Guided Capstone Project Report

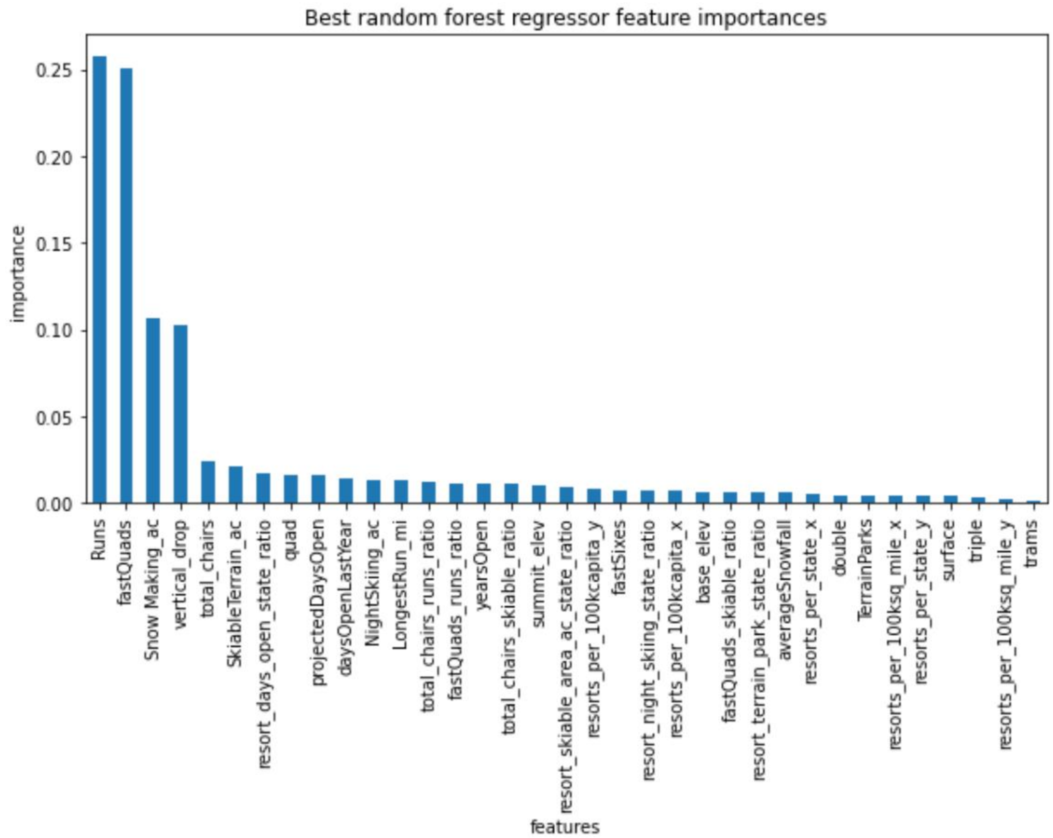
Big Mountain Resort, a ski resort located in Montana, suspects they may not be maximizing their returns relative to resorts in their market segment, so they wish to formulate a better pricing model.

We used ski data provided from the database manager, the data was cleaned up by: rearranging the columns and rows, ensuring that data did not overlap and were inputted correctly, and dropping rows with missing values. As we are curious in pricing, the AdultWeekday and AdultWeekend rows are observations of interest. Since there is a high correlation between both types of pricing, only the weekend prices were used for further analysis as it had less missing data.

Principal Cumulative Analysis (PCA) revealed that the first two components account for over 75% of the variance, so 2 components will be used in our model. Further analysis with a heatmap shows that ticket pricing is highly correlated with vertical drop, runs, fastQuads, total chairs, and snow making. These are variables of interest for predictor models.

The 70/30 train/test split method was incorporated to compare imputing missing values using median versus mean values and found that the results don't seem to differ much. Cross validation with a linear model showed that the value of k that gives the best performance was k = 8, with vertical drop , Snow Making_ac, total chairs, fastQuads and Runs being the top positive features, respectively.

The Random Forest Model was used and results show the dominant top 4 features are: fastQuads, Runs, Snow Making_ac, and vertical drop.



Between the best linear model and the best random forest model, the random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

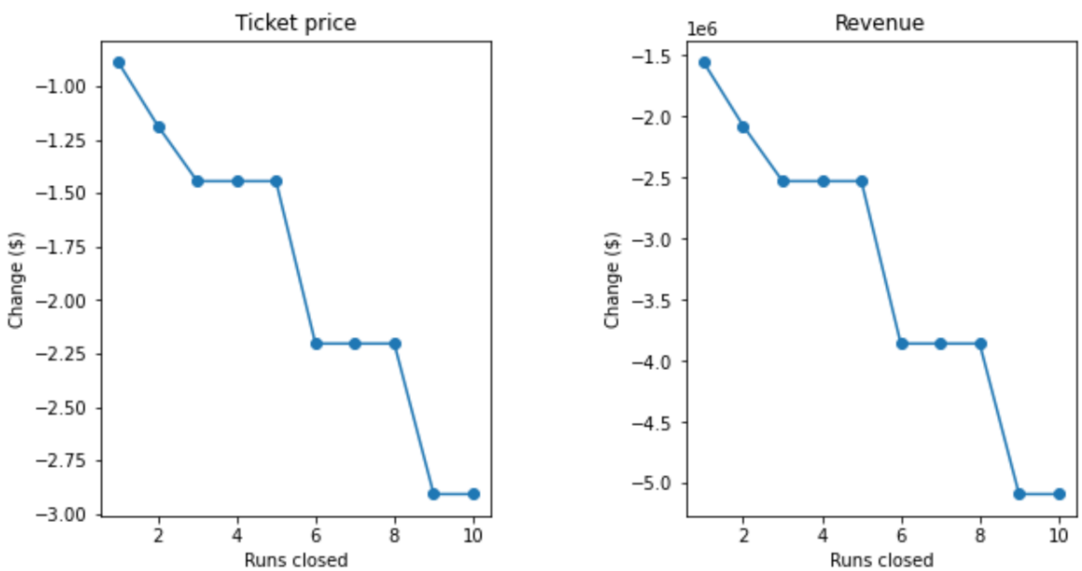
From the Refit Model: Big Mountain Resort's modelled price is $97.96, actual price is $81.00. Even with the expected mean absolute error of $10.36, this suggests there is room for an increase.

Big Mountain Resort has been reviewing potential scenarios for either cutting costs or increasing revenue (from ticket prices) Being able to sense how facilities support a given ticket price is valuable business intelligence. The business has shortlisted some options:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage

3. Same as number 2, but adding 2 acres of snow making cover
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

Modeling Scenario 1: The model says 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, it seems 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.



Modeling Scenario 2: This scenario increases support for ticket price by $2.22 Over the season, this could be expected to amount to $3888889

Modeling Scenario 3: This scenario increases support for ticket price by $2.22 Over the season, this could be expected to amount to $3888889

Modeling Scenario 4: 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 longest run way down in the feature importance list.