Big Mountain Resort, a ski resort in Montana, has recently installed a new chair lift to better distribute visitors across the mountain. This new chair lift increased their operating costs by $1,540,000 for the upcoming ski season. They would like to evaluate scenarios for raising prices or reducing operating costs to ensure that the new ski lift does not negatively impact projected earnings. We will determine which features could help drive a pricing increase to help drive a 10% revenue increase over the next year. We will deem the new pricing strategy a success if we attain our 10% increase objective. As far as potential constraints for this model, we have identified the possibility the pricing may be strongly impacted by things outside our control, such as weather conditions. Lastly, a potential constraint could be that increasing our prices may negatively impact attendance.

We obtained a report from Alesha Eisen, Database Manager, containing information on 330 ski resorts, including Big Mountain Resort, vying for the same market share. All ski resorts in the United States were considered as in scope for the resort market in which Big Mountain operates. We identified several features with significant numbers of null values. As such, we removed the fastEight feature (almost entirely null or 0) and all rows that were missing pricing information for the AdultWeekend ticket. This reduced the population of resorts in our dataset down to 277.

As part of the exploratory data analysis phase of this project, we first assessed some summary statistics of the numerical features. Many of the numerical features were related to the number of chairs / lifts in the parks. Beyond the number of chairs and lifts, there were also numerical features related to the amenities offered by the resorts. These include elevation, vertical drops, # of ski runs, skiable areas, night skiing areas, and terrain parks. There are also time related numerical features as well, such as # of days open in the prior year, number of years the resorts have been open and projected number of days the resort will be open. Lastly, and most importantly, we have the Adult weekend ticket pricing. Due to the sheer number of numerical features, we found it necessary to reduce the data using principal components analysis. With a result showing that components 1 and 2 accounted for nearly 75% of the variance and the first 4 components 95% of the variance.

A graph with a line

Description automatically generated

Next, we created a correlation heatmap and were able to see some correlation between ticket price and the following variables: fastQuads, Runs, and Snowmaking\_ac., resort\_night\_skiing\_ratio, runs and total\_chairs, and lastly vertical\_drop.

Before creating any models, we first calculated the mean as a predictor.  The resulting average price across our data set was $63.81.  We then calculated the MSE(mean squared error) to test the effectiveness of using the simple mean.   It shows that we can expect to be off by around $19 if we used averages of known values to guess our ticket price. Next, we leveraged a linear regression model imputing missing values first with median values and then with mean values. We found very little difference between the 2. Both resulting in mean squared error of around $9.00. Both would be a better price predictor than using the simple average ticket price, however there was a significant difference between our training sets and our test set, suggesting that we were overfitting. In an attempt to refine our linear regression model, we leveraged sklearn to identify the best number of parameters to use. Resulting in a value of 8, and determined that our most impactful features in order of importance for our dataset are: Vertical drop, Snow Making, total\_chairs, fastQuads, Runs, LongestRun\_mi, trams, skiableTerrain\_ac. Vertical drop was our biggest positive feature.

We moved from a linear regression model to a random forest model. We leveraged the preprocessing steps of imputing missing values using the mean & median and scaling the data. We found that imputing missing values was more successful than scaling and more specifically, imputing missing values with the mean. The random forest model also provides us with a top four features list including fastQuads, Runs, Snow Making, and Vertical drop. Lastly, we evaluated the mean absolute error for the linear regression model and that of the random forest model for our data. We found that the random forest model produces the lower cross-validation mean absolute error of $9.53 vs $10.50 for the linear regression model. Therefore, I would propose moving forward using the random forest model.

One of the options proposed, in addition to increasing prices, involved closing ski runs to reduce operating costs to offset the cost of the new lift. I would propose treading cautiously here. While closing a single run, is predicted to have no impact on ticket prices, anything more than that may result in a decline in ticket prices.

A graph of a price

Description automatically generated with medium confidence

In summary, we have modeled several pricing scenarios to cover the additional costs of the new ski lift for Big Mountain. Big Mountain currently charges $81 for a ski ticket. Given our current amenities at the resort, our model predicts that a ticket price of $95.87 could be supported in the marketplace. Even given an expected mean absolute error of $10.39, it would suggest there is room to increase our current pricing. At $81, we are moderately priced in the market, yet the most expensive ticket in the state of Montana.

A graph with blue lines

Description automatically generated

We expect roughly 350,000 visitors per year, who on average, purchase 5-day tickets.  The operating costs of the new ski lift of $1,540,000 annually means we are incurring an additional $0.88 in cost per ticket.  This means that even if we modestly raise our prices to $85.48 (our predicted price less the mean absolute error of $10.39), we would more than cover the operating costs of the new lift, while generating an incremental $3.60 per ticket over covering the cost of the new lift, or $6.3 million in revenues.

Lastly, the scope of our future work is to embed this pricing model within and Excel workbook for the financial and business analysts to use for what-if analysis in the future.