Guided Capstone Project Report

John Baselj

Problem Identification

This analysis and modeling aimed to evaluate how Big Mountain resort, a ski resort in Montana with 350,000 annual visitors, can optimize their ticket pricing strategy through market comparison to the pricing and facilities of other resorts in the United states. The model was deployed to estimate a competitive price considering Big Mountain's current facilities, and evaluate the expected revenue benefits of suggested facilities upgrades.

Data processing and analysis

The primary data source used for modeling was a database which contained information about 330 US ski resorts including their names and locations, ticket prices, facilities, and size. It was verified that all resort entries were unique, and then severe outliers (such as the skiable terrain area of one resort being 10x the second largest value) were identified, evaluated, and corrected. Columns that contained data for less than 5% of row entries were removed from the data set, such as the "Fast_Eight" column in which only one resort reported having a high speed 8-seat chairlift. Resort entries that did not contain ticket pricing data were removed from the primary dataset because the objective of this modeling is to optimize pricing strategy.

A second data set containing information on the size and population of each US state was merged with the resort data to investigate whether there is significant correlation between resort ticket prices and factors such as population density or resort density within a state, but no significant correlation was found. For this reason, the state and region for each resort entry was removed from the primary data set for modeling.

Resort prices were found to be most closely correlated with number of runs, number of fast quad lifts, and snowmaking ability, and were also positively correlated with total vertical drop, total number of chairs, and days open last season

Model development

A "control" model was developed for model performance comparison, which was simply a model that always predicts the dataset mean for price. A Linear regression model was then applied to a subset of the data chosen for model training, using an sklearn Pipeline. It was found that the most accurate linear regression model could be built using the best 8 parameters for predicting price, with the most important parameters being number of runs, fastQuads, snowmaking area, and vertical drop. Based on cross-validation, the linear regression performed with a mean absolute error of about $12, showing improvement compared to the $18 from the control model.

A random forest model was then applied to the training data subset, and tuned using GridSearchCV to optimize the number of trees and compare mean and median imputation (Figure 1). The improved RF model performed slightly better than the linear regression model, producing a mean absolute error of about $10 and less variability than linear when the model was applied to the test set. This model was selected for use in further pricing scenario modeling.

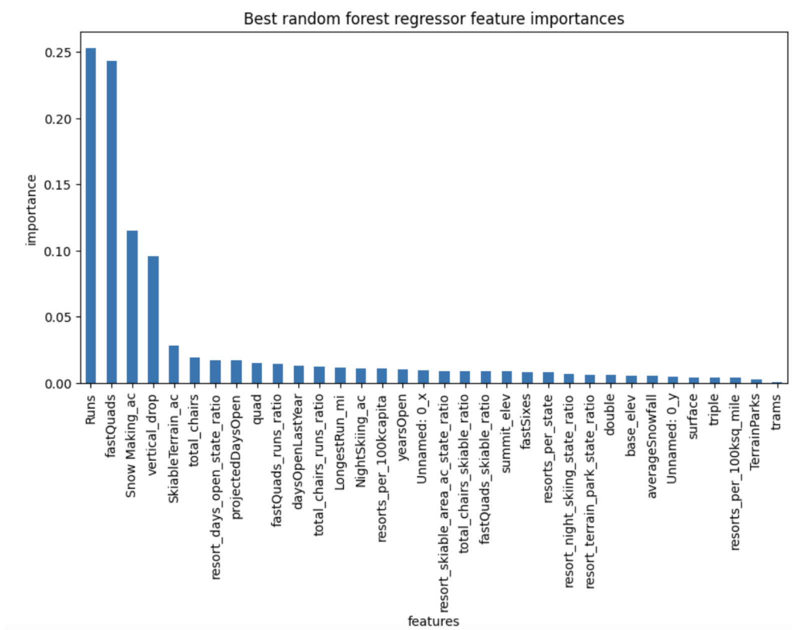

Figure 1: Model Feature Importances

Model deployment and recommendations

The Random Forest model was deployed to predict a market competitive ticket price for Big Mountain Resort's existing facilities of $92.83, compared to the previous price of $81.00. It was then used to model the ticket value impact of 4 proposed facility upgrades which yielded the following:

1. Closing 1 run would not result in a price decrease, closing 2-3 would decrease price significantly, but if at least 3 are closed then closing a 4th or 5th would not harm price (Figure 2).
2. The scenario of adding 1 run, increasing vertical drop by 150 feet, and adding a chair would bring an increase in revenue of about $17MM. This could be a strong investment to suggest for leaders to evaluate against the cost of these upgrades.
3. Adding snowmaking to the added terrain in scenario 2 does not lead to an increased prediction for revenue.
4. It is not expected that increasing the longest run will add to revenue.

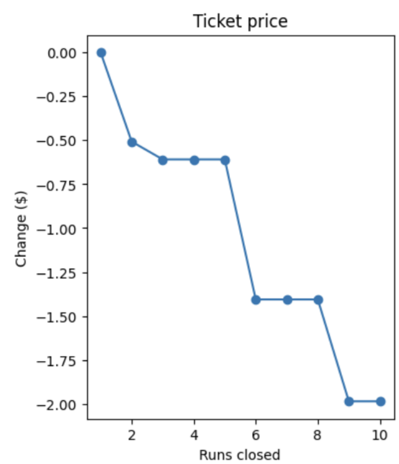

Figure 2: Impact of Removing Runs on Revenue

Conclusion

The modeled ticket price of $92.83 should come as good news to the business leaders at Big Mountain, because this could translate to increase in revenue to the magnitude of $20MM. It will be important for them, however, to utilize experience and domain expertise to consider factors outside of this model when finalizing and deploying a new pricing strategy.

This model can be used moving forward to evaluate future facilities upgrades, however users must be wary that it is trained on a time-snapshot of data and will not be updated with changing market conditions. Additionally, the model can be improved by including additional external factors for pricing such as skier/rider cost of travel, lodging, and equipment.