Guided Capstone Project Report:

The purpose of this project was to come up with a predictive pricing model for ski resort tickets in our market segment based on a number of facilities, or properties, boasted by other resorts at their resorts. Big Mountain, the resort in question, thinks it may not be maximizing its returns relative to its market position. They don’t currently have useful feedback on what facilities visitors are more or less likely to pay for.

The data is missing some information I’d have found useful in pricing, particularly operation costs and maintenance costs. For cleaning purposes, a number of rows and columns had to be dropped. The data is missing about 15% of the adult weekend ticket prices and about 16% of the adult weekday ticket prices, and about 18% of the rows total are missing ticket pricing information. About half the values for “fast eight” were missing as well. I had to make a distinction between regions and states because the names overlapped in some cases. They were not interchangeable. One resort was reported as being open for over 2000 years. All told, the rows I ended up working with numbered 277 from the original 330, and the columns 5 from the original 27. I did have Big Montana’s data though, so I can proceed.

An early observation was that weekend prices only exceed weekday prices on resorts that have tickets priced under $100. Some other observations: Montana is the third largest state in square miles, and has the fourth most resorts. Montana does not place in the top 5 for night skiing or days open. The heatmap shows that night skiing, runs, chairs, and snow closely correlate with price.

I went ahead and got going on model selection after exploring the data. The amount of terrain that was skiable came up as a negative since it stretches the service abilities of chairs and fast quads. The top 4 most positive in terms of pricing across the linear regression model used and the random forest model were: vertical drop, snow making, runs, and fast quads. I ended up sticking with the random forest model over the linear regression model despite the similar results because the random forest model had a lower cross-validation mean absolute error by almost $1 and because it had less variability. Verifying performance on the test set produced performance that was consistent with the cross-validation results.

For the model to work, you can’t include Big Mountain’s data with the rest. This introduces a bias to the data when you’re trying to find the best price. It’s contamination because Big Mountain isn’t competing with itself. The model, measuring the resorts that aren’t Big Mountain, shows that tickets can be priced at $95.87 instead of the actual price of $81. This should be taken with several assumptions though: the resorts are/aren’t pricing accurately based on the market, the resorts are/aren’t overpriced, that our resort and others are mispriced, and that we’re missing data. We can also see that, excluding outliers, Big Mountain is in the top for: total chairs, fast quads, runs, longest-runs, vertical drops, areas covered in snow, and amount of skiable terrain. Of the 4 scenarios presented, scenario 2 yielded the best results at a $1.99 ticket price increase yielding a projected $3,474,638 in revenue over the season. Scenario 3 had the same revenue projection, so the snow is a waste of money. Scenario 1 only results in losing money unless you close 1 run, and scenario 4 had no effect on the price. This was based on the assumption that 350,000 people visit during the season and that they ski for 5 days on average.

My suggestion for the model would be to try other variations of vertical drop, snow making, runs, and fast quads outside of the 4 highlighted scenarios. This is what we can recommend based on the data provided.