Report for Big Mountain Resort

Contents

1. Problem Statement

Big Mountain Resort recently installed an additional chair lift which increases the operating cost by $1,540,000. How can Big Mountain offset that increase of operating cost and capitalize better on its facilities? It is assumed that the current ticket prices are not capitalizing as much as the market would allow.

2. Data Wrangling

The Ski Resort data consists of 330 rows of unique ski resorts and 27 columns of potential features. Our goal is to create a model that can predict the better ticket price.

During the data wrangling step, we first checked the missing values in the data. We found that there were 633 missing or null values. Big Mountain Resort had no missing values. Then, we went ahead and checked all the features. We found that in **SkiableTerrain_ac** column, one resort had wrong entry about its skiable area which we fixed comparing it to their website. Furthermore, going through **SnowMaking_ac** column, another resort seemed to have a wrong entry and was missing ticket pricing, so we dropped that entry. Another column **fastEight**, seemed to have half of the values missing and the remaining half had 0, so we ended up dropping the column entirely. We also found that one resort seemed to have been open for 2019 years according to **YearsOpen** column which clearly seemed to be wrong, so we dropped this entry as well.

As our target variable of ticket price had two potential columns that could be used, **AdultWeekday** and **AdultWeekend**, we went ahead and focused on those columns for now. We found that 14% of the rows had no pricing data so, we decided to drop them. Then, we went ahead and checked the relationship between the two pricing features and found that even though some states showed variability between weekend and weekday prices, some states like Montana and South Dakota showed very small variability in the prices.

	AdultWeekend	AdultWeekday
141	42.0	42.0
142	63.0	63.0
143	49.0	49.0
144	48.0	48.0
145	46.0	46.0
146	39.0	39.0
147	50.0	50.0
148	67.0	67.0
149	47.0	47.0
150	39.0	39.0
151	81.0	81.0

Figure 1: AdultWeekend and AdultWeekday prices in resorts from Montana

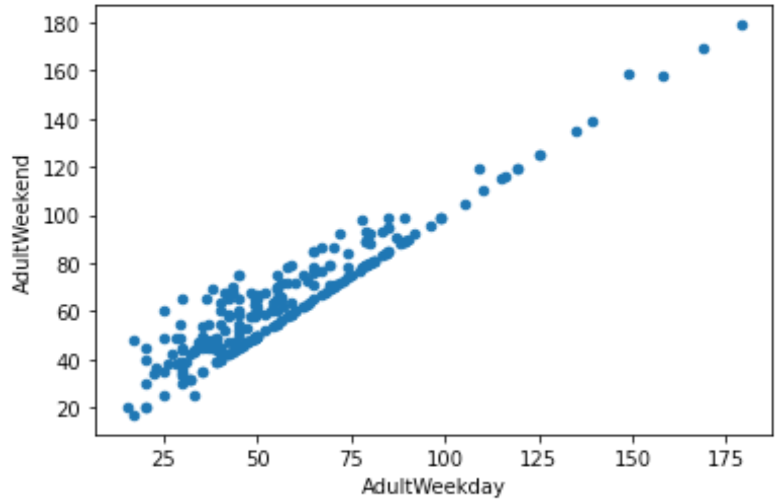

Figure 2: AdultWeekend vs AdultWeekday prices

We checked the relationship between the two potential columns and found that there is a collinear relationship between both ticket variables as seen in Figure 2, so we decided to use only one. As, the **AdultWeekday** column seemed to have more missing values than the **AdultWeekend** column, we decided to drop **AdultWeekend** column.

Finally, we stored this data frame in a CSV file with 277 rows and 25 columns/features. The categorical features in the data include **Name, Region, and state**. The numerical features in the data include **summit_elev, vertical_drop, base_elev, trams, fastSixes, fastQuads, quad, triple, double, surface, total_chairs, Runs, TerrainParks, LongestRun_mi, SkiableTerrain_ac, Snow Making_ac, daysOpenLastYear, yearsOpen, averageSnowfall, AdultWeekend, projectedDaysOpen, NightSkiing_ac**. We also created and saved a data frame with summary statistics of resorts. The numerical features of this data include **resorts_per_state, state_total_skiable_area_ac, state_total_days_open, state_total_terrain_parks, state_total_nightskiing_ac, state_population, state_area_sq_miles**. The categorical feature of this data is **state** which contains the names of all the states of the USA.

3. EDA

During EDA phase, we explored the data to check the statewide picture for our market and found that Montana comes in at third largest according to total state area covered by resort in sq miles. Montana seems to have fewer but larger resorts.

We then found that there was no clear pattern between state and the ticket price through PCA (Principal Component Analysis).

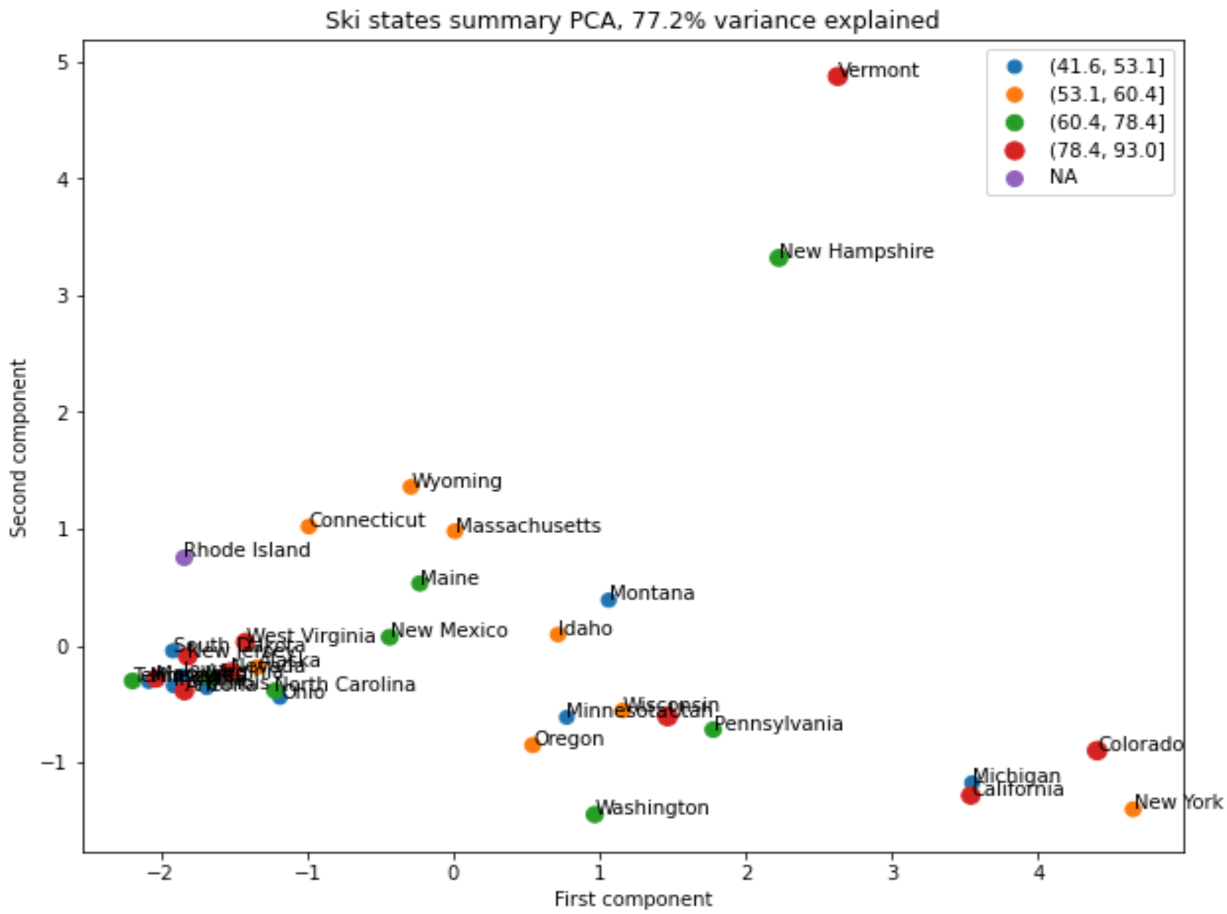

Figure 3: PCA of summary of Ski Resorts in relation to ticket pricing

We found that there isn't an obvious pattern. The colored points representing different quartiles of price could be seen scattered randomly.

Next, we investigated the feature correlation heatmap to check the relationship of the features.

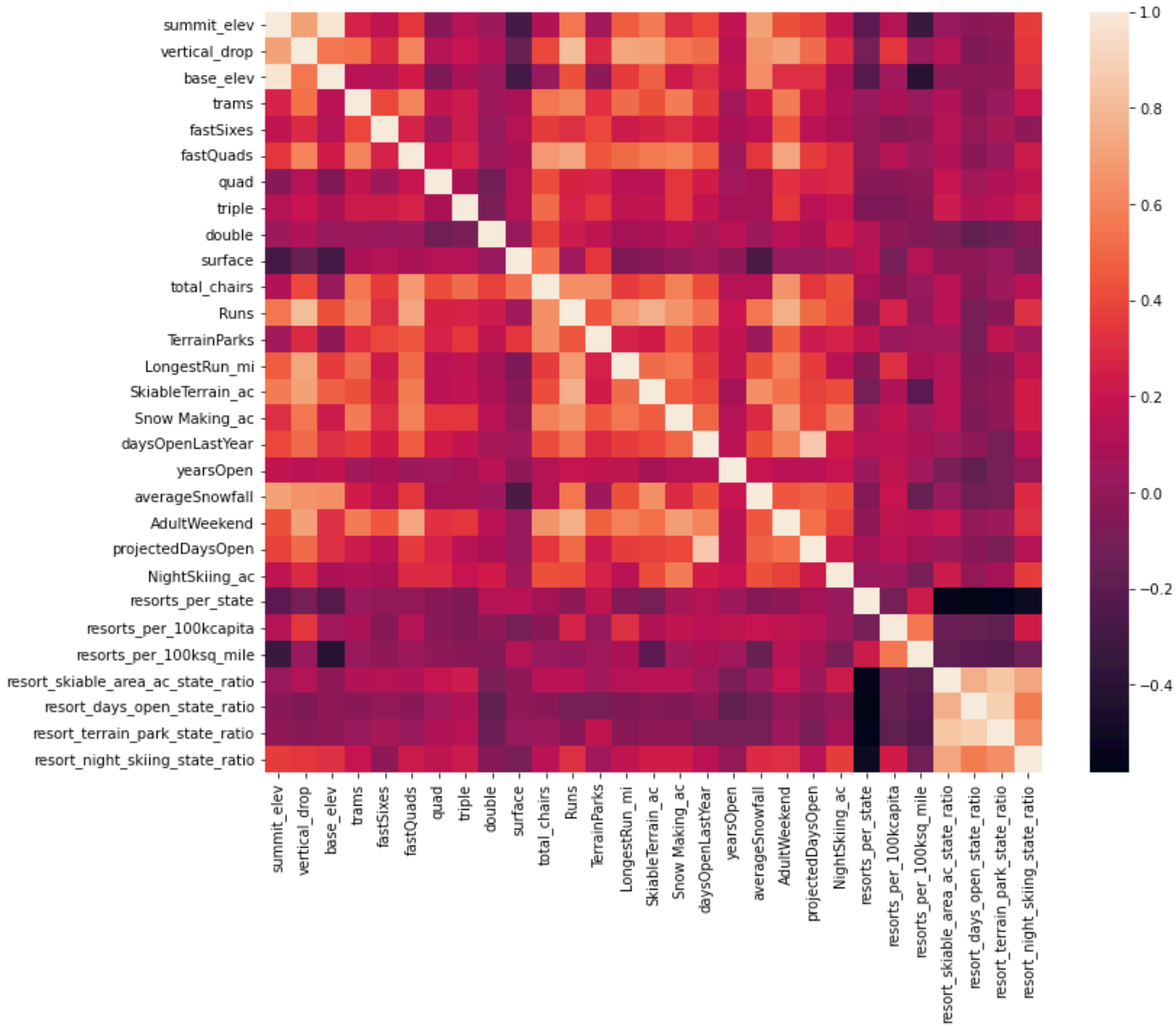

Figure 4: Correlation Heatmap

We found that there was some multicollinearity with the state ratio features that were created. We also found some obvious correlations for example **summit_elev** and **base_elev**. When we turned our attention at the target feature, **AdultWeekend**, we could see that **fastQuads**, **Runs**, **Snow Making_ac**, **vertical_drop**, and **total_chairs** stand out as features that could affect the ticket price.

We created a series of scatterplots to dive into how ticket price varied with other numeric features. We found a strong positive correlation with **vertical_drop**. Some features like **fastQuads, Runs**, and **total_chairs** seemed very useful.

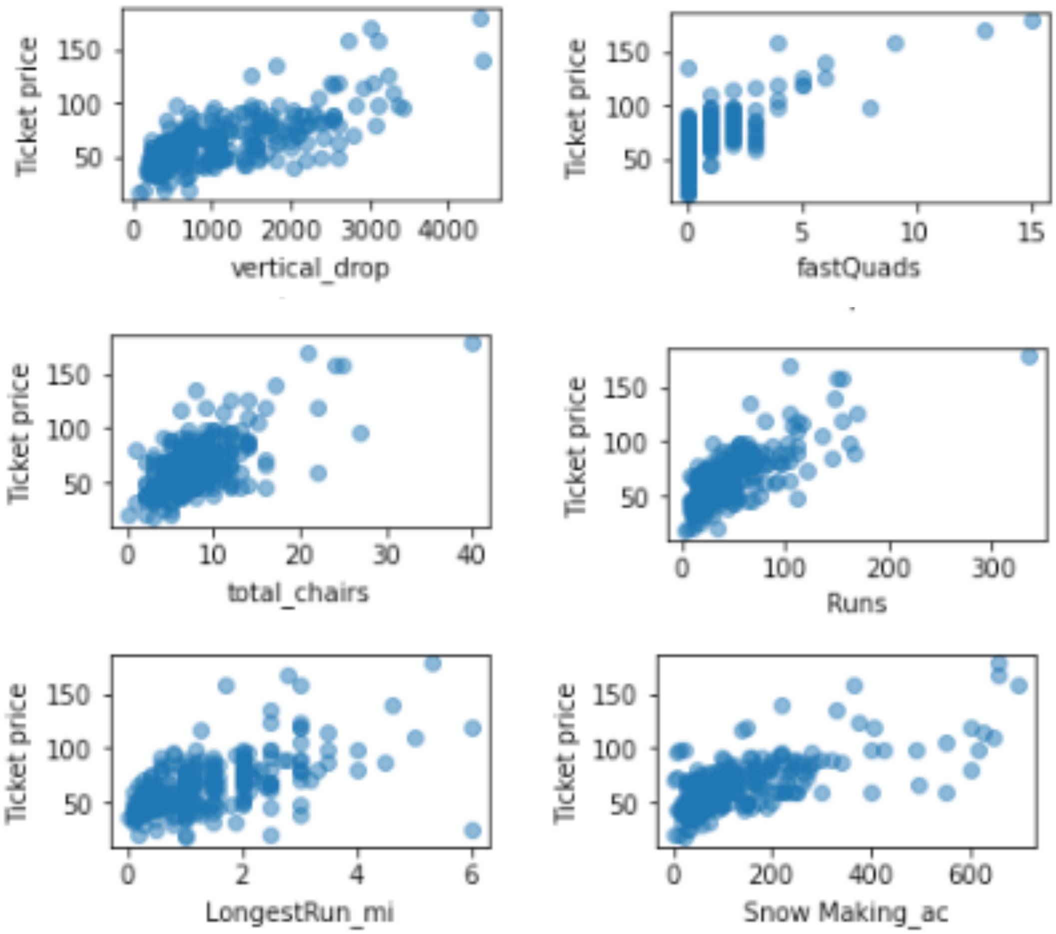

Figure 5: Scatterplots of some features that affect ticket price.

4. Preprocessing and Training

We then went ahead with building machine learning models. After separating the data into training and testing splits (70/30 train/test split), we had 193 rows for training and 83 for testing. First, for a baseline model, we used the mean of the ticket price (which was $63.81) as a prediction for our target. As expected, the model performed very poorly and was off by $19. Second, we tried the linear model and scaled the data imputing the missing data with the median. We used '**SelectKBest**' in a grid search and found that 8 features were optimal. The features are **vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi,**

trams, and SkiableTerrain_ac. This linear model was better than the baseline but there was inconsistency between the performance from cross validation. We also could improve the performance of the model in the test set and hence decided to try another model.

Finally, we tried the Random Forest model and found that using the median for imputing the data was better than using the mean and scaling of data was not helpful. We found that this model performed he best with a MSE oof 9.64 from the cross validation and 9.54 on the test set.

	Models	Cross-Validation MAE	Test MAE
1	Dummy Regressor (mean)	17.92	19.14
2	Linear Regression	10.5	11.79
3	Random Forest	9.64	9.54

Table 1: Accuracy of model based on Mean Absolute Error (MAE)

According to Random Forest, the top 4 features are **fastQuads, Runs, Snow Making_ac, and vertical_drop**.

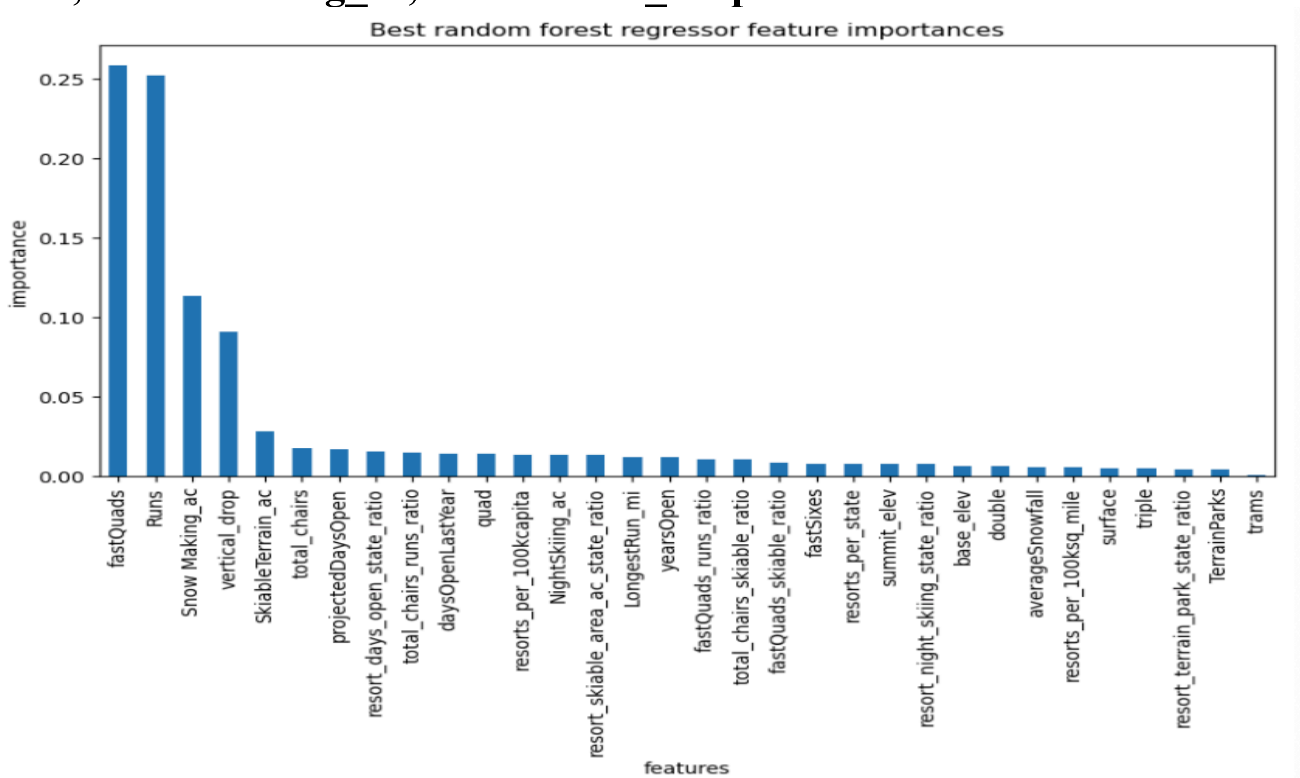

Figure 6: Best features using Random regressor.

5. Modeling

The current adult weekend ticket price for the Big Mountain Resort is $81.00. According to the model, they could charge up to $95.87 per ticket. We could calculate this price through our model using the ticket price of other resorts.

We also reviewed 4 other scenarios that was shortlisted by the Big Mountain for either cutting costs or increasing revenue from the ticket price. The first option was to close up to 10 of the least used runs as this doesn’t impact any other resort statistics. The model showed that closing 1 run made no difference. It also showed that closing 2 and 3 runs reduces support for ticket price and so revenue. If 3 runs are closed, it seemed to be better to close 4 or 5 as there was no further loss in ticket price. Increasing the closure down to 6 or more led to a large drop on the price.

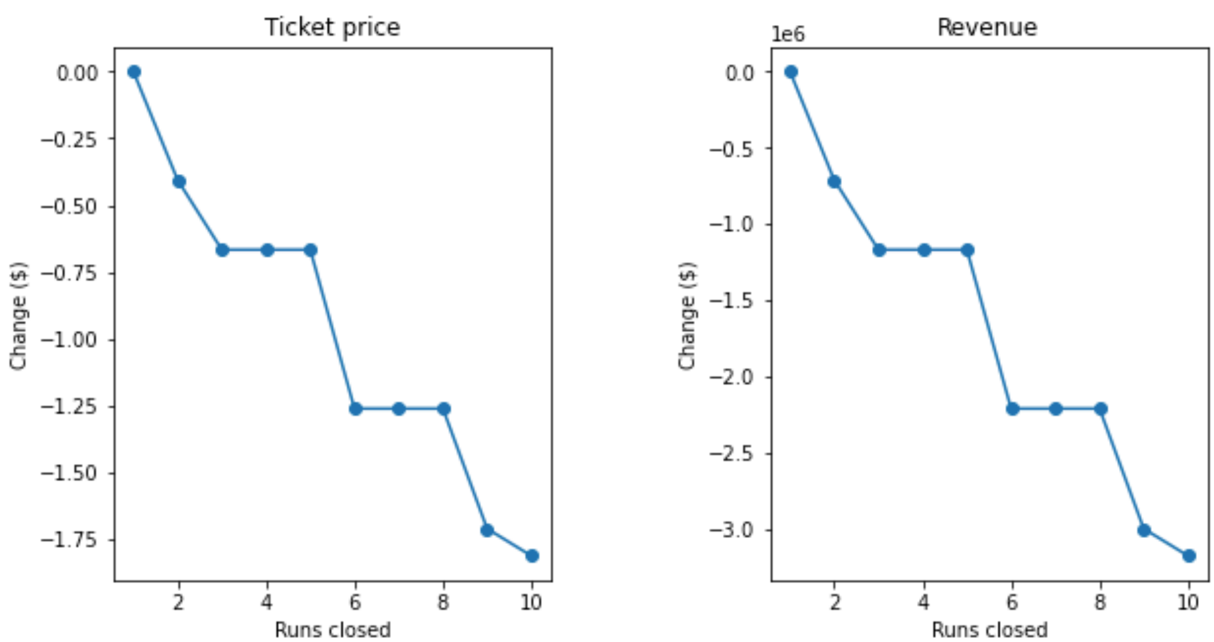

Figure 7: Scenario 1

The second scenario was to add a run, increase the vertical drop by 150 feet, and adding a chair lift. This scenario supported an additional increase of ticket price by $1.99 and is expected to amount to a $3,474,638.

The third scenario was the same as scenario 2 but adding 2 acres of snow making cover. This scenario had no difference compared to scenario 2.

The fourth and final scenario was to increase the longest run by 0.2 miles and increasing the snow making capability by 4 acres, but the model suggested that there is no difference in support for the ticket price.

	Scenarios	Description	Price	Difference with Predicted Price	Revenue ($Mi)	Revenue Diff from Modeled Price ($Mi)
0	Current Price	Current Price	81.00	-14.87	141.750	-26.02
1	Predicted Price	Predicted Price	95.87	0.00	167.772	0.00
2	Scenario 1a	Closing 1 run	95.87	0.00	167.772	0.00
3	Scenario 1b	Closing 2 run	95.46	-0.41	167.055	-0.72
4	Scenario 1c	Closing 3-5 runs	95.20	-0.67	166.600	-1.17
5	Scenario 1d	Closing 6-8 runs	94.61	-1.26	165.568	-2.20
6	Scenario 1e	Closing 10 runs	94.06	-1.81	164.605	-3.17
7	Scenario 2	adding 1 chair and 150ft vertical drop	97.85	1.99	171.238	3.48
8	Scenario 3	adding 2 acres of snow making	97.85	1.99	171.238	3.48
9	Scenario 4	extending longest run (0.2mi) & adding 4 acres...	95.87	0.00	167.772	0.00

Figure 8: Ticket prices and revenue of different scenarios

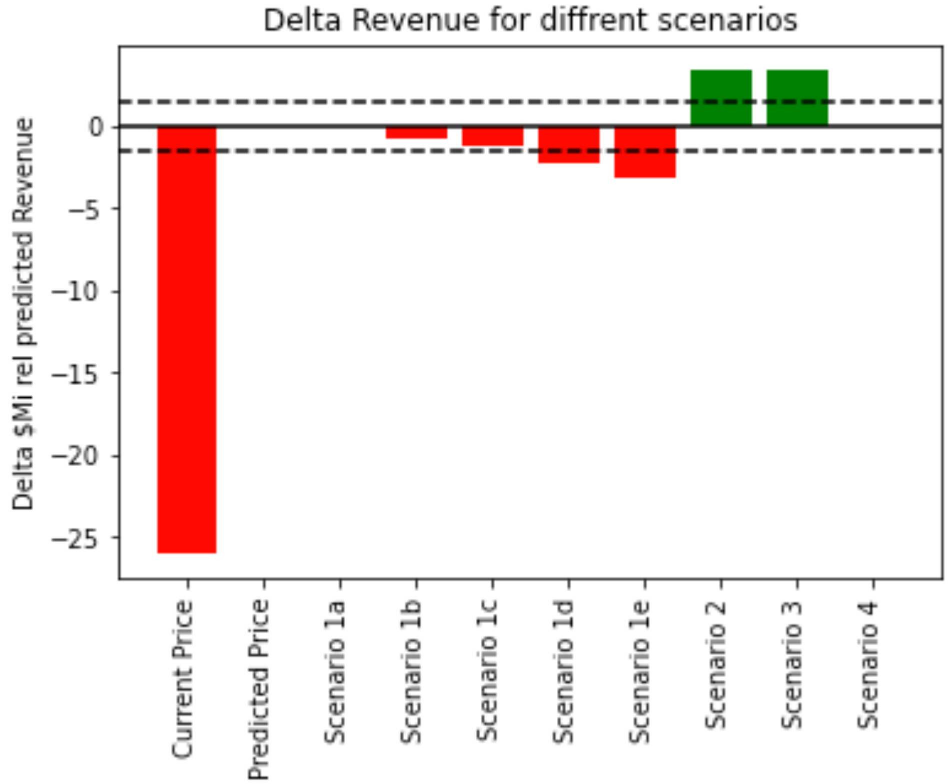

Figure 9: Revenue for the different scenarios given.

6. Recommendation

The current price for the Big Mountain Resort is $81, but the model predicted the price as $95.87. The additional operating cost for adding the new chair lift installed by the resort is $1,540,000. With 350,000 expected visitors and on the basis that each visitor on average buys 5-day tickets, the expected revenue increase from this $14.87 ticket price increase is $26,022,500. This revenue seems to be able to cover that additional cost easily.

For future improvement, I would suggest them to consider Scenario 2 which is to add a run, increase the vertical drop by 150 ft, and adding an additional chair lift. This scenario further supports on additional increase of the ticket price by $1.99, with the revenue increase of $3,474,638.

We could also consider scenario 1 but it is difficult to recommend without knowing the cost of maintaining the runs and what other consequences can happen due to run closure. The model suggests that closing 1 run will not affect the pricing, but we could possibly reduce the expenditure on maintenance of one of the least used run. The model also predicts that closing 5 runs would decrease the ticket price by $0.67 from the predicted price. This would amount to a revenue loss of $1,720,000 compared to predicted ticket price of $95.87. We would need further information about the operating costs to maintain those least used 5 runs, to check if the cost cut could offset that loss.

7. Conclusion

The data only provided information on weekend ticket prices. It would have been more beneficial to have no missing information of both weekend as well as weekday prices. Also, the price for other services provided by the resort like lodging, food, rentals, lockers, lessons etc would have added more information for evaluating the ticket price. Additionally, other operating costs for maintenance of runs, machines and facilities could have helped in more informed decision. Even though Big Mountain's ticket price was previously a premium above the market average, it is still less than the modeled price. Thus, they should consider all the facilities when deciding on the prices.