11/4/2024

Guided Capstone Project Report

Big Mountain Resort Price Analysis

Nilesh Suresh

Context

Big Mountain Resort is a ski resort located in Montana, offers spectacular views of Glacier National Park and Flathead National Forest with access to 105 trails. This resort can accommodate skiers and riders of all levels and ability and every year, around 350,000 people ski or snowboard here. These are serviced by 11 lifts, 2 T-bars and 1 magic carpet for novice skiers. Resort recently installed an additional chair lift to help increase the distribution of visitors across the mountain which increases their operating cost by $1,540,000 this season.

Problem Statement

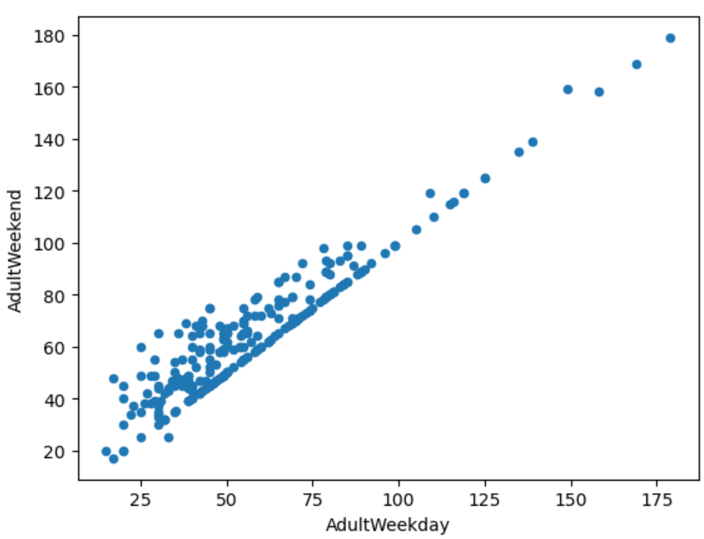
What strategies exists for Big Mountain Resort to reduce operating cost by 20% by the end of this year either by considering changes to cut down the costs or by increasing the ticket price.

Data Wrangling

Initial Ski Resort data contains 330 rows and includes our resort in interest 'Big Mountain Resort' which does not have any missing values. But there are missing values for other resorts. Looking at missing values by column shows:

- 'FastEight' column - just over 50% missing values. This column does not give any value for our model prediction and dropped the column.
- AdultWeekday and AdultWeekend ticket prices column - has 15-16% missing values. Identified resorts with missing both Weekday and Weekend prices and dropped those rows.

After the data correction and removing resorts with missing both AdultWeekday and AdultWeekend prices, now we have 277 rows. Most states had the same price for both Weekday and Weekend prices as seen by the chart below.

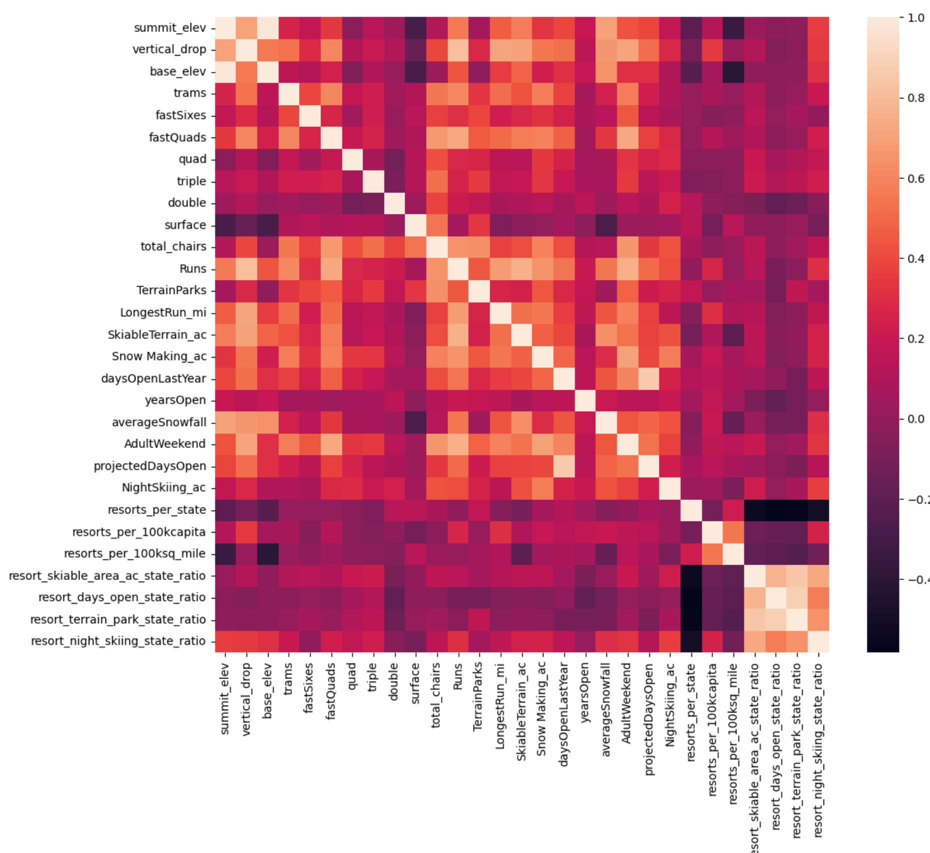


Again looked at the resorts with missing either AdultWeekday or AdultWeekend price found AdultWeekday prices have the least missing values of the two, so dropped the AdultWeekday prices column and then keep just the rows that have weekend price. Total number of rows after the data cleanup is 277.

Hence the target feature is AdultWeekend price for modelling.

Exploratory data Analysis

By exploring the state summary for each numerical features such as resorts per state, total skiable area, total days open, total terrain parks, total night-sking, state population and state area sq miles, trends seems various between states. In top 5, Montana is included only in total skiable area and total state area. In order to disentangle this interconnected relationships, need to use the technique PCA. This technique will find linear combinations of the original features that are uncorrelated with one another and order them by the amount of variance they explain. Cumulative variance ratio explained by PCA components for state/resort summary statistics shows that first two components accounts for 75% of the variance. Then for these first two components, plotted a scatter chart by state. In order to better visualize relationship between each feature, created a heatmap as shown below.



When look at correction between target feature (AdultWeekend ticket price) with others, these features shows a strong correlation - Vertical_drop, Trams, fastQuads, total_chairs, Runs, Snow Making_ac.

- Snow Making_ac shows that visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up.
- New feature resort_night_skiing_state_ratio seems the most correlated with ticket price. Then perhaps seizing a greater share of night skiing capacity is positive for the price a resort can charge.
- As well as Runs, total_chairs is quite well correlated with ticket price. This is plausible; the more runs you have, the more chairs you'd need to ferry people to them!
- The vertical drop seems to be a selling point that raises ticket prices as well.

Finally, looked at some further features that may be useful in that they relate to how easily a resort can transport people around. These relationships brings new features - total_chairs_runs_ratio, total_chairs_skiable_ratio, fastQuads_runs_ratio and fastQuads_skiable_ratio. It seems that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. It also appears that having no fast quads may limit the ticket price, but if your resort covers a wide area then getting a small number of fast quads may be beneficial to ticket price.

Model Preprocessing with feature engineering

Before performing the model, target feature 'Adult Weeked' is partition sizes with a 70/30 trin/split. Then calucated train mean and skylearn's dummy regressor, both has the value 63.8 which closely matches.

To perform the median/mean comparison, performed the sequence of steps:

1. impute missing values
2. scale the features
3. train a model
4. calculate model performance

Both comparison don't seem very different. This gives a good place to start is to see how good the mean is as a predictor.

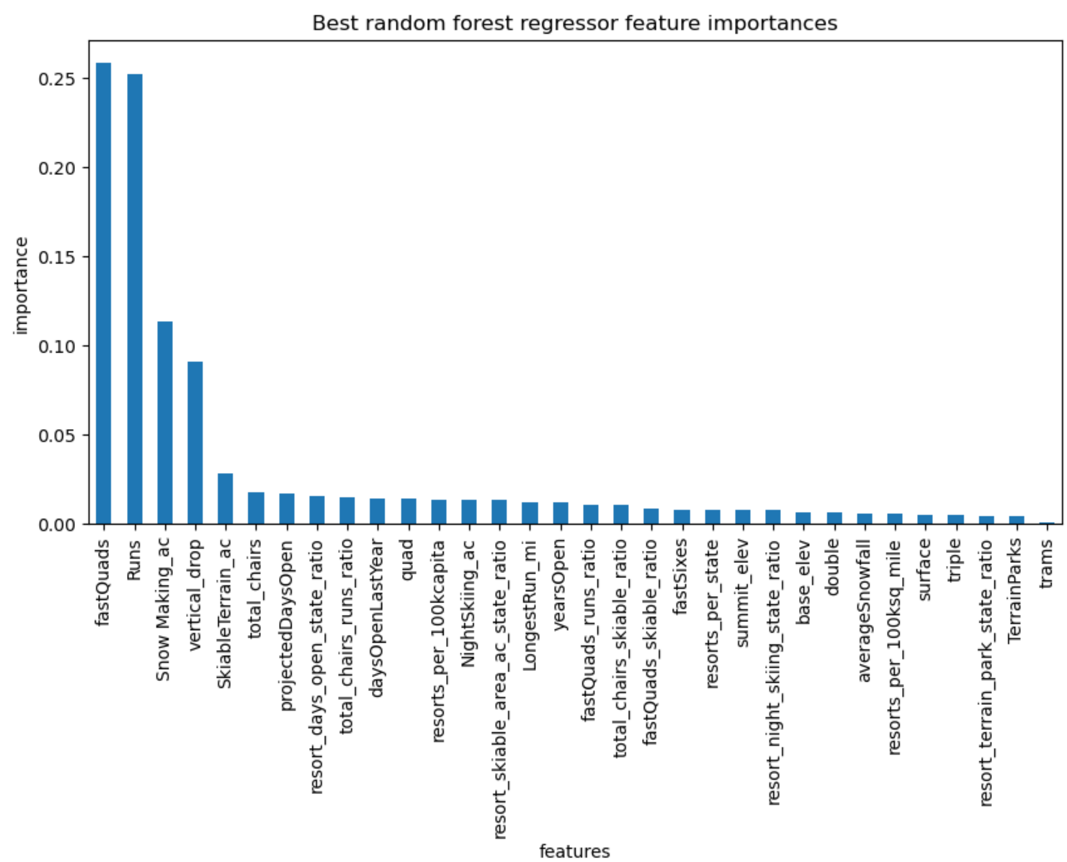
Based on this, created a Linear Model. This is performed by selected subset of features in the model. First performed with 10 features. Then performed with 15 features. These results highlight that assessing model performance in inherently open to variability. Thus used GridSearhCV to find the best k parameter which turns to be 8.

Out of top 8 features below, it shows that vertical drop is the biggest positive feature.

- vertical_drop - 10.767857
- Snow Making_ac - 6.290074
- total_chairs - 5.794156
- fastQuads - 5.745626
- Runs - 5.370555
- LongestRun_mi - 0.181814
- trams - -4.142024
- SkiableTerrain_ac - -5.249780

Cross validating gives a value of 11.793465668669324 for Linear Model.

Then performed the Random Forest Model. Below chart shows feature importance.



Based on Random Forest features, the dominant top four features are in common with linear model:

- fastQuads
- Runs
- Snow Making_ac
- vertical_drop

Cross Validating gives a value of 9.537730050637332 for 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.

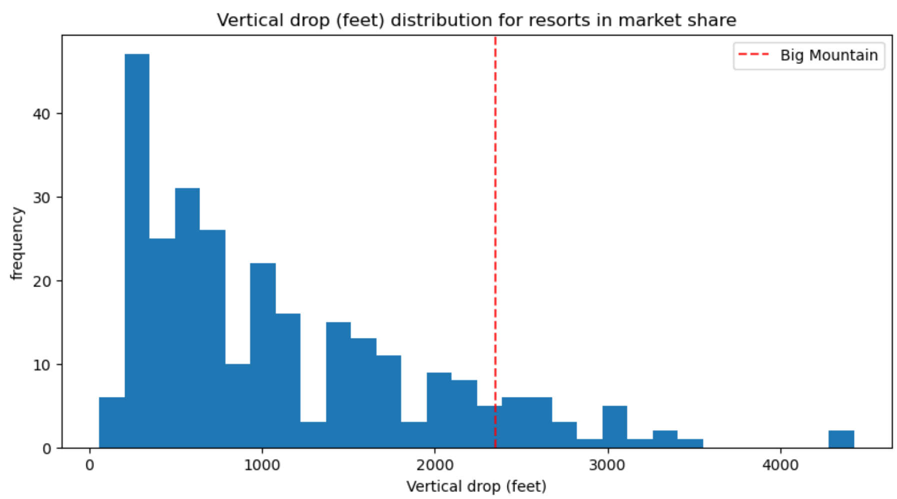
This decides to use going forward with Random Forest Model

Modelling

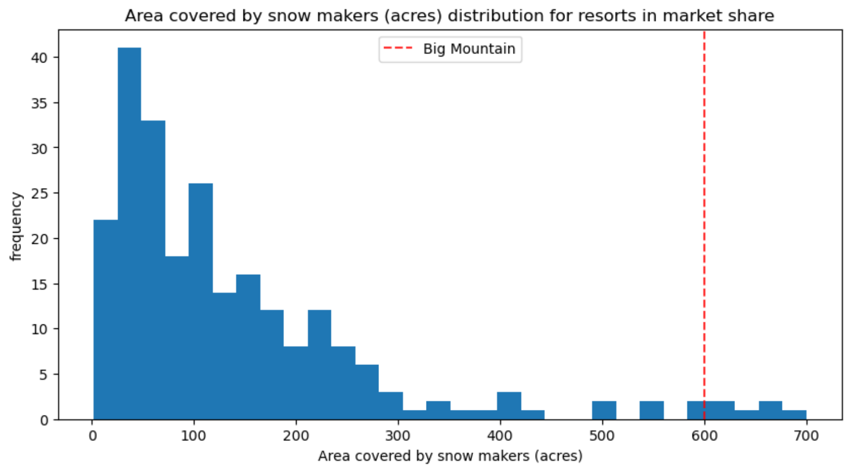
Current Adult Weekend ticket price for Big Mountain is $81.00. When we train the model to predict Big Mountain's ticket price based on data from all the other resorts, Big Mountain Resort modelled price is $95.87. With initial model, it shows a mean absolute error of $10.39 which suggests there is room for an increase.

If we compare Big Mountain resort with other resorts for the features that came up as important in the modeling

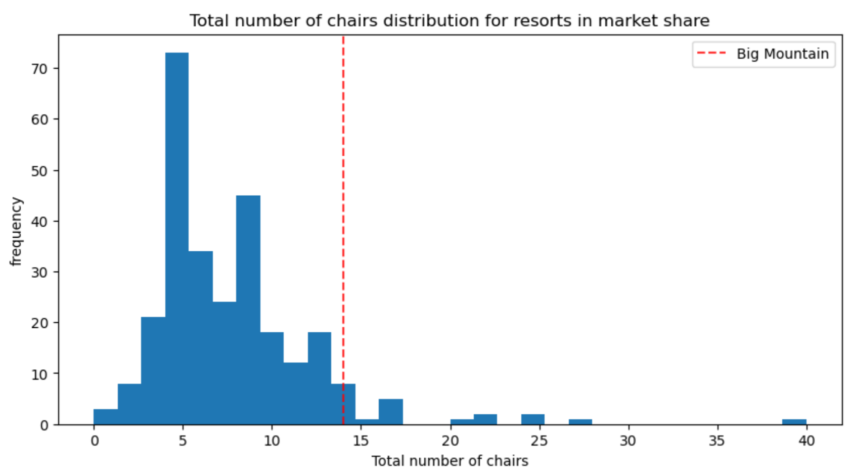
- vertical_drop - Big Mountain is doing well for vertical drop



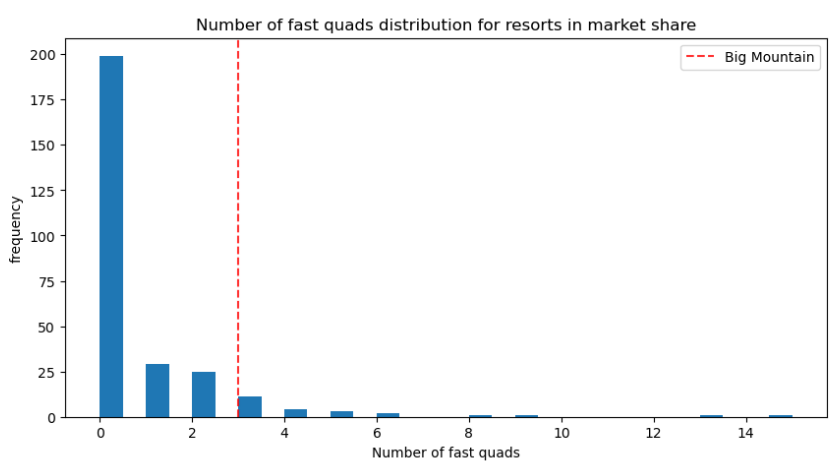
- Snow Making_ac - Big Mountain is very high up the league table of snow making area.



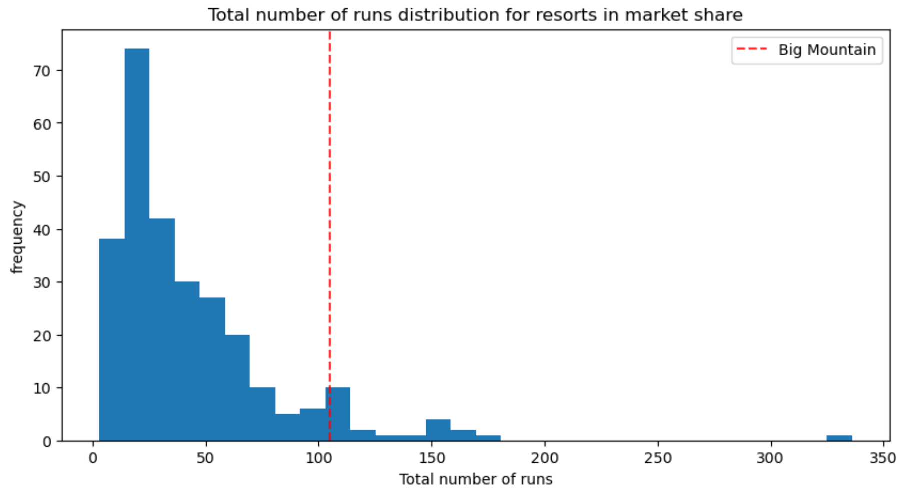
- total_chairs - Big Mountain has amongst the highest number of total chairs.



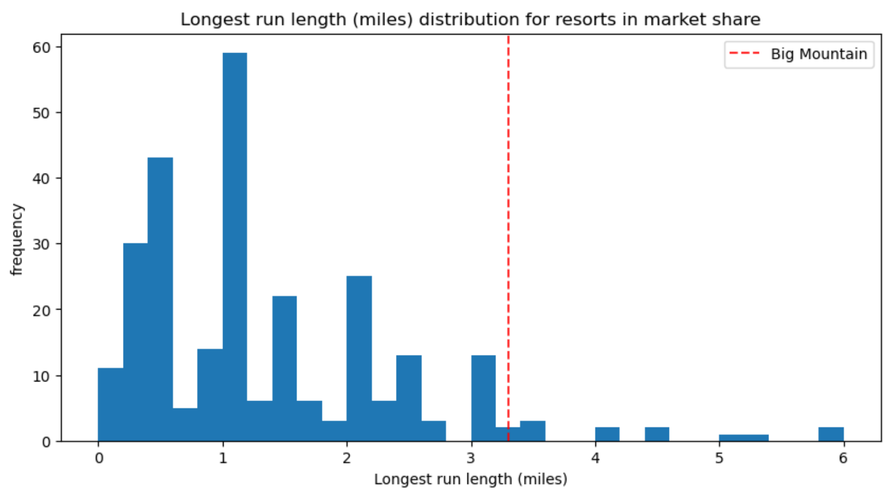
- fastQuads - Most resorts have no fast quads. Big Mountain has 3.



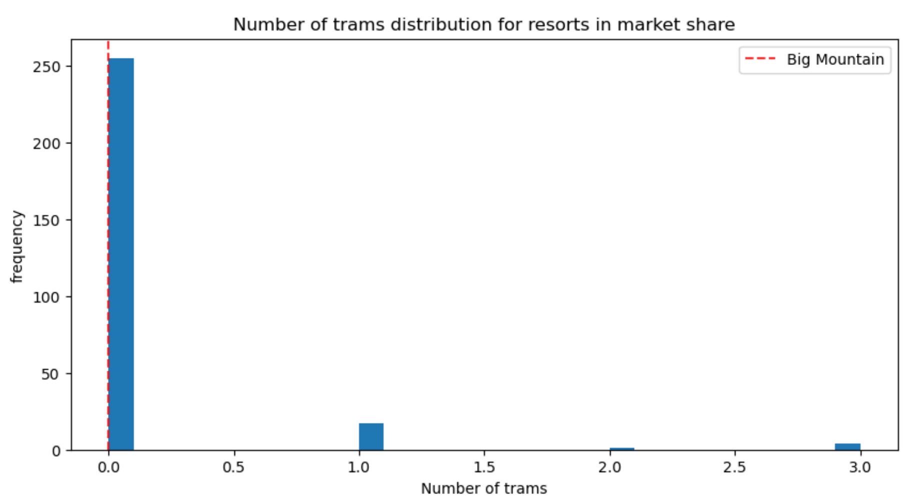
- Runs - Big Mountain compares well for the number of runs.



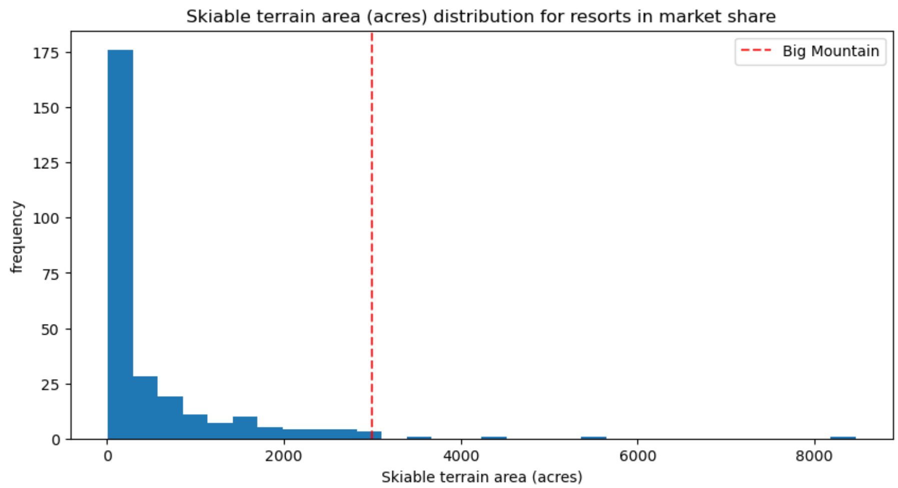
- LongestRun_mi - Big Mountain has one of the longest runs.



- trams - The vast majority of resorts, such as Big Mountain, have no trams

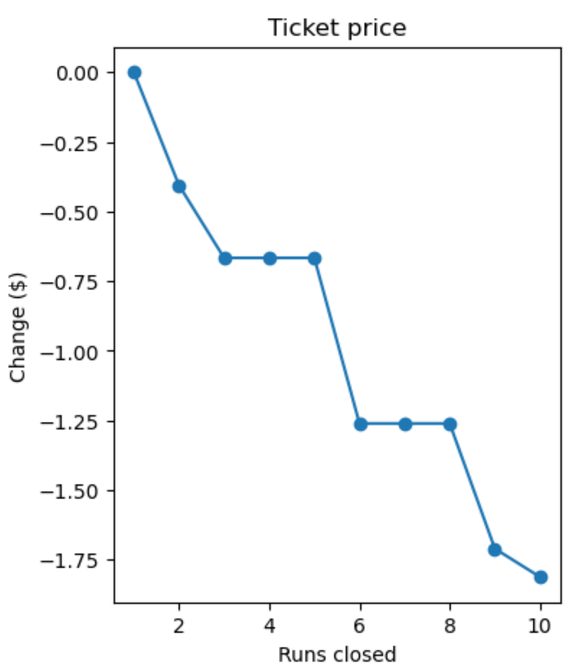


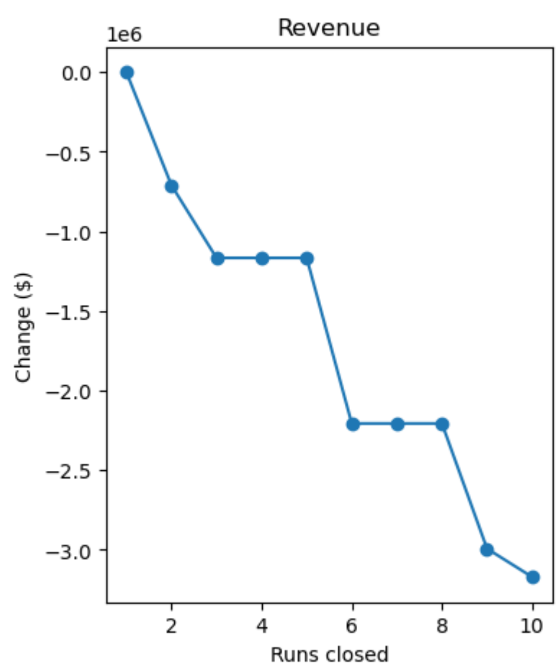
- SkiableTerrain_ac - Big Mountain is amongst the resorts with the largest amount of skiable terrain.



Then model some scenarios to see how it affects the price based on adding or reducing some of the features

Scenario 1 - Close up to 10 of the least used runs. 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.





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

Scenario 3 - Same as Scenario 2 but adding 2 acres of snow making. Result: Such a small increase in the snow making area makes no difference!

This helps business recommend more scenarios with different features to see how it impacts the ticket price.

Conclusion

According to this analysis, scenario 2 shows a clear indication that support for ticket price increase by $1.99. Over the season, this could be expected to amount to $3,474,638. Current model that build is useful to predict ticket prices based on the different features. But currently we don't have much info on how much each feature is used by customers. If we can get these info from business, then we can come up with more scenarios to add or remove features and validate impact on ticket prices. Even there is chance, we may be able to cut down some of the features and keeping the ticket price same.