Capstone 1: Guided Capstone, Big Mountain Ski Resort
Step 06: Documentation

The data set for Big Mountain Ski Resort and it's US market competitors included 330 ski resorts, and 26 unique variables, including the 2 response variables for weekend and weekday ticket prices. We would explore the data, cleaning it up and making observations about necessary exclusions and modifications, until we had a sufficiently clean set of data to run a series of regression analyses on.

To start with, we were missing many values from some variables, or those variables had entries that were inconsistent with the rest of the entries in that particular variable. For example, in the variable describing the number of years open, we found that the maximum value was 2,019. Since it is unlikely that a North American ski resort has been open for over 2,000 years, it became clear that instead of entering a count of years a particular facility was open, someone had entered the year in which it had opened. This was the sort of data we excluded or modified.

We also found that several variables had hundreds or dozens of values missing entirely. Some of them were easily understood, like when the variable describing the number of 8-person fast speed chairlifts at the resort had over 50% of it's values missing, or when the variable describing the number of acres for night-skiing had just over 43% missing. There is one highly probable explanation for this, which is that these resorts simply didn't have these features at their facilities, so they had left those questions blank. And, when 15% and 16% of Weekday and Weekend ticket prices were missing from the data set, the most likely explanation was that some resorts (roughly 15% of them) did not have different prices for weekends, but instead charged the same amount 7 days a week. 8 variables in total were missing more than 4% of their data, and all of those were missing 13% or more, with just two missing close to half of their data (Number of 8-person chairlifts, and number of acres for night skiing).

During our preprocessing, we found that a random forrest model suited our data set better than a linear regression model. It had a lower mean absolute error, and had less variability too. We also conducted a test where we compared cross-validation scores as we increased sampling sizes. We found that sampling sizes reached their greatest improvement in cross-validation scores somewhere between 40 and 65, with a long plateau and extremely slight increase again near 160. This would indicate that we have plenty of data to work with, and we can use sample sizes of n = 40-50.

We began by modelling the Big Mountain Resort prices against the market of other resorts, and we found that "Big Mountain Resort modelled price is $91.09, actual price is $81.00.
Even with the expected mean absolute error of $7.03, this suggests there is room for an increase." This suggests that the Big Mountain Resort is overpriced by about $10, however, because the validity of the model relies largely on the assurance that the market prices reflect a relationship between features and facilities of the competing resorts. In short, if our resort is "over-charging" fort tickets, how many other resorts are "under" and "over" charging for their

tickets, and what effect would this have on the model? And what other variables about visitor histories, visitor demographics, or features of the resorts not reported in the data set we have? For example, if a resort has a hotel on-site, does that have a material impact on ticket pricing? How does that effect size interact with other variables? Does it compound on them, or does it reduce the effect on resort ticket prices?

We looked into whether closing some runs would make a difference in the predicted value of a ticket based on the market. We found that closing between 1 and 5 runs made very little difference, but that closing 6 runs caused a sharp decrease in market ticket values, and therefore significantly decreased revenue in direct proportion. The interesting part of this is that Big Mountain resort can safely close up to 5 runs, thereby reducing the need for staff to operate lifts near them, or to operate at a lower capacity. In addition to saving money on the payroll by closing 1-5 runs, the resort could also potentially save money on maintaining the runs, and on maintenance for any nearby lifts or trams.

We then tested a series of scenarios where we made changes in the park and sought to get an idea of how that may affect the market value of a ticket. The only scenario which raised market values for Big Mountain Resort tickets was adding a run which increased the resort’s vertical drop by 150 feet, and installing an additional chair lift. This led to an $0.18 increase in the value of a ticket, which is at least $315K of extra revenue over the course of one year.