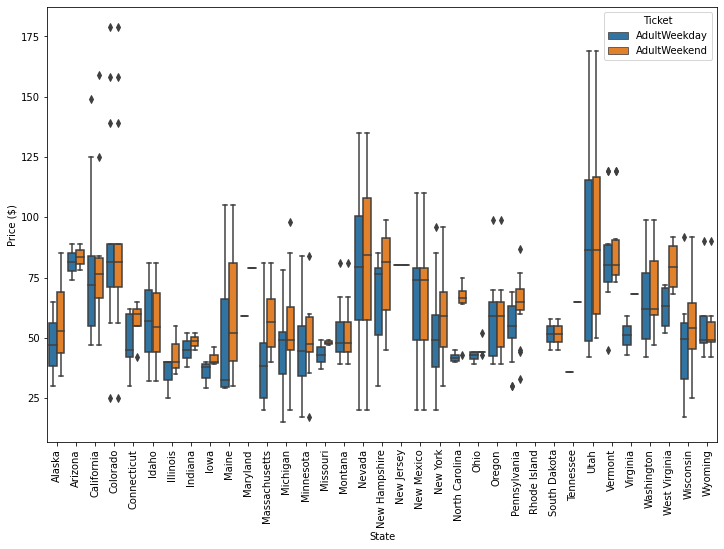
Guided Capstone Project Report

# Big Mountain Resort Recommendation

Big Mountain was looking for a better way to select its ticket prices by leveraging a data set from 330 ski resorts nationally in a similar market. The goal of the analysis was to gain a better understanding of what current features affect ticket prices and how Big Mountain’s features compare to see if a price change could be justified to increase revenue.

## Data Wrangling

All 330 ski resorts, including Big Mountain, had the same 27 columns and the data was presented in an excel file. The excel file contained null values, especially when looking at the fast eight features where half were missing. For that reason, the fast eight features column was dropped. A deeper dive into the values showed that one resort declared it was open for 2019 years. Without further information on this resort, it was dropped for consistency. Through fact checking, another resort had their skiable terrain corrected from 26819 acres to 1819 acres. State information from Wikipedia was merged as well.

The data set had ticket prices broken up into weekend and weekday. When the prices were looked at using box plot in Figure 1 and a scatter plot in Figure 2, two conclusions were drawn. One is Montana showed identical variance in the box plot with weekend and weekday prices yielding the same values. Two is there is a clear direct line in the weekend vs weekday prices that shows that in general, prices were roughly the same, with the exception being prices below $80. For these reasons, in addition to weekday prices have more missing values, weekday values were dropped, and weekend values remained as the target feature. At the end, the data consisted of 227 resorts and 25 columns.

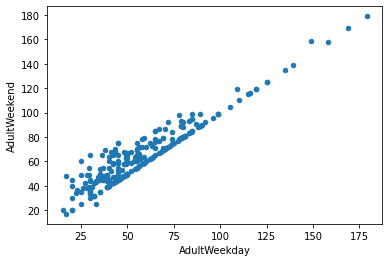


Figure 1. Scatter plot of Weekend Tickets vs Weekday Tickets

Figure 2. Box plot of the ticket prices per state

## Exploratory Data Analysis

Exploratory data revealed a few initial findings about the data set…

1. Montana is the third largest state by size
2. Montana is number 4 for skiable area
3. It is not in the top 5 for population, night skiing, or days open.
4. Resorts per 100k population and resorts per 100k sq miles are both right skewed.

After an initial look at the data, a principal component analysis (PCA) was performed and the findings shown in Figure 3 reveal that 75% of the variance comes from the first two features, resorts per state and state total skiable area. PCA values of the first two components were plotted on a scatter plot for each state and points were colored by their price quartile ranges. This is represented in Figure 4. Based on this figure, nothing discernable could be determined to lead to treating the states separate so going forward, all state data was treated the same. Positive correlations between prices and fastQuads, Runs, Snow Making ac, vertical drop resort\_night\_skiing\_state\_ratio also positively correlated

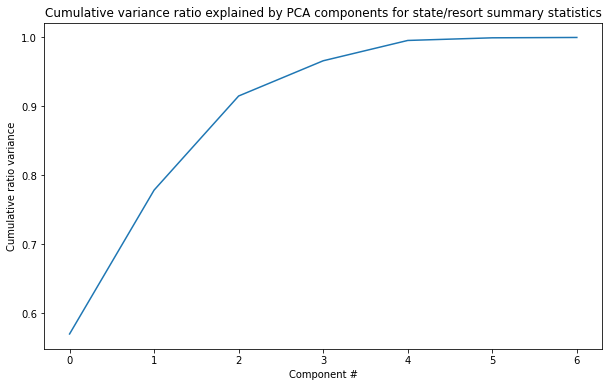


Figure . PCA of components for state/resort

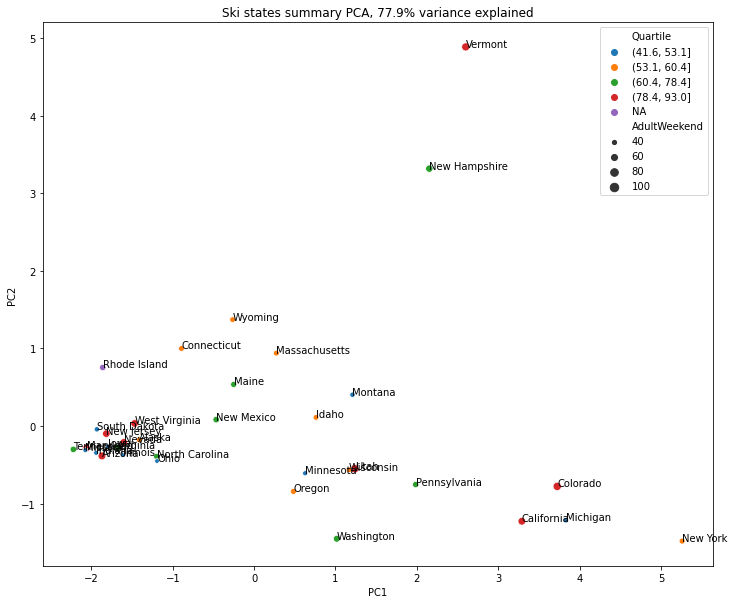


Figure . PC1 vs PC2 scatter plot

## Processing and Training

The mean, median, linear regression and random forest were utilized and compared to see which model would provide the most feasible use going forward. The findings are represented in Table 1. With Random forest providing less variability and a lower potential of error when determining ticket prices, Random Forest was selected as the model moving forward.

Table . MAE comparison of different models

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Mean | Linear Reg  Scaled  Median Imputed | Random Forest  Not scaled  Median Imputed |
| MAE | 19.1 | 10.5 | 9.6 |

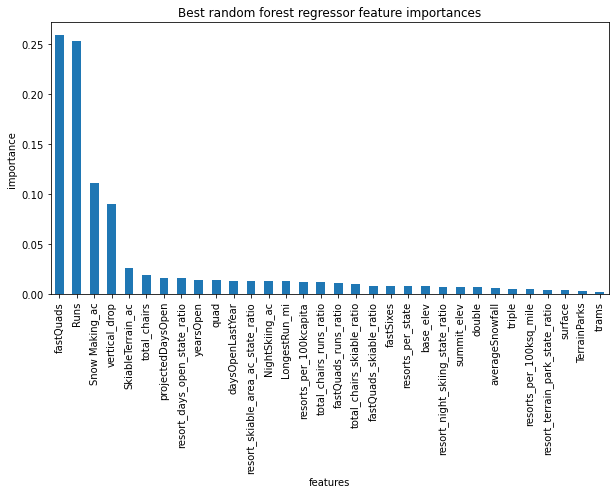


Figure . Random Forest Regressor feature importance

Plotting the feature importance from the random forest model as it relates to our target (ticket prices), the top four dominant features are fast Quads, Runs, Snow Making area, and vertical drop. These were focused on going forward to determine a better approximation of ticket prices.

## Modeling and Results

Looking the distributions of the major features (Table 2) and seeing where Big Mountain is compared to the rest of the resorts helps in determining if Big Mountain has the justification to increase ticket prices.

Table . Frequency of features and how Big Mountain compares

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Big Mountain currently charges 81 dollars for weekend ticket prices. Within the given data shown in Table 2 showing histograms of the most influential features as it relates to our target, Big Mountain is in the upper end of skiable terrain, longest run, number of runs, fastQuad numbers, total chairs, snow making area, vertical drop. Four of these, vertical drop, fastQuad number, number of runs, and snow making area are some of the influencing features to affect ticket price in the chosen model. In the given random forest model, by increasing runs by 1, vertical drop by 150 feet, total chairs by 1, and using our selected random forest model, ticket prices can be increased by $1.51. Other suggested changes did not yield a desirable increase in ticket prices in the model. Assuming 5 tickets sold and 350,000 visitors per year, this would generate roughly 2.64M dollars in revenue over the year and cover the opex of the added chair.

While run decreases resulted in less revenue as the run reduction increased, given that run count in the model was one of the features that had an influence on pricing, it may be worthwhile to consider various amounts of run numbers and their effects on pricing, while keeping in mind operating and fringe costs.

Data that would be a helpful addition would be the number of ticket sales sold in a year for the resorts and have that broken into day/night tickets. This would help better estimate actual revenue. I'd also suggest breaking down capex costs from investing into various features.

If the model was approved for future use, I'd suggest setting it up as an app that will allow main features that were shown to affect the ticket prices to be adjusted and allow the model to be re-run. I'd also suggest bringing in new data from competitors to see how the model output with changes done by the competitors.