

Analysis of Big Mountain Resort Pricing and Investment Strategy

Problem Statement

How can Big Mountain Resort (BMR) increase income by at least \$1.5M in the next season through changes in ticket pricing and/or facilities offerings?

Context

BMR is reevaluating its pricing and investment strategy for the upcoming ski season. It has recently installed an additional chair lift at an additional operating cost of \$1.5M for the season. It would like to know what level of ticket pricing its current facilities can support, and if any changes to its facilities could cut costs or increase revenue through support for higher ticket prices.

Data Wrangling

We built our model using data from 330 US ski resorts. Our data source contained 27 potential features for each resort. However, not all resorts had data for every feature. We dropped some features due to data sparsity and established adult weekend ticket price as our target feature for modelling. Additionally, we used state population and area data from Wikipedia to add features that track the share of statewide skiing assets that each resort captures.

Data was imported from `ski_resort_data.csv`, consisting of 330 rows and 27 columns detailing facilities and pricing data from many ski resorts across the US, including BMR.

The 'fastEight' column was dropped due to sparse data. The 'AdultWeekday' column was dropped, since all resorts in our state charge the same price for weekend and weekday tickets. Row 115 was dropped due to an erroneous value for 'yearsOpen'. After using their facilities data to generate a `state_summary` table, any rows that were missing price data were dropped from `ski_data`, as they are not useful for ticket price modeling.

Some other issues found with the data were:

- An incorrect value for `SkiableTerrain_ac` for Silverton Mountain. Corrected with data found via Google search.
- Heavily skewed distributions in columns 'fastQuads', 'fastSixes', and 'trams'. These columns were kept for now but may have limited use or require transformation in the modeling step.

The final ski_data table has 227 rows and 25 columns. Each row has a value for the target feature, 'AdultWeekend' ticket price. An additional table state_summary, was created with population data from Wikipedia, merged with aggregated ski resort facilities data from ski_data. Distributions for the cleaned features are shown in Figure 1.

Exploratory Data Analysis

First, we examined the state-wide summary data, looking for patterns and key features. We calculated ratios representing the density of resorts in each state by population and area. We scaled the data and applied a PCA transformation to extract the features that contribute most to the variance in the state data. We found that the first two principle components contribute 77% of the variance in the data (see Figure 2), and created a scatter plot visualization of these components with ticket price and quartile determining the point size and color, respectively. This plot showed two groupings split off from the main group of states: CA/CO/MI/NY separated by PC1, and NH/VT separated by PC2. An examination of the PCA components revealed resorts per capita and resorts per sqmi as having particularly large effects on ticket price.

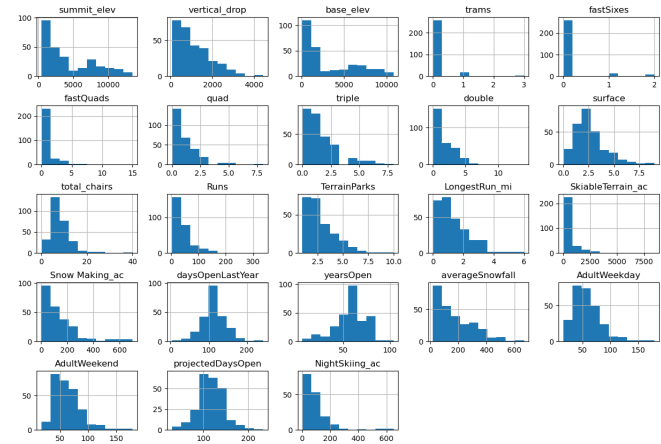


Figure 1

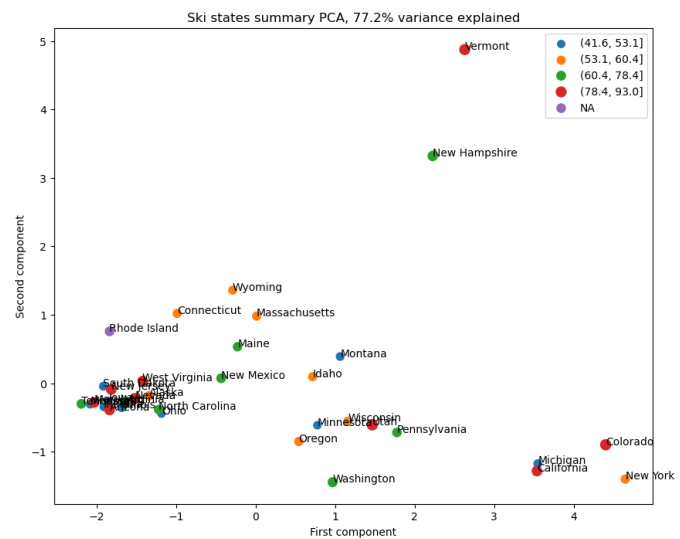


Figure 2

Second, we examined the ski resort data, adding the potentially useful features we generated from the state-wide data. We used these to generate new features representing each resort's share of the marketable ski resources in the state. Note that the validity of this data is dependent on the completeness of the provided ski resort data - if there are resorts missing from the initial data, these omissions could skew the values calculated here.

Our target feature, AdultWeekend ticket price, showed a number of correlations such as fastQuads, Runs, Snow Making_ac, vertical_drop, and resort_night_skiing_state_ratio. Several of the highly correlated features showed clear, useful trends in ticket price, such as vertical_drop and fastQuads. Resorts per capita showed a more complicated relationship with ticket price that could also be useful in future modeling.

Finally, we used the relationship between chairs and ticket price to generate some additional features, taking the ratios of total chairs and fastQuads to the number of runs and skiable area. We found evidence of a potential difference in pricing models between exclusive and mass market

resorts with varying chair/run ratios, but are unable to explore further here due to lack of data on the number of the visitors each resort receives.

Pre-processing and training

We explored several pricing models and evaluated their performance and created a train/test split to use for model evaluation.

For comparison, we began with a naive prediction using the mean ticket price. This produced a mean absolute error of 18.

We then built a linear model. For this model, we imputed missing values using the median and scaled the data using sklearn's StandardScaler (scales each feature to zero mean and unit variance). The linear model with 15 features ($k=15$) decreased the mean absolute error to 9.20 for the training set and 10.50 for the test set. Cross-validation assessment of this model showed some variability in R-squared, suggesting that the model may have been overfit at $k=15$. We performed a grid search for a range of k , finding that $k=8$ showed the best performance with the lowest variability. Some of the top features identified by the model were vertical_drop, Snow Making_ac, total_chairs, and fastQuads. Model performance using best parameters gave a mean absolute error of 10.50 via cross-validation, and 11.79 against the test set.

Next, we built a random forest model. We ran a grid search on several hyperparameters: number of trees, scaling, and imputer strategy. The best performing model used 69 trees, imputing using the median, and no scaling, which showed reasonable cross-validation results. The top features were fastQuads, Runs, Snow Making_ac, and vertical_drop. Model performance using best parameters gave a mean absolute error of 9.64 via cross-validation, and 9.54 against the test set.

Due to the lower cross-validation mean absolute error and variability, we selected the random forest model. An assessment of data quantity showed that the data set is of sufficient size for modeling, and no further data collection is advised.

Modelling

Big Mountain ski resort currently charges \$81 for an adult weekend ticket. Based on our best model (random forest retrained on the full data set minus Big Mountain), the predicted price for Big Mountain is \$95.87, with a mean absolute error of \$10.39. This suggests an increase in ticket price is supported while maintaining current facilities. This analysis assumes that other resorts across the country are pricing their tickets appropriately, and that we haven't left out any key features from our data analysis. It would be prudent to be conservative with any planned price increases and track any negative effects on park attendance.

An additional chair lift was recently installed, increasing operating costs by \$1,540,000 for this season. Given this season's expected visitor number of 350,000, with an average stay of 5 ski days, an increase in ticket price of \$0.88 would cover the additional expenses due to the new lift. An increase of this amount is more than supported by the model.

Several facility improvement scenarios were modeled. The addition of 150 feet of vertical drop via an additional lower chair lift would support a ticket price increase of \$1.99. We recommend further

consideration for this option, which would pay for itself if additional expenses could be kept under \$3.5M per year. Also modeled were a small addition to the snow making area and an increase in the length of the longest run, but these made no difference in the predicted ticket price. We recommend tabling these options.

Closing runs was found to have a non-linear, step-wise effect on predicted ticket price, with predicted price plateauing at 3-5 and 6-8 run closures. Closing 1 run was likewise not predicted to affect ticket price - a trial closure of 1 run could be a good test of the model.

Modelling and Pricing Recommendation

Big Mountain ski resort currently charges \$81 for an adult weekend ticket. Based on our best model (random forest retrained on the full data set minus Big Mountain), **the predicted price for Big Mountain is \$95.87**, with a mean absolute error of \$10.39. **This suggests an increase in ticket price is supported while maintaining current facilities.** This analysis assumes that other resorts across the country are pricing their tickets appropriately, and that we haven't left out any key features from our data analysis. It would be prudent to be conservative with any planned price increases and track any negative effects on park attendance.

An additional chair lift was recently installed, increasing operating costs by \$1,540,000 for this season. Given this season's expected visitor number of 350,000, with an average stay of 5 ski days, an increase in ticket price of \$0.88 would cover the additional expenses due to the new lift. An increase of this amount is more than supported by the model.

Several facility improvement scenarios were modeled. The addition of 150 feet of vertical drop via an additional lower chair lift would support a ticket price increase of \$1.99. We recommend further consideration for this option, which would pay for itself if additional expenses could be kept under \$3.5M per year. Also modeled were a small addition to the snow making area and an increase in the length of the longest run, but these made no difference in the predicted ticket price. We recommend tabling these options.

Closing runs was found to have a non-linear, step-wise effect on predicted ticket price (see Figure 3), with predicted price plateauing at 3-5 and 6-8 run closures. Closing 1 run was likewise not predicted to affect ticket price - a trial closure of 1 run could be a good test of the model.

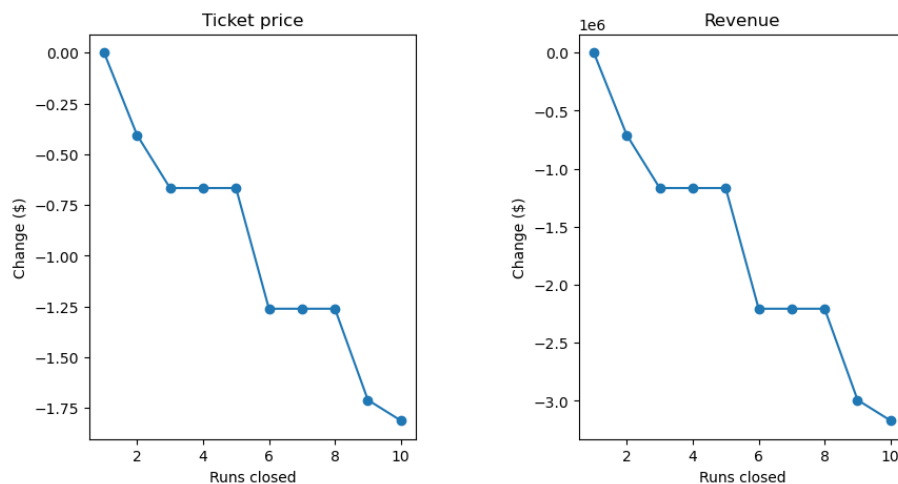


Figure 2

Further work

Additional data sources could improve this analysis, for instance operating cost and demand (attendance) data. If we had data on operating costs for various equipment in Big Mountain, we could calculate the effect of additions or closures on profit rather than just ticket prices. Attendance data for Big Mountain and other parks could be an important feature for predicting ticket price.

The modeling results should be submitted to business leaders for feedback. Any further insight or research they may have on ticket pricing strategies would be welcome in improving on or interpreting the modeling results. If the model is deemed useful, it could be put into production as a dashboard for use by business analysts.