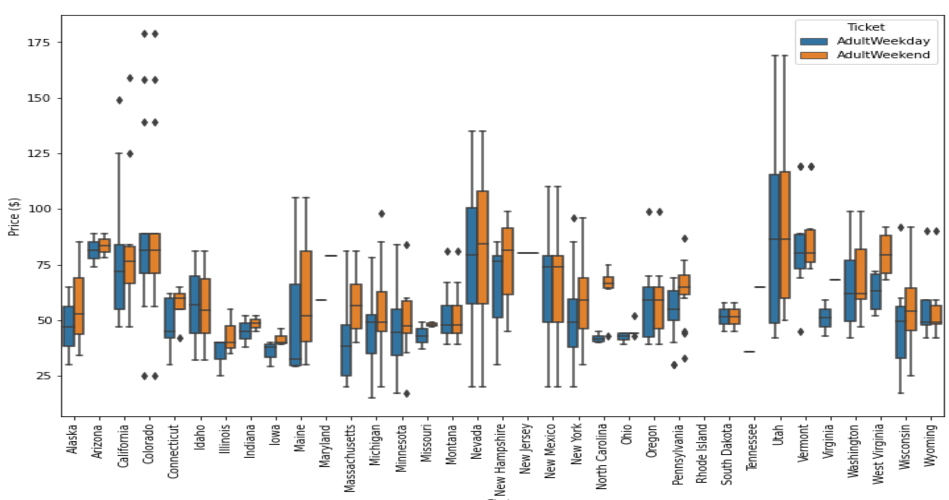
Guided Capstone: Big Mountain Resort
Mark Essery
June 4, 2021

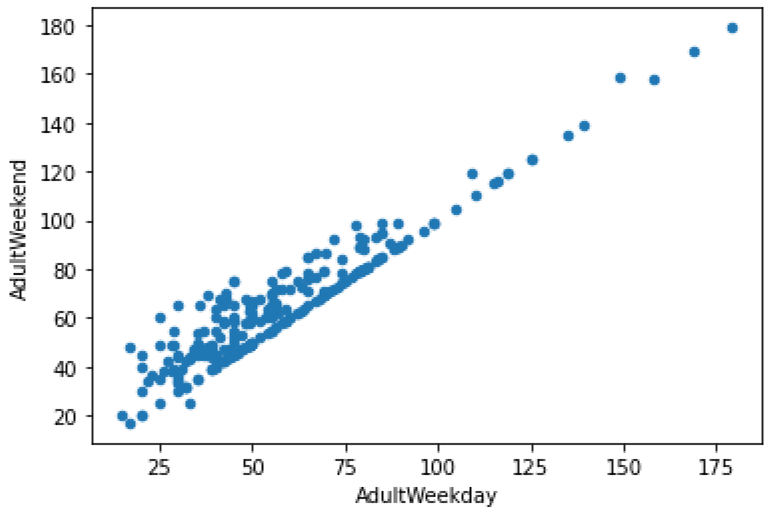
Overview:

Big Mountain Resort, a ski resort located in Montana. Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. This mountain can accommodate skiers and riders of all levels and abilities. These are serviced by 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. The longest run is named Hellfire and is 3.3 miles in length. The base elevation is 4,464 ft, and the summit is 6,817 ft with a vertical drop of 2,353 ft. Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season.
While the resort has been able to charge a higher premium than its competitors within Montana, there's a suspicion that Big Mountain is not capitalizing on its facilities as much as it could. This study aimed to remodel Big Mountain's adult ticket pricing structure through insightful analysis of the provided dataset of 330 resorts across the US and their 27 defining features.

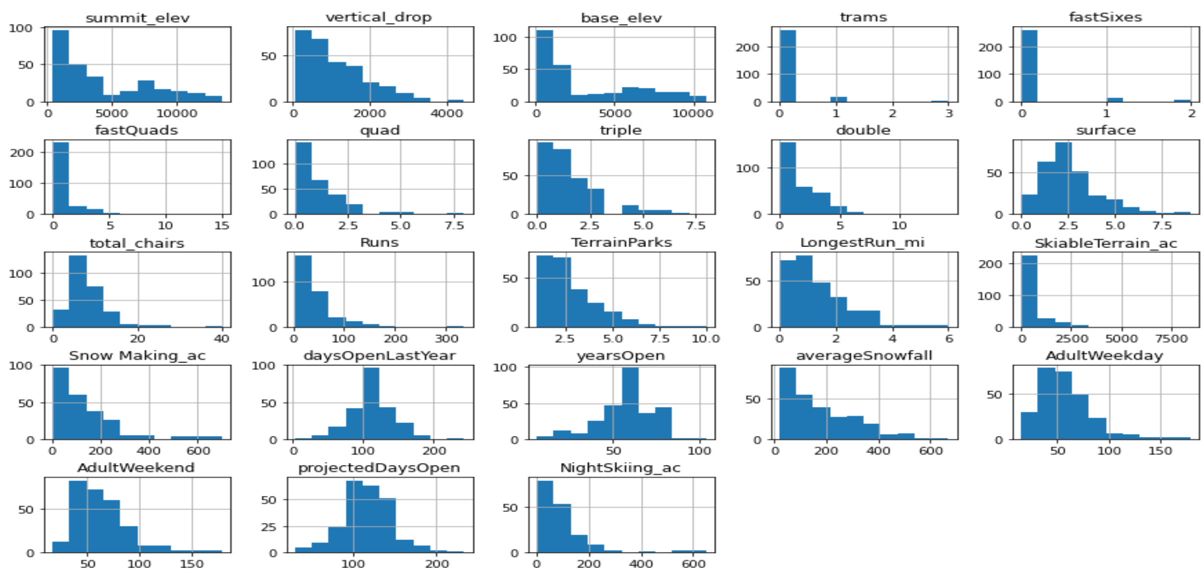
Data Wrangling:

Prior to making any calculations or models, we needed to clean the data for it had numerous null values. Most of the fastEight data was empty or missing so we removed the column entirely. Additionally, only 82% of resorts had no missing ticket price, 3% were missing either AdultWeekend or AdultWeekday tickets, and 14% were missing both. Since there were only 4 missing weekend ticket prices, compared to 7 for weekday tickets, we opted to drop the weekday ticket column and focused our attention on finding the most suitable weeked adult ticket price. It is worth noting that while prices varied greatly across the entire data set, Montana and its adjacent states had an average ticket price of $57 with Montana's mean price being $51.



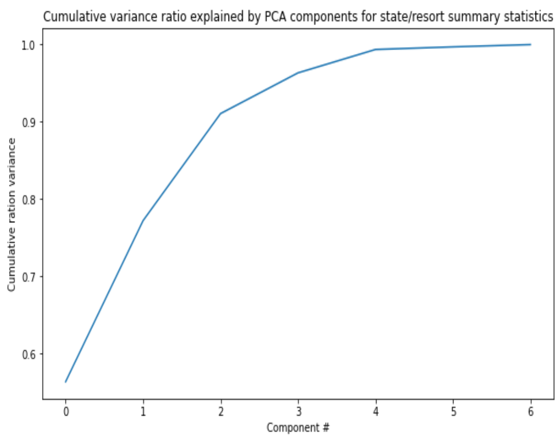


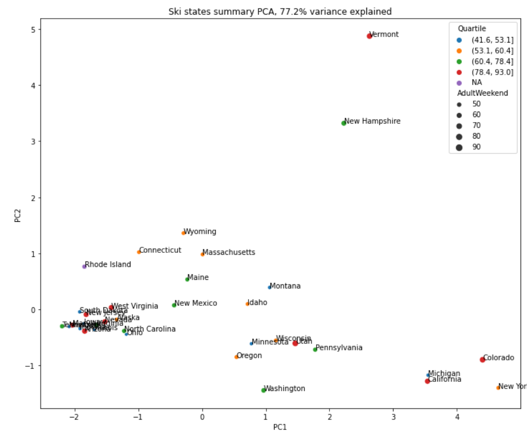
In terms of categorical variables, the region column was redundant so we opted for the state column instead for further insights. We performed feature engineering and created new features such as state population and total area for additional context.



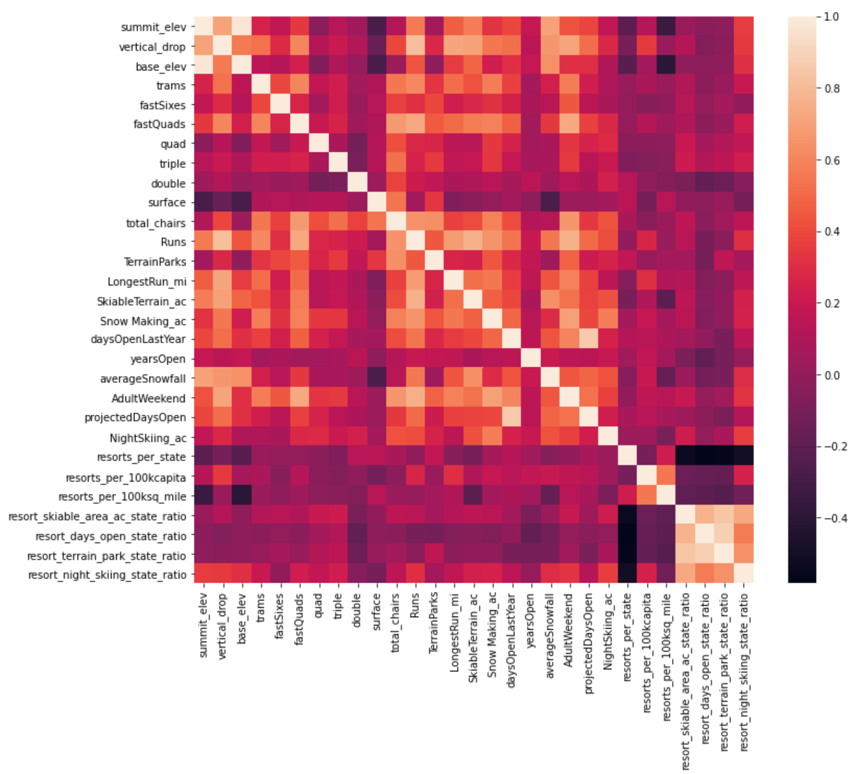
Exploratory Data Analysis:

This step was dedicated to finding patterns and trends within our data in order to provide practical, usable insights. Upon further inspection of our data, merged with information about each state, we found that Montana was 3rd in total state area, 30th in population, 11th in resorts per state, 4th in total skiable area, 8th in total night skiing, and 15th in total days open. One insight drawn from these numbers is that Montana seems to not be very densely populated (given its size and low population) and has fewer, larger resorts than its counterparts. Our next step was to deploy a principle component analysis (PCA) on our data to uncover relationships between our original and derived features, as well as sort them in order of the amount of variance they explain. In order to do this we needed to first scale the data. We see that the first two components account for over 75% of the variance, and the first four for over 95%.

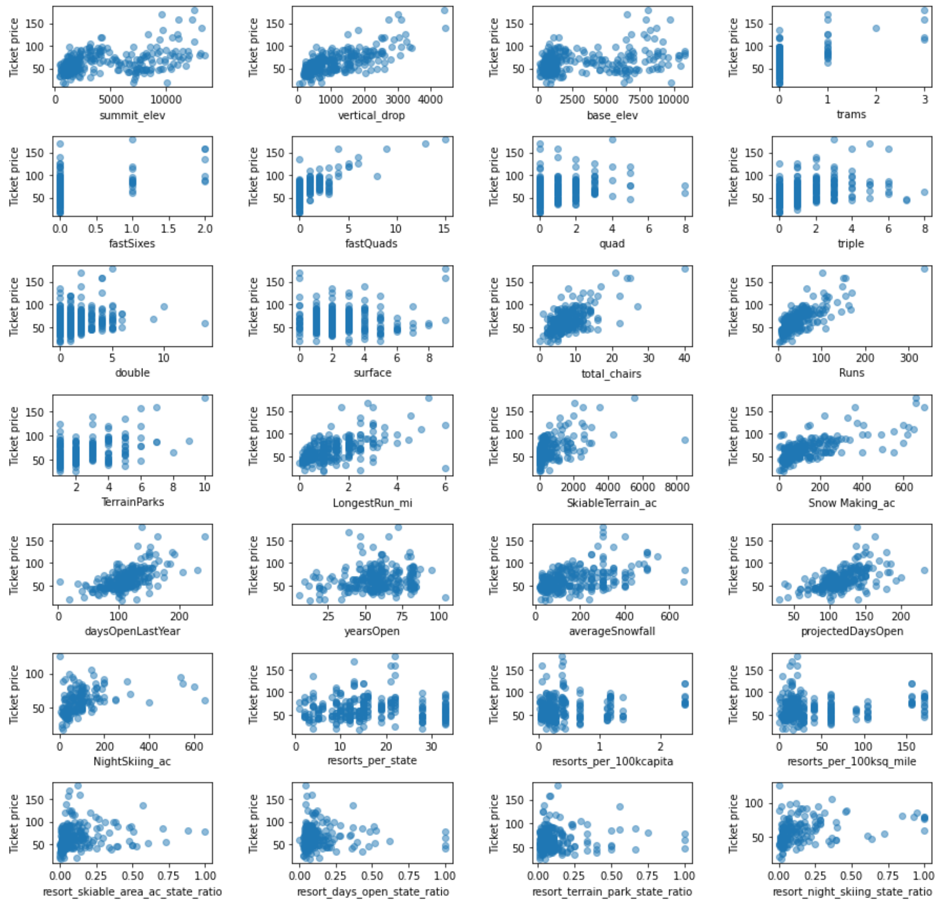




For a more clear view of the relationship between price and all the other features it derives from, we created a heatmap to visualize these relationships and their levels of correlation.



An interesting observation we found was that there is an apparent correlation between the ratio of night skiing area with the number of resorts per capita, which could be that more night skiing is provided in areas where resorts are more densely populated. Closer inspection reveals the strongest relationships associated with AdultWeekend are fastQuads, Runs, SnowMaking_Ac, vertical_drop, and resort_night_skiiing_state_ratio. For the last one, this could mean that seizing a greater share of night skiing capacity within the state could influence a higher price for a resort. With regards to runs and chairs, the more runs you have, the more chairs you would need to transport visitors. Also, guaranteed snow on the runs and larger mountains/higher runs seems to be important selling points that raise ticket prices.



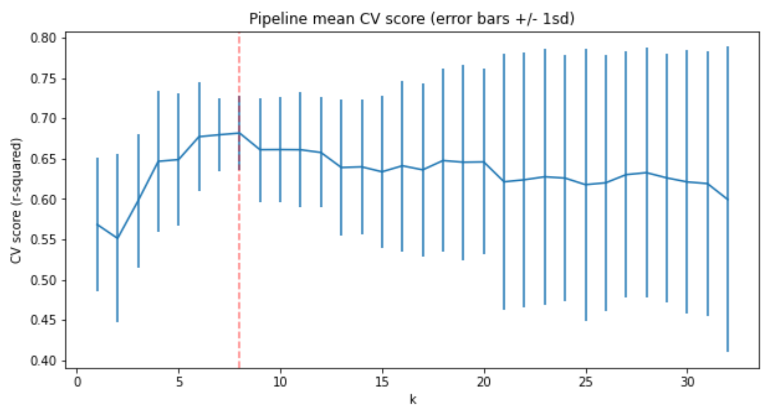
It is worth noting the interesting relationship between chairs, runs, and terrain. It seems that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. What we may be seeing here is an exclusive vs. mass market resort effect; if you don't have so many chairs, you can charge more for your tickets, although with fewer chairs you're inevitably going to be able to serve fewer visitors. Your price per visitor is high but your number of visitors may be low. It also appears that having no fast quads may limit the ticket price, but if your resort covers a wide area then getting a small number of fast quads may be beneficial to ticket price.

Preprocessing and Training Data:

This step in our analysis began with splitting our data set into 70% for training data and 30% for testing. We took the mean of our training set and used it as our basis for predicting price. This resulted in an initial price of $83.31 with a mean absolute error of $19 -- a value much too large to accept.

The next progression involved using linear regression with a standard scaler, building one model imputing missing values with the median and one with the mean. The resulting MAE and r2 scores show that there was no meaningful difference between either method. We arbitrarily chose to use the median moving forward. While linear regression produced a MAE of only $9, we aimed to improve this further.

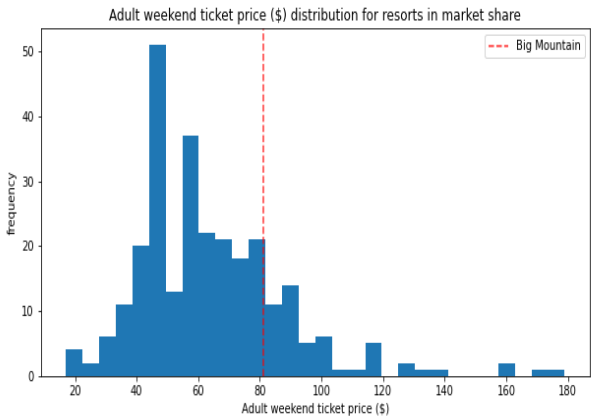
In the interest of efficiency of results and comparisons, we then constructed a data pipeline and used cross validation to test different random sample sizes and assess model performance (this also prevented us from implementing our own biases into the model and potentially creating a model that overfit the data). Using GridSearchCV, we were able to determine that the ideal amount of features the model needed for optimal results was 8: vertical drop, snowmaking, total chairs, number of fast quads, number of runs, longest run, number of trams, and skiable terrain. Vertical drop had a very high, positive influence, while trams and skiable terrain had a negative correlation. This negative relationship could be indicative that, if you kept the total number of chairs and fastQuads constant, but increased the skiable terrain extent, you might imagine the resort is worse off because the chairlift capacity is stretched thinner. In addition, the case could be made that people will pay more for a smaller resort with less people (and thus less crowded).

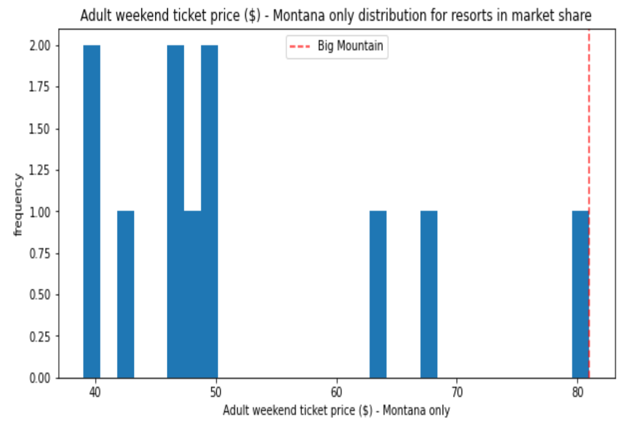


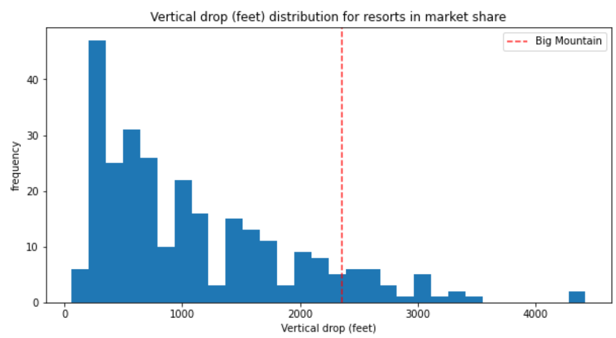
Further experimentation to this data led to the creation of a random forest model to compare against the linear model. The random forest validated the importance of fast quads, runs, snow making, and vertical drop, for GridSearchCV determined these to be the top features of this model. This model produced a MAE of $9.66 with a standard deviation of $1.35, while the linear model furnished $10.50 and $1.62, thus we determined that a random forest was the best option to model Big Mountain's price moving forward.

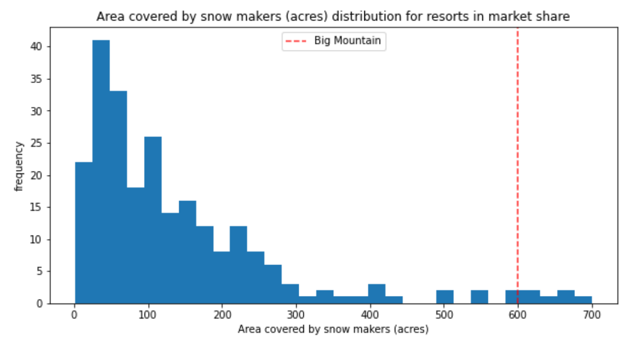
Modeling:

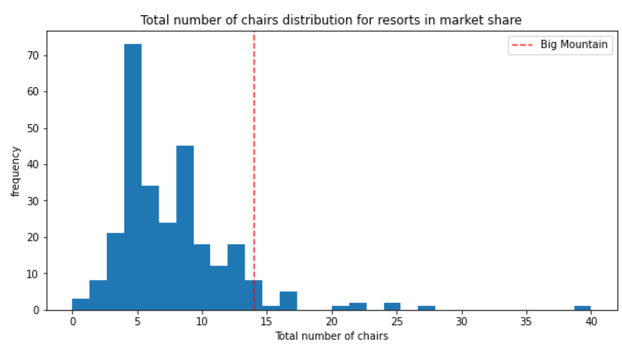
With the random forest model trained and tailored, we then refit the model on all available data, excluding Big Mountain. We determined Big Mountain Resort's modelled price to be $94.22 (its actual price is $81.00). Even with the expected mean absolute error of $10.39, this suggests there is room for an increase. For reference, we plotted adult weekend prices of all resorts against each of the most important features, noting BMR's location in each plot.

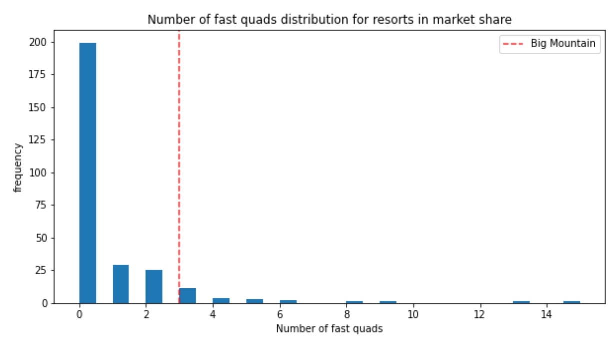

Adult weekend ticket price ($) distribution for resorts in market share
Big Mountain
frequency
Adult weekend ticket price ($)

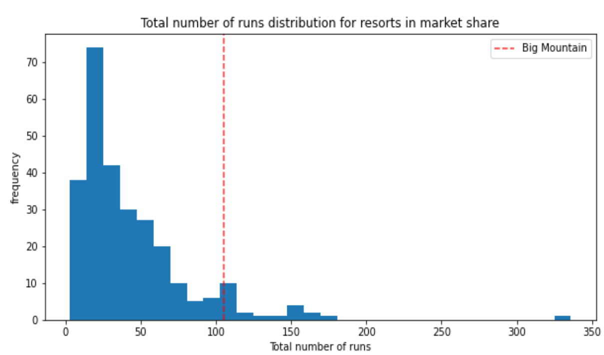

Adult weekend ticket price ($) - Montana only distribution for resorts in market share
Big Mountain
frequency
Adult weekend ticket price ($) - Montana only

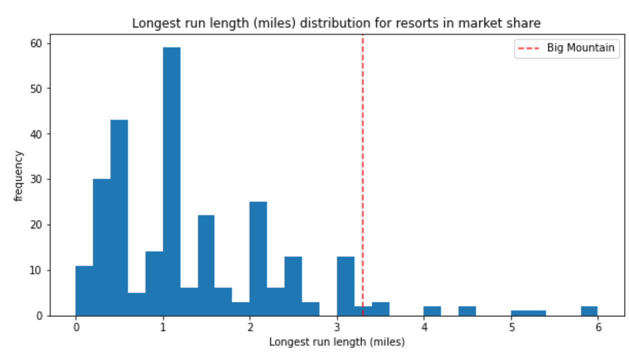

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet)

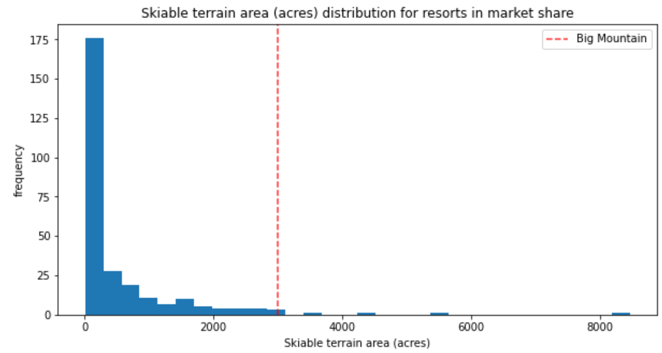

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
Area covered by snow makers (acres)


Total number of chairs distribution for resorts in market share
Big Mountain
frequency
Total number of chairs


Number of fast quads distribution for resorts in market share
Big Mountain
frequency
Number of fast quads


Total number of runs distribution for resorts in market share
Big Mountain
frequency
Total number of runs


Longest run length (miles) distribution for resorts in market share
Big Mountain
frequency
Longest run length (miles)



It is apparent that Big Mountain Resort ranks quite highly in each major feature in comparison to other resorts. The only category it does not dominate in is the number of trams, however most resorts do not have one anyways, so this is not very significant. Thus the higher modelled price is justified.

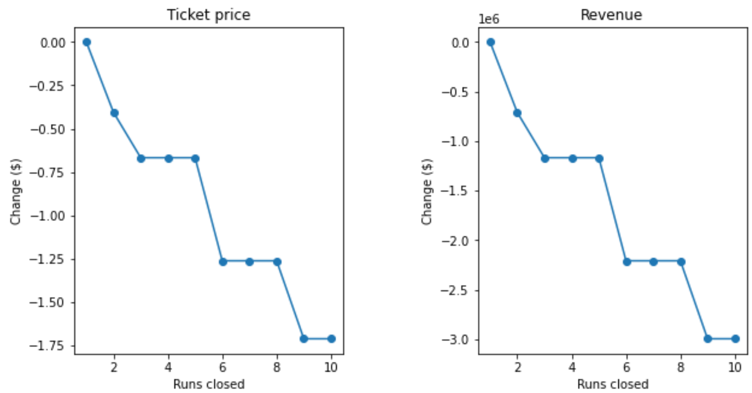
Conclusion:

Big Mountain Resort currently charges $81 per weekend ticket, although through analysis of its features and comparison to its competitors our model suggests a ticket price of $94.22. Our mean average error is $10.39 so this price increase is within reason, as is the possibility of even charging up to $105 per ticket. BMR was priced fairly high (most notably in Montana) and its modeled price is even higher, which could come as a surprise to business executives. Because there are so many variables to account for, it is hard to take everything into full consideration when determining prices with just the human eye. This model took into account all of the best attributes BM has to offer and priced its tickets accordingly, thus its suggested price range is actually quite reasonable given its top tier status for most features.

While leadership may be reluctant to raise prices, it is highly possible that BMR's competitors have mispriced tickets as well, so the fact that we have made these insights puts BMR a step ahead of the competition. However, considering how BMR is top tier in vertical drop, snowmaking area, number of chairs, number of fast quads (most have 0, BM has 3), number of runs, longest run, and skiable terrain, a price increase appears to be quite justifiable.

To increase profits even further by, the most notable candidates for reforming operational costs were: closing down a number of runs, adding 150 ft of vertical drop to a run, adding 2 acres of snow making, or lengthening its longest run by 0.2 miles. Upon further inspection, the data show that lengthening the longest run had zero impact on ticket price, while vertical drop and additional snow making would increase the ticket price by at most $1.99, leading to $347k in additional revenue yearly (each). Closing down 1 run produces no price change, and closing down up to 10 runs would at most drop ticket price by $1.75.

In the future, I would recommend closing the least used runs so as to conserve resources that could be better used elsewhere since its effect on price is negligible. As for how leadership should handle the vertical drop variable, they have options; on the one hand lowering vertical drop by 150ft would result in less resources needed to transport visitors, as well as an increase traffic on that run because there would be more people able to ski it due to faster transit times and more efficient lift coverage. This would at most suggest a price decrease of $1.99, but considering how BMR is already a top tier location for vertical drop we can dismiss this price change, or can off set it by adding 2 more acres of snow making.



Moving forward, other useful information could be average age(s) of visitors; the resort's proximity to civilization and places such as grocery stores, restaurants, and hotels; when the resort sees its highest volumes of customers; and how many customers are returning customers. In the future the more data we acquire will help build a stronger model and will allow us to investigate more features and relationships between them.