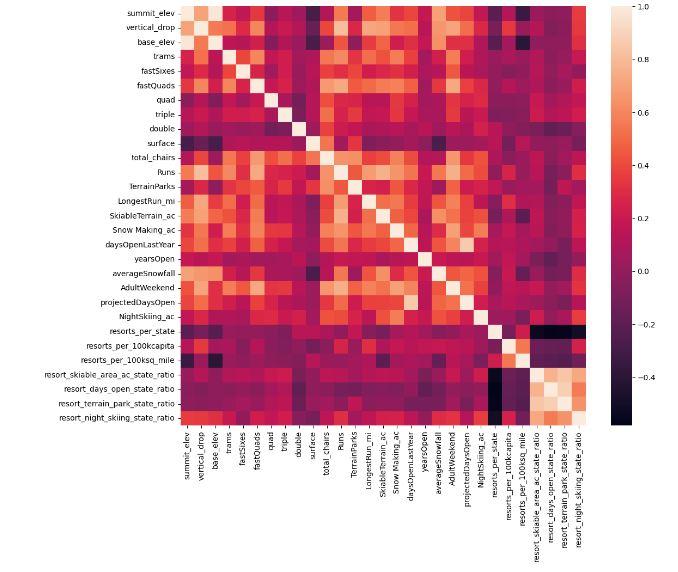
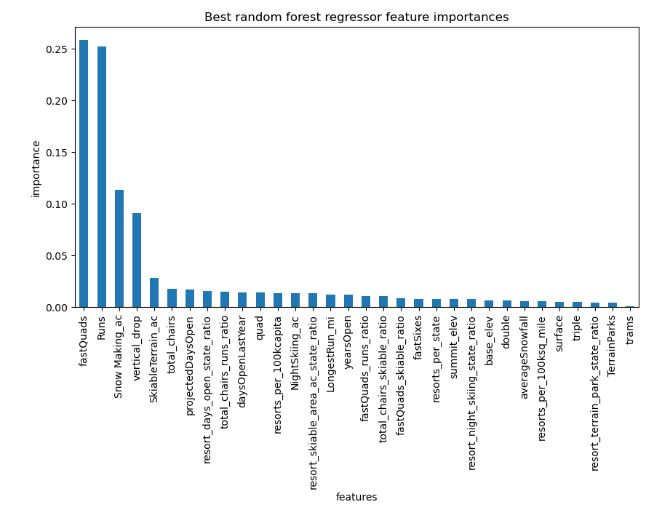
This project focused on creating a predictive ticket pricing model for Big Mountain Resort based on resort facilities in order to guide Big Mountain’s future ticket pricing strategy and facility investment plans. Big Mountain Resort was seeking to reevaluate their ticket pricing strategy to increase revenue in order to compensate for an added $1.54 million to their seasonal operating costs due to the installation of a new chairlift. They suspected that their resort facilities were not being fully optimized and provided data from ski resorts across the country in their market segment in order to assist in the creation of this ticket pricing model.

Data wrangling of the provided ski resort data helped identify the required target value of Adult Weekend Ticket Price as our dependent variable, and uncovered potentially useful features and fundamental issues with the data. This step was crucial in taking the raw provided data, becoming familiar with the numerical and categorical features, plotting distributions to begin to visualize data quality, and dealing with missing data and outliers. This step also gathered useful state-wide summary statistics for our market segment and joined in state population and total area data into our ski resort data. We were able to prepare the data for exploratory data analysis and begin to hypothesize about how resort features were distributed across states, how prices were distributed across states, and how weekday and weekend prices were related. It was determined that, in Montana (where Big Mountain is located), adult weekday prices and weekend prices were equal, and there were fewer missing data values for weekend prices overall, making Adult Weekend Price our target value.



Exploratory data analysis first focused on a lengthy exploration of the state-wide data that was summarized during data wrangling to determine if state-level analysis would be helpful. This helped us identify some potentially useful features, such as total state area and total state population, as well as measures of resort density per state. No clear pattern suggested a relationship between state and ticket price, so it was determined that all state labels would be treated equally when building the predictive pricing model. Using a seaborn correlation heatmap (left) on our features, we focused on our variable of interest for modeling, Adult Weekend Price, and saw that the four features with highest positive correlation with Adult Weekend Price were FastQuads, Runs, Vertical Drop, and Total Chairs.

To begin modeling the data, we began with a linear regression model on scaled data and applied it to a training set (70% of our dataset) and a test set (the remaining 30%), excluding data for Big Mountain Resort. Using pipelines, we were able to efficiently impute missing values, scale the data, and train the model. Model performance was assessed with R2 , median absolute error, and mean squared error. The model performed similarly when we imputed missing results first with the median of the training set, then with the mean, with a lower R2 for our test set, indicating an overfitting. This led to a refinement of the linear model with dimensionality reduction using SelectKBest, and then using cross-validation to estimate the performance. According to our linear regression model, Vertical Drop, Snow Making acreage, Total Chairs, FastQuads, and Runs were the five biggest positive features, which agreed with the conclusions drawn during EDA. 

We continued modeling by defining a pipeline for a Random Forest Model that imputed the mean and median, used the standard scaler or no scaler, and assessed performance using cross-validation. This allowed us to explore hyperparameters in our random forest in order to plot the importances of the features (right). FastQuads, Runs, Snow Making acreage, and Vertical Drop again proved to be important features. Ultimately after assessing and comparing performance of both models, we decided to use the Random Forest Model with the mean value imputed and no scaling, for its lower cross-validation mean absolute error and less variability between training and test set results.

Our last steps included refitting our model on all available data, excluding Big Mountain, so that we were left with a predictive model for Big Mountain’s ticket price based on features data from all other resorts. We then applied Big Mountain’s data and found our model predicted a ticket price of $95.87 (actual price of $81) and an expected mean absolute error of $10.39, indicating that there is room for a ticket price increase. However, we lacked a market context for this result, so plotted Big Mountain’s data against that of other resorts in Montana for ticket price, vertical drop, snow making area, total number of chairs, fast quads, runs, longest run, trams, and skiable terrain area. Big Mountain landed on the higher end of almost all of these feature distributions, indicating its above-average offerings compared to other resorts in the state. We then modeled four different scenarios for changes to facilities that Big Mountain suggested in order to determine how ticket price would be affected. Scenario 1 modeled how closing 10 of the least used runs would affect ticket price. Our model determined that closing one run makes no difference to ticket price, while closing 2-3 reduces support for ticket price increases, and anything below 6 closed runs leads to a large drop. This supported the business idea of closing at least 1 run to decrease operations costs, but gave no support for a ticket price increase. Scenarios 2 and 3 were very similar, calling for an increase of the vertical drop, installation of an additional chairlift (already completed at Big Mountain), and scenario 3 would also add 2 acres of snow making cover. Our model outputted identical results for both scenarios, with support for a $1.99 increase in ticket price to generate $3,474,638 in revenue over the season (assuming 350k visitors). Scenario 4 called for an increase in the longest run by 0.2 miles, requiring an additional snow making coverage of 4 acres. Because our random forest model did not consider the longest run as an important feature, the model did not predict any ticket price increase for this scenario.

Overall, our random forest model supports a $1.99 ticket price increase to $82.99 with the addition of the new chairlift at Big Mountain and an increase in vertical drop, but predicts that the facilities could support a $95.87 ticket. Additional recommendations include the closing of 1-3 of the least popular runs to decrease operating costs while not affecting ticket price, and possibly pursuing increased snow making acreage, which is an important feature and can be combined with other facilities changes to support increased ticket prices.

A better idea of operating costs, the true market context of resorts in Montana, and pricing strategies at other resorts would help us conclude if an even greater increase to approach, meet, or even exceed the predicted ticket price of $95.87 would be justified. Future work on this project could continue to test out different scenarios that would assist Big Mountain in decreasing operations costs and optimizing their facilities to further support ticket price increase. We can also collect more data on operating costs and resort pricing strategies to refine our model and gain more confidence in its predictions. There is also a lot to explore with the state-level data, which did not play a major factor in the development of our model, but raised many questions about how and why certain features are distributed across states. This information may be useful in determining pricing strategies for different amenities and different populations.