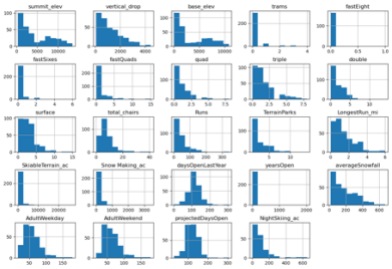
Tyler McRorie

Guided Capstone Project Report

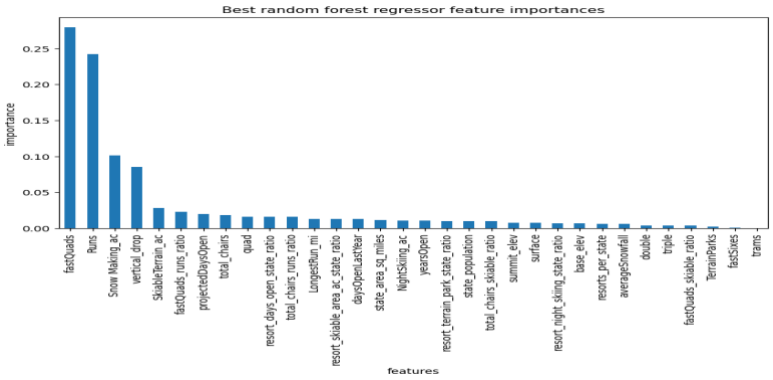
Big Mountain Resort recently installed a new chair lift to help increase the distribution of visitors across the mountain. In turn, this has increased their operating costs by $1,540,000. To recoup the increase in operating costs, leadership at Big Mountain Resort have invested in a data-driven business strategy to evaluate the resort's pricing strategy. The following are my team's findings with respect to the investment strategy.

The original data set was wrangled from a csv file and had 330 total rows, the first being the index, and 27 total columns. We were able to identify that approximately 16% of ticket prices were missing from the data set, though, we waited to identify distributions of the other 14% of ticket prices prior to dropping these values, they were eventually dropped. We then evaluated the amount of missing data in the rows, i.e. the resort data, we made ourselves aware of how much of that data was missing and what values were missing, but ultimately did not remove any of the rows or outliers. When wrangling the data for the target feature, ticket price, we needed to determine the relationship between weekend and weekday ticket prices among all resorts and then among resorts in Montana. A visualization of the distribution of target features can be seen below.



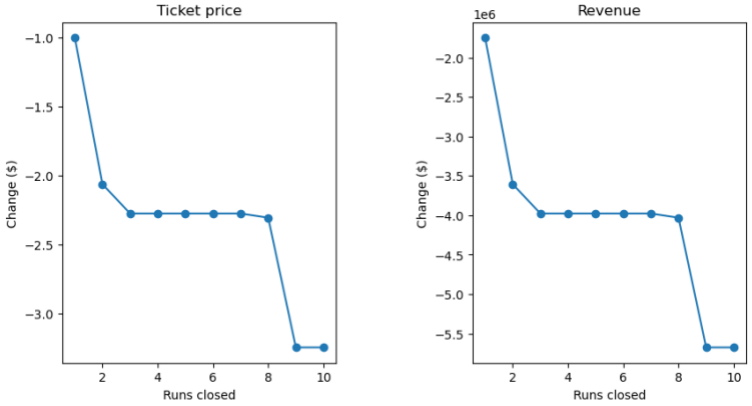
Once the data was in a format that we could evaluate, we performed exploratory data analysis, which revealed a great deal of numerical features in the data, which would prove to be helpful in comparison between resorts. When we were evaluating the ticket price with respect to state there were some justifiable outliers, but, overall, the state data can be treated equally. Thus, we can now look at the numerical features without sub-setting them at the state or regional level. When it comes to a target feature for modeling, ticket price will be our dependent variable and we will use some combination of the other features we have engineered and the natural data in the set.

Next, we preprocessed the data, first we attempted a simple linear regression model, but we felt that we could improve on its metrics. Instead, we attempted a RandomForest Regression model. Like the regression model, we defined a pipeline that showed us the median strategy was more effective than feature scaling. We performed cross-validation and showed an improvement using the Random Forest Regression model. This model identified the FastQuads, Runs, Snow Making, and vertical drop as the most reliable features as seen in the following graphic.



After comparing the two models, the random forest regressor has a lower absolute error which makes it more attractive. In addition, it shows less variability. It's performance on the test set is consistent with those from the cross-validation results.

Now that we found an effective model, we applied it to the data set and some business scenarios. Big Mountain currently charges $81.00 for tickets. Our model suggests that a ticket price of $95.97 could be supported by the resort's facilities. Though, we would like to caution that the model assumes that the fair price of tickets in general is assumed, and that it is possible that other resorts are mispricing their tickets. Recall that a new chair lift resulted in a significant increase in operating costs. As stated in the report an increase in ticket price of $1.45 per ticket would result in an annual increase of $2545455.00, which more than covers the new operating costs. I would recommend increasing the ticket price, and potentially closing some of the runs during non-peak portions of the year to reduce operating costs further. The business could start with closing one run and reevaluate, though the model suggests you could close as many as 6 without a significant change in revenue, seen below.



In all, we have taken the available data, wrangled it, processed it, analyzed it, and modeled it to find an improved ticket pricing strategy as well as other business scenarios that will result in reduced operating costs. The modeled price was higher than the current price because the model considered data from across the country. Big Mountain is a destination ski resort and should be competing at this level. Since this is the case, the model found that Big Mountain has similarities to other high-level resorts that charge more for their ticket prices and are successful in this. This price mismatch might come as a surprise, but again, I would urge the reader to remember we are thinking on a nationwide scale.

Moving forward, if the business wanted to make use of the model, then we could provide an API implementation that allows the user to use the model and draw new conclusions once new parameters are made available.