Big Mountain Ticket Prediction System

An Exercise in Price Modeling for Ski Resort Tickets

Guided Capstone Project
Kendall T. Herron

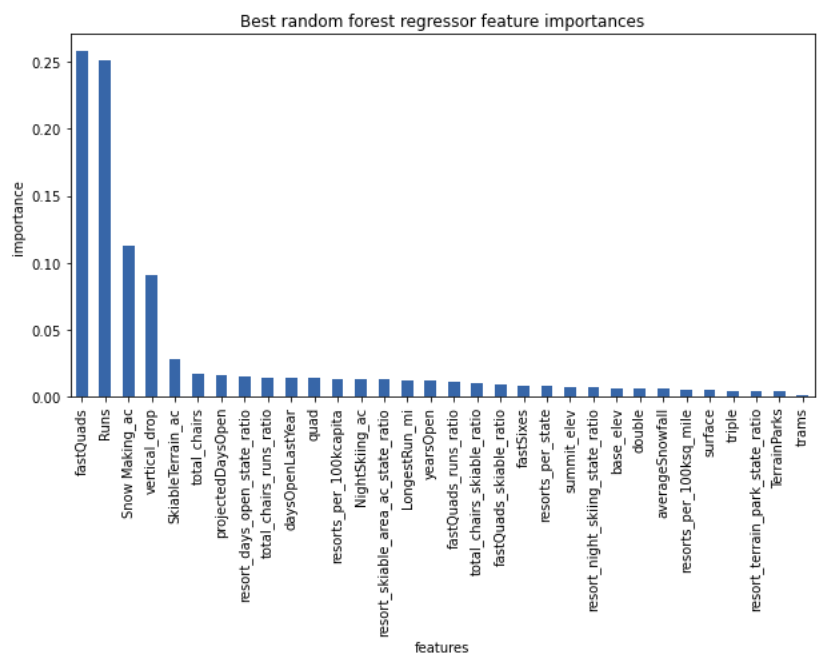
Big Mountain Resort is a well-known ski resort that offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. The resort has recently installed an additional chair lift that has increased there operating costs by $1,540,000 this season. Management has considered several changes that should cut the costs without undermining the ticket price and is willing to support a higher ticket price. Big Mountain is suspicious it may not be maximizing its returns, related to its position in the market.

Starting out with observing the data. The number of columns is 27 and the number of rows is 330. The Big Mountian Resort is present in this data. Observing duplicate names, these names are unique records. I dropped fastEight and AdultWeekday columns to narrow down my search. Removed all rows that had no price data as there is no use. I ended with the number of columns at 25 and the number of rows at 277. The target feature is AdultWeekend price.

There were three categorical features in the data, Name, Region, and state. The rest of the data was numerical. I removed two columns, state population and state size. I replaced the index with the state. There wasn't any specific pattern with the relationship of state and ticket price. I considered all states together when building a pricing model. Although, there was potentially relevant state data in features relevant to the business use case. There is a positive correlation between the ratio of night skiing areas and the number of resorts per capita.

Resorts that are more densely located with population tend to provide more night skiing. Three columns stand out with our target feature. fastQuads, Runs, and Snow Making_ac. People seem to put more value in guaranteed snow cover rather than more variable terrain area based on the correlation between the columns Snow Making_ac and SkiableTerrain_ac. Runs and total_chairs are correlated with ticket price. The more runs you have the more chairs you'd need to ferry people to them. The vertical drop seems to be a selling point that raises ticket prices as well.

If the deciding factor was to take the mean average as a predictor, it would cause the adult weekend ticket price to be way lower than the original listed adult weekend ticket price for the Big Mountain Ski Resort. Averaging around 19 dollars less compared to the original price. Using the average would not suffice in improving the business for the ski resort. If the deciding factor was a linear model it would lower the adult weekend ticket price 9 dollars, which would be a far better result than taking the mean average. When trying a different option, replacing the values with the median average predictor, it didn't make a difference from the mean average. Refining the linear model to select the k best features. Using cross-validation would partition the training set into k folds, calculate performance on the fold not used in training, and train the model on k-1 of those folds. Cross validation would build k models on k sets of data with k estimates of how the model performs on unused data without having to touch the test set. I will use cross-validation for multiple values of k and pick the value of k that gives the best performance. In this case k=8 has the best performance. Finding the linear model coefficients and matching the feature names provide enough evidence that the vertical_drop is the biggest positive feature. Random forest was used, the regressor provided a lower cross-validation mean absolute error by almost 1 dollar. It had less variability and the performance on the test set produced consistent results with cross-validation therefore I'm choosing to use the random forest model.



The Big Mountain Ski Resort weekend adult ticket price is listed at 81 dollars. The model created suggests four different scenarios the business can consider. First scenario, closing 1 run will make no difference. Closing 2 to 3 runs can successively reduce support for ticket price therefore revenue. Second scenario, adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. Increases support for ticket price to 8 dollars and 61 cents and over the season 15,065,471 dollars. Third scenario, repeat the second scenario and add 2 acres of snow. Increases support for ticket price to around 9 dollars and 90 cents and over the season 17,322,717 dollars. Fourth scenario, increase the longest run by .2 miles and guarantee its snow coverage by adding 4 acres of snow making capability. No difference to this scenario. However, the longest run feature was used in the linear model. We chose a random forest model for the better performance, and it had the longest run far down in the feature important list. Based on the model performance and the ticket price predictor. Scenario 2 would be the best business strategy to take due to installing the additional chair lift and the expected amount over the season being 15,065,471 dollars. In scenario 3, further consideration would be suggested as an increase of two million over the season can be made depending on the cost of making 2 acres of snow throughout the season.

Some hard takes that could happen next. I believe there could have been more done to explore possible hyperparameters. Possibly removing least useful features. Calculating and storing features adds cost and dependencies. If the features are not needed, they should be removed. Building a simpler model using the four features would have worked and would be easier to explain to stakeholders. The cost to make snow per acre over a season could have been useful for considering scenario 3. The big reason the modeled ticket prices were high was mostly due to the addition on the run, the increase of a vertical drop, and the install of an additional chair lift. These new resort features bring the model to increase the ticket per adult. The higher ticket price wouldn’t come as a surprise as it would be advertised based on the newly introduced features for the resort.