Guided Capstone Project Report:
Big Mountain Resort

INTRODUCTION

1.1 Background

Big Mountain Resort is a ski resort in Montana that offers wonderful views of Glacier National Park and Flathead National Forest. Approximately 350,000 yearly visitors enjoy access to 105 trails via 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers.

Taking visitors from a base elevation of 4,464 ft to the summit at 6,817, these features and trails also include a vertical drop of 2,353 ft.

Recently, Big Mountain Resort installed an additional chair lift to help increase the distribution of visitors across the mountain. While it does make it easier for visitors to fully enjoy the Resort, this additional chair increases operating costs by $1,540,000 for the season which has spurred a review of existing pricing strategies.

Currently, the ticket price is based on the market average and does not provide the business with a good sense of value offered compared to others in the market. Additionally, while they do charge a premium, there’s a suspicion that Big Mountain is not valuing its facilities as much as it could. Consequently, this ambiguity hampers investment strategy and how best to address the increased cost to operations.

1.2 Project Goal + Scope

This project analyzes and compares data looking at facilities and trail offerings within Big Mountain Resort’s US market segment to understand current positioning domestically and regionally. It seeks to answer the question: how can Big Mountain Resort best determine ticket price for 350,000 riders to cover an increase of $1,540,000 in operating expenses for this season through comparable market analysis of offerings and facilities? Does the data support increasing the ticket price by $4.40 to meet the increased operational break-even point?

The scope of this project is set to current facilities and trail offerings for in-season only. Off-season valuation is considered out of scope as is any marketing efforts to increase the number of yearly riders.

Additionally, at the request of management, the following scenarios are evaluated through modeling:

Scenario 1: Permanently close down up to 10 of the least used runs.

Scenario 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

Scenario 3: Same as number 2, but adding 2 acres of snow making cover.

Scenario 4: Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres .

1.3 Constraints

As these scenarios indicate, management is open to changes as long as cutting costs does not devalue the ticket price or, alternatively, any changes made support a higher ticket value.

While management is open to cutting costs, there is not any data available for rider usage by facility type so determining which costs are eligible to be cut or which facility type is being undervalued based on usage/popularity is not available.

1.4 Data Source

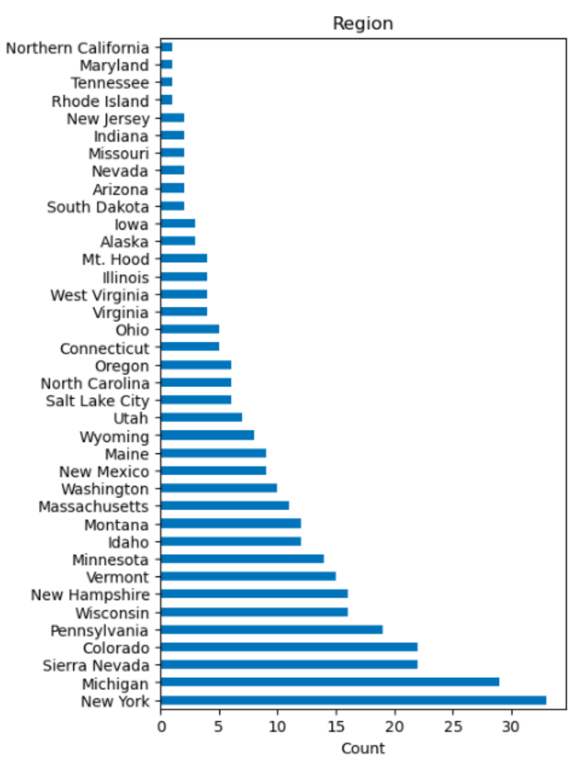
Data used for this project comes from a CSV File containing data for 330 resorts in the US within Big Mountain Resort’s market segment.

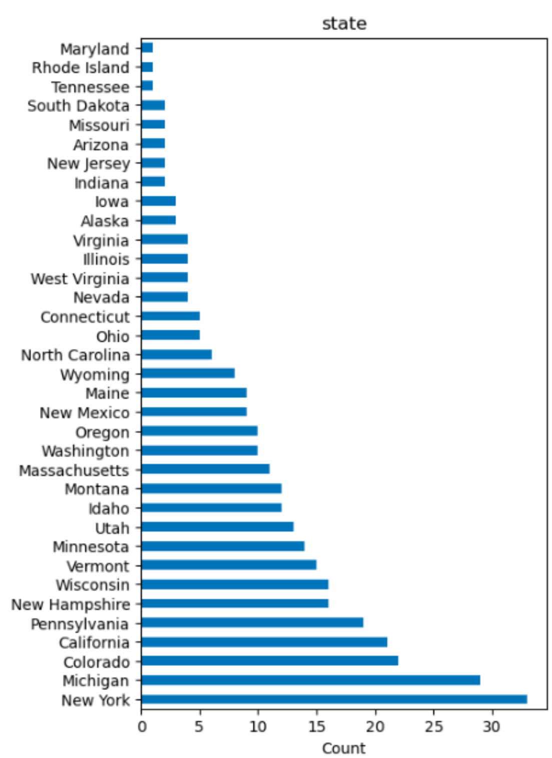
APPLYING DSC PROCESS

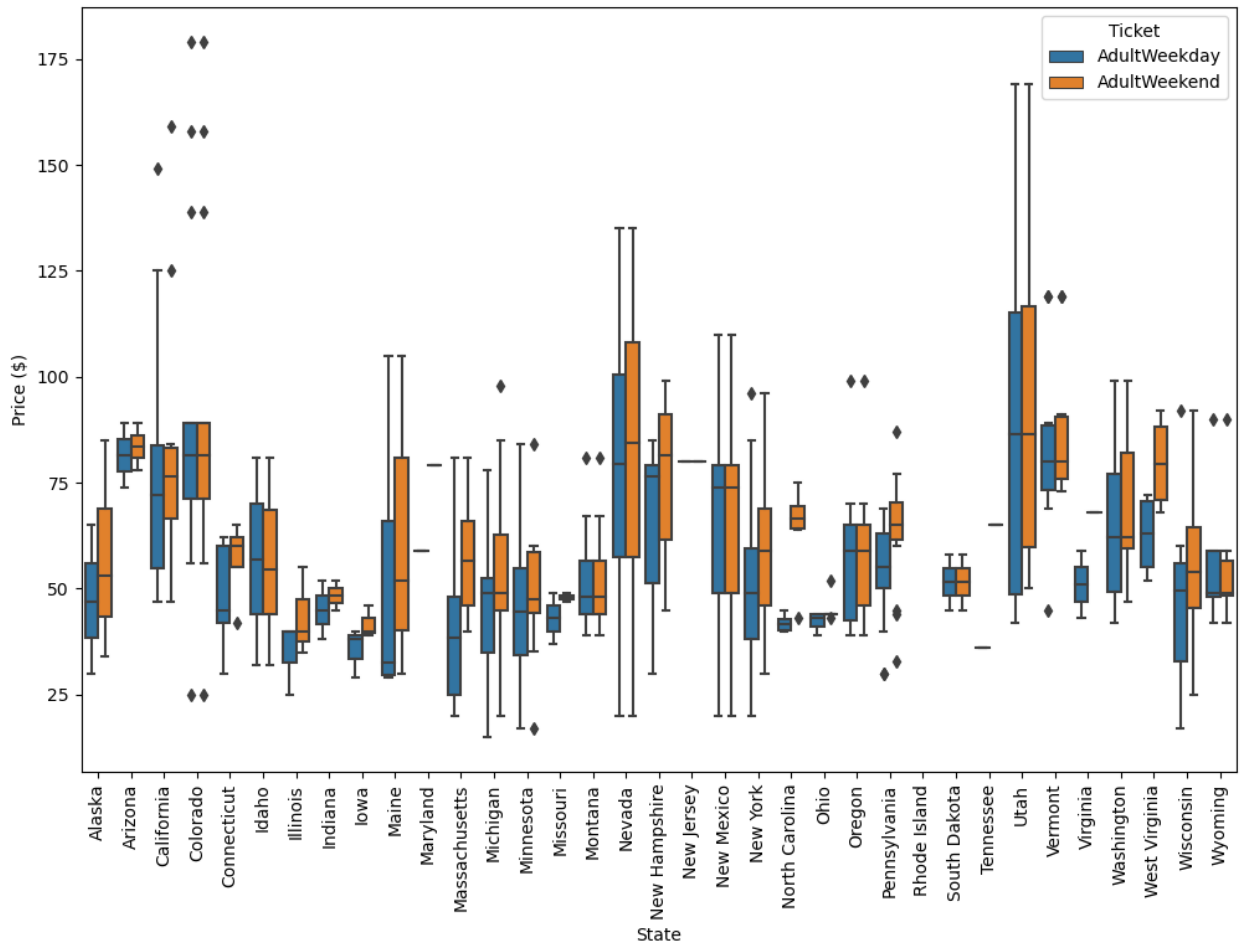
2.1 Data Wrangling

As an entry point, we start our exploration by looking at the distribution of resorts by region and state to get a general understanding of the markets' respective sizes and leaders. The bar graphs below show New York and Michigan topping off both lists with Montana as 11th largest when looking at the number of resorts in 38 regions and 12th across 35 states.

Big Mountain Resort seems relatively well positioned in terms of market health in Montana but what about ticket price? Spending some time with the boxplot below gives a sense that Montana is somewhere in the middle when it comes to ticket prices. There are states that have a much greater range and vice versa. While some are priced at much more luxury prices and some discounted, Montana seems to have a more specific market within this context.

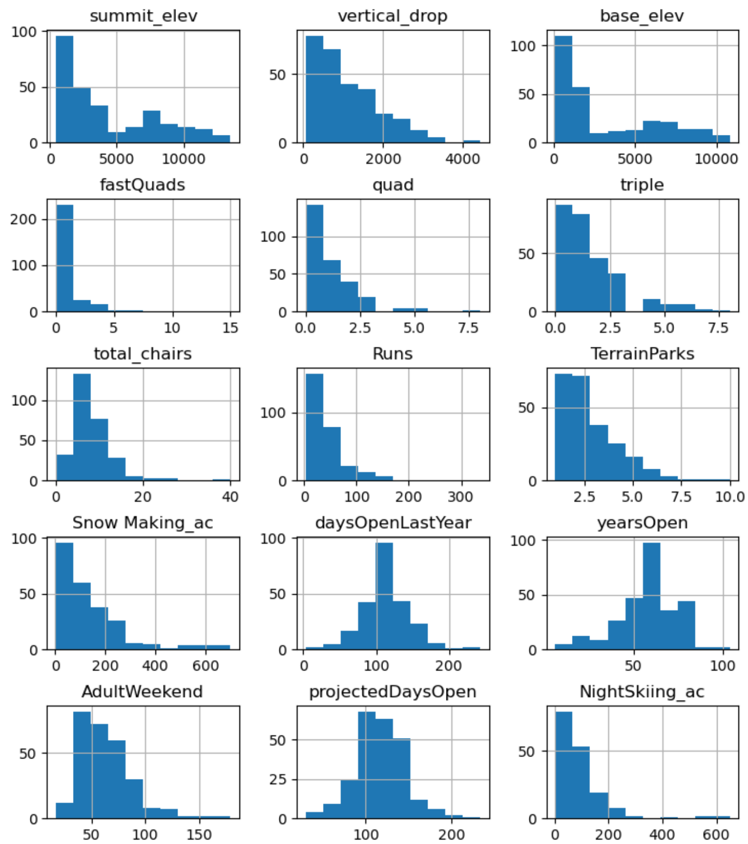






Of all 330 resorts, about 16% were missing price information with 14% missing values for both Weekday and Weekend prices. Even as such, with this initial context we have the broad strokes visualized. To get an understanding of what exactly is contributing to these prices, however, we need to explore what the distribution of additional features shows us.

First verifying the plausibility of our dataset, we can observe some questionable outliers and significant missing values in the features distribution graphs below. We drop the appropriate rows of resorts and exclude the Fast Eight feature moving forward since no resorts in the dataset have this equipment.



Now that we have the distribution of features across all resorts, we can add depth to our understanding by adding some context around population. First we imported the data by state, checking for missing states, and begin exploring the relationship of our target features: Weekday and Weekend prices. Weekend prices showed the least number of missing values for Montana so we're dropping weekday prices moving forward. The final shape of our wrangled data shows 277 rows and 23 features.

2.2 Exploratory Data Analysis

Now that we have a clean dataset, we can look at the national picture for the market by state, examining total state area, total state population, resorts per state, total skiable area, total night skiing area, and total days open. The top states for each are listed below with Montana appearing in the top state area and total skiable area.

Top State Area	
Alaska	665384
California	163695
Montana	147040
New Mexico	121590
Arizona	113990

Total State Population	
California	39512223
New York	19453561
Pennsylvania	12801989
Illinois	12671821
Ohio	11689100

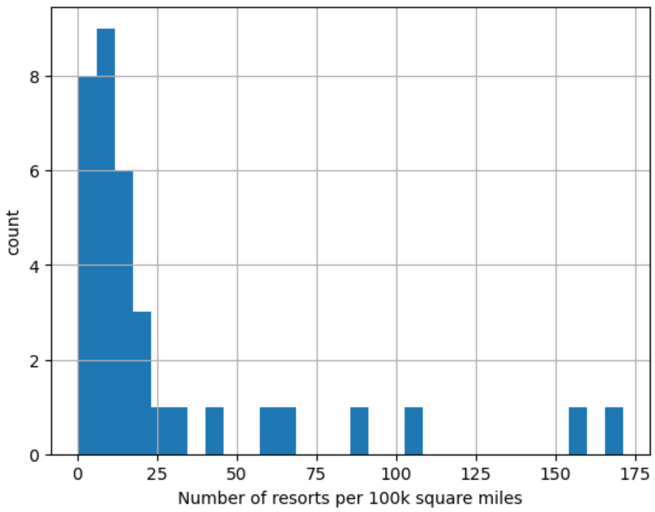
Resorts Per State	
New York	33
Michigan	28
Colorado	22
California	21
Pennsylvania	19

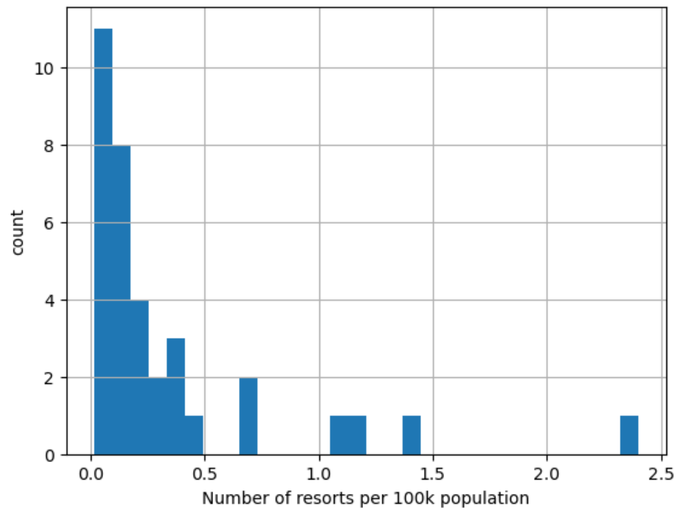
Total Skiable Area	
Colorado	43682.0
Utah	30508.0
California	25948.0
Montana	21410.0
Idaho	16396.0

Total Night Skiing Area	
New York	2836.0
Washington	1997.0
Michigan	1946.0
Pennsylvania	1528.0
Oregon	1127.0

Total Days Open	
Colorado	3258.0
California	2738.0
Michigan	2389.0
New York	2384.0
New Hampshire	1847.0

With this, we can now explore resort density per state, requiring the removal of some state-specific data. Looking at the distributions of resorts per capita and resorts per 100k square miles, we can observe outliers as we further validate the data. Top states for each summarized below as well with Montana appearing in the top states by resort per capita.





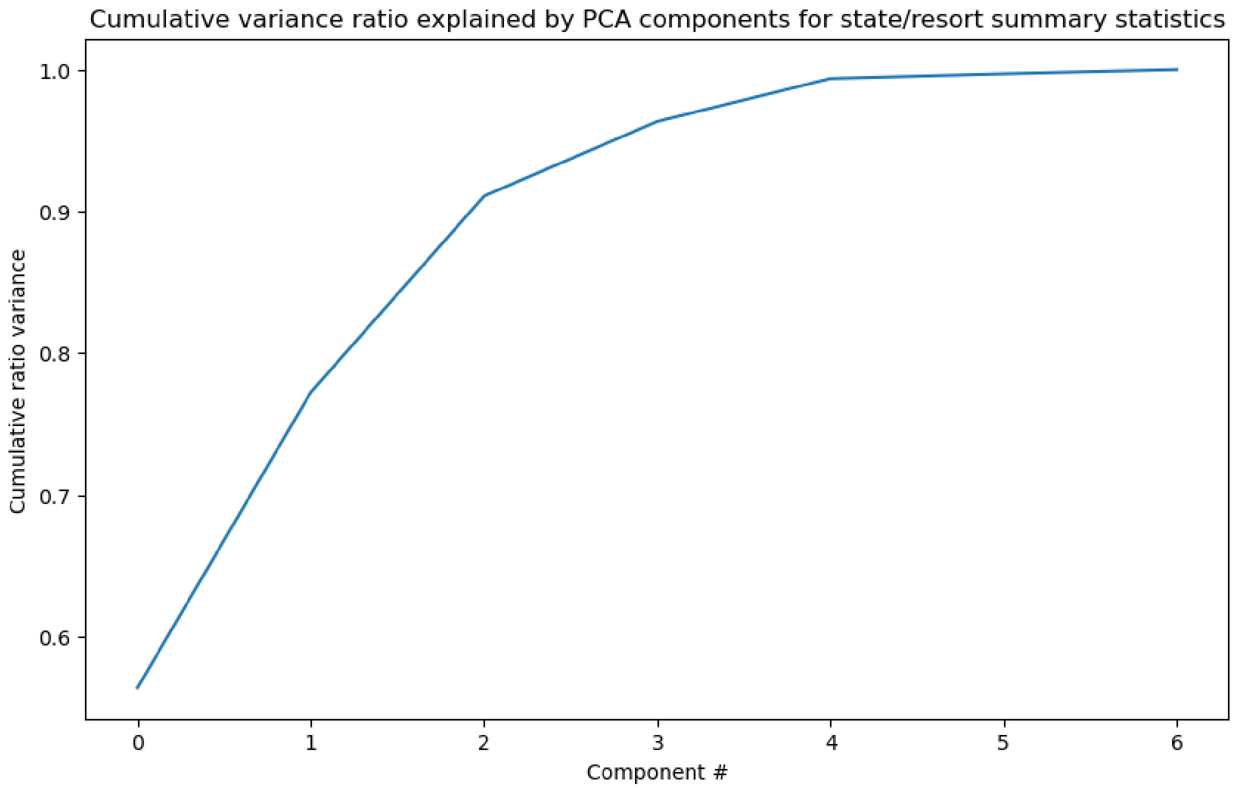
Top States By Resort Per Capita

Vermont	2.403889
Wyoming	1.382268
New Hampshire	1.176721
Montana	1.122778
Idaho	0.671492

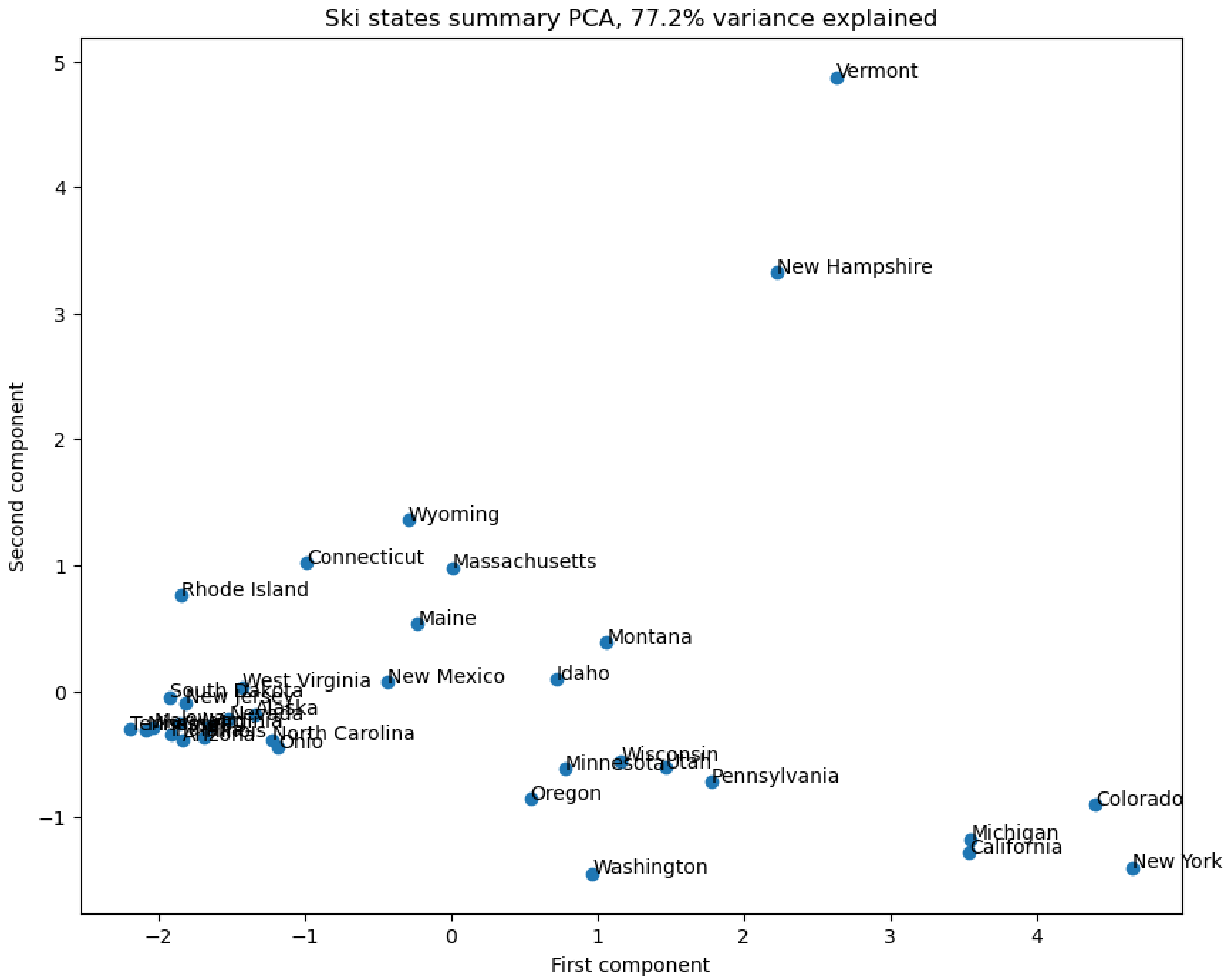
Top State By Resort Per Sq. Mi.

New Hampshire	171.141299
Vermont	155.990017
Massachusetts	104.225886
Connecticut	90.203861
Rhode Island	64.724919

Data exploration revealed a relatively high level of complexity so we can apply Principle Components Analysis to find linear combinations of original features that are uncorrelated with one and other, and order them by the amount of variance they explain. PCA transformation indicates that the first two components account for over 75% of the variance, and the first four with over 95%, as visualized below.

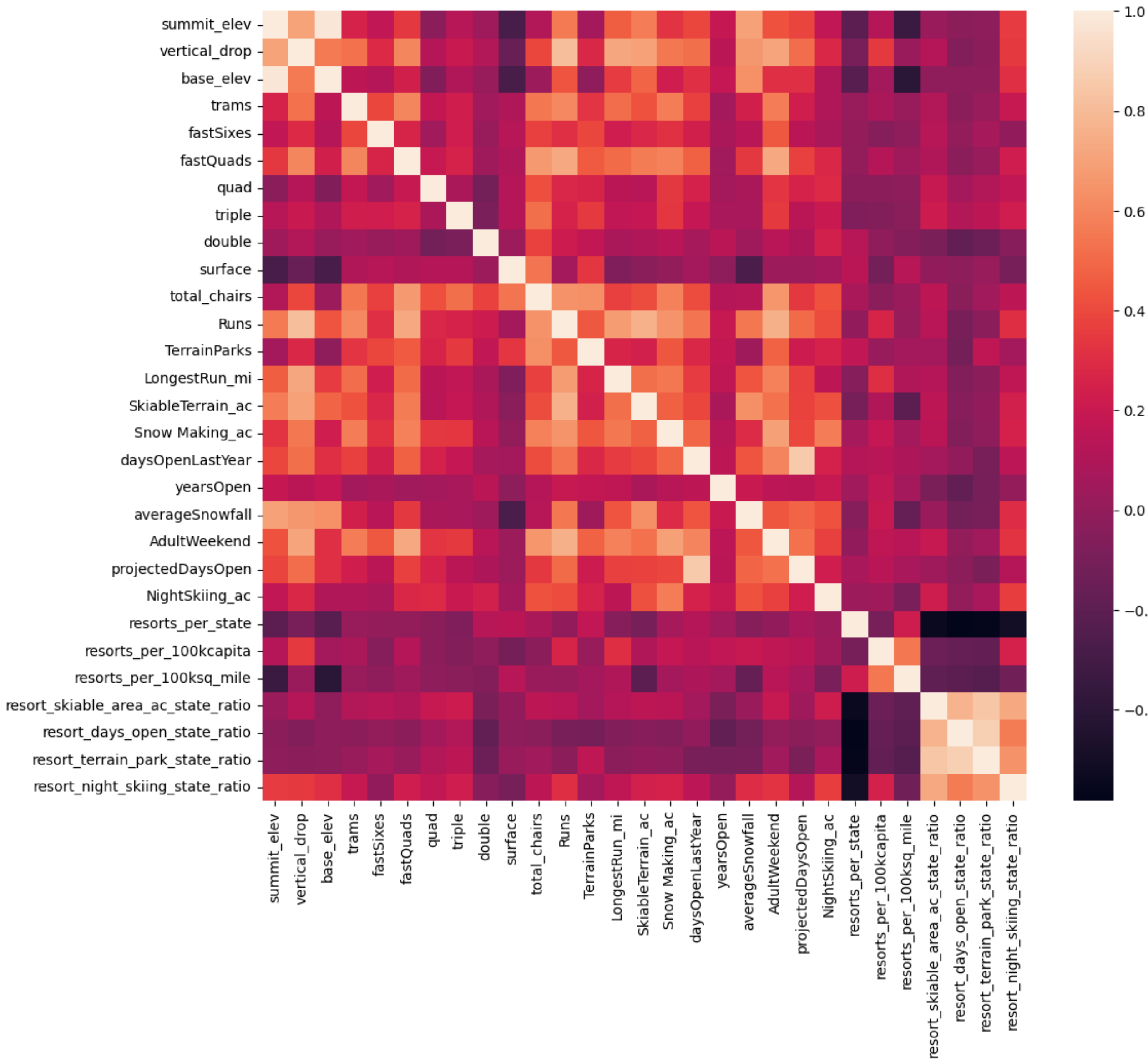


Obtaining derived features also through transformation, we can plot the first two components, as seen below. Concatenating these two components with average ticket price by state revealed no obvious pattern with price.

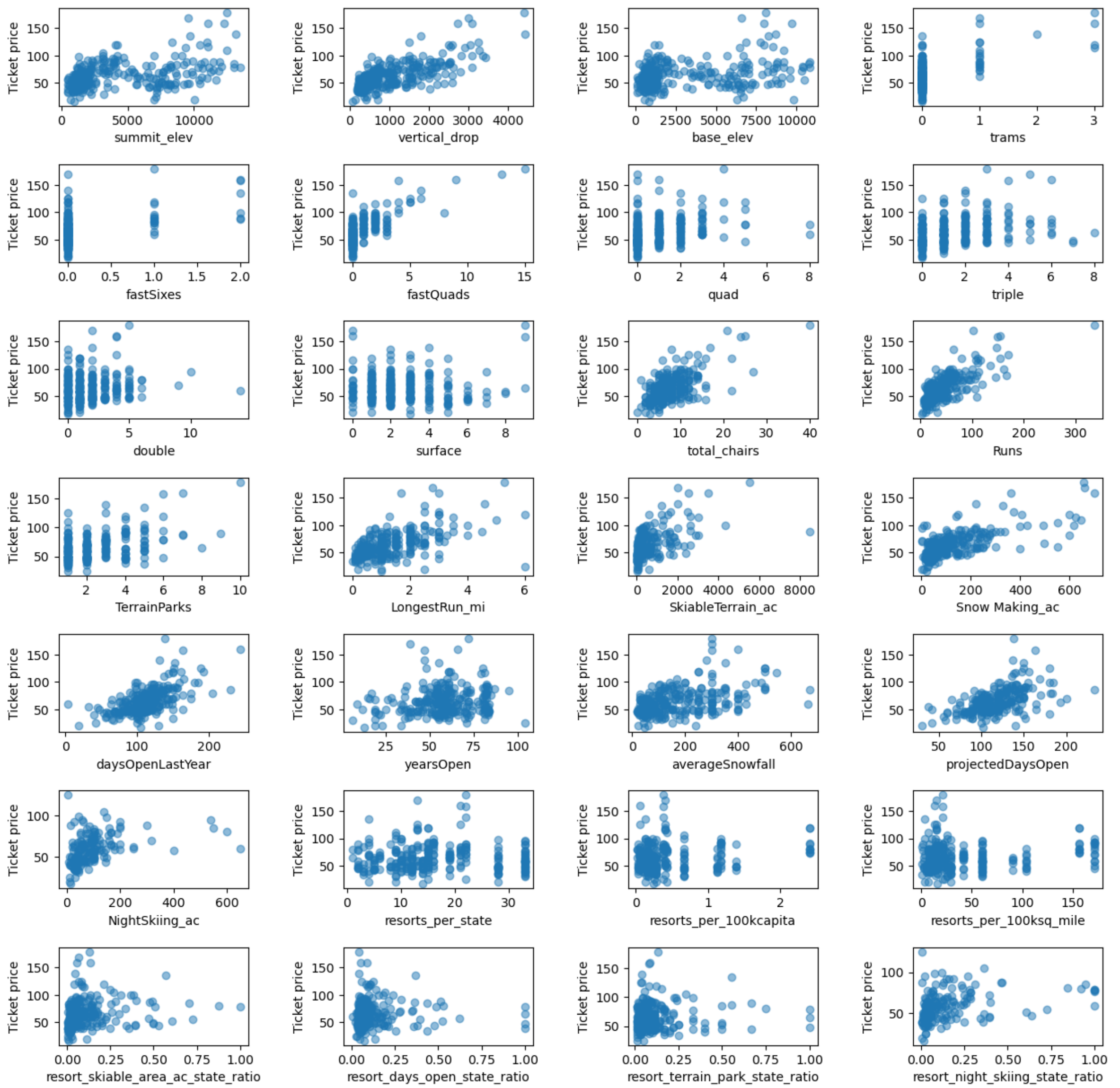


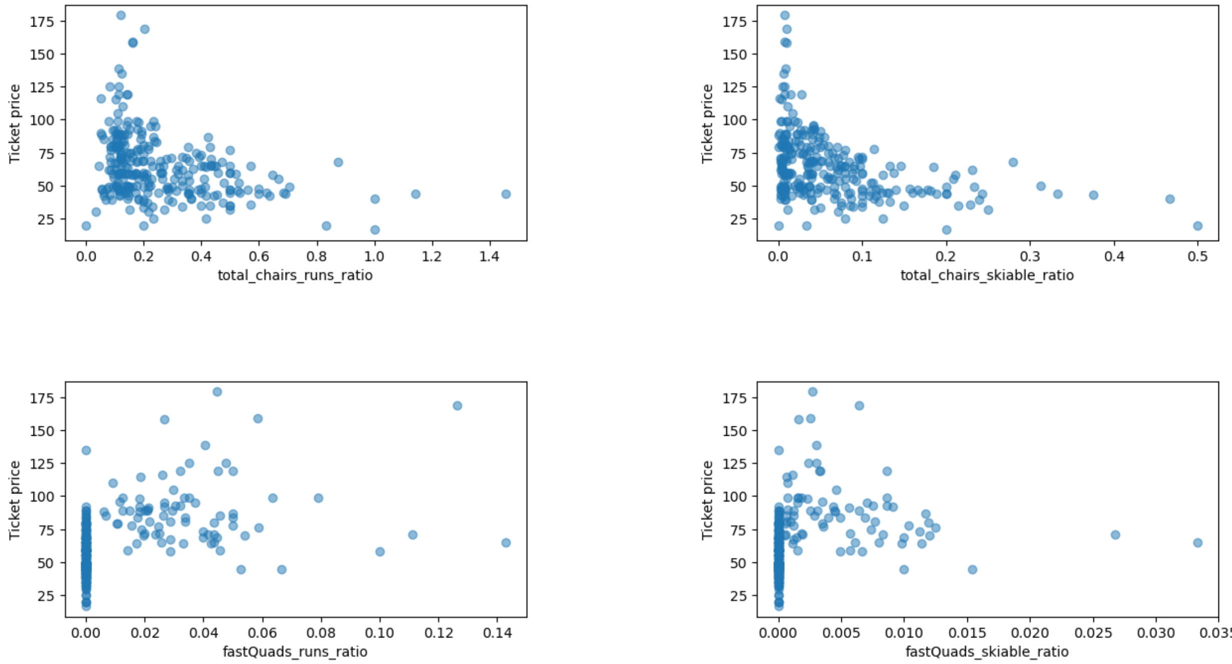
We can see that even with the two market leaders for density, there is no obvious grouping or patterns. As such, we can offer some justification for treating all states equally when building a pricing model that considers all states together.

After engineering additional features to round out our resort competition data, we can visualize the comparison of all features. This shows a reasonable correlation between Weekend ticket price and number of fast quads, runs, total chairs, vertical drop, and snow making acreage, with the ratio of a resort's night skiing to state availability most correlated with ticket price.



Next we can visualize numeric features against ticket price with a scatterplot to gain further understanding of the relationships. This shows a strong positive correlation with vertical drop, fast quads, runs and total chairs. Additional features exploring the relationship between price and number chairs and runs are also visualized below.


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



2.3 Preprocessing + Training

Now that we have a macro and micro understanding of our initial dataset, we can begin to prepare for modeling. Following best practices, we extract data for Big Mountain Resort before partitioning it for a 70/30 train/test split, and then set the model baseline by calculating average price, and regression metrics:

R squared = 0.0 train, -0.00312 test
Mean absolute error = 17.9235 train, 19.1361 test
Mean squared error = 614.1334 train, 581.43654 test
Root mean square error = 24.7817 train, 24.1130 test

In building the initial model, missing values are first imputed with the median because of the skew of many of the predictor feature distributions visualized in the previous section. Calculating regression metrics gives us the following:

R squared = 0.8178 train, 0.7210 test
Mean absolute error = 8.5478 train, 9.4070 test
Mean squared error = 111.8958 train, 161.7316 test

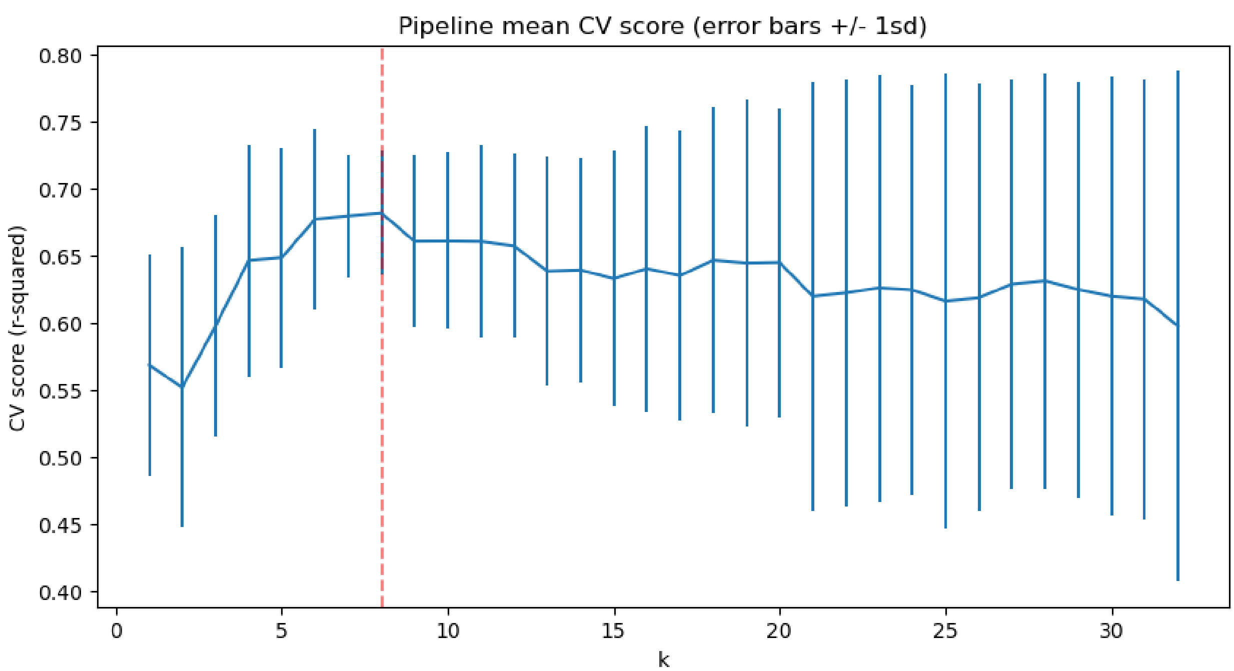
Immediately, we can observe a lower R-squared, suggesting slight overfitting, and a mean absolute error within about $9 of the real price. We can look at what happens when we impute the missing values with the mean and see no significant variation from imputing with the median.

R squared = 0.8170 train, 0.7164 test
Mean absolute error = 8.5369 train, 9.4164 test
Mean squared error = 112.3770 train, 164.3927 test

Next, we built a pipeline with k=10 and calculated regression metrics which indicated selecting a subset of features has a negative impact on model performance. Setting k=15 showed some improvement but it confirmed the need for cross-validation.

The resulting cross-validation mean score of 0.6327 with a standard deviation of 0.0950 highlights that the assessment is inherently open to variability. The estimate of variability calculated at (0.44, 0.82).

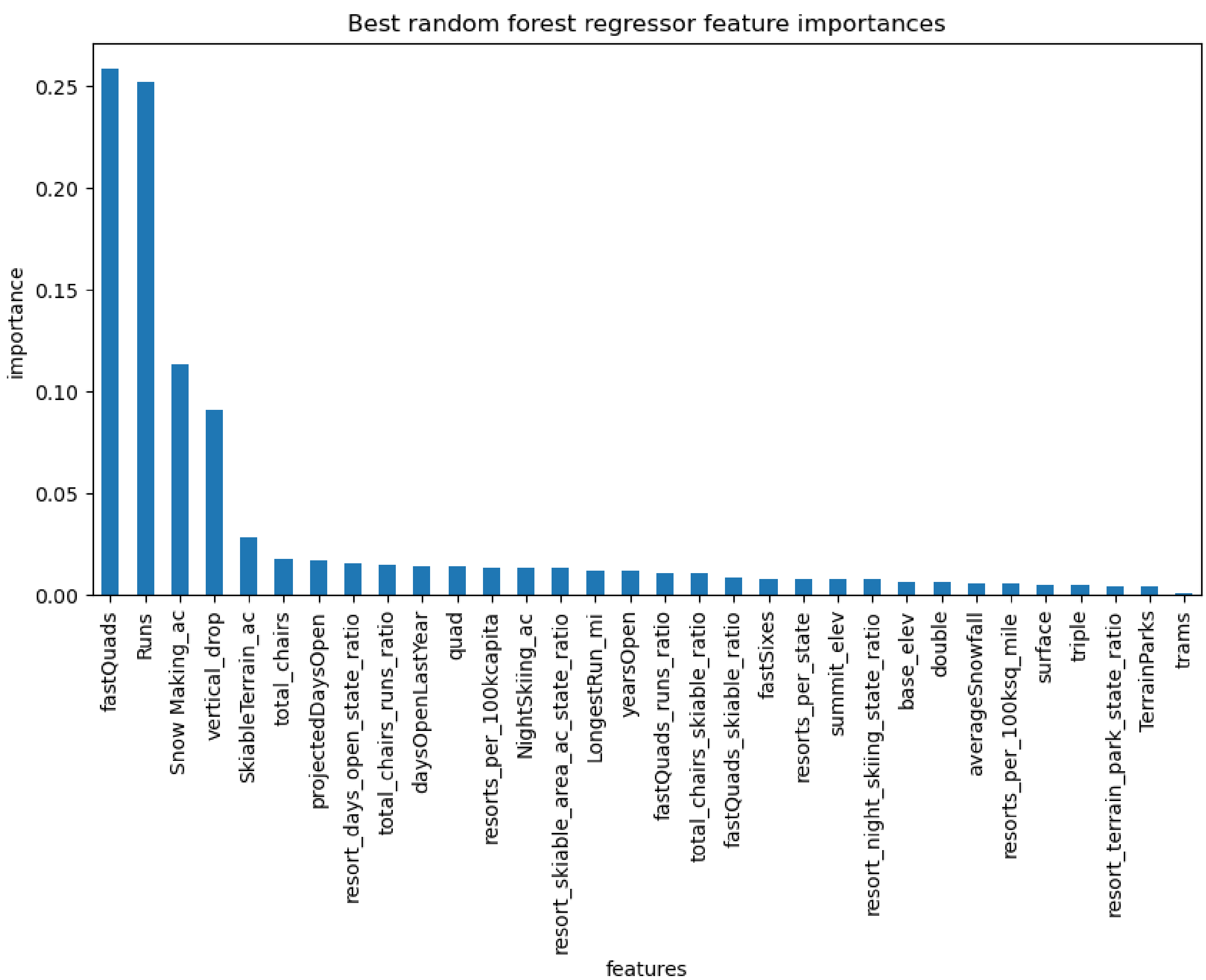
Conducting a hyperparameter search showed a best k of 8 with respective model results indicating that vertical drop is the biggest positive feature, consistent with EDA conclusions. Vertical drop is followed by snow making equipment, total chairs, fast quads and number of runs. Trams and skiable terrain showed negative results for this model.



Features By Importance

vertical_drop	10.767857
Snow Making_ac	6.290074
total_chairs	5.794156
fastQuads	5.745626
Runs	5.370555
LongestRun_mi	0.181814
trams	-4.142024
SkiableTerrain_ac	-5.249780

Following this, a random forest model was assessed using cross-validation as well, showing a mean score of 0.6976 with a standard deviation of 0.0709. Upon further exploration of hyperparameters showed that imputing median does help but scaling the features doesn't. Final scores were 0.7097, 0.0645. This assessment showed that the top four features in common with the linear model are fast quads, number of runs, snow making equipment and vertical drop.

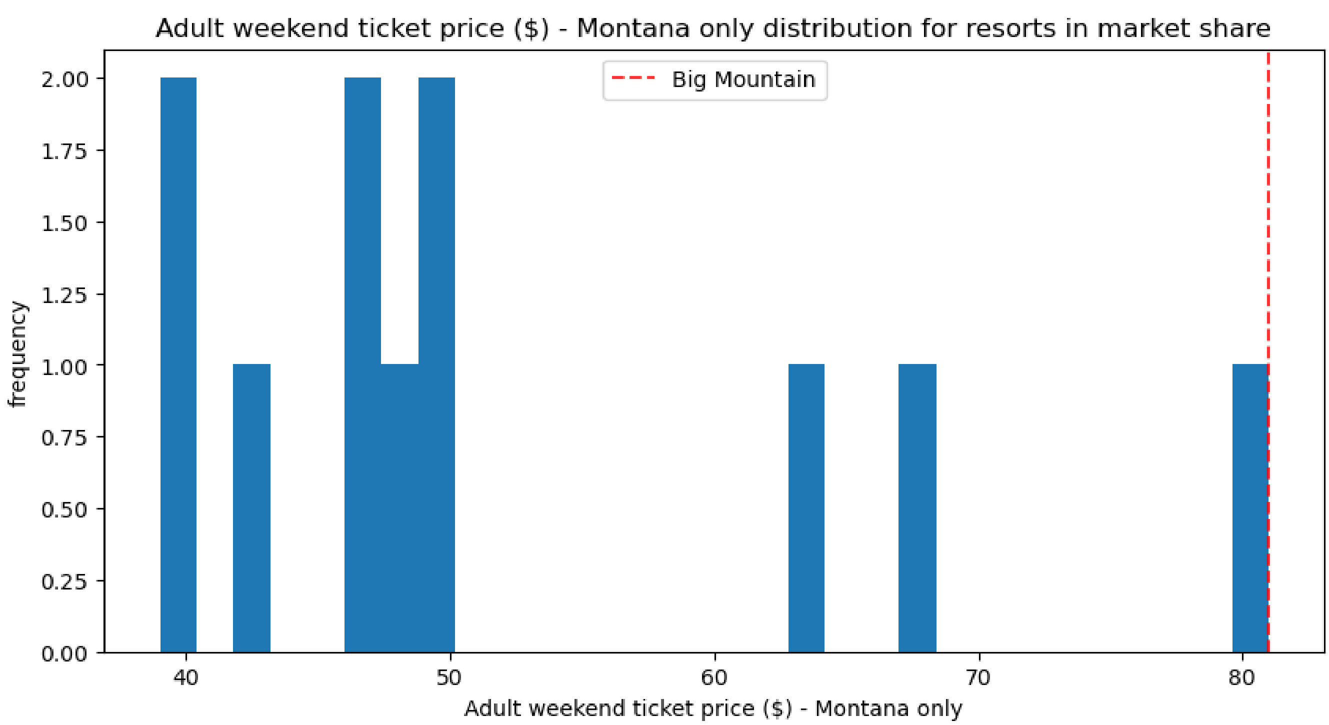


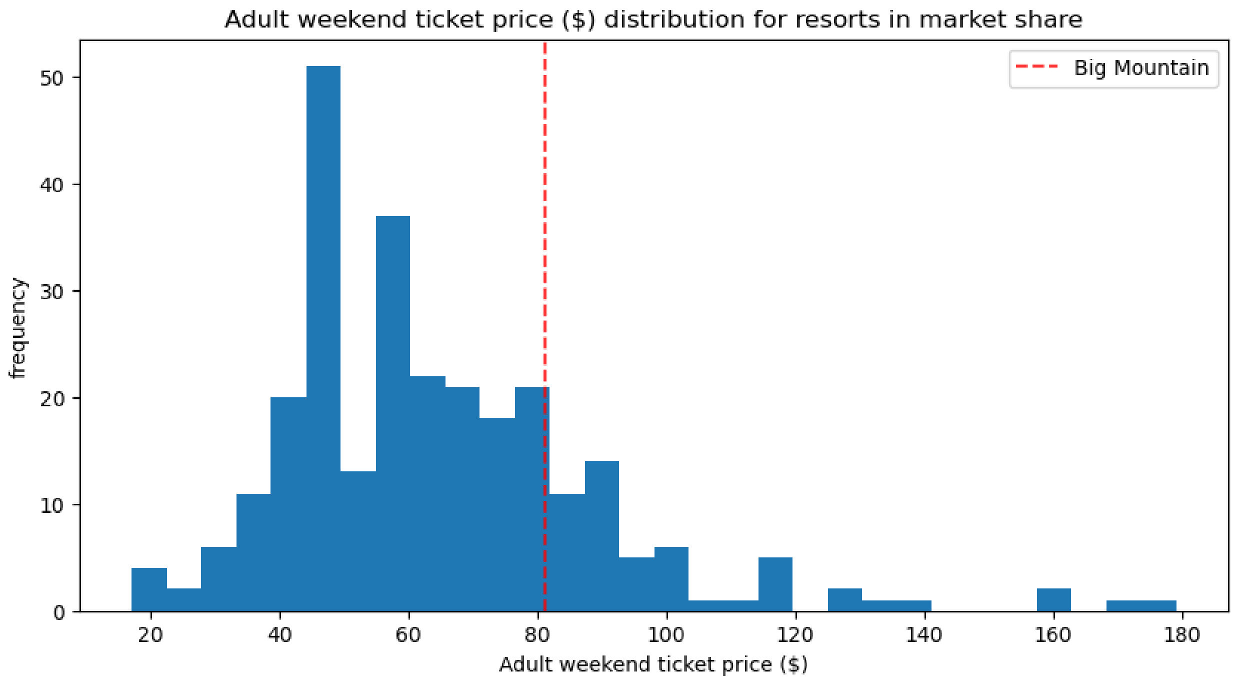
After calculating the mean absolute error for both regression models, the linear model gave us 11.7047 and the random forest model 9.5377. Here, the random forest model has a lower cross-validation mean absolute error by almost $1 while exhibiting less variability. We can verify this performance on the test set which produces performance consistent with cross-validation results.

2.4 Modeling

Big Mountain Resort modeled price is $95.87, compared to actual current price of $81.00. Even with the expected mean absolute error of $10.39, this suggests there is room for an increase. But what about in the context of competitors?

Looking at Montana only, Big Mountain Resort has the highest ticket price with two distant price competitors. From the market-perspective, however, increasing the ticket price is supported in relation to competitors.

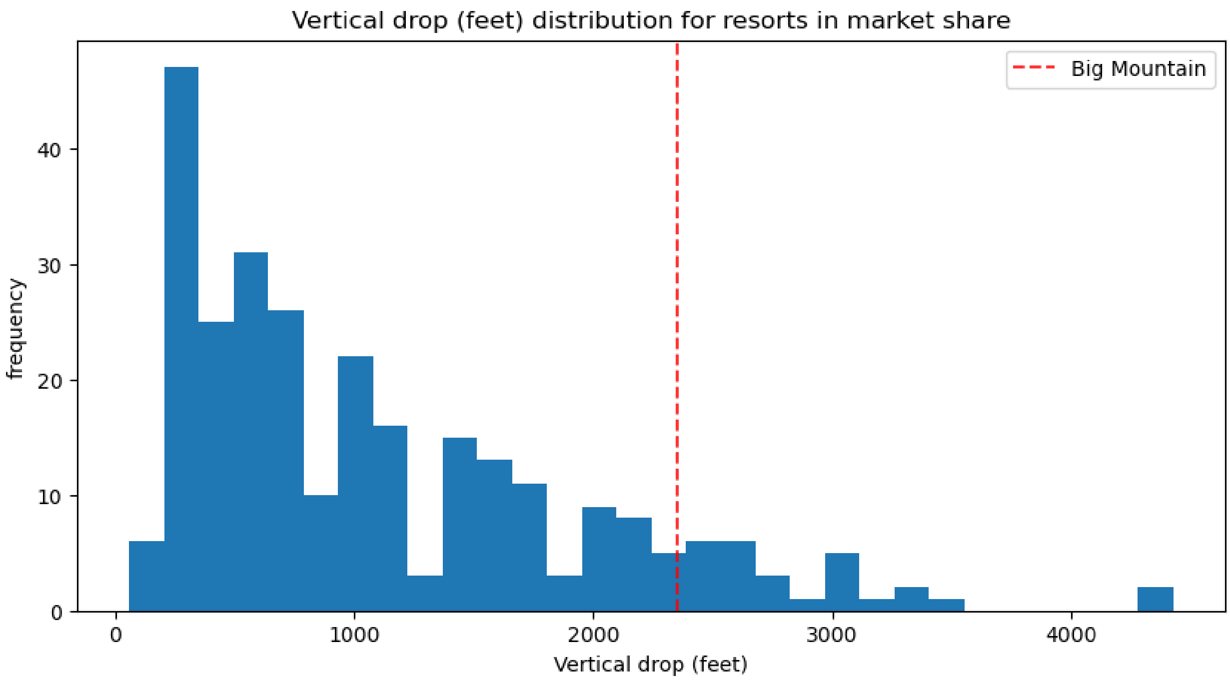


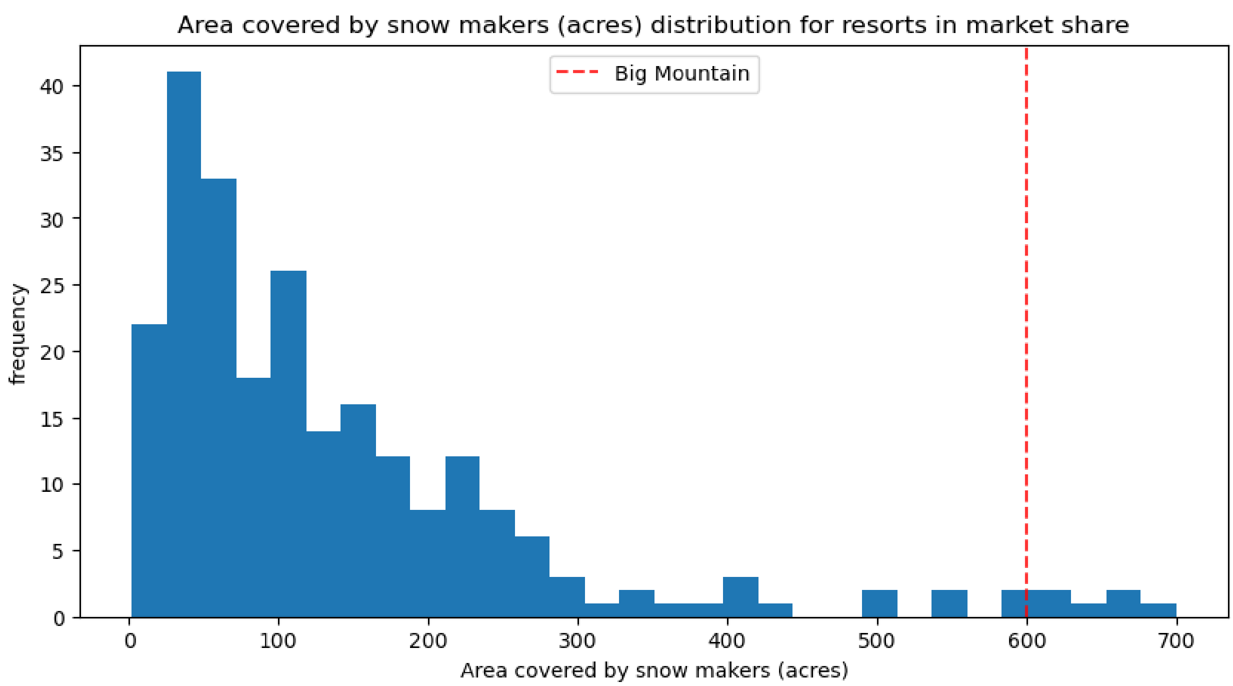


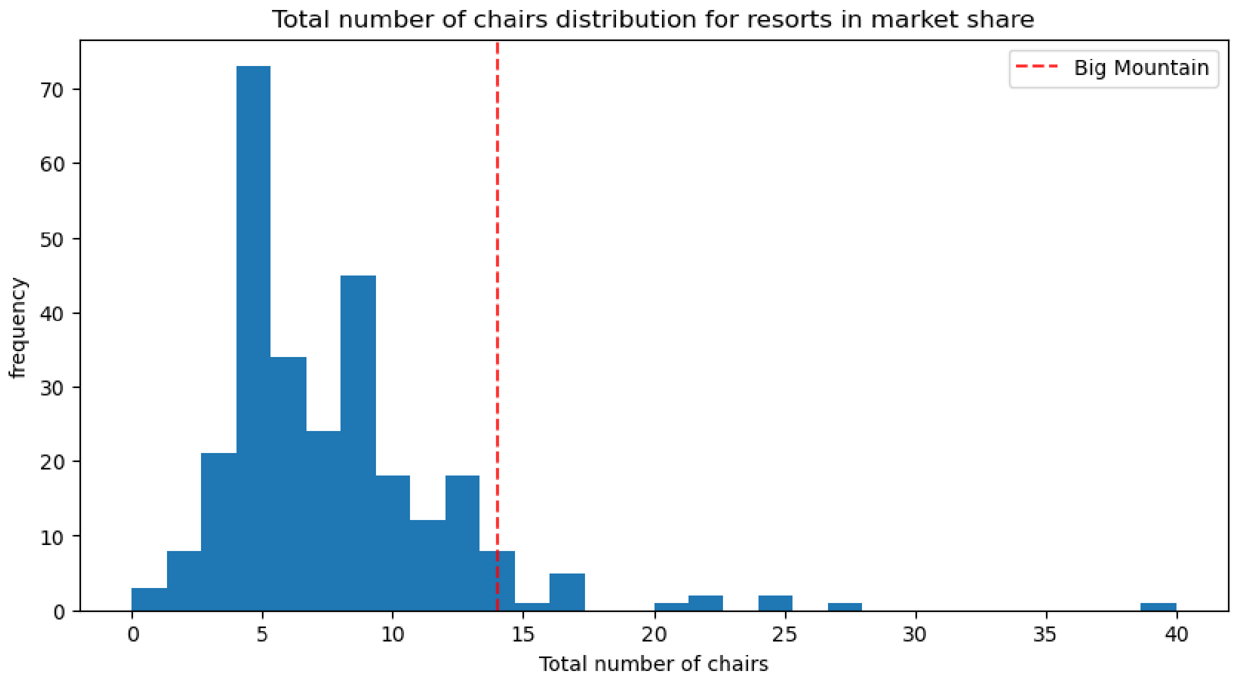
Big Mountain is very high up on the league table of snow making area and has amongst the highest number of total chairs. Resorts with more appear to be outliers, further reinforcing Big Mountain's positioning within the market.

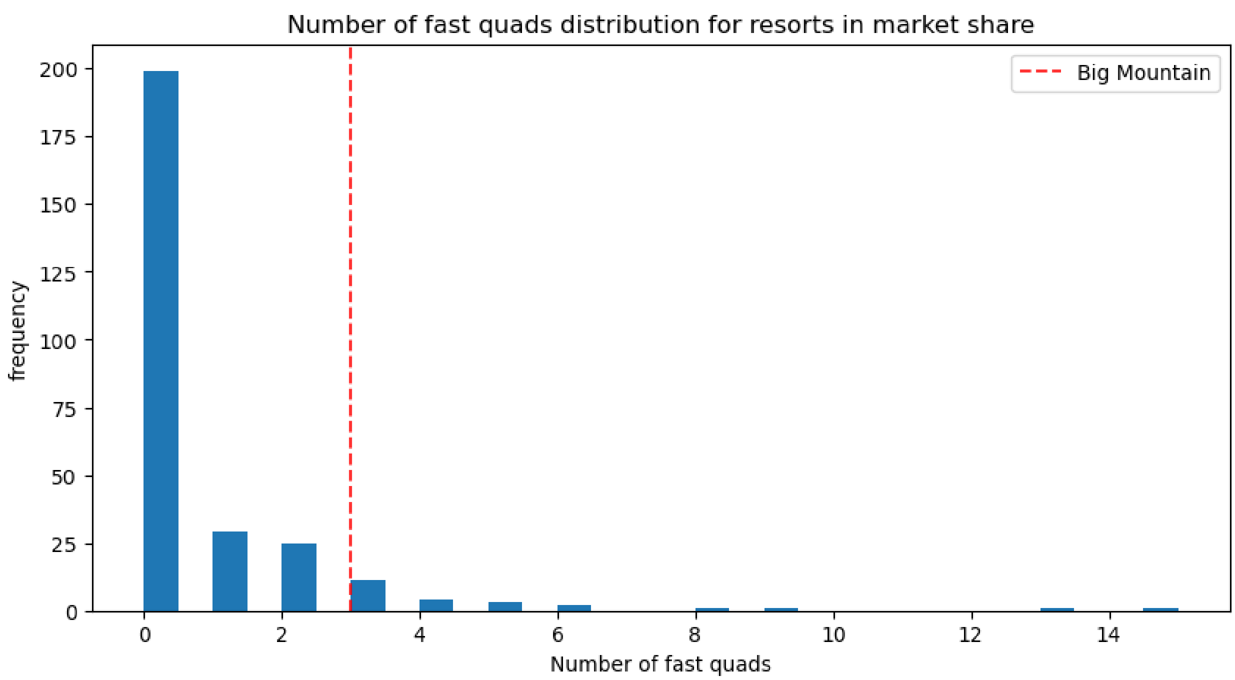
Additionally, most resorts have no fast quads. Big Mountain has 3, which puts it high up on the league table. There are some values much higher, but they are rare. When looking at the number of runs, Big Mountain also compares well. There are some resorts with more, but not many.

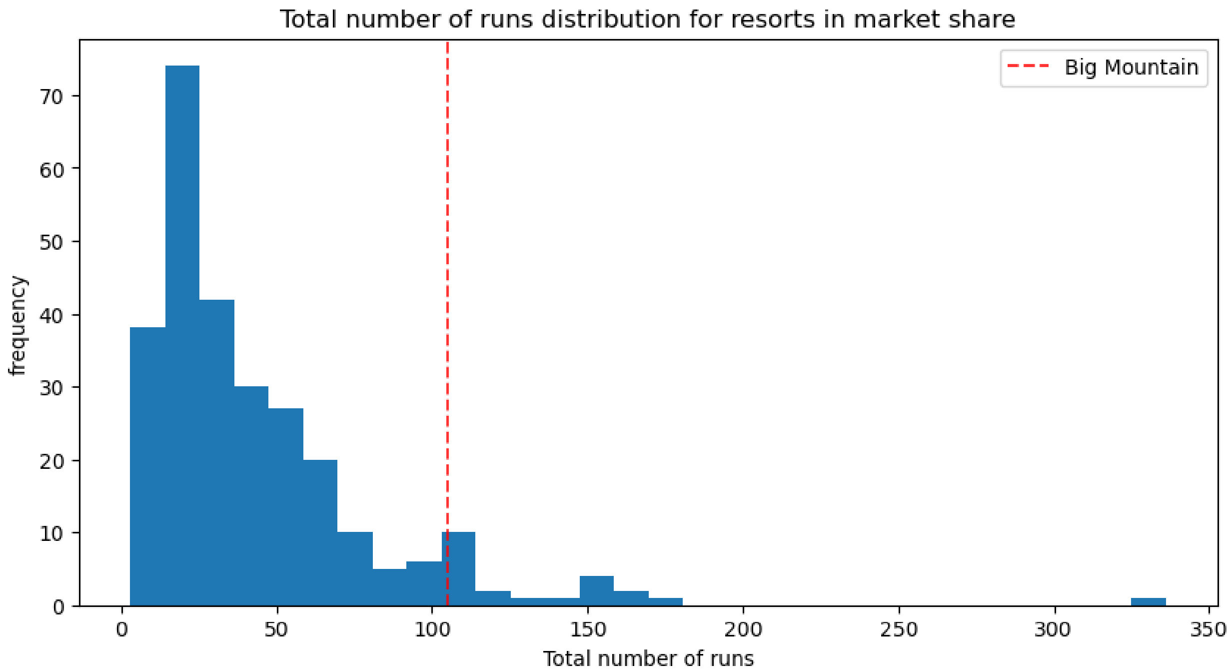
Finally, Big Mountain has one of the longest runs. Although it is just over half the length of the longest run, the longer ones are rare. As we can see, Big Mountain Resort offers a competitive value based on resort features.

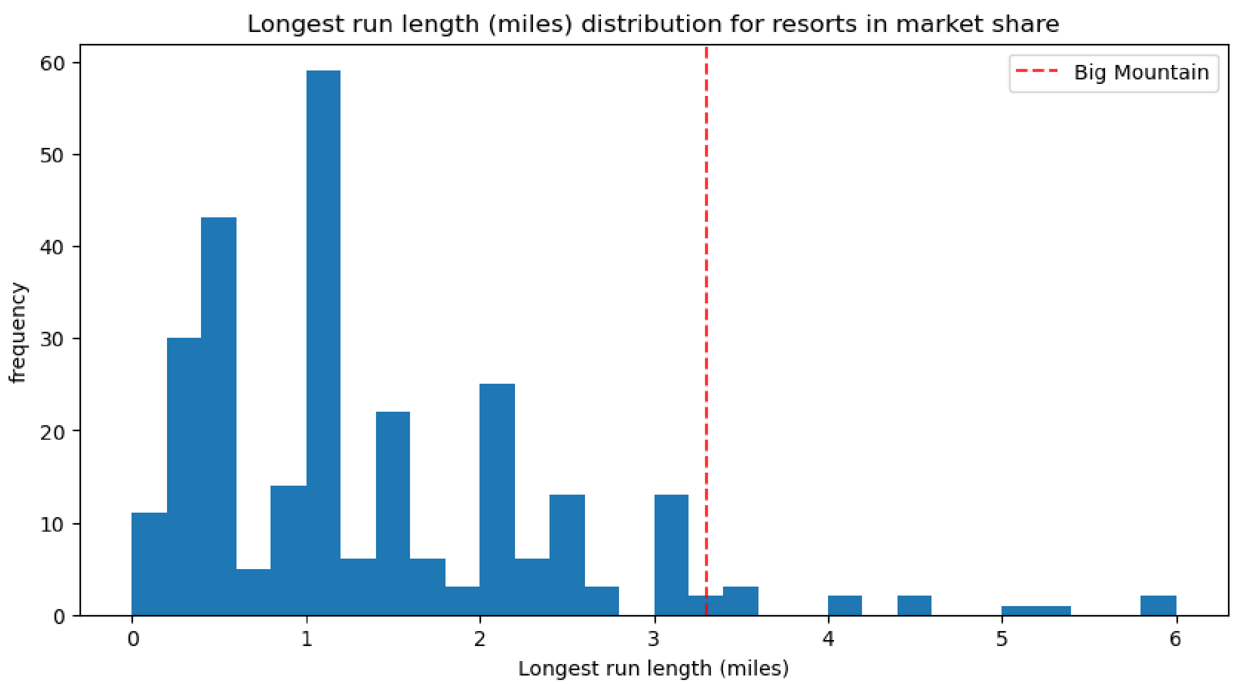

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
0
1000
2000
3000
4000
Vertical drop (feet)

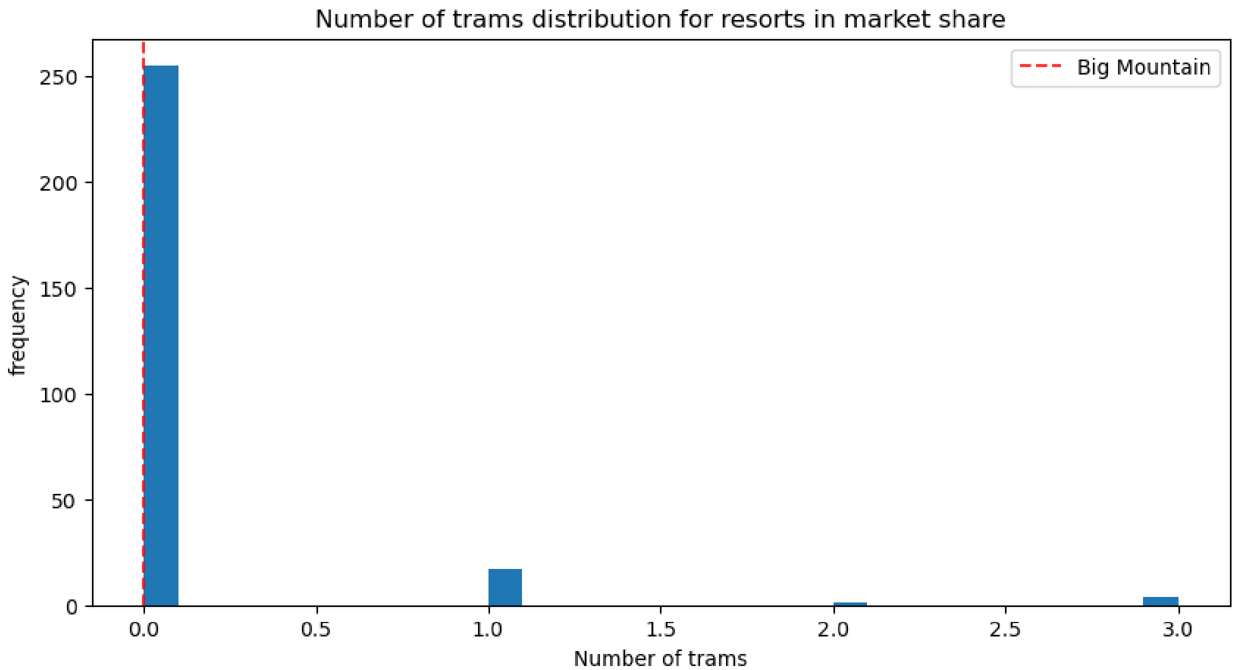

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
0
5
10
15
20
25
30
35
40
0
100
200
300
400
500
600
700
Area covered by snow makers (acres)

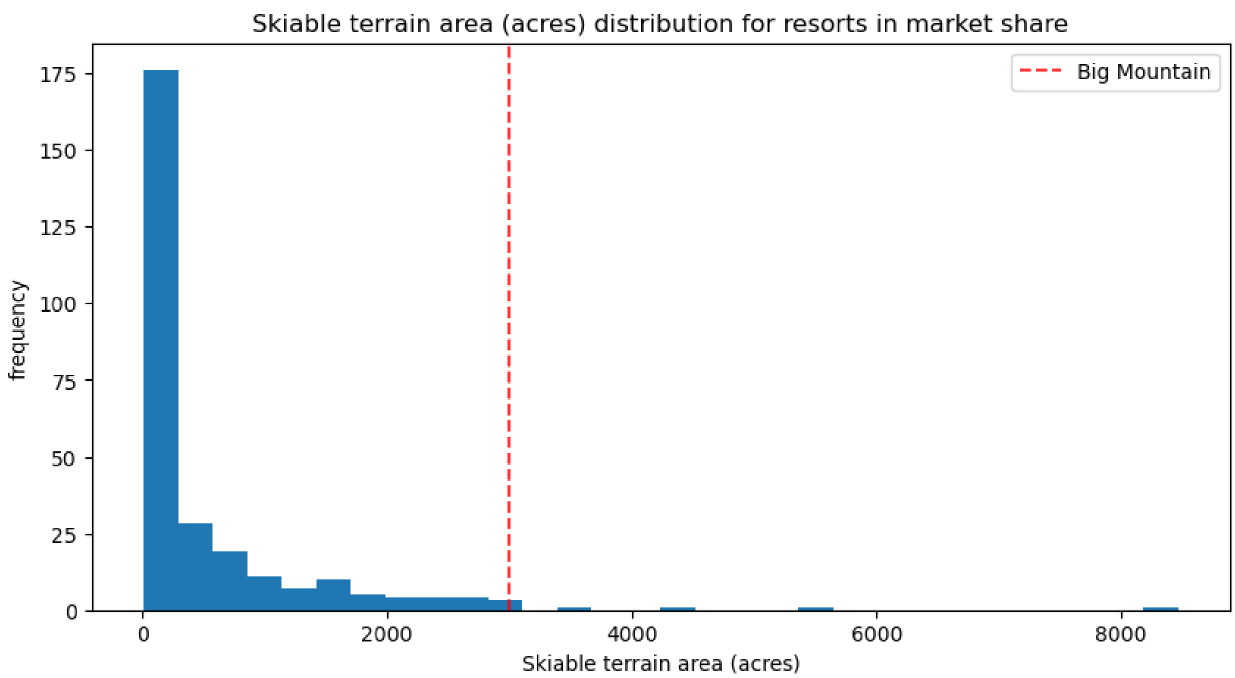

Total number of chairs distribution for resorts in market share
Big Mountain
frequency
Total number of chairs
0
10
20
30
40
50
60
70
0
5
10
15
20
25
30
35
40


Number of fast quads distribution for resorts in market share
Big Mountain
frequency
Number of fast quads
0
25
50
75
100
125
150
175
200
0
2
4
6
8
10
12
14


Total number of runs distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
50
60
70
0
50
100
150
200
250
300
350
Total number of runs


Longest run length (miles) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
50
60
0
1
2
3
4
5
6
Longest run length (miles)


Number of trams distribution for resorts in market share
Big Mountain
frequency
Number of trams
0
50
100
150
200
250
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Skiable terrain area (acres) distribution for resorts in market share
Big Mountain
frequency
Skiable terrain area (acres)
0
25
50
75
100
125
150
175
0
2000
4000
6000
8000

Scenario Modeling + Recommendations

Scenario 1: Permanently close down up to 10 of the least used runs Model Results

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and, subsequently, revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.

If management's business strategy is more focused on cutting expenses to meet the new break-even point, I would seriously reconsider closing anything more than 1. The impact to the ticket price even at only 2 closures is significant. With the total number of runs identified as a high value feature, I would suggest further analysis before making any decisions.

Scenario 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

This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3,474,638

Scenario 3: Same as number 2, but adding 2 acres of snow making cover Model

This scenario increases support for ticket price by $1.00. Over the season, this could be expected to amount to $3,474,638. As we can see, such a small increase in the snow making area makes no difference based on this model when compared to Scenario 2.

Recommendation**:** Looking at the market, Big Mountain is doing well for vertical drop, but there are still quite a few resorts with a greater drop. As Big Mountain is performing well for other features, Scenarios 2 or 3 would lead to increased support for a higher ticket price. This, of course, creates additional revenue to cover the new operating costs, but it also represents an opportunity to further increase competitive advantage.

Scenario 4: Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

Model results show no impact.

2.5 Further Work + Conclusions

There were a few assumptions made throughout this process that could impact accuracy. First was that we assumed the ticket prices listed were not being over- or under-valued. Second, that all equipment shares the same operating cost. It would be helpful to look at both variable operating costs per equipment type as well as fixed and overhead costs.

Executives already suspected that Big Mountain Resort was undervaluing its ticket price so while the size of the difference might be surprising, overall modeling results are in line with expectations. It is possible that because of the density ratios first explored in the EDA phase of this project, we might be valued correctly but when looking at the market as a whole, Big Mountain Resort is quite competitive in the features it provides.