Introduction:

Big Mountain Resort is a popular ski resort located in Whitefish, Montana. The resort offers fantastic views of the Flathead National Forest and Glacier National Park, in addition to several different types of ski runs. Its services include 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. The base elevation is 4,464 feet, the summit is 6,817 feet, and the longest run is 3.3 miles in length.

Big Mountain Resort has recently installed a new chairlift to improve visitor distribution across the mountain. This addition has increased their operating costs by $1.54 million this season. As a result, the business is rethinking its pricing strategy, which previously involved charging a slight premium above the average price of resorts in its market segment.

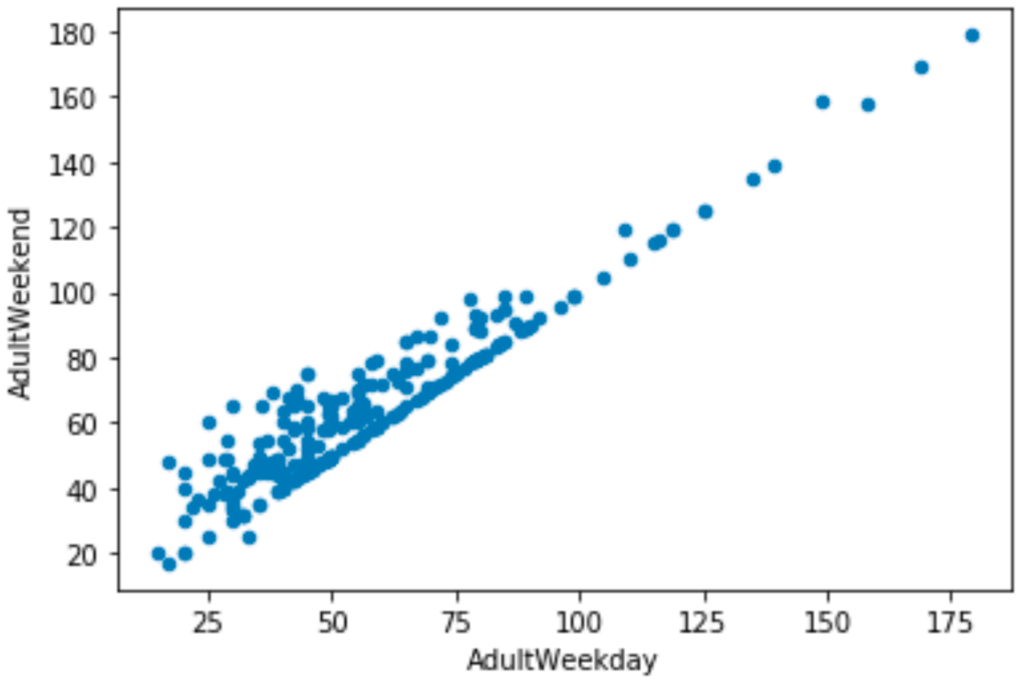
Problem:

Big Mountain Resort is in need of a new pricing strategy, one that is based on data gathered from numerous ski resorts across the country. How can we create a pricing model that determines a price that is competitive for customers while accurately reflecting the value of Big Mountain Resort's facilities?

Data Wrangling:

Our dataset includes several important values such as total vertical drop, number of lift chairs, weekday/weekend prices, and the total number of runs for each resort.

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, as shown in the chart below



Data Cleaning

Not only did the weekend and weekday values match for every entry in Montana, our resort's home state, but the Adult Weekend column also had several missing values. Because of this, the Adult Weekend column was removed.

In addition to Adult Weekend, the Fast Eight column was dropped because many of its values were null, and the rest were mostly 0.

Besides those two major columns, there were some smaller ones that had to be dropped as well, and a significant number of missing values that had to be addressed. Once this process was finished, we were left with 277 of the original 330 rows.

Exploratory Data Analysis

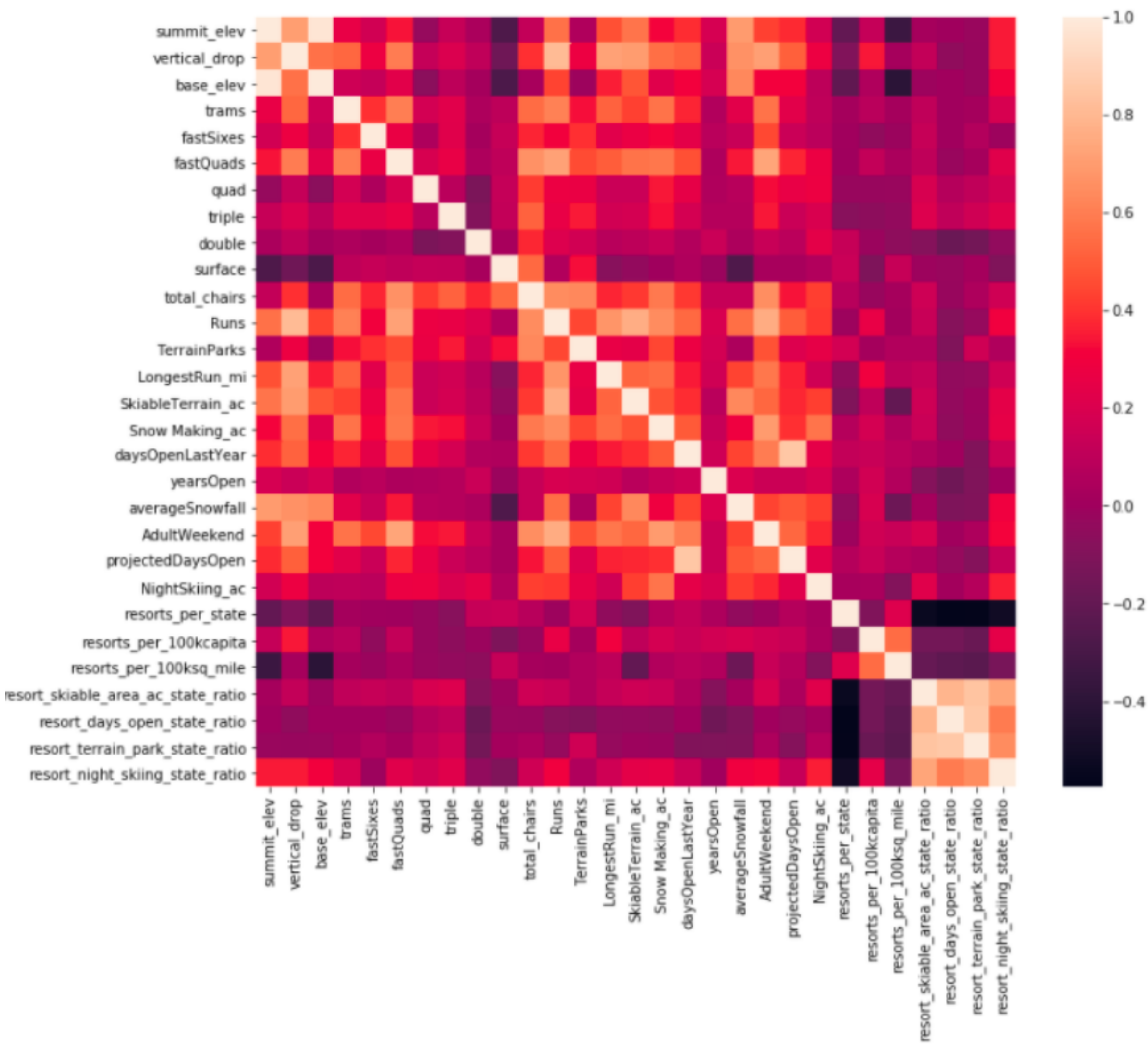
In order to find trends, patterns, and actionable insights in the data, we needed to explore and identify these relationships.

The first pattern explored was the relationship between the total number of resorts by population vs. the total number of resorts by area. This didn't yield much usable information for Big Mountain Resort, but it did help clear up some initial assumptions.

The next relationship examined was between components like vertical drop, years open, or skiable area versus the average price in each state. This required a Principal Component Analysis

(PCA), which showed that the first two components accounted for 75 percent of the variance, and the first four accounted for 95 percent.

Focusing on just the two primary components, I scaled the data and added the average ticket price to a scatter plot. However, I needed a clearer view of the relationship between price and each feature. To achieve this, I created a heatmap to better visualize these relationships.



Feature Relationships

Focusing on the relationships in the "AdultWeekend" row shows a clear positive correlation with the following features: fastQuads, Runs, SnowMaking_Ac, and resort_night_skiing_state_ratio.

These features can now be used to build a model that determines a new, data-based ticket price.

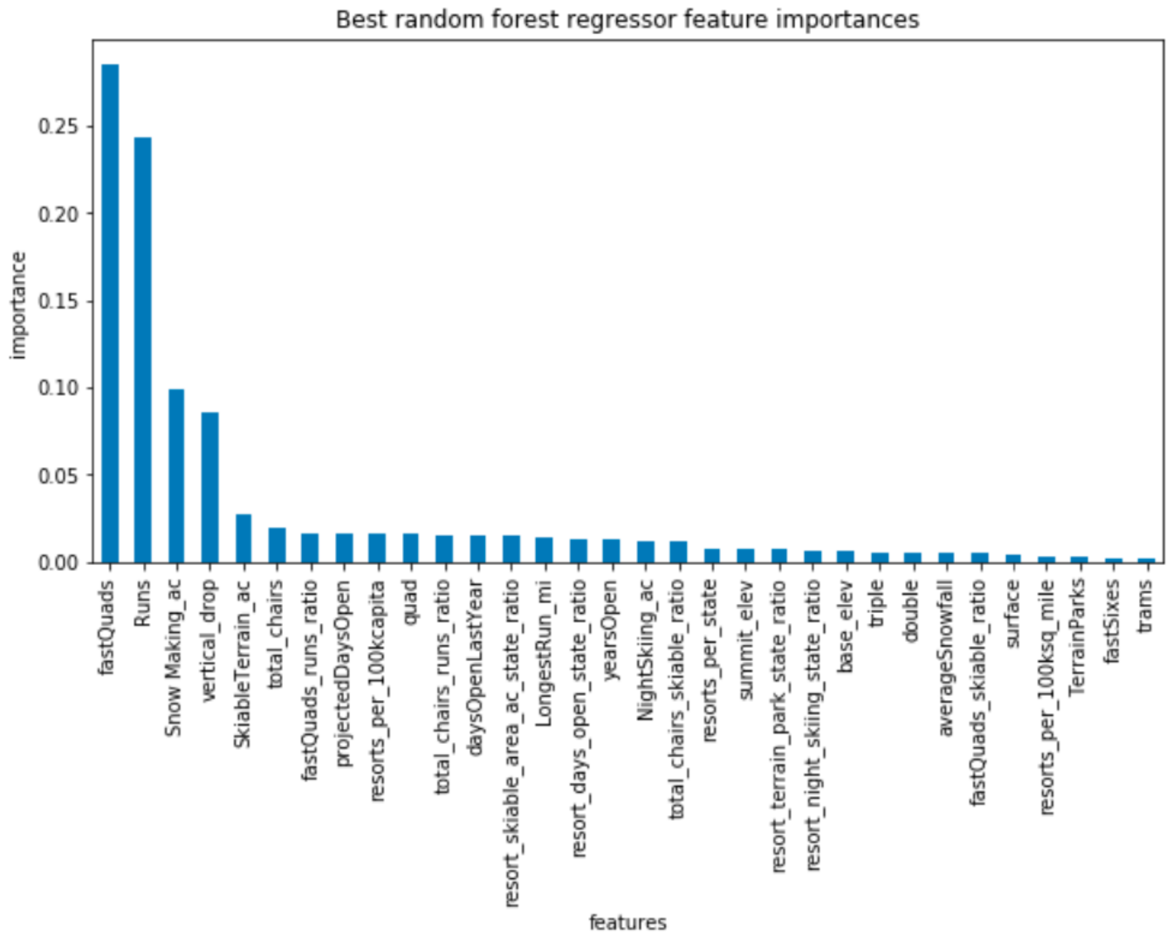
Pre-Processing and Training Data

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 an average price of $83.81, which served as the baseline for comparing future model outputs.

However, the mean wasn't ideal in this case, as the Mean Absolute Error (MAE) was off by about $19, which is too large for a practical pricing model. Instead, I performed a regression using the median between results. This brought the MAE down to about $9, but there was still room for improvement.

To streamline the process, I also created a data pipeline to efficiently reproduce results and make comparisons easier. The next regression was based on a Random Forest model, which helped confirm that imputing missing values using the median improved the MAE for our four core features.

During the analysis, I also discovered that vertical drop plays a significant role in determining ticket price. After adding vertical drop as a fifth feature to the Random Forest model, the Mean Absolute Error dropped to approximately $1, which is an acceptable level of variability for this use case.



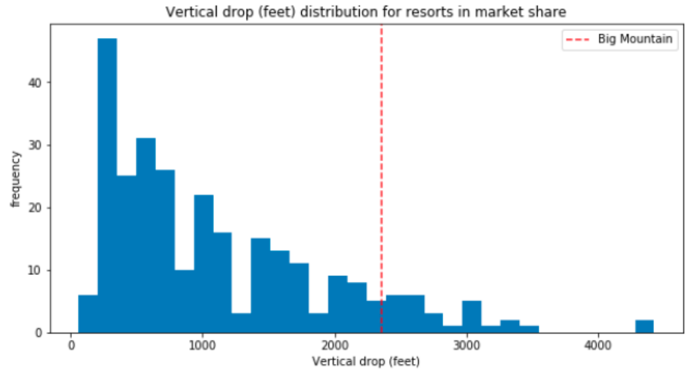
Modeling

At this point, I have selected the top components and a method of regression. I can use these together to create a model that provides a data-based ticket price.

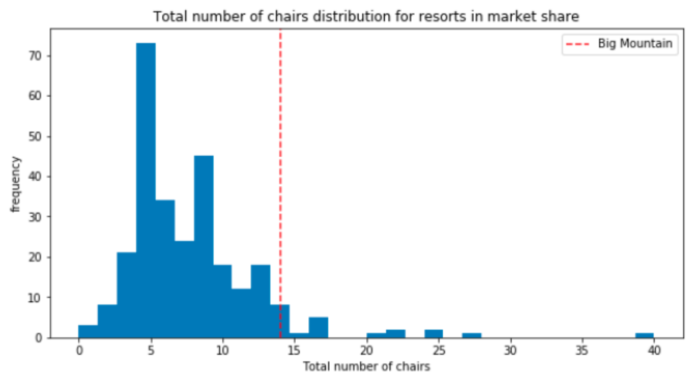
To make the model as accurate as possible, I increased the number of components to eight by adding total_chairs, LongestRun_mi, trams, and vertical_drop to the list.

To determine a fair price, I needed to see where Big Mountain Resort (represented by the dashed red line) ranked in these categories.

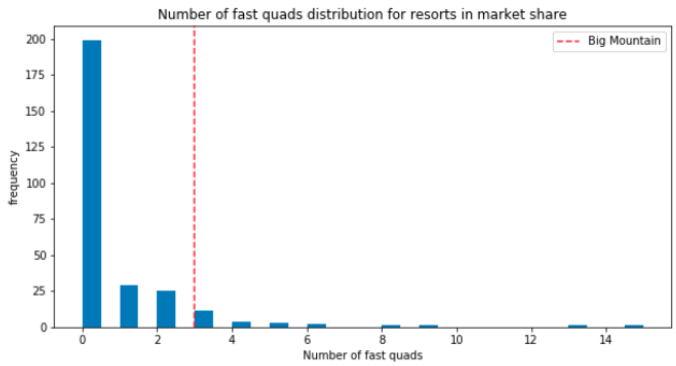
- **Vertical Drop**



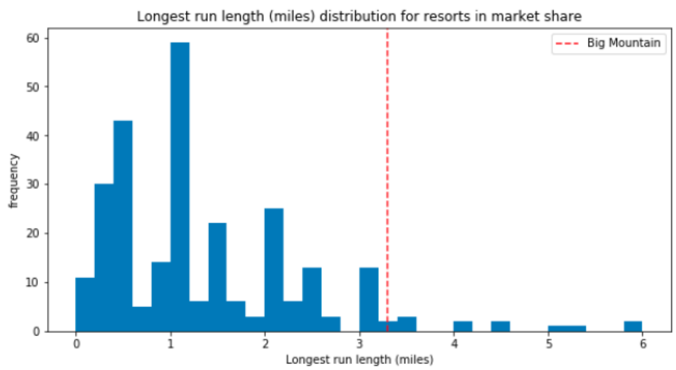
- **Total Chairs**



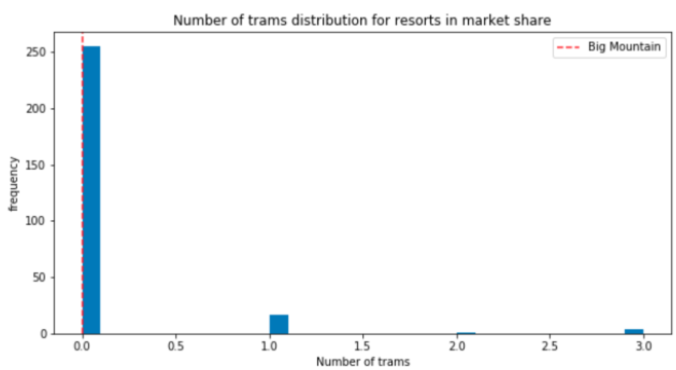
- **Fast Quads**



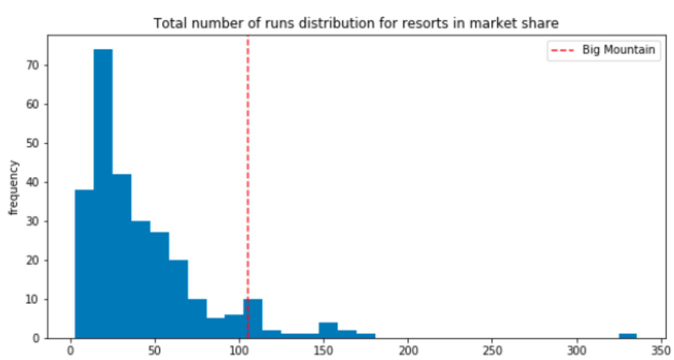
- **Longest Run**



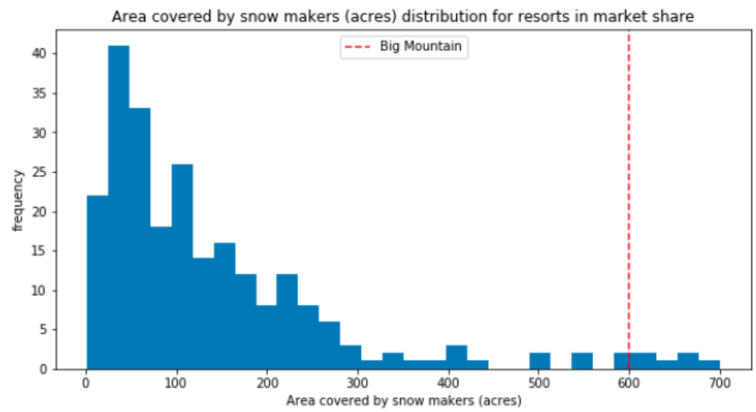
- **Number of Trams**



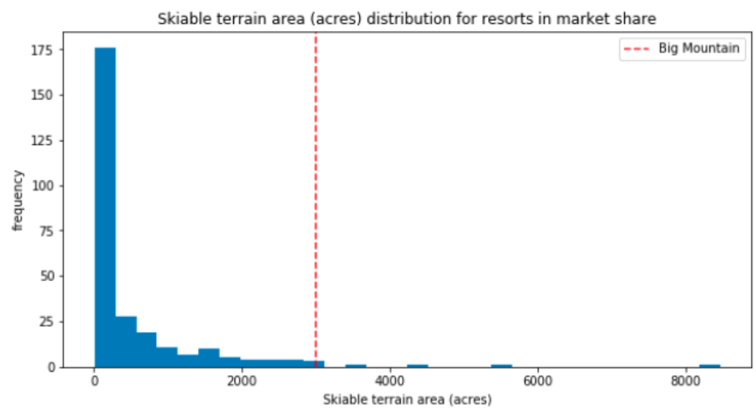
- **Number of Runs**



- **Area Covered by Snow Makers**



- **Skiable Area**



Results

As you can see, Big Mountain Resort either ranks high or well above average in each category, with the exception of trams, which most resorts don't have anyway.

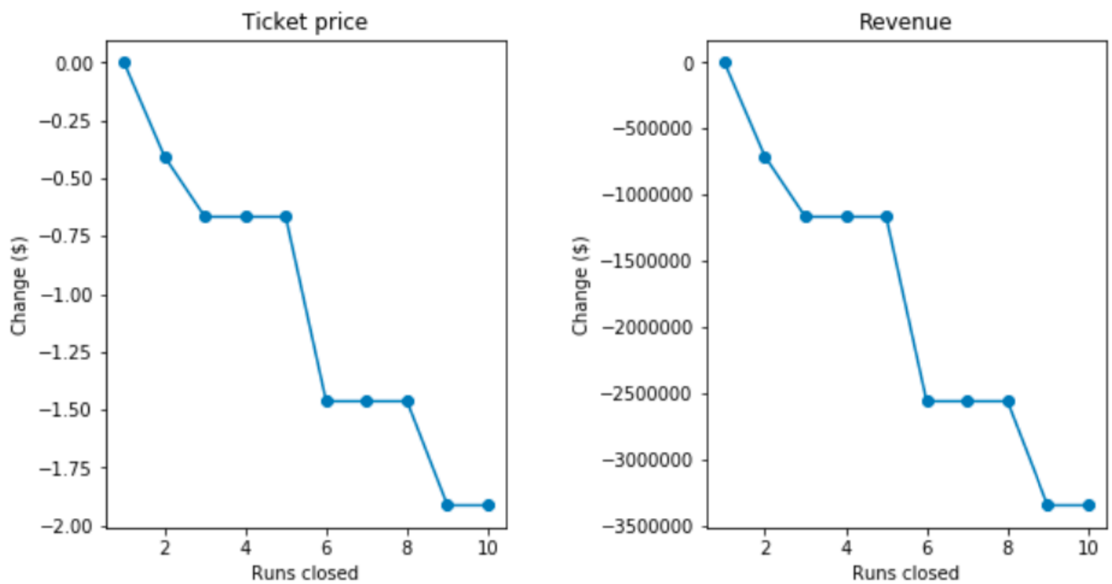
All of this shows that Big Mountain Resort is an exceptional resort with many high-quality facilities, and the price should reflect that. Based on the model, we get a predicted price of $96.62, which is well above their current price of $81.00.

Conclusion

Big Mountain Resort is currently undercharging for the services they offer. According to this analysis, they excel in seven of the eight most important price-determining features and should consider raising their ticket price by at least $10.00.

In addition, Big Mountain Resort can potentially save a significant amount of money by not having all of their runs open at once. While they receive approximately 350,000 visitors per year, not every run is fully utilized on a daily basis, and maintaining them can be 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 experiencing a major drop in revenue.



With all of this in mind, I believe we now have two very simple ways to keep Big Mountain Resort running while still providing the numerous amenities it offers to its customers. This data and model are good for now; however, in the future, I see opportunities for improvement, particularly in the area of prices that other resorts are charging.

The current data does not account for resorts that might be undercharging for tickets because they have expensive rental or hotel fees; we only have ticket prices. Once this data is acquired, the model can only improve from here.

I see a lot of potential for growth at Big Mountain Resort, and I hope the results of this analysis can be implemented so the resort can stay open for many years to come.