
count	604	604	604	539	543	98	
unique	600	598	600	497	506	88	
top	state_of_washington	Washington	http://dados.recife.pe.gov.br	African Development Bank Group	African Development Bank Group	27/08/2015	Gov
freq	4	4	2	21	22	2	

4 rows × 22 columns



In [31]: `size = df.size`

In [33]: `shape = df.shape`

In [35]: `size`

Out[35]: 13288

In [37]: `shape`

Out[37]: (604, 22)

In [39]: `df.ndim = df["name"].ndim`

```
In [42]: series_ndim = df["name"].ndim
```

```
In [44]: print("size = {}\nShape = {}\n Shape[0] * Shape[1] = {}".format(size, shape, shape[0] * shape[1]))

size = 13288
Shape = (604, 22)
Shape[0] * Shape[1] = 13288
```

```
In [46]: dataTypeSeries = df.dtypes
```

```
In [48]: print("Data Type of Each Column of DataFrame is : ")
print(dataTypeSeries)
```

```
Data Type of Each Column of DataFrame is :
name                object
title               object
url                 object
author              object
publisher            object
issued              object
publisher_classification object
description          object
tags                object
license_id           object
license_url          object
license_notes        object
place                object
location             object
country              object
language             object
status               object
metadatacreated      object
generator            object
api_endpoint         object
api_type             object
full_metadata_download object
dtype: object
```

```
In [50]: dummies = pd.get_dummies(df.author)
```

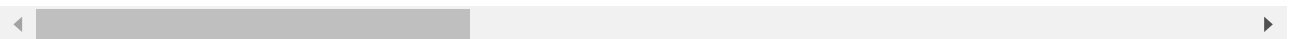
```
In [54]: merged = pd.concat([df, dummies], axis = 'columns')
```

```
In [56]: merged.drop(['author'], axis = 'columns')
```

```
Out[56]:
```

	name	title	url	publisher	issued	publisher_classi
0	a2gov_org	Ann Arbor, Michigan	http://www.a2gov.org/services/data/Pages/default...	City of Ann Arbor	NaN	
1	acikveri-sahinbey-bel-tr	Açık Veri Portali - Test Yayını	http://acikveri.sahinbey.bel.tr/dataset	SahinBey Belediyesi	31/01/2015	Govt
2	africa_open_data	Africa Open Data	http://africaopendata.org/	Africa Open Data	NaN	
3	ajuntament-de-tarragona	Open Data Tarragona	http://opendata.tarragona.cat/	Ajuntament de Tarragona	NaN	Govt
4	ajuntament-de-terassa	Open Data Terrassa	http://opendata.terassa.cat/	Ajuntament de Terrassa	NaN	Govt
...	...	...	...	...	...	...
599	open-data-nepal	Open Data Nepal	https://opendatanepal.com/	Open Knowledge Nepal	2018-03-03	Govt
600	stat-tj	Agency on Statistics under President of the Re...	https://www.stat.tj/en	The Statistical Agency under President of the ...	NaN	Govt
601	dadosjusbr	DadosJusBR	https://dadosjusbr.org/	Instituto Federal de Alagoas	2018-12-31	Govt
602	bratislava-opendata	Bratislava Open Data Portal	https://opendata.bratislava.sk	The city of Bratislava	NaN	Govt
603	sk-kosice-opendata	The city of Košice Open Data Portal	https://opendata.kosice.sk	The city of Košice	NaN	Govt

604 rows × 518 columns



In [57]: `print(merged)`

```

..          ...          ...          ...
599          0          0          0
600          0          0          0
601          0          0          0
602          0          0          0
603          0          0          0

http://garagelab.tumblr.com/ (http://garagelab.tumblr.com/) pinardag regione sardegna Ör
ebro Kommun
0          0          0          0          0
1          0          1          0          0
2          0          0          0          0
3          0          0          0          0
4          0          0          0          0
..          ...          ...          ...          ...
599          0          0          0          0
600          0          0          0          0
601          0          0          0          0
602          0          0          0          0
603          0          0          0          0

```