Big Mountain Resort Pricing Project
Erdal Terkivatan

Problem Identification Overview

Context

Big Mountain Resort, located in northwestern Montana, offers spectacular views of Glacier National Park and Flathead National Forest. Big Mountain Resort was opened in 1947 with an annual snowfall of 333 inches and 3,000 acres of skier and rider accessible terrain. Annually 350,000 skiers and snowboarders ride Big Mountain Resort's 105 named trails, vast bowl, and tree skiing.

Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This extra chair increases their operating costs by $1,540,000.00 this season. With the new lift, Big Mountain management needs to know how much they can raise the adult weekday and weekend lift tickets to maintain their profit margin of 9.2% this season.

Problem

How much will Big Mountain Resort need to increase their weekday and weekend adult lift tickets to maintain a minimum profit margin of 9.2% for its investors during the upcoming season, while covering additional operating costs of $1,540,000.00 for the new chair lifts?

To generate a solution to the problem, we looked at the current adult weekday and weekend lift ticket prices and other relevant resort data from across the country. The goal of this project was to develop models that predict the expected adult weekday and weekend prices to increase revenue in the upcoming season to account for the increased operating cost of $1,540,000.00 for a new chair lift.

Data

The data includes 27 features for 330 ski resorts all around the U.S. including Big Mountain Resort. The variable of interest is the ticket price. There are two types of ticket prices: 1) Adult weekday('AdultWeekday') and 2) Adult weekend ('AdultWeekend')

Data Wrangling

We calculated the NaN rate for each feature. fastEight which is the number of 8-person lifts is the feature that has the most missing values with a rate of 50%. As a result, we dropped it as half of its values are missing and the remaining values are mostly zero.

The target variables 'AdultWeekday' and 'AdultWeekend' have the missing value rate of %16.36 and %15.45 respectively. 'AdultWeekday' is missing in a few more records less than 'AdultWeekend'. We looked at the price information for Big Mountain Resort specifically to decide what price type to keep. We considered only the state of Montana as it is our target state. The average price for weekends and weekdays are the same. Since weekend price has the least missing values we decided to keep AdultWeekend price as our target variable.

We performed feature engineering and created the new features such as the total number of terrain parks, the total skiable area, the total number of days open, and the total area available for night skiing as it makes more sense to work with the totals for these features. We also dropped the rows with no price information.

Exploratory Data Analysis

The top States By Order Of Each Of The Summary Statistics

What does the state-wide picture for the market look like?

1. Total State area (sq miles)

Alaska	665384
California	163695
Montana	147040
New Mexico	121590
Arizona	113990

Montana comes in as the third-largest state.

2. Total State Population

California	39512223
New York	19453561
Pennsylvania	12801989
Illinois	12671821
Ohio	11689100

California dominates the state population figures despite coming in second behind Alaska in size (by a long way). The resort's state of Montana was in the top five for size but doesn't figure in the most populous states. Thus, Montana is less densely populated.

3. Resorts per State

New York	33
Michigan	28
Colorado	22
California	21
Pennsylvania	19

New York comes top in the number of resorts in our market. We do not know whether this is because of its proximity to wealthy New Yorkers wanting a convenient skiing trip or is it simply that its northerly location means there are plenty of good locations for resorts in that state?

4. Total skiable area

Colorado	43682.0
Utah	30508.0
California	25948.0
Montana	21410.0
Idaho	16396.0

New York state may have the most resorts, but they don't account for the most skiing area. In fact, New York doesn't even make it into the top five of skiable area. Montana makes it into the top five, though. One may start to think that New York has more, smaller resorts, whereas Montana has fewer, larger resorts. Colorado seems to have a name for skiing; it's in the top five for resorts and in the top place for the total skiable area.

5. Total Night Skiing Area

New York	2836.0
Washington	1997.0
Michigan	1946.0
Pennsylvania	1528.0
Oregon	1127.0

New York dominates the area of skiing available at night. Looking at the top five in general, they are all the more northerly states.

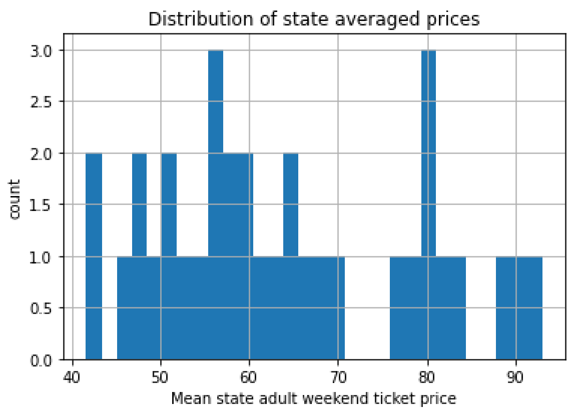
6. Total days open

Colorado	3258.0
California	2738.0
Michigan	2389.0

New York	2384.0
New Hampshire	1847.0

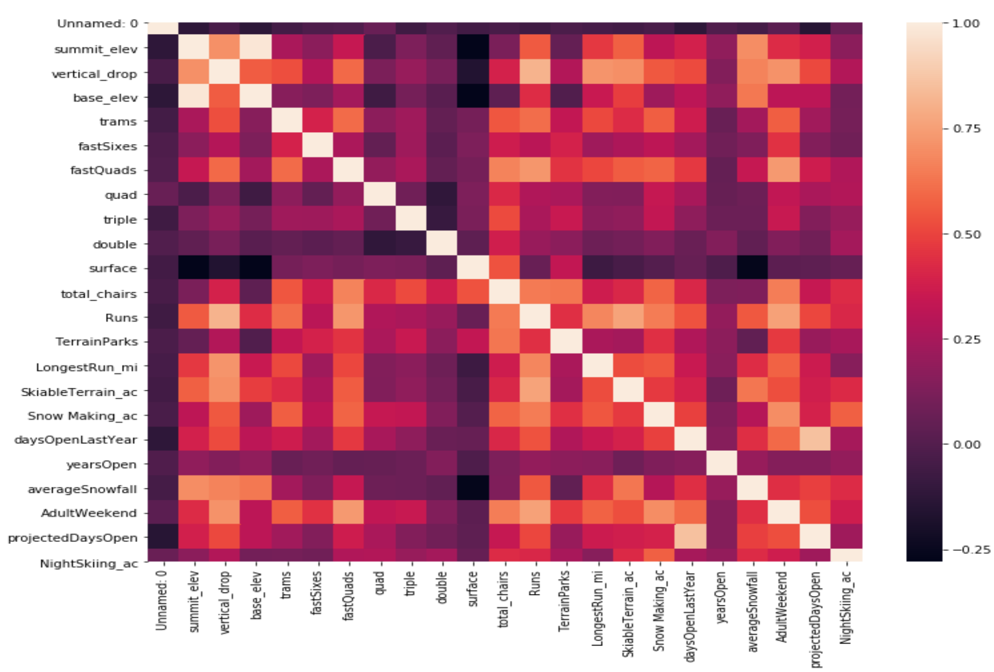
The total days open seem to bear some resemblance to the number of resorts. This is plausible. The season will only be so long, and so the more resorts open through the skiing season, the more total days open we'll see. New Hampshire makes a good effort at making it into the top five, for a small state that didn't make it into the top five of resorts per state

7. Average ticket price by state



The average ticket price is mostly between $50 and $70

Correlation Analysis

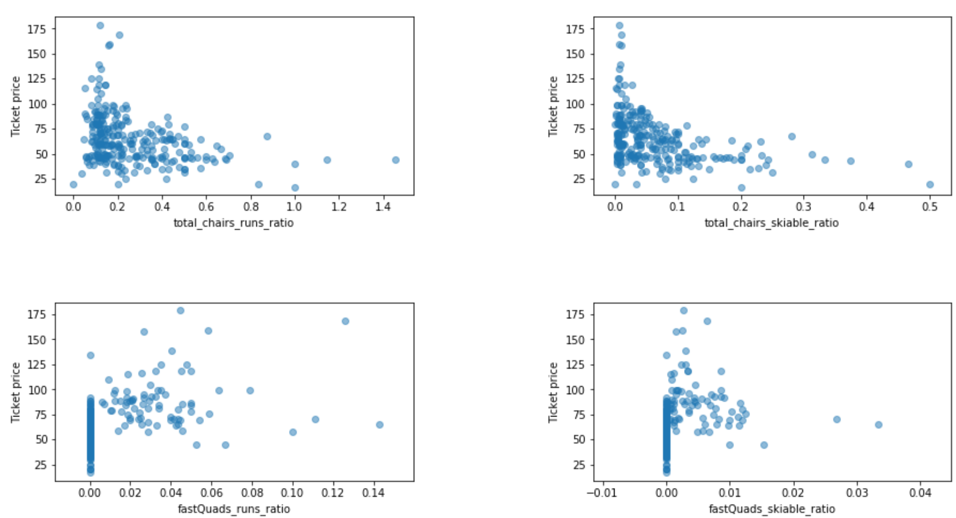


First, summit and base elevation are quite highly correlated. This isn't a surprise. You can also see that you've introduced a lot of multicollinearity with your new ratio features; they are negatively correlated with the number of resorts in each state. This latter observation makes sense! If you increase the number of resorts in a state, the share of all the other state features will drop for each. An interesting observation in this region of the heatmap is that there is some positive correlation between the ratio of night skiing area with the number of resorts per capita. In other words, it seems that when resorts are more densely located with a population, more night skiing is provided.

Turning our attention to the target feature, AdultWeekend ticket price, we see quite a few reasonable correlations. fastQuads stands out, along with ‘Runs’ and Snow Making_ac. Visitors would seem to value more guaranteed snow, which would cost in terms of snowmaking equipment, which would drive prices and costs up. Of the new features, resort_night_skiing_state_ratio seems the most correlated with the ticket price. If this is true, 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 the ticket price. This is plausible; the more runs you have, the more chairs you'd need to ferry people to them! Interestingly, they may count for more than the total skiable terrain area. For sure, the total skiable terrain area is not as useful as the area with snowmaking. People seem to put more value in guaranteed snow cover rather than more variable terrain area.

Scatter Plots



It seems that the more chairs a resort has to move people around, relative to the number of runs, the ticket price rapidly plummets and stays low. What we may be seeing here is an exclusive vs. mass-market resort effect; if you don't have so many chairs, you can charge more for your tickets, although with fewer chairs you're inevitably going to be able to serve fewer visitors. Your price per visitor is high but your number of visitors may be low. Something very useful that's missing from the data is the number of visitors per year.

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 the ticket price.

Preprocessing and Training Data

The following steps are performed for preprocessing and training the data

1. Imputation: We imputed the missing values using both median and mean
2. Used linear regression model to train the data
3. To evaluate the model performance , R2, mean absolute error and mean square error are used

Model Performance

	Imputed by Median		Imputed by Mean	
	Test	Train	Test	Train
R2	69.64%	80.61%	69.64%	80.61%
MAE	8.6291	9.6797	9.703	8.62
MSE	119.071	175.96	119.054	177.359

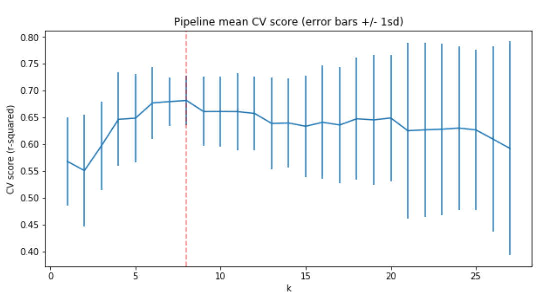
These results don't seem very different to when you used the median for imputing missing values.

Assessing Model Performance Using Cross Validation

Cross validation is performed with 5 folds to assess the model. The average CV score of %63.47 is obtained.

Hyperparameter Tuning with GridSearchCV

We want to find the optimal number of features to be included in the model.



The above suggests a good value for k is 8. There was an initial rapid increase with k, followed by a slow decline. Also noticeable is the variance of the results greatly increase above k=8

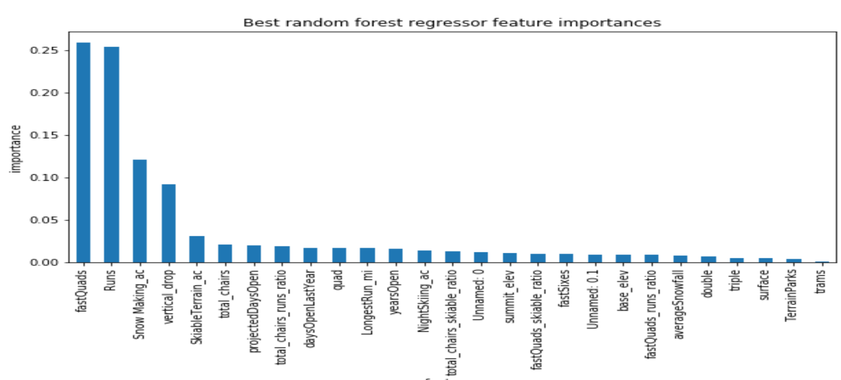
Now, we want to know which features are most useful?

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

These results suggest that vertical drop is your biggest positive feature. This makes intuitive sense and is consistent with high correlation with the ticket price that we observed in the EDA part. Also, you see the area covered by snow making equipment is a strong positive as well. People like guaranteed skiing! The skiable terrain area is negatively associated with ticket price! This seems odd. People will pay less for larger resorts? There could be all manner of reasons for this. It could be an effect whereby larger resorts can host more visitors at any one time and so can charge less per ticket. As has been mentioned previously, the data are missing information about visitor numbers. The coefficient for skiable terrain is negative *for this model*. For example, if you kept the total number of chairs and fastQuads constant, but increased the skiable terrain extent, you might imagine the resort is worse off because the chairlift capacity is stretched thinner.

Random Forest Model

The following graph shows the most important features that are suggested by the random forest regressor model.



Encouragingly, the dominant top four features are in common with our linear model:

- fastQuads
- Runs
- Snow Making_ac
- vertical_drop

Final Model Selection

Based on the table below , we concluded that random forest regressor performs better than linear regression.

	Linear Regression	Random Forest
Mean Absolute Error	11.794	9.537

Calculate Expected Big Mountain Ticket Price From the Model

Big Mountain Resort modelled price is $96.65, actual price is $81.00. Even with the expected mean absolute error of $10.23, this suggests there is room for an increase.

This result should be looked at optimistically and doubtfully! The validity of our model lies in the assumption that other resorts accurately set their prices according to what the market (the ticket-buying public) supports. The fact that our resort seems to be charging that much less that what's predicted suggests our resort might be undercharging. But if ours is mispricing itself, are others? It's reasonable to expect that some resorts will be "overpriced" and some "underpriced." Or if resorts are pretty good at pricing strategies, it could be that our model is simply lacking some key data? Certainly we know nothing about operating costs, for example, and they would surely help.

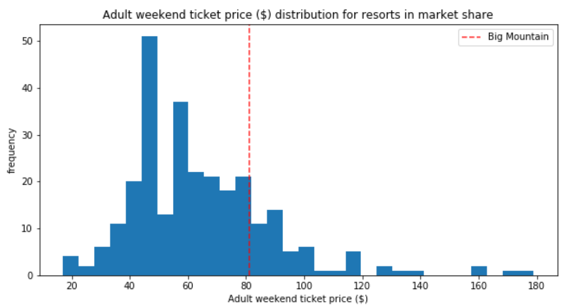
Big Mountain Resort in Market Context

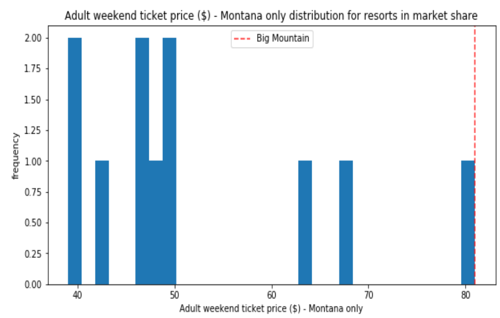
Features that came up as important in the modeling (not just our final, random forest model) included:

- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

The following graphs gives us an idea where Big Mountain Resort sits in the market in terms of the features above.

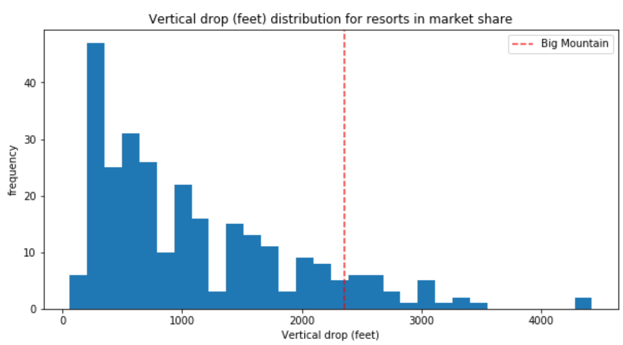
Ticket Price





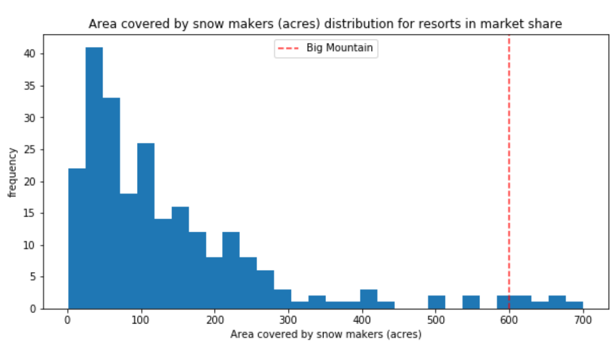
Big Mountain Resort ticket price is above most of the resorts both nationwide and in Montana.

Vertical Drop



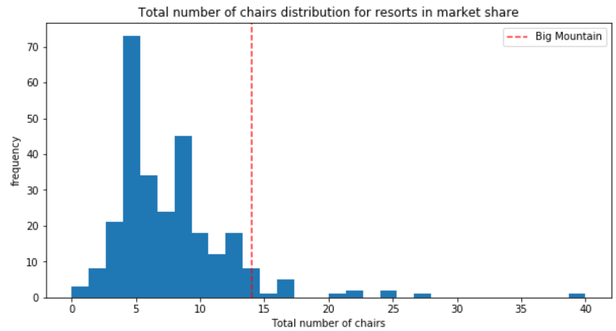
Big Mountain is doing well for vertical drop, but there are still quite a few resorts with a greater drop.

Snow Making Area



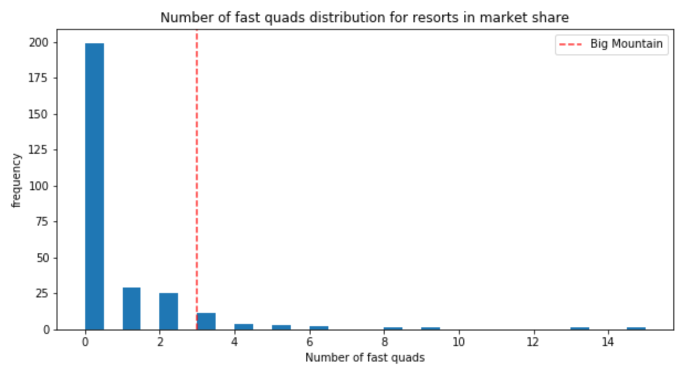
Big Mountain is very high up the league table of snowmaking area

Total Number of Chairs



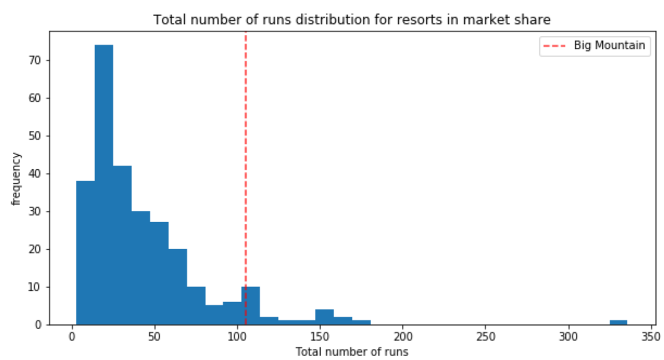
Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers.

Fast Quads



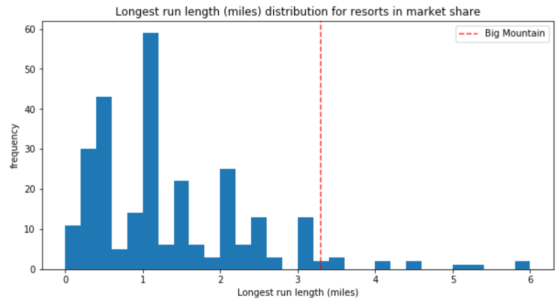
Most resorts have no fast quads. Big Mountain has 3, which puts it high up that league table. There are some values much higher, but they are rare.

Runs



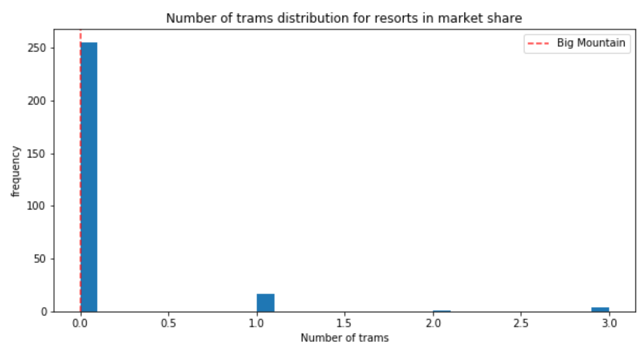
Big Mountain compares well for the number of runs. There are some resorts with more, but not many.

Longest Run



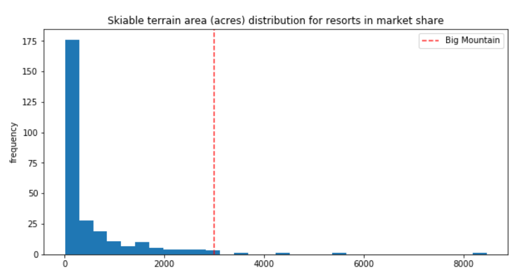
Big Mountain has one of the longest runs. Although it is just over half the length of the longest, the longer ones are rare

Trams



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

Skiable Terrain Area



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

Model Scenarios

Big Mountain Resort has been reviewing potential scenarios for either cutting costs or increasing revenue (from ticket prices). Ticket price is not determined by any set of parameters; the resort is free to set whatever price it likes. However, the resort operates within a market where people pay more for certain facilities, and less for others. Being able to sense how facilities support a given ticket price is valuable business intelligence. This is where the utility of our model comes in.

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

The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days. Assume the provided data includes the additional lift that Big Mountain recently installed.

Scenario 1: Close up to 10 of the least used runs. The number of runs is the only parameter varying.

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.

Scenario 2: In this scenario, 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 $1.61
Over the season, this could be expected to amount to $2815217

Scenario 3: In this scenario, you are repeating the previous one but adding 2 acres of snow making.
This scenario increases support for ticket price by $1.61
Over the season, this could be expected to amount to $2815217

Such a small increase in the snow making area makes no difference!

Scenario 4: This scenario calls for increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability.

0.0
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.

Next Steps

1. Do some more research into gathering more accurate data to reduce the number of initial NaN values.

2. Look into obtaining additional data such as other revenue streams, the number of customers per weekday/weekend, and other expenses(i.e.COGS, advertising, lawsuits) for each resort to improve the models.

3. Try to obtain the data over the last 30 years for each resort to see how the price increase as the resorts expand.

4. With COVID-19, how can this affect the analysis?
a. Many people are working from home and learning from home which means traveling is easier to schedule. This could provide more skiers.

b. Many people are not financially stable right now and are they going to have the money to go on a ski trip? Will they try and go somewhere where it’s cheaper? Will they try and go on weekdays instead of weekends? So there could also be fewer skiers or fewer skiers on weekends.