Pricing Model for Big Mountain Resort's Ticket Price

- **Problem Statement**

The problem we are trying to solve is to find the best value for Big Mountain Resort's ticket price to increase their revenue by approximately $1.54M in less than one year to compensate for the additional operational costs of their newly installed chair lift. For this purpose, we have worked through each step of the data science method as described below to come up with the best pricing model that Big Mountain Resort can utilize.

- **Data Wrangling**

We have done the initial data wrangling to clean and organize data. After exploring the dataset we have found that it contains two types of ticket price: an adult weekday price and an adult weekend price. We checked the data for missing values and we found out that 82% of resorts had no missing price information, 3% were missing just one value and 14% of resorts were missing both price values. We dropped those 14% of records which had no information on pricing data.

We have also cleaned the rest of the data by doing some research and correcting some suspicious values for some of the features (e.g., SkiableTerrain_ac) with other values that we were able to verify. Another example of cleaning data was related to a resort that stated to have been open for 2019 years, and we were not sure about the correct value for it therefore, the whole row was dropped. The histogram of these features are shown in Figure 1, before and after cleaning the data to show how data cleaning can help to get better distributions of the features.

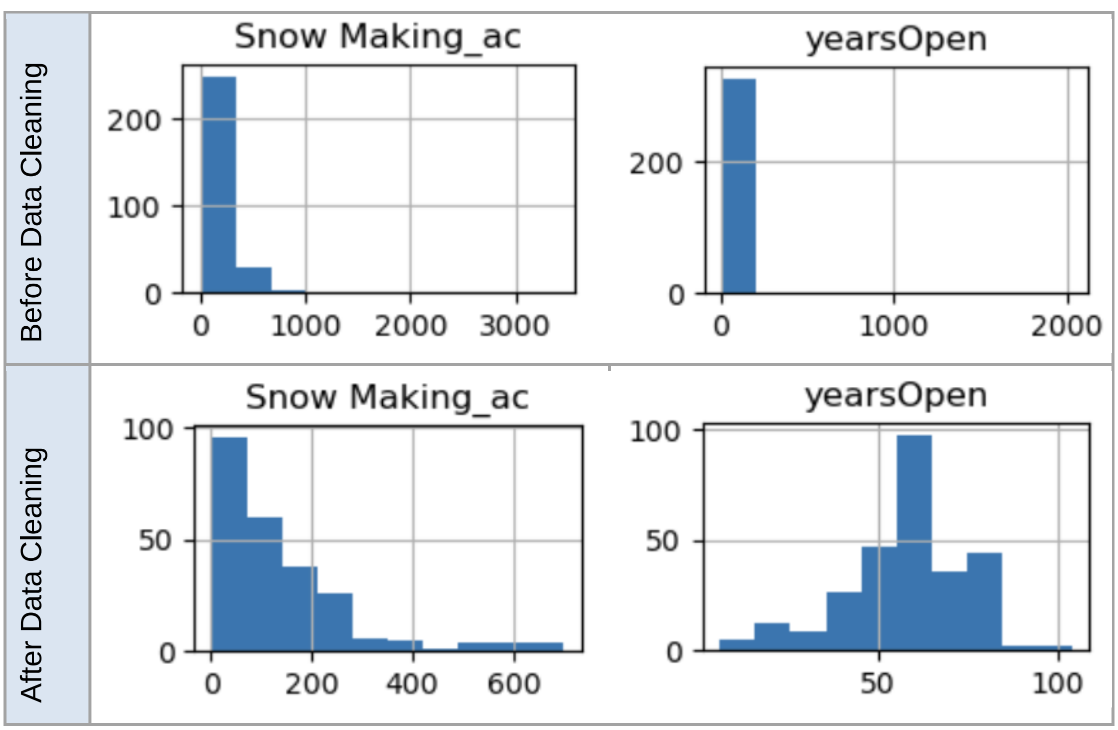

Figure 1- Histogram of features before and after data cleaning

Some of the key takeaways from data wrangling were as below:

- As shown in Figure 2, the overall data analysis showed that once the weekday price approached $100 from below, weekend and weekday prices were typically the same. Perhaps once a resort starts charging that price, it's regarded as a premium resort and their pricing models don't distinguish between weekend and weekday.

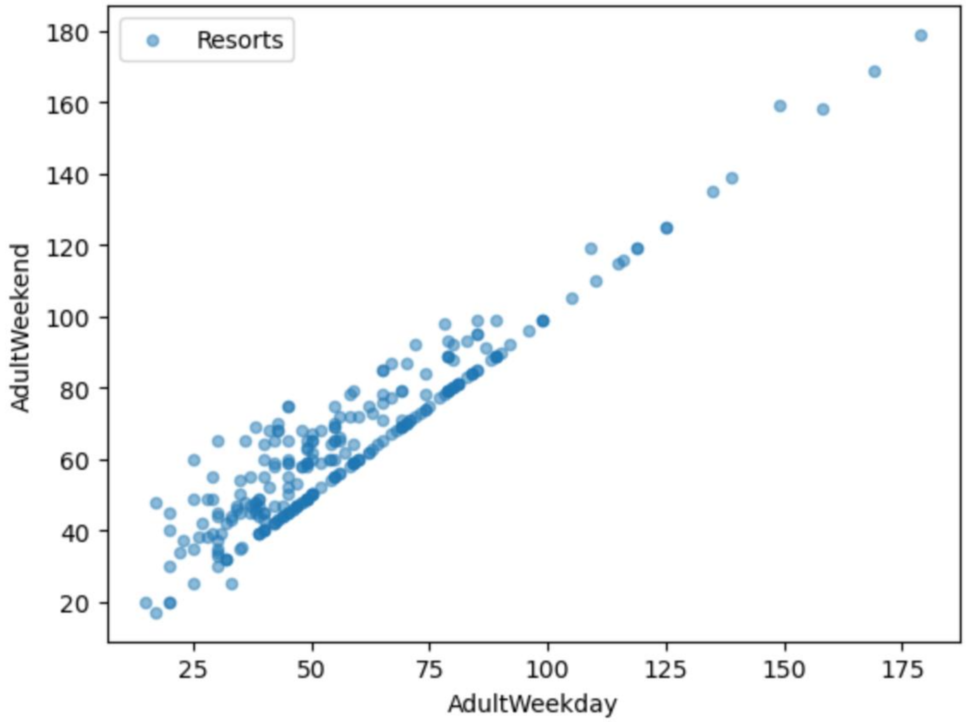

Figure 2- Relationship between weekend and weekday ticket prices

- The current weekday and weekend prices for Big Mountain resort were seen to be equal, at $81. This equality between weekend and weekday prices was seen for all resorts in Montana; whether this is because of some state policy or just what the market in Montana dictates.
- Since we had a few more weekend prices than weekday prices in the dataset, we decided to drop the weekday prices, and choose weekend prices as our target ticket price.

After cleaning and organizing the ski data to the best of our knowledge, the data was augmented with some state-wide information such as state population, size, and number of resorts in the market segment in each state from a third party which may be useful for including state-wide competition amongst resorts in this market segment.

- **Exploratory Data Analysis**

After careful analysis of the data based on their associated state using Principal Component Analysis (PCA) method, it was concluded that there has been no clear pattern suggestive of a relationship between state and ticket price. Due to this, state labels were

treated equally towards building a pricing model that considers all states together, without treating any one particularly specially.
The key takeaways after working with PCA is that we are able to identify which states appear to be outliers in the dataset. In this instance those were Vermont, New Hampshire, New York, and California. All had either a huge number of resorts per capita or per square mile relative to the remaining states being examined. However those outliers don't appear to weight the ticket price at all, so this isn't useful for the purposes of modeling ticket price in later steps. Real insights were gained after calling a seaborn correlation heatmap (Figure 3) on the original dataset (rather than PCA scaled dataset) that identified which features were more positively and negatively associated with ticket price.

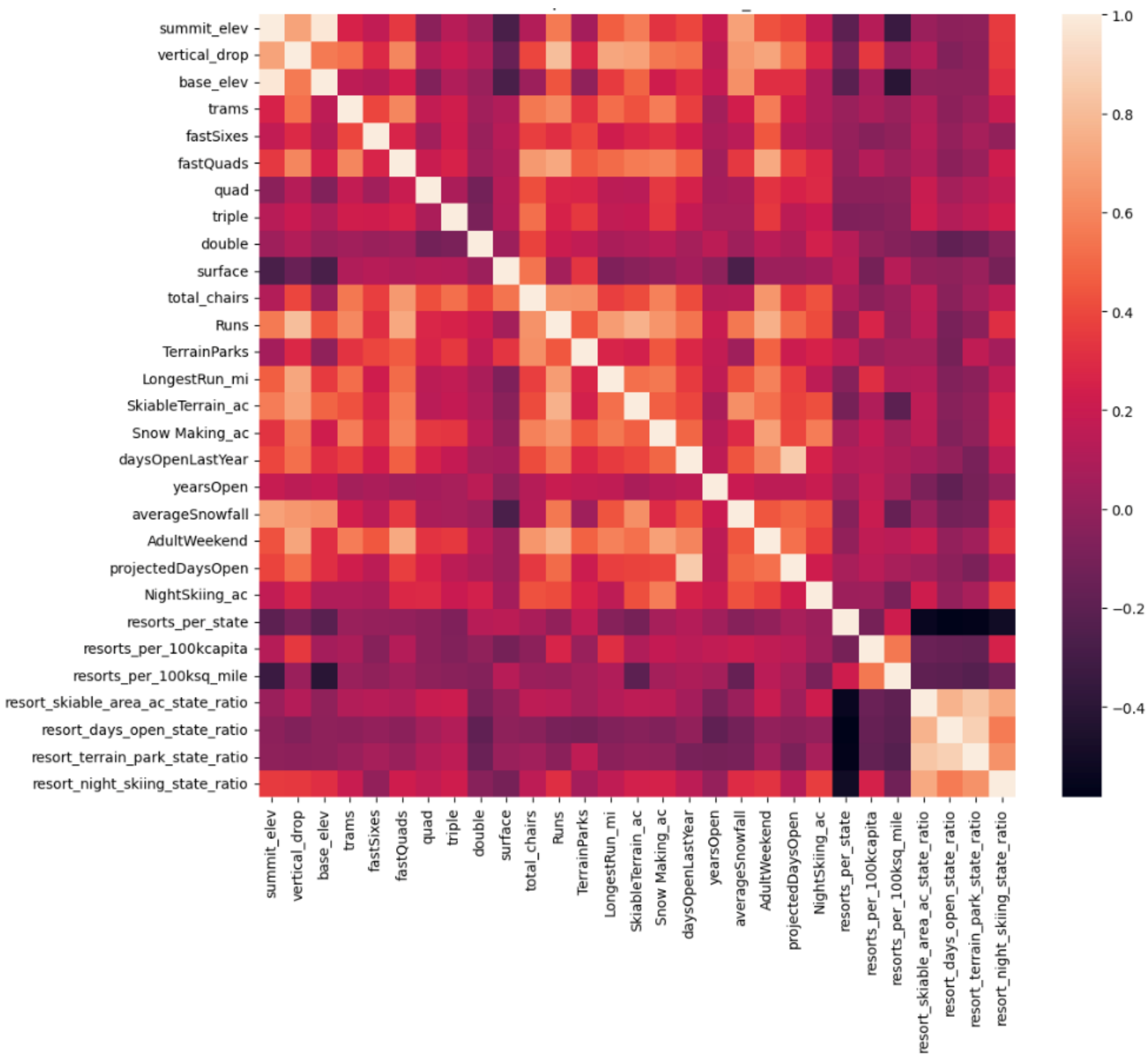

Figure 3- Heatmap of Correlations in ski_data

From here it is immediately clear that ticket price is heavily impacted by these primary features: FastQuads, Runs, Vertical Drop, Total Chairs; and then moderately impacted by the secondary features: Longest Run, Skiable Acres, Snowmaking Acres, and Night Skiing.

It seems that the four primary features are in fact best for modeling ticket price. Runs and skiable acres are similar and positively correlated. Vertical Drop would also seem to be inclusive of the longest run. Without a high vertical drop there isn't space for an exceptionally long run. Similarly FastQuads captures both snowmaking acres and night skiing. Scatterplots below (Figure 4) are showing in more detail how ticket price varies with different values of the important features.

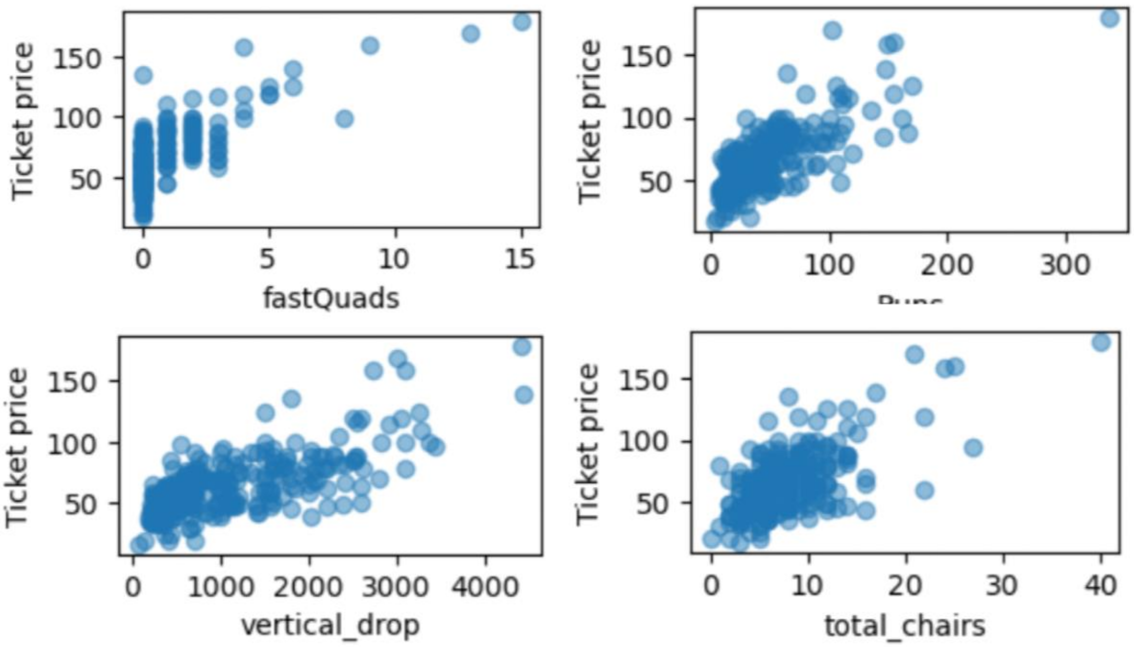

Figure 4- Scatterplots of important features against ticket price

- **Model Preprocessing and Training Data**

In the next step, we started building machine learning models. The Big Mountain Data was separated into a different variable from the rest of the data, and the rest of the data was split into a training set and a testing set using the train_test_split function.
We have used pipeline to build two different models, a linear model and a random forest model. The linear model was refined using the SelectKBest function which selected the 15 'best' features to predict the ticket price. According to R-squared and mean absolute error, the linear model using 15 'best' features seemed to perform slightly worse on the testing set than the training set, but not very drastically worse.
In terms of preprocessing steps, imputing the missing values in the data set using either the mean or the median did not seem to matter or significantly change the results.
The results of the linear model suggested that vertical_drop is our biggest positive feature which is consistent with what we saw during the EDA work. Also, the next strong positive feature is Snow Making_ac which shows that people like guaranteed skiing! The skiable terrain area is negatively associated with ticket price. It could be because larger resorts can serve more visitors at any one time and therefore charge less per ticket. However, the information about visitor numbers are missing.

Next, a random forest model was built. As shown in Figure 5, the results showed that the dominant top four features are in common with our linear model which are fastQuads, Runs, Snow Making_ac, vertical_drop. Comparing the model performance for the two models, showed that the random forest model had a mean absolute error that was lower by almost $1 versus the linear regression model. The random forest regressor had a lower mean absolute error across cross validation folds than the linear regression model, and the random forest regressor mean absolute errors had a lower standard deviation. Performance of the random forest regressor on the test set is consistent (as measured by mean absolute error) to performance on the cross validation sets. The mean absolute error of the random forest regressor on the test set was within less than one standard deviation from the mean of the cross validation MAEs (~9.53 for test vs ~9.64 for CV, std of CV MAEs was ~1.35). Based on this you have decided to use the random forest regressor model going forward because it seems to consistently produce a noticeably lower mean absolute error than the linear regression model.

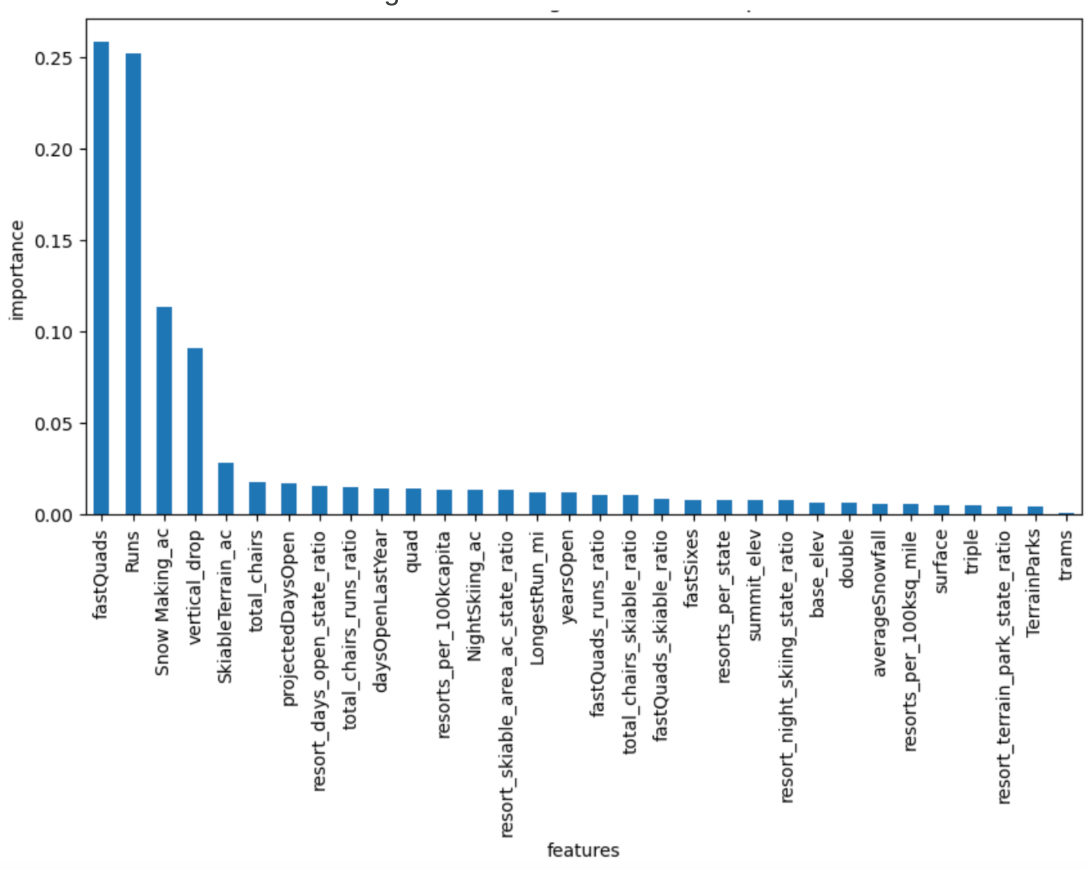

Figure 5- Best Random Forest Regressor Feature Importances

- **Winning Model and Scenario Modeling**

The random forest model was then used to gain some insights into what price Big Mountain's facilities might actually support as well as explore the sensitivity of changes to various resort parameters. For this purpose, we refitted the model on all data excluding the

Big Mountain data. The results showed that, Big Mountain Resort modeled price is $95.87 while the actual price is $81.00. Even with the expected mean absolute error of $10.39, this suggests there is still room for an increase. However, this result is based on some initial assumptions that were made. For example, we assumed all other resorts set their prices accurately according to the market. However, since the Big Mountain might be mispricing their tickets, it's reasonable to expect that some resorts will be "overpriced" and some "underpriced." To better visualize the important features, we created a function to see where Big Mountain sits in terms of price among all other resorts (Figure 6), and also among other resorts in Montana (Figure 7). The results showed that overall Big Mountain is charging more than many of the other resorts but there are quite a few resorts with higher ticket prices. Among the other resorts in Montana, Big Mountain is charging at the upper end of the ticket prices with $81 per ticket.

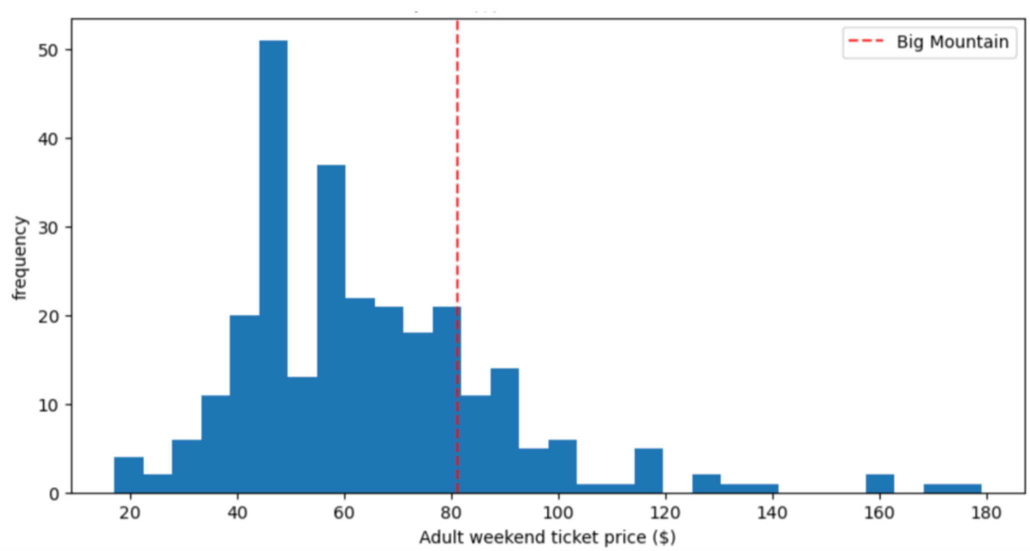

Figure 6- Adult weekend ticket price ($) distribution for resorts in market share

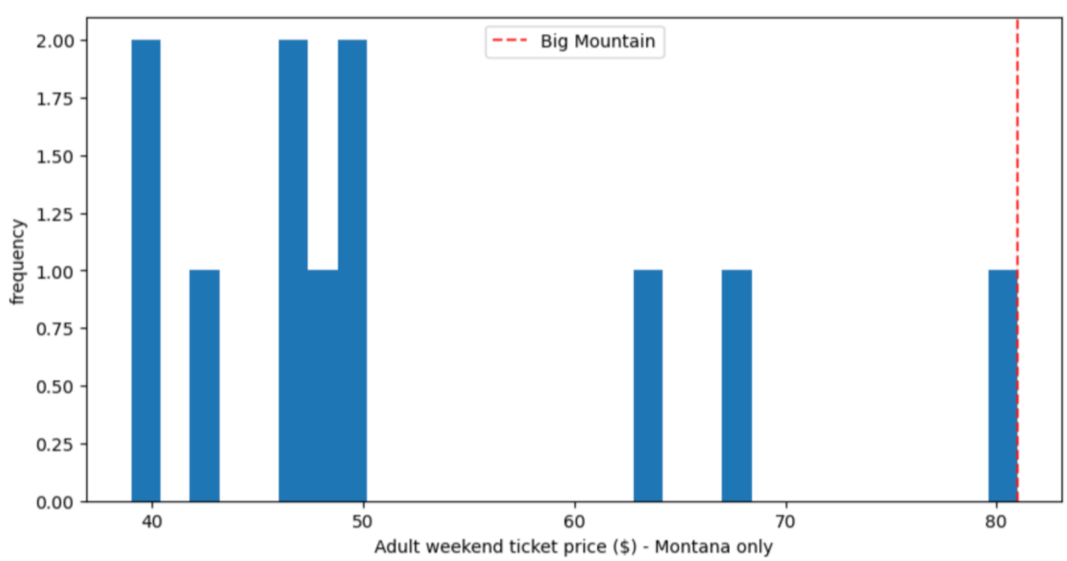

Figure 7- Adult weekend ticket price ($) - Montana only distribution for resorts in market share

The other important features that were plotted included: Vertical drop, Snow making area, Total number of chairs, Fast quads, Runs, Longest run, Trams, and Skiable terrain area.

Different modeling scenarios were considered assuming the expected number of visitors in Big Mountain resort over the season is 350,000 and on average, visitors ski for five days. We also assumed the provided data includes the additional lift that Big Mountain recently installed. To run different scenarios, a new function was first created to introduce a delta (changes from the original value of each feature) and calculate the difference between the scenario's prediction and the current prediction. The scenarios were as below:

1. We explored the option to permanently close up to 10 of the least used runs. Price deltas were predicted based on closing 1, 2, 3, … up to 10 least used runs. Then we made two plots. One for the predicted ticket price change for each number of closed runs, and one for the associated predicted revenue change based on the assumption that each of the visitors buys 5 tickets. The results of the plots showed that closing one run makes no difference. Closing 2 and 3 runs reduces support for ticket price and revenue. However, from closing 3 runs to 5 runs there is no further loss in ticket price or revenue. Closing more than 5 runs will drastically reduce the ticket price and revenue.
2. In this scenario we explored adding a run, increasing the vertical drop by 150 ft, and installing an additional chair lift. The results showed that this scenario increases support for ticket price by $1.99. Assuming 350,000 visitors during the season, and 5 days per visitor, this could add up to approximately $3.5 M.
3. This scenario is similar to scenario 2, but also adding 2 acres of snow making. The results showed that adding a small increase in the snow making area does not make any difference and the support for ticket price still increases by $1.99.
4. The scenario increases the longest run by 0.2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capabilities. The results showed that this scenario does not increase the support for ticket price as all. This could be because the longest run feature is down the importance list in the random forest model.

- **Pricing Recommendation and Conclusion**

We may suggest the leadership take a closer look at scenario 2 which can potentially result in additional of about $3.5M in revenue, however, we don’t have enough information on operating costs. Based on what they estimated, if the additional chairlift installed will cause about $1.5M in operating cost, then adding another chairlift will double this operating cost to about $3M and still there will be about $0.5M increase in revenue.

- **Future Scope of Work**

More information on operating costs of different facilities could definitely be helpful to have a better estimation of the total costs and revenues at the end. Also, more information on weekday ticket prices, seasonal pass, ticket price for kids, etc. could help us better understand the market price and how to support higher ticket prices. It was noted that Big Mountain offers magic carpet, however, there was no information on this facility for the other resorts and whether they offer it or not. It is possible that beginner skiers will tend to pay more for such facilities and therefore this could have a significant impact on the ticket price. Because this information was missing for other resorts we were not able to evaluate the impact of this feature on ticket price. Collecting more relevant data can help us to include more appropriate features in our modeling and improve the accuracy of our results.