



YData Profiling Report

Overview

Brought to you by [YData](#)

- Overview
- Alerts 13
- Reproduction

Dataset statistics

Number of variables	20
Number of observations	4803
Missing cells	3941
Missing cells (%)	4.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	750.6 KiB
Average record size in memory	160.0 B

Variable types

Numeric	7
Text	10
Categorical	2
DateTime	1

Variables

Select Columns

**budget**

Real number ( $\mathbb{R}$ )

High correlation   Zeros

Distinct436

Distinct (%)9.1%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean29045040

Minimum0

Maximum3.8 × 1

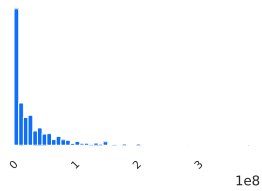
Zeros1037

Zeros (%)21.6%

Negative0

Negative (%)0.0%

Memory size37.7 Ki



More details

genres

Text

Distinct	1175
Distinct (%)	24.5%
Missing	0
Missing (%)	0.0%
Memory size	37.7 KiB



More details

homepage

Text

Missing

Distinct	1691
Distinct (%)	98.8%
Missing	3091
Missing (%)	64.4%
Memory size	37.7 KiB

http://www.thehobbit.com  
http://www.howtotrainyourdragon.com  
http://www.thehungergames.movie  
http://www.lordoftherings.net  
http://www.romainelapied.com  
http://www.transformersmovie.com  
http://www.kungfupanda.com  
http://www.workandtheglory.com http://www.indianajones.com  
http://www.missionimpossible.com  
http://www.atlsshruddedmovie.com  
http://www.teenagemutantninja turtlesmovie.com  
http://www.themazingspiderman.com  
http://www.sandwichmovie.com http://www.11amovies.com  
http://www.twilightthemovie.com http://www.sandwichmovie.com  
http://disney.go.com/disneypictures/pirates  
http://www.katzenbergentertainment.com http://www.gijomovie.com http://www.mankyouself.com

More details

id

Real number ( $\mathbb{R}$ )

Unique

**Distinct** 4803

**Minimum** 5

**Distinct (%)** 100.0%

**Maximum** 459488

**Missing** 0

**Zeros**      0

**Missing**  
**(%)** 0.0%

**Zeros (%)** 0.0%

Infinite 0

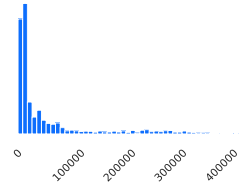
**Negative**      0

**Infinite**  
**(%)** 0.0%

**Negative**  
**(%)** 0.0%

<b>Mean</b>	57165.484
-------------	-----------

Memory size	37.7 KiB
-------------	----------

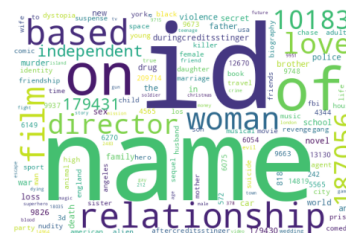


More details

## keywords

Text

<b>Distinct</b>	4222
<b>Distinct (%)</b>	87.9%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB



More details

original\_language

## Categorical

## Imbalance

<b>Distinct</b>	37
<b>Distinct (%)</b>	0.8%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB

en	4505
fr	70
es	32
zh	27
de	27
Other v...	142

More details

original\_title

Text

<b>Distinct</b>	4801
<b>Distinct (%)</b>	> 99.9%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB

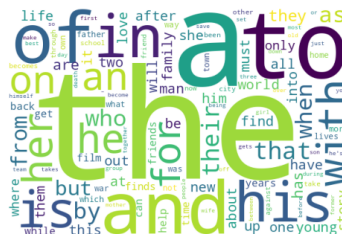


More details

## overview

Text

<b>Distinct</b>	4800
<b>Distinct (%)</b>	100.0%
<b>Missing</b>	3
<b>Missing (%)</b>	0.1%
<b>Memory size</b>	37.7 KiB



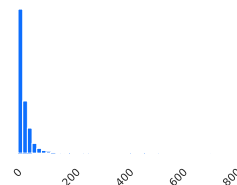
More details

popularity

Real number ( $\mathbb{R}$ )

## High correlation

<b>Distinct</b>	4802	<b>Minimum</b>	0
<b>Distinct (%)</b>	> 99.9%	<b>Maximum</b>	875.58
<b>Missing</b>	0	<b>Zeros</b>	1
<b>Missing (%)</b>	0.0%	<b>Zeros (%)</b>	< 0.1%
<b>Infinite</b>	0	<b>Negative</b>	0
<b>Infinite (%)</b>	0.0%	<b>Negative (%)</b>	0.0%
<b>Mean</b>	21.492301	<b>Memory size</b>	37.7 Ki

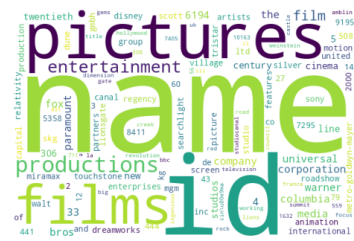


More details

## production\_companies

Text

<b>Distinct</b>	3697
<b>Distinct (%)</b>	77.0%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB



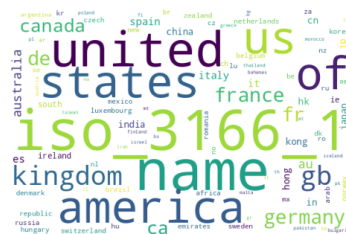
More details

production\_countries

Text




<b>Distinct</b>	469
<b>Distinct (%)</b>	9.8%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB

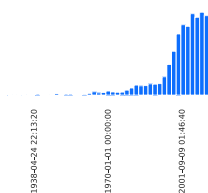


More details

release\_date

Date

Distinct	3280	Minimum	1916-0
Distinct (%)	68.3%	Maximum	2017-0
Missing	1	Invalid dates	0
Missing (%)	< 0.1%	Invalid dates (%)	0.0%
Memory size	37.7 KiB		



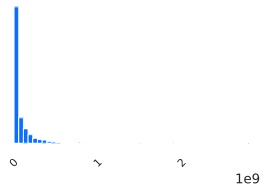
More details

revenue

Real number (ℝ)

High correlation   Zeros

Distinct	3297	Minimum	0
Distinct (%)	68.6%	Maximum	2.7879
Missing	0	Zeros	1427
Missing (%)	0.0%	Zeros (%)	29.7%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	82260639	Memory size	37.7 Ki

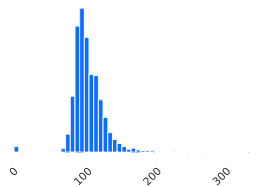


More details

runtime

Real number (ℝ)

Distinct	156	Minimum	0
Distinct (%)	3.2%	Maximum	338
Missing	2	Zeros	35
Missing (%)	< 0.1%	Zeros (%)	0.7%
Infinite	0	Negative	0
		Negative (%)	0.0%



Infinite  
(%) 0.0%

Memory size	37.7 KiB
-------------	----------

**Mean** 106.87586

More details

spoken\_languages

Text

Distinct	544
----------	-----

Distinct (%)	11.3%
--------------	-------

Missing 0

Missing (%)	0.0%
-------------	------

Memory size	37.7 KiB
-------------	----------



More details

status

## Categorical

## Imbalance

Distinct	3
----------	---

<b>Distinct (%)</b>	0.1%
---------------------	------

Released 4795

Rumored 5

Post Pr... 3

<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB

More details

## tagline

Text

**Missing**

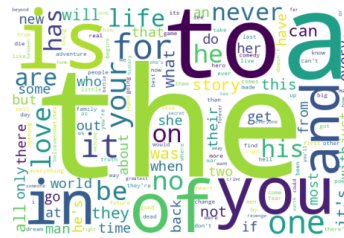
Distinct	3944
----------	------

<b>Distinct (%)</b>	99.6%
---------------------	-------

Missing 844

**Missing (%)** 17.6%

**Memory size** 37.7 KiB

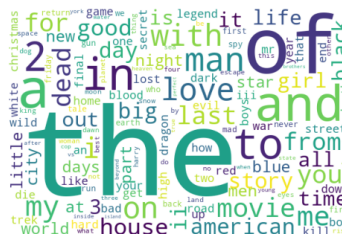


More details

title

Text

<b>Distinct</b>	4800
<b>Distinct (%)</b>	99.9%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	37.7 KiB



More details

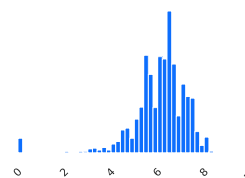
vote\_average

Real number ( $\mathbb{R}$ )

## Zeros

Distinct	71
Distinct (%)	1.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	6.0921716

Minimum	0
Maximum	10
Zeros	63
Zeros (%)	1.3%
Negative	0
Negative (%)	0.0%
Memory size	37.7 K



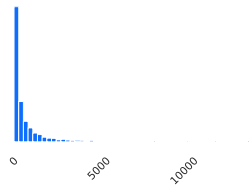
More details

vote\_count

Real number ( $\mathbb{R}$ )

High correlation   Zeros

Distinct	1609	Minimum	0
Distinct (%)	33.5%	Maximum	13752
Missing	0	Zeros	62
Missing (%)	0.0%	Zeros (%)	1.3%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	690.21799	Memory size	37.7 Ki



More details

Interactions

budget

id

popularity

revenue

runtime

vote\_average

vote\_count

vote\_count

budget

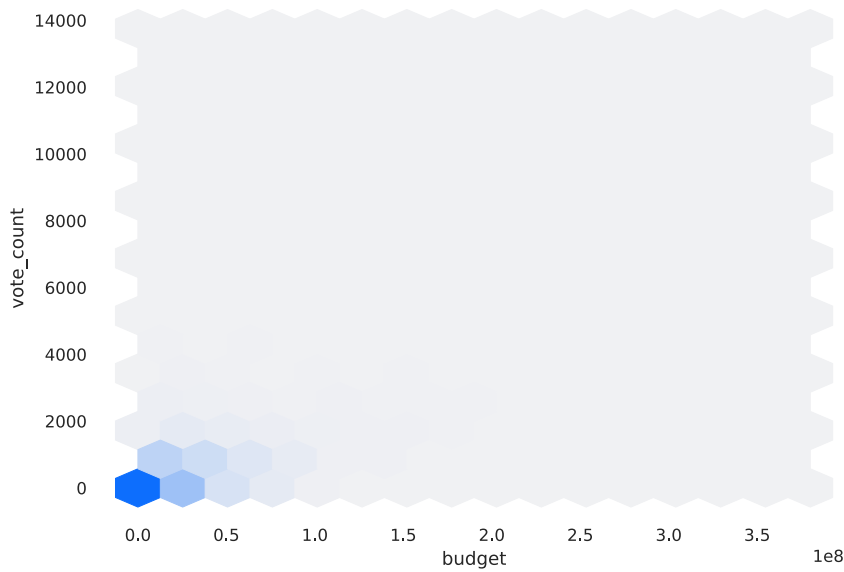
id

popularity

revenue

runtime

vote\_average

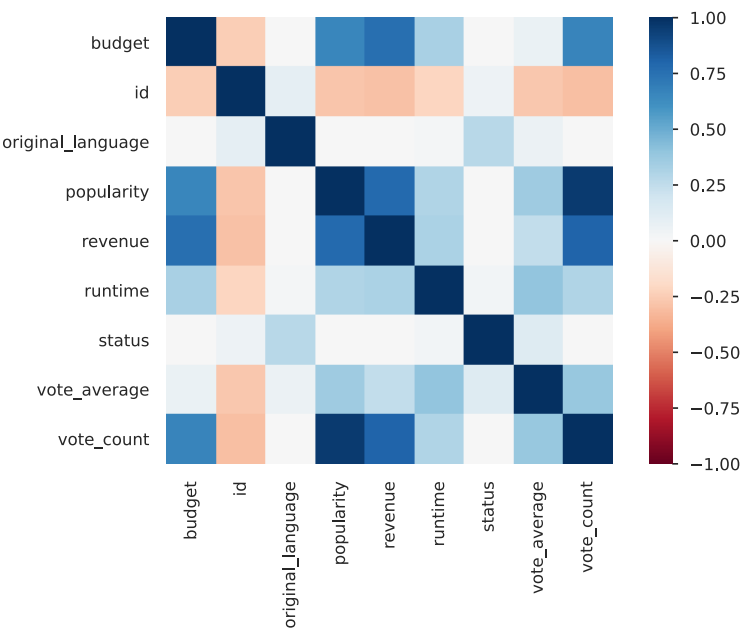


# Correlations

Auto

Heatmap

[Table](#)



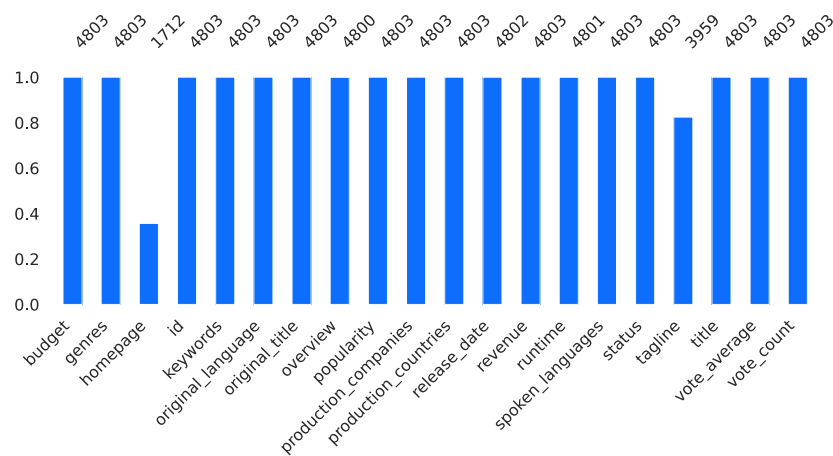
# Missing values

Count

Matrix

Heatmap





A simple visualization of nullity by column.

# Sample

First rows

Last rows

budget		
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Ad
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name":
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Ad
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Cri
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Ad
5	258000000	[{"id": 14, "name": "Fantasy"}, {"id": 28, "name": "A

budget		
6	2600000000	[{"id": 16, "name": "Animation"}, {"id": 10751, "name": "Animation"}]
7	2800000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]
8	2500000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Adventure"}]
9	2500000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]

Report generated by [YData](#).

