**Springboard Guided Capstone Documentation**

First we did a number of various steps to clean and wrangle the data. This included removing data with missing prices as that data was very little use to us. We also investigated whether the state and region had much affect on ticket price but decided that the state was not a large factor in pricing. Some states like North Carolina and Virginia had a high difference in weekend prices vs weekday prices but we later found that this was actually a result of resorts priced under $100 used this strategy in general. We decided to treat all states equally and build a pricing model that takes all states into account. After cleaning our final data contained 277 rows and 25 columns. We created a heatmap to look at correlation between different factors and found that obvious things like summit and base elevation, or days open and projected days open were highly correlated. We then created a number of scatterplots with ticket price against other features we saw a strong positive correlation with vertical drop, runs, fastquads, and total chair lifts vs the ticket price. 

To get a baseline of performance we took the mean of Adult Weekend price to see if it was a good predictor and found if we used the mean absolute error and averaged all the known features we would be off by about $19 on ticket price.

We then created linear models by imputing missing values with both the mean and median and saw that the results were very similar to each other. We also used sklearn and pipeline to check the same results. We had overfitting concerns as we were using all the features to build our models. While the default k value of 10 was a bad indicator for performance, by using cross-validation technique we found that a good value for k was 8 using the features vertical drop, snow making acreage, total chairs, fast quads, runs, longest run, trams and skiable terrain. Vertical drop had the largest positive association with ticket price whereas trams and skiable terrain acreage had the largest negative association with ticket price. We used random forest regressor to find that imputing with the median is better than the mean and the four dominant features were fast quads, runs, snow making acreage, and vertical drop.



We then modeled a number of scenarios to create recommendations. We found that while Big Mountain is currently priced at $81 for an adult weekend ticket there is support for $95.87 a ticket and even if the expected mean absolute error of $10.39 was fully used we could still raise ticket prices. Big Mountain is ahead of other resorts in important features like vertical drop, snow making acreage, total chairs, fast quads, runs, longest run, and skiable terrain acreage. We found that we could close one run and that would have little to no affect on ticket price. But without knowing how much it costs to run each run we cannot recommend the closure given the missing information. Closing more runs does reduce support for ticket price so we certainly recommend against that.



Our most successful scenario was when we added a run, increased the vertical drop by 150 feet, and installed a new chair lift, this created support for an addition $1.99 in ticket price and could bring in an addition $3.4 million over the course of a ski season. We saw that adding additional snow making did not affect that same scenario and neither did adding length to the longest run so we don’t recommend expanding these features.