Capstone Project Report

Hio Wa Mak

04/09/2024

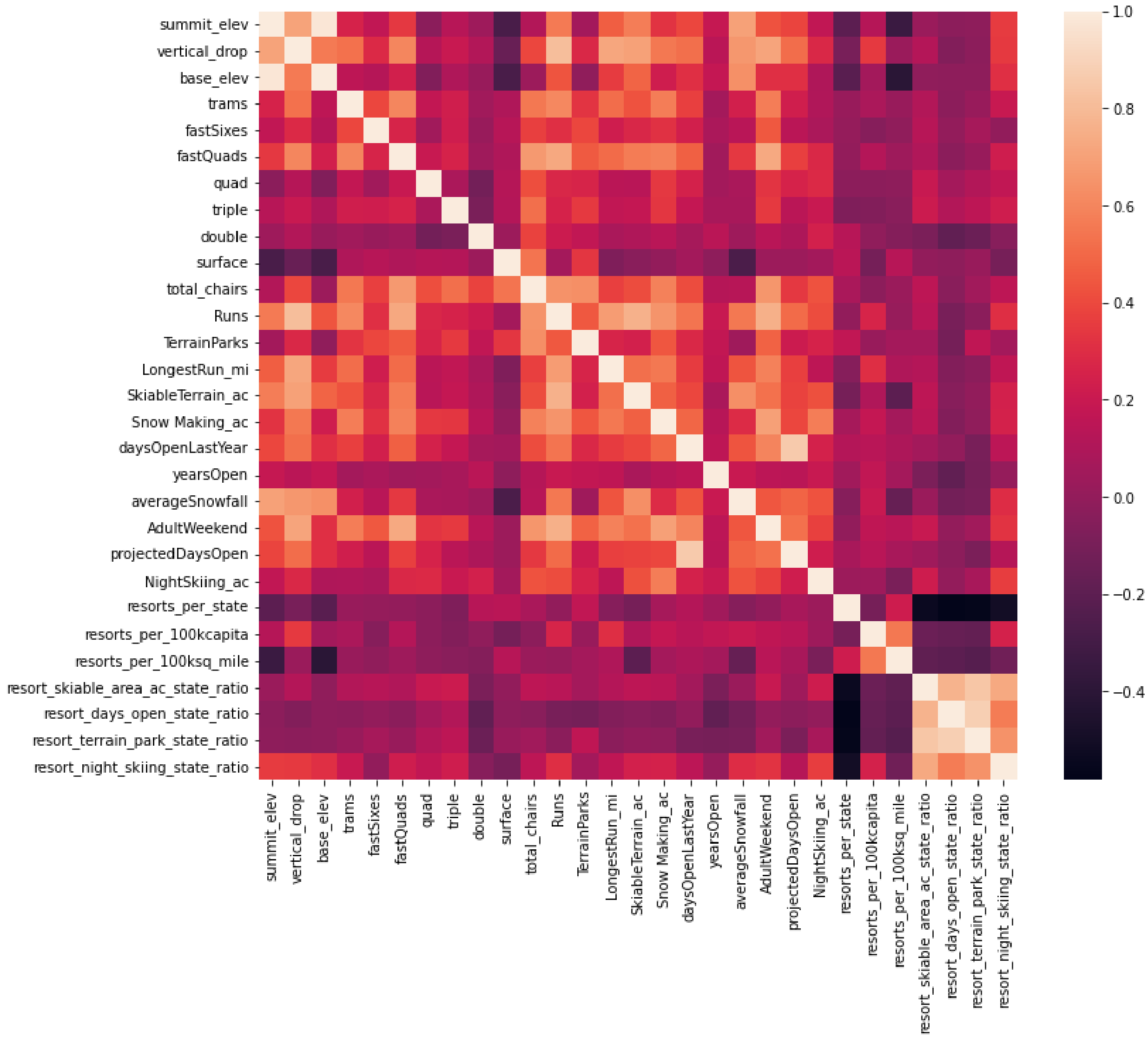
Big Mountain Resort (BMR) has recently installed an additional chair lift that cost 1,540,000 in operating cost this season. The company wants to improve its pricing strategy to increase revenue, including a better pricing strategy and find areas to cut costs.

The big question for Big Mountain Resort is: How can Big Mountain Resort improve pricing strategies to generate at least $2,000,000 revenue by the next season through identifying specific facilities that are the most attractive to customers (i.e., the most predictive of ticket price)?

I received a dataset from Alesha Eisen, the Data Manager, with information on ski resorts across the United States. The dataset contains information regarding resort facilities and two types of ticket price: adult weekday and adult weekend ticket price. I conducted some data wrangling and examined the two types of ticket price. They are highly positively correlated, and I decided to use the adult weekend ticket price as the target feature to predict because it has less missing data. The goal of this project was to identify features that are the most predictive of ticket price. I examined the distributions of certain features including resort facilities and skiable areas and corrected some data entry errors. I also imported a dataset from Wikipedia that contained state-level information and merged the two datasets for subsequent processing and modeling.

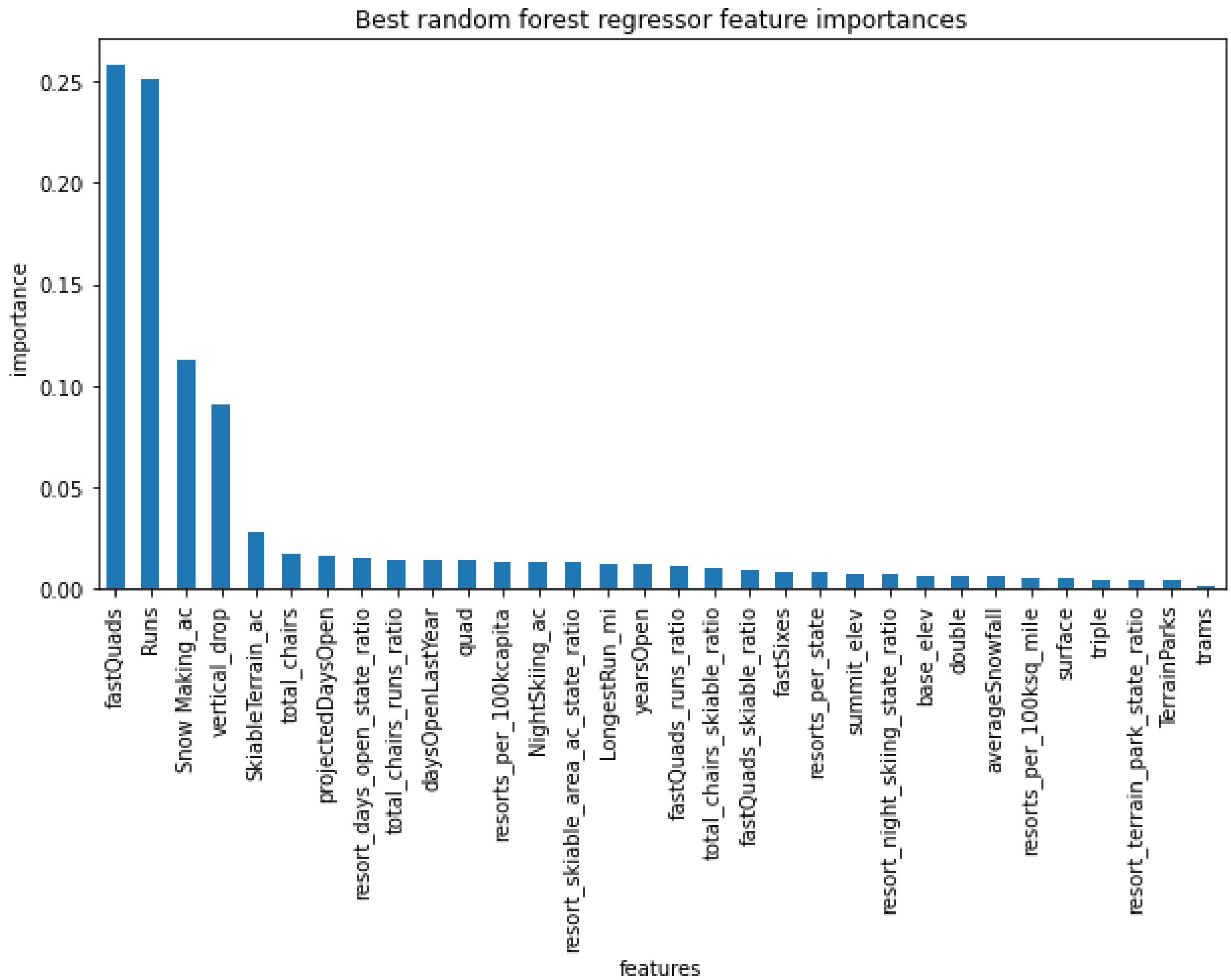
In the next step, I ran a principal component analysis (PCA) to simplify the state-level information and summarize them using two components. However, there was not a clear pattern of how the two components were related to ticket price, therefore, it seems reasonable to treat all the states equally. Regarding the resort-level data, I created additional variables that indicate resort characteristics relative to the respective state and then I correlated all the variables with ticket price (see Figure 1). There were some key features that were quite correlated with ticket price, including the vertical drop, the number of fast four person chairs (fastQuads), the number of runs, the number of trams, the total number of chairlifts (total chairs), skiable area, area covered by snow making machines, and night skiing area. Nevertheless, I do not have information regarding the number of tickets sold per year at each resort, which would be very useful information on revenue.

Figure 1. Heatmap of correlations in the ski resort data



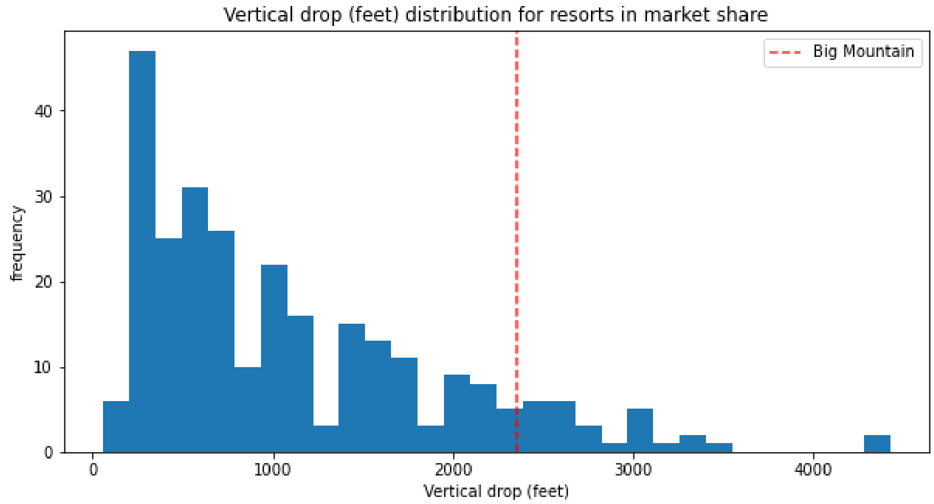
Before applying machine learning models to predict ticket price, I conducted a baseline model for comparison using the mean ticket price as the predictor of ticket price. Results suggest that using this model to predict ticket price is expected to be off by $19. Then I fitted two models using pipelines: a linear model and a random forest model. The linear model, with 5-fold cross-validation, and tuning hyperparameter using GridSearchCV, suggests that there are 8 important features in predicting ticket price: the vertical drop, the area covered by snow-making machines, total number of chairlifts, the number of fast four person chairs (fastQuads), the number of runs, the longest run, the number of trams, and the area of skiable terrain. The mean absolute error was used to assess model performance, which was about $10.5. The performance on the test set was consistent with the training set. For the random forest model, with 5-fold cross-validation, median imputation, and GridSearchCV for tuning hyperparameter, I found the following features most important in predicting ticket price: fastQuads, the number of runs, the area covered by snow-making machines, and the vertical drop (Figure 2). The mean absolute error was about $9.6. As the random forest model has a smaller mean absolute error by almost $1, is more parsimonious (fewer features to target), and has less variability, I chose this model as the final model. Results suggest using this model to predict ticket price is expected to be off by $9.5, which is a huge improvement compared to the baseline model with a mean absolute error of $19.

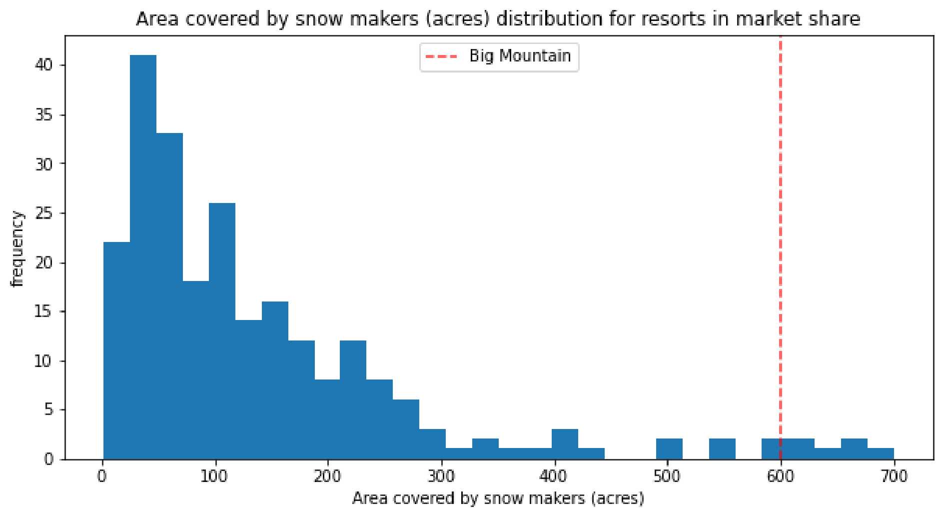
Figure 2. Feature importances based on the random forest model.

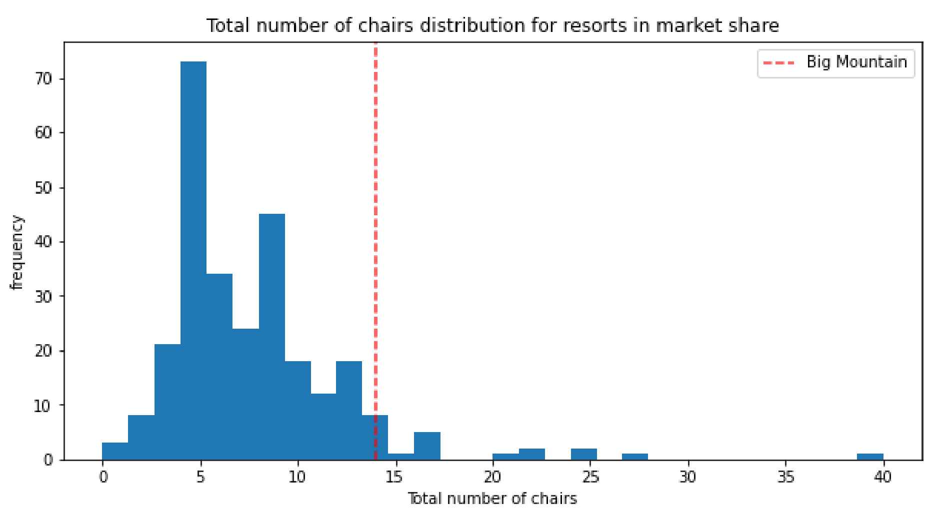


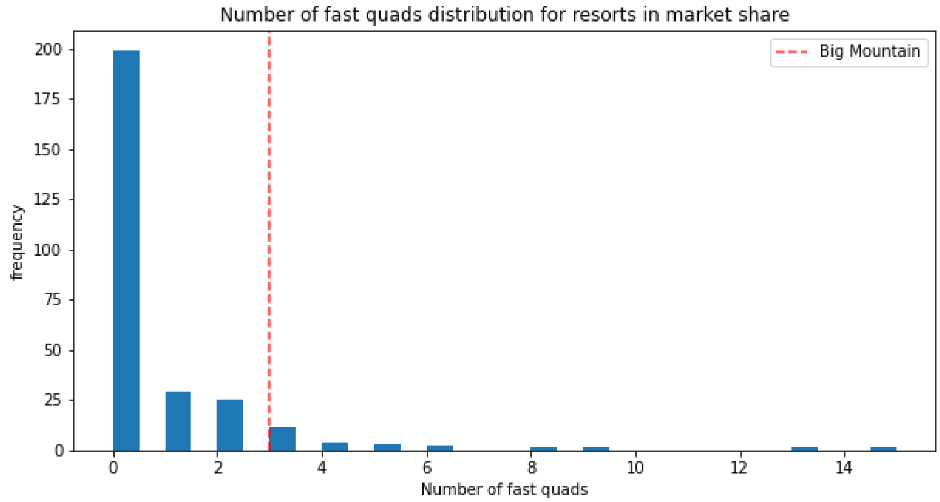
Big Mountain Resort is currently charging $81 for an adult weekend ticket. I used the random forest model to predict Big Mountain Resort ticket price and results suggest that there is room for an increase to $95.87, with an expected mean absolute error of $10.39 based on the current facilities in Big Mountain Resort. Big Mountain's current facilities (e.g., the vertical drop, the area covered by snow-making machines, the total number of chairs, the number of fast quads, the number of runs, the longest run, and the area of skiable terrain) are much above average compared to their competitors in the market (Figure 3).

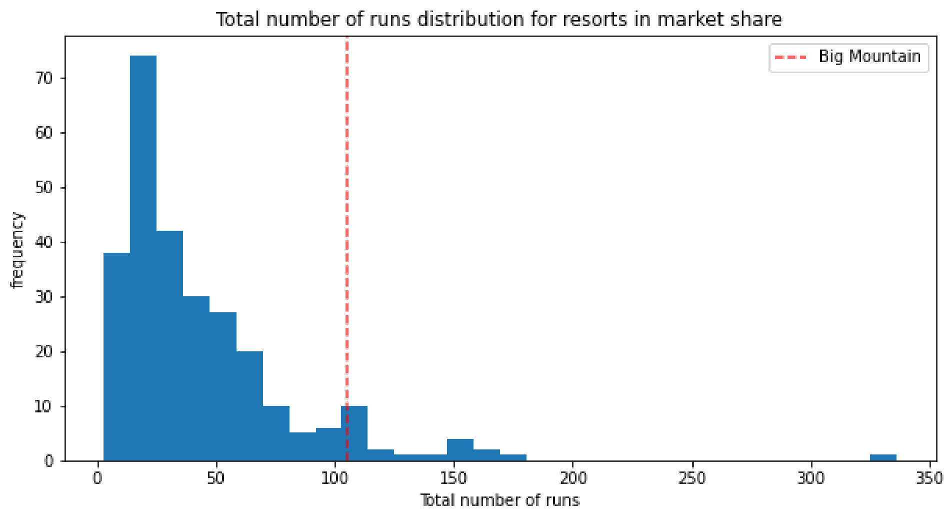
Figure 3. Big Mountain Resort facilities compared to competitors in the market.

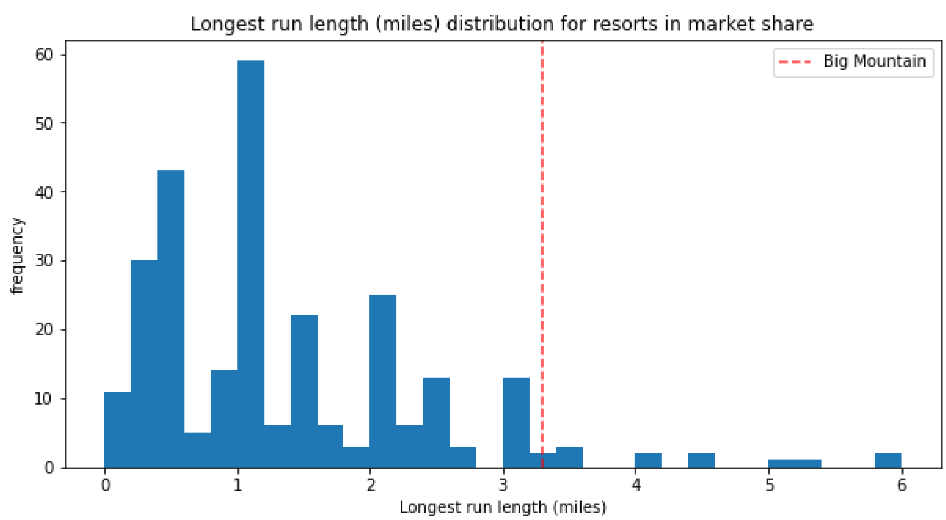

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet)

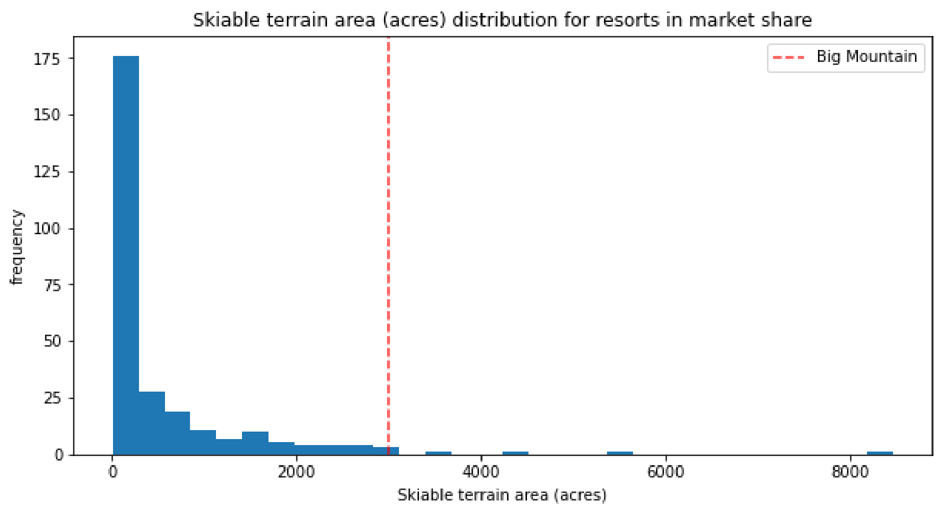

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
Area covered by snow makers (acres)


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
0
2
4
6
8
10
12
14
0
25
50
75
100
125
150
175
200

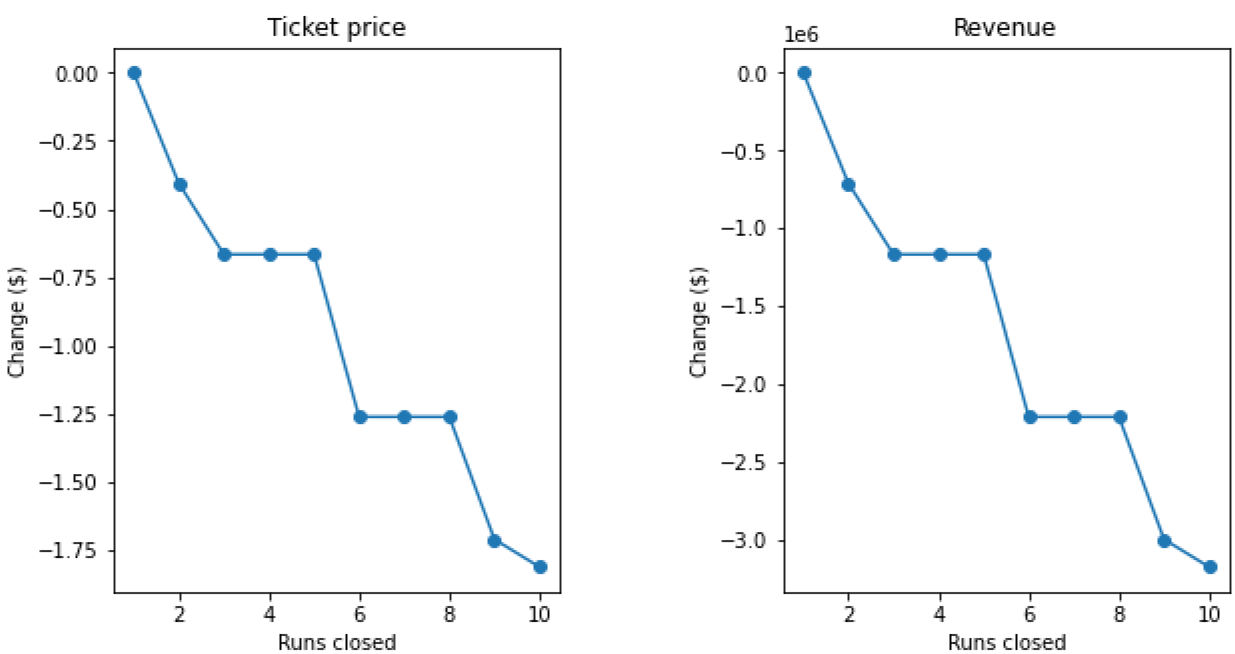

Total number of runs distribution for resorts in market share
Big Mountain
frequency
Total number of runs
0
50
100
150
200
250
300
350
0
10
20
30
40
50
60
70


Longest run length (miles) distribution for resorts in market share
Big Mountain
frequency
Longest run length (miles)
0
1
2
3
4
5
6
0
10
20
30
40
50
60



Based on the four shortlisted options for future improvement, I modeled the change in ticket price and the associated change in revenue. As for Option #1, based on the model, closing one run will not affect ticket price and revenue, so this could be done to reduce cost. However, the model suggests that closing 2 and 3 runs will reduce revenue, but if one decides to close 3 runs, one may just close down 5 without further loss in revenue (Figure 4). This decision will depend on the cost in maintaining each run.

Figure 4. Closing runs in predicting change in ticket price and revenue.



As for Option #2 (Increase the vertical drop by adding a run to a point 150 feet lower down and the installation of an additional chair lift), it would support an increase in ticket price by $1.99, bringing about revenue increase to $3,474,638 over the season (based on the assumption that there are 350,000 visitors and each visitor on average buy 5 tickets), which could cover the additional operating cost of $1,540,000 for a new chair. Option #3 (Same as Option #2, but add 2 acres of snow making cover) and Option #4 (increase the longest run by .2 miles) both do not affect revenue compared to Option #2, thus do not appear to be good investment. If Big Mountain Resort decides to take one or more actions based on the suggested options, continued monitoring of ticket price and revenue is recommended. I have developed modeling pipelines and visualization codes that could easily be updated with future data from Big Mountain Resort and other competitors to inform Big Mountain Resort business decisions.

It should be noted that there are some limitations to this predictive model. First, there is an assumption that the ticket price in other resorts was based on how much customers valued certain facilities. Second, the weekend ticket price is the only useful target feature to model in the dataset, we have no information regarding how many tickets were sold for each resort in the market. Further, I only have information regarding additional operating cost of a new chair lift, but do not have information on the operating cost for keeping a run open, which would be an important cost information to decide whether cutting down more than one run is worth doing. Moreover, I only have information regarding the total number of runs, thus, the model assumed that all the runs are equally important to customers and that they are equal in cost to manage. Although I suggested an increase in ticket price, it should be noted that Big Mountain Resort's ticket price is already the highest one in Montana, so the cost of living in Montana as well as the competition within the state should also be considered. I have developed the modeling pipeline and visualization for modeling ticket price, this tool could be used by business analysts to explore different scenarios in the future to inform Big Mountain Resort business decisions.