Big Mountain Pricing Analysis Report

Problem statement

What opportunities exist for Big Mountain Resort to gain $1,540,000 in additional revenue in the upcoming season by a) increasing ticket price and/or b) reducing costs?

Strategies evaluated

1) Raise ticket price.
2) Permanently close up to 10 of the least used runs.
3) Increase the vertical drop by adding a run to a point 150 feet lower down and installing an additional chair lift.
4) Increase vertical drop as well as adding 2 acres of snow making cover.
5) Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres.

Data quality and processing

Raw Data:

Data was provided by Alesha Eisen with information on 330 other resorts. The raw data provided contained 330 rows, representing resorts, and 27 columns, representing variables that describe the resorts.

Changes made to the dataset:

1. Skiable area for Silverton Mountain was corrected to 1819 from 26819 by finding the correct value with an online search.
2. The "fastEight" column was removed due to half the data being missing and all other values being zero
3. Weekend price was selected for the target variable due to having more data available, and the weekday price column was removed.
4. 53 rows were removed: The Pine Knob resort row was removed due to the "yearsOpen" variable value (2019) being a clear error; 52 additional rows were removed because of missing Weekend price data.

The cleaned dataset has 277 rows and 25 columns.

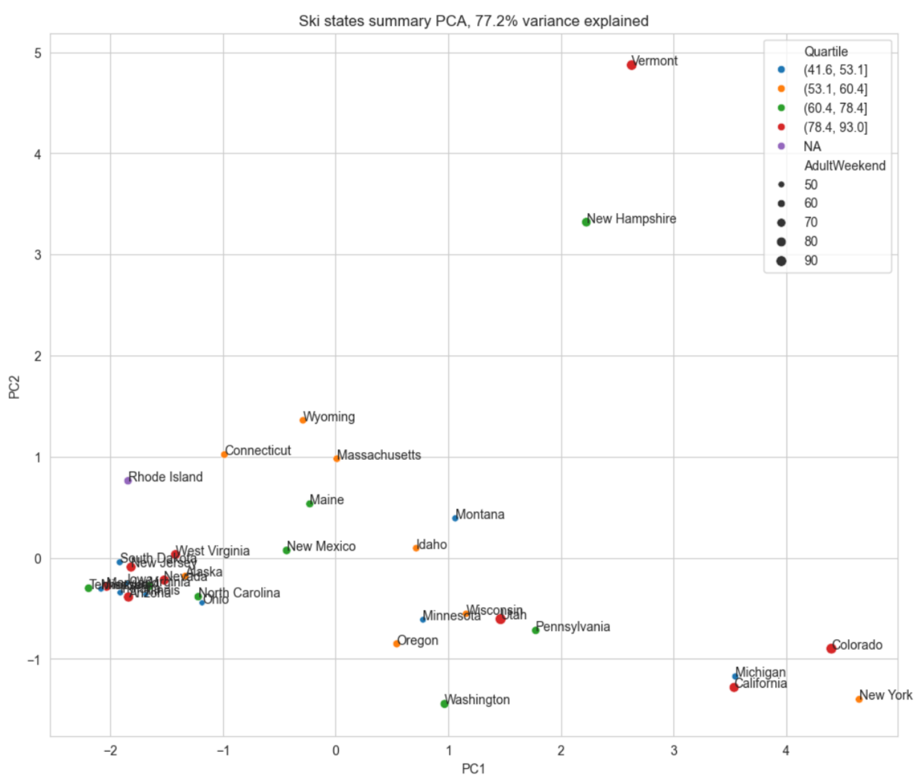
State data added:

Population and area data for states was retrieved from https://simple.wikipedia.org/wiki/List_of_U.S._states for future analyses.

Exploratory data analysis

Dividing data by state (see figure below):

A principal components analysis using all numeric features determined that there was no clear relationship between state and ticket price. Treating states equally currently appears justified in future analyses, since it is not a major predictor of price and will retain a significant amount of data compared to restricting analyses to Montana.



Selecting features to model:

Examination of correlations and scatterplots both suggest the importance of vertical drop, number of Fast Quads, number of runs, and number of chairs. Second priority features that were notably correlated with price are snowmaking area and night skiing ratio.

Model preprocessing

Train-test split:

This analysis used a randomly selected 70% of the data as a training or model fitting set, and the remaining 30% of the data as the test set.

Baseline estimation:

First, to create a baseline comparison for Mean Absolute Error (MSE) between training and test sets, the mean of the training set resulted in a MSE in the test set of a $19.13 average difference from the true values.

Missing data imputation:

The following analyses compared strategies of imputing missing the data using the median and the mean.

Algorithms used

Linear model:

A linear model was created using Ordinary Least Squares regression, with predictors standardized to remove differences in measurement units. Models including all available predictors were compared using either the median or the mean for imputation. Both resulted in similar MAE fit in the test set (median = 9.41; mean = 9.42), therefore the median was used for imputation in the next analyses as a best practice to address skew in several predictors.

To select the number of predictors that would maximize R-squared in a linear regression, cross-validation with 5 folds was used within the training set. After testing numbers of predictors ranging between 1 predictor and all available predictors, R-squared was maximized on average across the cross-validation folds when using only the 8 strongest predictors (R-squared = 0.68).

A linear regression model using the 8 strongest predictors was therefore applied to the complete training set, with predictors in order of importance being Vertical Drop (10.77), Snow Making Area (6.29), Total Chairs (5.79), Fast Quads (5.74) Runs (5.37), Skiable Terrain (-5.25), Trams (-4.14), and Longest Run (0.18). Using an additional 5-fold cross validation to test model fit in the training data, this model had a MAE mean of 10.50 and a MAE standard deviation of 1.62. In the test data, this model had a MAE of 11.79.

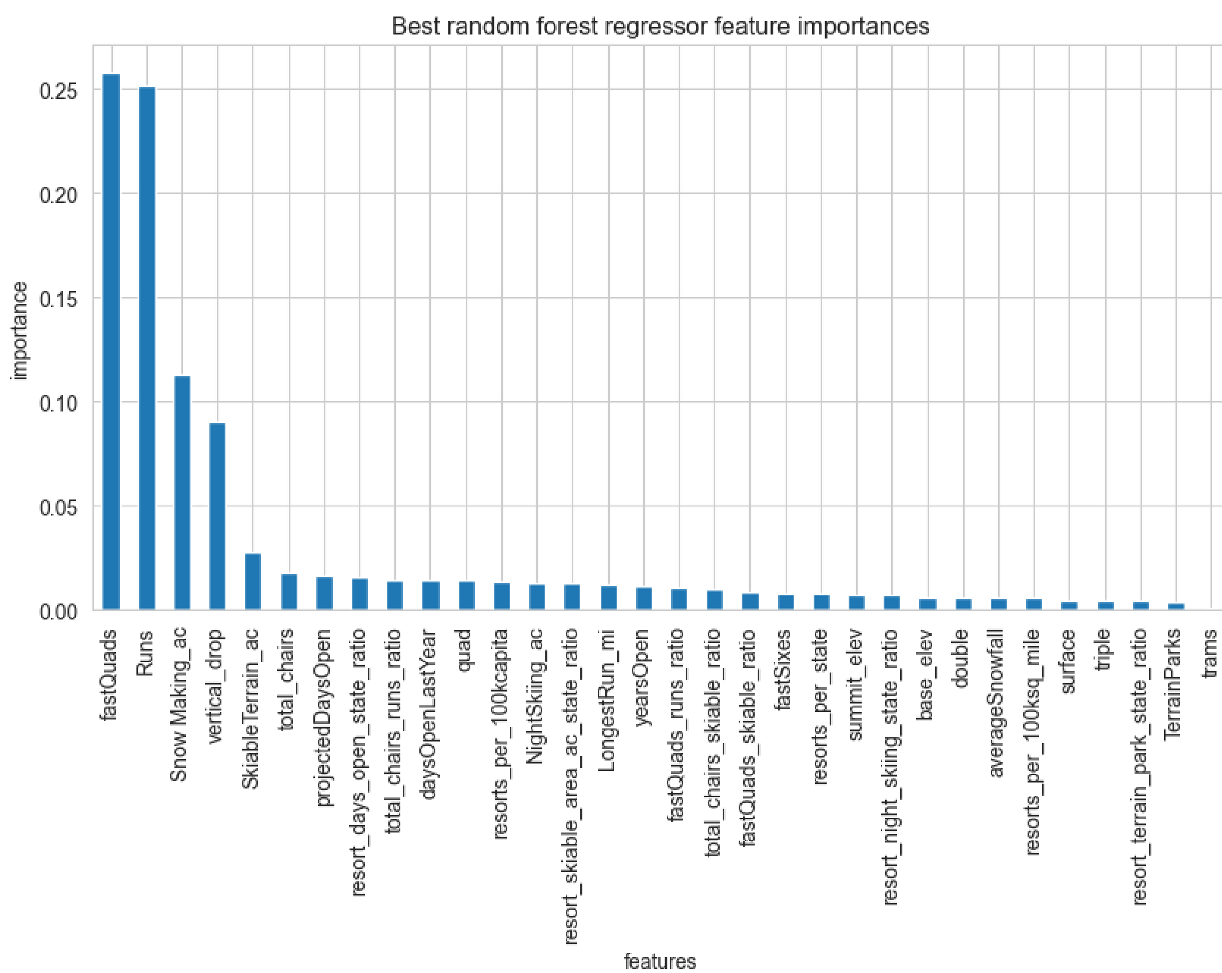
Random forest model:

Next, the Ordinary Least Squares regression model was compared with a Random forest regression model. In the random forest model, both mean and median were compared to impute missing data, and standardized and non-standardized scale for predictors was compared. Using 5-fold cross-validation, the model that maximized average R-squared across the folds (R-squared = 0.71) included 69 estimators, used the median for imputation, and did not standardize predictors.

When applied to the complete training dataset, this model resulted in top predictors in order of importance of Fast Quads, Runs, Snow Making Area, Vertical Drop, and Skiable Terrain. Using an additional 5-fold cross validation to test model fit in the training data, this model had a MAE mean of 9.64 and a MAE standard deviation of 1.35. In the test data, the model had a MAE of 9.54.

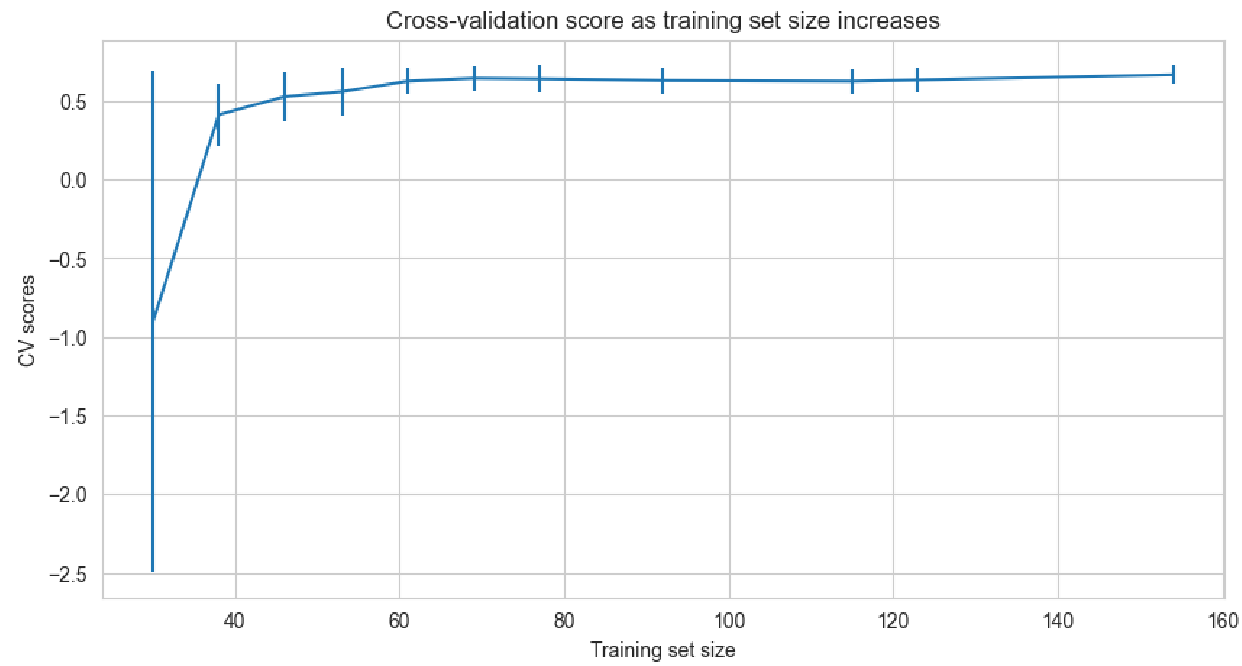
Model selection (see figure below):

Due to a lower MAE in the training data and test data, and lower MAE standard deviation in the training data, the random forest regression model is better fitting.



Data quantity assessment (see figure below):

Finally, data quantity was assessed by testing model fit with different training set sizes. While performance improves up to about a sample size of about 50, change is minimal after this, showing that collecting more data would not likely improve the model.



Scenario modelling

Predicted and actual ticket price:

The final model was refit to all available data, excluding Big Mountain, with 5-fold cross-validation fit measures of MAE mean 10.34 and MAE s.d. of 1.47. The modelled price based on the resort's current features is $95.87, compared to the actual price of $81.00. Assuming that 350,000 visitors each buy 5 tickets per season, for a total of 1,750,000 tickets sold, this price change ($14.87) would increase revenue by $26,023,000 per season.

Effects of proposed operating changes:

1. Permanently closing down up to 10 of the least used runs.

   Closing one run has a negligible negative effect in the model. Closing two would reduce the supported ticket price by $0.41, and closing three through 5 would reduce the supported ticket price by $0.66.

2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift.

This proposal increases support for ticket price by $1.99, increasing total revenue by about $3,475,000. Therefore, after compensating for the $1,540,000 cost of the recently installed lift, this increase would represent about $1,935,000 of overall new revenue. Therefore, this value must be compared with the cost of the new chairlift and renovating the run.

3. Same as number 2 but adding 2 acres of snow making cover.

   This relatively small increase in the snow making area has no positive effect in the model and is therefore not recommended.

4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres.

   This change also has no positive effect in the model and is therefore not recommended.

Pricing recommendation

Summary of proposed operating changes:

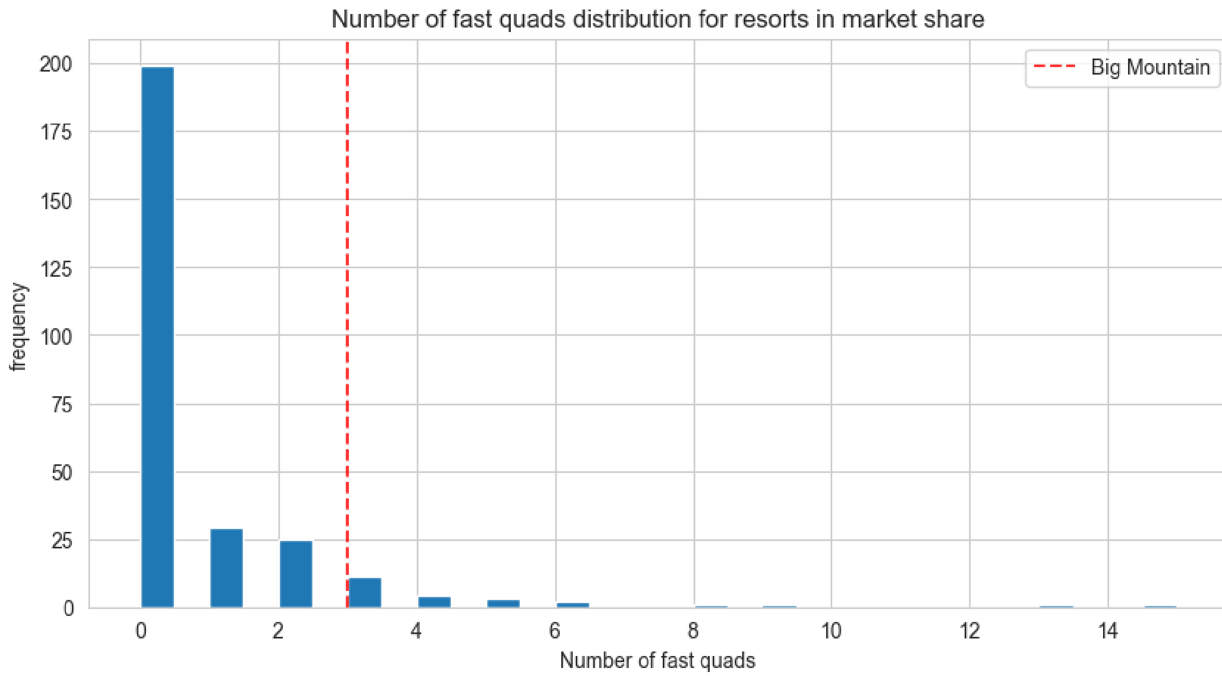
The only recommended operating change is to close the least popular run, which would not negatively affect ticket price. However, savings from this change must be compared with the costs of changing advertising and informational material to alter the number of runs.

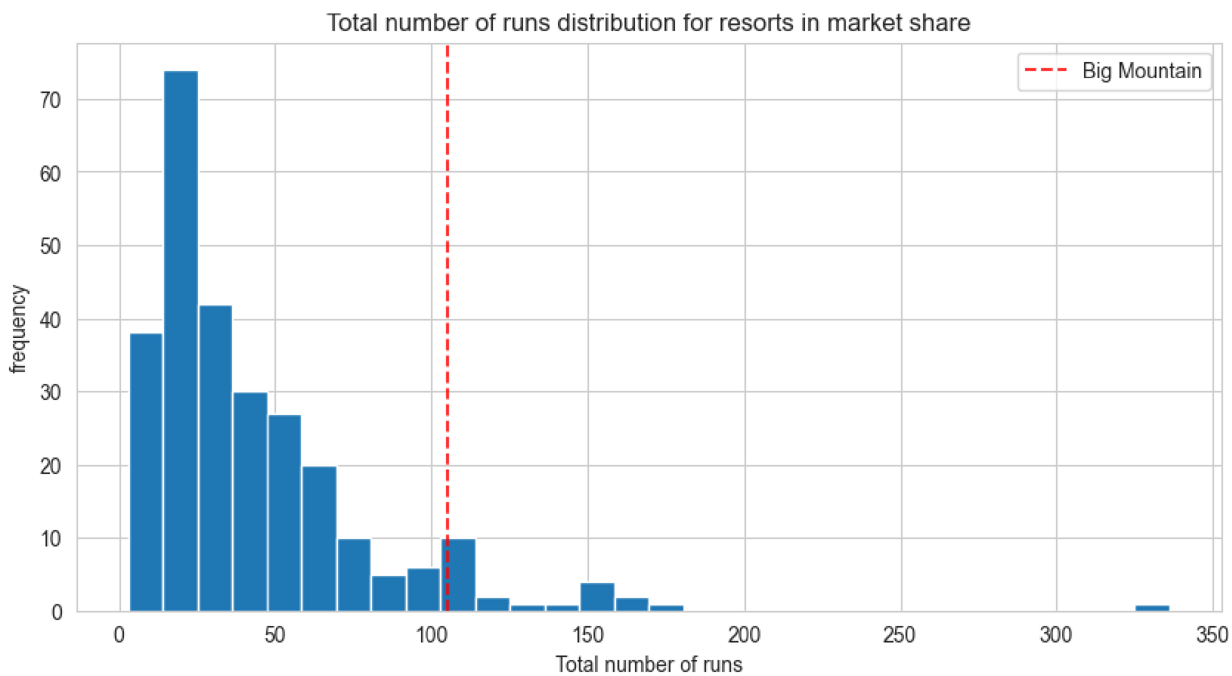
Summary of support for changes in ticket price:

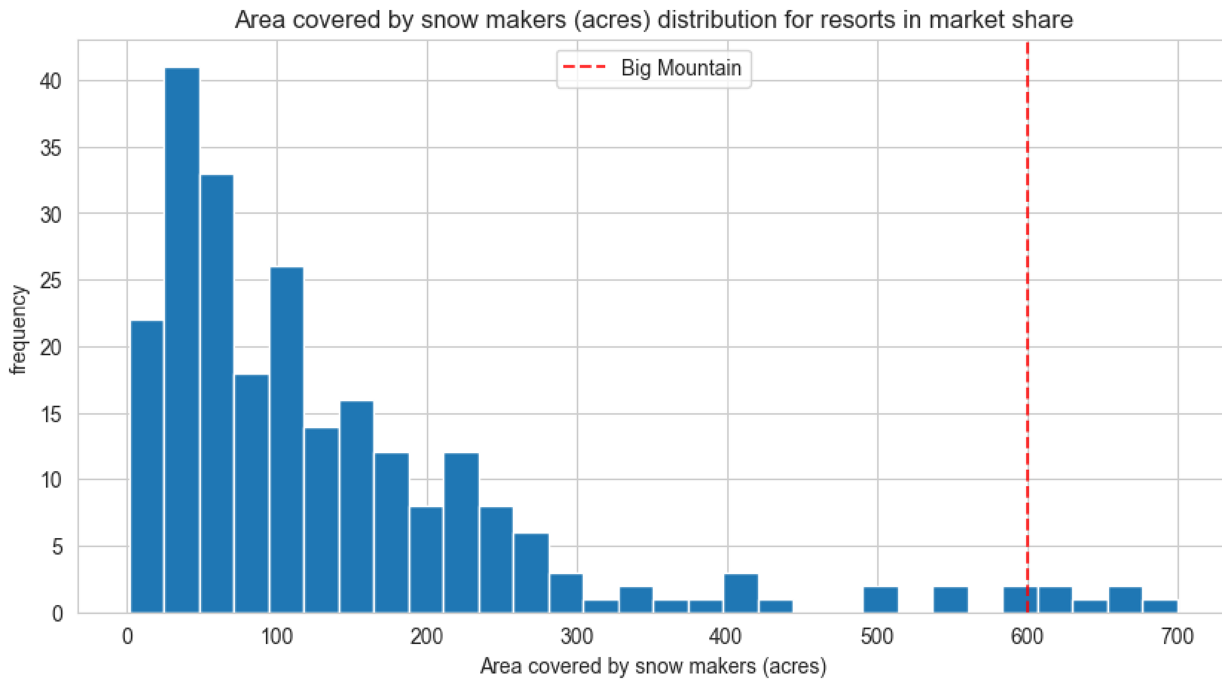
Based on current features with no operating changes, the model supports raising the ticket price to about $96.

Conclusion

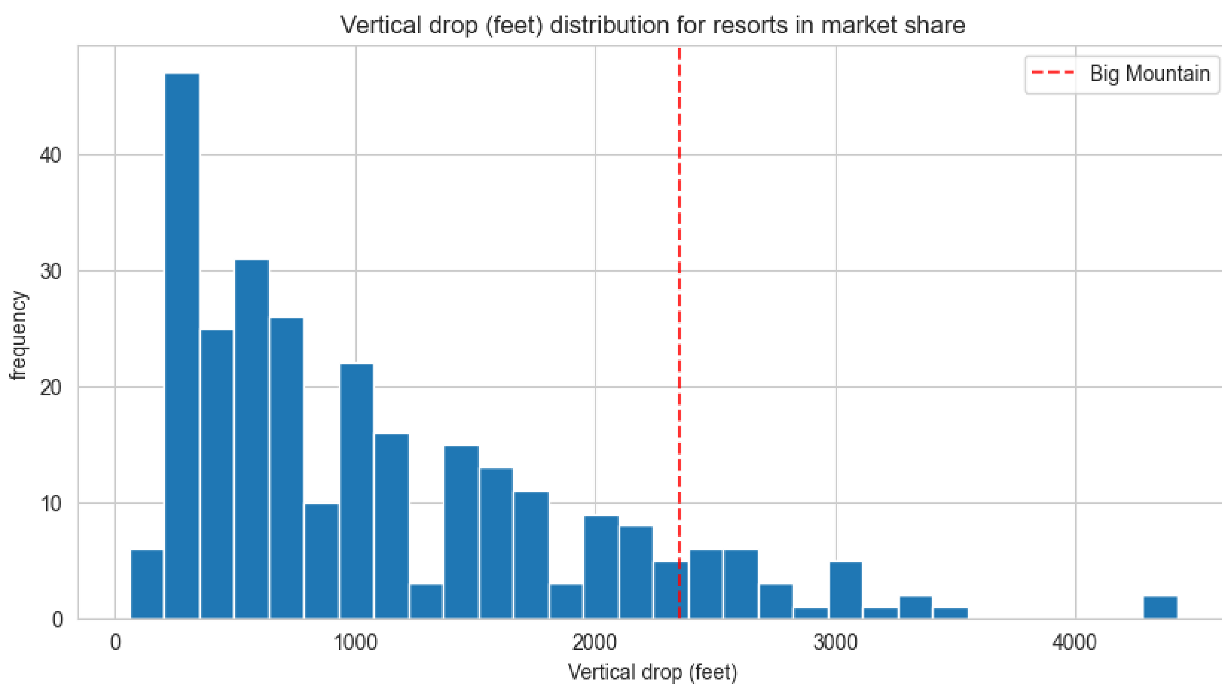
The suggested pricing change from $81 to $96 would increase revenue by $26,023,000 per season. This strategy is supported because Big Mountain is already at the extreme upper range of the four features that were notably the most important predictors in the pricing model: Fast Quads, number of runs, snowmaking area, and vertical drop (see figures below). Highlighting these features in advertising material is highly recommended.


Number of fast quads distribution for resorts in market share
Big Mountain
frequency
Number of fast quads


Total number of runs distribution for resorts in market share
Big Mountain
frequency
Total number of runs



Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
0
5
10
15
20
25
30
35
40
0
100
200
300
400
500
600
700
Area covered by snow makers (acres)



Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
0
1000
2000
3000
4000
Vertical drop (feet)

Future scope of work

Possible data limitations to address to improve the model:

- 52 resorts were missing price data.
- The percentage of missing values per row appear in multiples of 4, suggesting unknown editing of the data.
- Estimated number of visitors was not available for other resorts.
- Area of origin of visitors was not available. Since ticket prices increase with the number of resorts serving a population, this could imply that visitors are unwilling to travel very far, which might affect the success of changes to ticket price.

Additional financial data that could improve the analysis:

- Cost of increasing snowmaking area, which was an additional important feature in the model.
- Savings of closing the least popular runs.
- Cost of the proposed additional new chairlift, altering the run that improves vertical drop.
- Cost of changing advertising material based on changes in runs or highlighting price-linked features.

Future steps:

To allow analysis of further scenarios, the final random forest regression could be included in an interactive dashboard with price, feature, and projected number of tickets as inputs.