

```
#Profiling
```

```
df.shape
```

```
(3037, 62)
```

Validando la cantidad de registros y variables, se confirma que hay 3.037 registros y 62 variables

```
df.sample(n=10)
```

	id	url \
2020	1984466	https://www.tvmaze.com/episodes/1984466/new-ja...
951	2000049	https://www.tvmaze.com/episodes/2000049/ultima...
2153	1992679	https://www.tvmaze.com/episodes/1992679/rhyme-...
628	1971206	https://www.tvmaze.com/episodes/1971206/the-ru...
177	1945144	https://www.tvmaze.com/episodes/1945144/nobles...
2965	1998673	https://www.tvmaze.com/episodes/1998673/louder...
1768	1990512	https://www.tvmaze.com/episodes/1990512/black-...
2884	2007690	https://www.tvmaze.com/episodes/2007690/my-bes...
734	1976033	https://www.tvmaze.com/episodes/1976033/twiste...
1307	1972058	https://www.tvmaze.com/episodes/1972058/red-vs...

	name	season
number \		
2020	Road to Tokyo Dome Night 1	2020
88.0		
951	Episode 2	1
2.0		
2153	The Teaspoon Dance / Jaime's Bridge Has Fallen...	1
8.0		
628	Episode 20	1
20.0		
177	Blood Contract / Devote	1
9.0		
2965	Stuck In The Middle With You	3
1.0		
1768	Jatin Must Die	1
9.0		
2884	Episode 6	1
6.0		
734	Episode 22	1
22.0		
1307	Shattered	18
6.0		

	type	airdate	airtime	airstamp	runtime
...					
2020	regular	2020-12-21	18:30	2020-12-21T09:30:00+00:00	120.0
...					

951	regular	2020-12-10		2020-12-10T12:00:00+00:00	45.0
2153	regular	2020-12-22		2020-12-22T12:00:00+00:00	13.0
628	regular	2020-12-07	20:00	2020-12-07T12:00:00+00:00	60.0
177	regular	2020-12-02	00:00	2020-12-03T05:00:00+00:00	25.0
2965	regular	2020-12-31		2020-12-31T12:00:00+00:00	30.0
1768	regular	2020-12-18		2020-12-18T06:30:00+00:00	30.0
2884	regular	2020-12-30		2020-12-30T12:00:00+00:00	45.0
734	regular	2020-12-08	20:00	2020-12-08T12:00:00+00:00	45.0
1307	regular	2020-12-14		2020-12-14T12:00:00+00:00	10.0
...					

	_embedded.show.network.country.code	\
2020	NaN	
951	NaN	
2153	NaN	
628	NaN	
177	JP	
2965	NaN	
1768	NaN	
2884	NaN	
734	NaN	
1307	NaN	

	_embedded.show.network.country.timezone	\
2020	NaN	
951	NaN	
2153	NaN	
628	NaN	
177	Asia/Tokyo	
2965	NaN	
1768	NaN	
2884	NaN	
734	NaN	
1307	NaN	

	_embedded.show.network.officialSite	\
2020	NaN	
951	NaN	
2153	NaN	
628	NaN	
177	None	
2965	NaN	

1768	NaN
2884	NaN
734	NaN
1307	NaN

	_embedded.show._links.nextepisode.href
_embedded.show.webChannel \	
2020	<a href="https://api.tvmaze.com/episodes/2353245">https://api.tvmaze.com/episodes/2353245</a>
NaN	
951	NaN
NaN	
2153	NaN
NaN	
628	NaN
NaN	
177	NaN
NaN	
2965	NaN
NaN	
1768	NaN
NaN	
2884	NaN
NaN	
734	NaN
NaN	
1307	NaN
NaN	

	_embedded.show.image	_embedded.show.webChannel.country \
2020	NaN	NaN
951	NaN	NaN
2153	NaN	NaN
628	NaN	NaN
177	NaN	NaN
2965	NaN	NaN
1768	NaN	NaN
2884	NaN	NaN
734	NaN	NaN
1307	NaN	NaN

	_embedded.show.dvdCountry.name	_embedded.show.dvdCountry.code \
2020	NaN	NaN
951	NaN	NaN
2153	NaN	NaN
628	NaN	NaN
177	NaN	NaN
2965	NaN	NaN
1768	NaN	NaN
2884	NaN	NaN
734	NaN	NaN

1307 NaN NaN

```
_embedded.show.dvdCountry.timezone
2020 NaN
951 NaN
2153 NaN
628 NaN
177 NaN
2965 NaN
1768 NaN
2884 NaN
734 NaN
1307 NaN
```

[10 rows x 62 columns]

Se observa en una muestra de los datos la información con la que se cuenta

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3037 entries, 0 to 3036

Data columns (total 62 columns):

#	Column	Non-Null Count	Dtype
0	id	3037 non-null	int64
1	url	3037 non-null	
object			
2	name	3037 non-null	
object			
3	season	3037 non-null	int64
4	number	2956 non-null	
float64			
5	type	3037 non-null	
object			
6	airdate	3037 non-null	
object			
7	airtime	3037 non-null	
object			
8	airstamp	3037 non-null	
object			
9	runtime	2810 non-null	
float64			
10	image	0 non-null	
float64			
11	summary	890 non-null	

object		
12 rating.average	301 non-null	
float64		
13 _links.self.href	3037 non-null	
object		
14 _embedded.show.id	3037 non-null	int64
15 _embedded.show.url	3037 non-null	
object		
16 _embedded.show.name	3037 non-null	
object		
17 _embedded.show.type	3037 non-null	
object		
18 _embedded.show.language	3003 non-null	
object		
19 _embedded.show.genres	3037 non-null	
object		
20 _embedded.show.status	3037 non-null	
object		
21 _embedded.show.runtime	2039 non-null	
float64		
22 _embedded.show.averageRuntime	2877 non-null	
float64		
23 _embedded.show.premiered	3037 non-null	
object		
24 _embedded.show.ended	1331 non-null	
object		
25 _embedded.show.officialSite	2652 non-null	
object		
26 _embedded.show.schedule.time	3037 non-null	
object		
27 _embedded.show.schedule.days	3037 non-null	
object		
28 _embedded.show.rating.average	423 non-null	
float64		
29 _embedded.show.weight	3037 non-null	int64
30 _embedded.show.network	0 non-null	
float64		
31 _embedded.show.webChannel.id	2957 non-null	
float64		
32 _embedded.show.webChannel.name	2957 non-null	
object		
33 _embedded.show.webChannel.country.name	1462 non-null	
object		
34 _embedded.show.webChannel.country.code	1462 non-null	
object		
35 _embedded.show.webChannel.country.timezone	1462 non-null	
object		
36 _embedded.show.webChannel.officialSite	1769 non-null	

object		
37 _embedded.show.dvdCountry	0 non-null	
float64		
38 _embedded.show.externals.tvrage	56 non-null	
float64		
39 _embedded.show.externals.thetvdb	2116 non-null	
float64		
40 _embedded.show.externals.imdb	1597 non-null	
object		
41 _embedded.show.image.medium	2879 non-null	
object		
42 _embedded.show.image.original	2879 non-null	
object		
43 _embedded.show.summary	2725 non-null	
object		
44 _embedded.show.updated	3037 non-null	int64
45 _embedded.show._links.self.href	3037 non-null	
object		
46 _embedded.show._links.previousepisode.href	3037 non-null	
object		
47 image.medium	1011 non-null	
object		
48 image.original	1011 non-null	
object		
49 _embedded.show.network.id	206 non-null	
float64		
50 _embedded.show.network.name	206 non-null	
object		
51 _embedded.show.network.country.name	206 non-null	
object		
52 _embedded.show.network.country.code	206 non-null	
object		
53 _embedded.show.network.country.timezone	206 non-null	
object		
54 _embedded.show.network.officialSite	7 non-null	
object		
55 _embedded.show._links.nextepisode.href	163 non-null	
object		
56 _embedded.show.webChannel	0 non-null	
float64		
57 _embedded.show.image	0 non-null	
float64		
58 _embedded.show.webChannel.country	0 non-null	
float64		
59 _embedded.show.dvdCountry.name	27 non-null	
object		
60 _embedded.show.dvdCountry.code	27 non-null	
object		
61 _embedded.show.dvdCountry.timezone	27 non-null	

```
object
dtypes: float64(16), int64(5), object(41)
memory usage: 1.4+ MB
```

Verificamos las variables, la cantidad de valores no nulos y su tipo de dato, donde se detalla que hay varios campos con valores faltantes, por esto revisamos el valor porcentual de los mismos.

```
dfr=pd.DataFrame(round((df.isna().sum()/df.shape[0])*100,2),
columns=['Missings'])
df2 = dfr[dfr['Missings'] >= 1]
df2
```

	Missings
number	2.67
runtime	7.47
image	100.00
summary	70.69
rating.average	90.09
_embedded.show.language	1.12
_embedded.show.runtime	32.86
_embedded.show.averageRuntime	5.27
_embedded.show.ended	56.17
_embedded.show.officialSite	12.68
_embedded.show.rating.average	86.07
_embedded.show.network	100.00
_embedded.show.webChannel.id	2.63
_embedded.show.webChannel.name	2.63
_embedded.show.webChannel.country.name	51.86
_embedded.show.webChannel.country.code	51.86
_embedded.show.webChannel.country.timezone	51.86
_embedded.show.webChannel.officialSite	41.75
_embedded.show.dvdCountry	100.00
_embedded.show.externals.tvrage	98.16
_embedded.show.externals.thetvdb	30.33
_embedded.show.externals.imdb	47.42
_embedded.show.image.medium	5.20
_embedded.show.image.original	5.20
_embedded.show.summary	10.27
image.medium	66.71
image.original	66.71
_embedded.show.network.id	93.22
_embedded.show.network.name	93.22
_embedded.show.network.country.name	93.22
_embedded.show.network.country.code	93.22
_embedded.show.network.country.timezone	93.22
_embedded.show.network.officialSite	99.77
_embedded.show._links.nextepisode.href	94.63
_embedded.show.webChannel	100.00
_embedded.show.image	100.00
_embedded.show.webChannel.country	100.00

```

_embedded.show.dvdCountry.name          99.11
_embedded.show.dvdCountry.code           99.11
_embedded.show.dvdCountry.timezone       99.11

```

```
df.describe(include='all')
```

```

              id
url \
count    9.800000e+01
98
unique          NaN
98
top          NaN  https://www.tvmaze.com/episodes/1979824/sim-
fo...
freq          NaN
1
mean    2.018019e+06
NaN
std     8.446022e+04
NaN
min     1.939481e+06
NaN
25%     1.976152e+06
NaN
50%     1.978707e+06
NaN
75%     2.033287e+06
NaN
max     2.336906e+06
NaN

```

```

              name      season      number      type      airdate
airtime \
count          98    98.000000    98.000000          98          98
98
unique          92          NaN          NaN          1          1
15
top    Episode 21          NaN          NaN    regular    2020-12-01
freq          2          NaN          NaN          98          98
64
mean          NaN    208.081633    25.612245          NaN          NaN
NaN
std          NaN    613.949420    48.059378          NaN          NaN
NaN
min          NaN     1.000000     1.000000          NaN          NaN
NaN
25%          NaN     1.000000     4.250000          NaN          NaN
NaN
50%          NaN     1.000000    11.000000          NaN          NaN

```



NaN					
75%	NaN	2.000000	21.750000	NaN	NaN
NaN					
max	NaN	2020.000000	328.000000	NaN	NaN
NaN					

		airstamp	runtime	...	\
count		98	93.000000	...	
unique		20	NaN	...	
top	2020-12-01T12:00:00+00:00		NaN	...	
freq		64	NaN	...	
mean		NaN	32.11828	...	
std		NaN	22.96777	...	
min		NaN	5.000000	...	
25%		NaN	13.000000	...	
50%		NaN	26.000000	...	
75%		NaN	45.000000	...	
max		NaN	120.000000	...	

	_embedded.show.network.country.code	\
count	13	
unique	7	
top	RU	
freq	3	
mean	NaN	
std	NaN	
min	NaN	
25%	NaN	
50%	NaN	
75%	NaN	
max	NaN	

	_embedded.show.network.country.timezone	\
count	13	
unique	7	
top	Asia/Kamchatka	
freq	3	
mean	NaN	
std	NaN	
min	NaN	
25%	NaN	
50%	NaN	
75%	NaN	
max	NaN	

	_embedded.show.network.officialSite	\
count	0.0	
unique	NaN	
top	NaN	
freq	NaN	

mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

_embedded.show.links.nextepisode.href	
_embedded.show.webChannel \	
count	5
0.0	
unique	5
NaN	
top	<a href="https://api.tvmaze.com/episodes/2309426">https://api.tvmaze.com/episodes/2309426</a>
NaN	
freq	1
NaN	
mean	NaN
NaN	
std	NaN
NaN	
min	NaN
NaN	
25%	NaN
NaN	
50%	NaN
NaN	
75%	NaN
NaN	
max	NaN
NaN	

_embedded.show.image		_embedded.show.webChannel.country \	
count	0.0		0.0
unique	NaN		NaN
top	NaN		NaN
freq	NaN		NaN
mean	NaN		NaN
std	NaN		NaN
min	NaN		NaN
25%	NaN		NaN
50%	NaN		NaN
75%	NaN		NaN
max	NaN		NaN

_embedded.show.dvdCountry.name	
_embedded.show.dvdCountry.code \	
count	1

unique	1	1
top	Japan	JP
freq	1	1
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

	<code>_embedded.show.dvdCountry.timezone</code>
count	1
unique	1
top	Asia/Tokyo
freq	1
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

[11 rows x 62 columns]

A nivel general podemos observar el comportamiento estadístico de cada una de las variables