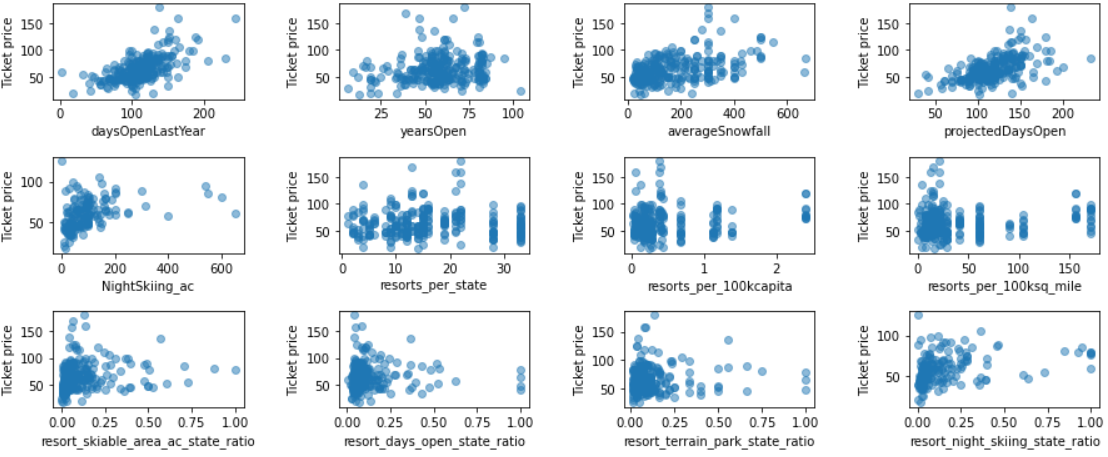
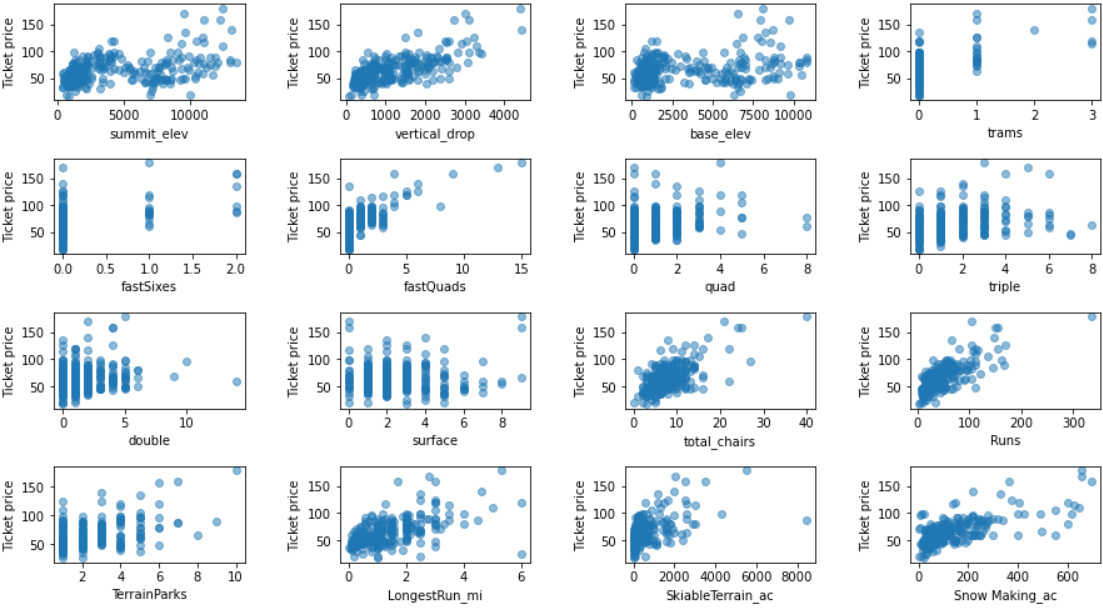
Guided Capstone Project Report

Big Mountain Resort tasked us with determining a ticket price for the upcoming ski season based on ski resort data and ticket price data from around the country. The big mountain board also wanted to identify what facilities they could upgrade or close to improve their resort income and maintenance costs. We completed data analysis on the data given and will give the results along with a recommendation based on this data.

First we cleaned the data, dropping any resorts without a ticket price and took a look at the values in each data column to check for any errors and looked for duplicate rows. We combined our data with state population and size. We separated our resorts by state to determine if there was any correlation between state and ticket price. We created new columns to show resorts per capita and resorts per area in each state.

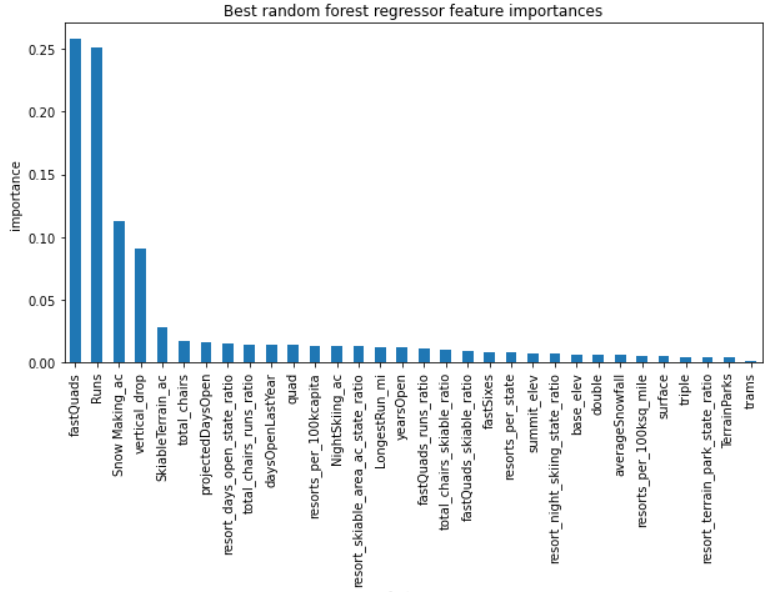
While the data didn’t provide any strong links between state and ticket price, we created scatterplots to show how the other data related.

From these plots, we can see that most do not show great correlation to price. Vertical Drop, Fast Quads, Total chairs, and Runs show the best correlation. Even though we didn’t see great significance in them, we kept all columns to run through our predictive models.

We then removed Big Mountain Resort from our data and split the rest into a 70/30 train/test data sets. To get a base accuracy level, we replaced all values with the mean for that column feature and calculated the r2 and MAE for the train/test set. With this base mean test, our predictions were within $19 dollars of the actual price. Not very good, as our potential ticket price could be anywhere in a range of $38.

Putting our data into a Linear regression model, filling our missing data with the median of that column, and scaling our data, we lowered our MAE to 9.40 which decreases our target ticket price range to under $20. This is a great improvement, but this MAE come from just testing our one test data set. We also used all of our values which could actually hurt or overfit our model. Because of this issue, we did a gridsearch function to find the best parameters for our model. This let us know, we only needed eight of our variables to make the best and most consistent predictions. We also used cross validation, which splits and retestsour data a number of times. Form this, we can calculate the average of how our model performed. The variables that made up the model are vertical drop, snow making ac, total chairs, fast quads, runs, longest run, trams, and skiable terrain. This refined linear regression model had an average MAE of 10.49 or a prediction within $10.49 of the actual price.

Trying to improve our accuracy, we created a Random Forest regression model. This model used all our data columns. Again, we used gridsearch to acquire the best parameters. In this model, the variables were not scaled, and median was used for missing data. This model used all variables but only four seemed of significance.



After cross validating this model, we get an average MAE of 9.64, almost a whole number better than the Linear model, and with a MAE standard deviation of 1.35, it also showed less variability in its performance. This is our recommended model moving forward.

After plugging in all data to train our model, our model gave us a final MAE or 10.39. It predicted a ticket price of $95.87 +- 10.39. With this estimate, i would recommend a ticket price of 87.00. This is currently on the low end of our predicted price, but our close competition in the state will all have lower prices. Data from our visitors about where they are from and other ski resorts they visit. Knowing what percentage of our visitors are local could give us a good indication of who and where is our main competition. This increase in price will provide an increase in overall income of $10.5 million.

The board also listed potential scenarios to improve ticket prices or lower maintenance fees. Some information is needed for the first scenario. Is there any maintenance cost of keeping runs open? If so, closing a couple runs may help with costs. Of the other scenarios, scenario #2 is the best option to improve ticket price while keeping cost down. This scenario would add a predicted $2 million in income after extra costs are deducted.