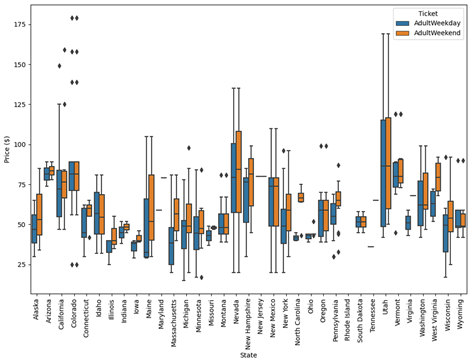
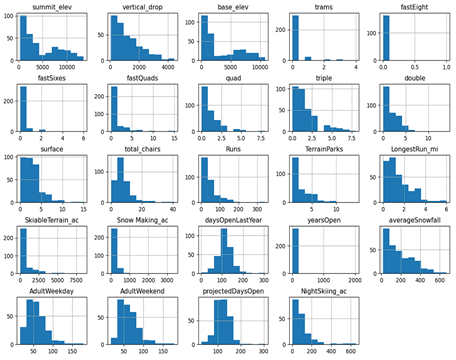
Guided Capstone Project Report

Big Mountain Resort is looking for some guidance on how to better price tickets compared to other resorts in its market segment to help determine increased profits by 6% to offset the new $1.54MM operating cost of the new chair lift in the next 12 months.

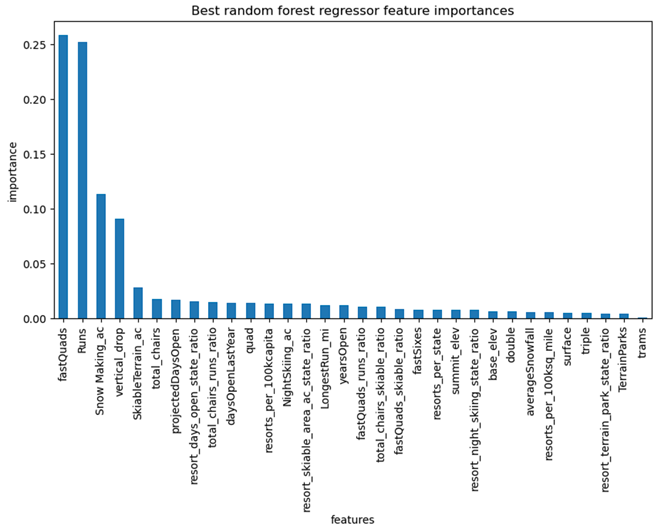
I started by looking at various statistics and information regarding the data in general. Some key areas of interest were if there were duplicate entries for ski resorts and missing data values. We used several plotting tools to look at the number of resorts per state, the relationship between 'state' & 'Region' and price differences between 'AdultWeekday' & 'AdultWeekend' prices of our dataset. The seaborn boxplot was particularly illuminating on the distribution of pricing in each state.

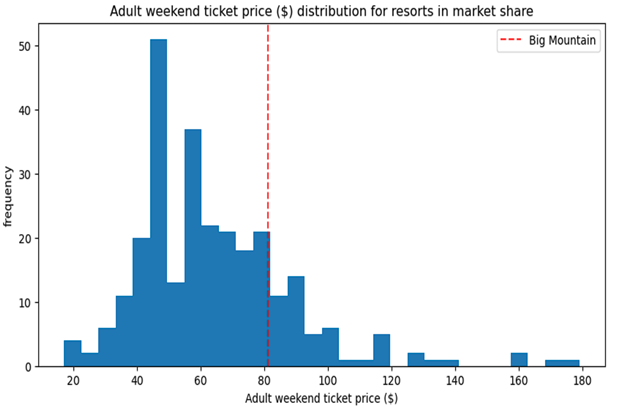
    Through a series of filtering of our data we explored these columns to determine if there were problems that could interfere with good data. After correcting a typo we found, we moved on to some exploration of the abundance of missing data in the 'fastEight' column so we proceeded to drop it. In the next steps we started a new data frame based on a summary of resorts in each state and some basic aggregate functions to help summarize. We then proceeded to explore missing prices and found several resorts with no price in both 'AdultWeekday' and 'AdultWeekend' so we dropped these rows as well. Next we started to examine population data on the internet to add to our state summary data frame. After some careful comparison between both data frames, we cleaned up the state summary, set state names and concatenated the population data to it. Finally we started to explore the pricing with a scatter plot and filters for our state in question to help determine which price to use for our further exploration. Since 'AdultWeekend' had the least number of missing values it was chosen and the 'AdultWeekday' column was dropped.

Next, we started off by creating a training set and testing set of data to begin the process of training a model. We used an average of the data set to get a baseline for checking our model's performance. We used r\_squared functions and the built-in options for linear regression. After calculating some Mean Absolute Errors and Mean Squared Errors we were able to start with using some of the built-in tools of sklearn to gain basic metrics. Next we moved on to establishing the best method for filling in missing values, mean or median. With these tests we came closer to an estimated ticket price of $9 versus the $19 from just an average standard deviation. We used pipelines to help automate some of the training processes and test different methods for better results with linear regression. Eventually we determined the best approach going forward was to use cross-validation to assess performance.

This eventually allowed us to use multiple values for our folding process to determine the best features tied to the ticket price. Using the model's coefficients we were able to pull the best performing features from our data set. Some of the best performing features from our dataset were 'vertical\_drop', 'Snow Making\_ac', 'total\_chairs', 'fastQuads', and 'Runs'.

Another good process for testing regression was the Random Forest Model. This allowed us to go straight from the testing process to cross-validating. Here we were able to use mean and median to determine out best testing methods. Through testing we reduced our predicted standard deviation and were able to determine the best features affecting ticket price were 'fastQuads' and 'Runs'. Other notable features were 'Snow Making\_ac' and 'vertical\_drop'. Using cross-validation to compare r2\_scores we were able to determine that random forest regression model performance was better than linear regression model performance. We were able to narrow down the predicted ticket price over $1 using this method. With a quick check on data quantity we determined we had enough data to proceed with the next steps.





Currently the weekend ticket price for Big Mountain is $81. According to the important facilities of Big Mountain and other resorts, ticket prices could increase as much as $14.87. There is a margin of error of $10.39 but still suggests an increase of $4.48 is viable.

When comparing the important features of Big Mountain to other resorts like vertical drop, snow making coverage, total number of chairs, number of fast quads, total number of runs, longest run length, and skiable terrain, Big Mountain sits near the top of features affecting ticket price. Using the model and running a few scenarios, it's predicted the following results:

1. Permanently closing down up to 10 of the least used runs:  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, 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 in revenue. Some additional data might further explore this option.

2. Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift:  This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3,474,638 additional revenue.

3. Big Mountain is repeating the previous one but adding 2 acres of snow making as well:  Such a small increase in the snow making area makes no difference according to our model!

4. Increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability:  The model predicts no difference whatsoever. 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 the longest run way down in the feature importance list.

Based on the 4 scenario model results, scenario 2 seems to yield the best results for overall profitability. Based on the modeling, Big Mountain resort seemed to score high in the desirable features for ticket price change. The model predicted a higher price due to the 'weight' of these features. Big Mountain could use this data to safely predict an increase based on only the current change of the additional lift by a small increase in ticket price. This model was saved for use with changing parameters to help analysts maintain further predictions in the coming years based on new data.