## Academic Excellence: A machine learning-based deep dive of RateMyProfessors reviews


## Authors:

- Sam Cowan
- Donghwan Kwon
- Cyrus Sadr
- Andry Rubio


## Code available at: https://github.com/samc5/RMP_110

## Introduction:

RateMyProfessors, also known as RMP, is a website created in 1999, consisting of a database of ratings given to professors over the years by their students. Most professors throughout the US have at least one review, so there is a vast amount of data available. Traditionally, students use the site as a method to decide whether to take a class, often opting out of a professor with poor reviews in favor of a popular one.

Several people we've talked to have noticed a strange phenomenon: an oddly high number of reviews talk about a professor's attractiveness, which is not something that any of us would have considered an important factor. We began to wonder if professors who were considered attractive received higher ratings, which led us to ask: what features contribute significantly to a RateMyProfessors review? Our main goal in this study was to determine the most important factors in a RMP review, and our secondary goal was to determine if attractiveness has any impact on RMP ratings.

## Previous Work:

- Professor, Student, and Course Attributes that Contribute to Successful Teaching Evaluations
    - This research paper connects to our central question on what aspects of a professor/class play the biggest role in a student's RateMyProfessors review by showing the results of a study on student evaluation on the surveys of professors. It includes various contributing factors like professor characteristics as well as course factors with student details such as major, year, or other details included in the research that tie into the student's final evaluation. Through this work, the authors were able to show that students and professors in higher learning, students with high GPAs, and elective classes were able to have higher student evaluation of teaching (SET) scores.
- Beauty Premiums Among Academics
    - This research paper ties to our central question of the contributing factor of attractiveness to the overall rating of college professors. It utilized multiple attractiveness measures, including facial symmetry software, subjective evaluations, and a novel, proxy methodology. It was shown that attractive faculty members were given much higher SETE (student evaluation) scores, but also, the teaching environment would affect the professor. Compared to an in-person learning environment, an online environment turned out to negatively affect the attractive instructor.

- Characteristics of an Effective University Professor From Students' Perspective: Are the Qualities Changing?

    - This study was conducted at a state university in Iran. It was based on the characteristics of an effective university professor who is evaluated by students in BA, MA, and Ph.D. Through this study, many subcategories affected a student's evaluation of certain professors. Based on Table 1, the subcategory of the "suitable exam" was considered more crucial than the other two, which were "scoring procedure" and "course work required". Other qualities that seemed to stand out in the study were generosity with course grades and fairness in a professor. Fairness was considered because students were afraid that the teacher's view toward different students would affect their scores in general. Personality in general was also a huge aspect of a professor; for example, "respect toward students" was considered the second most important and positive quality, and the category "sensitivity and strictness about students' class attendance" was considered one of the most negative characteristics in their professors. In the end, whether the different aspects of a professor are a negative or positive attribute, they affect a professor's evaluation from a student's standpoint.
- STUDENT EVALUATIONS OF TEACHING ARE MOSTLY AWFULLY WRONG
    - This paper corroborated many others in finding that student evaluations of professors are not very related to teaching outcomes. The authors mention several biases that students have towards professors, including race, gender, and most pertinent to our paper, attractiveness. They also mention that almost all academic literature about student evaluations of teachers suggests they are unreliable and should not be used. It is also mentioned that 1-5 scales, like that on RMP, are often used for student evaluations of professors. Because this scale can be interpreted drastically differently by different students, it is not particularly useful. The paper claims that student evaluations can be easily biased against a teacher's appearance
- Variables that Can Affect Student Ratings of Their Professor
    - This study analyzes three factors for evaluating professors: students' grades, perceptions of the professors as caring, and time spent working on the course. The results indicate that there was a three-way interaction effect of these factors on students' satisfaction with a professor (a proxy for rating), showing that the grade a student receives influences their overall rating of a professor. The paper explains that student ratings of professors were used as metrics by universities even before RateMyProfessors existed. It also connects professor ratings with psychology, in that students' expectations of how well they did play a large role in how they view a professor. It also connects to our paper in its suggestion of "caring" as a value important to a student's rating of a professor.

## Dataset:

- In 2018, Dr. Jibo He of Tsinghua University scraped 9.5 million RateMyProfessors reviews over two months. While we were unable to procure the full dataset, he released publicly a 20,000-line demo dataset which we analyzed. Each line comprises a single review of a professor. It identifies the professor and their university, as well as the comment and star rating out of 5 of the review, and a list of tags assigned to each professor. The reviews span from 2000 to 2018 and each professor included had up to 20 ratings scraped, but some had less than 20. Unfortunately, we had to throw out about 40% of professors in the dataset because they did not come with tags.

## Methodology:

### Data Preparation

After attempting several variations of NLP using embeddings, as outlined later in this paper, we decided to create an attractiveness column with simple regex instead. For each of the 20,000 reviews, the column has a 1 if its comment contains any of a list of keywords like “hot” or “sexy,” and a 0 otherwise. We added a column to the dataset with the total reviews of the professor and set the attractiveness column to the percentage of the specified professor’s reviews which were caught by our regular expression. We then cut each professor to one line by dropping all reviews for a professor after the first. Since our previous changes made it so all rows of the same professor are identical (for the columns we care about), we don’t lose any data. Ultimately, 13% of professors have at least one review caught by our regular expression. The data was now all boolean values other than attractiveness and star rating.

For statistical tests, we don’t care about star rating and can treat attractiveness as a boolean value based on whether it is equal to or greater than 0. With this data, we ran the t-tests and chi-square tests. For machine learning methods, we wanted more detailed tag data, which existed hidden as a string in the tag_professor column. By using a second regular expression, we could extract the total number of each tag that each professor received, instead of a boolean value for if any single reviewer had assigned that tag. We divided this number by the total number of reviews so that every value was scaled between 0 and 1 (and wasn’t skewed by the number of reviews). We also multiplied the attractiveness column by 4 so it was moderately close to the same scale as the other features.

The features used in statistical tests and machine learning are as follows:

| | | |
|---|---|---|
| Accessible outside class | Amazing lectures | Attractiveness |
| Beware of Pop Quizzes | Caring | Clear grading criteria |
| Extra credit | Get ready to read | Graded by few things |
| Group project | Hilarious | Inspirational |
| Online course | Lecture heavy | Lots of homework |
| Participation matters | Respected | Skip class? You won’t pass |
| So many papers | Test heavy | Tough grader |

Difficulty index was used for correlation heatmap but was taken out for the remainder of our analysis.

We used star rating (star_rating), an average rating out of 5, for most outputs.

## Statistics: Regarding Methodology and Results

As part of our exploratory data analysis, we created a heatmap showcasing the correlation between different features.

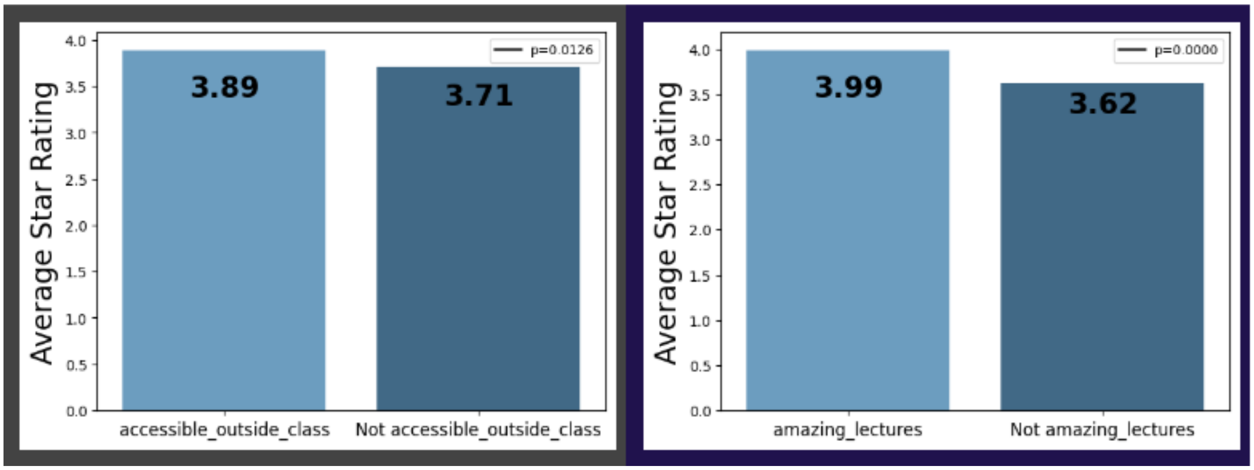
According to the visualization:

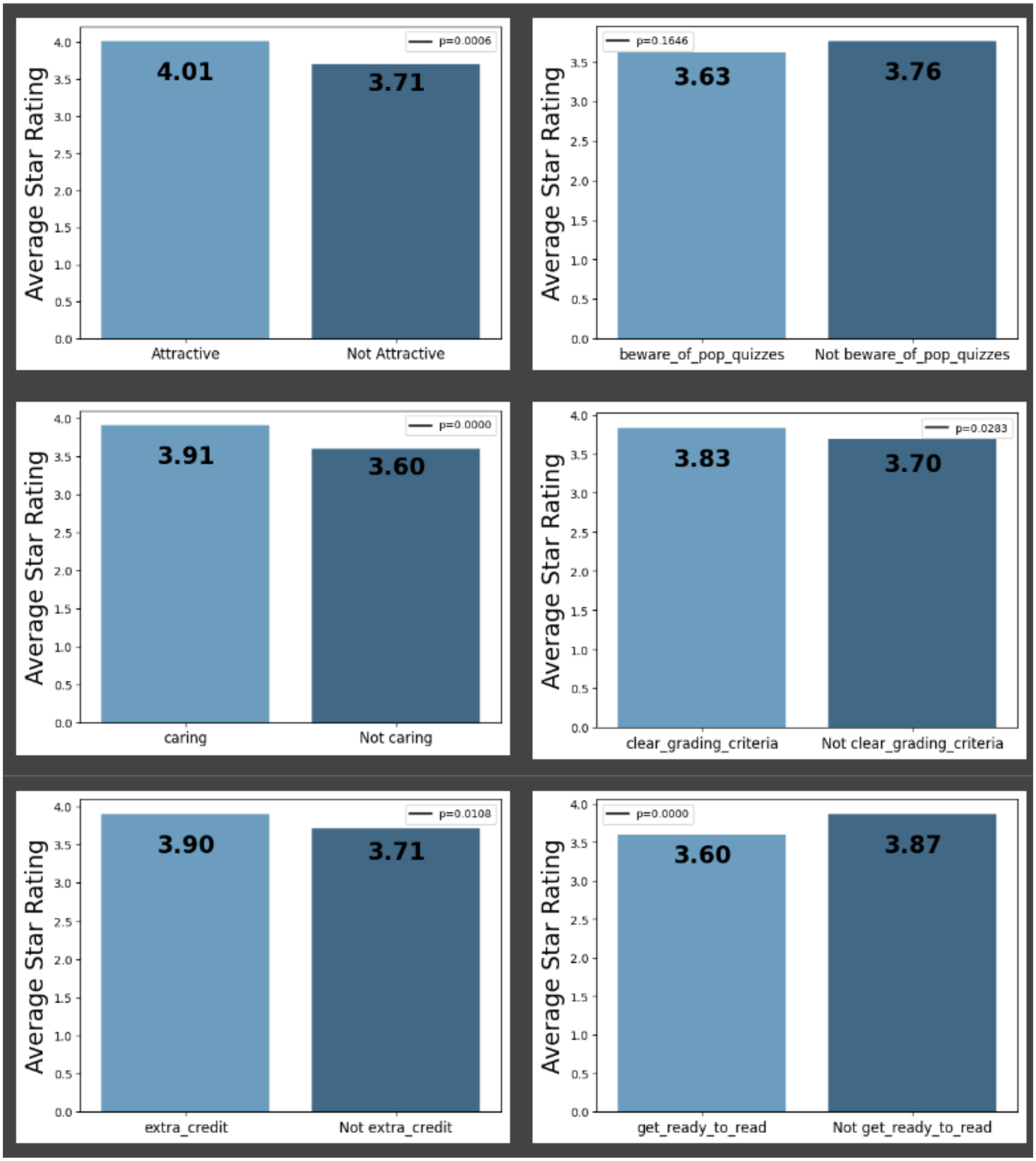
- The largest correlation is a -0.5 correlation between difficulty and star rating.
- Difficulty also correlates closely (0.4) with tough graders, which makes sense
- Amazing lecturers correlate strongly with respected, inspirational, and hilarious professors
- Attractiveness correlates a little with star rating and not with anything else

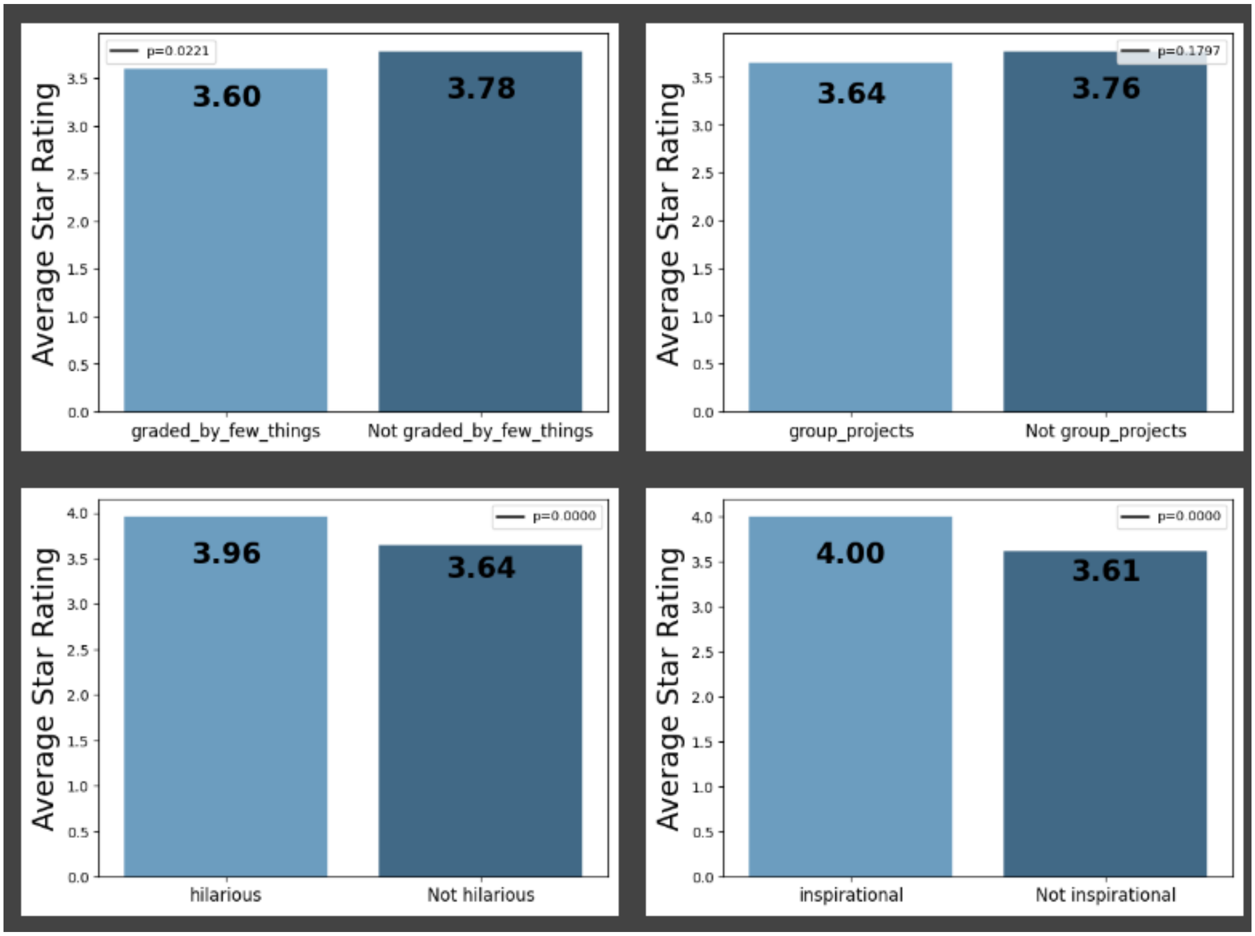
- T-tests were used to check if the difference in average star rating between professors who had and did not have a tag was statistically significant. We looked at all features listed in the data preparation section above

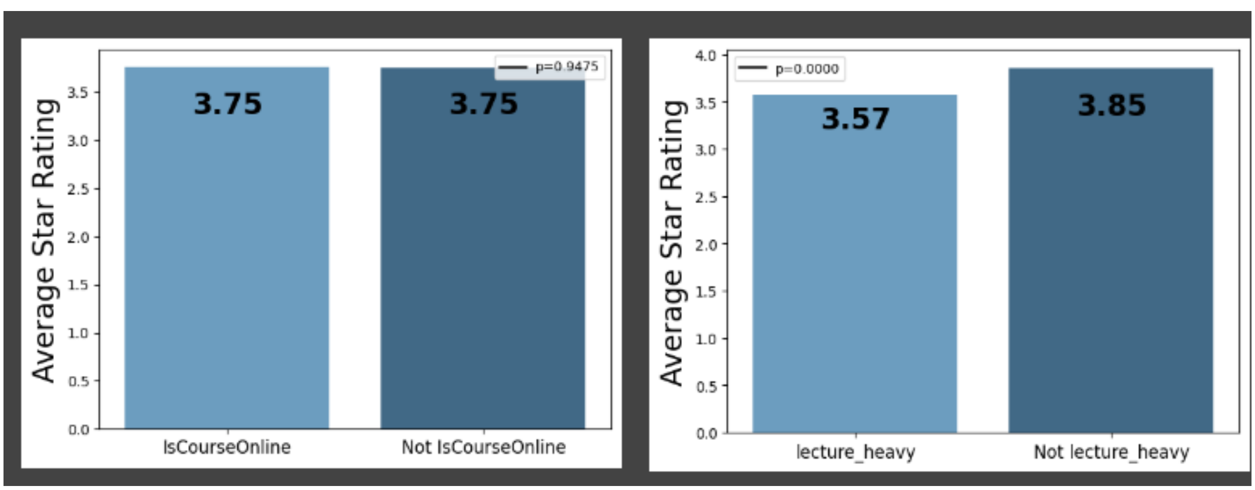
    Significant differences ($p < 0.05$) were found for:
        - Attractive
        - Caring
        - Gives good feedback
        - Respected
        - Clear grading criteria
        - Skip Class? You won't pass
        - Amazing Lectures
        - Inspirational
        - Tough grader
        - Hilarious
        - Get ready to read
        - Lots of homework
        - Accessible outside class
        - Lecture heavy
        - Extra credit available
        - Graded by few things
        - Group Projects
        - Test heavy
        - So many papers
    - Insignificant differences ($p > 0.05$) were found for:
        - Participation Matters
        - Group Projects
        - Beware of pop quizzes
        - Online Course
- Some takeaways from a student's perspective:
    - When choosing a professor/class based on Rate My Professor reviews, students should consider other factors before taking into account participation, group projects, or pop quizzes
    - There was a significant difference between the average rating of professors thought to be attractive (4.01 stars) and professors not thought to be attractive (3.71 stars), as $p = 0.0006$. This

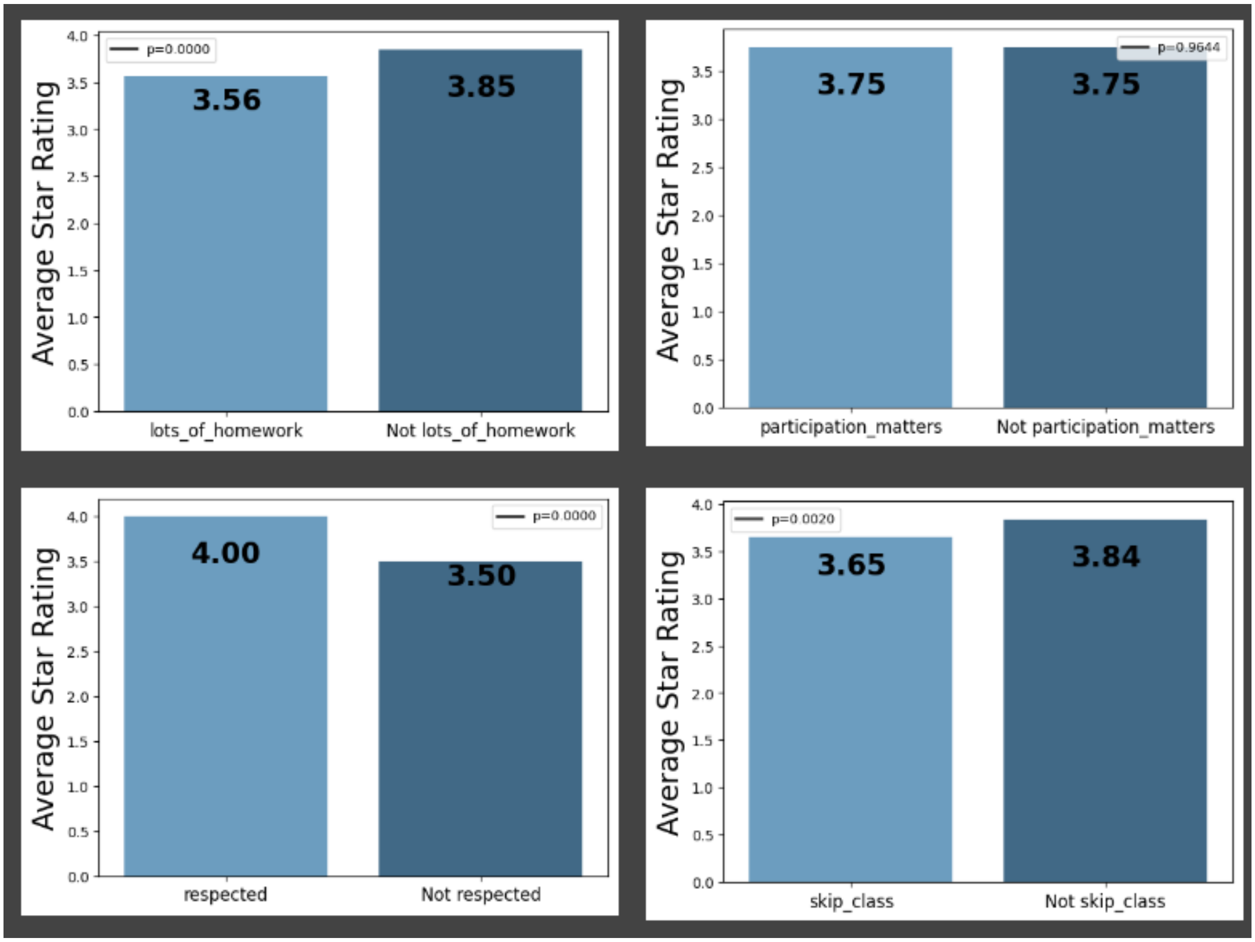
result suggests that teachers seen as attractive receive notably higher RMP ratings, yet teachers not thought to be attractive only have marginally lower ratings.

- We graphed the average of the professor populations for all 21 tags used, along with the p-value of its t-test
- 

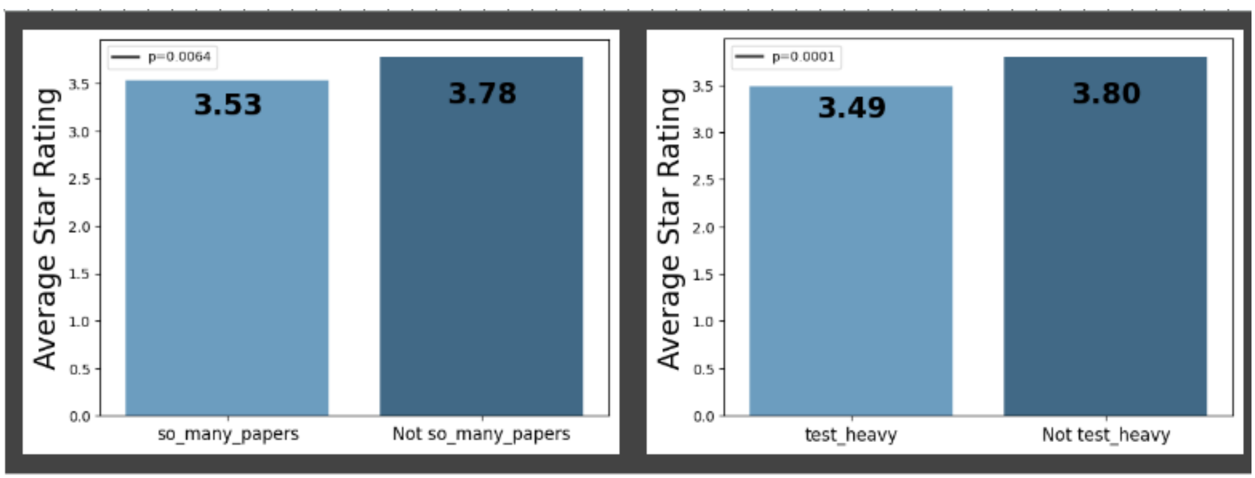

Average Star Rating
p=0.0006
4.01
3.71
Attractive
Not Attractive
p=0.1646
3.63
3.76
beware_of_pop_quizzes
Not beware_of_pop_quizzes
p=0.0000
3.91
3.60
caring
Not caring
p=0.0283
3.83
3.70
clear_grading_criteria
Not clear_grading_criteria
p=0.0108
3.90
3.71
extra_credit
Not extra_credit
p=0.0000
3.60
3.87
get_ready_to_read
Not get_ready_to_read


p=0.0221
Average Star Rating
3.60
3.78
graded_by_few_things
Not graded_by_few_things
p=0.1797
Average Star Rating
3.64
3.76
group_projects
Not group_projects
p=0.0000
Average Star Rating
3.96
3.64
hilarious
Not hilarious
p=0.0000
Average Star Rating
4.00
3.61
inspirational
Not inspirational

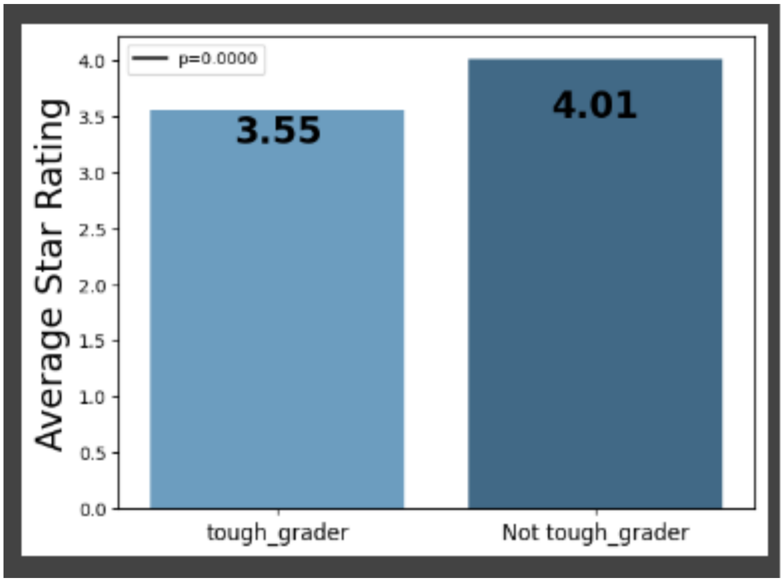

p=0.9475
Average Star Rating
3.75
3.75
IsCourseOnline
Not IsCourseOnline
p=0.0000
Average Star Rating
3.57
3.85
lecture_heavy
Not lecture_heavy

- 

Average Star Rating
p=0.0000
3.56
3.85
lots_of_homework
Not lots_of_homework
Average Star Rating
p=0.9644
3.75
3.75
participation_matters
Not participation_matters
Average Star Rating
p=0.0000
4.00
3.50
respected
Not respected
Average Star Rating
p=0.0020
3.65
3.84
skip_class
Not skip_class

- 

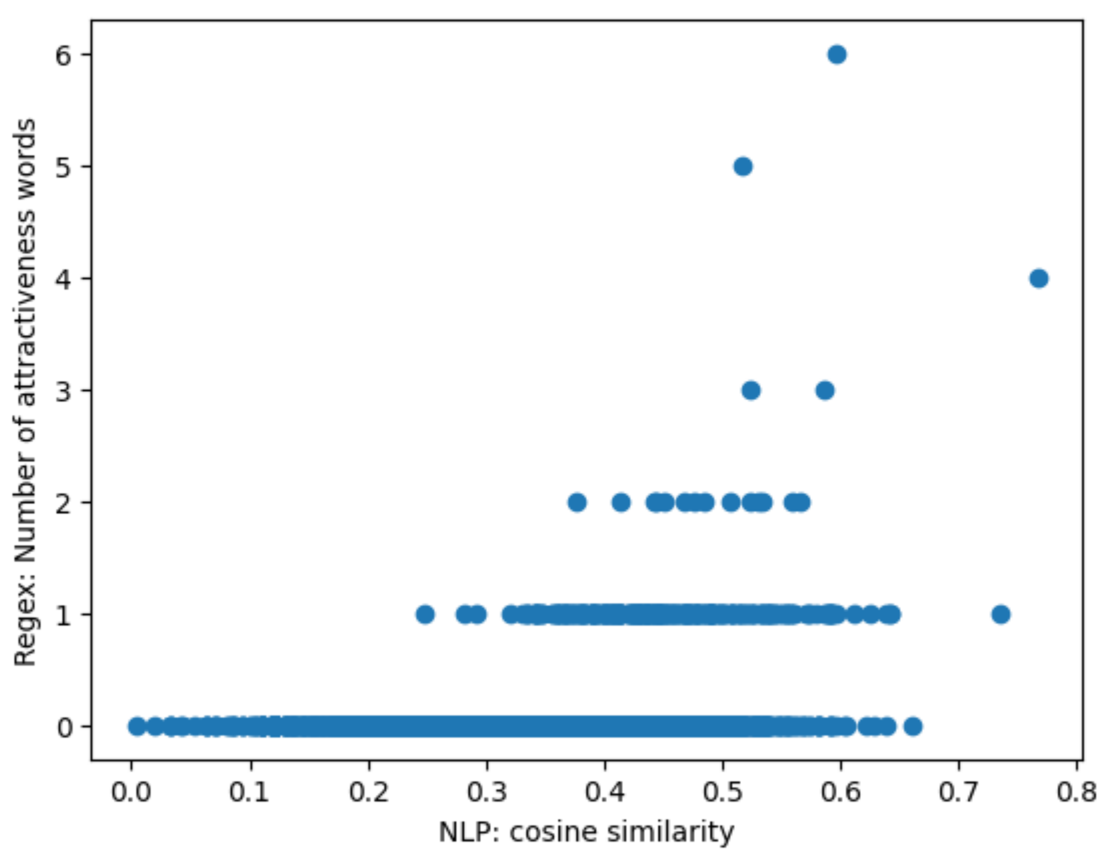
Average Star Rating
p=0.0064
3.53
3.78
so_many_papers
Not so_many_papers
Average Star Rating
p=0.0001
3.49
3.80
test_heavy
Not test_heavy

- 

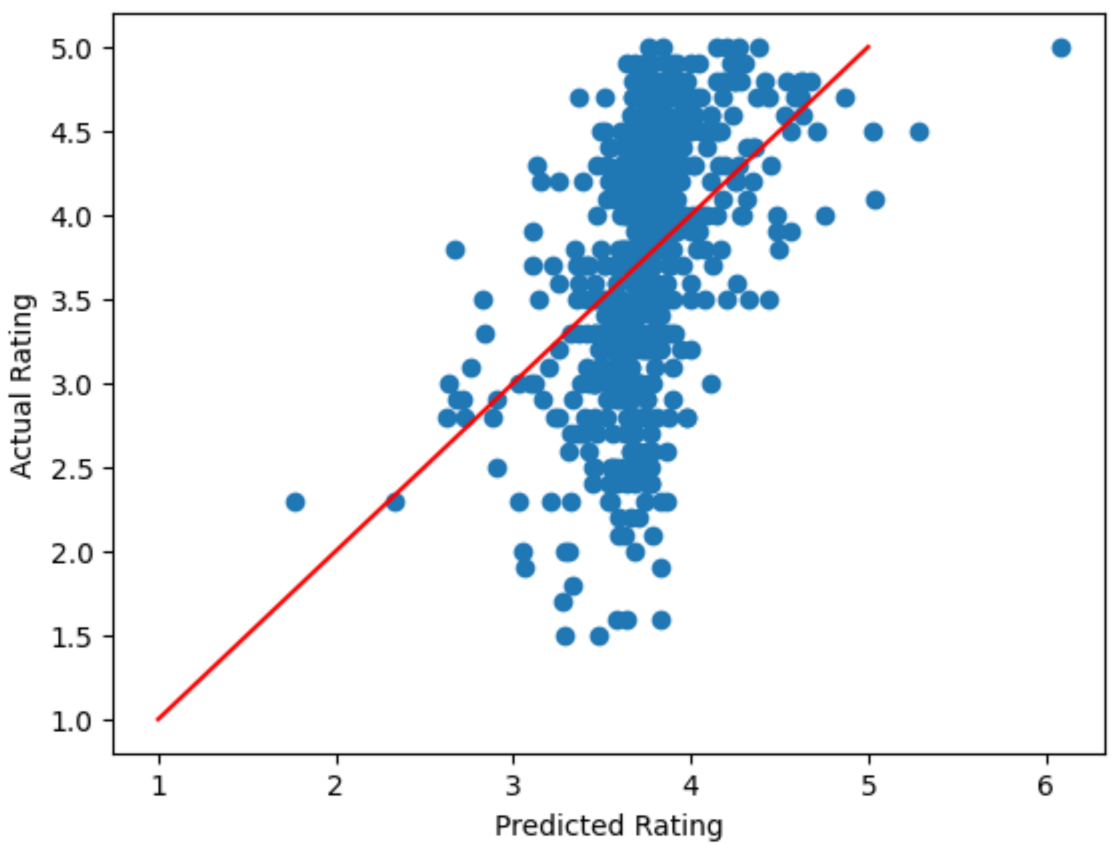

- Chi-Square tests were used on every possible pair of the features listed in the data preparation sentence to test whether any features depended on each other to a significant degree. 133 dependencies were found to exist. All such combinations are listed in Appendix B.
    - Attractiveness was the only variable that did not have a single p-value < 0.05 when compared with every other variable. All other features had at least five such dependencies, and most had over 10
        - This suggests that there is no specific feature of a professor, such as being respected, an amazing lecturer, or hilarious, which correlates with students finding them attractive. Students arrive at the judgment of attractiveness by an unrelated process

- **Machine Learning: Regarding Methodology and Results**
    - **Natural Language Processing:**
        - RateMyProfessors doesn't come with a rating for attractiveness. To figure out how much of an impact attractiveness plays on ratings, we needed to create our measure. For the dataset we used, starting with 20,000 lines, it probably would have been possible to label this manually with nearly 100% accuracy. However, because our original plans were to conduct this on a much larger scale (9 million lines), we wanted to work on a solution that could realistically work on a larger dataset.
        - We spent a lot of time working with NLP and embeddings before ultimately deciding that it would be best to use Regex instead to determine if a comment thought a professor was attractive.
        - We iterated through each comment in the dataset, found its average vector embeddings, and checked its cosine value with the word2vec vectors of model "attractive" sentence, such as "The Professor is attractive" or "cute, attractive, hot, handsome, good-looking, gorgeous, adorable, stunning, sexy, eye-candy, beautiful, pretty."

- Therefore, the input was the comment column, and the output was a cosine value displaying its similarity to our model sentence
- Unfortunately, the results were not great. With a filter set to catch all comments with a cosine value above 0.5, it caught about half of the comments in the dataset about attractiveness. Additionally, about half of the comments it flagged were simply positive sentiments (which word2vec somewhat similar to attractiveness) and had nothing to do with attractiveness. This seemed to be a common problem with the embeddings.
- We attempted to average only the embedding vectors of the adjectives in a review, by using an nltk feature suggested by ChatGPT. While our comments to analyze were now more similar to our model sentence, this did not improve our results
- We tried instead using regular expressions with the pattern`"cute|attractive|hot|handsome|good-looking|gorgeous|adorable|stunning|sexy|eye-candy|beautiful." This produced much more accurate results: We estimated that 75-80% of the comments it identified were correct and that it identified 80-90% of the total comments about the professor's attractiveness. The main issue here was that a simple regex pattern lists comments like "not hot" as "hot." Additionally, some people would use more unusual words to describe their professors as attractive. Still, this was as close as we could get to success.
- Ultimately, we used the regex results as our attractiveness column.

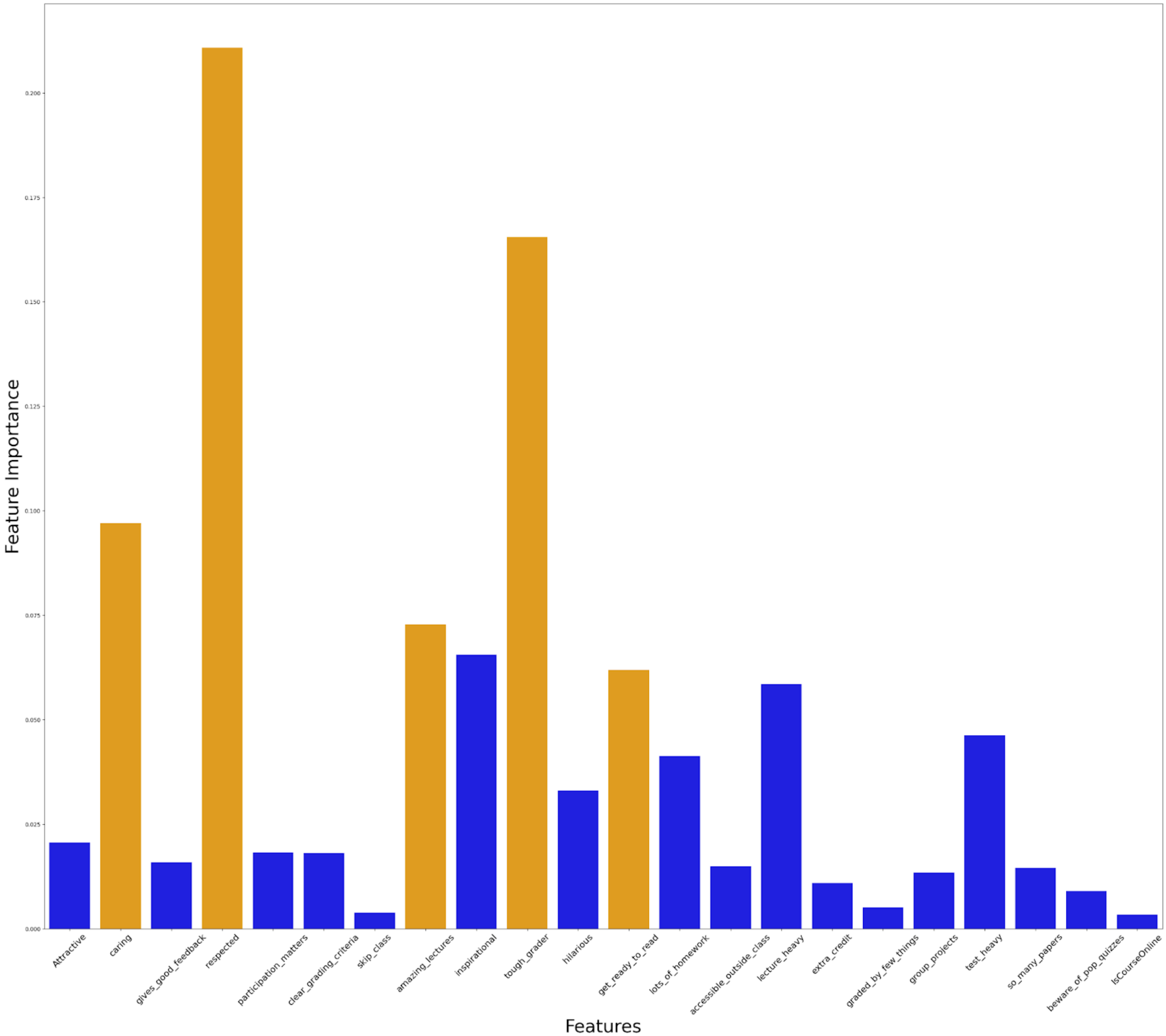


- This graph compares the cosine similarity of the reviews and the sample "attractiveness" comment to the number of words found matching the regular expression used
- For sentences with a cosine < 0.3, the embeddings matched up well with the regex. However, any higher and the same number of keywords found by the regular expression would result in very different cosine similarity values. Even comments with 2 keywords, which in reality almost always referred to the professor as attractive, were rated anywhere from 0.35 to 0.6 by the NLP
- The regex was not incredibly accurate but was at least a lot less volatile than the NLP

- **Regression:**
  - Our plan to determine which factors contributed most to professor ratings was to conduct an ElasticNet regression model on an updated dataset.
  - Our input was the features listed above in data preparation, for each professor. The expected output was a predicted star rating.
  - The regression, like the NLP, did not work well, which was surprising to us, as there are legitimate correlations between several of the variables and star ratings.
  - The R^2 stayed around 0.2 on the validation data. This would indicate that the overwhelming majority of the rating is not explained by the data
  - 

  - The model's problem is obvious from looking at the chart. Almost all of the model's predictions are clustered between 3 and 4, while the actual ratings were well spread out between 2 and 5. By making a less conservative model, it might be possible to fit this line, given all the correlations that exist. However, it does not currently appear to be in our skill set. Further hyperparameters could fix this.

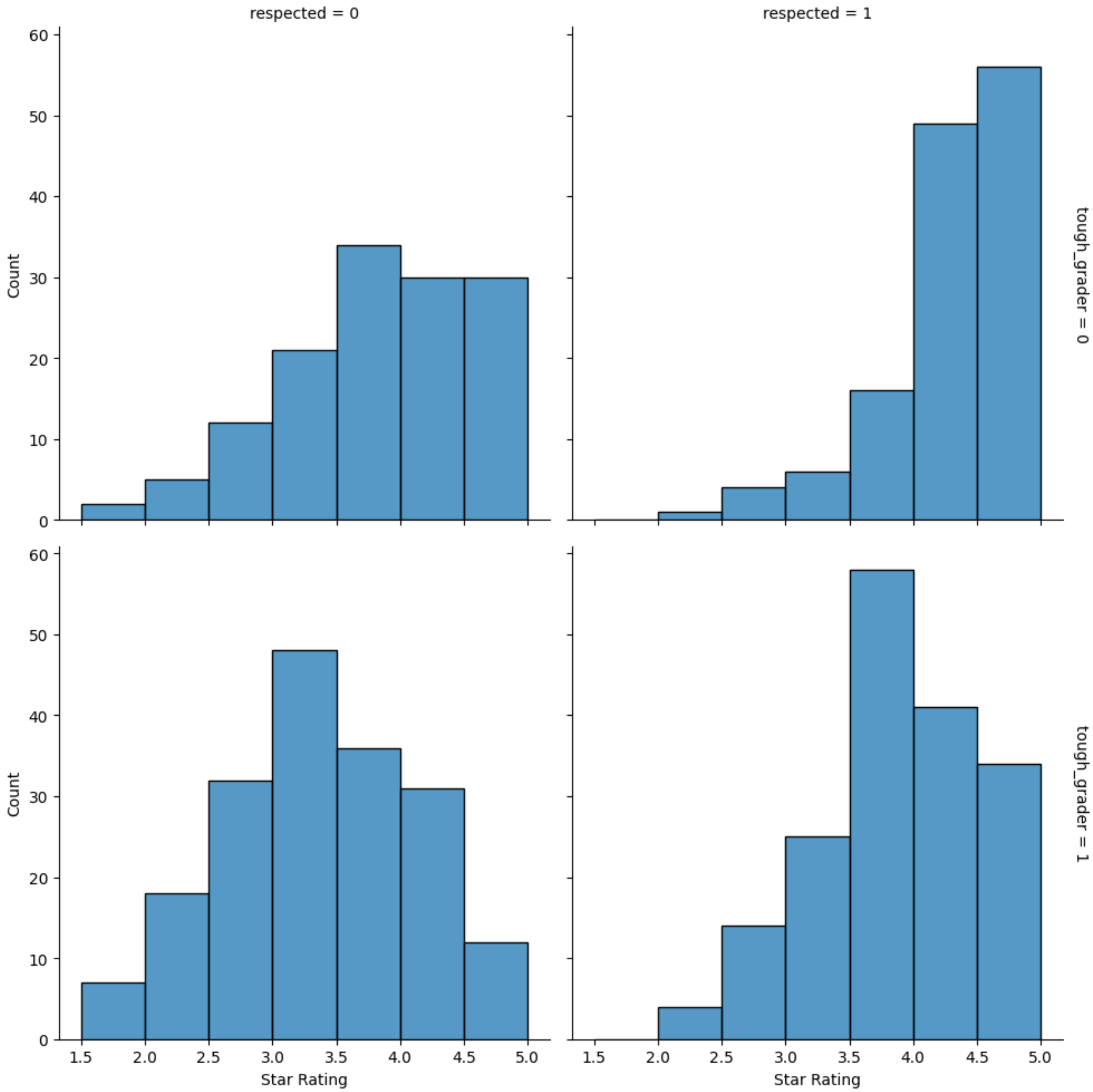
        - One possible explanation for the failure of the regression model is that the 1-5 scale for ratings, as pointed out by Noel Otu and Ntiense E. Otu, is subjective in that different people see the scale differently. Two professors in our data may have vastly different average ratings despite having almost identical tags.
        - Note: we tried out training the model with fewer columns to avoid multicollinearity, but this did not improve it significantly.
    - **Random Forest Classification**
        - For Random Forest classification, we split the rows into 2 groups, ratings > 3.7 and ratings <= 3.7. The input had the same features as the regression, but the labels were instead the boolean values for star_rating as created above. The expected output was a prediction of whether the professor's rating would be above or below 3.7. We constructed a hyperparameter grid using scikit-learn's RandomizedSearchCV function. Despite having options of max_depth 3, 5, or 7, the best performer had a depth of 3. This suggests that any higher depth led to overfitting significant enough to lower the score of the model. Our final model had a test accuracy of about 0.65. This is still not great, but it's better than a guess.
        - Although the model is not good enough to use for predictions, we can still review the most important features detected by the model. Among the most important features were:
            - Respected
            - Tough grader
            - Caring
            - Amazing lectures
            - Get ready to read
        - Among the least important were:
            - Skip class? You won't pass
            - Graded by few things
            - Online course
        - Notably, respected and tough grader are significantly more important than any other values




- This bar chart visualizes the importance of each feature of the dataset

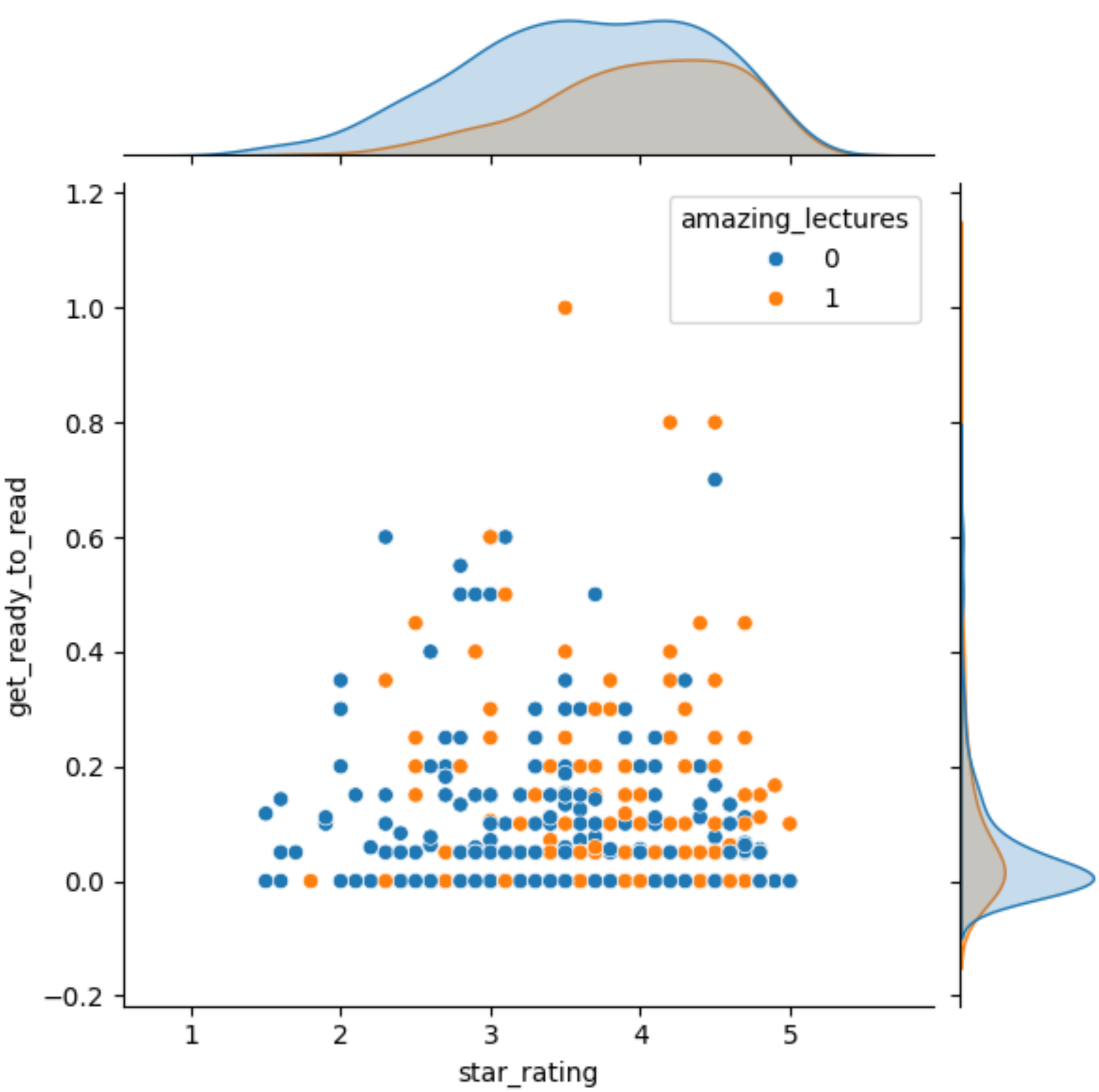
## Additional Visualizations

Figure A:



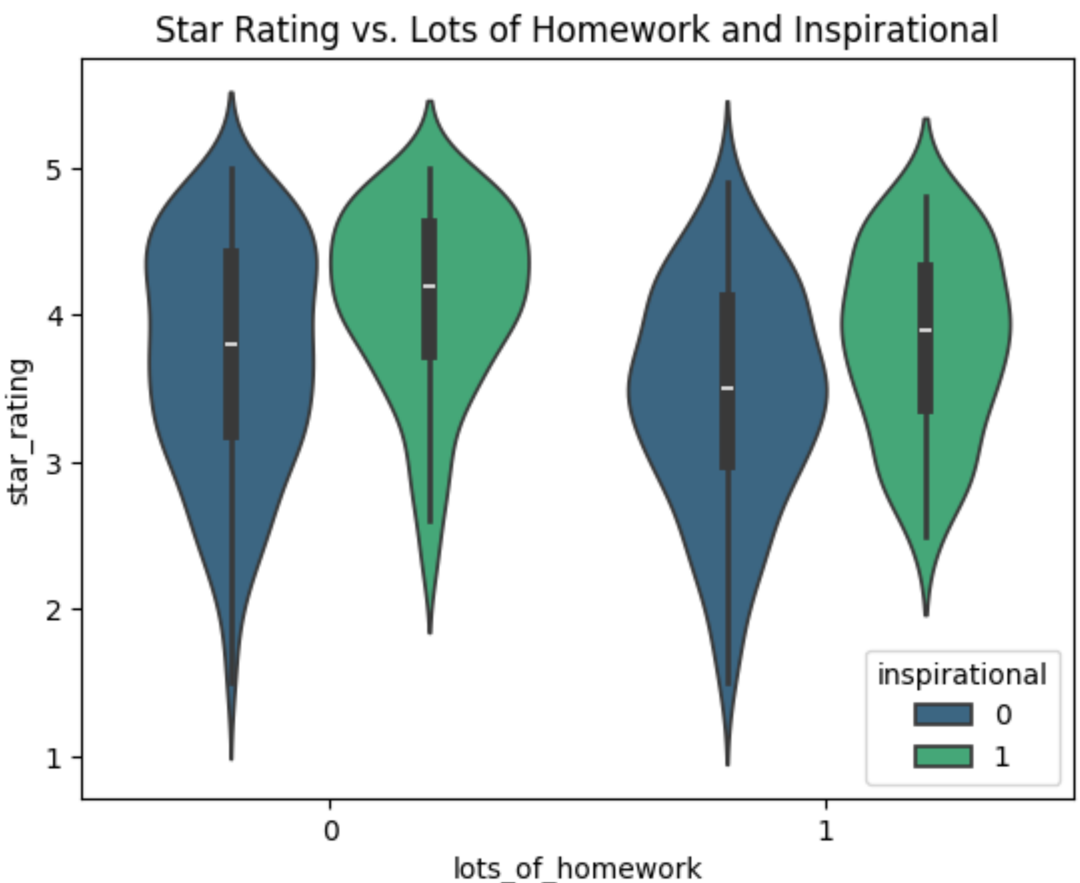

- Figure A is a Seaborn histogram that compares the star ratings of the two most important features we found: respected and tough grader.
    - Star ratings are lower for tough graders who are also not respected. However, professors who are not tough graders and are respected have significantly higher ratings.

Figure B:




- Figure B is a seaborn joint plot which shows the relationship between professors who give amazing lectures and professors who assign a lot of reading (get_ready_to_read), with respect to star rating
    - Amazing lecturers have on average a higher star rating than non-amazing lecturers, as expected.
    - However, amazing lecturers were not more likely to assign lots of reading homework.

Figure C:




- Figure C is a Seaborn violin plot, which shows that professors giving less homework are clustered at higher star ratings, while inspirational professors do not tend to reach as low ratings as non-inspirational professors

## Conclusions:

The most important factors of an RMP review include:

**Difficulty decreases ratings, such as:**

- Tough grading
- Heavy homework load

**Positive Personality traits increase ratings, such as:**

- Respected
- Amazing lectures

**Attractiveness** likely has a **slightly positive impact** on the rating of a professor.

However, our results should not be fully trusted, as the feature importance is based on a poor model, and the attractiveness data is somewhat inaccurate. As a team, we learned how to incorporate all the machine learning methods to understand the data set and create relationships from it.

## Bibliography:


Babin, J. Jobu, et al. "Beauty Premiums Among Academics." *Economics of Education Review*, Elsevier Ltd., 25 Aug. 2020, www.sciencedirect.com/science/article/pii/S0272775719307538?ref=pdf_download&fr=RR-2&rr=831634dd7ccc6ac5.

Gotlieb, Jerry. "Variables That Can Affect Student Ratings of Their Professors." Journal of Learning in Higher Education, Vol. 9, Issue 1, Spring 2013, https://files.eric.ed.gov/fulltext/EJ1144100.pdf

Nushi, Musa, et al. "Characteristics of an Effective University Professor from Students' Perspective: Are the Qualities Changing?" *Frontiers*, Frontiers Medias, 7 Mar. 2022, www.frontiersin.org/articles/10.3389/feduc.2022.842640/full.

Otu, N., & Otu, N. E. (2023). Student Evaluations of Teaching Are Mostly Awfully Wrong. *Universal Journal of Educational Research,* 2(2), 168-183.

Seiler, Michael J., et al. *Professor, Student, and Course Attributes That Contribute to ... - FMA*, www.fma.org/assets/docs/JAF/1999/fpe1999_9_2_11.pdf. Accessed 6 Dec. 2023.


## Appendix A: ChatGPT Citations

**Help with determining keywords for regex to find attractiveness**
https://chat.openai.com/share/d694f1ec-f037-41b7-836d-185dafa3b51d

**Converting sentences to the vectors of just adjectives**
https://chat.openai.com/share/0db218af-5861-4259-aca8-9364f170857f

**Syntax error in dataframe filtering**
https://chat.openai.com/share/19238d47-8732-4d22-946f-66d1f2b71d89

**Hyperparameter grid help**
https://chat.openai.com/share/7d52faf1-81b6-4402-aad0-c3affa94fa36

**.notna() to filter out n/a values from a column**
https://chat.openai.com/share/1f058e1b-e3ae-42e8-bdf3-86688c81b211

**Regex to convert the tag_professor string into columns for the total of each tag**
https://chat.openai.com/share/7299f3b7-949a-48c3-a941-711a8d17f4b5

**SNS heatmap formatting**
https://chat.openai.com/share/913dd1ae-0c3d-4df8-be13-0bb92ec3b2ed

**SNS bar graph formatting**
https://chat.openai.com/share/1dc09d72-3f82-4089-a631-359d905212e1

## Appendix B: List of all dependent variable pairs according to Chi-Square test

**Total dependent variable pairs for each feature**

caring | 14
respected | 10
participation_matters | 17
clear_grading_criteria | 17
skip_class | 18
amazing_lectures | 15
inspirational | 11
hilarious | 15
lots_of_homework | 15
accessible_outside_class | 14
lecture_heavy | 14
extra_credit | 15
graded_by_few_things | 13
group_projects | 11
test_heavy | 13
beware_of_pop_quizzes | 12
tough_grader | 13
get_ready_to_read | 13
so_many_papers | 11
IsCourseOnline | 5

**List of all dependent variable pairs**

caring and respected | p-value: 3.730849110588085e-13
[[189 102]
 [119 216]]

caring and participation_matters | p-value: 0.0005468174698148895
[[145 146]
 [120 215]]

caring and clear_grading_criteria | p-value: 2.429070132714339e-10
[[148 143]
 [ 87 248]]

caring and skip_class | p-value: 5.265034951086794e-06
[[166 125]
 [129 206]]

caring and amazing_lectures | p-value: 2.4319385242123334e-08
[[134 157]
 [ 82 253]]

caring and inspirational | p-value: 1.9989267571514376e-09
[[138 153]
 [ 81 254]]

caring and hilarious | p-value: 3.792135599478641e-10
[[134 157]
 [ 74 261]]

caring and lots_of_homework | p-value: 0.000388953288732961
[[120 171]
 [ 92 243]]

caring and accessible_outside_class | p-value: 6.4365800170211e-11
[[ 99 192]

[ 40 295]]

caring and lecture_heavy | p-value: 0.005724894314348595
[[125 166]
[107 228]]

caring and extra_credit | p-value: 9.845801883765693e-09
[[ 85 206]
[ 36 299]]

caring and graded_by_few_things | p-value: 2.782736105512288e-05
[[ 71 220]
[ 38 297]]

caring and group_projects | p-value: 0.0023808237546294616
[[ 53 238]
[ 32 303]]

caring and test_heavy | p-value: 0.002242058314858917
[[ 62 229]
[ 40 295]]

respected and participation_matters | p-value: 0.0033745281712206604
[[149 159]
[116 202]]

respected and clear_grading_criteria | p-value: 2.230078578110117e-08
[[150 158]
[ 85 233]]

respected and skip_class | p-value: 2.4305827597006037e-05
[[172 136]
[123 195]]

respected and amazing_lectures | p-value: 7.015389343995988e-18

[[158 150]
 [ 58 260]]

respected and inspirational | p-value: 1.854144397893113e-21
[[165 143]
 [ 54 264]]

respected and hilarious | p-value: 1.7864805188118892e-14
[[148 160]
 [ 60 258]]

respected and accessible_outside_class | p-value: 1.9041807929310562e-10
[[102 206]
 [ 37 281]]

respected and extra_credit | p-value: 1.2210428358371463e-06
[[ 84 224]
 [ 37 281]]

respected and beware_of_pop_quizzes | p-value: 0.02903965507277243
[[ 43 265]
 [ 26 292]]

participation_matters and clear_grading_criteria | p-value: 0.00023466560525855004
[[122 143]
 [113 248]]

participation_matters and skip_class | p-value: 1.249018664650054e-07
[[158 107]
 [137 224]]

participation_matters and amazing_lectures | p-value: 0.00033832974911802944
[[113 152]
 [103 258]]

participation_matters and inspirational | p-value: 0.012107013398848412
[[108 157]
 [111 250]]

participation_matters and tough_grader | p-value: 0.005127805829510825
[[170  95]
 [190 171]]

participation_matters and hilarious | p-value: 0.00022998333543914113
[[110 155]
 [ 98 263]]

participation_matters and get_ready_to_read | p-value: 4.8722605098743675e-08
[[153 112]
 [128 233]]

participation_matters and lots_of_homework | p-value: 0.0003885493864500122
[[111 154]
 [101 260]]

participation_matters and accessible_outside_class | p-value: 0.0011861084861389992
[[ 76 189]
 [ 63 298]]

participation_matters and extra_credit | p-value: 0.0017494580034802987
[[ 67 198]
 [ 54 307]]

participation_matters and graded_by_few_things | p-value: 0.027139490602218385
[[ 57 208]
 [ 52 309]]

participation_matters and group_projects | p-value: 0.003117084645676459
[[ 49 216]
 [ 36 325]]

```
participation_matters and so_many_papers | p-value: 0.009977585277995395
[[ 45 220]
 [ 35 326]]


participation_matters and beware_of_pop_quizzes | p-value: 7.825792923735215e-05
[[ 45 220]
 [ 24 337]]


participation_matters and IsCourseOnline | p-value: 0.029617026963655285
[[ 26 239]
 [ 18 343]]


clear_grading_criteria and skip_class | p-value: 8.633916825081298e-07
[[141  94]
 [154 237]]


clear_grading_criteria and amazing_lectures | p-value: 2.5744237532571744e-10
[[118 117]
 [ 98 293]]


clear_grading_criteria and inspirational | p-value: 0.008149722180266961
[[ 98 137]
 [121 270]]


clear_grading_criteria and tough_grader | p-value: 0.01652507000273325
[[150  85]
 [210 181]]


clear_grading_criteria and hilarious | p-value: 1.879992260761496e-05
[[103 132]
 [105 286]]


clear_grading_criteria and get_ready_to_read | p-value: 0.0008965863558412967
[[126 109]
 [155 236]]
```

clear_grading_criteria and lots_of_homework | p-value: 6.970103604352402e-08
[[111 124]
 [101 290]]

clear_grading_criteria and accessible_outside_class | p-value: 1.7236929641748466e-07
[[ 79 156]
 [ 60 331]]

clear_grading_criteria and lecture_heavy | p-value: 7.980104179296483e-08
[[119 116]
 [113 278]]

clear_grading_criteria and extra_credit | p-value: 1.0566427216424236e-12
[[ 80 155]
 [ 41 350]]

clear_grading_criteria and graded_by_few_things | p-value: 2.8551350549282333e-07
[[ 65 170]
 [ 44 347]]

clear_grading_criteria and group_projects | p-value: 0.005223924066581949
[[ 44 191]
 [ 41 350]]

clear_grading_criteria and test_heavy | p-value: 2.134540019967039e-07
[[ 62 173]
 [ 40 351]]

clear_grading_criteria and IsCourseOnline | p-value: 0.0037278829034799853
[[ 26 209]
 [ 18 373]]

skip_class and amazing_lectures | p-value: 5.610022013397705e-07
[[132 163]

[ 84 247]]

skip_class and inspirational | p-value: 0.016381538576448598
[[118 177]
[101 230]]

skip_class and tough_grader | p-value: 6.4215388692775e-19
[[225  70]
[135 196]]

skip_class and hilarious | p-value: 6.567193537762678e-05
[[122 173]
[ 86 245]]

skip_class and get_ready_to_read | p-value: 0.0006901755329916031
[[154 141]
[127 204]]

skip_class and lots_of_homework | p-value: 2.0045108791172957e-16
[[149 146]
[ 63 268]]

skip_class and accessible_outside_class | p-value: 2.0603998184761042e-10
[[ 99 196]
[ 40 291]]

skip_class and lecture_heavy | p-value: 3.0067571499583116e-05
[[135 160]
[ 97 234]]

skip_class and extra_credit | p-value: 6.920182922754478e-07
[[ 82 213]
[ 39 292]]

skip_class and graded_by_few_things | p-value: 1.0362022808423335e-06

[[ 75 220]
 [ 34 297]]

skip_class and group_projects | p-value: 0.014639543011191516
[[ 51 244]
 [ 34 297]]

skip_class and test_heavy | p-value: 2.041194594897536e-12
[[ 81 214]
 [ 21 310]]

skip_class and so_many_papers | p-value: 0.009592270605778762
[[ 49 246]
 [ 31 300]]

skip_class and beware_of_pop_quizzes | p-value: 3.232353124504809e-07
[[ 53 242]
 [ 16 315]]

amazing_lectures and inspirational | p-value: 6.3214037782109566e-18
[[125  91]
 [ 94 316]]

amazing_lectures and hilarious | p-value: 1.530059134168707e-22
[[127  89]
 [ 81 329]]

amazing_lectures and get_ready_to_read | p-value: 0.013969485711208655
[[112 104]
 [169 241]]

amazing_lectures and accessible_outside_class | p-value: 7.94399620708831e-08
[[ 75 141]
 [ 64 346]]

amazing_lectures and lecture_heavy | p-value: 0.0007100349616417615
[[100 116]
 [132 278]]

amazing_lectures and extra_credit | p-value: 1.2749171751899535e-06
[[ 65 151]
 [ 56 354]]

amazing_lectures and graded_by_few_things | p-value: 0.008387104327002397
[[ 50 166]
 [ 59 351]]

amazing_lectures and test_heavy | p-value: 0.0011274387310665638
[[ 50 166]
 [ 52 358]]

amazing_lectures and so_many_papers | p-value: 0.0027375323627489567
[[ 40 176]
 [ 40 370]]

amazing_lectures and beware_of_pop_quizzes | p-value: 0.0006560944662670457
[[ 37 179]
 [ 32 378]]

inspirational and hilarious | p-value: 6.338022862088251e-11
[[110 109]
 [ 98 309]]

inspirational and accessible_outside_class | p-value: 6.150191908303466e-05
[[ 69 150]
 [ 70 337]]

inspirational and extra_credit | p-value: 0.0026372126611631844
[[ 57 162]
 [ 64 343]]

inspirational and group_projects | p-value: 0.03203651364684728
[[ 39 180]
 [ 46 361]]

inspirational and beware_of_pop_quizzes | p-value: 0.04885812507999906
[[ 32 187]
 [ 37 370]]

tough_grader and get_ready_to_read | p-value: 2.8939256167265415e-13
[[207 153]
 [ 74 192]]

tough_grader and lots_of_homework | p-value: 6.8281997084773936e-15
[[168 192]
 [ 44 222]]

tough_grader and accessible_outside_class | p-value: 0.014520727405705166
[[ 93 267]
 [ 46 220]]

tough_grader and lecture_heavy | p-value: 4.7562954368619496e-07
[[164 196]
 [ 68 198]]

tough_grader and extra_credit | p-value: 0.025420519020445217
[[ 81 279]
 [ 40 226]]

tough_grader and graded_by_few_things | p-value: 0.0015830805376926304
[[ 78 282]
 [ 31 235]]

tough_grader and group_projects | p-value: 8.769192898591409e-05
[[ 66 294]
 [ 19 247]]
tough_grader and test_heavy | p-value: 5.710809564848323e-07

[[ 82 278]
 [ 20 246]]

tough_grader and so_many_papers | p-value: 5.11185860439141e-08
[[ 69 291]
 [ 11 255]]

tough_grader and beware_of_pop_quizzes | p-value: 0.0052166590151910665
[[ 51 309]
 [ 18 248]]

hilarious and lots_of_homework | p-value: 0.03060551796203989
[[ 83 125]
 [129 289]]

hilarious and accessible_outside_class | p-value: 6.053311545073421e-10
[[ 77 131]
 [ 62 356]]

hilarious and lecture_heavy | p-value: 0.018439454289626536
[[ 91 117]
 [141 277]]

hilarious and extra_credit | p-value: 4.738146517514129e-06
[[ 62 146]
 [ 59 359]]

hilarious and graded_by_few_things | p-value: 1.5980913043092656e-05
[[ 56 152]
 [ 53 365]]

hilarious and test_heavy | p-value: 0.014791398887233098
[[ 45 163]
 [ 57 361]]

hilarious and so_many_papers | p-value: 0.02340867755382575
[[ 36 172]
 [ 44 374]]

hilarious and beware_of_pop_quizzes | p-value: 0.020179841859801095
[[ 32 176]
 [ 37 381]]

get_ready_to_read and lots_of_homework | p-value: 7.414754071960305e-05
[[119 162]
 [ 93 252]]

get_ready_to_read and lecture_heavy | p-value: 2.8499666839290447e-08
[[138 143]
 [ 94 251]]

get_ready_to_read and graded_by_few_things | p-value: 1.3049349145174211e-05
[[ 70 211]
 [ 39 306]]

get_ready_to_read and group_projects | p-value: 0.015235180332210963
[[ 49 232]
 [ 36 309]]

get_ready_to_read and test_heavy | p-value: 0.00011603764897954347
[[ 64 217]
 [ 38 307]]

get_ready_to_read and so_many_papers | p-value: 2.0322900228131542e-07
[[ 58 223]
 [ 22 323]]

get_ready_to_read and beware_of_pop_quizzes | p-value: 0.0005185463497108081
[[ 45 236]
 [ 24 321]]

get_ready_to_read and IsCourseOnline | p-value: 0.005954153064572711
[[ 29 252]
 [ 15 330]]

lots_of_homework and accessible_outside_class | p-value: 1.933586550938973e-06
[[ 71 141]
 [ 68 346]]

lots_of_homework and lecture_heavy | p-value: 2.8550132342125397e-05
[[103 109]
 [129 285]]

lots_of_homework and extra_credit | p-value: 7.450376701972199e-05
[[ 60 152]
 [ 61 353]]

lots_of_homework and graded_by_few_things | p-value: 0.0011602735934549012
[[ 52 160]
 [ 57 357]]

lots_of_homework and group_projects | p-value: 3.954841490849481e-06
[[ 48 164]
 [ 37 377]]

lots_of_homework and test_heavy | p-value: 7.071088541592793e-10
[[ 62 150]
 [ 40 374]]

lots_of_homework and so_many_papers | p-value: 2.440049096315418e-07
[[ 48 164]
 [ 32 382]]

lots_of_homework and beware_of_pop_quizzes | p-value: 0.0010684250111597422
[[ 36 176]
 [ 33 381]]

accessible_outside_class and lecture_heavy | p-value: 0.005357629544335878
```
[[ 66  73]
 [166 321]]
```

accessible_outside_class and extra_credit | p-value: 0.0001406665019155542
```
[[ 43  96]
 [ 78 409]]
```

accessible_outside_class and graded_by_few_things | p-value: 0.0002882921221353868
```
[[ 39 100]
 [ 70 417]]
```

accessible_outside_class and test_heavy | p-value: 9.731795555175313e-11
```
[[ 48  91]
 [ 54 433]]
```

lecture_heavy and extra_credit | p-value: 1.4977969821766218e-05
```
[[ 66 166]
 [ 55 339]]
```

lecture_heavy and graded_by_few_things | p-value: 1.4394862517235067e-07
```
[[ 65 167]
 [ 44 350]]
```

lecture_heavy and group_projects | p-value: 0.0016885764087748033
```
[[ 45 187]
 [ 40 354]]
```

lecture_heavy and test_heavy | p-value: 2.837719560527065e-11
```
[[ 68 164]
 [ 34 360]]
```

lecture_heavy and so_many_papers | p-value: 0.028233493694339553
```
[[ 39 193]
```

```
 [ 41 353]]


extra_credit and test_heavy | p-value: 0.0031202356079937047
[[ 31  90]
 [ 71 434]]


extra_credit and beware_of_pop_quizzes | p-value: 0.04638638940569451
[[ 20 101]
 [ 49 456]]


extra_credit and IsCourseOnline | p-value: 0.04794794802903692
[[ 14 107]
 [ 30 475]]


graded_by_few_things and group_projects | p-value: 0.002841651702349435
[[ 25  84]
 [ 60 457]]


graded_by_few_things and test_heavy | p-value: 5.495915338314928e-10
[[ 40  69]
 [ 62 455]]


group_projects and so_many_papers | p-value: 1.0035480609131272e-05
[[ 24  61]
 [ 56 485]]


test_heavy and beware_of_pop_quizzes | p-value: 0.030592406792022207
[[ 18  84]
 [ 51 473]]


so_many_papers and beware_of_pop_quizzes | p-value: 0.029848966450468854
[[ 15  65]
 [ 54 492]]
so_many_papers and IsCourseOnline | p-value: 0.022375015717905518
[[ 11  69]
 [ 33 513]]
```