Big Mountain Resort Price Analysis Report

Sepehr Hajibikloo

November 2023

Data Sourced from: Springboard Data Science Career Track

Problem statement:

The Big Mountain resort is a ski resort located in Montana, 350,000 people visit every year to ski or snowboard the big mountain. They recent addition of chair lift increasing their operation cost to $1.54M per season. Need to find a way to reduce the operation cost, and project the revenue of the season based on the reduced operation cost.

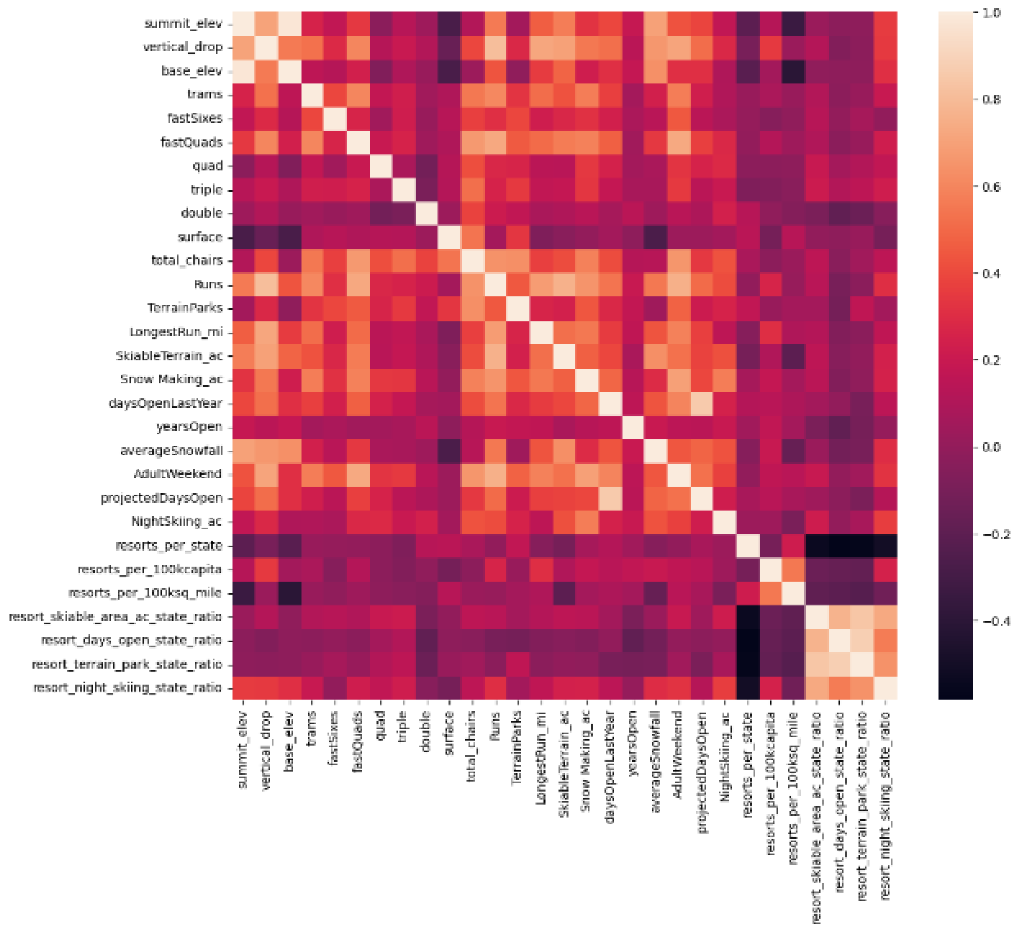
Data Wrangling:

The first values inspected were the Adult Weekend vs Adult Weekday prices, was it actually advantageous to have a different price for the weekend? Most states had the same price for both.

Adult Weekend had several missing values. Because of this, the Adult Weekend column was removed. In addition to Adult Weekend, the fastEight column was dropped because value of it's values were null and the other half were mostly 0. Besides those two big ones, there were some smaller columns that had to be dropped and a lot of missing values that had to be taken care of. Once this was finished, we were left with 277 of the original 330 rows.

Exploratory Data Analysis:

In order to find trends, patterns, and actionable insights in the data, we need to explore and find these patterns. The first pattern explored was the relationship between the total number of resorts by population vs total number of resorts by area. This didn't yield much usable information for Big Mountain Resort, but it did clear up some initial thoughts. The next relationship I wanted to look into was the relationship between the components like vertical drop, years open, or skiable areas versus the price in each state. This required a Principal Cumulative Analysis (PCA), which showed that the first two components account for 75% of the variance, and the first four account for 95%. Focusing 3 on just the two components, I scaled the data and added the average ticket price to a scatter plot. I needed to see a clearer relationship between price and components, so instead I created a heatmap to better visualize the relationship between each feature.



Focusing on the relationships in the “Adult Weekend” row shows that there is clear positive correlation associated strongly with fastQuads, Runs, SnowMaking_Ac, and resort_night_skiing_state_ratio. Now we can use these features to build a model that can determine a new data-based ticket price.

Pre-Processing and Training Data:

After identifying the four categories with the strongest correlation to price, the first step was to take an initial average as a “best guess” for pricing. This resulted in $83.81 initially which was the baseline I used to compare other prices to moving forward. Mean wasn’t the best for this because our Absolute Error was off by around $19, much too large for a price like this. Instead, I performed a regression using the median between results. The Mean Absolute Error was only off by $9 this time, but I knew this could still be improved. I also decided to create a data pipeline in order to efficiently produce identical results to make comparisons easier. The next regression was based on a Random Forest Model which helped identify that imputing the median value helps with the MAE of our four components. In addition to our four components, during the analysis I found that vertical drop also plays a big role in determining ticket price. Once this new component was added to the Random Forest Model, the Mean Absolute error was down to around $1 which was an acceptable amount of variability.

Modeling:

At this point I have picked the top components and a method of regression. I can use these together to create a model that actually gives us a data-based ticket price. To make the model as accurate as possible I boosted the number of components to eight by adding total_chairs, LongestRun_mi, trams, and vertical_drop to the list. To determine a fair price, I needed to see where Big Mountain Resort (represented by the dashed red line) ranked in these categories. - Vertical Drop - Total Chairs 6 - Fast Quads - Longest Run - Number of Trams - Number of Runs 7 - Area Covered by Snow Makers - Skiable Area.

As you can see, Big Mountain Resort either ranks high or is well above the average in each category. With the exception of trams which most resorts don't have anyway. All this shows us that Big Mountain Resort is an exceptional resort with many great facilities, and the price should reflect that. With that we get a modelled price of $96.62, which is well above their current price of $81.00.

Algorithms used to build the model with evaluation metric:

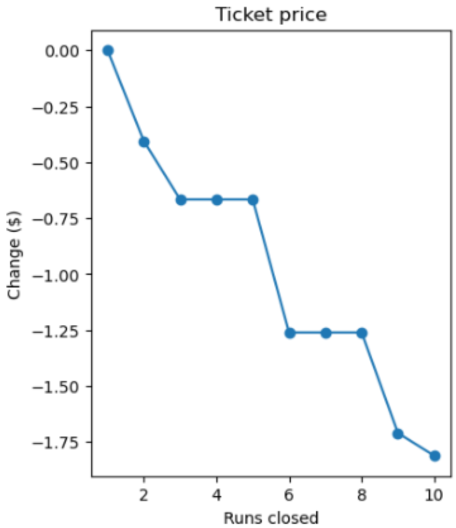
I decided to try two different types of models, a linear regression imputed with the median and mean values, and a random forest model by adding additional classifications.

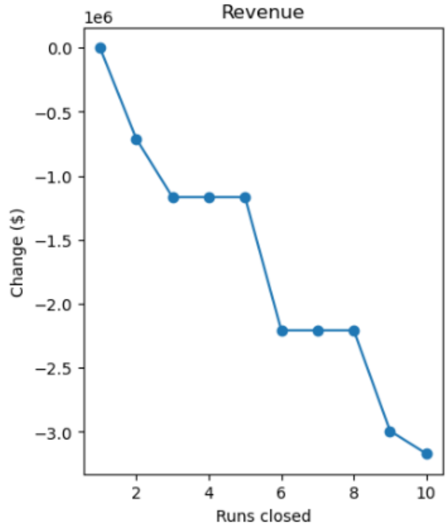
- **Linear Regression:** As stated above, this model worked by imputing missing values with median and mean values based on the data we had. With the four features from the heatmap used, this model had a Mean Absolute Error of $9 which was much too high for this scale.

- **Random Forest Model:** Adding all eight features gave us a model that was much more accurate, with a Mean Absolute Error around $1. This was the model that was used to determine the final ticket price for the resort. After cross validating the two models, this one produced more consistent results and a lower MAE.

Conclusion:

Big Mountain Resort is currently charging very little for all the services they are offering. According to this analysis, they excel at seven of the eight most important, price-determining features and therefore should raise their ticket price by at least $10.00. In addition to this, Big Mountain Resort can also save a lot of money by not having all of their runs open at once. While they do get 350,000 visitors per year, day-to-day not each run is being fully utilized and this can be very expensive. According to a predictive analysis based on the Random Forest Model (shown below), Big Mountain Resort can keep up to 5 runs closed without a huge drop in revenue.





With all of this in mind, I think we now have two very simple ways to keep Big Mountain Resort running while still providing the numerous amenities it has to offer to its customers. This data and model are good for now, however in the future I can see improvements being made particularly in the category of prices that other resorts are charging. The data fails to account for resorts that might be undercharging their tickets because they have expensive rental or hotel fees, we only have the ticket price. Once this data is acquired however, the model can only get better from here. I see a lot of opportunity for growth at Big Mountain Resort and I hope that the results of this analysis can be implemented so that the resort can stay open for many years to come.

Future scope of work:

I think some other expenses exist that we can add to our data. Its modeled price was so much higher than its current price, because experts couldn't predict process correctly. This mismatch would surely come as a surprise to the business executives, because they can earn a lot more. I guess business leader will use it easily. They can quickly change their ideas to make a perfect plan. I believe they absolutely come to me test new combination of parameters in scenario. The price predicting model suggests the potential for raising the price above $83 (potentially much higher), but this would mean making additional enhancements to the facility to support the price increase. Since Big Mountain has a geographical advantage to compete in the premium market, upgrading the facility to charge a higher price while attracting more visitors would increase the resort's completive advantage in a pro market cycle. Other consideration, invite a cross-functional team of business experts to try the model and test the business assumptions.