

mdata

Autogenerated data summary from dataMaid

2019-09-18 13:57:09

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	231
Number of variables	14

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Summary table

	Variable class	# unique values	Missing observations	Any problems?
Movie	character	231	0.00 %	×
Year	numeric	2	0.00 %	
Ratings	numeric	45	0.00 %	×
Genre	numeric	11	0.00 %	
Gross	numeric	215	0.00 %	×
Budget	numeric	105	0.43 %	
Screens	numeric	201	4.33 %	×
Sequel	numeric	7	0.00 %	×
Sentiment	numeric	36	0.00 %	×
Views	numeric	231	0.00 %	×
Likes	numeric	227	0.00 %	×
Dislikes	numeric	203	0.00 %	×
Comments	numeric	213	0.00 %	×
Aggregate Followers	numeric	191	15.15 %	

Part 3

Variable list

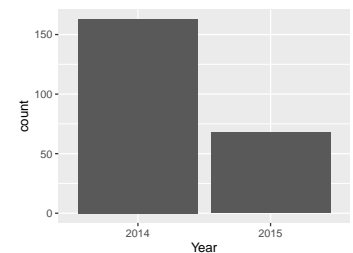
Movie

- The variable is a key (distinct values for each observation).

Year

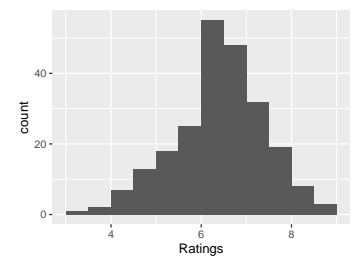
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	“2014”
Reference category	2014



Ratings

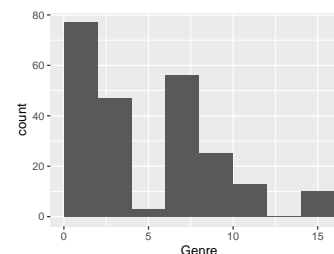
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	45
Median	6.5
1st and 3rd quartiles	5.8; 7.1
Min. and max.	3.1; 8.7



- Note that the following possible outlier values were detected: "3.1".

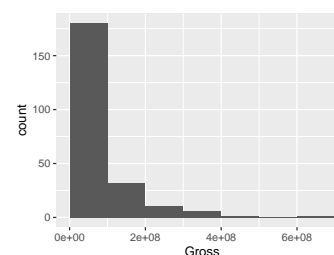
Genre

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	3
1st and 3rd quartiles	1; 8
Min. and max.	1; 15



Gross

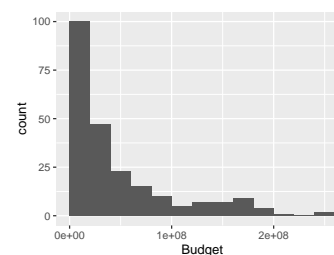
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	215
Median	37400000
1st and 3rd quartiles	10300000; 89350000
Min. and max.	2470; 6.43e+08



- Note that the following possible outlier values were detected: "6.43e+08".

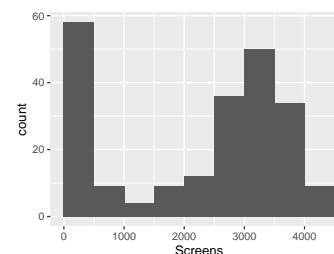
Budget

Feature	Result
Variable type	numeric
Number of missing obs.	1 (0.43 %)
Number of unique values	104
Median	2.8e+07
1st and 3rd quartiles	9e+06; 6.5e+07
Min. and max.	70000; 2.5e+08



Screens

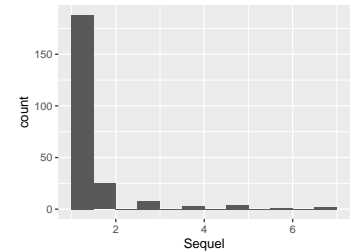
Feature	Result
Variable type	numeric
Number of missing obs.	10 (4.33 %)
Number of unique values	200
Median	2777
1st and 3rd quartiles	449; 3372
Min. and max.	2; 4324



- Note that the following possible outlier values were detected: "4080", "4151", "4233", "4253", "4274", "4276", "4301", "4324".

Sequel

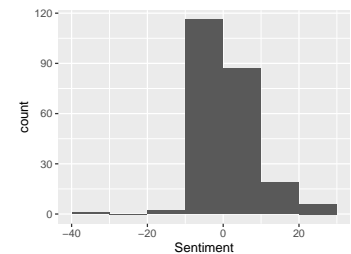
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	7
Median	1
1st and 3rd quartiles	1; 1
Min. and max.	1; 7



- Note that the following possible outlier values were detected: "2", "3", "4", "5", "6", "7".

Sentiment

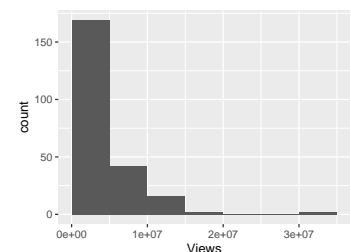
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	36
Median	0
1st and 3rd quartiles	0; 5.5
Min. and max.	-38; 29



- Note that the following possible outlier values were detected: "-38", "-17", "-11", "-9", "-8", "-6", "-5", "-4", "-3", "-2" (1 additional values omitted).

Views

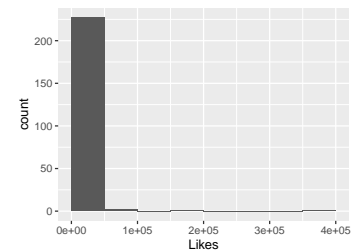
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	231
Median	2409338
1st and 3rd quartiles	623302; 5217379.5
Min. and max.	698; 32626778



- Note that the following possible outlier values were detected: "31859569", "32626778".

Likes

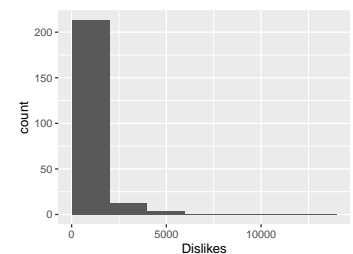
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	227
Median	6096
1st and 3rd quartiles	1776.5; 15247.5
Min. and max.	1; 370552



- Note that the following possible outlier values were detected: "187162", "370552".
-

Dislikes

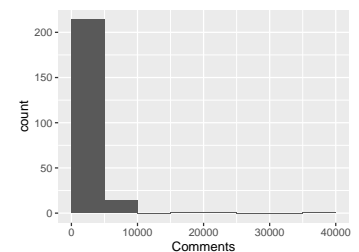
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	203
Median	341
1st and 3rd quartiles	105.5; 697.5
Min. and max.	0; 13960



- Note that the following possible outlier values were detected: "3439", "3524", "3565", "3812", "4245", "4382", "4752", "5746", "13960".
-

Comments

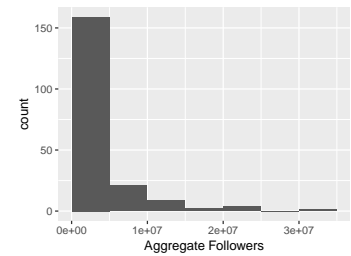
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	213
Median	837
1st and 3rd quartiles	248.5; 2137
Min. and max.	0; 38363



- Note that the following possible outlier values were detected: "18077", "24919", "38363".
-

Aggregate Followers

Feature	Result
Variable type	numeric
Number of missing obs.	35 (15.15 %)
Number of unique values	190
Median	1052600
1st and 3rd quartiles	183025; 3694500
Min. and max.	1066; 31030000



Report generation information:

- Created by Blake Holmes (username: **bjholmes**).
- Report creation time: Wed Sep 18 2019 13:57:09
- Report was run from directory: /Users/bjholmes/Library/Mobile Documents/com~apple~CloudDocs/DataScience/Sta
- dataMaid v1.3.2 [Pkg: 2019-07-27 from CRAN (R 3.5.2)]
- R version 3.5.1 (2018-07-02).
- Platform: x86_64-apple-darwin15.6.0 (64-bit)(macOS 10.14.6).
- Function call: `makeDataReport(data = mdata, replace = TRUE)`