Big Mountain Report

Contents

1 Problem Statement

How can Big Mountain Resort offset the $1,540,000 increase in operating costs for this season by cutting costs and selecting a better value for their ticket prices?

2 Data Wrangling

The dataset we are analyzing consists of 330 rows each representing a unique ski resort and 27 columns of potential features. Our goal is to create a model that predicts ticket price.

During the data wrangling phase of this project, we found that about 14% of the resorts had no ticket price information, so these rows were dropped. We also found that one resort had an error in the **yearsOpen** column, and so we decided to drop this row as well. Leaving us with 277 rows. For the columns, **fastEight** had mostly missing data, so we dropped it. Our target variable of ticket price had two potential columns that could be used, **AdultWeekend** and **AdultWeekday**. We can see from figure 1 that there is a collinear relationship between both ticket variables, so we should be justified in using either one. Since **AdultWeekend** has fewer missing entries, we chose to go with it as our target. We also created a new data frame consisting of statewide statistics and will be merged with the resort data.

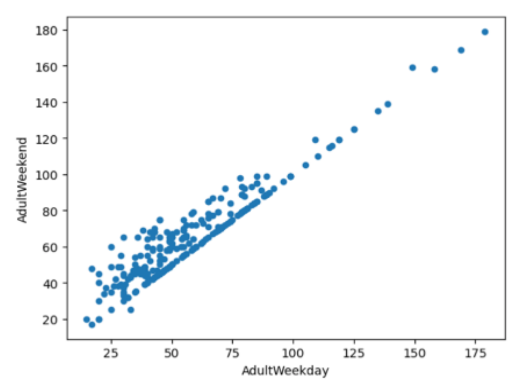

Figure 1: AdultWeekend vs. AdultWeekday

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, projectedDaysOpen, NightSkiing ac, resorts per state, resorts per 100kcapita, resorts per 100ksq mile, resort skiable area ac state ratio, resort days open state ratio, resort terrain park state ratio, resort night skiing state ratio, total chairs runs ratio, total chairs skiable ratio, fastQuads runs ratio, fastQuads skiable ratio.** Categorical features in the data include **Name, Region, and state**.

3 EDA

Exploratory data analysis revealed that there was no clear pattern between **state** and **AdultWeekend**. This was discovered by using a PCA on our statewide statistics data. Because of this, we will not be including **state** (and **region**, since it is essentially the same as **state**) as features in our model. Next, we will take a look at a feature correlation heatmap to get a sense of which features have a relationship. There is a lot going on in this heatmap, but what stands out the most is there is a lot of multicollinearity with the state ratio features that were created, as well as some obvious correlations like **summit elev** and **base elev**. When we look at our target, **AdultWeekend**, we can see that **fastQuads, Runs, Snow Making ac, vertical drop,and total chairs** stand out as features that could effect ticket price.

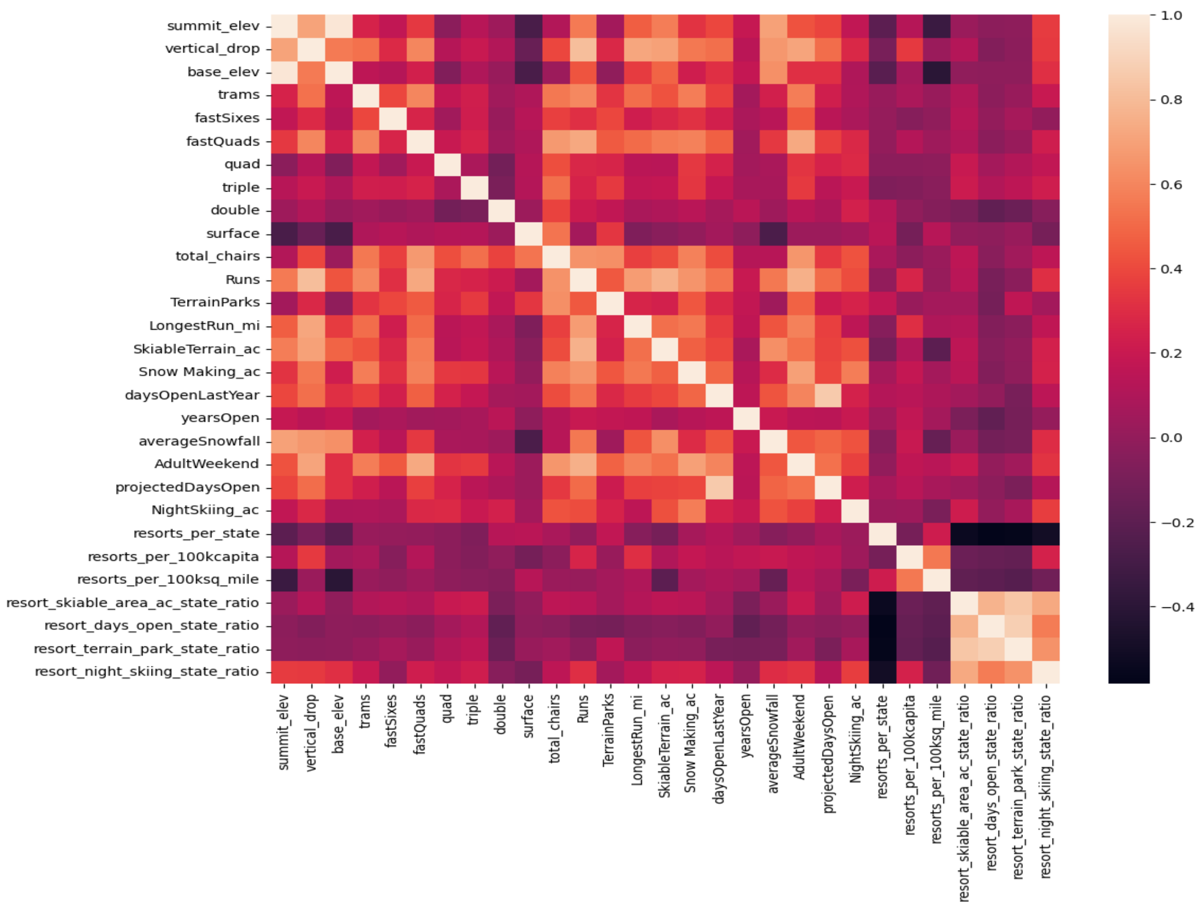

Figure 2: Correlation Heatmap

4 Preprocessing and Training

For a baseline model we used the mean ticket price as a prediction for our target. As expected, this performs poorly and is off by about $19. To improve on this we can try using a linear model. We can get our optimal results by using **SelectKBest** which revealed that the 8 best features are **vertical drop, snow making ac, total chairs, fastQuads, Runs, LongestRun mi, trams, and SkiableTerrain ac**. The linear model improved on this, but the inconsistency between the performance from cross-validation and the performance on the test set should be improved. The final model we tried was a random forest, and this performed the best with a MSE of 9.6 from cross-validation and 9.5 on the test set. The top 4 features according to the random forest are **fastQuads, Runs, Snow Making ac, and vertical drop**. We will be using the random forest model for further business modeling, and we have a good idea for which features impact ticket price the most.

Models		
Model	Cross-Validation MSE	Test MSE
Dummy Regressor (mean)	17.9	19.14
Linear Regression	10.5	11.8
Random Forest	9.6	9.5

Table 1: Accuracy of Models

5 Modeling

Big Mountain resort modeled ticket price is **$95.87** and the actual price is **$81.00**. We have 4 possible scenarios that the business has shortlisted to either cut costs or increase revenue from ticket price. The first scenario is to close down up to 10 of the least used runs. The model suggest that closing one run will not make a difference, while closing 2 and 3 reduces support for ticket price and revenue. Closing 4 and 5 will not reduce it anymore, but closing 6 results in a big drop. The results can be seen in figure 3.

Scenario 2 is to add a run, increase vertical drop by 150 feet, and installing an additional chair lift. This scenario increase support for ticket price by **$8.61** and is expected to amount to a **$15,065,471** increase in revenue. Scenario 3 is the same as scenario 2 but also adding 2 acres of snow making. This increases support for ticket price by **$9.90** and increases revenue by **$17,322,717**. The last scenario calls for increasing the longest run by 0.2 miles and increasing snow making capability by 4 acres, but the model suggests that there is no difference in support for ticket price.

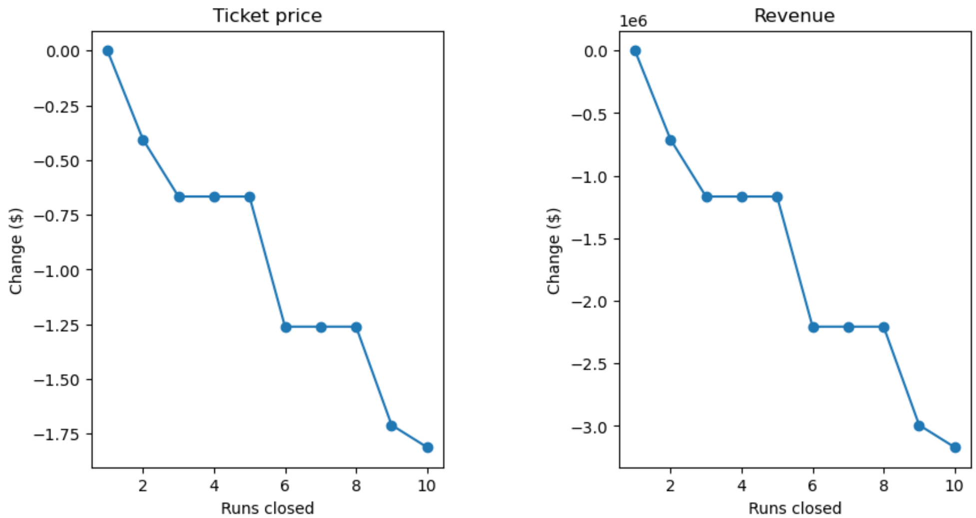

Figure 3: Scenario 1

6 Recommendation and Conclusion

The current ticket price for Big Mountain is $81, and the modeled price is $95.87. With 350,000 expected visitors, and assuming that each visitor purchases 5 day tickets on average, the expected revenue increase from this $14.87 ticket price increase is $26,022,500. This would more than cover the addition operating cost of $1,540,000 for the new chair lift.

For future improvement, I would suggest scenario 2 which is to add a run, increasing the vertical drop by 150 ft, and adding an addition chair lift. This scenario supports a ticket price increase of $8.61, with a revenue increase of $15,065,471.

Scenario 1 is something to consider, but it is difficult to say what costs will be cut without knowing the cost of maintaining the runs. The model suggest that closing runs supports reducing ticket price. For example, the model predicts that closing 5 runs would decrease the ticket price by approximately $0.75. This would amount to a revenue loss of $1,312,500. I would suggest collecting data on the operational costs to maintain the 5 least used runs and see if cutting those costs would offset that loss.

The data only provided information on ticket prices, it would be beneficial to also have information on prices for other goods and services that the resort might provide. These can include data on lodging, ski equipment rentals, lockers, lessons, etc. Likewise we were only provided with the operating cost of 1 additional chairlift, but it would be useful to know about other operational costs, like general maintenance of the resort facilities or the snow making machines. The cost to pay employees could also be useful information to have.

Big mountain's ticket pricing strategy was previously a premium above the market average. This is a naive approach and it is expected to be less than the modeled price since it doesn't take into account the importance of the facilities of the resort. When looking the the distribution of our important features (vertical drop, snow making area, total number of chairs, fast quads, runs, longest run, trams, and skiable terrain), big mountain is above the average for most of them. Thus, they should be taken into account when deciding on ticket prices.