How can Big Mountain Resort increase revenue by 2% in the next year through higher ticket price, advertisement, or adding/closing some facilities such as spa, lift chairs, or runs?

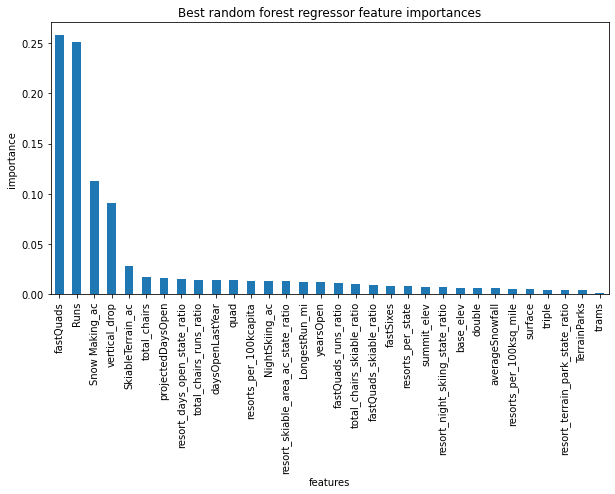
In the first step we focused on collecting the data, organizing it, and making sure it's well defined. We checked the data for the missing values, and removed 'fastEight' and 'AdultWeekday' features. 'fastEight' has the most missing value in the data. Weekday and weekend prices in the Montana state are equal and weekday prices have more missing values of the two. Therefore, 'AdultWeekend' consider as the target feature to predict ticket price and the records with the missing values are removed. We also looked at distributions of features to check if there are any possible outliers. There were some obvious issues with some of the features in the data that led to a data error being corrected, and some other records dropped. The distribution is much better after this cleaning. The cleaned data has 277 records and 25 features.

We also created the statewide summary before removing any missing record. TerrainParks, SkiableTerrain\_ac, daysOpenLastYear, NightSkiing\_ac are the selected features for this purpose. In addition, state population and state area were added by extracting data from online sources.

In the next step, we explored the statewide summary data to see if there is any pattern between the states. We didn’t see any and decided to treat all states equally. We also merged the statewide summary data into the ski resort data and calculated the ratio of each resort to the total state for the statewide summary features: skiable area, days open, terrain park count, and night skiing area. The correlation between variables and the ticket price are also checked.

There are quite a few reasonable correlations with the ticket price. Regarding the state features, ‘resort\_night\_skiing\_state\_ratio’ seems the most correlated with ticket price. Moreover, ‘resorts\_per\_100kcapita’ shows something interesting. When the value is low, there is quite a variability in ticket price, although it's capable of going quite high. Ticket price may drop a little before then climbing upwards as the number of resorts per capita increases. Ticket price could climb with the number of resorts serving a population because it indicates a popular area for skiing with plenty of demand. The lower ticket price when fewer resorts serve a population may similarly be because it's a less popular state for skiing. For the other features, there's a strong positive correlation with ’vertical\_drop’. Also, ‘fastQuads’ seems very useful. ‘Runs’ and ‘total\_chairs’ appear quite similar and also useful. At first these relationships are quite counterintuitive. It seems that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. What we may be seeing here is an exclusive vs. mass market resort effect; if you don't have so many chairs, you can charge more for your tickets, although with fewer chairs you're inevitably going to be able to serve fewer visitors. Your price per visitor is high but your number of visitors may be low. Something very useful that's missing from the data is the number of visitors per year. The total skiable terrain area is not as useful as the area with snow making ‘Snow Making\_ac’. People seem to put more value in guaranteed snow cover rather than more variable terrain area. It also appears that having no fast quads may limit the ticket price, but if your resort covers a wide area then getting a small number of fast quads may be beneficial to ticket price.

Furthermore, we started building machine learning models. To evaluate the model performance, we split the data into training and testing sets. For the first model, the average price is considered as our prediction. MAE is 17.9 for training data and 19.1 for the testing data. It means, in this model we expected to be off by around $19. Next, we built a linear regression model. The missing data was imputed with the median. ‘Vertical\_drop’, ‘Snow Making\_ac’, ‘total\_chairs’, 'fastQuads,’ ‘Runs’, ‘LongestRun\_mi’, ‘trams’, ‘SkiableTerrain\_ac’ are the significant features. These results suggest that vertical drop is the biggest positive feature. The cross-validation MAE for the training set is around 10.5, and for the testing set is 11.8. This means the model is off by around 11.8 dollars, which is almost 7 dollars less compared with the average price model. We also tried random forest model. We tried the model with and without feature scaling, and tried both the mean and median to impute missing values. The results show that imputing with the median helps, but scaling the features doesn't. The MAE is 9.6 for the training set and 9.5 for the testing set. We can verify performance on the test set produces performance consistent with the cross-validation results. Random forest model has lower cross-validation mean absolute error by almost $2 compared with the linear regression model. It also shows less variability by comparing the std of the mean absolute error of the models. Therefore, we decided to use “Random Forest” model. Below is the bar plot of the random forest’s feature importance.

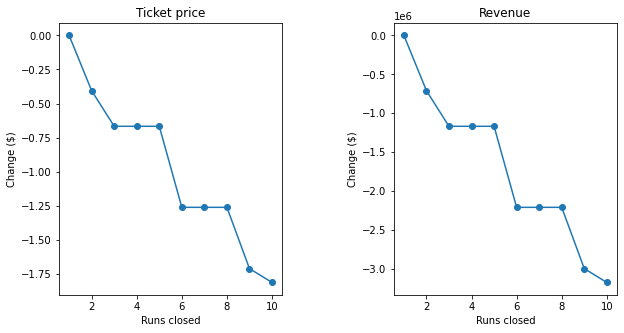


With using the random forest model explained above, the modeled price is 95.87 dollar. Big Mountain Resort currently charges 81 dollars for the ticket price. Even with the expected mean absolute error of 10.39 dollar, this suggests there is room for an increase. Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases the operating costs by 1,540,000 dollar this season. Increasing the ticket price to 85 dollars (95.87-10.39=85.48) can increase the revenue by almost 7,000,000 dollars for the whole year and cover the cost of the additional chair lift.

The resort operates within a market where people pay more for certain facilities, and less for others. Below shows four scenarios that can help business to improve the revenue:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.

* As shown in the following figure, the model shows closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.



1. 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.

* This scenario increases support for ticket price by 1.99 dollar over the season, this could be expected to increase revenue by amount of 3,474,638 dollar.

1. Same as number 2, but adding 2 acres of snow making cover

* The results are exactly like scenario 2. Such a small increase in the snow making area makes no difference.

1. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

* The price change is zero. Although the longest run feature was used in the linear model, the random forest model (the one we chose because of its better performance) only has longest run way down in the feature importance list.

Scenarios one with 5 runs and scenario two seems more profitable and are recommended for future consideration.