Big Mountain Resort: A Ticket to Ride

Problem Statement

Big Mountain Resort is a popular ski resort in Montana, with stunning views of Glacier National Park and Flathead National Forest. The resort hosts about 350,000 skiers and snowboarders each season. To help improve distribution of visitors across the mountain, Big Mountain added a new chair lift. While this is expected to improve our guests' experience, the additional chair lift increases operating costs by $1,540,000 this season. We are thus tasked with finding a solution to offset this cost. Our options for doing so are to increase ticket prices from our current price of $81, to reduce costs by cutting down on other facilities, or some combination therein.

Data Wrangling

To solve this problem, I worked with a data frame containing data on ski resorts throughout the US, including Big Mountain, so as to compare ticket prices and various features. A fair share of data cleaning took place, reducing the initial data frame from 330 rows x 27 columns to 277 rows x 25 columns. This was primarily due to the removal of missing information and entailed the correction of a few extreme values. Rows corresponded with ski resorts and columns corresponded with features: ticket price, number of runs, vertical drop, number of chairs, etc. Some initial considerations included amassing some state-wide statistics (Figure 1) and looking at distributions of various features across all resorts (Figure 2). Upon calculating state-wide summaries, I combined ski data with additional state data including population and total area, for a broader view of each state. While some resorts charge different ticket prices depending on whether it is a weekend or weekday, I chose to universally use weekend ticket price as a price metric, as this indicates the higher threshold that resorts charge.

Exploratory Data Analysis

I delved further into analyzing state-wide summaries, and created various ratios for each state, such as the number of resorts per state relative to the state's population, and the amount of skiable area per state relative to total state area. I conducted a principle components analysis to explore features that explain the most of the variance in the state summary dataset, as indicated by components that represented consolidation of the most influential features. I scaled all data values to unit variance from the zero mean, and found that the cumulative sum of the first two components accounts for 77% of the variance among state summaries. Upon looking at the distribution of states along each of the first two components, and their relative average ticket price ranges (Figure 3), there was not a discernible pattern. I thus moved on from the state-based summaries and turned to an analysis of direct correlations between features and ticket price across all resorts (Figure 4). This view demonstrated that the features that had the strongest correlation with ticket price were fast quads, runs, vertical drop, area covered by snow making machines, and total chairs.

Model Preprocessing

Before training models, I divided the data up. I separated Big Mountain into its own set, as models were to be trained and tested on data from the rest of the resorts. I then divided the data from the remaining resorts into a train / test split on a 70 / 30% ratio of all the data. This would enable training on one set of data, and testing on data the machine hasn't yet seen, to ensure its generalizability. I then created a baseline estimate of ticket price based on mean.

A model's ability to predict ticket price can be measured by its coefficient of determination, R^2, an indicator of the proportion of variance explained by the independent variables. This was used to help fine tune model parameters to most effectively predict ticket price. I created pipelines for a linear regression model and a random forest regressor. The pipeline would as needed impute missing values based on median, scale data to zero mean and unit variance to compensate for different magnitudes based on different measurement units for different features (note that the linear regression model used feature scaling while a random forest model did not), incorporate a selector for the most influential features, train the model, and cross validate over multiple folds of the training data set.

The linear regression model suggested 8 features that strongly correlate with ticket price: vertical drop, area covered by snow making machines, number of chairs, number of fast quads, number of runs, longest run, total area of skiable terrain, and number of trams. The random forest model suggested 4 primary features that correlate with ticket price: fast quads, runs, area covered by snow making machines, and vertical drop (Figure 5). I compared the prediction results of the two models by looking at each model's mean absolute error, which demonstrates the margin of error of the model's prediction. The random forest model had a mean absolute error that was more than $2 less than the linear regression model, and it also demonstrated lesser variability. Thus, the random forest regressor was the choice model for predicting ticket price based on features.

Modeling

After fine-tuning our random forest model and fitting it to the entire ski data set except for Big Mountain, I used it to predict Big Mountain's ticket price based on our features and, low and behold, the estimated ticket price is $95.87. Even considering the model's mean absolute error of $10.39, this puts our estimated ticket price well above our current price of $81. Our current underpricing might be due to the fact that we are already among the most expensive resort in Montana, and among the higher end of the distribution for ticket prices among all resorts (Figure 6). Nevertheless, we are also toward the top of the distribution for key features that increase a ticket's value, including vertical drop (Figure 7), area covered by snow making machines, number of chairs, number of fast quads, number of runs, and longest run. Therefore, a ticket price increase is justified.

Using the model with a function that adds or subtracts counts of a given feature from that feature's current count, enables the customization of a given standard for each feature. This allows us to compare how increasing or decreasing given features will affect an estimated ticket price. Considering the options that the business suggested of (1) closing up to 10 runs, (2) increasing vertical drop by 150 feet and thus installing an additional chair, (3) same as option (2) plus adding 2 additional acres of snow making coverage, and (4) increasing the longest run by 0.2 miles and thus requiring an additional 4 acres of snow making coverage, I modeled all options and found that the first two may be valid, while the latter two are not.

For option (1), though closing runs will indeed lead to a lower estimated value of ticket price (Figure 8), the subsequent loss in seasonal revenue (considering 350,000 visitors who each buy 5 tickets) of $1,116,667 for closing 5 runs and $2,206,521 for closing 8 runs, might be offset by the money saved from lesser operating costs. As indicated in Figure 8, closing 5 or 8 runs were chosen as the most advantageous numbers of runs to close, as neither had an associated drop in ticket price from 3 or 6 runs respectively, while other numbers of run closures were associated with further price drops. It should also be noted that if run closures lead to reduced operating costs by enabling the elimination of chairs or snow making area etc., further modeling will be necessary to factor in the lesser number of chairs, snow making area, etc. in determining the subsequent ticket price estimate.

For option (2), the model determined an increased ticket price by $1.99, resulting in a seasonal ticket revenue increase of $3,474,638. As this is over twice the operating cost of a new chair, this might also be a viable option.

Conclusion

This analysis compared ticket prices and features of resorts across the nation, and used a random forest regressor to estimate an appropriate ticket price based on a resort's given features. This approach works under the assumption that ticket price is indeed based on valuation of features. Given that Big Mountain is at the higher end of the distribution for the features that both the linear regression and random forest models deemed most relevant for predicting a ticket price, it is reasonable to consider increasing ticket prices. The business proposed a couple a valid solution for cutting costs—reducing the number of runs, or for increasing ticket prices—increasing our vertical drop. While I would suggest looking into either of these possibilities, I also emphasize that the random forest regressor indicates that given our current features, our current ticket estimate based on our features is almost $15 more (give or take $10) than our current ticket price of $81. My primary suggestion is that regardless of what feature changes we consider, we can justifiably raise our ticket prices now, given our current features.

Future Scope of Work

This model provides valuable insight into how a ticket might be priced based on features. By passing the model into a function that compares current features with changes in features, the business can explore various feature / subsequent ticket price combinations. For instance, the business can use the model to predict how the ticket price might change if more chairs were installed, and compare the resultant revenue given from an expected number of ticket sales in a season with the operating cost of additional chairs. This allows for a multitudinous view of scenarios, determining which feature combinations might lead to the greatest revenue increases. There is further information that could help to better design future models, such as number of guests per season and average stay per guest. Should a broader span of data be extracted, more modelling can be done accordingly. Nevertheless, our current model based on the data we have, supports a robust exploration of feature and pricing options.

Appendix

Figure 1: Average ticket price of resorts per state

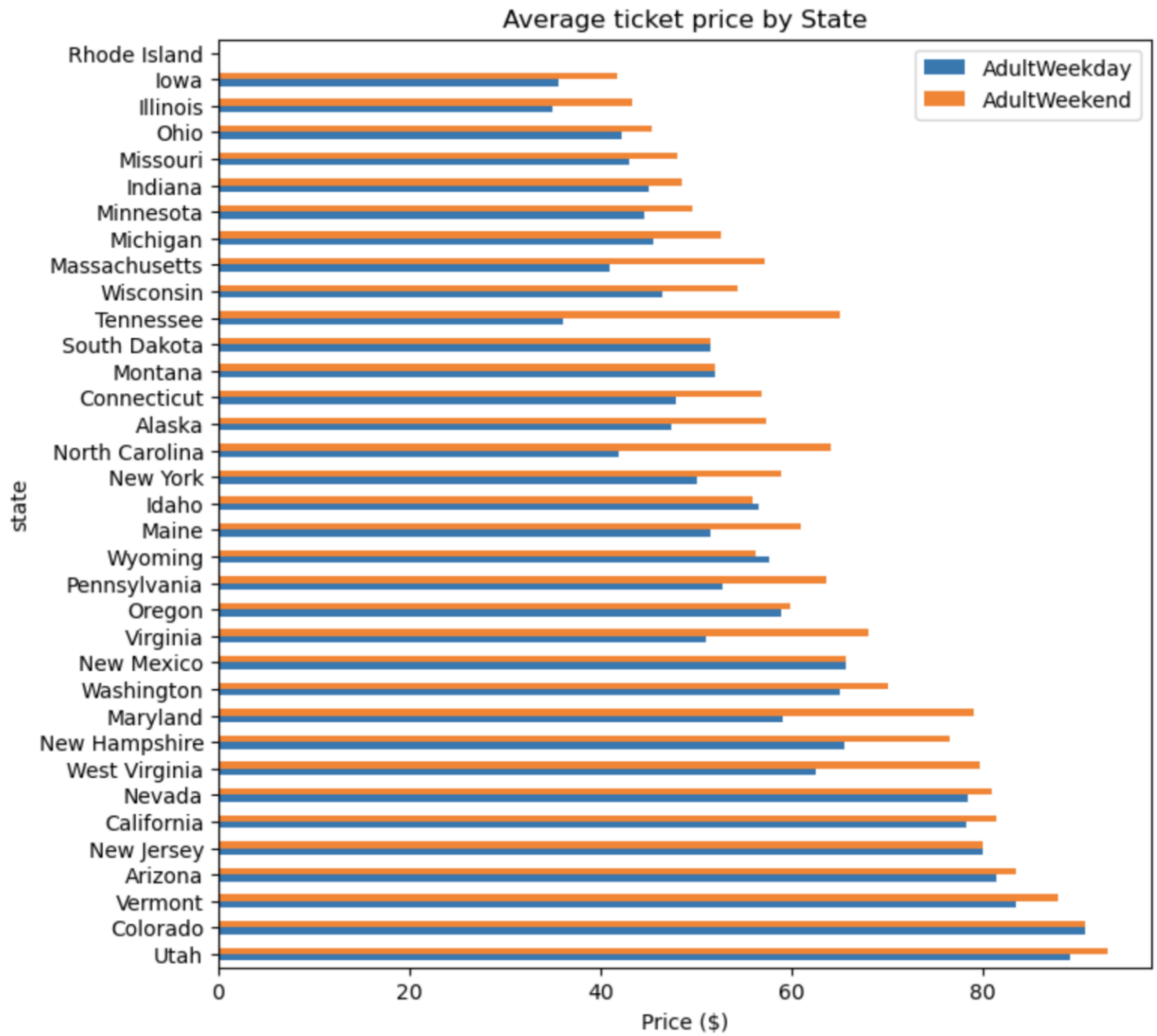

Figure 2: Distribution of features across resorts (after some initial cleaning of outliers)

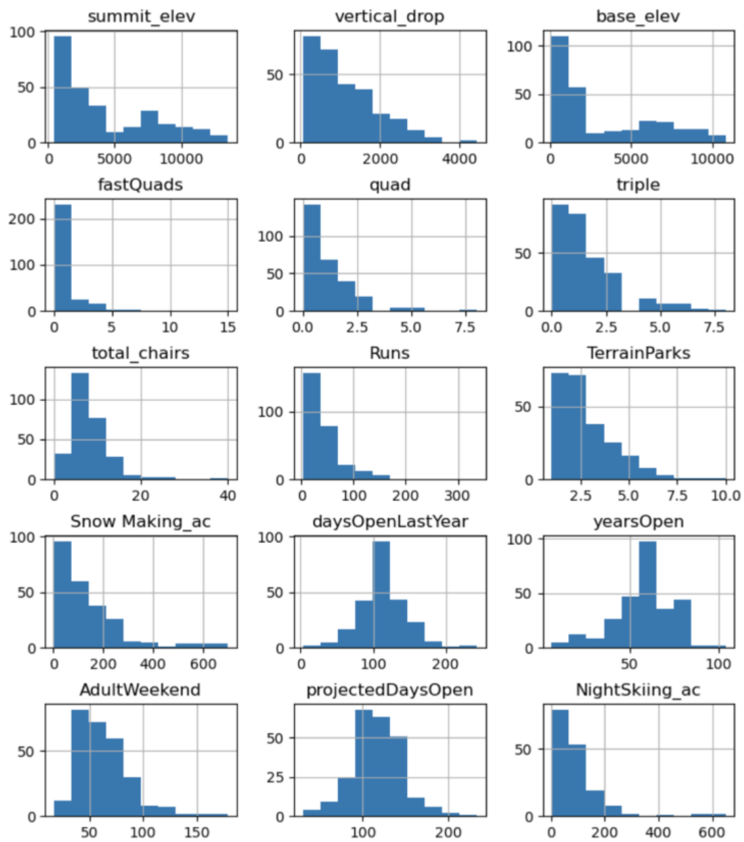

Figure 3: Principle component analysis: feature variance and ticket price quartile per state

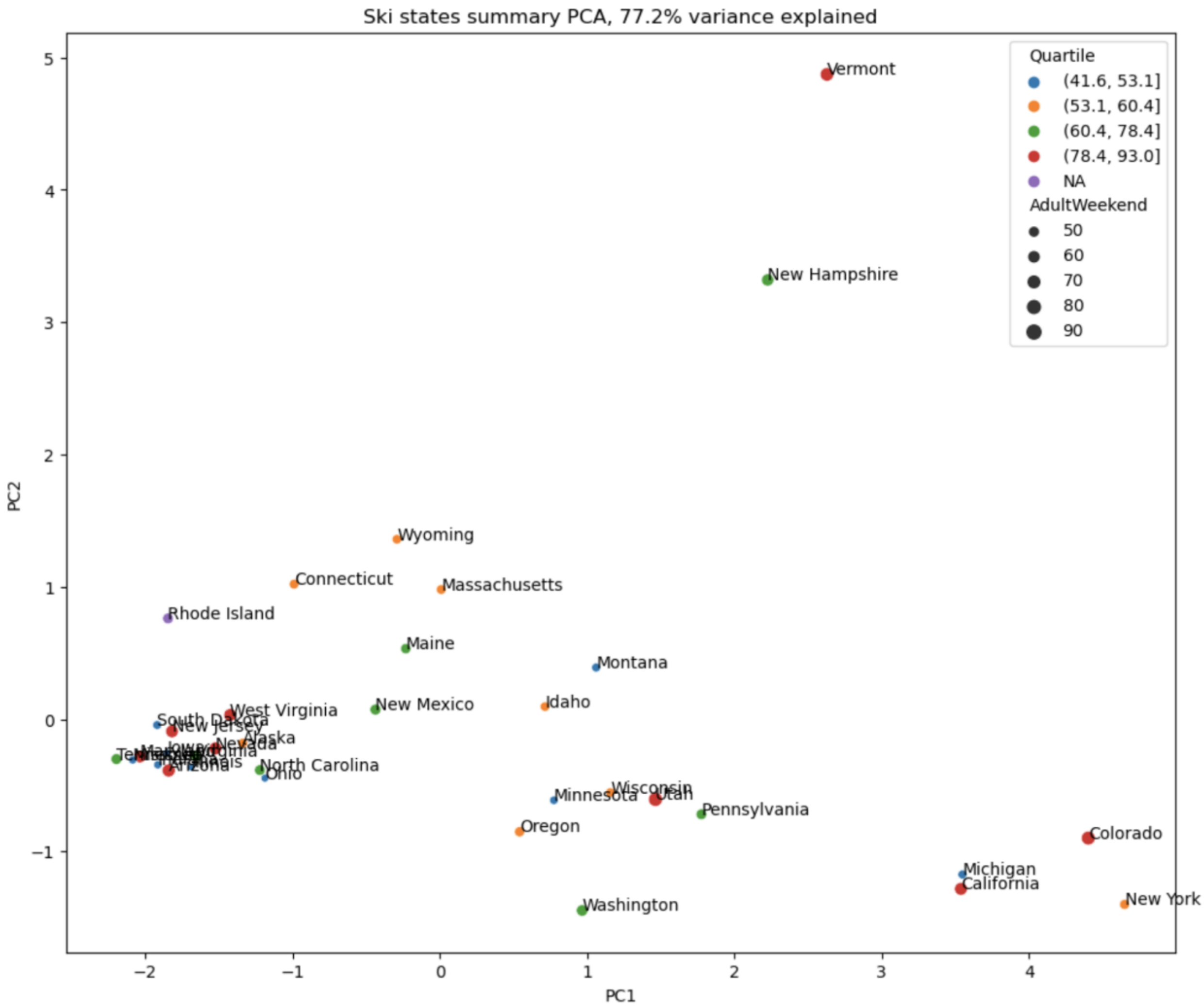

Figure 4: Heat map depicting strength of correlations between feature and ticket price

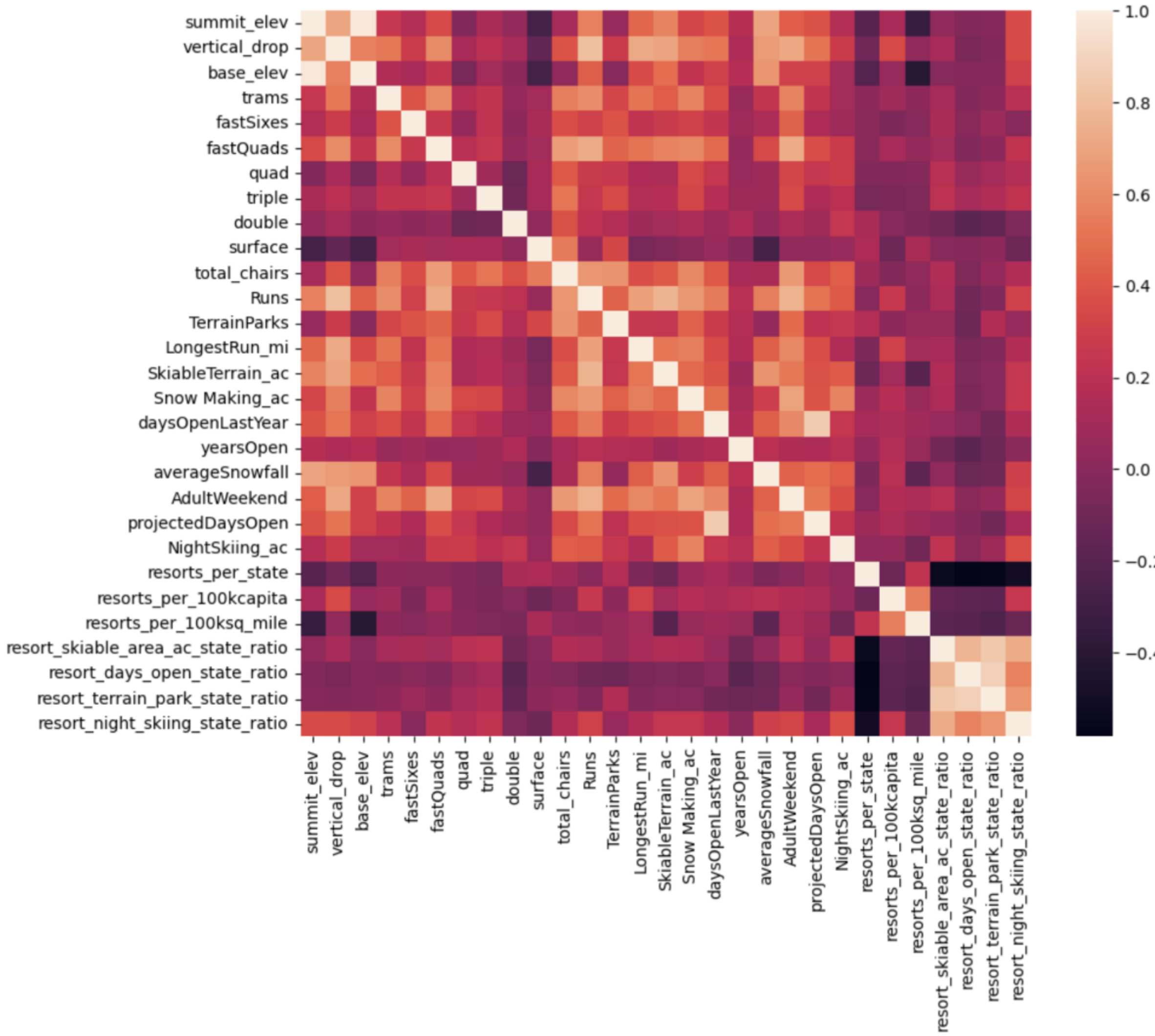

Figure 5: Random forest regressor strength of feature correlations with ticket price

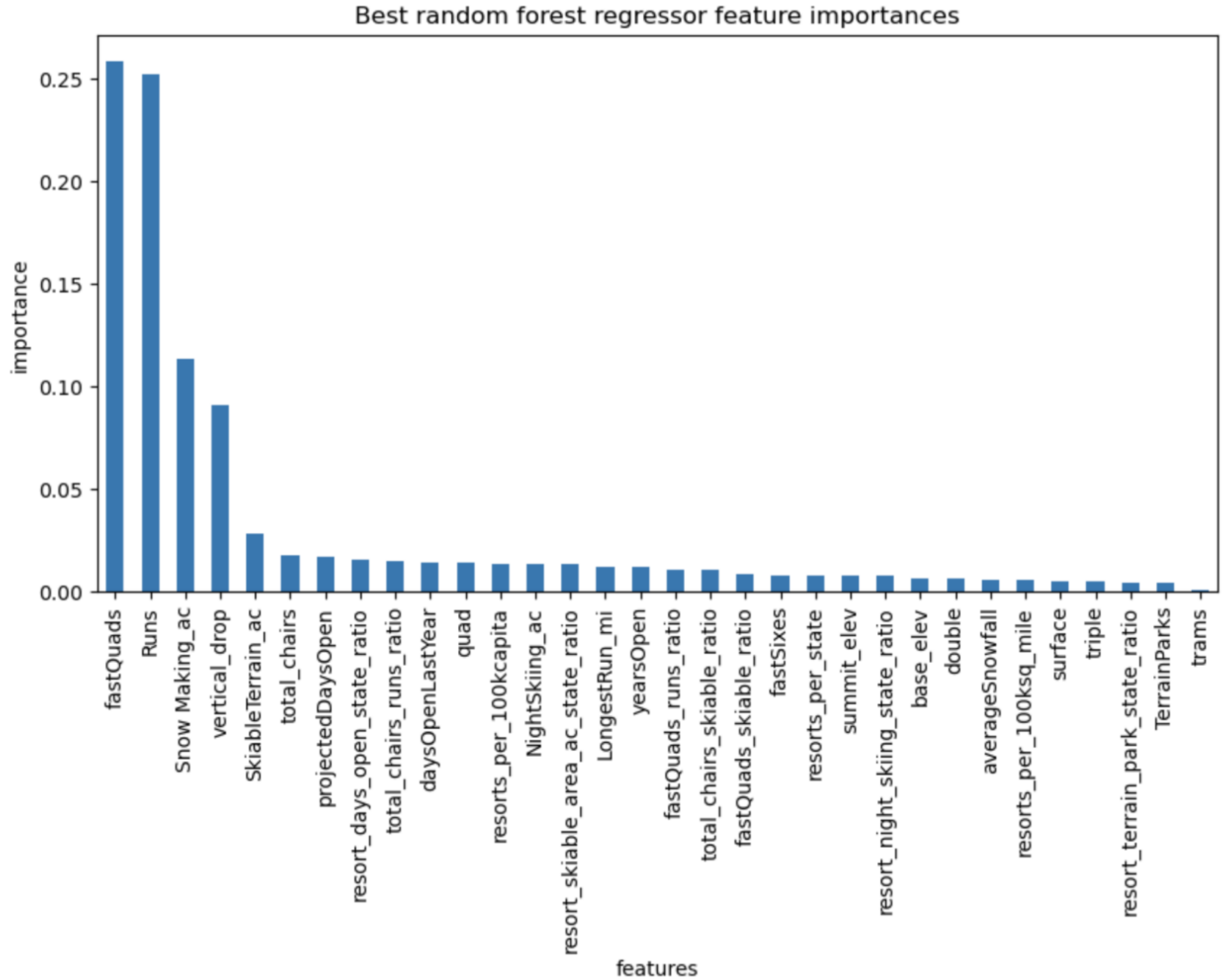

Figure 6: Big Mountain ticket price relative to ticket price distributions of all resorts

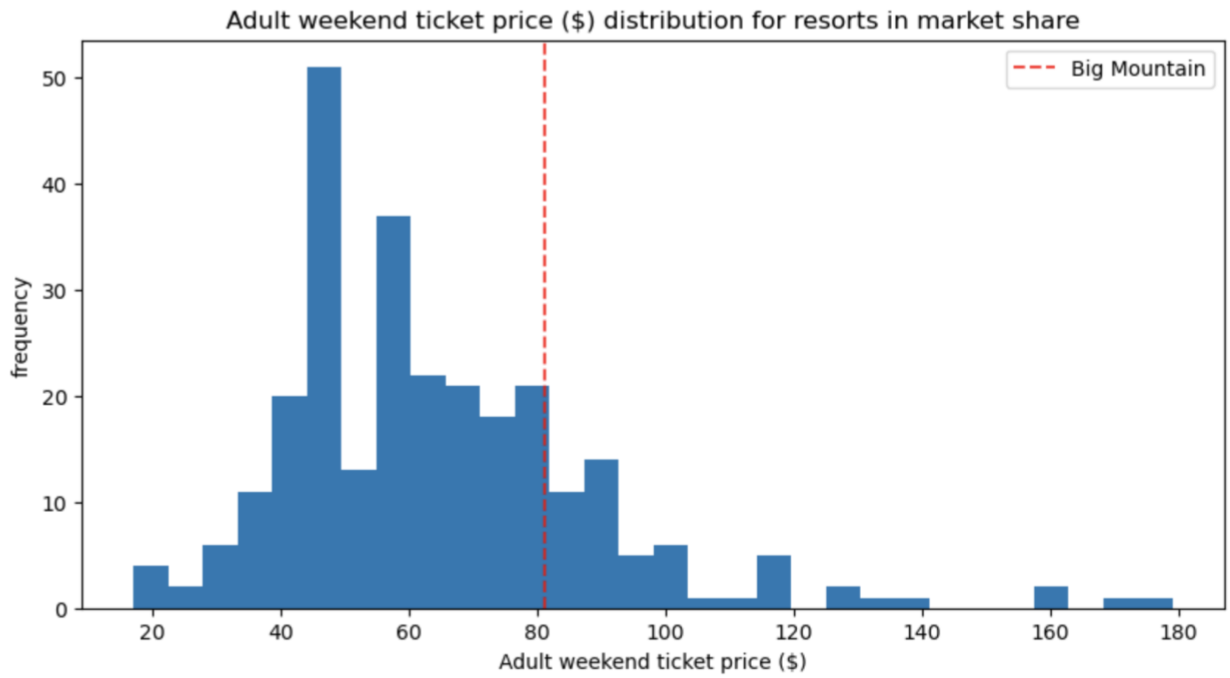

Figure 7: Big mountain vertical drop relative to vertical drop distributions of all resorts

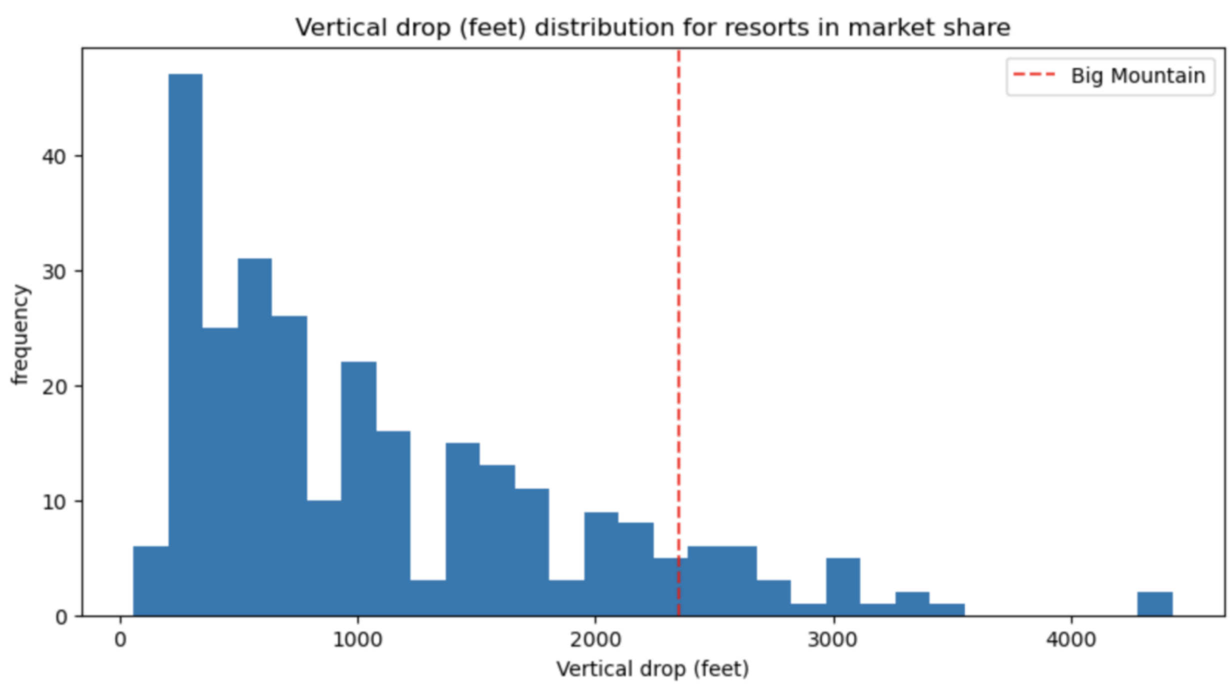

Figure 8: Changes in ticket price and revenue if runs are closed