Big Mountain Resort Pricing Report

Big Mountain Resorts (BMR) recently purchased and deployed a new chair lift which increases annual operating costs by $1.54M. Our hypothesis is that BMR's pricing, which was found by simply taking the average of all resort ticket prices, was undervalued. I was tasked with analyzing all resorts in the US to find the optimal pricing for BMR so as to offset this $1.54M increase in operating costs and increase profitability.

I was given the "Resort.csv" dataset by our Database Manager. There were a number of measurable features in the dataset and the majority of the data was present although there were some missing values, particularly, missing pricing data. Data for BMR was complete & missing no values. The data was cleaned to produce a dataset that would be workable into a pricing model that could answer our initial question, but before resorts with missing pricing data were dropped, I found it to be useful to see the distribution of resort features that I would be using to train the model (Figure1).

The dataset features inherently have numerical values on different scales (ie vertical drop in feet and absolute value of runs) so the data was scaled and I used principal component analysis (PCA) to determine the percentage of the variability that the top few components accounted for. The top two accounted for 77% of the variability in the data. My exploratory data analysis showed that there were very specific features that had a high positive correlation to a resort's price (Figure 2). Runs, vertical drop, snow making, total chairs, & fast quads appear to have the strongest positive correlation to pricing.

Starting with our ski_data, I first extracted our resort data for later use and then determined a 70/30 train/test split to train and test the models. I established a baseline by taking the mean of our pricing data and found the average price to be $63.81 and that (given the mean absolute error) on average, you might expect to be off by around $19 if you guessed ticket price based on an average of known prices.

I then built a linear regression model that initially produced a mean absolute error of about $9.40 using all features in the data in the model but it appeared to be overfit to the test data. The model was then refined and cross validated. It produced average prices distributed around $63 which is in line with our test data as well as an updated mean absolute error of about $10.50. It was determined that using 8 features in the model produced the highest cross validation score with the lowest variance. The model showed that vertical drop had by far the most significant positive effect on price, followed by snow making, total chairs, fast quads, and runs.

I also produced a random forest model. Running for best parameters, it was found that imputing with the median was best although scaling didn't help. This model produced a distribution of prices around $70 which was higher than our baseline average. The features it found to be the most significant were the same four features found in our linear regression, however fast quads and runs were found to be most important, followed by snow making and vertical drop (Figure 3).

After comparing the two models side by side, I determined to use the random forest model going forward as it produced a cross validation mean absolute error $1 lower than the linear regression model and showed less variability.

Before fully deploying our model, I found that BMR is on the higher side of all pertinent features which would suggest its predicted price should be much higher than the mean. BMR's current pricing is $81/ ticket. Our model shows that BMR's pricing should be at $95.87, and even with an expected mean absolute error of $10.39, our hypothesis was confirmed that BMR's pricing is too low. My recommendation is to increase BMR's ticket price to $95-$96.

BMR also has four possible courses of action that the executive team is entertaining that our model can predict pricing for.

Scenario 1: Close up to 10 of the least used runs. The model shows there is no predicted change in price in closing 1 run. Closing 2 or 3 runs reduces the predicted price per ticket, however, if 3 runs are closed, Big Mountain may as well close 5 runs as there is no price difference between closing 3, 4, or 5 runs. There is a similar price plateau for 6, 7 or 8 run closures (Figure 4). 1 run should definitely be closed as it will not affect price, but depending on the operating costs to maintain those runs, it may be recommended to close a total of 5 or 8 runs as it would only decrease price by about $.70 cents or $1.40 per ticket respectively, which reduces annual revenue by only about $1.2M or $2.4M respectively. This could actually aid in reducing perceived price shock felt by customers when raising the price per ticket. To test this, I would first close 1 run, then 5, then 8 only if you do not see a dropoff in ticket sales at each step.

Scenario 2: Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. The model shows this change in features could justify an increase in ticket price by an additional $1.99. Over the season, this could be expected to add an approximate 3.5M to annual revenue. This option should be implemented.

Scenario 3: Same as scenario 3 but adding 2 acres of snow making. The model shows no additional increase in price from scenario 2. This scenario should not be implemented.

Scenario 4: Increase the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability. This does not change the predicted price at all so this option should not be implemented.

Big Mountain was already priced higher than other Montana resorts, however the predicted price was much higher than the previous "model" used (calculated on average price of all resorts alone) because the prior model did not factor in a resort's features, of which, Big Mountain has more than most to offer. This may come as a surprise to executives because without a predictive model such as the one implemented here, it would be very difficult to ascertain the true value of these features. As demonstrated, this model can be used to explore predicted changes in price given a proposed change in features. This model has been saved and can be readily accessed by the business analyst team for future use in experimenting with how future proposed changes would be expected to impact price.

Figure 1

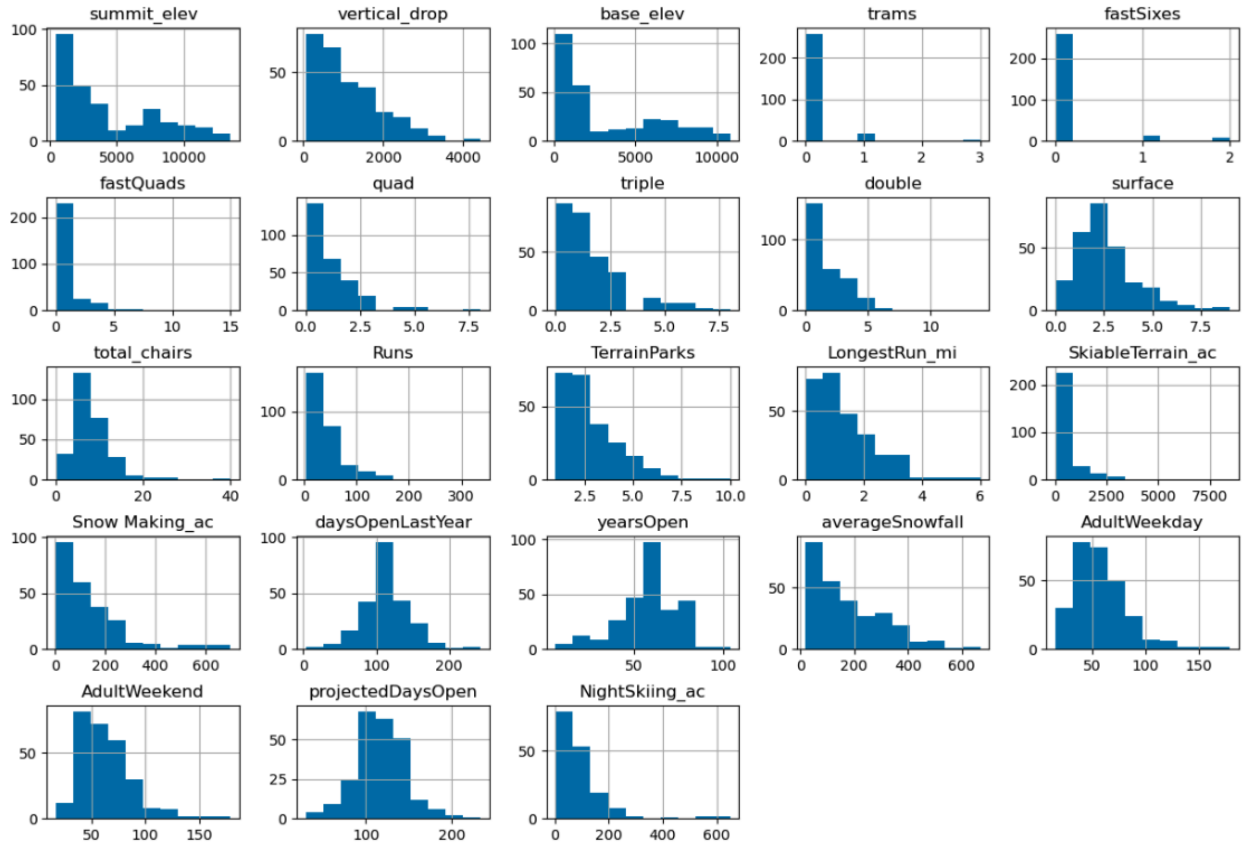

summit_elev
vertical_drop
base_elev
trams
fastSixes
fastQuads
quad
triple
double
surface
total_chairs
Runs
TerrainParks
LongestRun_mi
SkiableTerrain_ac
Snow Making_ac
daysOpenLastYear
yearsOpen
averageSnowfall
AdultWeekday
AdultWeekend
projectedDaysOpen
NightSkiing_ac

Figure 2

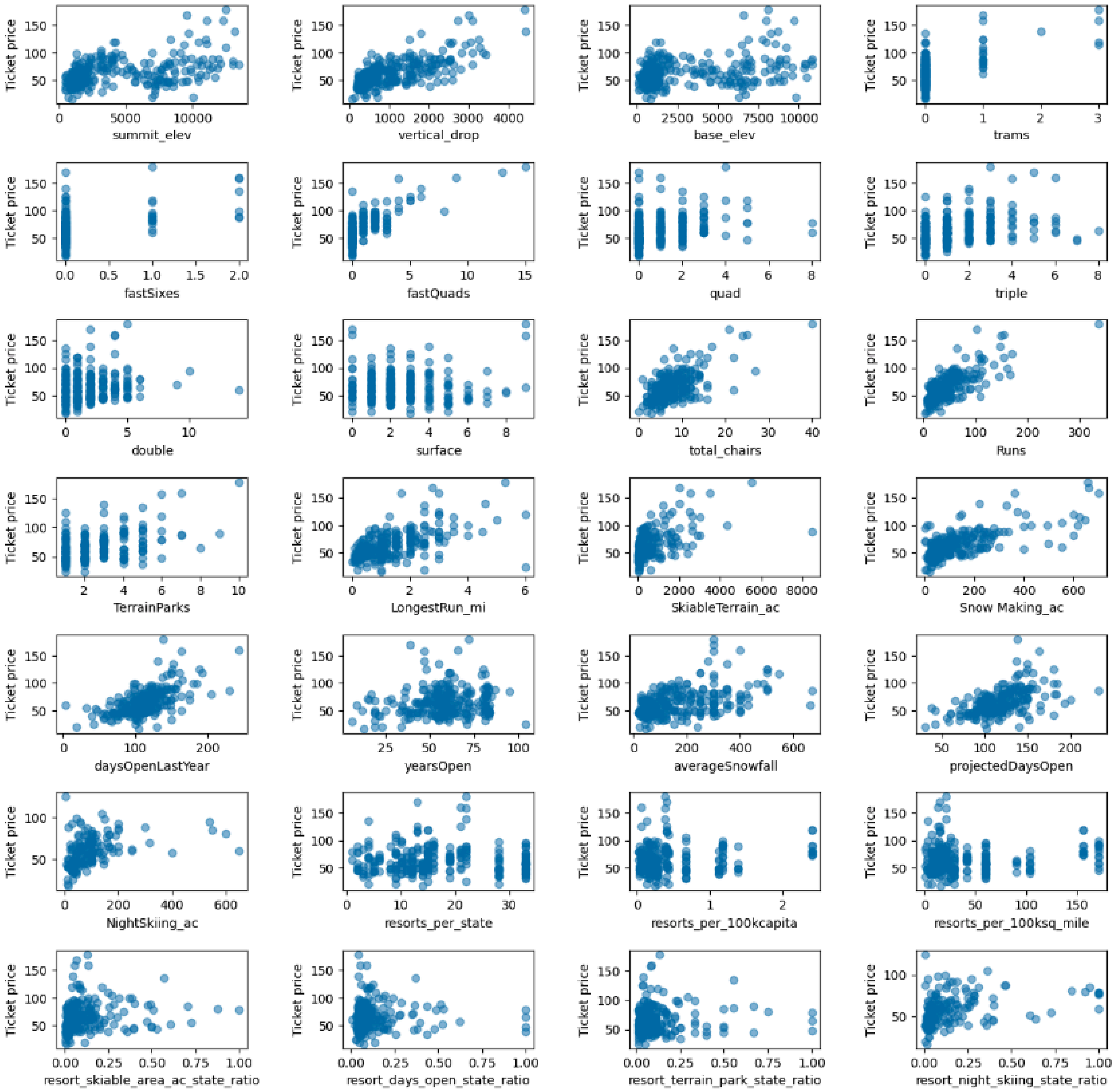

Ticket price
summit_elev
vertical_drop
base_elev
trams
fastSixes
fastQuads
quad
triple
double
surface
total_chairs
Runs
TerrainParks
LongestRun_mi
SkiableTerrain_ac
Snow Making_ac
daysOpenLastYear
yearsOpen
averageSnowfall
projectedDaysOpen
NightSkiing_ac
resorts_per_state
resorts_per_100kcapita
resorts_per_100ksq_mile
resort_skiable_area_ac_state_ratio
resort_days_open_state_ratio
resort_terrain_park_state_ratio
resort_night_skiing_state_ratio

Figure 3

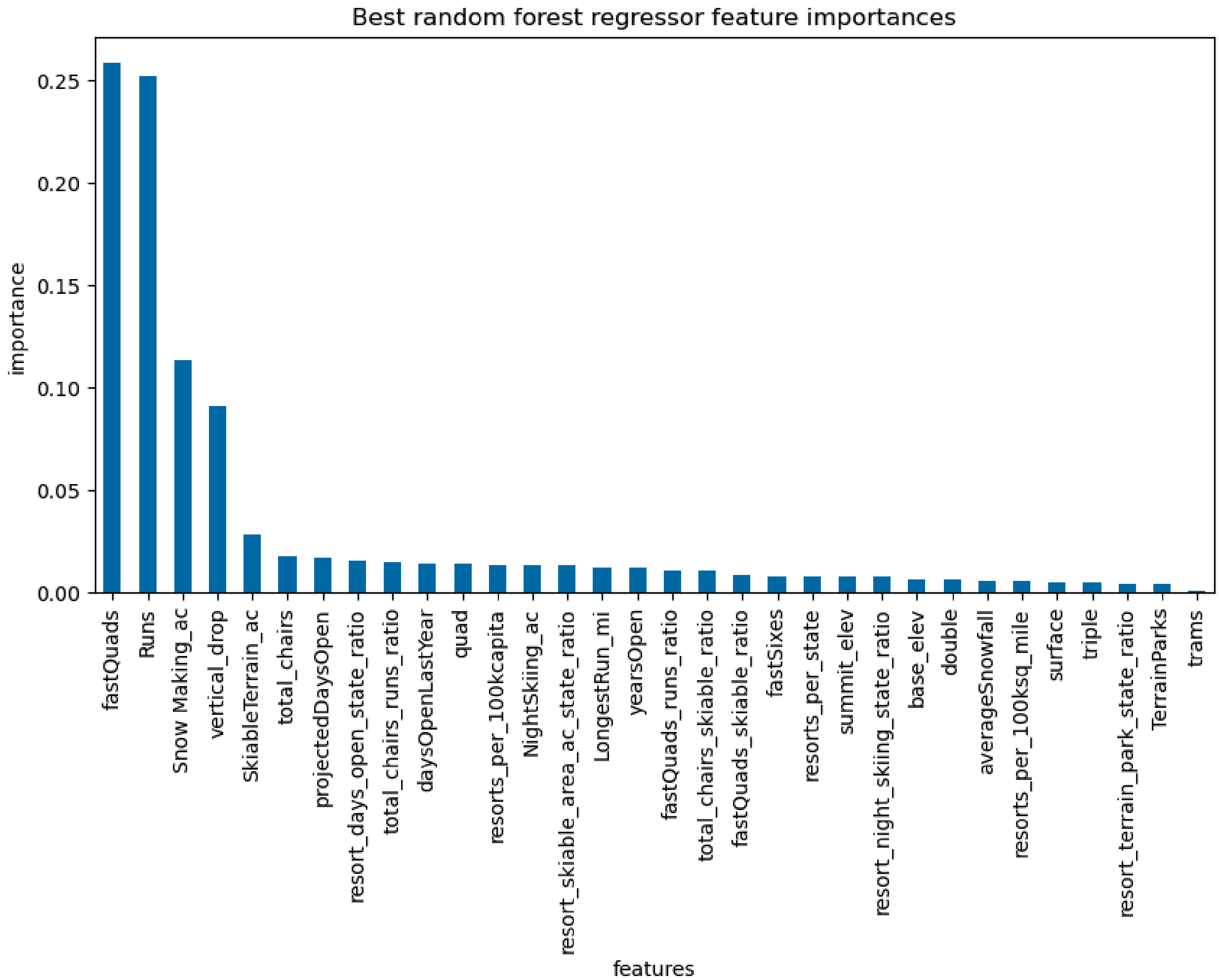
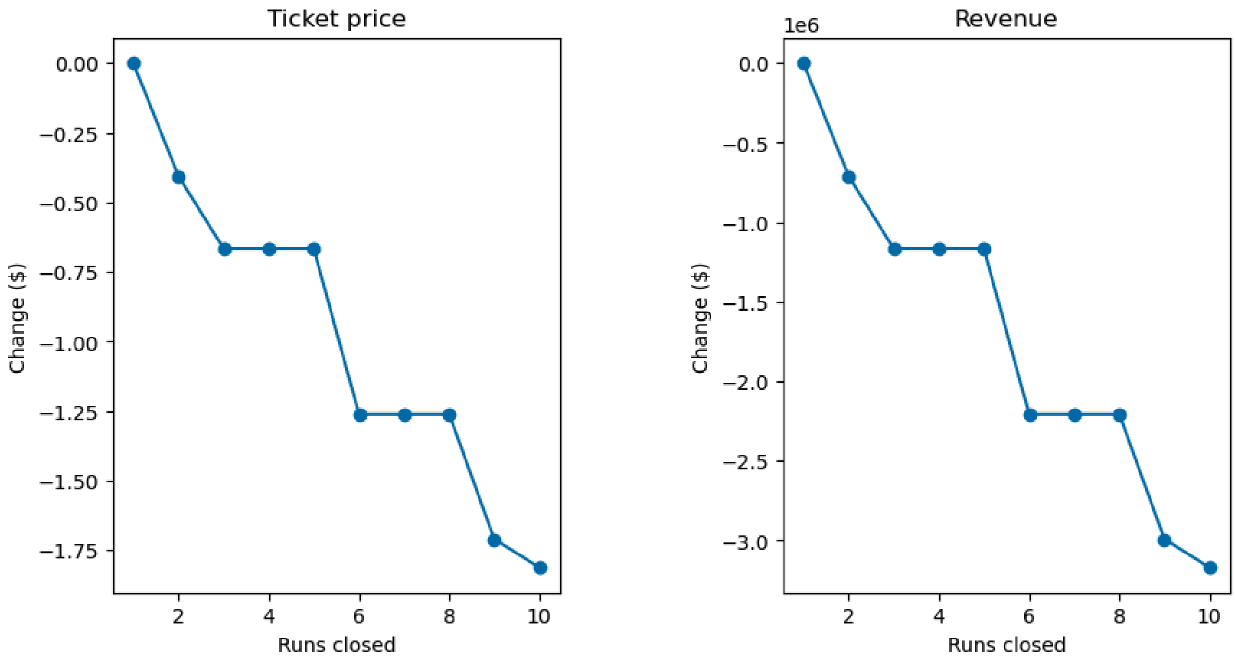

Best random forest regressor feature importances
importance
0.25
0.20
0.15
0.10
0.05
0.00
fastQuads
Runs
Snow Making_ac
vertical_drop
SkiableTerrain_ac
total_chairs
projectedDaysOpen
resort_days_open_state_ratio
total_chairs_runs_ratio
daysOpenLastYear
quad
resorts_per_100kcapita
NightSkiing_ac
resort_skiable_area_ac_state_ratio
LongestRun_mi
yearsOpen
fastQuads_runs_ratio
total_chairs_skiable_ratio
fastQuads_skiable_ratio
fastSixes
resorts_per_state
summit_elev
resort_night_skiing_state_ratio
base_elev
double
averageSnowfall
resorts_per_100ksq_mile
surface
triple
resort_terrain_park_state_ratio
TerrainParks
trams
features

Figure 4



Ticket price
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
Change ($)
2
4
6
8
10
Runs closed
Revenue
1e6
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
Change ($)
2
4
6
8
10
Runs closed