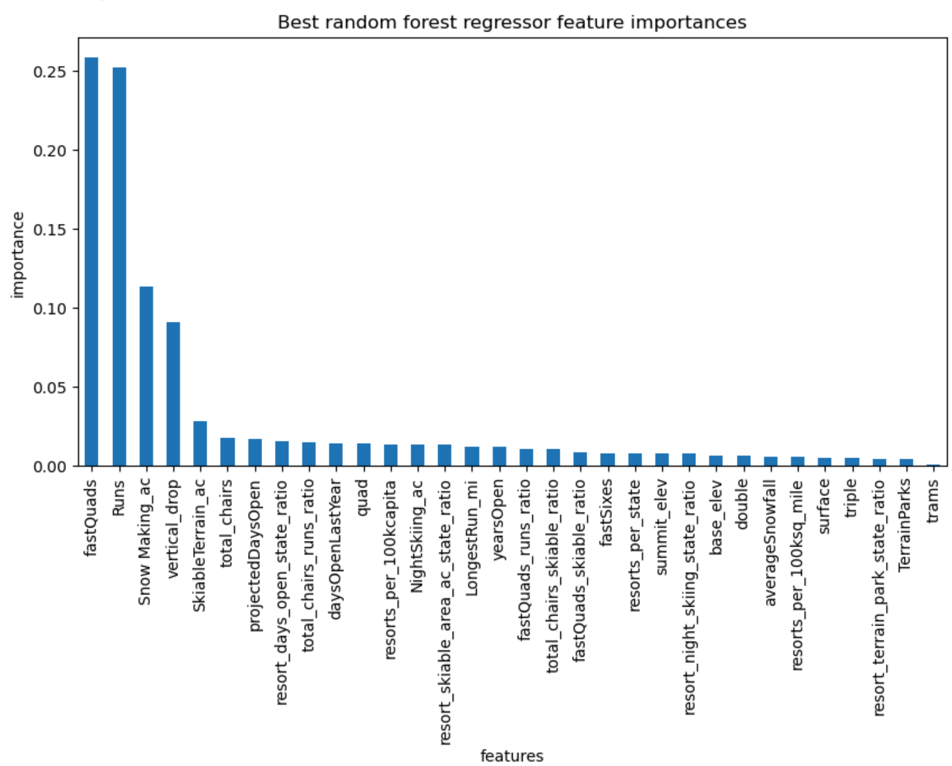
By analyzing the data regarding the ski resorts in the United States, we want to predict how a resort's facilities affect operating costs and ticket prices so we can decide how Big Mountain Resort can cut costs and raise ticket prices over the course of this season.

Our first challenge was to find the target feature and prepare the ski resort data for modeling. There was no operating cost data, so the feature we naturally targeted was ticket price, which the data represented as adult weekday and adult weekend ticket prices. Before deciding which to use, we had to deal with lots of missing values (even dropping the fastEight column entirely for lack of data) and remove questionable outliers that would have skewed the analysis. After all this, we decided to target adult weekend ticket prices because there were fewer missing values.

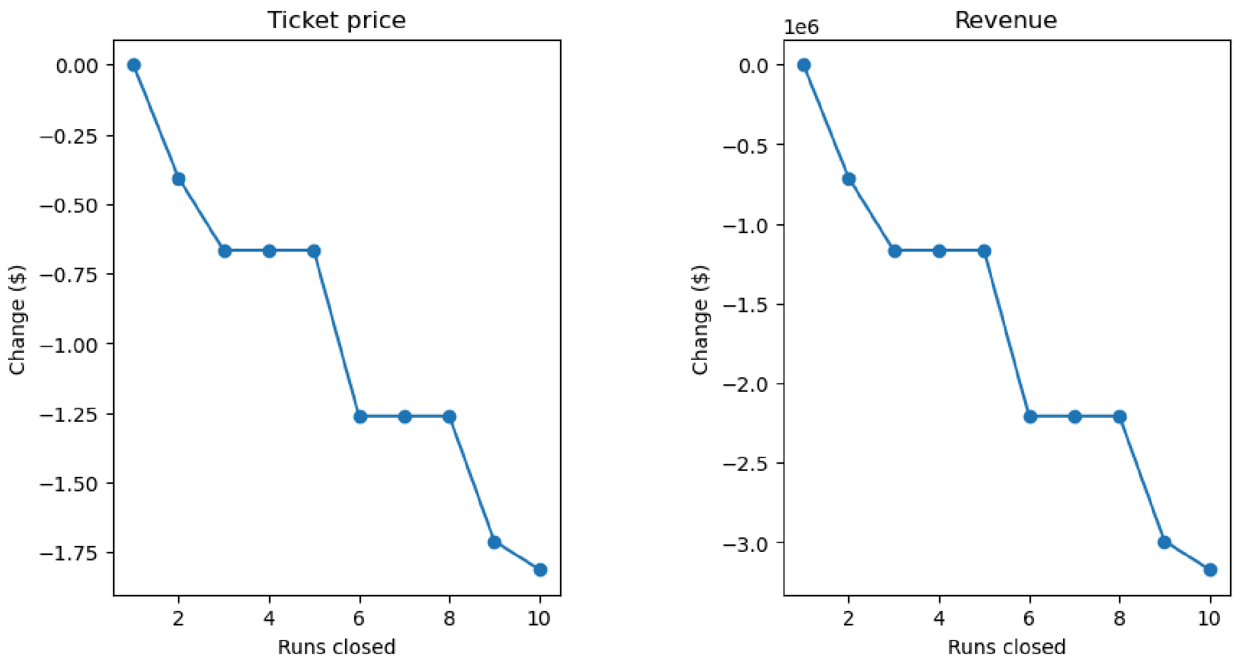
We were left with various numeric data and each resort's state, and we explored how these affected ticket prices. Because a resort's state is not numeric, we added new features based on state data (e.g. resorts per 100k capita, resorts per 100k sq miles) to represent it numerically. With all the data represented numerically, we engineered some feature ratios and analyzed the relationships between the features via PCA, feature correlation heatmaps, and scatterplots. We found that a resort's state had no clear relationship with ticket price (and could be safely ignored), but other features (e.g. vertical drop, number of fast quads, number of runs, total number of chairs) were more correlated.

After cleaning and exploring the numeric and categorical data for patterns, we built a few machine learning models to predict ticket prices. We first used the average ticket price to establish a baseline for model performance. We then used pipelines to impute the median for, scale, and train a linear regression model and a random forest model, which we then evaluated via cross-validation. We found that the random forest model performed the best and did not overfit to our training data, and we used it to model our cost-cutting and price-raising scenarios. But in both models, we found that fastQuads, Runs, snow making, and vertical drop were the most important features.



We trained the random forest model on all the data (excluding Big Mountain) before using it to model Big Mountain's ticket price. The predicted price was $95.87, whereas the actual price was $81. Even with the expected mean absolute error of $10.39, this suggests there is room for a price increase.

Assuming 350,000 visitors and that each buys 5 tickets, we model the four scenarios: closing up to 10 least used runs, increasing the vertical drop by 150 feet (which requires an additional chair lift), adding 2 acres of snow making cover in addition to the vertical drop, and increasing the longest run from 3.3 miles to 3.5 miles (which requires increasing snow making coverage by 4 acres). A graphic of the first scenario is shown below; closing some runs had no effect on ticket price/revenue, while closing other runs dropped the ticket price/revenue by up to $0.75/$1.2M. In the second scenario, ticket price/revenue increases by $1.99/$3474638. The last two scenarios had no effect on ticket price, so they aren't worth considering.



To conclude, we can follow the direction of the first scenario to cut costs, and we can follow the direction of the second scenario to raise ticket prices/revenue. From the first scenario graphic, we know we can close at least one run because closing just one run doesn't affect ticket price/revenue. And we may be able to use the second scenario as cover to raise ticket prices toward $100.

But before enacting any of these changes, we still need to know Big Mountain's operational costs and how these scenarios will affect them. We may not be able to reduce the number of runs if we lose more revenue than operational costs, and we may not be able to increase the vertical drop and add the additional chair lift if the extra operational costs exceed the revenue gained. A complete analysis demands that we also model the costs and not just the revenue. And to model the costs, we need data; not just from Big Mountain, but also from all the other ski resorts in the United States.