Report on Whether Big Mountain Resort's Facilities Support an Increase in its Ticket Price

Problem description

Big Mountain Resort has been a popular ski resort in the state of Montana for many years, drawing about 350,500 visitors every year. So far, its pricing strategy has been to charge a premium above the average price on the market. However, this was not a data-driven decision, so there is a suspicion that this current price does not adequately represent the value of the services provided. It also does not reflect the fact that some of their facilities are more valuable to customers than others. Therefore, this project was undertaken to utilize readily available data on numerous ski resorts across the USA to make a prediction on which ticket price for Big Mountain is supported by its facilities.

Description of source data

The source data, provided by Big Mountain Resort database manager Alesha Eisen, contains data on 330 ski resorts in the United State. Each contains 27 features, such as the state where it is located, summit elevation height, number of trams, skiable terrain area, price of an adult ticket for weekdays and weekends, the area covered by snow-making machines, and many others.

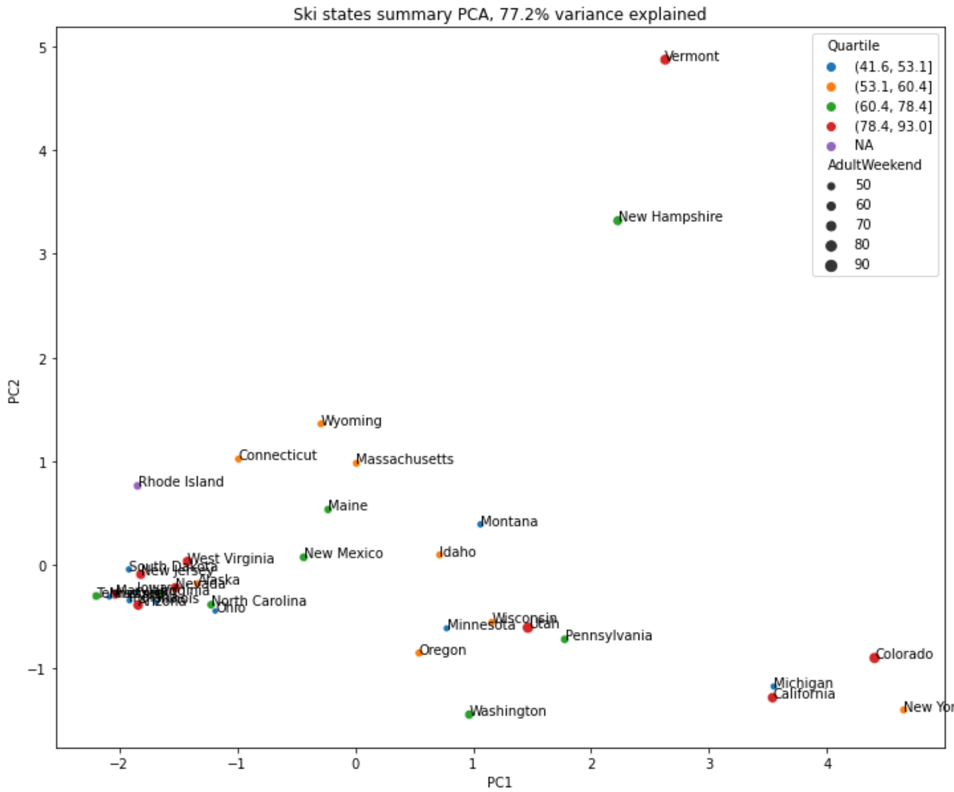
Conclusions from exploratory data analysis

The usefulness of the "state" label

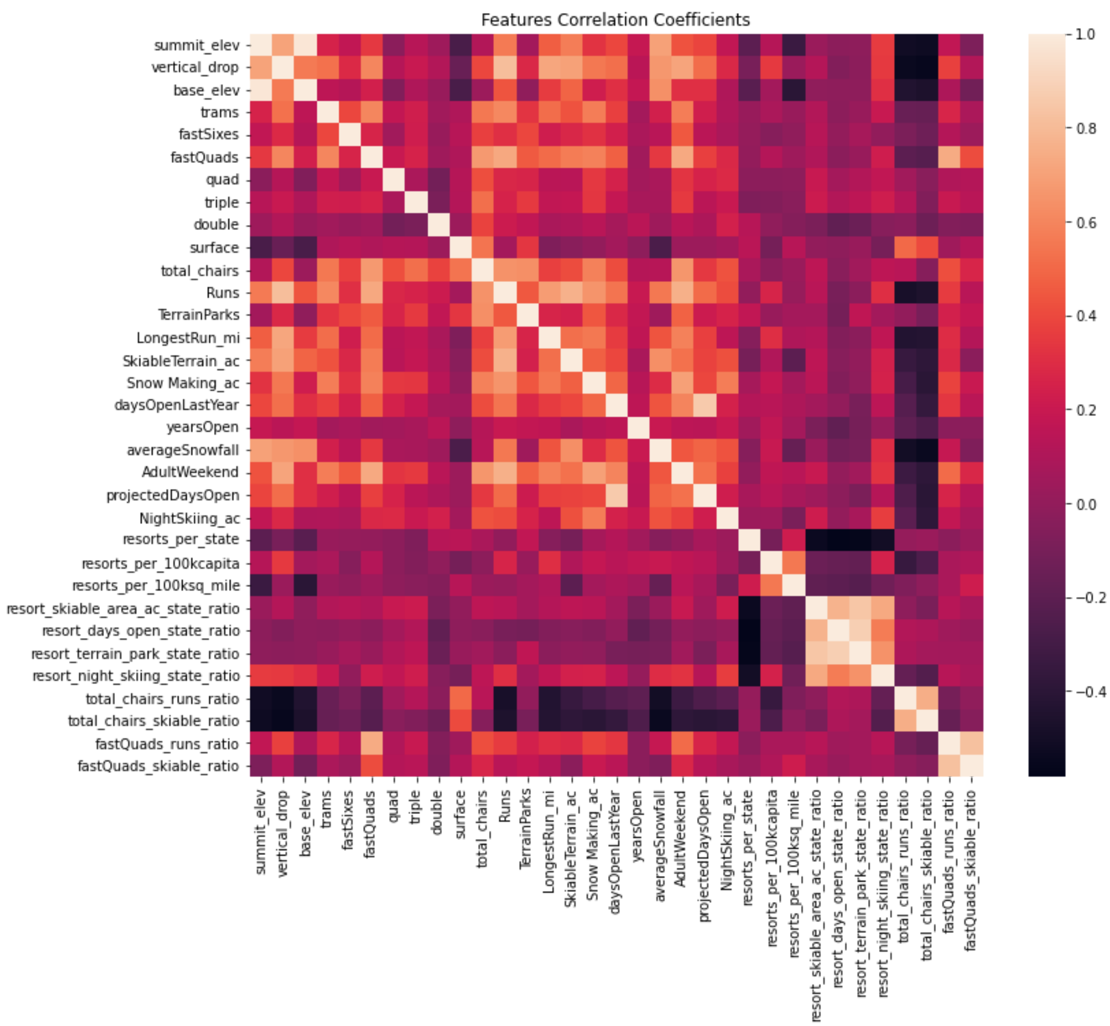
We wanted to see if the state where each resort is located should be considered an important feature for predicting ticket price. We used Principal Component Analysis (PCA) to see if different states "clustered" into different groups. The following first plot of the next page shows the first two components of the PCA; the size of each of the points corresponds to the state's average ticket price, and the points are also color-coded based on which of the four quantiles the resort lies on the distribution of ticket price. Other than the fact that Vermont and New Hampshire appear as outliers, no obvious patterns can be seen. **We therefore conclude that the state where a resort is located is not a good predictor of its ticket price, and it was therefore not used in the model.**

Correlation Coefficient Heat Map

The second figure on the next page shows a heat map of the correlation coefficients of all of the features. It shows that some of the features that correlate strongly with the target feature of ticket price are number of fast four-person chair lifts, total number of runs, snow making area, total runs, and total chairs. The feature that correlates the most is the night skiing area ratio (to the total night skiing area for the state).



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
50
60
70
80
90
Vermont
New Hampshire
Wyoming
Connecticut
Massachusetts
Rhode Island
Maine
Montana
New Mexico
Idaho
West Virginia
North Carolina
Ohio
Minnesota
Wisconsin
Utah
Pennsylvania
Oregon
Colorado
Michigan
California
New York
Washington
PC1
PC2



Features Correlation Coefficients
summit_elev
vertical_drop
base_elev
trams
fastSixes
fastQuads
quad
triple
double
surface
total_chairs
Runs
TerrainParks
LongestRun_mi
SkiableTerrain_ac
Snow Making_ac
daysOpenLastYear
yearsOpen
averageSnowfall
AdultWeekend
projectedDaysOpen
NightSkiing_ac
resorts_per_state
resorts_per_100kcapita
resorts_per_100ksq_mile
resort_skiable_area_ac_state_ratio
resort_days_open_state_ratio
resort_terrain_park_state_ratio
resort_night_skiing_state_ratio
total_chairs_runs_ratio
total_chairs_skiable_ratio
fastQuads_runs_ratio
fastQuads_skiable_ratio

Model development

Three models were explored during this project: a dummy model that always predicts the mean target feature regardless of input, a linear model, and a random forest model. A train/test split of 70/30 percent was applied in their development. Various ways of building these models were considered, such as whether missing values should be imputed using the mean of the median, whether the data should or should not be scaled to zero mean and unit standard deviation, how many total features should be used, and — for the random forest model — the number of trees used. Also, cross-validation was used throughout to ensure that the performance was not measured exclusively on an arbitrary subset of the data.

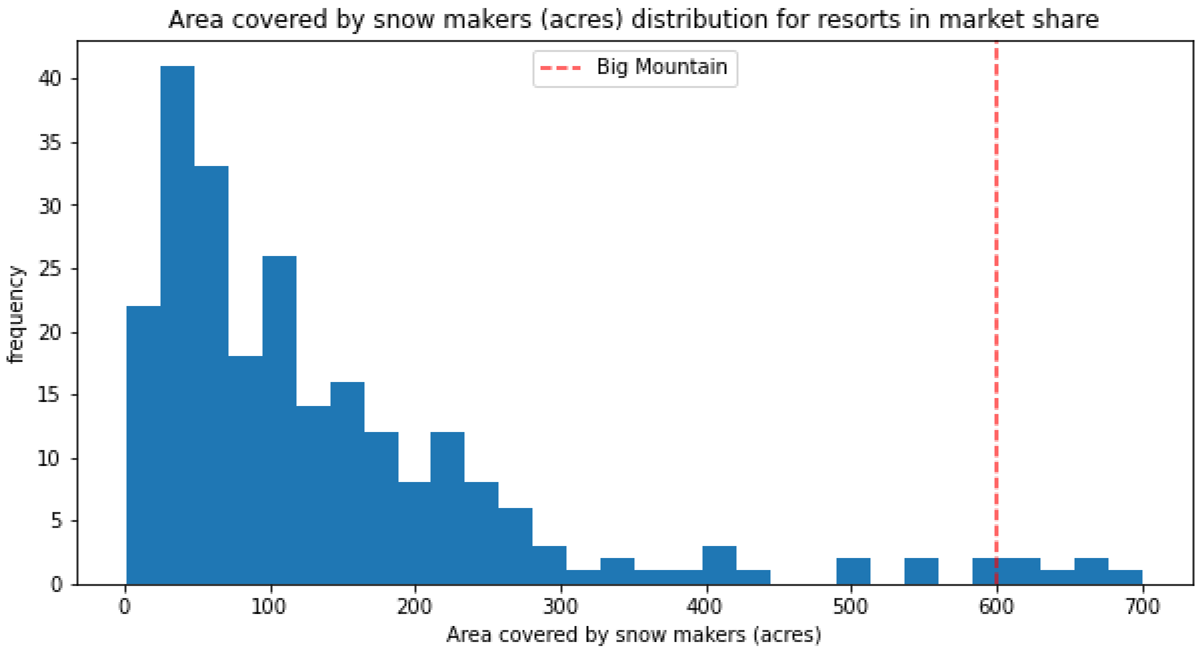
It became clear that the performance of the linear and random forest models performed far better than the dummy model. **The best version of the random forest model performed marginally better than the linear model, and it was therefore chosen as the one used for solving the business problem.**

Conclusions from the model

The chosen model predicted that the current facilities of Big Mountain resort can support a ticket price of $95.87. Since the expected mean absolute error is $10.39, and the current ticket price is $81.00, **there is evidence that the ticket price should be increased.**

One interesting aspect of note is that, when it comes to many of the features that are considered important (as determined during the model development stages), Big Mountain Resort has those features in higher-than-average-quantities. For example, the histogram on the next page shows where Big Mountain Resort lies on the histogram for the area covered by snowmakers; it shows that Big Mountain Resort's value is much higher than the average. This could be interpreted to mean that, in terms of the important features, Big Mountain Resort ranks among the best and, therefore, the ticket price should be increased to reflect that.

Another question that can be answered by the model is which of the proposals that have been suggested by the business for cutting costs and increasing revenue would lead to an increase in revenue. **The chosen model**, combined with the assumption that on average each of the 350,000 expected visitors will purchase 5 tickets, **leads to the conclusion that the only viable proposal is to increase the vertical drop by 150 by adding a new run and installing a new chair lift. The model suggests that this supports an increase in the ticket price of $1.99, leading to an increase in revenue of $3474638.**


Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
0
5
10
15
20
25
30
35
40
0
100
200
300
400
500
600
700
Area covered by snow makers (acres)