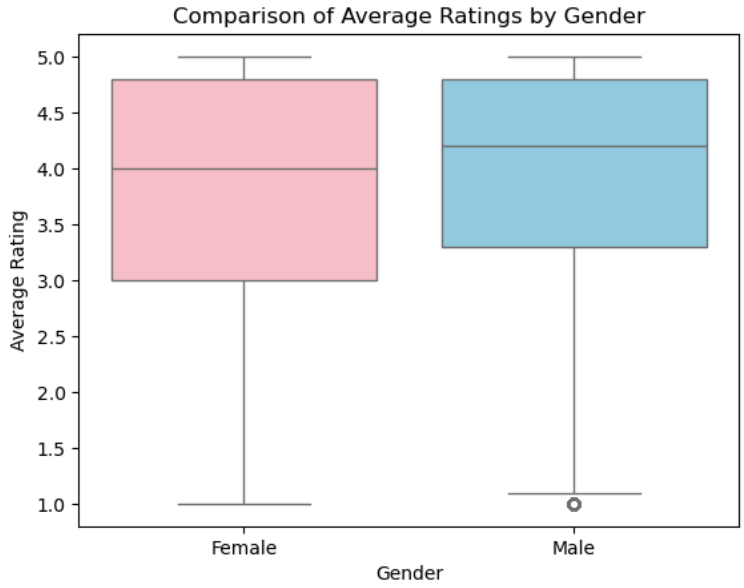
Project Report

Before doing the data analysis, for easier visual representation locating the data, I added the names to each one of the columns and rows for both sets of data. For questions that contain uneven number of rows and ones that contain nulls, I used dropna() to clean the data and even the rows.

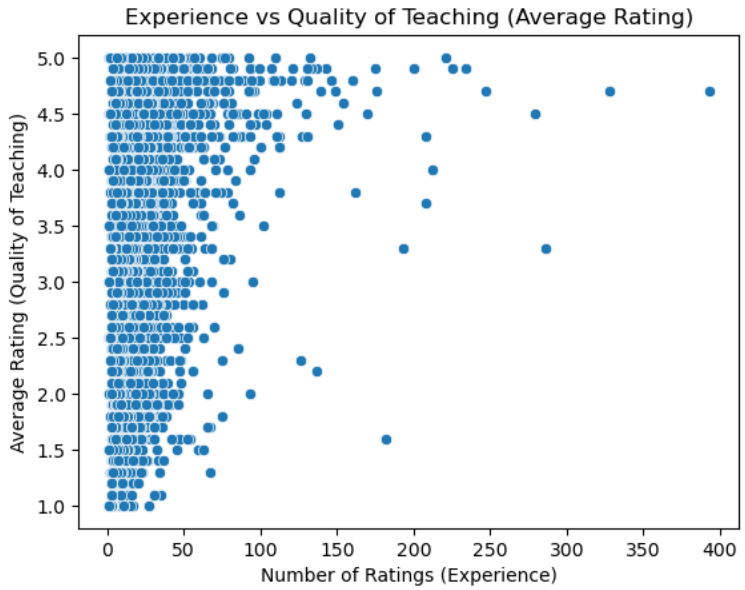
Question 1: Is there evidence of a pro-male gender bias in the dataset?

Since the dataset does not follow a normal distribution, I used the mann-whiteney-u test as the significance test. After applying the mann-whiteney-u test, the U statistics was 406757411.5 and the p_value was 2.443998308992039e-05. Base on the numbers, there is evidence supporting that male professors tends to have higher ratings than female professors. And a p_value below 0.05 means that we can reject the null hypothesis.



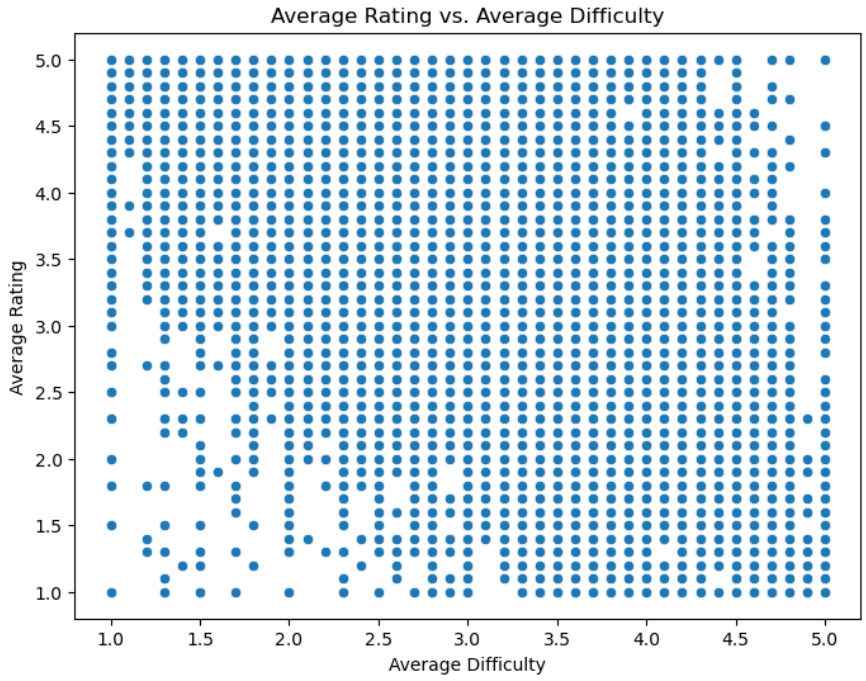
Question 2: is there an effect of experience on the quality of teaching.

For the experience of teaching, I used the "num_ of_ratings" column and for the quality of teaching, I used the spearman's correlation coefficient. The pearson coefficient is 0.037360079371875 8 and the p-value is 4.688152363959901e-23. There seems to be a slight positive relationship between quality of teaching and the effect of teaching, that the more experience a professor has, the higher the rating they tend to have. The p-value below 0.05 means that we can reject the null hypothesis and that it is statistically significant. However, since you can only conclude a causal relationship through an experiment, there is not enough evidence to say that the more years of teaching leads to higher quality of teaching.



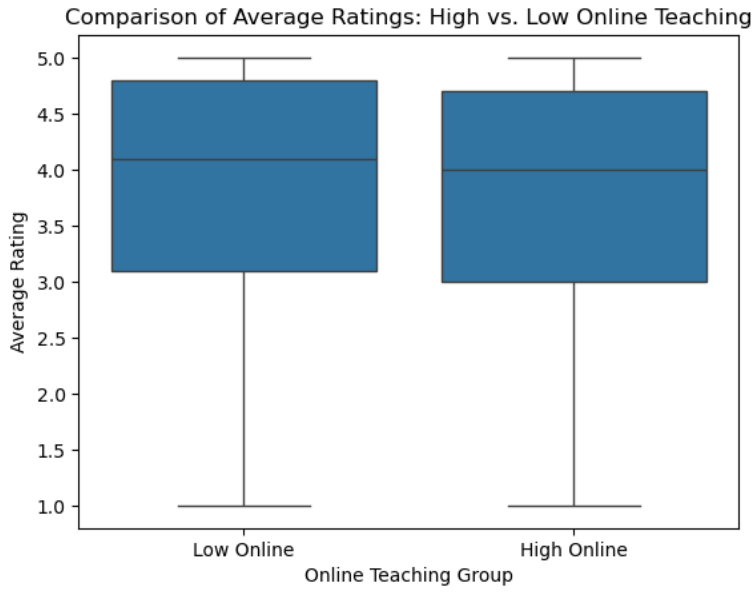
Question 3: whats the relationship between average rating and average difficulty.

For this question, I used the mann-whitney-u test. The correlation is -0.5367776383854018 and the p-value is 0.00. There is a negative relationship between the average rating and difficulties. Because the p-value is 0.00, it is also statistically significant.



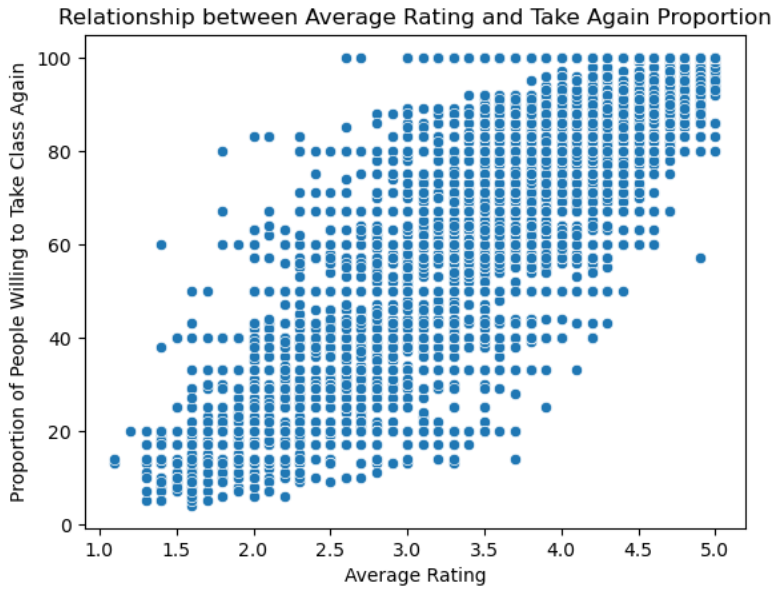
Question 4: Do professors who teach online classes get rated higher or lower than those who don't?

I split the data by the median if "num_of_online." The ones greater than the median is classified as high online, and the ones below is classified as low online. I used the t-test to test the significance. The t-statistics is -13.3398 and the p-value is 0.0000. This show that there is a significant difference between the ratings of high online group and low online group. This is also statistically significant, showing that high online classes are associated with lower ratings.

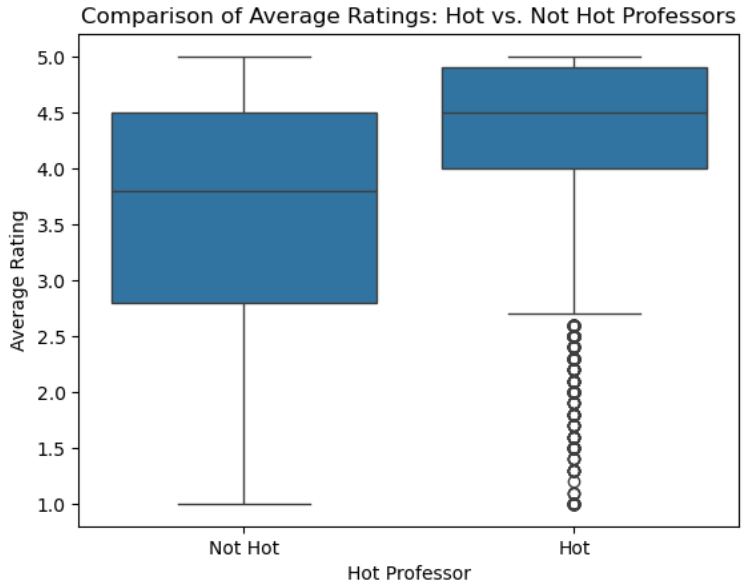


Question 5: What's the relationship between average rating and the proportion of people who would take the class again?

To find the relationship between average rating and the proportion of people who would take again, I used Pearson's correlation. After dropping the null data, I set X as the average rating and y as the proportion of students. The correlation is -0.5367776383854018 and the p value is 2.3185881975435266e-40. The correlation is negative which means the higher the rating, the less likely people would take the class again and vice versa. A p-value of 2.3185881975435266e-40 which is less than 0.05 shows it is statistically significant.



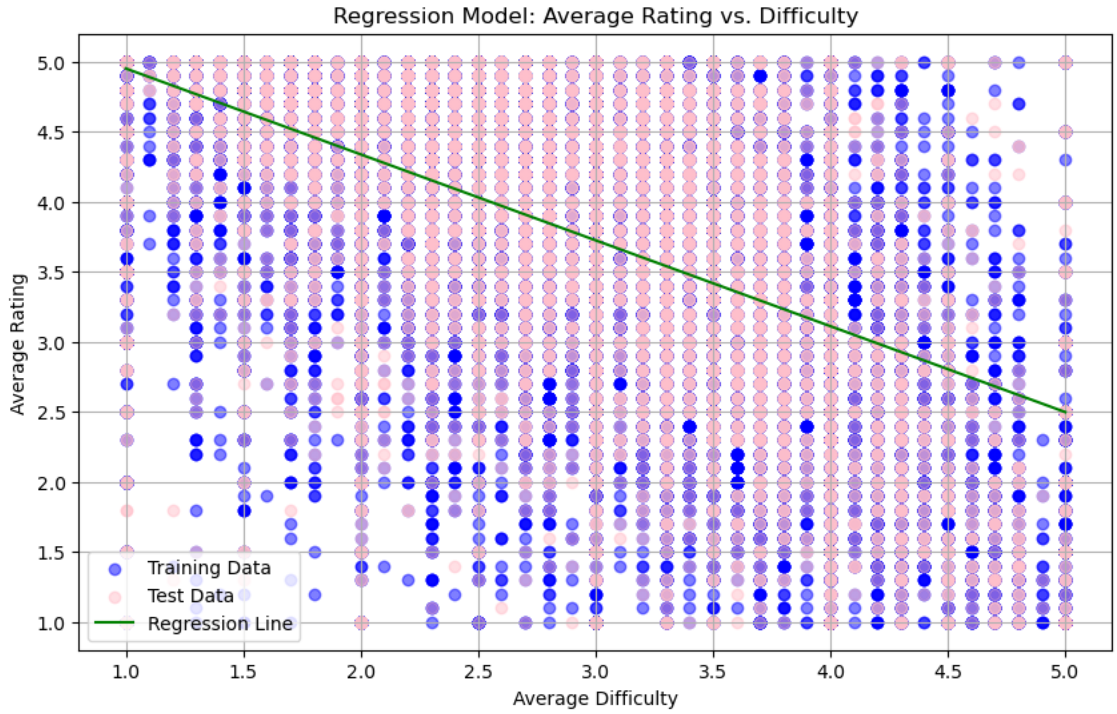
Question 6: Do professors who are “hot” receive higher ratings than those who are not?



I used mann-whitney-u test for this. The u-statistics is 192001608.0, and the p-value is 1.0. Since the p-value is less than 0.05, that means we do not reject the null hypothesis and there is not significant relationship between average ratings and whether the professor have received a pepper.

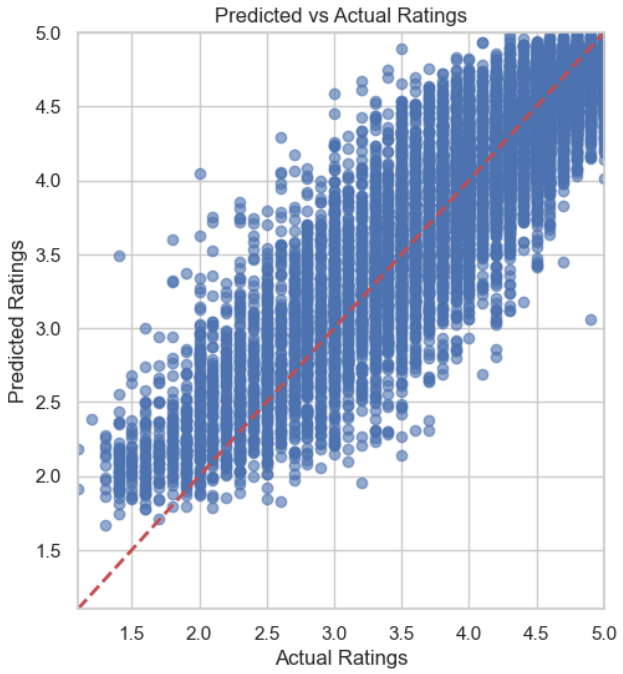
Question 7: Build a regression model predicting average rating from difficulty (only). Make sure to include the R2 and RMSE of this model.

I reshaped the data into a 2D array for it to be consistent with the data shape, and split the data into tests and trains. After fitting it into linear regression, the MSE is 0.92, which indicates that the predicted ratings and actual ratings squared is 0.92.



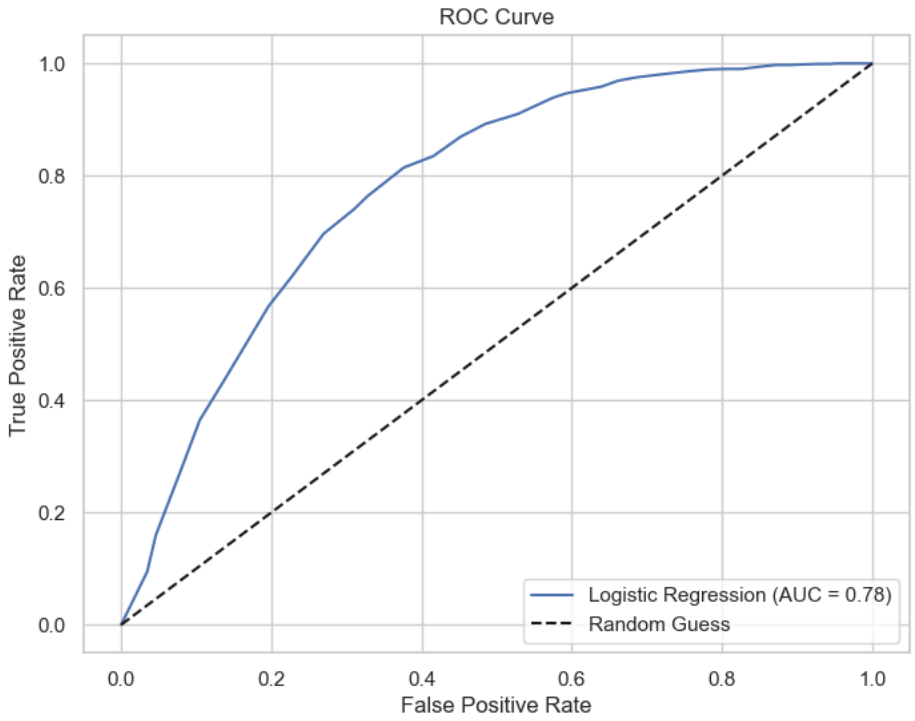
Question 8: Build regression model predicting average rating from all available factors.

For this question, I set X data as all other factors and y as the avg_rating. I fitted the model using OLS model fitting and evaluated the mean square error. The R^2 is 0.81, the RMSE is 0.3687, and the F-statistics is 7406.



Question 9: build a classification model that predicts whether the professors receives pepper from average rating only.

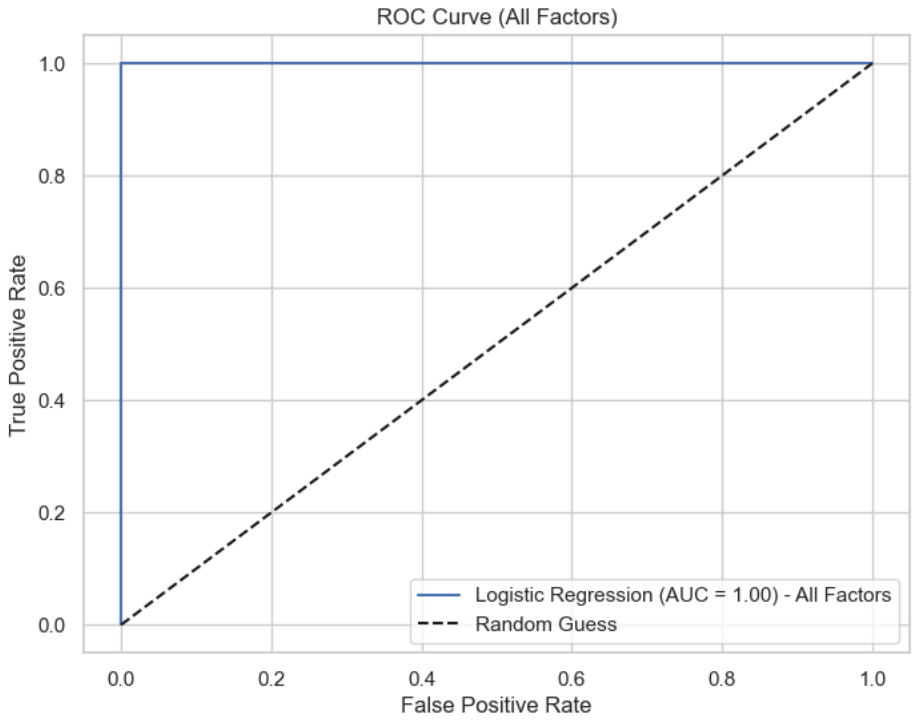
After dropping the null datas, I set X1 as the average rating and y1 as pepper. I split the data into test and train and used SMOTE to balance out the class distribution. I fitted the data into logistic regression and the calculated the AUC. 0.0 represent the negative class and 1.0 represent positive, which is when the professor received a pepper. The 0.0 class has a precision of 0.77, recall of 0.68, f1-score of 0.72, and support of 1299. The 1.0 class has precision of 0.68, recall of 0.77, f1-score of 0.72, and support of 1133



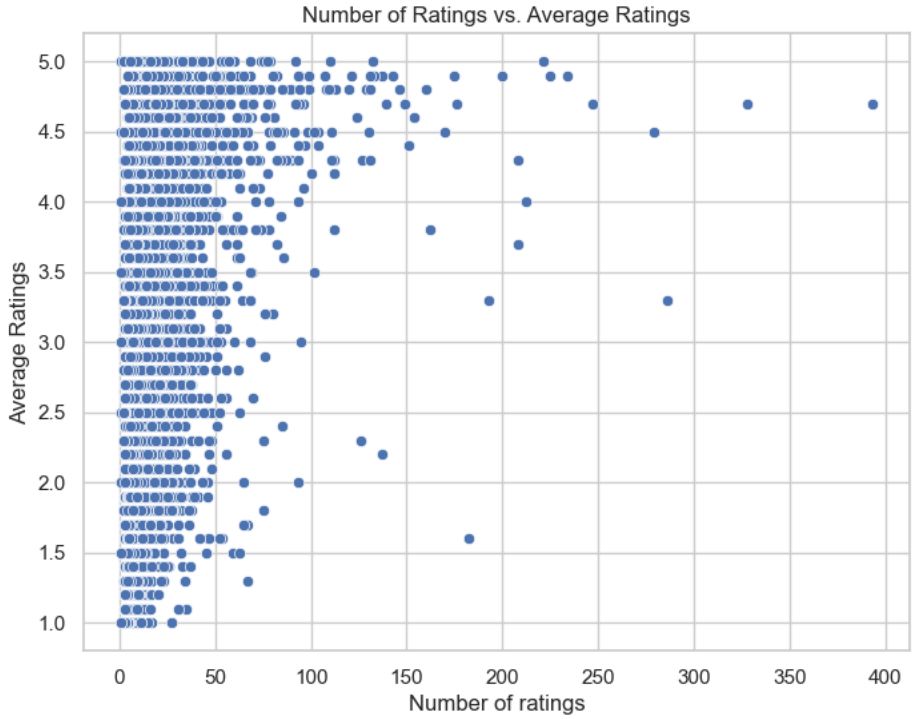
.

Question 10.

build a classification model that predicts whether the professors receives pepper from all factors. For 0.0m negative class, the precision is 1.00, recall is 1.00, fl-score is 1.00, and support is 1299. For 1.0 class, precision is 1.00, recall is 1.00, fl-score is 1.00, and support is 1133. The accuracy of the model is 1.00, which means it correctly predicts all the instances in the test set. Macro and weighted average are also 1.00. The models presents a high accuracy for prediction. Compared to the average rating only, the all factor model has a higher accuracy in prediction.



Extra Credit: Is there a relationship between average ratings and number of ratings?



Using pearsons coefficient, the correlation is 0.0373600793718758 and the p value is 4.688152363959901e-23, showing there is a slight correlation and that we can reject the null hypothesis.