When beginning this project we started with the problem statement “How can Big Mountain Resort improve profits by more than $770,000 (50% of cost) in this upcoming winter season to offset the costs of a new chair lift? Initial thoughts are to increase ticket prices (assuming 350,000 patrons each purchase 1 ticket this year we would need to increase price by $2.20/ticket) and/or lower expenses by 5-10% by closing underused or unused facilities. In order to make sure we are on track we will be continuing to monitor the situation through financial reports and customer surveys through revenue/profit numbers and customer satisfaction(determined by whether customers would be happy to return).” This allows us to have a clear goal which we can work towards throughout the project.

To begin analysis we had to clean the data making it usable. Initially the data had 330 rows and 27 columns. We removed the fastEight column as it had no data in just over 50% of the rows and in the remaining 50% only had one non-zero entry making. Next we removed a row which had “2019” as the value in the Years open column as we are not sure if that means the resort opened in 2019 or is opening in 2019 as we are not sure when this data was collected. Finally we removed the rows which had no data in both the AdultWeekend and AdultWeekday columns as without either of these we have no clue on the price of the resorts making them useless data. At the end we are left with 277 rows and 26 columns of data to work with going forward.

To begin EDA we scaled and fitted the data showing that the first two components(the two most important features) accounted for 75% of the variance and the first four accounted for over 95%. With this knowledge we plotted a graph showing where each state’s scaled value for the first two components in Figure 1. Based on figure 1 we can see that there does not seem to be a correlation between state and price as there are states in the top quartile for price in 3 of the 4 quadrants of the graph. This leads us to build a pricing model that considers all states. Next we created a correlation heatmap (Figure 2) so we could find which features had high correlations to our target feature(AdultWeekend) and concluded fastQuads, Runs and Snow Making were the most important.

In order to begin working on creating a model we first had to split the dataset into two parts, a test set and a training set. Using the training set we worked to create a model which would predict the best model for figuring out what the optimal price would be. After that we would test against the test set to see how the model performed.

The evaluation metrics we used were R2, MAE and MSE. Mainly with a focus on MAE as the lower the MAE the more confident we could be in our model. Our first algorithm was to test the test set against the mean of the training set so we could get a baseline for the three metrics. Next we used linear regression (lm.predict()) to create a model and our final choice for the model was using a random forest regressor. By doing these we found the most important features to be vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, Runs, LongestRun\_mi, trams and SkiableTerrain\_ac.

In the end we ended up choosing the random forest regression as it had a lower MAE (10.499032338015294 for linear regression vs 9.644639167595688 for the random forest regression). After choosing this model we found that based on the current situation Big Mountain should be charging $95.87 however our model has a MAE of $10.39 which is less than the gap between the modeled price and current price meaning we can raise the price. Following this we were told to investigate 4 scenarios:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
3. Same as number 2, but adding 2 acres of snow making cover
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

To investigate these scenarios we wrote a function which would allow us to change the values of the features then plug in these new values into the model to see if the ticket price increased.

For scenario 1 our revenue would be slightly lowered but at the same time expenses would be lowered. Scenario 2 would increase revenue but also increase costs. Scenario 3 should not be considered because in our model it does not increase revenue when compared to scenario 2 but increases costs by adding another change. Scenario 4 should also not be considered as it does not increase revenue but increases costs.

Going forward we should be increasing ticket prices by at least $0.88 (total of $81.88 assuming 350,000 customers each buying 5 tickets and a new lift cost of $1,540,000) to cover costs of the new lift. However it is likely that we can continue to increase it to $85.48 considering our base model price of $95.87 and MAE of $10.39. Next we should consider which, if any of the scenarios are worth considering. Based on the data gained we should probably consider Scenario 1 and Scenario 2. If we are implementing Scenario 1 it should be in phases of 1 run closed, 2 runs closed, 5 runs closed, 8 runs closed, 9 runs closed, 10 runs closed. This is because with 3,4, and 5 runs closed the ticket price and therefore revenue decrease is projected to be the same. A similar situation occurs with 6,7 and 8 runs closed thus there is no point in closing only 3,4,6 or 7 runs when we could close more to further lower expenses while losing the same amount of revenue.

Finally we should have a conversation with finance to make sure the numbers are correct and based on the expenses and projected revenue generate Big Mountain Resort a greater profit going forward.

Assuming this model is useful we should continue to use it at an annual strategy meeting to decide whether new runs be added or old runs removed or kept the same, and whether adding more facilities would be profitable. I would expect a business analyst or perhaps someone in the finance team to learn to use the model and test the various combinations of parameters which would be discussed in the resort's plans going forward. Probably the best way to make this model available to others would be to create a user guide on how to change variables in the model and what exactly the various outputs mean. Then record a quick video going over it for those who would prefer to watch instead of reading a guide.

Figure 1

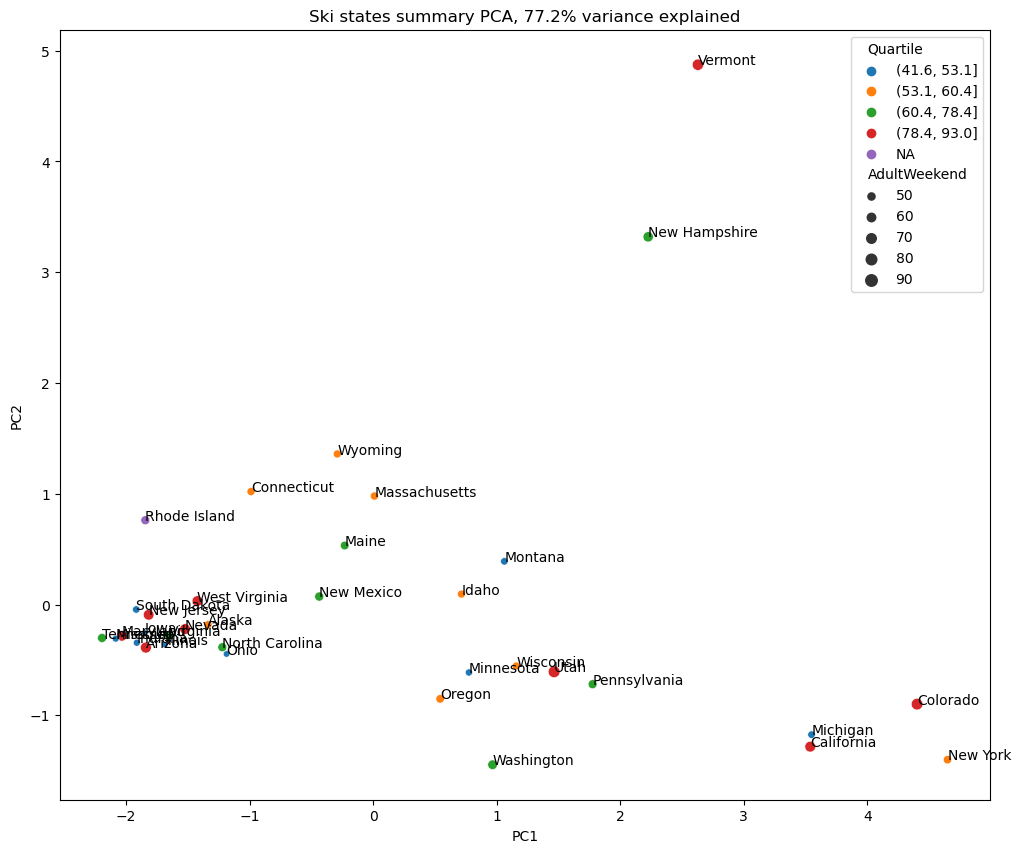


Figure 2

