Author: Jay Laplante

Problem Statement

Big Mountain Resort needs to know which facilities at ski resorts across the country are most valuable to the market—as indicated by competitors’ ticket prices—so that they can streamline their financial strategy and choose a more lucrative ticket price. They need to increase profits by $1,540,000 over the coming season to balance out new operating costs.

Big Mountain Ski Resort

Ticket Price Model & Analysis

# Step 1: Data Wrangling

This data was composed of comparative feature information for a number of ski resorts across the US, including Big Mountain Resort. It started with 330 entries, each a unique resort. By the end of wrangling, it had been reduced to only 277 entries; 47 were dropped because they had no price data, 4 because they were missing Weekend price data, and 2 for missing or unreasonable year data. 2 unnecessary columns were also removed: fastEight was removed because the data was largely incomplete, and AdultWeekday was removed to focus on the more complete (and more applicable) AdultWeekend ticket price data. Finally, population data by state was acquired to add additional context to analyses.

***Target Feature***

*AdultWeekend, which contained adult weekend ticket prices, was chosen as the target feature for future ticket price analysis.*

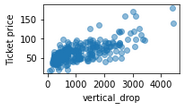
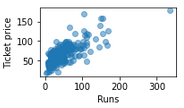
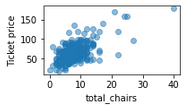
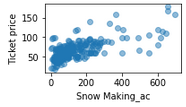
# Step 2: Exploratory Analysis

This exploration started with a look at the numerical state data, finding states which stood out on values like size, population, resorts per state, skiable area, nightskiing, and days open. Resort density (both resorts per 100k capita and resorts per 100k square miles) was calculated to find additional patterns. To better understand how much each feature impacted the states' data variances, a principal components analysis was performed. This analysis showed that resort density (both by area and population) strongly affected the variance between states. However, it showed no strong relationship by state between resort density and ticket price, indicating that all states could be treated equally for the model.

Next, resort data was explored more closely. Ratios, highlighting how individual resorts compare to their state market, were calculated and added for comparison. (These were: skiable area, days open, terrain park count, and night skiing availability.) Vertical drop, counts of resort runs and lifts, and total snow making area all emerged as features with a strong correlation with ticket price.

They also highlighted an important absence in the dataset: though it contains ticket price data, it does not include more holistic measures of park success, such as number of visitors or season profit. This means that it doesn't account for whether a park markets itself as exclusive (charging more per ticket, but not needing as many features) or mass market (charging less per ticket, but perhaps requiring a higher volume of features to serve a higher volume of visitors). It may be impossible to account for this during the modeling process, but it should be kept in mind.

### Scatterplots showing correlation between ticket price and key features.

# Step 3: Model Preprocessing

At this step in the data analysis process, it was finally time to review attributes and determine which ones were the most valuable predictors of price. This involved analysis of the attributes across multiple models—the models were trained on a subset of the data (the training data, approximately 70% of available data), and then tested on their ability to accurately predict values for the remaining data (the test data).

Mean price was assessed as a baseline model. (i.e. Is the mean a good predictor of price? How good?) The coefficient of determination (R2), mean absolute error (MAE), and mean squared error (MSE) were all calculated to measure goodness of fit. The R2 value measures whether the model predicts more closely than the mean, so it returned an expected value of 0 for the training data, with a close but slightly worse accuracy level for the training data. The MAE shows the approximate difference of a model’s prediction from the true value, indicating here that a prediction based purely on the mean would be off by about $19.

Results of R2, MAE, and MSE tests when price is predicted using a **simple average**.

|  |  |  |
| --- | --- | --- |
|  | Training Data | Test Data |
| R2 | 0.0000 | 0.0031 |
| MAE | 17.9235 | 19.1361 |
| MSE | 614.1334 | 581.4365 |

Next, a linear model was built based on all available features. Missing values were imputed using the median (assessment showing no great difference between mean and median.) Data was scaled to a consistent order of magnitude and then modeled using simple linear regression. This performed substantially better than simply using the mean as the predictor; the linear model explained about 80% of the training data’s variation and about 70% of the test data’s. This indicated some overfitting to the training data, but was still a good start. Cross validation was employed to hone the model down to only the most influential features: vertical\_drop, Snow Making\_ac, total\_chairs, and fastQuads.

Results of R2, MAE, and MSE tests when price is predicted using a **linear model**.

|  |  |  |
| --- | --- | --- |
|  | Training Data | Test Data |
| R2 | 0.8178 | 0.7210 |
| MAE | 8.5478 | 9.4070 |
| MSE | 111.8958 | 161.7316 |

The third model generated was a random forest regression model. Again, cross-validation was used during model creation. After testing and comparing different versions of the model, it was determined that imputing with the median worked more effectively than imputing with the mean, and that feature scaling didn’t make a substantial difference. This model indicated that fastQuads, Runs, Snow Making\_ac, and Vertical\_drop were the most influential features, aligning with the findings from the linear model.

In the end, the final linear model scored a MAE of 11.79 on the test set of data, while the random forest scored slightly better at 9.54. Therefore, the **random forest model was chosen to proceed**.

# Pricing Recommendation

Big Mountain currently charges $81.00 for each adult weekend ticket. The model, based on other ski resorts around the US, indicates that Big Mountain’s current facilities and attributes would support a **ticket price of about $95, ± $10**. The most influential factors appear to be vertical drop, snow making area, total number of chairs, total number of runs, longest run, and skiable terrain area.

Given that this would be a substantial increase, we would recommend instituting it gradually, monitoring customer response over time. A conservative starting place would be $85, the lower end of the spectrum. After that, the business might experiment with increased prices in a controlled, testable way; for example, shifting to $90 or $95 tickets for the busiest days of the season. If customers accept these new prices, and if the increase doesn’t harmfully decrease visitor counts, then additional rollout of the new price is worth pursuing.

# Scenario Modeling

We modeled 4 scenarios which would either cut costs or increase revenue. In each case, we explored how that would affect potential ticket price and overall revenue.

Scenario 1: Permanently closing down up to 10 of the least used runs. The model found that closing 1 run would be a safe change, with very little impact on potential ticket price. Closing up to 5 reduces the potential price by a little more than a dollar, and increasing closures after that begin to have substantial impact on the ticket price. If the operating costs of maintaining those runs is very high, it may be worth closing them. However, more specific data on operating costs us needed for that cost-benefit analysis.

Scenario 2: Add 1 additional run and 1 additional chair lift, but no additional snow coverage. This scenario would support an increased price of $1.99, which (assuming consistent expected visitor counts) would amount to $3.4 million over the course of the season. If this is enough of an increase to balance the cost of installing and operating the new run and lift, this scenario should be pursued.

Scenario 3: Add 1 run, 1 lift, 2 additional acres of snow coverage. This scenario returns results identical to scenario 2; such a small change in snow acreage does not significantly impact the ticket price.

Scenario 4: Increase the longest run by 0.2 miles, adding an additional 4 acres of snow coverage. This change has no effect on potential ticket price. The change would be an unnecessary expense.

# Future Work

This pricing model is based exclusively on prices and features, and so can only make recommendations with regard to revenue. With data about installation and operating costs, we could offer a more nuanced set of recommendations, balancing ticket price increase against those costs. For example, information about the cost to install and maintain a run would have been extremely useful in analysis of scenarios 2 and 3. Likewise, understanding the operating costs of a run would help with recommendations around scenario 1. If runs are very expensive to maintain and keep open, it may be an enormous benefit to close some; how many depends entirely on that cost-benefit analysis.

The modeled price may come as a surprise, given how high it is compared to the current price. This is where careful testing and industry knowledge are paramount. There are a number of factors at play for which no data was available, such as expected daily visitors and popularity of other resorts. If there are additional factors which other resorts are using to leverage their price, factors not included in this data, then the model will have been artificially inflated. If stakeholders at the business know of any such factors and can provide additional data, it may be possible to create an updated model. Alternatively, a slow and careful implementation of price changes with a close eye on KPIs will allow the business to learn about their customers real behavior in real time.

If the business found this model useful, we would like to generate an interactive applet with which they could experiment. Perhaps exploring different combinations of features would allow them to find new opportunities for refinement and growth.