Background

This report summarized data exploration and modeling efforts undertaken at the direction of Big Mountain Resort leadership to explore opportunities to increase profits generated from ticket sales by 5% over the next season, either by identifying drivers to support even higher ticket prices or areas to cut costs without undermining the current ticket price customers are willing to pay. Big Mountain is a premier ski resort located in Montana. Big Mountain has 105 trails, 2,353 feet of vertical drop, and is served by 14 lifts of varying types. Big Mountain has historically charged a premium over the average price of tickets for resorts in its market segment and had made recent lift upgrades to continue to support this premium price - though this has also increased their operating costs.

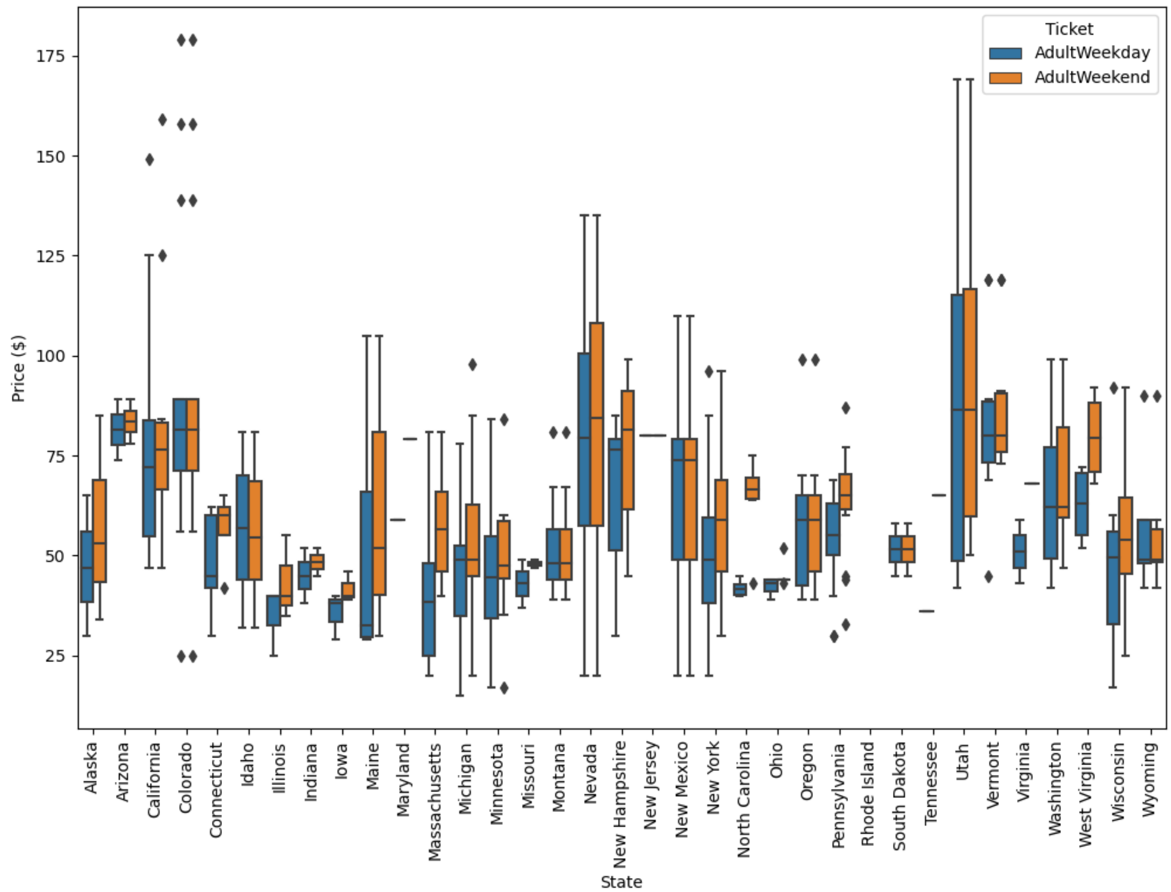
This effort focused specifically on the drivers for ticket price. This meant working to identify the amenities that support a higher ticket price relative to other resorts, or identifying opportunities for cost savings by eliminating/reducing amenities or infrastructure that are not relevant to customers when determining the price they are willing to pay for tickets. The analysis focused on ticket pricing, and the factors which contribute to this value, to drive additional profit.

The key data source for analysis was a .csv file containing information on key characteristics of 330 resorts in the US (including Big Mountain), all considered part of the same market share. Some of the characteristics included information on topography, runs, detailed information on charilifts, ticket prices, and days open per year.

Data Wrangling

At this point the analysis shifted to initial data wrangling, looking at things like missing values, any duplicated values, and relationships between certain value types (for example States and Regions). We had to make some decisions around how to address outlier values for data, including dropping some data due to unreliable values, or updating others.

One thing that jumped out of the data up front, was that our data source contained two ticket price values, one for Adult Weekday Tickets and one for Adult Weekend Tickets. Since our central goal was to explore pricing, we needed to determine which of these we would base the eventual model off. We explored the values with a box plot to better understand the variations across states and between nights and weekends:



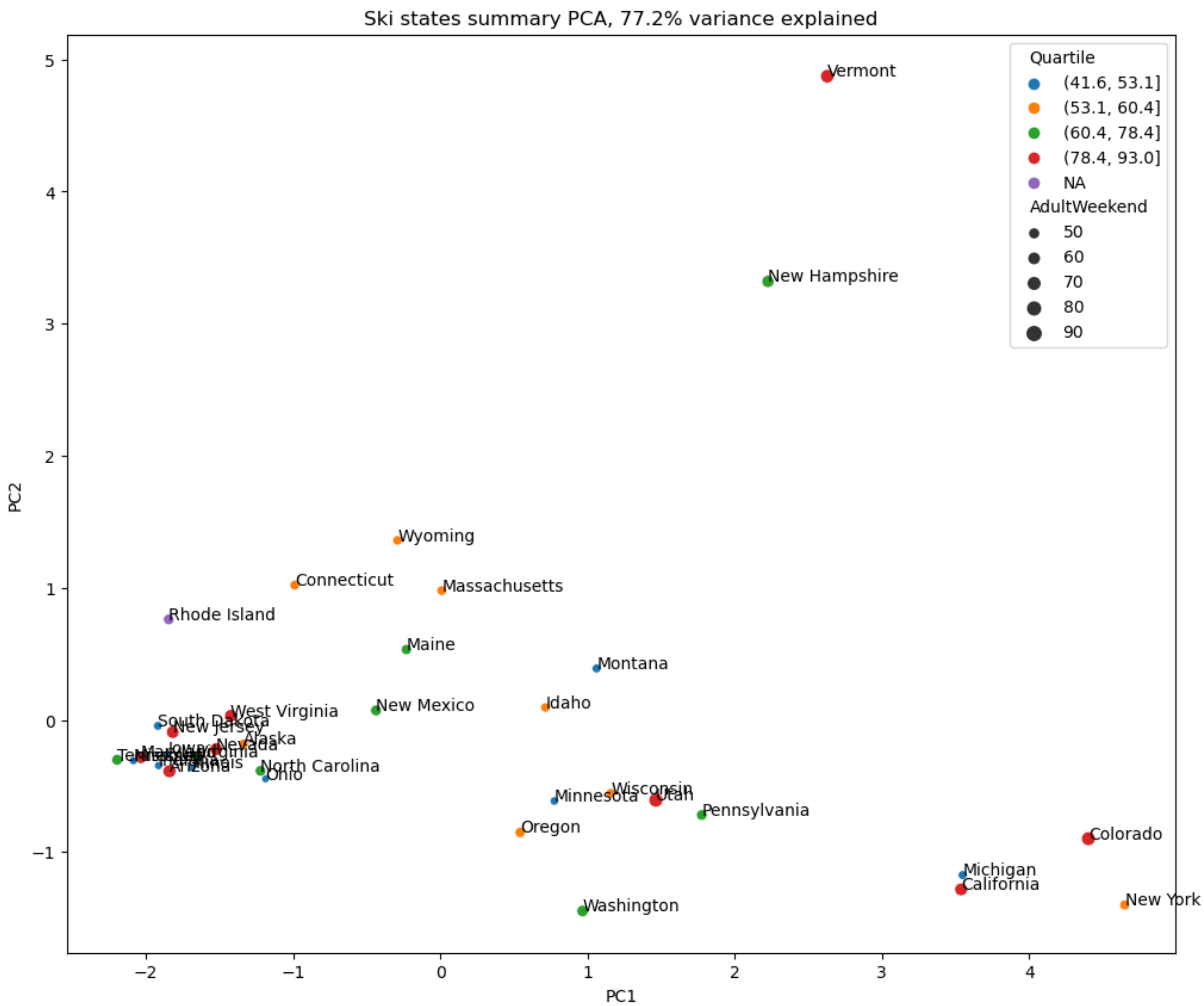
Looking at the data for each State, we noted that for many resorts the value was the same, and that variability was concentrated in the sub-100 dollar/ticket range. A review of Montana's data specifically showed there was no variability for Montana resorts between Weekend and Weekday. We also realized there were fewer missing values for the Weekend prices, so we decided to proceed with those values for our analysis.

We then created some summary statistics for each of the states, summing things like nightskiing acreage, total days open for the state, number of resorts, and some other features. We also pulled down general population and total area information for all of the states from Wikipedia (https://simple.wikipedia.org/w/index.php?title=List_of_U.S._states&oldid=7168473), cleaned/verified some of the data, and merged this data with our state summary statistic information from the resort data.

Very cursory examination showed that New York state had the most resorts per state, but interestingly this did not correlate directly to total skiable area. Montana, our target state, did show up in the top five of total skiable area, despite not being top five for total resorts. These insights led us to look at resort density measures as perhaps more insightful, for example resorts per 100k population and resorts per 100k square miles. Per 100k of population, our target state Montana was in the top five, but not in the top five of resorts per 100k sq miles (which intuitively makes sense). Vermont and New Hampshire showed up in the top five for each of these density measures.

Exploratory Data Analysis

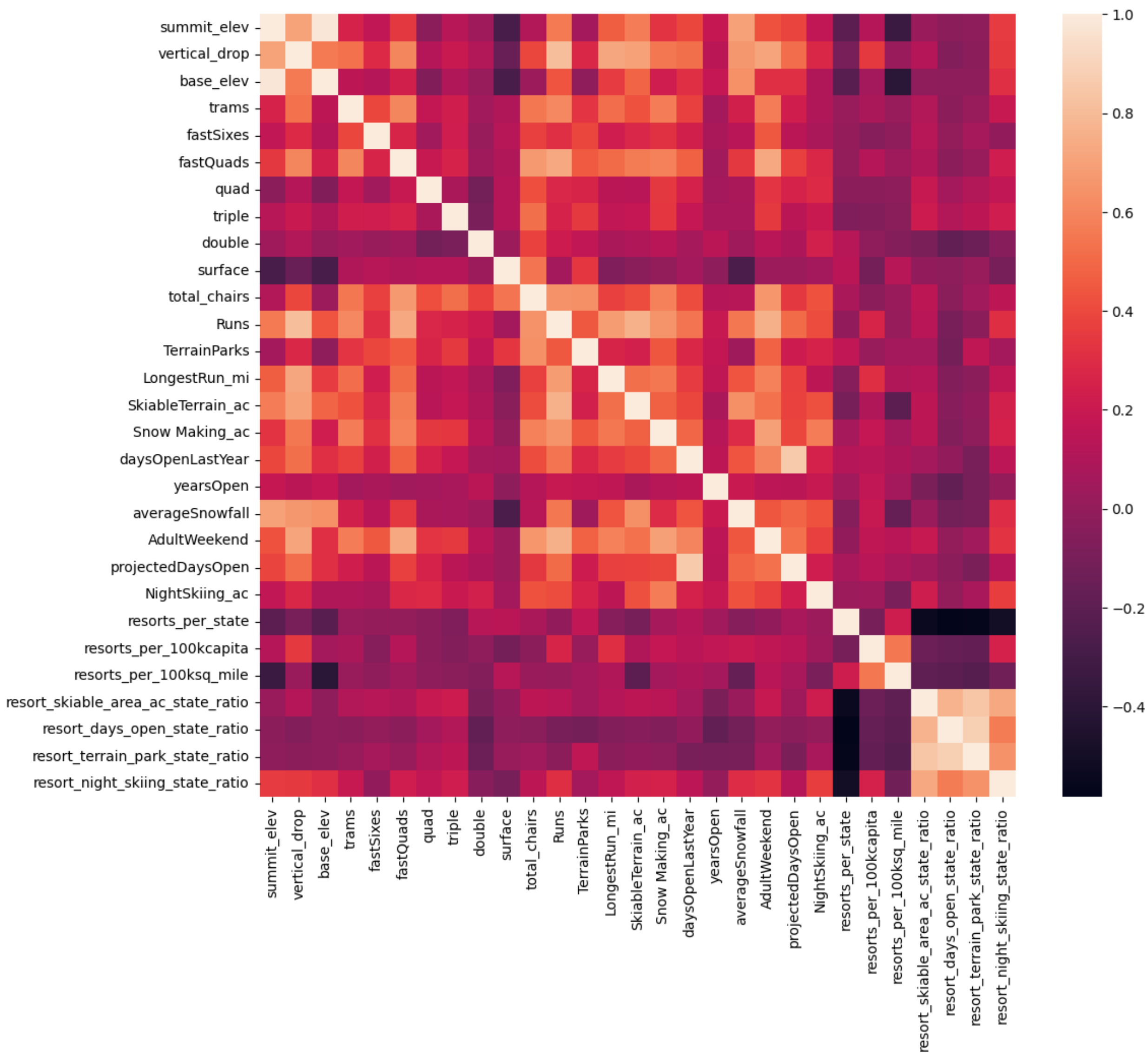
We then shifted over to Principal Component Analysis (PCA) to help disentangle the relationships in our highly dimensional data and identify the most important components driving price. This showed us that our first two principal components (PCs) accounted for 75% of the variance in price. We then moved forward with visualizing these two PCs, in a scatterplot with each point representing a state, its average weekend ticket price, and the quartile that average price fell into:



As can be seen in the plot above, there is not an obviously discernable pattern with ticket pricing relative to the identified PCs. This helped inform our conclusion that really states should probably be treated equally in our eventual model.

With this conclusion reached, we merged our state summary data into our original resort data set. We took some of these state features and used them to generate competitive features for the resorts relative to other resorts in their state.

This merging allowed us to generate a helpful feature correlation heatmap to visualize correlations, particularly with our target “AdultWeekend” ticket price value. Some of the initial ones which stood out were “fastQuads”, “Runs”, and “Snow Making_ac”. There was also a correlation with one of our new features “resort_night_skiing_state_ratio” and to a degree with “vertical_drop”.



We then further explored the relationship of “AdultWeekend” ticket price with features that involved moving people around the mountain, including Chairs - Runs, Chairs - Skiable Terrain, and the same for Fast Quads. This produces somewhat counterintuitive results, that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. We posited that this may be an exclusive vs. mass market resort effect related to ticket price relative to number of visitors a year, but we lack data on the number of visitors a year for each resort to further explore this relationship. Lastly, it appeared that having no fast quads hurt ticket prices, and that for resorts with lots of skiable terrain, adding a small number of fast quads could increase ticket prices.

<u>Model Choice</u>

We started out by loading in our data and extracting a copy of the data for Big Mountain, which is our target resort. We then created a 70/30 train/test split to be able to objectively evaluate models/approaches.

We started by looking at how well the simple avearge functioned as a predictor – this produced a Mean Absolute Error (MAE, the average of all absolute errors) which indicated our predicted price fell within about 19 dollars of the expected price. We then explored a linear regression model, which performed far better than the mean alone and after tuning got us to an MAE predicted ticket price within about 12 dollars of the expected price.

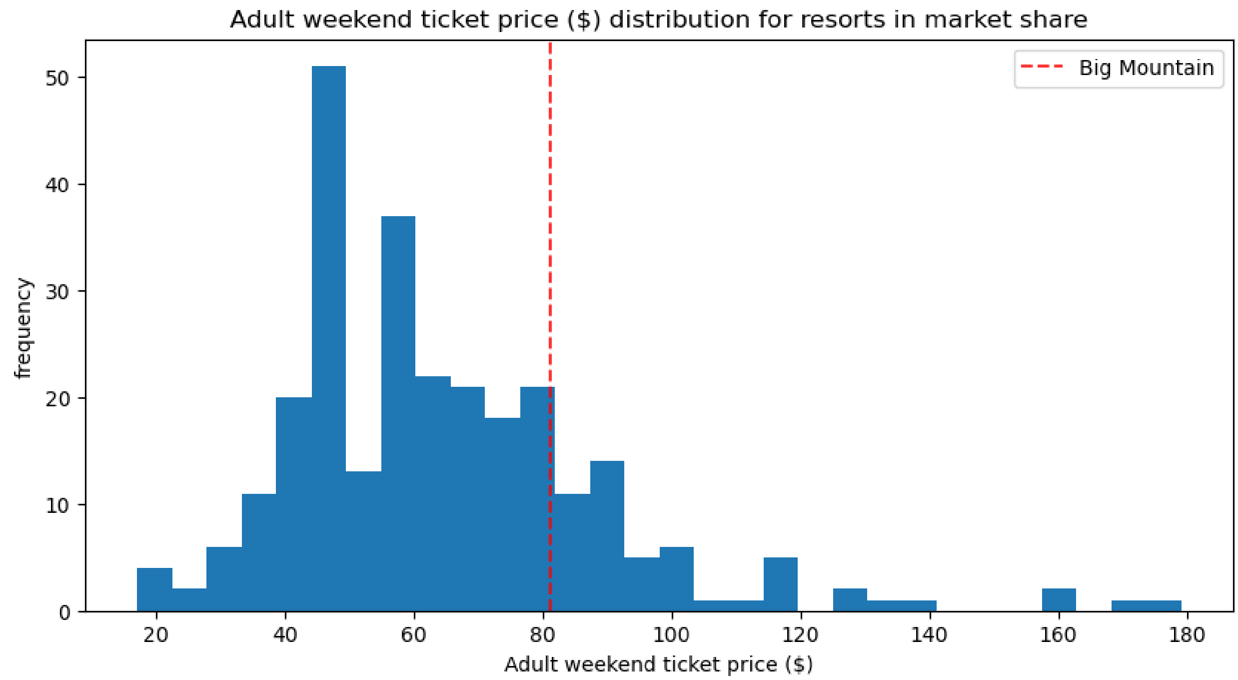
We explored the possibility that the linear regression model was over-fitting to the training data, and ultimately opted to pare down the number of features we were using to train the model. This led us to eight key features, with vertical drop, snow making coverage, total chair lifts, fast quads, and runs as some of the most predictive.

We also explored a Random Forest model as another potential model to apply to our data. The Random Forest with some tuning of parameters did marginally improve our MAE with a predicted ticket price within about 10 dollars of the expected price. The Random Forest model also exhibited less variability. Another encouraging aspect was that many of the most predictive features were in common with the linear regression approach - namely fast quads, runs, snow making coverage, and vertical drop. Based on the better, less variable performance, we opted to move forward with the Random Forest model.

Results

Our initial prediction for Big Mountain's ideal ticket price after fitting the model suggested that Big Mountain could be undercharging about 15 dollars - even with an expected absolute error of about 10 dollars, this initially suggested that Big Mountain has some room to increase prices. This is a rough approach and doesn't consider the possibility that other resorts are either over or under priced. To address this, our model could have used some additional information, such as operating costs for various resorts, to better understand and model the disparate drivers of price.

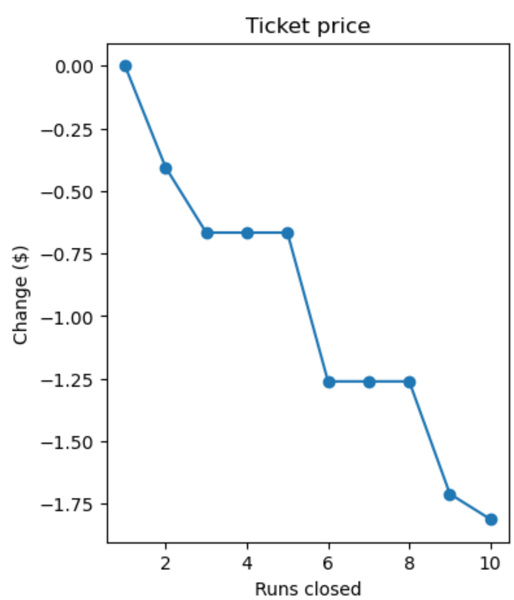
We then looked at Big Mountain's position on key features (including price) relative to its peers. This indicated that Big Mountain's ticket price was somewhat higher compared to the market, and at the top end for Montana; however, when you look at Big Mountain’s positioning on other measures of facilities offered, the resort remains on the high end, in many cases the extreme high end of the distribution:

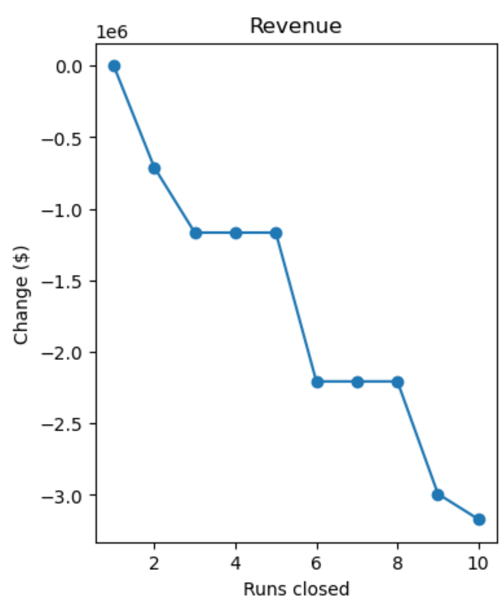


At this point we modeled some options put on the table as possible scenarios, to see how they influenced predicted pricing.

Scenario 1

The first of these was looking at closing up to 10 of the least popular runs. Based on our modeling, closing one run would make no difference to ticket pricing. Closing 2 and 3 runs decreased predicted ticket price by roughly 0.40 cents and 0.70 cents respectively. Decreases leveled off between 3 and 5 runs, which suggests if the resort decided to close 3 of the least used runs, they should go ahead and close 5 for additional savings with no impact on pricing. Prices dropped considerably again between 5 and 6 runs closed.





Scenario 2

The second scenario looked at increasing vertical drop by 150 feet, which would also require the addition of another chair lift. This scenario would increase the predicted ticket price by 1.99, which would amount to roughly 3474638 in additional revenue each year.

Scenario 3

Scenario three was the same as scenario two, but looked at adding 2 acres of snowmaking capacity as well. This change resulted in no meaningful impact on the ticket pricing per the model.

Scenario 4

Scenario four looked at increasing the longest run by .2 miles, which also required a 4 acre increase in snowmaking capacity to ensure coverage. This change also resulted in no change to the ticket price. (This could be a result of the fact that the random forest model we elected to use considered longest run to be lower in the hierarchy of important features).

Conclusions

The second scenario looks quite promising in terms of potential for profit generation. We had been told up front that the addition of another chairlift this year increased operating costs by 1540000 - but if another chairlift could increase ticket revenues by 3474638, that is a considerable profit margin. A factor that would need to be explored is the cost of the lift itself, and how that may factor into Return on Investment (ROI) longer term.

Closing less used runs could also be an area for exploration, but its utility will be highly dependent on the expected savings in operational costs that would be achieved by closing the runs, and how this interacts with the decrease in predicted ticket price.

In terms of areas for further exploration/expansion, we could benefit from additional datapoints/information. One item we homed in on here is the interaction between operating costs and ticket pricing. Also, while we have information on the operating cost of the new chairlift, it would be helpful to know the upfront installation costs for the chairlift and the business's perspective on necessary profit to achieve an acceptable ROI. Previously we've noted that number of visitors per year would also be a helpful metric to better predict ticket pricing. It would also be helpful to know the breakdown of in state vs. out of state visitors, given that Big Mountain was at the top-end of pricing within Montana.