Background:
Recent install of an additional ski lift increased the operating cost for the season by $1540000. This install was done to increase the distribution of visitors across the mountain.

Our goal is to cut operating costs by $1540000 without change in ticket price and revenue or justify an increase in ticket prices to offset the increase in operating costs.

Initial EDA summary:

“Big Mountain Resort” in Montana did not have any missing data in the original data set compared to other resort data. There are no duplicate rows. There are 2 unique “Crystal Mountain” resorts in 2 different states (Michigan and Washington). Some states have multiple regions. There are 38 unique regions and 35 unique states. Rows with both AdultWeekday and AdultWeekend ticket prices missing (14.24% of original data). fastEight column was removed for having more than 50% of values missing. AdultWeekday column was removed after initial EDA (results listed below). There were 3 more entries in the weekend column than in the weekday column.

Weekend prices were higher than weekday prices for resorts with sub $100 pricing. Big Mountain resort currently prices both weekday and weekend tickets at $81.
Montana is 13th in the states with the most number of resorts. New York is 1st.

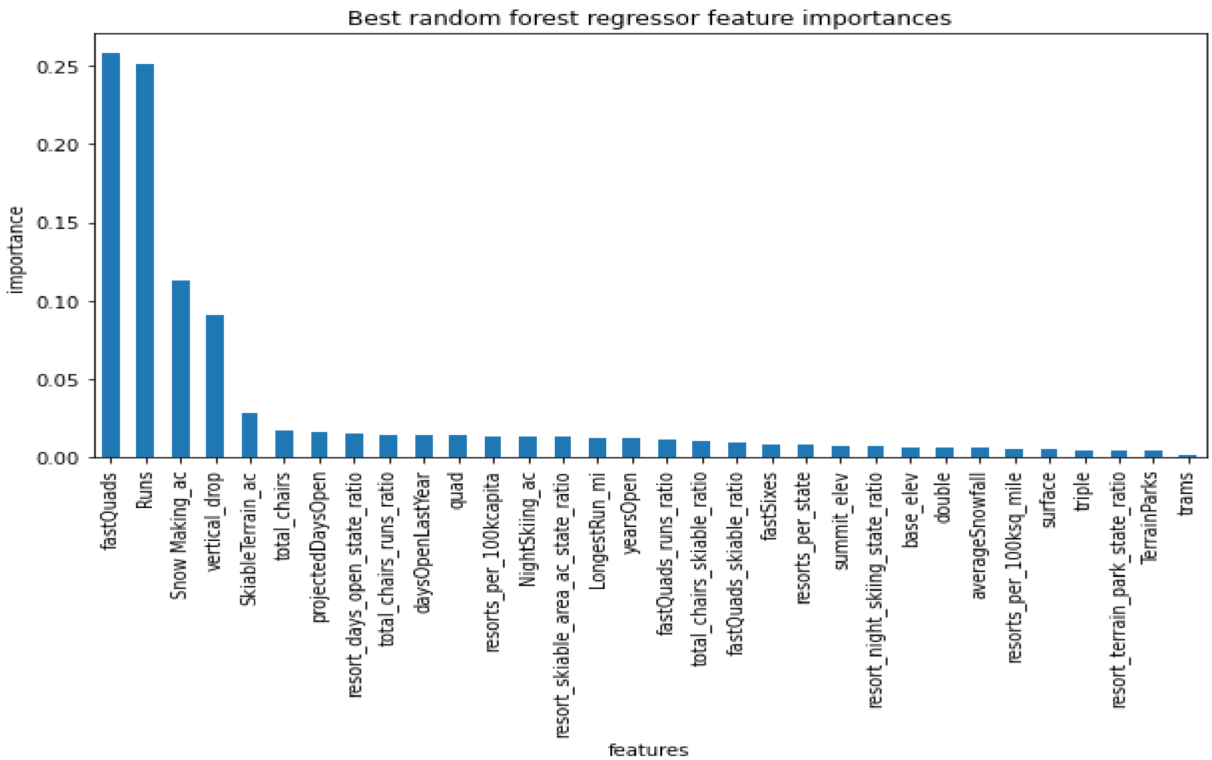
Categorical features present in the original data: Resorts per state, total skiable area, total days open, total terrain parks, total night skiing area, number of resorts per 100K population, number of resorts per 100K square miles.

There are 3 clusters of state vs ticket prices. Montana is on the low side of the cluster distribution. During EDA vertical drop, fast quads and numbers of runs are seen to have a significant correlation to ticket prices based on the correlation heat map.

Modeling:
Baseline average ticket price in the market is $63.81

A linear regression modeled with 70/30 training/test data split resulted in a 0.71 coefficient of determination. A hyper parameter search using a cross validation of 5 resulted in 8 best parameters (vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams, SkiableTerrain_ac). The model performed with a prediction R2 score of 0.79 on the training data and 0.64 on the testing data.

Using a random forest regression resulted in the dominant features (also present in linear regression) to be fastQuads, Runs, Snow Making_ac, vertical_drop. The mean absolute error is at 9.53 compared to linear regression mean absolute error at 11.79. Based on this information, Random Forest model is chosen for further analysis.



Initial findings based on model:

Scenario#1: Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
Impact: Permanently closure of least runs make no difference to negatively impacting revenue.

Scenario#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
Impact: This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3474638

Scenario#3: Same as number 2, but adding 2 acres of snow making cover
Impact: This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3474638

Scenario#4: Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres
Impact: Made no difference to revenue.

Next Steps:
1.) Data on operating costs of snow making and installation costs of an additional chair lift will allow for ROI evaluation of the above proposed scenarios.
2.) If we move forward with a ticket price increase, we need to evaluate if this needs to be done for weekday ticket, weekend ticket for both.
3.) Build a front end for business analysts to explore different scenarios.