As mentioned in the methodology section, the linear regression backwards elimination model was used to find the features that best predicted hotel scores. A p-value less than 0.05 was the threshold for finding these features significant. These features were found to be ‘Dummy Pool’, ‘Dummy Free Internet’, ‘Saturday’, ‘China’, ‘Business’, and ‘Excalibur Hotel and Casino’, as previously mentioned. This is a worthwhile result because the basis of the remaining results and scores are contingent upon these findings. Part of the goal of this study was to find the best prediction features from TripAdvisor reviews, which were not dependent upon a single categorical predictor.

The three models with the highest accuracy, F1 score, and mean absolute error (MAE) values were the decision tree, random forest, and linear regression models. These models were run on the original imbalanced data, an oversampled dataset, and a weighted data set. Each balance of the dataset used was modelled from the same split of the original data.

For the imbalanced dataset, the decision tree (Figure 3) had a cross-validated accuracy score of 0.503, a cross-validated F1 score of 0.25, and a MAE value of 0.729. For the oversampled dataset, the decision tree model had a cross-validated accuracy score of 0.397, a cross-validated F1 score of 0.399, and a MAE of 1.517. For the weighted dataset, the decision tree model had a cross-validated accuracy of 0.481, a cross-validated F1 score of 0.409, and a MAE of 0.729. A visual comparison of these metrics can be seen in Table 1. For the decision tree model of the imbalanced dataset, very few scores of two and three were predicted, and no scores of one were predicted. For this same model of the oversampled dataset, there were values of one and three predicted, but not two. None of the lower predicted hotel score values were accurately predicted for the oversampled dataset. For the weighted dataset, the hotel score predictions had

correctly forecast values of two, three, four, and five, and incorrectly predicted some scores of one.

For the random forest model, the imbalanced dataset had a cross-validated accuracy score of 0.491, a cross-validated F1 score of 0.399, and a MAE of 0.746. For the oversampled dataset, the random forest model had a cross-validated accuracy score of 0.404, a cross-validated F1 score of 0.335, and a MAE of 1.472. The weighted dataset had a cross-validated accuracy of 0.481, a cross-validated F1 score of 0.412, and a MAE of 0.893. A visual comparison of these metrics can be seen in Table 2. For the imbalanced dataset random forest model, again there were no hotel scores of one predicted, and very few of two or three predicted (Figure 4). With the oversampled dataset, there were many correct predictions of fives, but very few correctly predicted four values, and no values predicted correctly lower than four. There were attempted predictions of one and three, similar to the decision tree model, but no predictions of scores of two, and none of the predictions for scores of one or three were correct. For the weighted data set, values had been predicted for all five possible scores, but scores of one and three were not accurately predicted (Figure 5).

For the linear regression model, only the MAE was calculated for the imbalanced and oversampled data sets. The imbalanced dataset had a MAE of 0.759, and the oversampled data set had a MAE of 1.592. The weighted data set was unable to have a MAE calculated for the linear regression model because the code cell was unable to be run. This is due to the code used not being able to use the weighted data to run a linear regression model. All of the predicted values were predicted to be decimals for the imbalanced and oversampled datasets. The linear regression model for the imbalanced dataset again only predicted a few values of scores below four. For the oversampled data, the linear regression similarly predicted mostly higher values,

and not always correctly. For the weighted data, the linear regression model did not work very well, and no predictions were able to be made.

One issue that was prominent throughout the course of the study was that models did not frequently predict hotel scores lower than four correctly, especially when using the imbalanced data. The imbalanced dataset tended to not predict lower values at all, whereas the weighted or oversampled datasets tended to just inaccurately predict lower scores.

Of the three main models used, the random forest model tended to have the highest cross-validated accuracy scores for all three balances of data. The imbalanced dataset had a score of 0.491, while the weighted data set had a score of 0.481. These two values are close enough that the difference is not very significant, and thus both the imbalanced and weighted data sets could be used interchangeably to predict hotel scores with similar accuracy results.

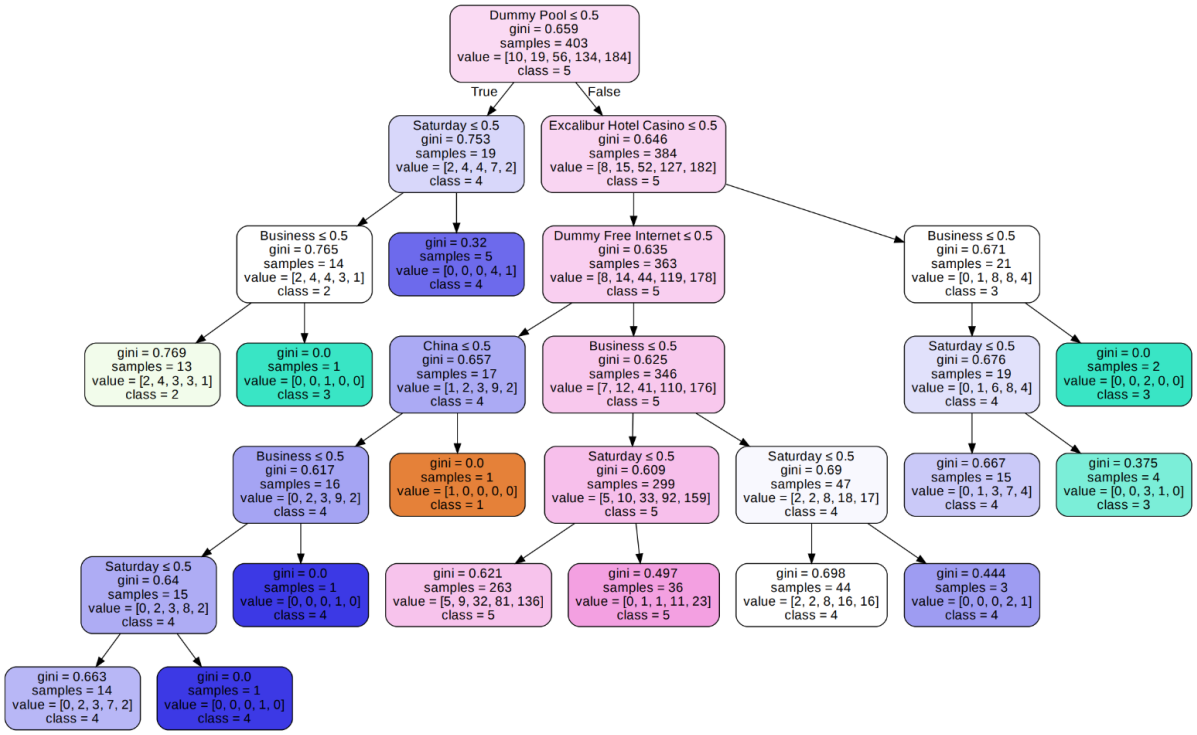

Figure 3. A decision tree using the features ‘Dummy Pool’, ‘Saturday’, ‘Excalibur Hotel and Casino’, ‘Business’, ‘Dummy Free Internet’, and ‘China’ attempts to predict the TripAdvisor score given by users. It is to be read as a flow chart, following a true or false pattern that leads to a prediction about the TripAdvisor score, following the arrows that lead to a true statement for each data point.

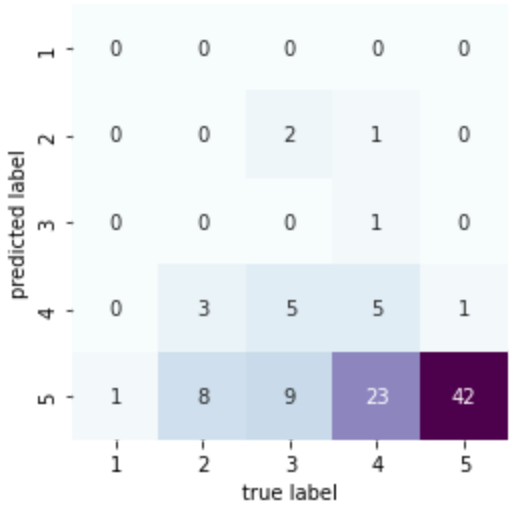

Figure 4. A confusion matrix for the random forest model of the imbalanced data set shows there are very few scores lower than four being predicted. The bottom axis shows the true value of the score being predicted, whereas the left side axis shows the predicted value of the score. A highly accurate model has highlighted colors along the left diagonal, in boxes where predicted and true scores meet. A confusion matrix can also show how many values are being predicted for each score value, which can be valuable when ensuring the model is running correctly.

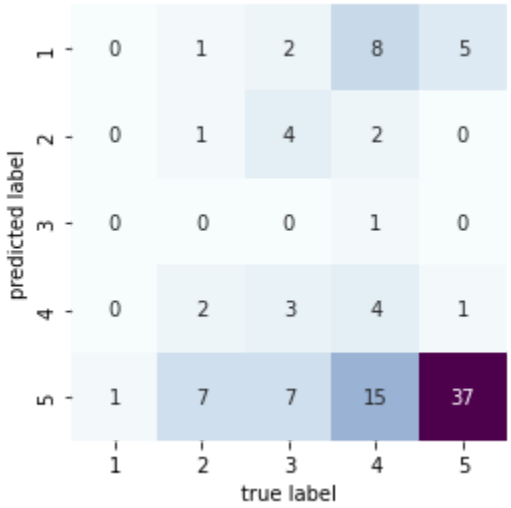

Figure 5. Same as Figure 3, but this confusion matrix is for a weighted random forest model.

	Original Data	Oversampled Data	Weighted Data
Accuracy	0.503	0.397	0.481
F1 Score	0.25	0.399	0.409
Mean Absolute Error	0.729	1.517	0.729

Table 1. The accuracy, F1 score, and mean absolute error are shown for the decision tree classification model for the original dataset, the oversampled dataset, and the weighted dataset.

	Original Data	Oversampled Data	Weighted Data
Accuracy	0.491	0.404	0.481
F1 Score	0.399	0.335	0.412
Mean Absolute Error	0.746	1.472	0.893

Table 2. The accuracy, F1 score, and mean absolute error are shown for the random forest classification model for the original dataset, the oversampled dataset, and the weighted dataset.