

20220816_Psychometrics

Sidhant Chaudhary

Sys.Date()

Reading the datasets

```
# uncomment to download data
tuesdata <- tidyuesdayR::tt_load('2022-08-16')
```

```
Downloading file 1 of 3: `characters.csv`
Downloading file 2 of 3: `myers_briggs.csv`
Downloading file 3 of 3: `psych_stats.csv`
```

```
characters <- tuesdata$characters %>% janitor::clean_names()
myers_briggs <- tuesdata$myers_briggs %>% janitor::clean_names()
psych_stats <- tuesdata$psych_stats %>% janitor::clean_names()
```

```
skim(characters)
```

Table 1: Data summary

Name	characters
Number of rows	889
Number of columns	7
Column type frequency:	
character	6
numeric	1
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
id	0	1	2	7	0	889	0
name	0	1	2	44	0	885	0
uni_id	0	1	1	6	0	100	0
uni_name	0	1	2	33	0	100	0
link	0	1	56	61	0	889	0
image_link	0	1	74	79	0	889	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
notability	0	1	67.63	17.39	18.4	55.6	71.7	81	96.9	

```
skim(myers_briggs)
```

Table 4: Data summary

Name	myers_briggs
Number of rows	14224
Number of columns	7
Column type frequency:	
character	5
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
char_id	0	1	2	7	0	889	0
char_name	0	1	2	44	0	885	0
uni_id	0	1	1	6	0	100	0
uni_name	0	1	2	33	0	100	0
myers_briggs	0	1	4	4	0	16	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
avg_match_perc	0	1	54.62	8.03	28.6	49.30	55	60.7	75.6	
number_users	0	1	5605.75	6164.75	547.0	1406.75	2142	9324.5	23344.0	

```
skim(psych_stats)
```

Table 7: Data summary

Name	psych_stats
Number of rows	356489
Number of columns	10
Column type frequency:	
character	6
numeric	4
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
char_id	0	1	2	7	0	889	0
char_name	0	1	2	44	0	885	0
uni_id	0	1	1	6	0	100	0
uni_name	0	1	2	33	0	100	0
question	0	1	3	44	0	400	0
personality	0	1	1	22	0	780	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
avg_rating	0	1	68.44	11.43	50.1	58.7	67.4	77.1	100.0	
rank	0	1	537.32	373.15	1.0	225.0	483.0	793.0	1901.0	
rating_sd	0	1	25.36	6.00	0.0	21.8	26.2	29.6	46.7	
number_ratings	0	1	131.04	185.82	3.0	28.0	58.0	142.0	2940.0	