**Guided Capstone Project Report**

Introduction

The Big Mountain Resort is a ski resort located in Montana. It offers many different amenities include 11 chair lifts, 2 T-bars, and a 3.3-mile-length run named Hellfire. Every year approximately 350,000 skiers and snowboarders of all levels visit the resort. For better distribution of guests through the resort, Big Mountain added another chair lift. This lift increases their operating costs by $1,540,000 for this season. Normally, they use the average ticket price of resorts in its market segment and charge a premium over it for their ticket prices. However, they could be capitalizing on its amenities further and charge more for them, especially to offset the price of the new lift. This report will detail how different features could be more important than others and how that could support increasing the ticket prices enough to make a profit for this season.

Data

The dataset provided by the database manager includes information on 330 different ski resorts, including Big Mountain. For each resort, the following features are utilized in the analysis: region, state, summit elevation, vertical drop from the summit to the base, elevation at the base of the resort, numbers of the different trams and chair lifts, number of runs, number of terrain parks, length of the longest run, total skiable area, total snow-making area, number of terrain parks, total days open in the last year, total years open, average snowfall, weekend ticket prices, projected days open for the upcoming season, and total skiable area suitable for night skiing. Population and area data of US states were also used to generate summary statistics for each relevant state.

Methodology

Principal component analysis was used to see how much variance the different features caused, which showed that there is no specific pattern among the states, and therefore will all be treated equally. Using Seaborn’s heatmap, it was possible to see how each feature correlated with each other, specifically considering how each feature correlated with Weekend Price. The features that are positively correlated would seem to be more highly valued by guests of the resorts. Using scatterplots of each feature against the ticket price illustrated these correlations further, by showing the extent of the correlations.

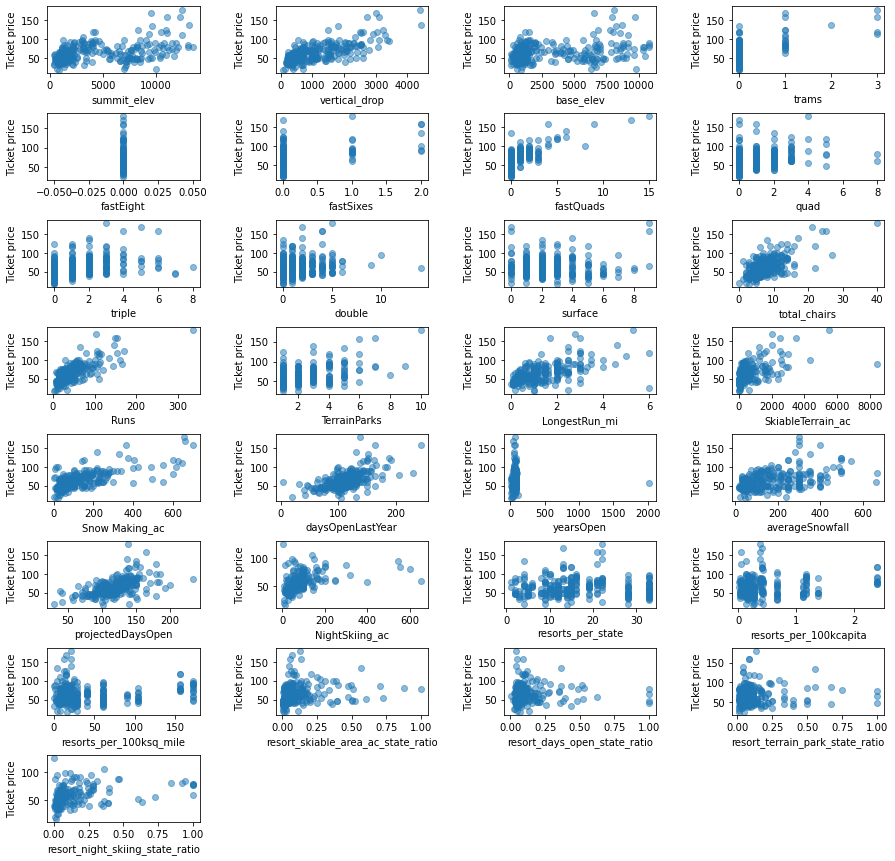


Figure 1

From these scatterplots, the features that have the highest correlation with ticket price are vertical drop, number of fast quad chairs, number of runs, total number of chairs, and longest run. These may be used in the modeling to determine if and how the ticket price should be adjusted.

To generate a model, a 70/30 train/test split of the data was used. There is a possibility that the average ticket price across all resorts could be the best price to charge. A baseline linear regression model was created that used the average ticket price as a prediction. Then, other features of the data were used. However, in the data there are a lot of missing values across the features. To fix this and train a better model, the missing data was imputed. First, the medians of each feature were used for the imputation, then their means were used, and a linear regression model was fit in both scenarios. It was determined that some specific features would be more valuable in choosing ticket price. A pipeline was used to impute data, fit a random forest regression, and choose the better features to influence the regression. Cross-validation was done to assess model performance and use the best one.

Analysis

When the average ticket price was used as a prediction for the baseline model, its coefficient of determination, or R2, on the training set was 0; the testing set had an R2 of -0.00072, which is not a good sign. Further, the mean absolute error, or MAE, is determined to be approximately 18, meaning if the average ticket price is used, on average, we would miss a better price by around $18. This all verified that just using the average price would be ill-advised.

Then the medians and means of each feature were used. A linear regression model was fit on the data in both scenarios. Using the median, the R2 is 0.838 on the training set and 0.692 on the test set; the MAE is 8.062 and 10.448, respectively. Using the mean, the R2 is also 0.838 on the training set and 0.692 on the test set; the MAE is 8.095 and 10.352, respectively. These results are better than just using the average price, but the model can be refined even more.

The cross-validated random forest regression pipeline that was generated shows the features with highest importance, which are vertical drop, number of fast quad chairs, number of runs, and snow-making area, in agreement with the EDA. This model’s MAE is 10.271 and 9.04 on the training set and the test set, respectively. This is the model chosen to predict what Big Mountain’s ticket price should be.

Results

Using the random forest model, Big Mountain the ticket price should be $92.63. It was also used to tell whether raising the price further is justified when different features are changed. Assuming the new chair lift was included in the data, based on the expected number of visitors (350,000) and an average 5-day stay, the different scenarios provided were:

1. Permanently closing up to 10 of the least used runs; this doesn't impact any other resort statistics. The model showed that closing one run makes no difference but closing two or three runs would reduce support for the ticket price. If three runs are closed, then closing four and five would not reduce the ticket price further. Closing 6 or more runs would decrease price, and subsequent revenue, dramatically.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage. The model supports ticket price increase by $18.07, leading to $31,629,630 in revenue over the course of the season.
3. Same as number 2, but adding 2 acres of snow making cover. The model supports ticket price increase by $19.31, leading to $33,800,926 in revenue over the course of the season.
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres. The model would not support a change in ticket price in this scenario.

Conclusions

Big Mountain Resort recently installed a new chair lift that increases their operating costs by $1,540,000 for this season. Normally, they charge a premium over the average ticket price of resorts in its market segment. However, they could be capitalizing on Big Mountain’s amenities further and charge more for them, especially to offset the price of the new lift. According to a random forest model, ticket prices at $92.63 would be justified. The model also determined that some features are more valuable than others, and, when adjusted, the ticket price could be changed. For example, increasing vertical drop by adding a run that would require another chair lift would justify raising price by $18. Some additional data that could be useful, other then values for missing data, is operating costs of other facilities. This could possibly make a model more robust in the future.