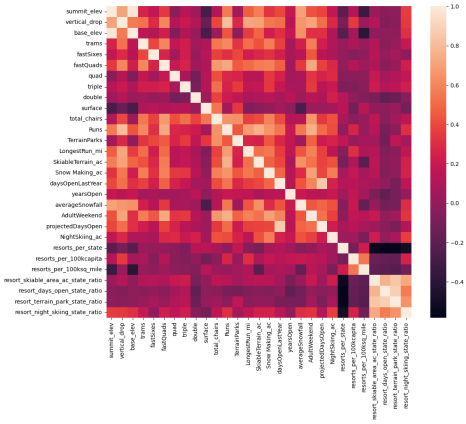
Problem Statement:
What types of changes should Big Mountain Resort make to select a better value for their ticket prices so that their revenue increases by 10% within a year?

Data Wrangling:
We decided on dropping two columns, FastEight and AdultWeekday, because FastEight contained too many missing values and the prices are the same for the weekend and weekday in Montana. We also dropped rows that did not have an AdultWeekend ticket price. Then we made sure that every row had a unique ski resort, which was confirmed. Merged the table that contained information of the state population and square miles to the original table. Right now we have a clean dataset, which we can use to figure out what features of other resorts we can use to base the ticket prices for Big Mountain Resort off of.

Exploratory Data Analysis:


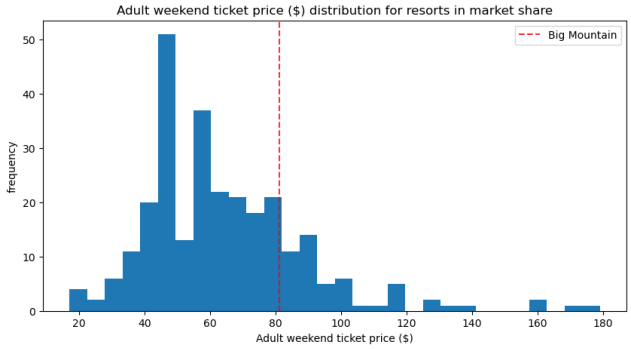
First, we looked at other data, like state population and state sq. miles, to look for any patterns. We did not find any, so we treated each state as the same. Using a heatmap, we tried to see if there were any relationships between the features. It looks like the features that had the highest correlation with AdultWeekend ticket price were fastQuads, Runs, Snow Making_ac, total_chairs, resort_night_skiing_state_ratio, and vertical drop.

Model Preprocessing with Feature Engineering:
First, we partitioned the data into training and testing splits. We then used the mean to see how good of a predictor it is for ticket prices. Then we used R-squared, mean absolute error, and mean squared error as metrics to see how good one set of values agrees with another.

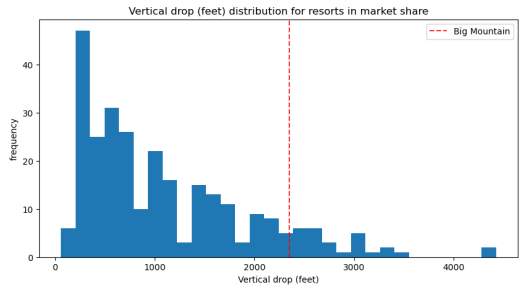
Algorithms used to build the model with evaluation metric:
We first trained a linear regression model. Using cross validation on the linear regression model showed that the mean absolute error was 10.5 on the training data, while on the test data it was 11.8. We then used the Random Forest Model and found that the top 4 features using this model are fastQuads, Runs, Snow Making, and vertical drop. The estimated performance via cross validation for random forest was found to be 9.6 for the mean absolute error which is consistent with the test set.

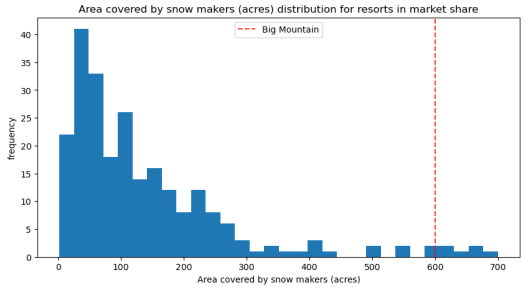
Winning model and scenario modeling
The random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

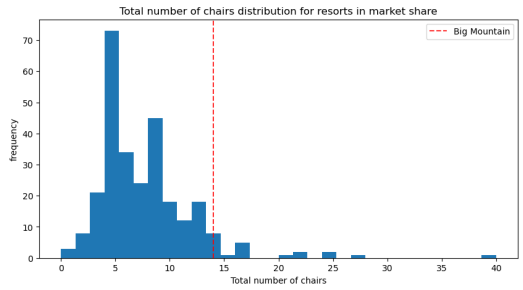
Pricing Recommendation:

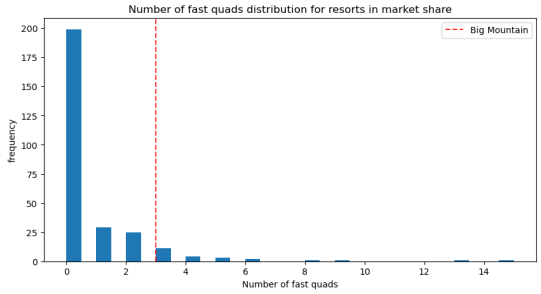


Currently, Big Mountain charges $81, but according to the model the suggested ticket price should be $95. I would tell leadership that Big Mountain Ski resort should increase their ticket prices. In the model, some of the important features that came up were vertical drop, snow making, total chairs, and fast quads. In all of these 4 features Big Mountain is near the top compared to other ski resorts, while their ticket price is not near the top compared to other ski resorts.









Conclusion:

Big Mountain Ski Resort should increase their ticket prices by at least 10%. This is supported by the fact that based on the features Big Mountain has it is not maximizing its returns relative to its position in the market.

Future scope of work:

It would be helpful if we knew more about what the overall maintenance costs are and what the payroll costs are. This way we can use this information and compare it to other ski resorts to see if any adjustments can be made to those costs. In the future we can test a new combination of parameters to see if that has any effect on ticket prices. Also, we can try out other models to see if there is any better predictor for ticket prices.