In this project I looked at a dataset that included information on the amenities and prices of ski resorts around the country to make a recommendation for how Big Mountain could more appropriately price its tickets in comparison to its competitors. An initial look at the dataset revealed that while there were some variables that were missing data (as much as 50% for fast Eights), the majority were not missing any, and the most important column (resorts) was missing none. I started by cleaning the data by determining if there were duplicates and removing them if necessary. I then sorted it and created plots to look at a number of pricing differences, such as the relationship between a resort’s weekend vs weekday pricing. I also created visualizations of each of the numeric features to see where there were normalized and non-normal distributions as seen below.

Diagram

Description automatically generated

Ticket prices for days vs weekends in Montana is the same, and the majority of locations have greater weekend costs than weekday costs.

I then conducted an exploratory analysis of the dataset to identify the most relevant variables. Using principle components analysis I found linear combinations of the original features that are uncorrelated with one another and ordered them by the amount of variance they explain. I also used a feature correlation heatmap which is a way to conduct a high level view of the relationship between he different resort features (seen below).

Chart

Description automatically generated

Using this heat map we see that fastQuads, Runs, and Snow\_Making\_ac are the three most correlated variables to AdultWeekend ticket prices. Of the new features, resort\_night\_skiing\_state\_ratio seems to be the most correlated with ticket price. I also created scatterplots of ticket prices against the desired columns as seen below.

Diagram

Description automatically generated with low confidence

We see again the strong positive correlation between fastQuads and runs, as well as vertical\_drop, total\_chairs, and resorts\_per\_100k\_capita.

Two different types of models were used: a linear regression model using median and mean values and a random forest model that had additional variables added. The dominant four features in the random forest model were also common with the linear model; fastQuads, runs, snow\_making\_ac, and vertical\_drop. The Random Forest model had a lower cross-validation mean absolute error by almost $1 and exhibits less variability. If prices were predicted using the linear model, they would be off by about $9 . The random forest model says that the ideal ticket price for Big Mountain would be $95.20 even though the actual price is $81. Even with the expected mean absolute error of $10.45, it suggests that there is room for an increase.

I would make the following recommendations:

By adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift, the support for ticket price would increase by 83 cents per ticket which would amount to 1458333 dollars over the course of the season. Adding an additional 2 acres of snow makes no difference in the estimated revenue increase. Closing a single run would reduce the ticket price by $0.4. Closing a second would reduce the price by $0.69, and a third by $0.30 but there is no difference in reduced revenue and ticket price for the 4th and 5th runs closed. There is a steep dropoff for the 6th and 7th runs closed but closing the 8th – 10th would have no impact on prices and revenue.

Chart, line chart

Description automatically generated

The best recommendation I could make would be to increase the vertical drop by 150 ft and installing an additional chair lift, which increased ticket price by 83 cents and would amount to a total of $1,4585,333 over the course of the year. The scenarios where we add either 2 acres of snow or add 4 acres of snow and increase the longest run by 0.2 miles each have no impact on ticket pricing whatsoever so I would recommend against that. While we know that removing runs may reduce operating costs, we do not know what these costs are and are unable to compare them to the estimated reduction in revenue to determine if shutting them down would be overall profitable.