BIG MOUNTAIN SKI RESORT - PROJECT REPORT

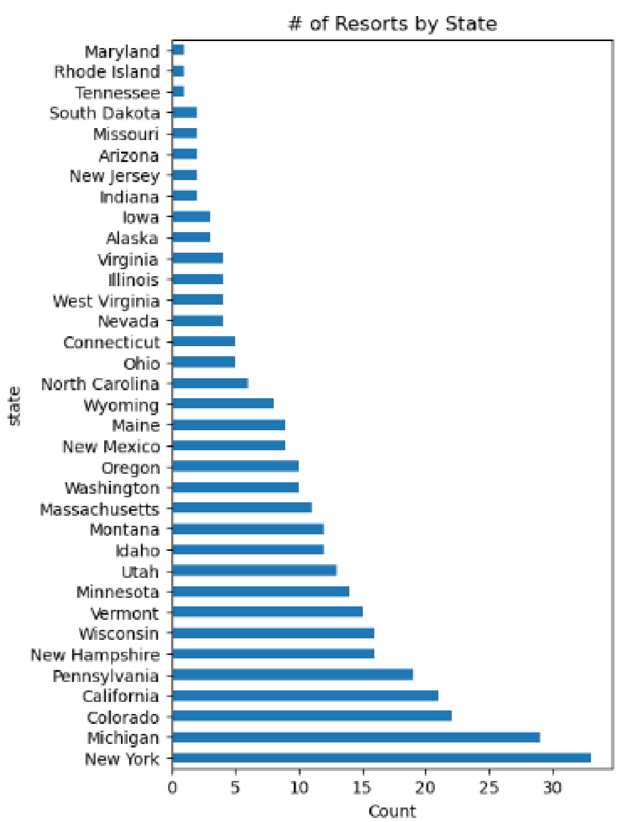
https://wallpaperaccess.com/full/959907.jpg

Big Mountain is located in the old railroad town of Whitefish, Montana, just south of the Canadian border, and attracts approximately 350,000 visitors each year. The leadership team would like to develop a pricing strategy that capitalizes on its facilities. Historically, they simply charged a premium above the average price of other resorts. Moreover, with the recent addition of another chair lift, their operating costs will increase by $1,540,000 each season so they also want to cut costs.

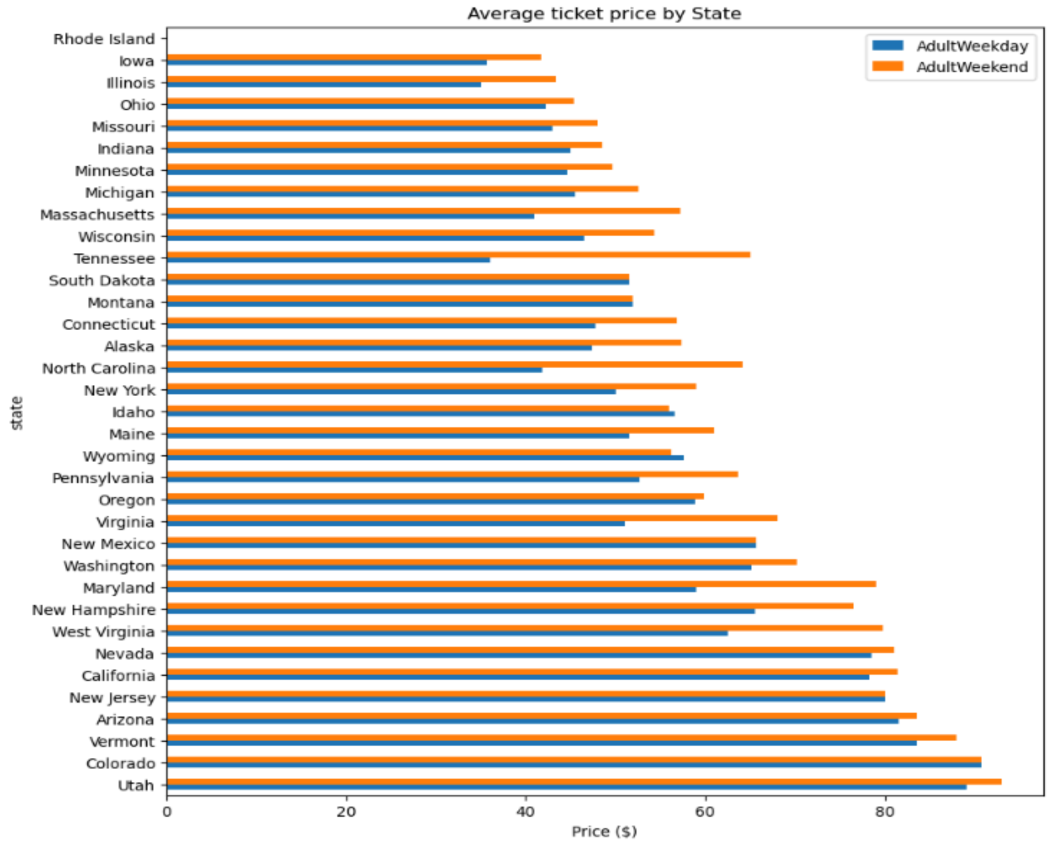
We hypothesized that Big Mountain should focus on reducing its costs by 20% over the course of 6 months and likewise increase its ticket price by 15% to recoup the cost of the additional chair lift.

We were provided with data that included resorts, their locations by state and region, and various features such as number of runs, lifts, acreage, snow making coverage and whether or not night skiing was offered. While we knew our focus would be to model ticket price, we needed to determine whether to use the weekday or weekend price in our model.

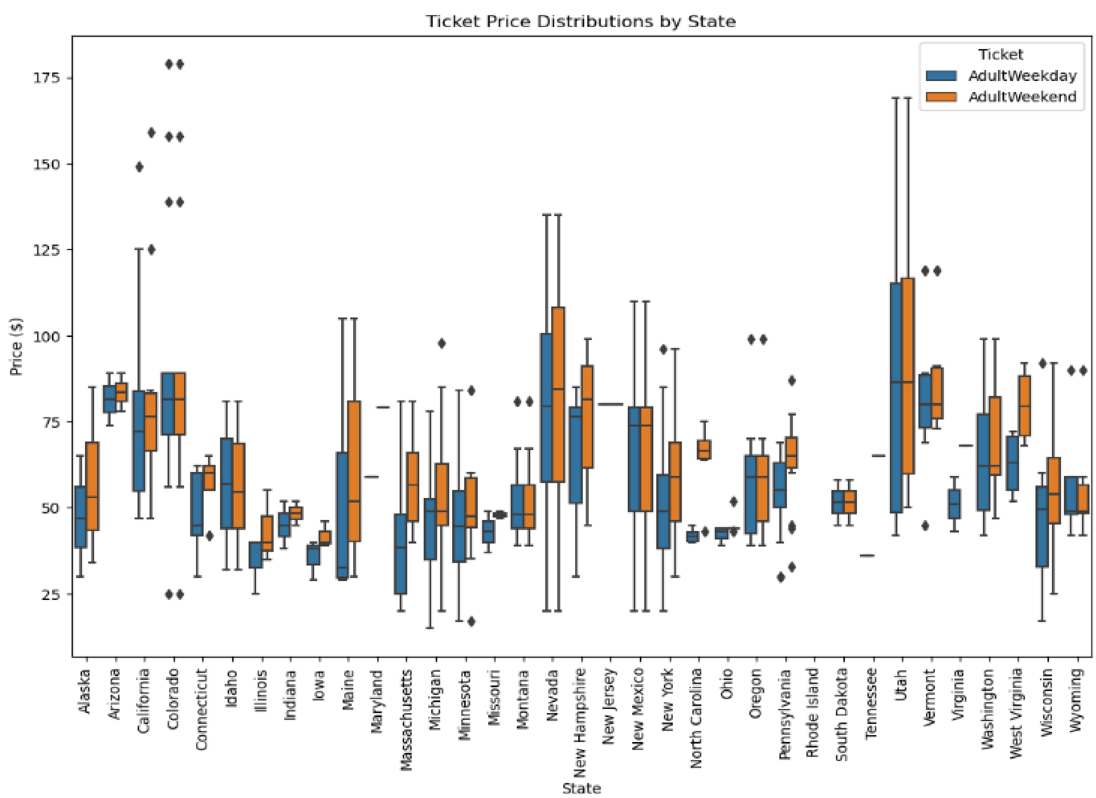
To begin, we saw that most of the resorts were located in New York.



We looked at the distribution of average ticket prices:



Then went deeper by visualizing the ticket price distributions for each ticket type by state:

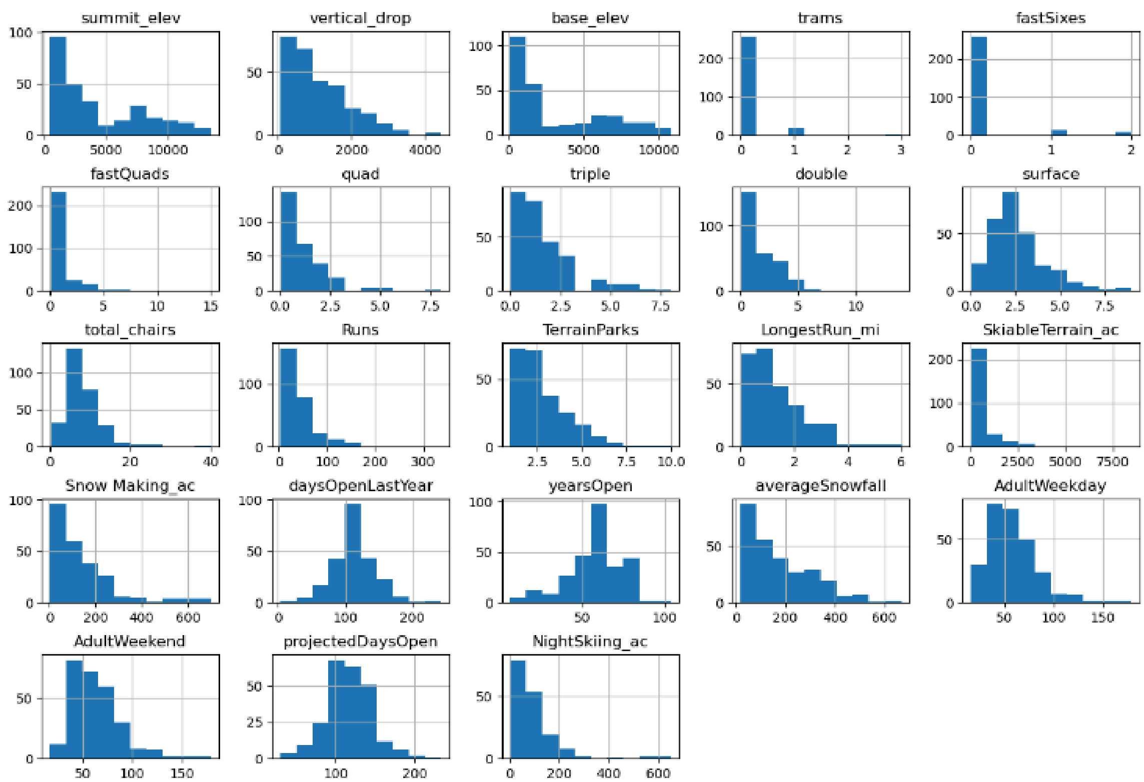


While we knew Big Mountain, and 82% of the resorts overall had no missing ticket price data, we were missing prices for some of the other resorts. Before we deleted these, we wanted to see if we could glean insights from the distributions of their features and subsequently dealt with several outliers and missing values:

- We replaced a plausibly incorrect skiable area at Silverton Mountain of 26,819 acres with the skiable area stated on its website: 1,819.
- Heavenly Mountain resort had an extremely high acreage for snow making and because it didn't have any ticket price data at all, we dropped it.
- We dropped the 'Fast Eight' lift data since all but one value was 0 and nearly half the values were missing.
- We had one resort whose number of years open was 2019. While we could attribute this to someone recording a calendar year rather than the number of years, we couldn't be sure how long this resort had been open, so we dropped it.

Then, before removing anything else, we aggregated certain data relevant to our analysis by state: number of resorts, number of terrain parks, skiable acreage, previous year days open and night skiing acreage, and saved these to a new 'State Summary' database.

Then we removed all the resorts in our original database that were missing all price data and visually reviewed the distributions of the features of the remaining resorts.



We were able to see that some features were still skewed so we made note to be sure the model would not subsequently become overly influenced by those features.

To further enrich our 'State Summary' database, we created summary statistics by state for population and area per square miles (population data downloaded from Wikipedia [List of U.S. states](#)). At this point, we concluded we could drop the ticket price with the most missing data from our original database. Our target feature would be the Adult Weekend ticket price. We then dropped all of the resorts that were missing Adult Weekend ticket prices.

Next, we explored our ‘State Summary’ data and discovered the following:

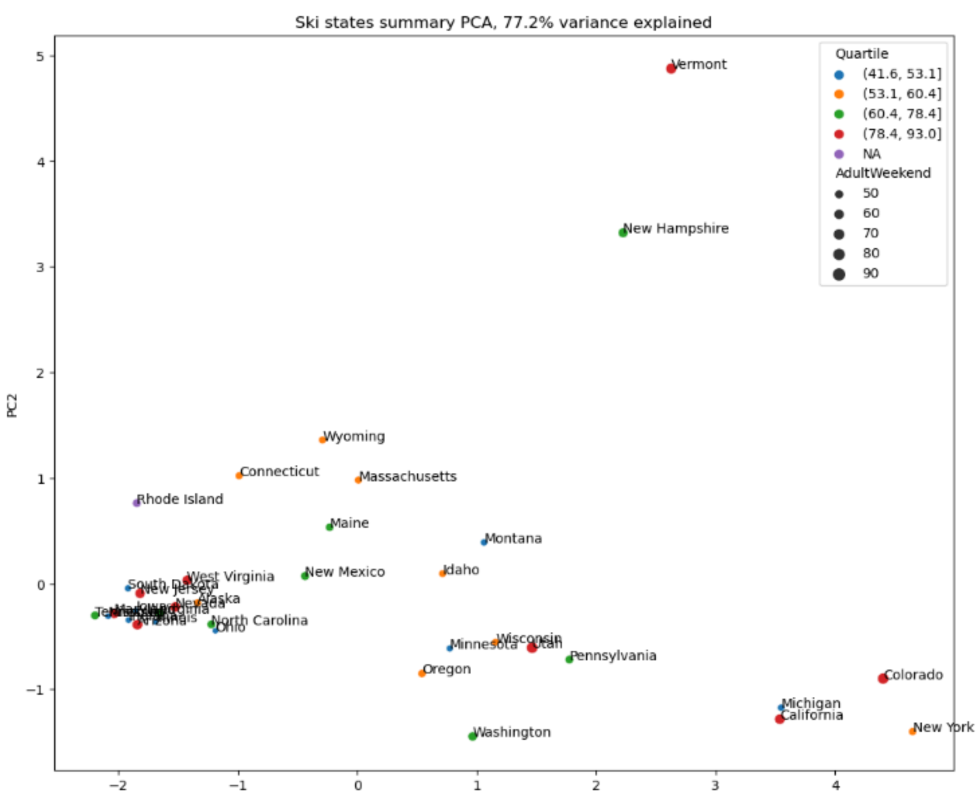
- Colorado had the largest amount of skiable area, most days open and ranked third for number of resorts.
- New York had the most resorts but was not in the top five for skiable area, and offered the most night skiing.
- California had the largest population and ranked fourth for the number of resorts.
- Alaska held the top post for largest area.
- Montana, where our resort in question resides, came in third for the largest area and fourth for total skiable area but did not rank in the top five for any of the other categories.
- New Hampshire made the top five for days open but did not have a large number of resorts.
- Northern states offered the most night skiing.

We were curious why New York had the most resorts and night skiing but not the largest skiable area so we decided to add two new columns to our ‘State Summary’ database, the ratio of resorts to population (per 100k capita) and area (100k square miles) as a means of measuring resort density, and discovered the following:

- Vermont had the highest number of resorts per capita.
- New Hampshire and Vermont had the highest number of resorts per 100k square miles.
- New York did not rank in the top five of either list.

This confirmed that although New York had the most resorts and night skiing compared to the other states, it was not due to high population density or a high number of resorts within a small area.

At this point we were still unable to draw any clear relationships between state and ticket price, so we looked for linear combinations of the original features. We did this using PCA analysis, first scaling the data via standardization. We then visualized this, reducing ticket price into quartiles to display as different point sizes. However, before visualizing, we replaced missing prices for Rhode Island with ‘NA’, rather than removing it, as we were focusing on clusters or trends related to the other features, not price.



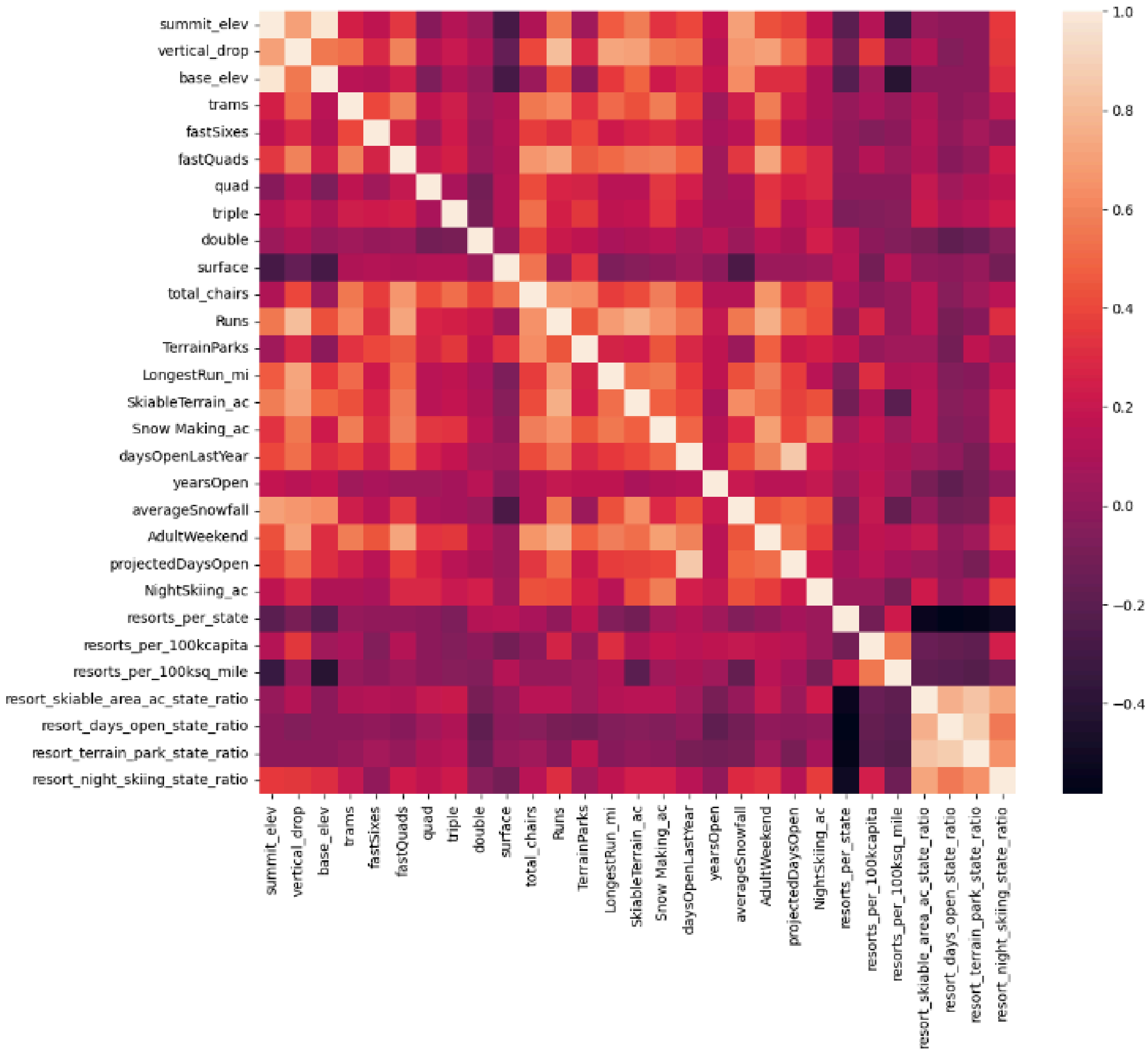
We saw that the states are spread out across component one, again with Vermont and New Hampshire standing alone further out in terms of component two, but no more extreme than Colorado and New York as they relate to component one.

We couldn't find any way to group any of the states together so we decided to treat them all equally in our model. That said, we did discover some potentially relevant state data regarding certain features.

We then merged our original dataset and our 'State Summary' dataset and added the following ratios to understand a resort's share of the supply of certain features for a given state:

- Resort skiable area to total state skiable area
- Resort days open to total state days open
- Resort terrain park count to total state terrain park count
- Resort night skiing area to total state night skiing area.

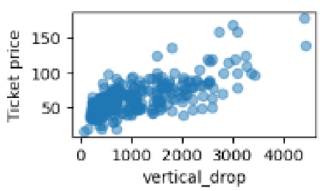
As a first pass at identifying patterns, we visualized the relationships among the features using a heatmap:

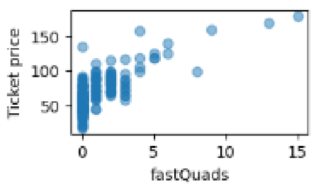


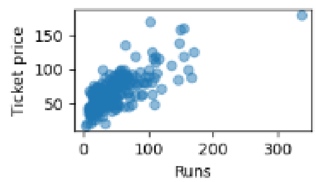
We saw that the following features were highly correlated with ticket price:

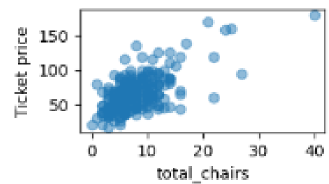
- Vertical Drop and Fast Quads were positively correlated: The higher the drop, the higher the ticket price.
- Runs was closely correlated with Total Chairs: More lifts are needed to move people across more runs.
- Of the ratioed features, Resort Night Skiing to State seems to be the most correlated with ticket price; capturing a greater share could justify a higher ticket price.
- Resort Night Skiing to State also showed some positive correlation with Resorts per 100k capita; it seems more night skiing is provided in more densely populated areas.
- Snow Making Acreage was correlated; higher prices would support making more snow.
- Snow Making Acreage proved more valued than Skiable Terrain Acreage, presumably people valued guaranteed snow over more terrain.

We also visualized how ticket price varied in terms of the other features using scatter plots. Again we see Vertical Drop, Fast Quads, Runs, Total Chairs show strong correlations with ticket price:

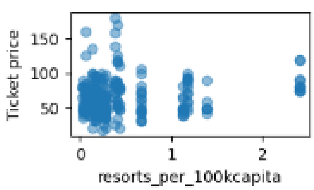






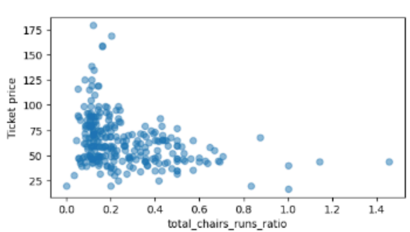


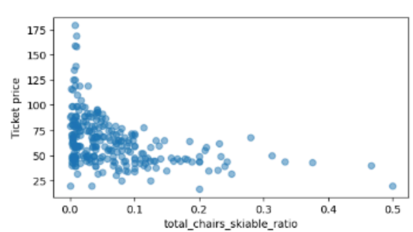
Not seen in the previous heatmap was the variability in ticket price when there are fewer resorts per 100k capita:

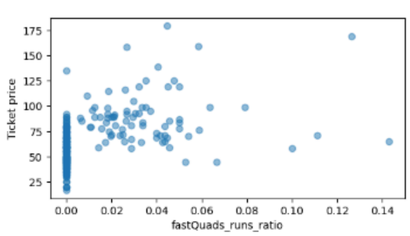


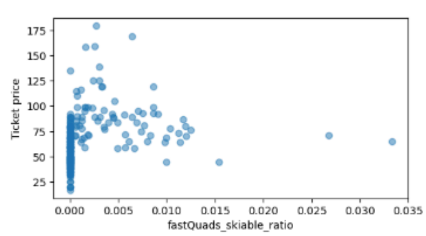
A lower-priced ticket in an area with fewer resorts may be because skiing isn't as popular there. A higher priced ticket in an area with fewer resorts may be because those resorts benefit from a monopoly effect.

Finally, we thought that looking into how the ratio of chairs to runs might affect ticket price might be helpful so we added these to our dataset and displayed them as scatterplots.









These plots showed that more chairs to relative runs did not always mean a higher ticket price. Fewer visitors might cause higher ticket prices in some areas, but we lack data on visitors per year to confirm this. Lastly, fewer fast quads seemed to limit ticket prices but having a few seemed to raise prices if the resort had a larger skiable area.

Next, to gain a baseline idea of the performance of our first model (our comparator for our subsequent models), we determined how useful the mean value would be as a predictor and then used R^2 to measure the amount of variance explained: Performance on the test set was slightly worse than on the training set.

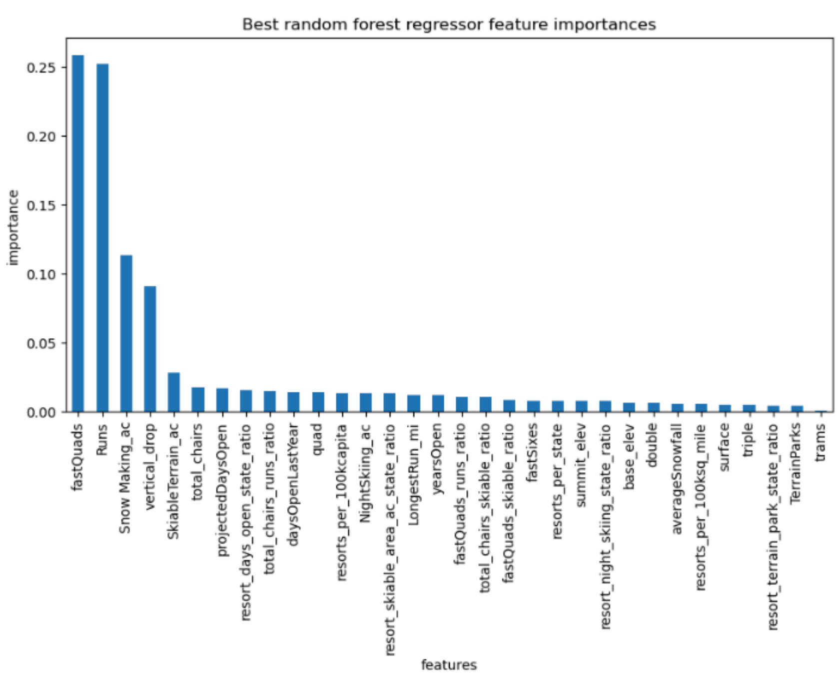
Then, to understand how close our predictions were to the true values, we used the mean absolute error and the mean squared error. The mean absolute error revealed that, on average, we could expect to be off by around $19. The mean squared error, which we converted back to our measurement space by taking the square root, was slightly better on the test set than on the train set.

Next, we imputed missing values using the median and scaled our features to make them consistent. We used linear regression to train the model and R^2 to assess performance: This explained the variance well but the much lower value for the test set suggested some overfitting. We then calculated the mean absolute error score: With this model, we'd generally expect to estimate a ticket price within $9 or so of the real price.

We then imputed missing values using the mean, with results similar to those of the median, and tried feature selection to choose the best 'k' functions, as well as 'score function' with 'f_regression' and used SelectKBest, first with 10, then 15, with similar outcomes: We were tuning the model to the arbitrary test set.

We then ran Linear Regression using cross-validation and saw that Vertical Drop was the top positive feature, just as we saw during our EDA work. Next best was Snow Making Acreage. Skiable Terrain Acreage was negatively correlated: Ticket price was lower for larger resorts, possibly because larger resorts can host more visitors and can therefore charge less but we don't have data on number of visitors so we can't confirm this.

Then we tried a Random Forest Regressor with cross-validation and various hyperparameters to explore using different values for the number of trees, with and without scaling and both mean and median to impute missing values: Median imputing helped but scaling the features didn't. This bar plot revealed the top features, which correspond to those of our linear model:



Because the random forest model had a lower mean absolute error by almost $1 compared to the linear regression model and exhibited less variability, we decided to proceed with the random forest model for further business modeling. Performance on the test set was consistent with the cross-validation results. Collecting more data was not advisable since the learning curve function leveled off quickly at a sample size of 40-50.

We were now ready to use our model to see what price Big Mountain's facilities might support and to explore making changes to various resort parameters.

Big Mountain currently has an $81 ticket price. Our model predicted a ticket price of $95.87. The mean absolute error of $10.39 implies that Big Mountain currently has their ticket price set too low.

To recap, the following features were deemed important during our modeling:

- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

We therefore wanted to know Big Mountain's position with respect to the other resorts for each of these features so we visualized this in a set of histograms and were able to draw the following conclusions:

- Big Mountain is on the higher end of the distribution for price and the highest priced resort in Montana.
- It has a higher vertical drop than most resorts.
- It ranks high compared to the other resorts regarding snow making.
- It sits at the high end for the number of chairs (there are resorts with more chairs but these are outliers).
- Most resorts have no fast quads but Big Mountain has three.
- It has a lot of runs with only a few resorts having more.
- Big Mountain has one of the longest runs.
- It has no trams like most of the resorts (there are only three with trams).
- It is one of the resorts with the most skiable terrain.

We then learned that visitors generally ski for five days and that Big Mountain leadership has four possible scenarios for cutting costs or increasing revenue:

1. Permanently close up to 10 of the least used runs with no impact to any other resort statistics.
2. Increase the vertical drop by 150 feet and add a chair lift, without additional snow making coverage.
3. Same as number 2, but add 2 acres of snow making coverage.
4. Increase the longest run by 0.2 miles to achieve 3.5 total, with added snow making coverage of 4 acres.

Our model determined the following outcomes for each scenario:

1. Closing one run makes no difference but closing more runs created variable loss in ticket price support.
2. Increase ticket price by $1.99, which would amount to $3,482,500 in additional revenue over the season, more than enough to cover the added operational cost of another chair lift.
3. Same as number 2 but with the added cost of making snow.
4. No difference in ticket price.

We can therefore discuss our findings with the business leadership: Our model's predicted ticket price of $95.87, with a mean absolute error of $10.39, means that even a conservative increase in price to $85 could be considered feasible.

The annual cost of $1,540,000 for the newly installed chair lift could be recovered by simply increasing the ticket price to $82, generating annual revenue of $1,750,000 (350,000 visitors per year skiing an average of five days).

Looking forward, increase the vertical drop by 150 feet and add a chair lift, with a ticket increase of $1.99, which would be expected to amount to $3,482,500 over the season, more than covering the chair lift's operational cost.

Finally, to reduce expenses, one run could be closed, leaving the ticket price as is, with no impact on revenue.

To improve on our analysis, additional information would be helpful: While we do know the recent installation of an additional chair lift will increase operating costs by $1,540,000, we don't have information on total operating costs.

Additionally, Big Mountain's modeled price was substantially higher than its current price. Our modeling included prices set by other resorts and assumed these were set accurately. That said, the higher price may have been due to scoring fairly high on multiple features compared to other resorts. This might come as a surprise to the business executives and increase their confidence that implementing a price increase could produce a successful outcome.

Finally, given that we have come up with a model we believe could be useful to the business going forward, we could explore the possibility of incorporating this model into a dashboard so future combinations of parameters could be run as needed.