Big Mountain Project Report

The initial question I sought to answer with this project was how to measure the value of Big Mountain's facilities compared to it's ticket price in relation to the national market. Working within scope of the dataset provided I began by taking a broad look at the overall picture, and making sure to patch up any obvious problems. Specifically I made sure that each resort was unique, fixed a few erroneous entries, and dropped the row for "Silverton Mountain" in Colorado because it's data couldn't be verified. I also pulled state population and area data from Wikipedia (see documentation) and added it to the general dataset. Since there were two separate ticket price types available for each resort I confirmed that there was a close linear relationship between the two prices, resorts in Big Mountain's price range tended to have little variance in ticket price, and chose Adult Weekend Ticket price as my target feature since it had the most complete information across all resorts. *Figure 1

In the process of performing exploratory data analysis I discovered several trends that might have an effect on ticket price including state resorts per capita as well as per 100K square miles to see what effect competition had on ticket price. However there did not seem to be a consistent relationship between resort density and price.

A principal components analysis revealed that four features accounted for 95% of the variance in ticket price *Figure 2-A. Looking deeper, I was able to model a summary of the PCA color coded by quintile of the ticket price *Figure 2-B. While there are a few outliers, Vermont, New Hampshire, New York, and Colorado, I could find no pattern to justify treating any state differently and determined to treat all states equally in future modeling. There was also no clear grouping between features and ticket price quintile.

With no clear pattern present, I created a heatmap of the features to see how they related to each other, and to ticket price in particular *Figure 3. A few features stood out immediately: fastQuads, Runs, Snow Making_ac, vertical_drop, and resort_night_skiing_state_ratio, the last item being a summary statistic of an individual resort's share of it's state's night skiing capacity *Figure 3-A.

At this point I moved on to preprocessing and training. I removed Big Mountain from the set and split the remainder into a training and testing portion with a 70/30 ratio. I used the mean ticket price to establish a baseline metric to compare my models against and when tested this produced a mean absolute error of near 19. I started with two linear regression models, filling missing numerical values with the respective means and medians. When tested, both models produced results with a mean absolute error of near 9.

I suspected that the linear regression models had been overfitted here, so in the next step I used SKLearn's SelectKBest function to reduce the number of features used. This, in combination with GridSearchCV, reduced the features down to just the 8 most important features which were: vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, and trams.

With the optimum features chosen I created a random forest regression model and cross-validated it against the linear regression model. The random forest regression model outperformed the linear regression model by nearly $1 and had a lower standard deviation. I also tested the random forest model by imputing the median for missing values, which helped, and scaling the features, which did not. I decided that moving forward I would use the random forest model.

Before making formal predictions I took a moment to confirm that I had enough data for optimum performance. I used the learning_curve function to cross validate the training scores as set sizes increased and concluded that the amount of data included was satisfactory *Figure 4.

The initial result of the model illustrated that Big Mountain's current price of $81 per ticket contrasted a modeled price of $94.22 balanced against an absolute error of $10.39. This is without any changes to the resort's current features. Several changes were shortlisted including:

1) Permanently closing down up to 10 of the least used runs.
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 basic assumptions used when modeling the effectiveness of each of these scenarios is that Big Mountain receives 350,000 visitors and that each visitor buys 5 tickets.

In the first scenario the model indicates that one run can be closed without any changes in revenue or ticket price, but that closing ten runs would result in a ticket price that models over $1.50 lower than projections and brings in a loss of nearly 3MM in revenue. The best value appeared to be either 5 runs at a ticker price decrease of $0.75 or 8 runs at a ticket price decrease of $1.25. *Figure 5

The second scenario supported a ticket price increase of $8.46 and is projected to bring in an increase of revenue of greater than 14.5MM, and although this is balanced against the 1,540,000 cost of operating the machinery, the 13MM difference is clear. These numbers are based on the assumptions listed above.

The third and fourth scenarios were modeled to make no significant improvements to ticket price or revenue.

In summary I recommend options 1 and 2 as well as an increase in ticket price to around \$102 +/- \$9. More information may be used to refine this recommendation including data on the number of visitors to the resort.

Figure 1 (Ticket Prices across all resorts)

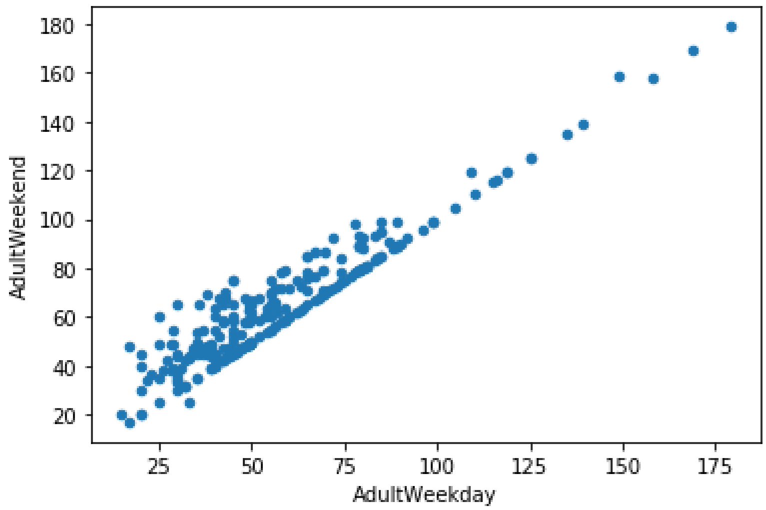


Figure 2-A

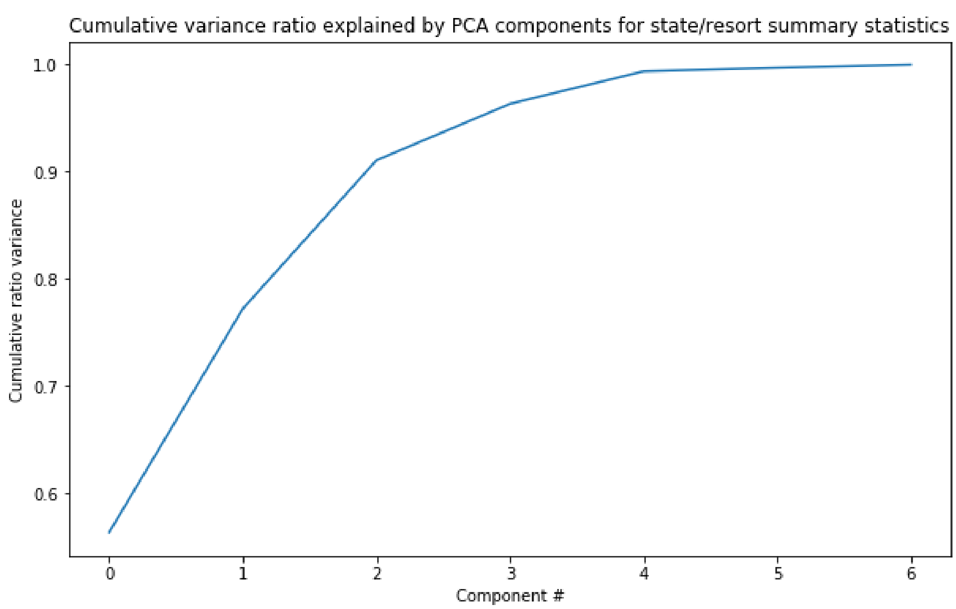
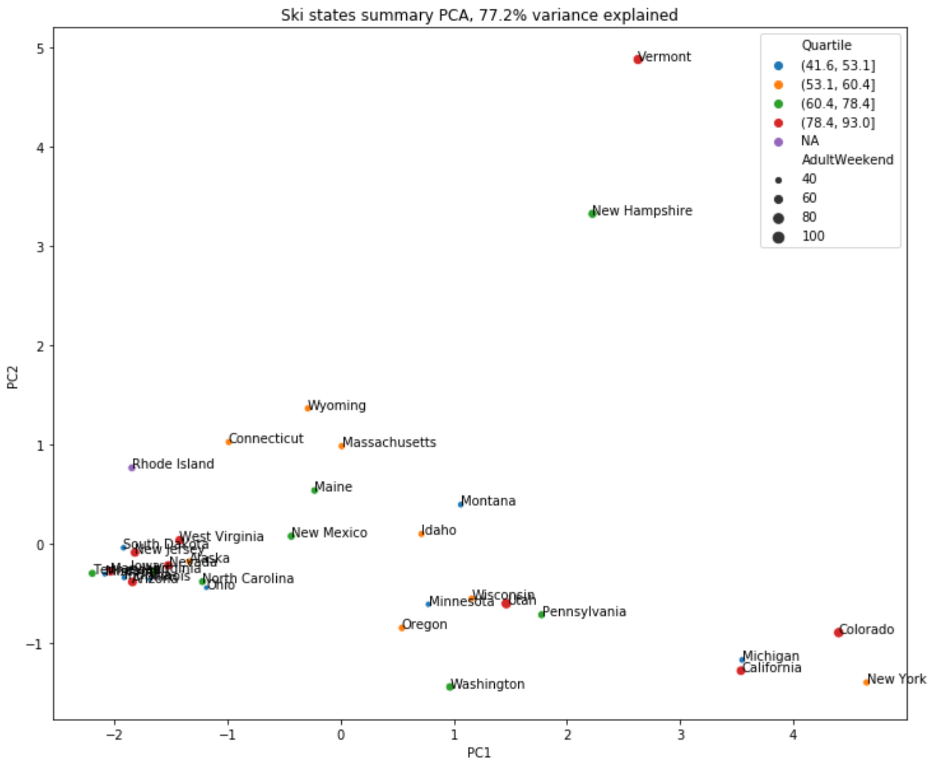


Figure 2-B



Ski states summary PCA, 77.2% variance explained
Quartile
(41.6, 53.1]
(53.1, 60.4]
(60.4, 78.4]
(78.4, 93.0]
NA
AdultWeekend
40
60
80
100
Vermont
New Hampshire
Wyoming
Connecticut
Massachusetts
Rhode Island
Maine
Montana
Idaho
New Mexico
West Virginia
North Carolina
Ohio
Minnesota
Wisconsin
Utah
Pennsylvania
Oregon
Colorado
Michigan
California
New York
Washington
PC1
PC2

Figure 3

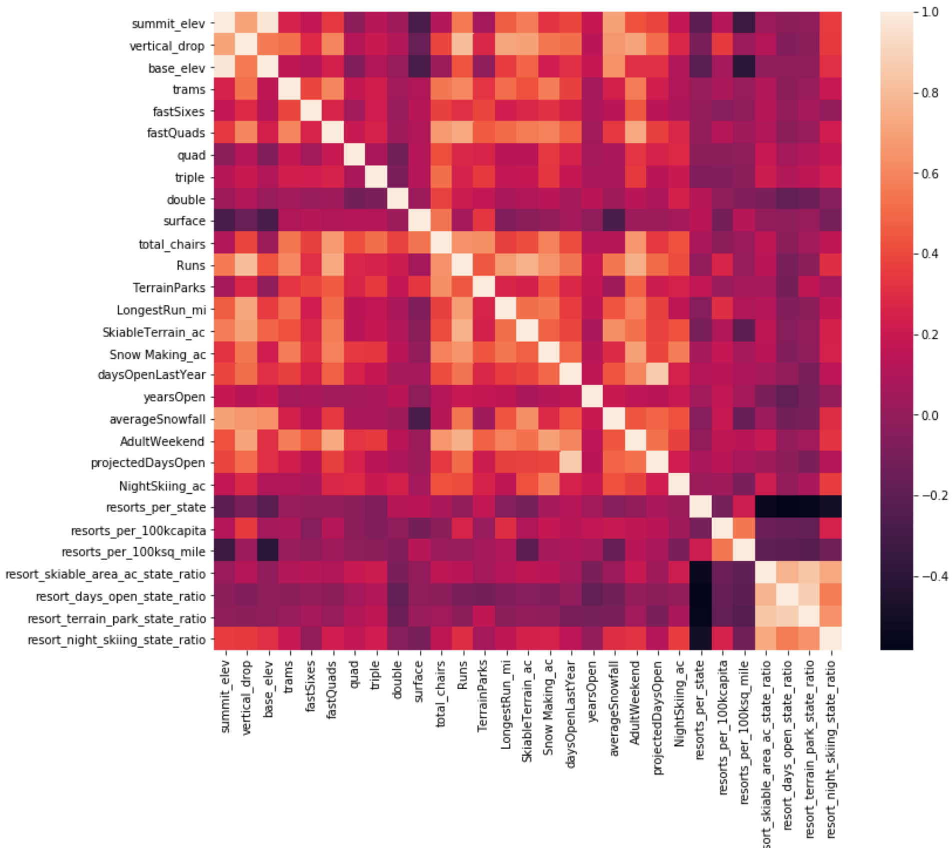


Figure 3-A

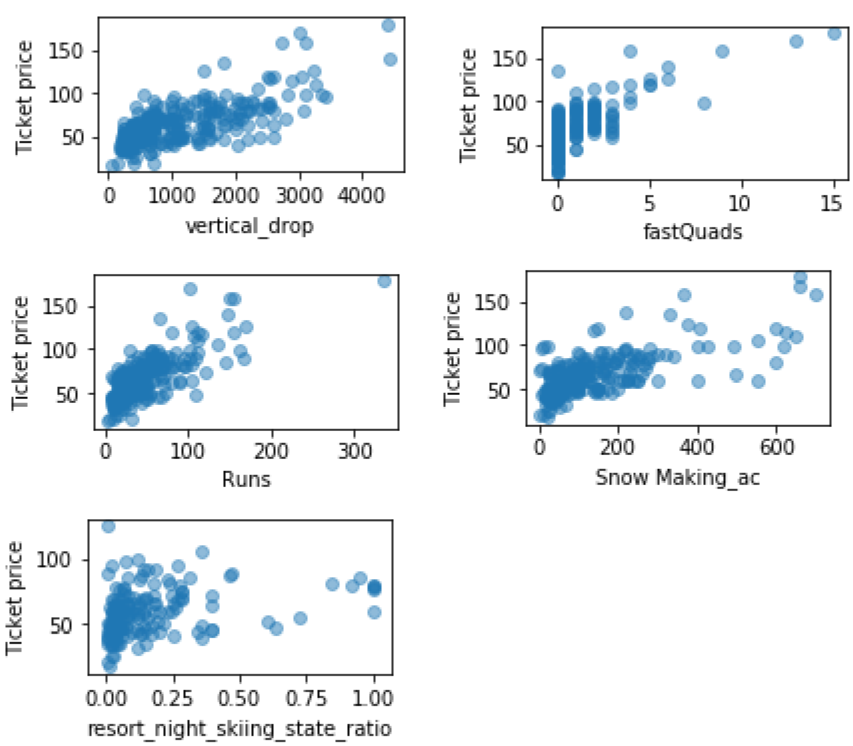


Figure 4

Figure 5