Project Report - Big Mountain Resort Pricing Model

Problem Statement

Can Big Mountain Resort (BMR) establish a pricing strategy that accurately reflects the value of our facilities and services without losing our customer base? And can we implement it by July to allow time for marketing to launch campaigns prior to ski season?

Recommendation

Based on Big Mountain's features, our model suggests a price of $95.87. With a mean absolute error of $10.39, I feel BMR has a strong opportunity to improve its pricing model. They should consider the price points of $85-$96, taking into account how bullish they are on demand.

The Model

Our model uses random forest regression, which we assessed with cross-validation. We fine-tuned our parameters using GridCVSearch, finding the median was the best for imputing missing values, features should not be scaled, and 69 was the ideal n_estimator value (trees in the forest).

Before implementing, we refit the model using all of our available records (besides BMR itself). Our cross-validation resulted in a mean absolute error of $10.39, with a standard deviation of $1.47.

In our model, the most important features were number of fast quads (fastQuads), number of runs (Runs), acres covered by snow making (Snow making_ac), and vertical drop (vertical_drop).

The Process

Data Wrangling

Our source data came from a .csv file provided by BMR. We first validated there were no duplicate categorical features before moving on to understand and correct for issues in the data. By plotting histograms for each feature we were able to quickly spot areas of potential concern for further research. To supplement the analysis we created a dataframe of state summary statistics and loaded in population and land area data.

Exploratory Data Analysis

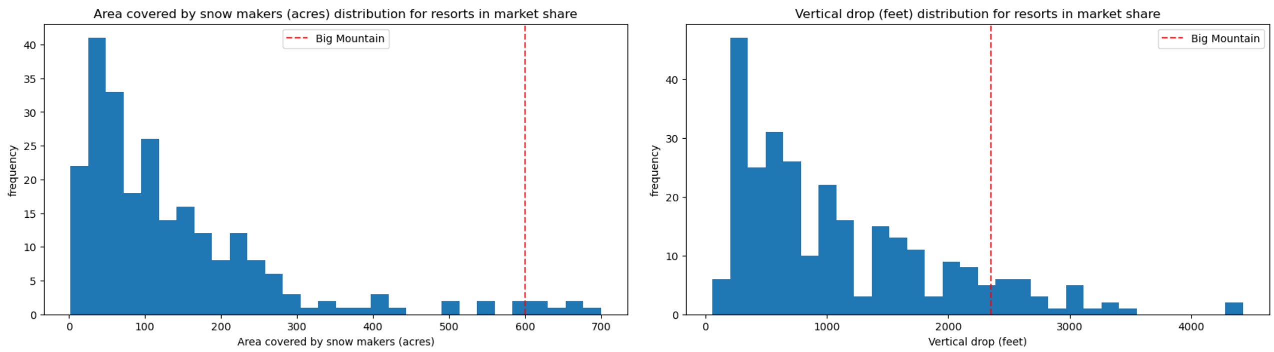
We first explored our state dataframe and decided not to pursue any grouping/categorizing to supplement our data, and build our model off of the whole dataset. Our main tool in this research was a Principle Component Analysis which allowed us to trim the data down to two components to help our understanding. At last, we shifted focus to our feature target, weekend price, using heatmaps and scatter plots to visually assess correlation.

Pre-Processing & Training

We split our data into 70% training and 30% testing and started with a linear regression. The first pass showed some promise but also alluded to overfitting. By refining and comparing results of different models, we settle on the random forest regression model highlighted above. The last remaining question before modeling is if we had enough data. With the learning_curve() function we could see adding more records would not add any value.

Modeling

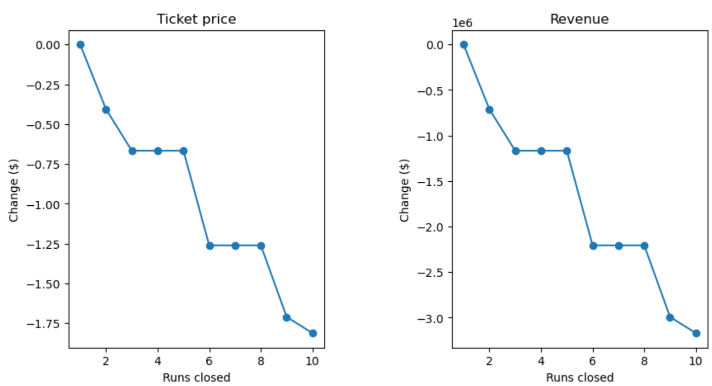
As stated above, our model recommends increased pricing due to BMR's quality of features compared to the market. Two of their standout features are visualized here:



Future State

Scenarios Modeling

Of the scenarios presented, BMR should further explore closing runs or increasing their vertical drop with another run and lift. There is a balance to this, and they should carefully weigh the cost with the expected price/revenue impact. For closing runs, they could close one without impacting their suggested price, but further changes start to increase in a step fashion:



Conclusion

Further refinements to the model are possible. It would be ideal to set up access to the model in a format that doesn't require python knowledge to give BMR's business analysts access to make projections. On the data side, expanding our market data to include visitors at each resort could be a worthwhile investment to continue to fine-tune the model. But even without this, BMR has a strong reason to believe it should adjust its pricing strategy to better align with the value of its facilities.