The data for resorts across the country has been examined and modeled. The two main methodologies being used to model the relations between ticket price and the different attributes of the resort are linear regression and random forest. The random forest model was selected as the best model for the dataset and using that model it was found that if the ticket price for Big Mountain would be priced in a similar manner to the ticket price of other resorts based on all the used attributes we would expect a much higher price for the ticket, nearly $20 higher. Based on this finding it is quiet likely that increasing the ticket price could lead to an increase in revenue as the ticket currently seems significantly underpriced compared to what we would expect for a resort with feature values like that of Big Mountain. Every case is different so this is purely based on statistical averages, so it would make sense to start raising the price gradually and seeing how that would affect the revenue.

We looked at the distributions of multiple features which we would normally expect to correlate positively and price and the values of Big Mountain’s features seemed well on the right side of almost all of the distributions we looked at, so just based on that we would expect the ticket price for Big Mountain to be bigger than the average for all the resorts in the dataset. The features included total number of runs, number of fast quads, area covered by snow makers and vertical drop. When looking at the average price of 70% of randomly selected resorts from the dataset we can see that Big Mountain’s price is more than $15 greater but it is almost two mean absolute errors lower than the price we would expect for Big Mountain based on the data for all the other resorts.

The point of the project was to determine which price the resort should charge in order to maximize profit. To achieve this goal a dataset with hundreds of rows was read in, where each row gave data for one of the resorts. The dataset was examined for data types, sorted tables, distributions and correlations. Also, a good deal of rows were discarded as they were missing the data required for the analysis. After ensuring the dataset is in proper form we proceeded to perform the analysis.

We wanted to model the price for Big Mountain based on what we would expect from examining the relations between feature values and prices for other resorts, so for that reason we removed the row for Big Mountain so it wouldn’t bias the data. We later split the data into a training set and a testing set with a 70% to 30% split. To determine how good our model was we employed three metrics which were R-squared, mean absolute error and mean squared error. The modeling methodology employed was training mean, linear regression and random forest. As would be expected training mean was the worst of the three with random forest slightly beating out linear regression. After employing grid search on both linear regression and random forest models it turned out the best model was random forest with 33 estimators and using median and not the mean as a means to impute missing values. Not scaling the data was better than scaling. When examining feature importance the four features that were by a long margin more important than the rest were fastQuads, Runs, Snow Making\_ac and vertical\_drop.

After determining the best model we finally predicted the price for Big Mountain based on the values of its independent variables. It’s important to keep in mind that there are a lot of other modeling methodologies which we haven’t tested out here which might give more accurate results. Although random forest is a reliable and well-established algorithm for future scope of work it might be helpful to test out other algorithms and see if they do better. Although it probably is not going to make a big difference perhaps the staff might also be interested to perform a similar analysis on weekday prices, since this analysis is only concerning weekend prices. Overall, although there is definitely more to explore with this dataset the fine-tuned random forest model provides a solid framework for predicting weekend ticket prices and seeing how changes in a feature’s value are predicted to affect that price.

Graph of importance of features:

A graph with blue and white text

Description automatically generated

Graph of the distribution of total number of runs per resort with dashed line showing the value for Big Mountain:

A graph with numbers and a line

Description automatically generated with medium confidence