

Data Dictionary

Here is a data dictionary for the Rate My Professor data set which I have tweaked and cleaned to my needs. In R, I called this data frame `rmp_clean`. Each observation in this data set represents one review made by a student of a certain professor for a certain course that they took with that professor. There are 20,000 reviews by students in this dataset about professors across the United States. The time frame is from 2001 to 2020.

Link to my data which also has some variable descriptions:

<https://data.mendeley.com/datasets/fvtfjyvw7d/2>

Data Dictionary:

Variable Name	Description	Type	How I Would Classify
professor_name	Professor's Name	character	categorical
school_name	Professor's school	character	categorical
department_name	Professor's department of study	factor	categorical
state_name	Professor's state	factor	categorical
year_since_first_review	How many years since the Professor's first review	numeric	quantitative
star_rating	Professor's overall quality rating	numeric	quantitative
diff_index	Professor's overall difficulty rating	numeric	quantitative
post_date	When the student posted their review	numeric	quantitative
student_star	Individual student rating of the professor's overall quality in a single review	numeric	quantitative
student_difficult	Individual student rating of professor difficulty in a single review	numeric	quantitative
attence	Whether or not attendance is mandatory	character	categorical
for_credits	Whether or not the student took the class for credits	factor	categorical
grades	The grade the student received in the class	factor	categorical
gender	Professor gender	character	categorical

comments	A student's full comment for a professor	character	word analysis
department_fewer	Department by discipline of study/college	factor	categorical
us_region	States organized into 4 US regions	factor	categorical
attendance	Whether or not attendance is mandatory	factor	categorical

Screenshot of Data Loaded into R Environment:

```

rmp_clean 20000 obs. of 18 variables
 $ professor_name      : chr  "Leslie Looney" "Leslie Looney" "Leslie Looney" ...
 $ school_name         : chr  "University of Illinois at Urbana-Champaign" "Unive...
 $ department_name     : Factor w/ 101 levels "Accounting","Accounting & Finance"...
 $ state_name          : Factor w/ 46 levels "AL","AR","AZ",...: 14 14 14 14 14 14...
 $ year_since_first_review: num  11 11 11 11 11 11 11 11 11 11 ...
 $ star_rating         : num  4.7 4.7 4.7 4.7 4.7 4.7 4.7 4.7 4.7 4.7 ...
 $ diff_index          : num  2 2 2 2 2 2 2 2 2 ...
 $ post_date           : Date, format: "2017-06-27" "2017-04-16" "2016-12-07" "20...
 $ student_star        : num  5 5 4 5 5 5 5 5 5 ...
 $ student_difficult   : num  3 2 3 3 1 2 2 2 1 2 ...
 $ attence             : chr  "Mandatory" "Not Mandatory" "" "Mandatory" ...
 $ for_credits         : Factor w/ 3 levels "NA","No","Yes": 3 1 3 3 1 1 1 1 1 ...
 $ grades              : chr  "B" "A+" "" "A" ...
 $ gender              : Factor w/ 3 levels "unknown","female",...: 1 1 1 1 1 1 1 ...
 $ comments            : chr  "This class is hard, but its a two-in-one gen-ed kn...
 $ department_fewer    : Factor w/ 4 levels "Business","Humanities",...: 3 3 3 3 3...
 $ us_region           : Factor w/ 5 levels "South","West",...: 5 5 5 5 5 5 5 5 ...
 $ attendance          : Factor w/ 3 levels "NA","Mandatory",...: 2 3 1 2 2 1 2 1 ...

```