Summary Report for Big Mountain Resort Ticket Pricing Model

The purpose of this model is to solve the following problem. How can Big Mountain Resort leverage data-driven ticket pricing strategies to either raise ticket prices or reduce operating costs for less impactful facilities (supporting a lower ticket price) by determining the importance of features and quantizing their potential for profit with relation to market segment competitors? The context of the problem is that Big Mountain Resort prices its tickets by evaluating the market segment avg. ticket price and charging a premium. Using this strategy completely disconnects the success of ticket sales with the importance of specific facilities. Big Mountain Resort would like to price their tickets in a data-driven fashion, so that the price of tickets is connected to the resort’s features. This would have the effect that the company could cut costs without needing to lower the price or implement features that would support a higher ticket price. Big Mountain Resort recently installed a new chair lift, that will increase operating costs by $1.54mil for the season, so success in this initiative would include meeting either of the two following requirements. The ticket price increases at a rate with which gross sales (calculated from projected number of ticket sales based on 2021) increases by $1.54mil with no projected increase in number sales, or operational costs are cut by removing features, justifying a ticket price that outweighs the new operational cost. In the initial problem statement, I recommended limited the scope of the solution space to geographically related states, however, further analysis indicated that the solution space should be adjusted to account for all states equally. A major constraint in the solution space, now considering all states, is that the success of other resorts is not measured, so the model may be affected by the over- or underpricing of resort’s models.

Chart, bar chart, histogram

Description automatically generatedBuilding an effective model would require wrangling and cleaning of the raw ski resort data. Counting the number of missing values highlighted some immediate issues. The raw ski resort data contained entries for 330 resorts with 27 columns of feature data. Some of these features, particularly number of fast Eights with 166, had many missing values. More importantly, our target feature ticket prices had 51 missing values for weekend price and 54 for weekday price. It was essential to determine which of these rows and columns needed to be drop for the model to be effective. One observation was that the state and region column were the same for nearly every state, and the distributions of number of resorts (“Count”) was nearly identical, as shown below. So, the decision was made to prioritize the state column and categorize the resorts by state only. Grouping resorts by states, a series of box plots was used to visualize the Chart, bar chart

Description automatically generateddistribution of ticket prices across states, shown here. Still no decision has been made regarding how to use the categorical state data for each resort.

Timeline

Description automatically generatedFollowing this observation, our attention was turned to categorical features, and after reviewing feature distributions, certain individual values were identified to be erroneous and were corrected. To further clean these distributions, resorts with missing ticket price data were to be removed, as these resorts would not provide any valuable information for our model. Before removing any resorts though, some state-wide summary statistics were calculated as even resorts with missing ticket data could provide state wide information. The aggregate totals: 'state\_total\_days\_open', 'state\_total\_terrain\_parks', and 'state\_total\_nightskiing\_ac' were collected and added to a `state\_summary` DataFrame. This DataFrame would be used when evaluating how to handle the state feature of resorts in the final model. Then, rows for resorts with no pricing data at all were dropped and the feature distributions were reviewed. There are shown here. These distributions were determined to be acceptable to continue. Some distributions much variance away from 0. These features would be considered with care when determining feature importance in the final model. Population and state area data were then sourced from wikipedia and the data joined to the `state\_summary` DataFrame. To finalize the DataFrame containing resort data, it was determined that since resorts in the state of Montana priced weekend and weekday tickets the same, the model would need not consider two separate values for tickets, and the weekend ticket prices were determined to be retained, and the weekday ticket prices column dropped since it contained more missing values. Next, the a decision needed to made regarding whether to treat all states equally, and the exploratory data analysis step was initiated.

First, in the state summary data, the columns for population and state area were replaced with columns representing the number of resorts per 100k captia and number of resorts per 100k square miles. This was a better indicator of each state’s share of the skiing market. A principal comonents analysis was conducted to “untangle” these 7 state features and look for any notable distinguishing patterns in certain states that may indicate the importance of seperating the market by state in our model. The data in the DataFrame was scaled to 0 mean and unit standard Chart, line chart

Description automatically generateddeviation. The scaled data was fit to a PCA model, which revealed that the first 2 prinicipal compenents described over 75% of the variation. Considering this to be a sufficient proportion to justify reducing the data to 2-dimensions. Each states value for the frist and second component were plotted on a against each other. Mean state ticket price was also segmented into quartiles to add color and size to the individual points on the scatterplot to provide more clarity to the significance of each states position along these two Chart

Description automatically generatedcomponents. The scatterplot is shown here. The spread of the data across quartiles does not show any particular pattern with pricing, offering the final justification that states should be treated as equal in the final model. However, Vermont’s and New Harmphire’s position on PC2 are notable. Evaluating the features of the components confirms something we observed before: Vermont and New Hamsphire are dominant states regarding the number of resorts per 100k capita and the number of resorts per 100k square miles. Now, having evaluated the importance of state features, the state summary data was merged with the resort data, however, the capture the relevance of that data regarding each specific resort, the totals columns for these features were replaced with the proportion of each feature the individual resorts accounted for by state.

Chart

Description automatically generatedTo return to the original business question, we were now interested in determining which features were most highly correlated with ticket price. A heatmap of the correlation matrix is shown to display these relationships. The following features showed promising correlations with ticket price: ‘fastQuads’, ‘Runs’, ‘Snow Making\_ac’, ‘resort\_night\_skiing\_state\_ratio’, ‘total\_charis’, and ‘vertical\_drop’.

The data now needed to be pre-processed for the model. Effectively, a linear regression and random forest pipeline were tested and refined for the model. A 70/30 train/test split was produced, separating adult weekend ticket price as the target feature. Categorical non-numeric features were dropped. This was acceptable because state information was decidedly disregarded in the previous step. Using the `GridSearchCV` and `make\_pipeline` functions of the `sklearn` module to determine the best paramenters of each pipeline and using a 5-fold cross-validation for each, it was determined that the random forrest model had a higher mean cross validation score with lower variance. This was determined, therefore, to be the best model. It’s best parameters were extracted and saved to the variable `best\_model`, which would be used in the next step. The most useful features for predicting ticket prices identified by the model were ‘fastQuads’, ‘Runs’, ‘Snow Making\_ac’, and ‘vertical\_drop’. ‘SkiableTerrain\_ac’ and ‘total\_chairs’ ranked 5th and 6th.

Chart, histogram

Description automatically generatedFor the final modeling step, the model was refit to the entirity of the available data, and the ticket price for Big Mountain Resort was predicted to be roughly $95. This was a significant increase over its current price of $81. The following distribution shows where Big Mountain currently sits in the ticket price distribution. While you might note that Big Mountain sits on the higher end of the distribution. Big Mountain also ranked extremely high in the features that were determined by the model to be highly correlated with higher prices, indicating that this price is not only justified, but that a higher price of $95 would still be justified!

Now that we have a modeled ticket price, we must also consider the alternate strategies proposed of increasing revenue by either cutting costs or adding new features that justify higher ticket prices. These scenarios are:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage.
3. Same as number 2, but adding 2 acres of snow making cover.
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres. To evaluate these revenue increases, a `predict\_increases()` function was created which would return the difference between the updated ticket price predictions and original model price predictions, allowing us to easily extrapolate the predicted venue increase on the assumption that the average guest (350,000 predicted guests) visit the resort 5 times per operating season.

Here are the model predicted summaries of the revenue effect of these proposed strategies.

1. Reducing the number of runs results in a predicted ticket price decrease of maximally $1.80. On the assumption that the average visitor returns to the resort 5 times per season, reducing the cost of removing, say, 10 run did not reduce the cost of operations by by roughly $3.2 mil this season, this would have a negative impact on net profit compared to the model's predicted optimal ticket price. The plot below provides an even more in depth description of this situation. The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop. So, it the company decides to move forward with the strategy of closing runs, it may be optimal to close 6 runs if the reduced operational cost is greater than roughly $1.25 mil or 8 runs if the reduced operational cost is greater than roughly $2.25 mil. Otherwise, it is nearly essential to close exactly 1 run as this is guaranteed to reduce operational cost without any projected revenue decrease.
2. This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3,474,638. Only an increase in operational cost less to this value would justify this.
3. This scenario supported the exact same increase in ticket price. However, it would definitely increase operational cost, so it should almost definitely not be considered.
4. This scenario increased support for ticket price by no difference. This is likely a result of the model, but again, this model was selected to be the best model for a reason, so this scenario should probably not be considered.

I included a retrospective scenario 5. What if the new chair lift recently installed was never installed in the first place. Let’s imagine it was never installed, and see how revenue decreases to determine if it will justify its projected $1.5 mil operational cost this season. It appears that removing the newly installed lift would only account for a revenue decrease of $0.6 mil, but the increased operating cost is forecasted at $1.5 mil. This certainly creates pressure to maximize ticket price increase, and highlights the importance of data-driven decisions in the future. Perhaps we can reevaluate scenarios 2 and 3, without the addition of a new lift to offset this profit loss. This alternative to scenario 2 increases support for ticket price by $1.70. Over the season, this could be expected to amount to $2967391. This alternative to scenario 3 expectedly produces the same results. Once again, the extra 2 acres of snow making capability make no difference in the pricing forecast, but adding a new run and increasing the total drop still yield a healthy revenue increase. This would definitely offset the cost of the recently install lift. Could it be possible to simply modify one of the existing chair lists to accommodate the extra 150 ft of height, without adding an entire new lift?

To summarize, Big Mountain Resort is in a dominant position in the market segment regarding features that support a premium price. With the right marketing, theoretically consumers could quickly adjust to the price increase of roughly $15 per pass if they were made aware of the significance of the resort’s premium features. The need to increase the ticket prices as imminent, as evidenced by the fact that the operational cost of the new lift does not offset the ticket price increase that the feature justifies. I recommend implementing strategy 2 to offset this cost as much as possible. Additionally, if it is possible to modify an existing lift without creating a new lift to serve this run, I would recommend this as it would yield a net profit (rather than simply a loss reduction).