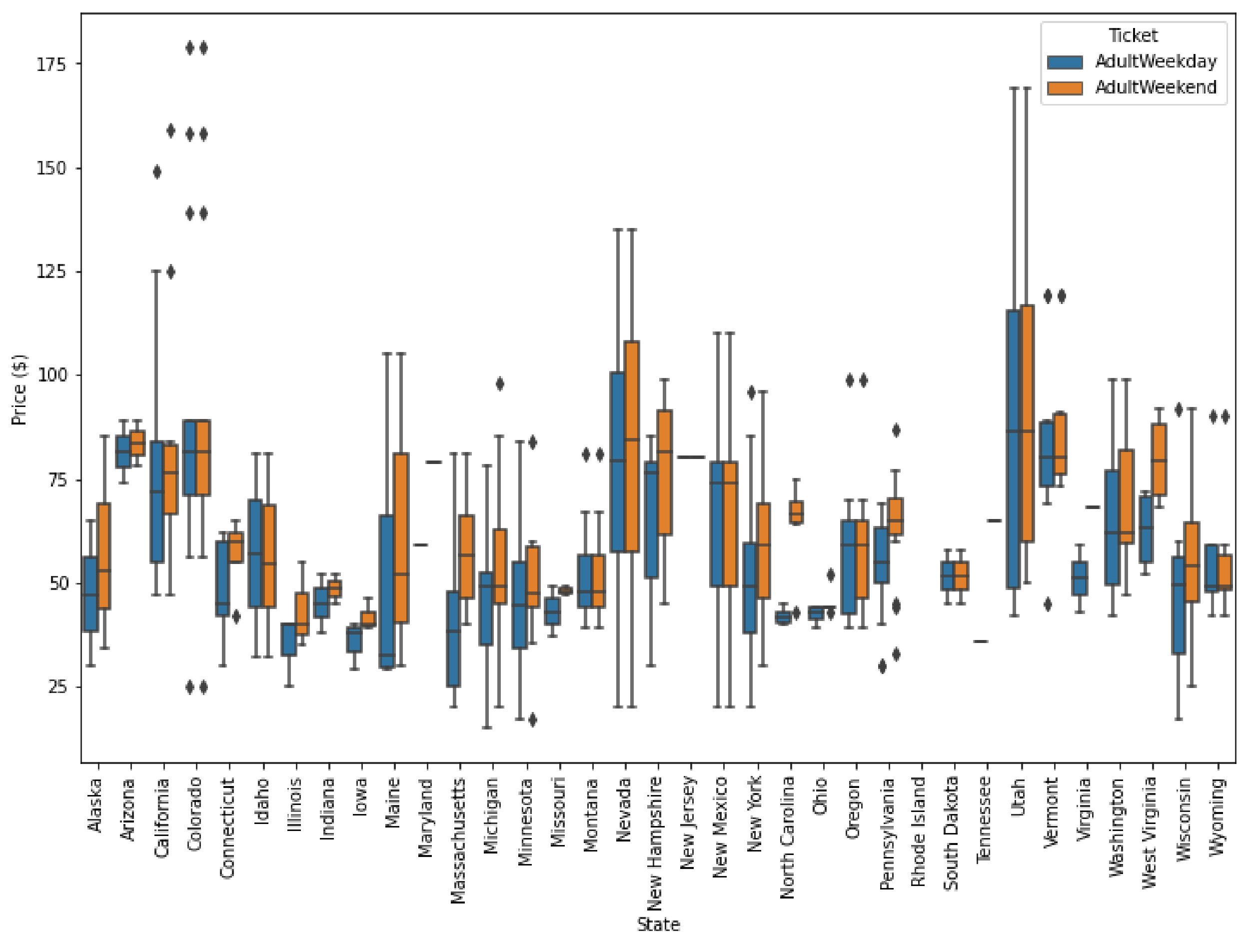
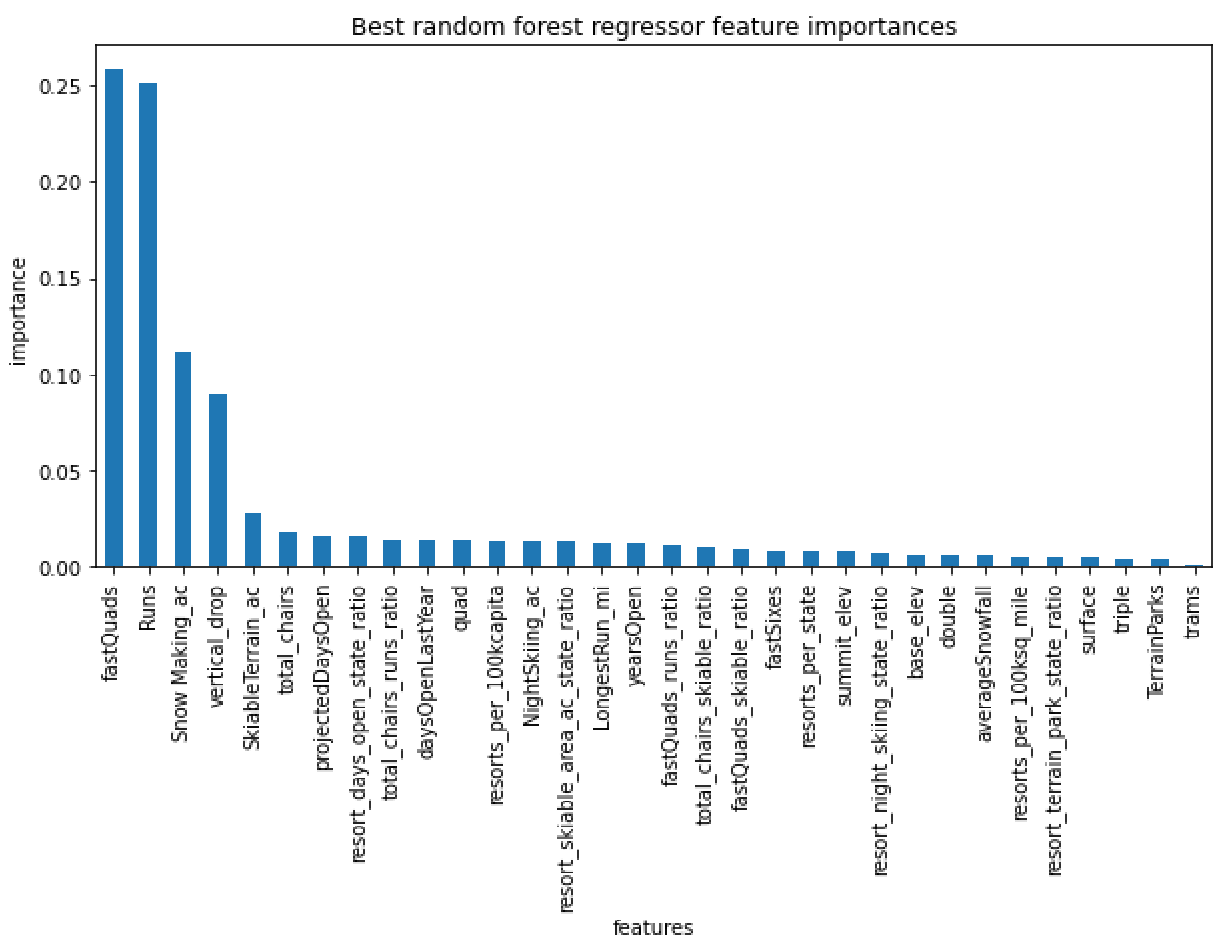
The purpose of this project is to determine how Big Mountain Resort can use a data driven approach to select an appropriate ticket price and better capitalize on their facilities. The dataset used in the project contained 330 rows and 27 columns, with each row being a particular ski resort, and each column representing a specific feature such as name, state, prices, or information about the facilities. The project also incorporated basic data on states such as population and area. The analysis explores how different features affect ticket price. The finished product is a machine learning model that takes features of a ski resort as its parameters and returns an expected ticket price. This model is used to calculate a price for Big Mountain resort, which turns out to be considerably higher than the current price. The model also allows the user to change the values for different features, and see how this affects the modelled ticket price.

The plot below shows that average price varies quite a bit from state to state, as well as the spread of the prices, and the difference between weekday and weekend prices. Montana has no average difference between weekday and weekend prices, which suggests that it is not terribly important which of these we choose as our target value.

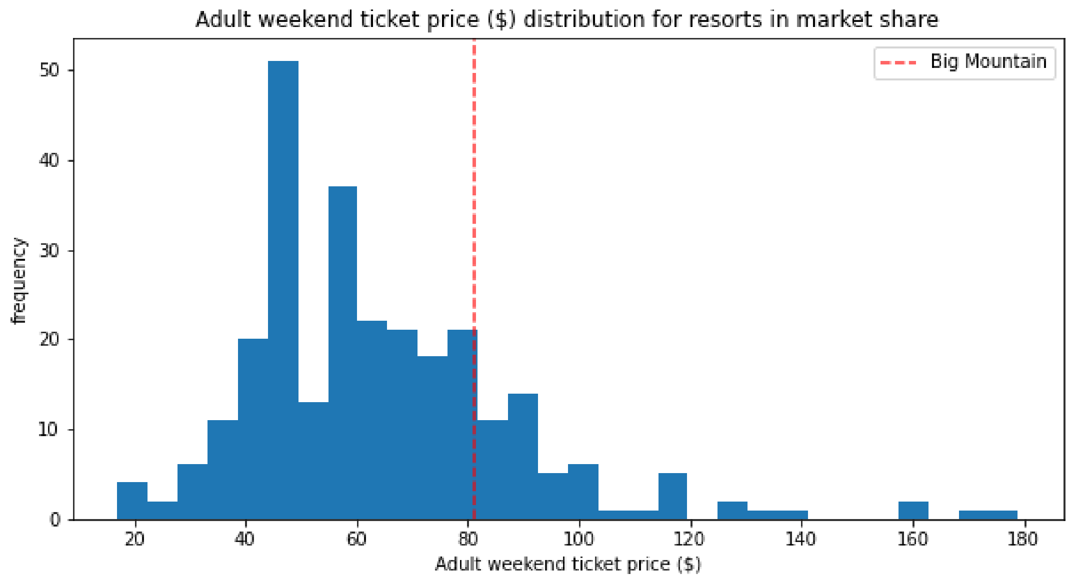


Ultimately weekend price is a better choice for target value since it has fewer missing values than weekday price. Also, there does not seem to be a clear pattern between state and ticket price, so it is better to treat states equally. State data can however be used to create new features which show the ratios of certain features accounted for by each resort. For example, we first calculate total skiable area by state, and then use that to determine what portion of the state's total skiable area belongs to a given resort in that state. Further analysis shows that the features that correlate most with price (AdultWeekend) are fastQuads, Runs, Snow Making_ac, and total_chairs.

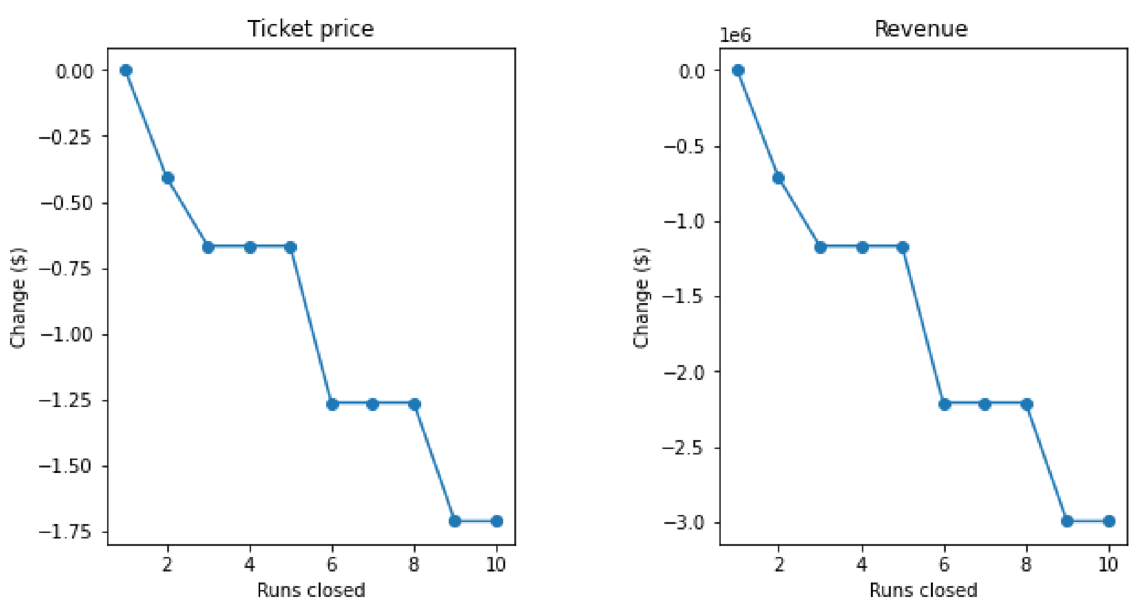
In the machine learning phase of the project, the data is split into training and test sets, with 70% and 30% of the data respectively. The AdultWeekend column is separated out as the target value. We create two different models to choose from, one using Linear Regression and the other a Random Forest Regressor. Both of these models were built into pipelines, trained using cross-validation, and fine tuned with a grid search to find the best values for the parameters. In the end, the Random Forest model returned the best results, with a mean absolute error of about $9.50. This was better than the Linear Regression model by roughly one dollar. The Random Forest was also slightly more consistent in its results, so it is the best model to use going forward. The following plot shows which features this model found to be most important.



For the next step, the model is trained on the entire dataset except for Big Mountain Resort. Using all of the dataset causes the mean absolute error to go up to around $10.40. Finally, when the model is used to calculate the expected ticket price for Big Mountain, it returns a price of $94.22. This is significantly higher than the current price of $81.00, even considering the expected error. Big Mountain is already one of the more expensive resorts in the market, as can be seen in the plot below.



However, Big Mountain also has some of the highest values for features that were given high importance by the model. It is also possible to experiment by changing the values for different features to see how the expected ticket price is affected. The plots below show the expected effects of closing different numbers of runs on both ticket price and total revenue for the season.



The model suggests that Big Mountain could close one run without affecting ticket price. Closing two or three runs causes a decrease in modelled ticket price, while closing four or five causes no additional decrease. This suggests that the resort should either close one run and maintain the same ticket price, or close five runs and decrease the price by around $0.70 per ticket.

The next scenario experimented with adding one run, one chair lift, and 150 feet of vertical drop, and found that this would support a price increase of $1.99 per ticket, which would translate to a $3,474,638 increase in total revenue. Repeating these conditions, but also adding two additional acres covered by snow making machines resulted in the same modelled price, suggesting there would be no benefit from the additional use of snow machines. The final scenario was to add 0.2 miles to the longest run, and increase the area covered by snow machines by four acres. This showed no predicted increase in price and would therefore probably not be advised.

This model can easily be used to explore other scenarios involving adding or removing features to see the effect on the expected price. Of course, the model is not without its limitations. There was some potentially useful information missing from the dataset, such as the average number of visitors for resorts other than Big Mountain, or operating costs of all of the facilities. There is also an implicit assumption that other resorts are setting the correct price given their facilities, or at least that they are correct when averaged together. It is entirely possible that all the resorts are following some conventional wisdom that could turn out to be erroneous. Also, the mean absolute error of the model could be the result of either random differences in price or perhaps additional factors that are not visible in the data.