Clement Chen, Big Mountain Resort Findings Summary

Chart

Description automatically generatedBig Mountain Resort believes that they are not currently pricing their tickets properly and want a more data backed pricing strategy that will reflect the facilities and accommodations at the resort. Big Mountain also wishes to know what changes or additions/removals to facilities would be most appealing to visitors and maximize profits.

Chart, histogram

Description automatically generatedThe data wrangling step included looking for missing values and their distribution within each feature as well as nonsensical values such as a resort being open for 2019 years (Fig 1a and 1b). Each row should represent a unique resort so duplicate keys were checked for. We also imported some state level statistics such as population and size because these factors might be relevant in predicting ticket prices for Big Mountain.

Figure 1a

Figure 1b

In the EDA step, we still weren’t sure if skiing features differed state to state and if these metrics would affect ticket prices, so we merged the external data with internal data and feature engineered new columns such as resorts per 100k people. We also looked at how ticket price correlated with all the features and saw some strong correlations with vertical drop and longest run features.

In the model preprocessing step, we used the simple model of average ticket price to predict future ticket prices. This baseline model is a starting point, and any future model should aim to do better than this. We first tried a linear regression model which performed significantly better than the simple model. We then constructed a pipeline which did all the preprocessing and modeling steps in one line of code. This makes it easier to test different parameters as well as enable future analysts to perform the same steps I took. We used a function called SelectKBest that would select features most relevant to predicting ticket price. We also used k fold cross validation to ensure we weren’t overfitting on the same training data. We then tried using a different model using a random forest regressor which showed a little more promise with slightly better scoring over our linear model. The random forest selected best features that coincided with the best features from the linear regressor so we might be onto something (Figure 2).

To evaluate these models, we used several metrics: R^2, mean absolute error, and mean squared error. R^2 tells us how much of the variance in output is due to the features in our model. Mean absolute error tells us how off our model predictions were from actual values, in this case it would be ticket prices. Mean squared error is similar to mean absolute error but since the errors are squared, it punishes a model more from making predictions that are way off from actual values.

Chart, histogram

Description automatically generated

Figure 2

After deciding on using the random forest regressor over our linear regressor, we used our model to predict how certain scenarios would affect ticket prices. We found out that removable of runs didn’t quite reduce expected ticket price linearly. Increasing vertical drop by 150 ft increases support for an increase in ticket price of $2. Similarly adding 2 acres of snow did not seem to affect ticket price. Adding 4 acres of snow along with .2 miles to longest run also did nothing to change ticket price.

Our pricing recommendation is to raise ticket prices by about $10 with the current facilities. Adding new chairlifts or runs or increasing vertical drop will also enable higher ticket prices but without more information on operational costs, a recommendation to add more of these cannot be made.

In conclusion, Big Mountain should be able to price their tickets $10 higher than the current price. Facilities most important in determining ticket price are number of fast quads, number of runs, number of snowmaking acres, and tall vertical drops. Addition to the resort should start with these 4 variables but not before further assessment is made with expected operational cost information.