Guided Capstone Project Report:

Big Mountain Resort

INTRODUCTION

1.1 Background

Big Mountain Resort has installed a new chair lift to increase the distribution of riders across the mountain. This lift has led to an increase of $1.54 million in operating expenses this season, spurring a review of existing pricing strategies.

Current ticket price is based on the market average and does not provide the business with a good sense regarding value offered via facilities compared to others in the market. Management is open to changes as long as cutting costs does not devalue ticket or, alternatively, any changes made are to support a higher ticket value.

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 before the season begins 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?

Scope 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 Model Results

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 Model

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

While management is open to cutting costs, data is not 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.

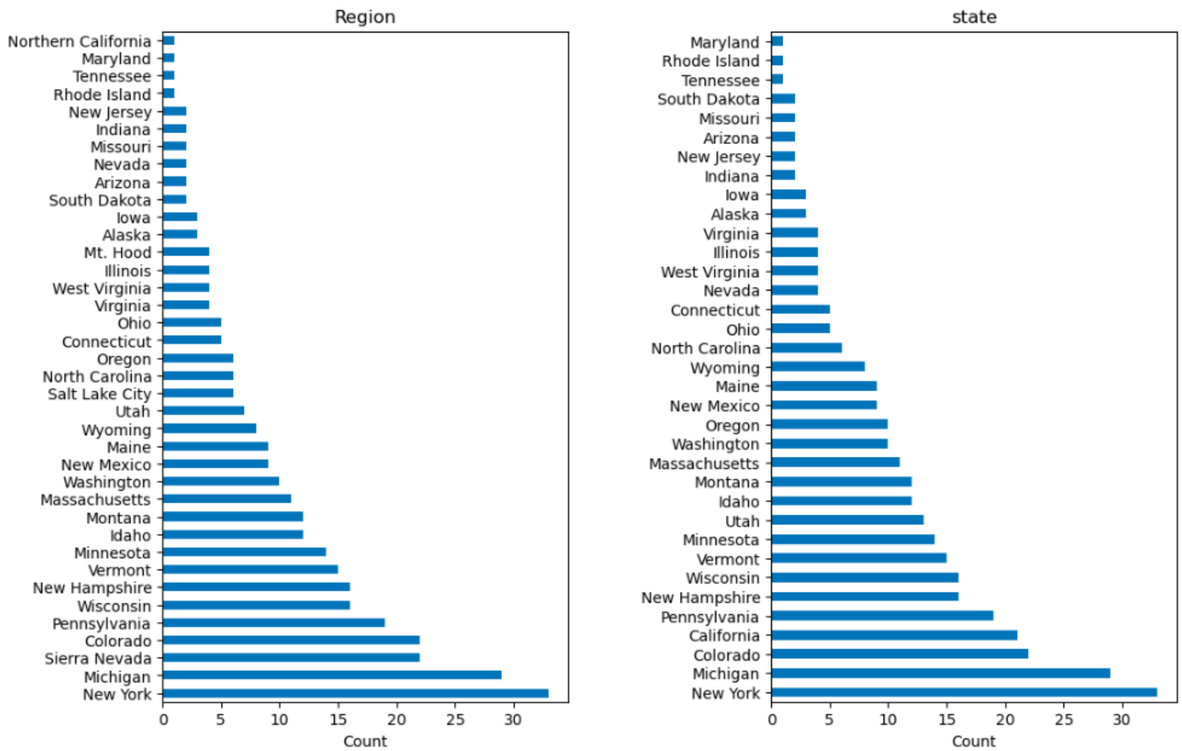
1.4 Data Source

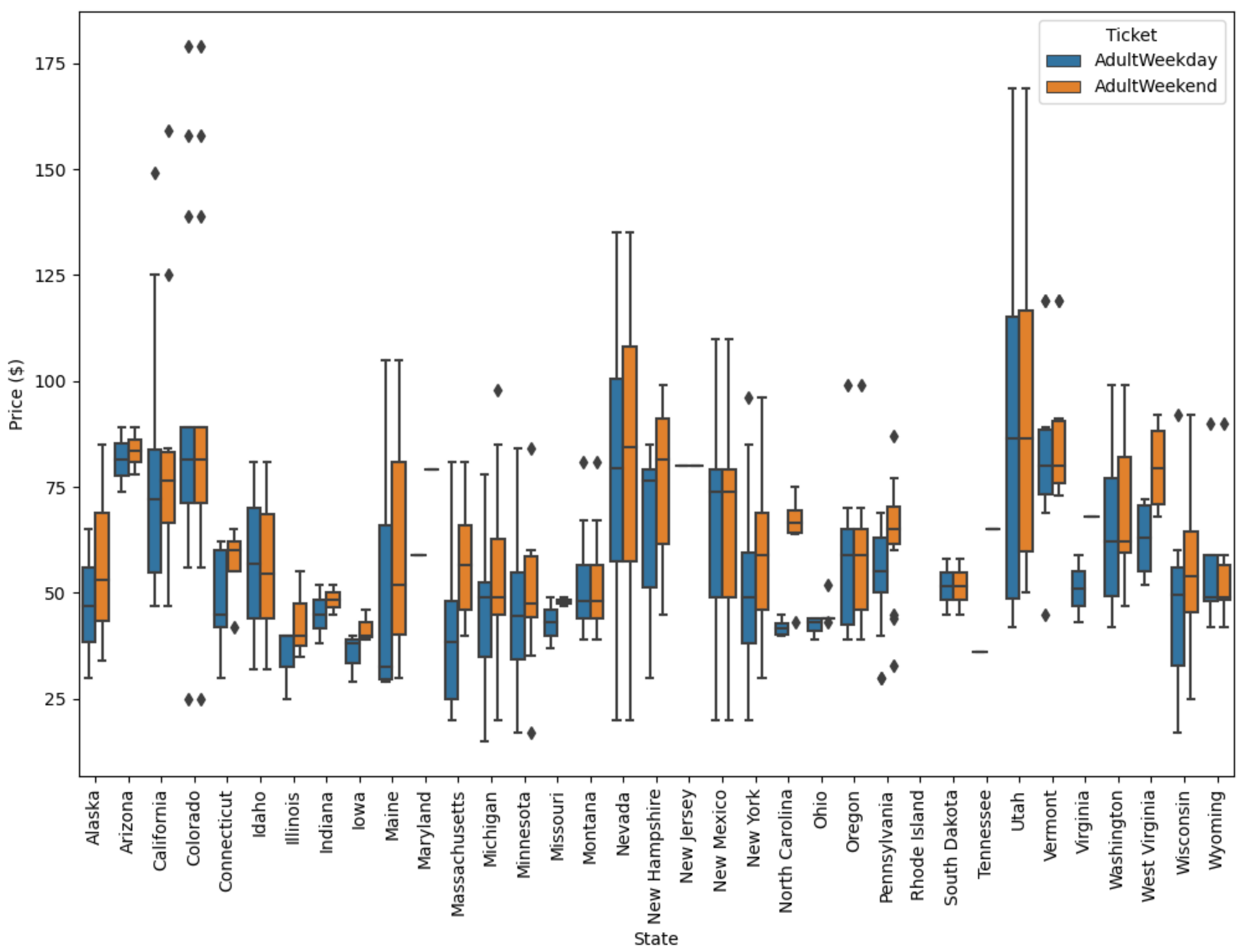
CSV File containing data for 330 resorts in the US within Big Mountain Resort's market segment.

APPLYING DSC PROCESS

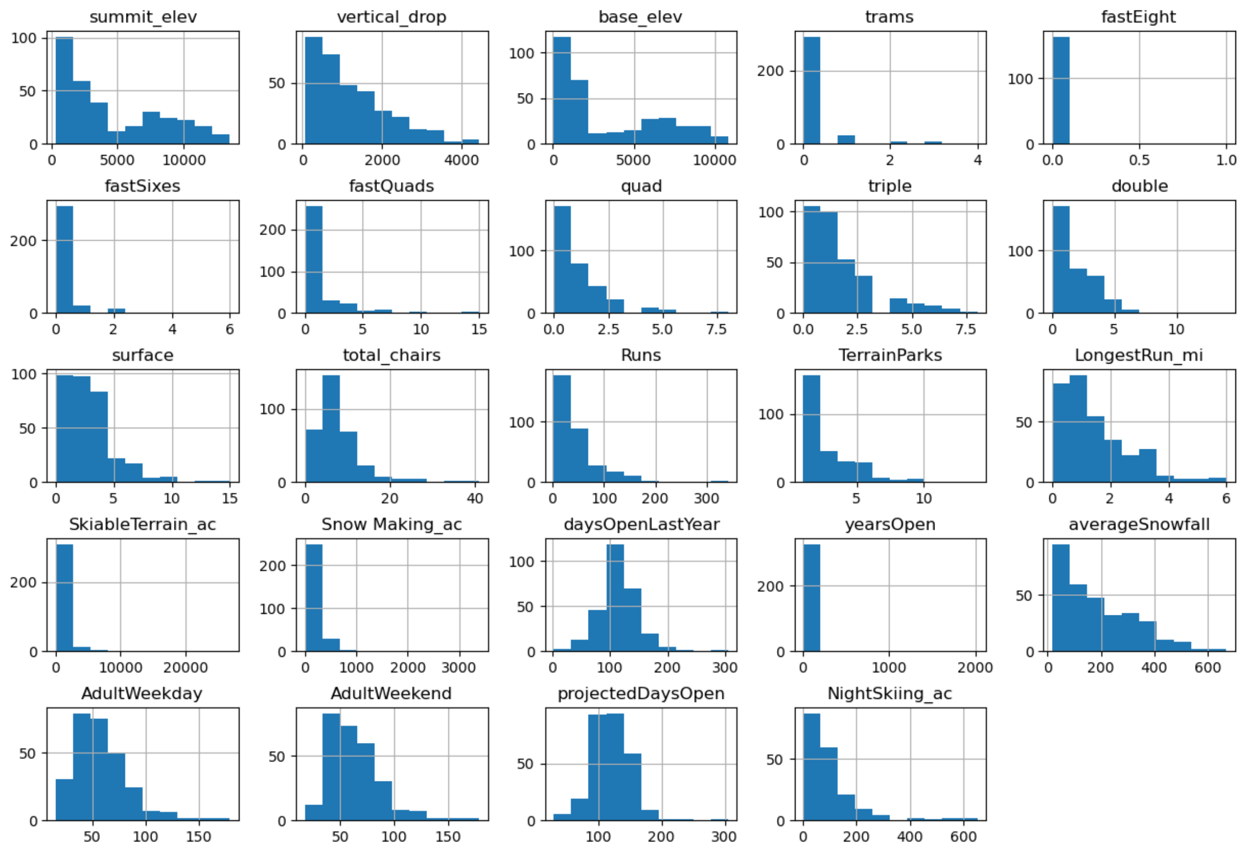
2.1 Data Wrangling

Explored raw data for 330 unique Resorts, confirming the inclusion of Big Mountain Resort and performing an initial check for missing values by column. Visualized distribution of resorts by region and state as well as average ticket price by state for weekday and weekend prices (below):

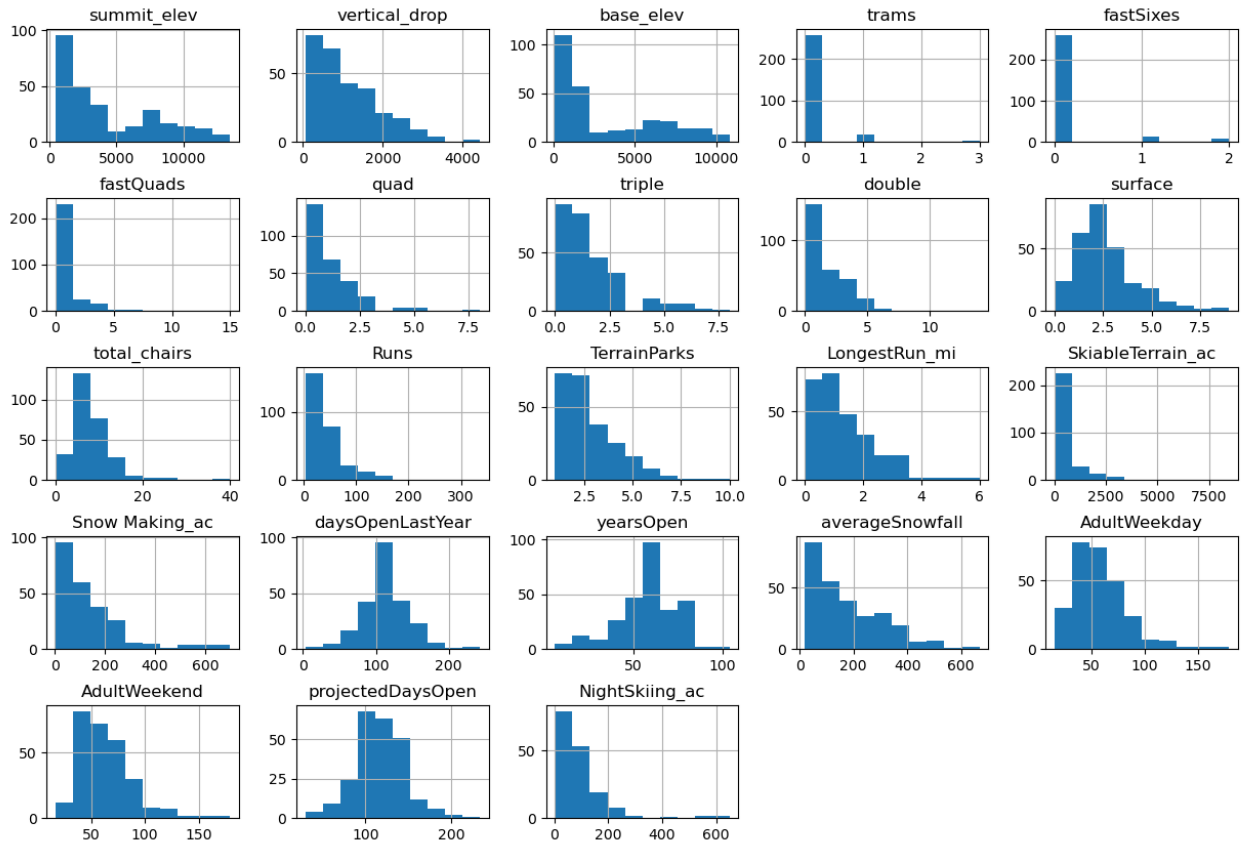

Region
Northern California
Maryland
Tennessee
Rhode Island
New Jersey
Indiana
Missouri
Nevada
Arizona
South Dakota
Iowa
Alaska
Mt. Hood
Illinois
West Virginia
Virginia
Ohio
Connecticut
Oregon
North Carolina
Salt Lake City
Utah
Wyoming
Maine
New Mexico
Washington
Massachusetts
Montana
Idaho
Minnesota
Vermont
New Hampshire
Wisconsin
Pennsylvania
Colorado
Sierra Nevada
Michigan
New York
0 5 10 15 20 25 30
Count
state
Maryland
Rhode Island
Tennessee
South Dakota
Missouri
Arizona
New Jersey
Indiana
Iowa
Alaska
Virginia
Illinois
West Virginia
Nevada
Connecticut
Ohio
North Carolina
Wyoming
Maine
New Mexico
Oregon
Washington
Massachusetts
Montana
Idaho
Utah
Minnesota
Vermont
Wisconsin
New Hampshire
Pennsylvania
California
Colorado
Michigan
New York
0 5 10 15 20 25 30
Count


Ticket
AdultWeekday
AdultWeekend
Price ($)
175
150
125
100
75
50
25
Alaska
Arizona
California
Colorado
Connecticut
Idaho
Illinois
Indiana
Iowa
Maine
Maryland
Massachusetts
Michigan
Minnesota
Missouri
Montana
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Ohio
Oregon
Pennsylvania
Rhode Island
South Dakota
Tennessee
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State

Discovered missing price information for about 16% of resorts with 14% missing both values for Weekday and Weekend prices. Distributions of feature values explored to verify plausibility of data, revealing questionable outliers and significant missing values (below). Appropriate rows of resorts dropped or excluded as well as FastEight feature column dropped.



Derived state summary statistics for market segment for Terrain Parks, Skiable Terrain, days open, and night skiing availability, dropping rows with no price data.



Imported population data by state, checking for missing states. Began exploring the relationship of target features: Weekday and Weekend prices. Weekend prices showed the least number of missing values for the target state, thus dropped weekday prices. Final shape of data shows 277 rows.

2.2 Exploratory Data Analysis

Looked at the national picture for market by state, examining total state area, total state population, resorts per state, total skiable area, total night skiing area and total days open. Top states for each listed below.

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

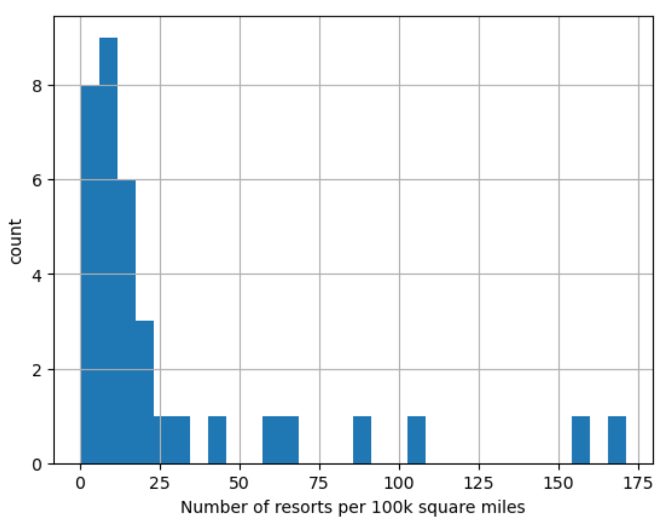
Total Night Skiing Area

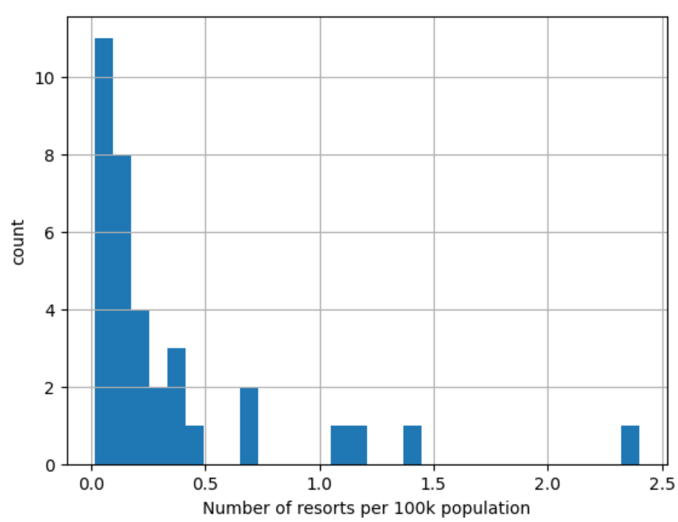
New York	2836.0
Washington	1997.0
Michigan	1946.0
Pennsylvania	1528.0
Oregon	1127.0

Total Days Open

Colorado	3258.0
California	2738.0
Michigan	2389.0
New York	2384.0
New Hampshire	1847.0

Then explored resort density per state, requiring the removal of some state-specific data. Looking at the distributions of resorts per 100k capita and resorts per 100k square miles, available below, showed outliers while further validating the data. Top states for each summarized below as well.



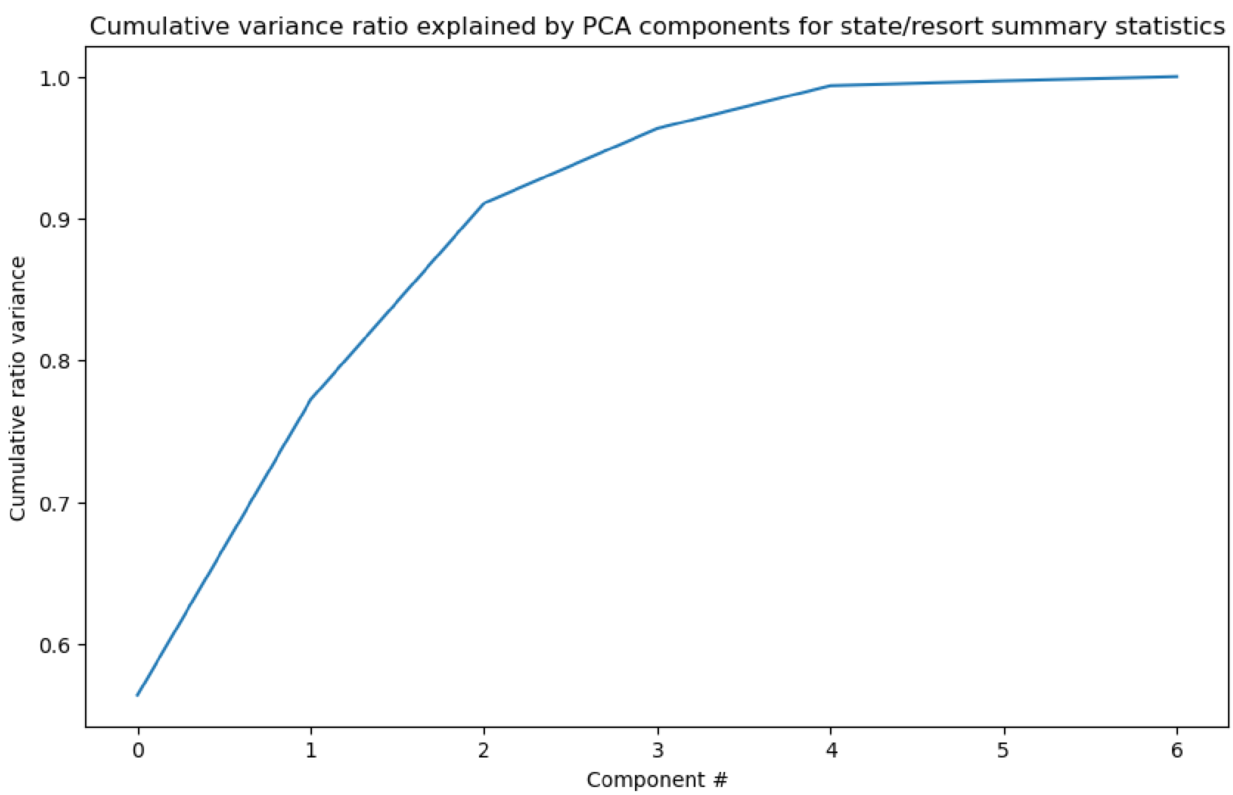


Top States By Resort Per Capita	
Vermont	2.403889
Wyoming	1.382268
New Hampshire	1.176721
Montana	1.122778
Idaho	0.671492

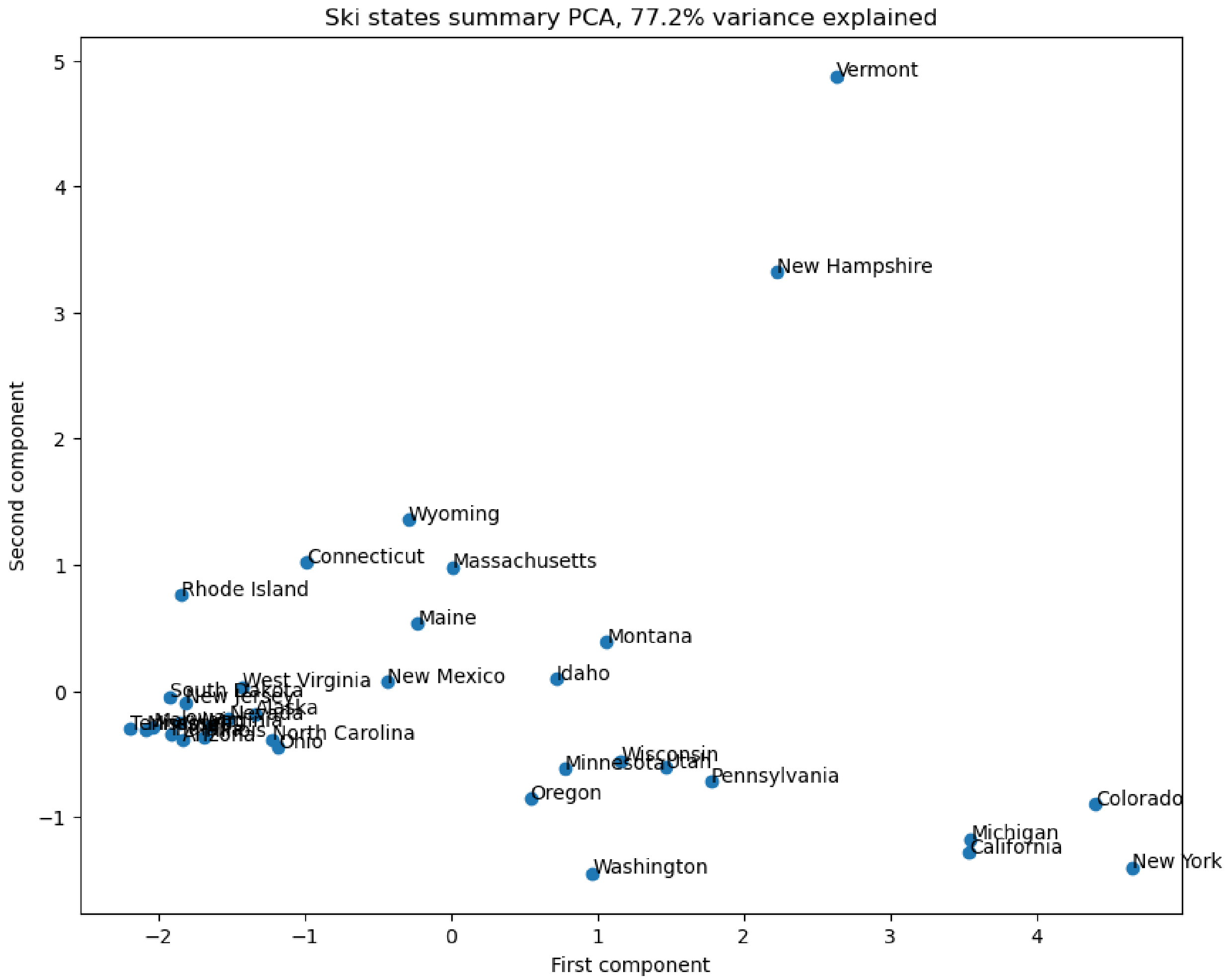
Top State By Resort Per Sq. Mi.	
New Hampshire	171.141299
Vermont	155.990017
Massachusetts	104.225886
Connecticut	90.203861
Rhode Island	64.724919

Data showed a relatively high level of complexity so Principle Components Analysis was applied to find linear combinations of original features that are uncorrelated with one and other, and order them by the amount of variance they explain.

Data was scaled and verified, looking at mean and standard deviation, adjusting for biased estimate. PCA transformation was calculated: the first two components accounting for over 75% of the variance, and the first four with over 95%.



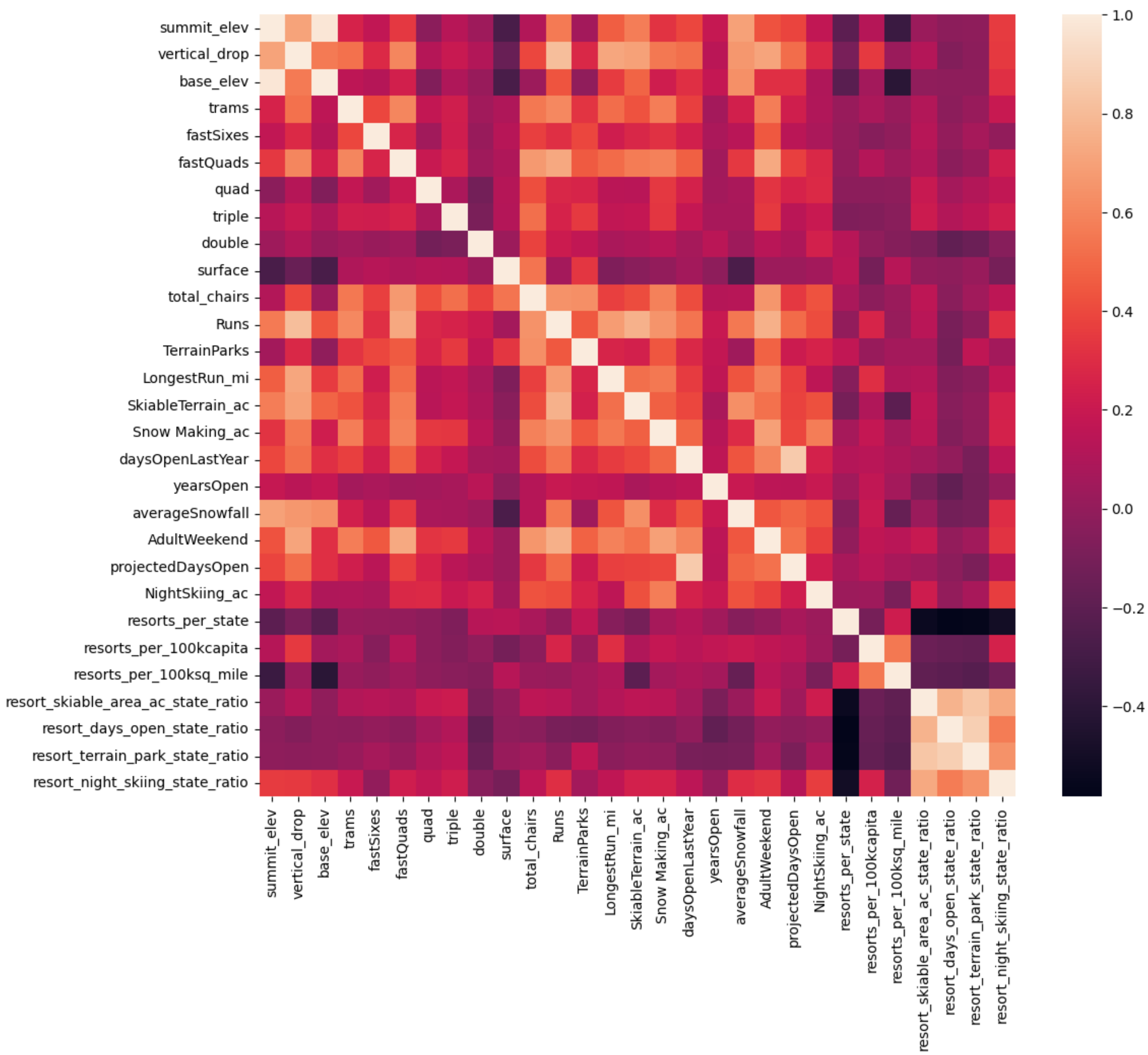
Transformation as applied to obtain derived features, plotting the first two.



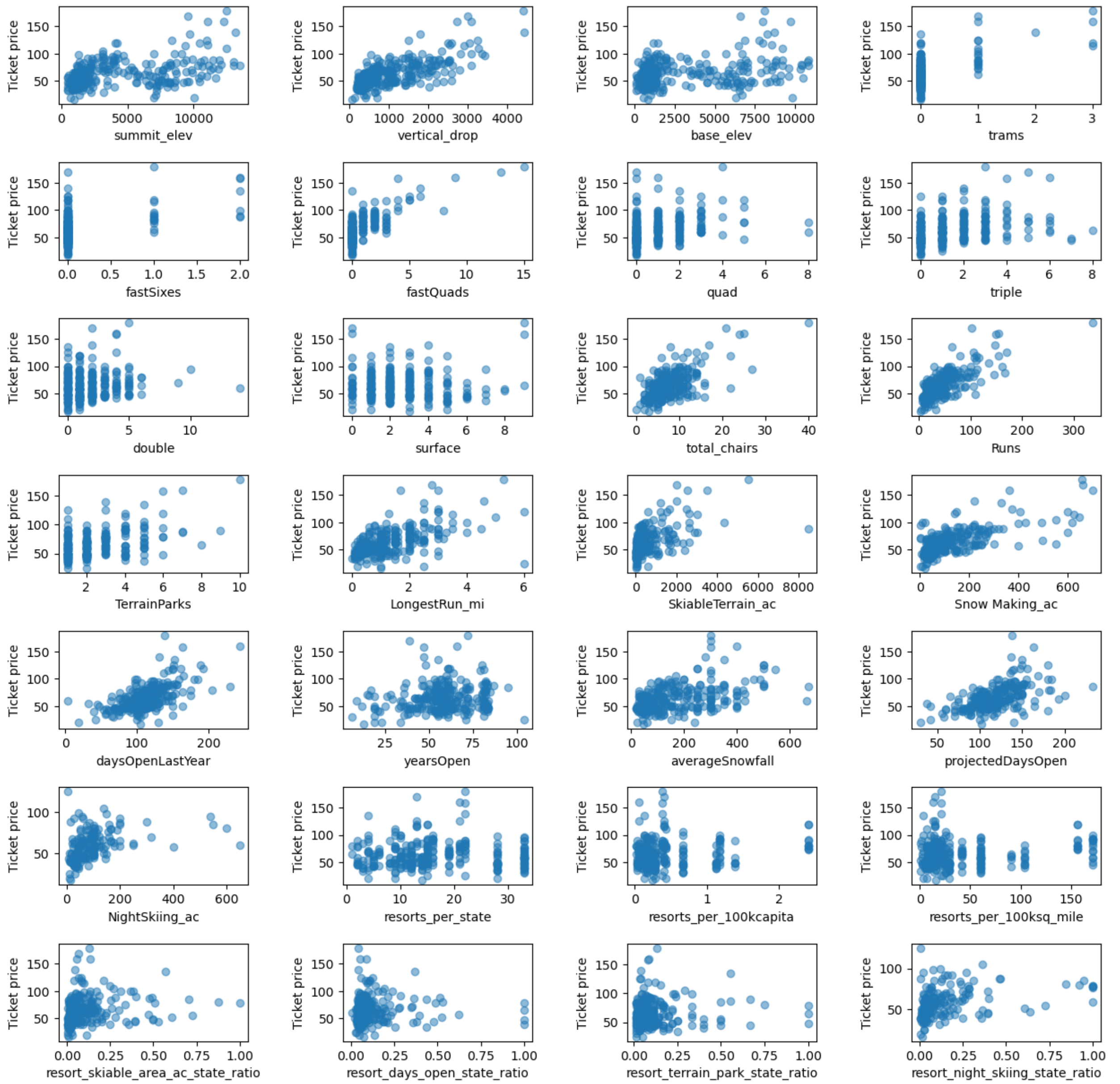
Concatenating these two components with average ticket price by state, plotting again with marker size for ticket price and color for discrete quartile, revealed no obvious pattern with price.

Even with two market leaders for density, there was no obvious grouping or patterns so we can offer some justification for treating all states equally when building a pricing model that considers all states together.

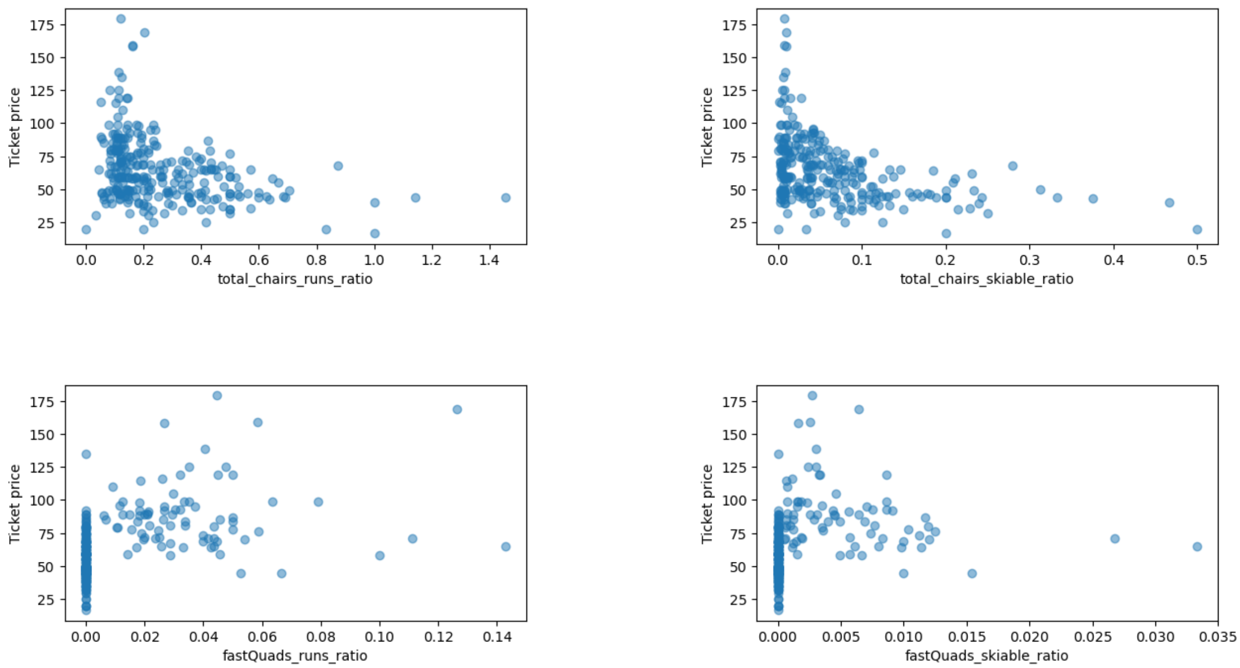
Engineered additional features for resort competition data. Visualizing comparison of all features show a reasonable correlation between AdultWeekend ticket price and number of fast quads, runs, total chairs, vertical drop, and snow making acreage, with the ratio of a resort's night skiing to state availability most correlated with ticket price.



Numeric features against ticket price visualized with a scatterplot to gain further understanding of relationships showing a strong positive correlation with vertical drop, fast quads, runs and total chairs.



Resorts per capita revealed interesting pricing behavior. Additional features exploring relationship between price and number chairs and runs visualized.



2.3 Preprocessing + Training

After extracting data for Big Mountain Resort, partitioned data for a 70/30 train/test split, dropping name, state, and region columns.

Set 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

Built initial model, imputing missing values with median, and calculated regression metrics. Started with medians because of the skew of many of the predictor feature distributions.

R squared = 0.8178 train, 0.7210 test (lower suggests slight overfitting)
Mean absolute error = 8.5478 train, 9.4070 test (estimate within $9 of real price)
Mean squared error = 111.8958 train, 161.7316 test

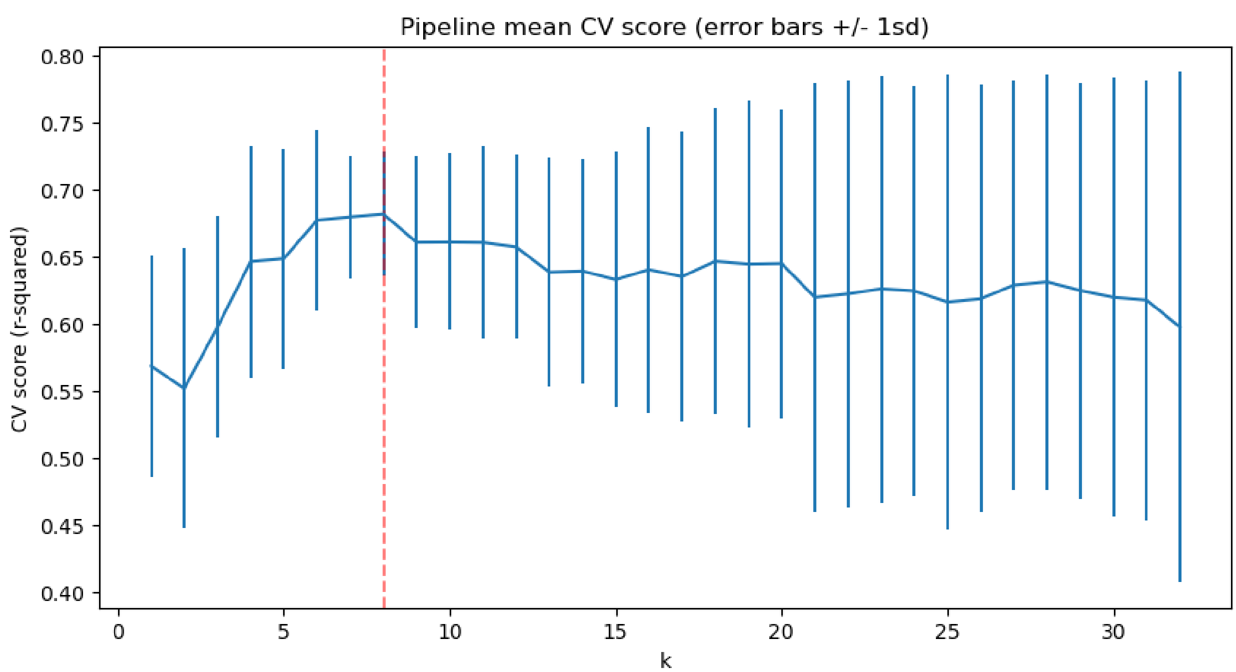
Imputing missing values with mean didn't show significant variation from median:

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

Built 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 the need for cross-validation confirmed.

Cross-validation mean score of 0.6327 with a std = 0.0950 highlights assessment is inherently open to variability. Estimate of variability calculated at (0.44, 0.82).

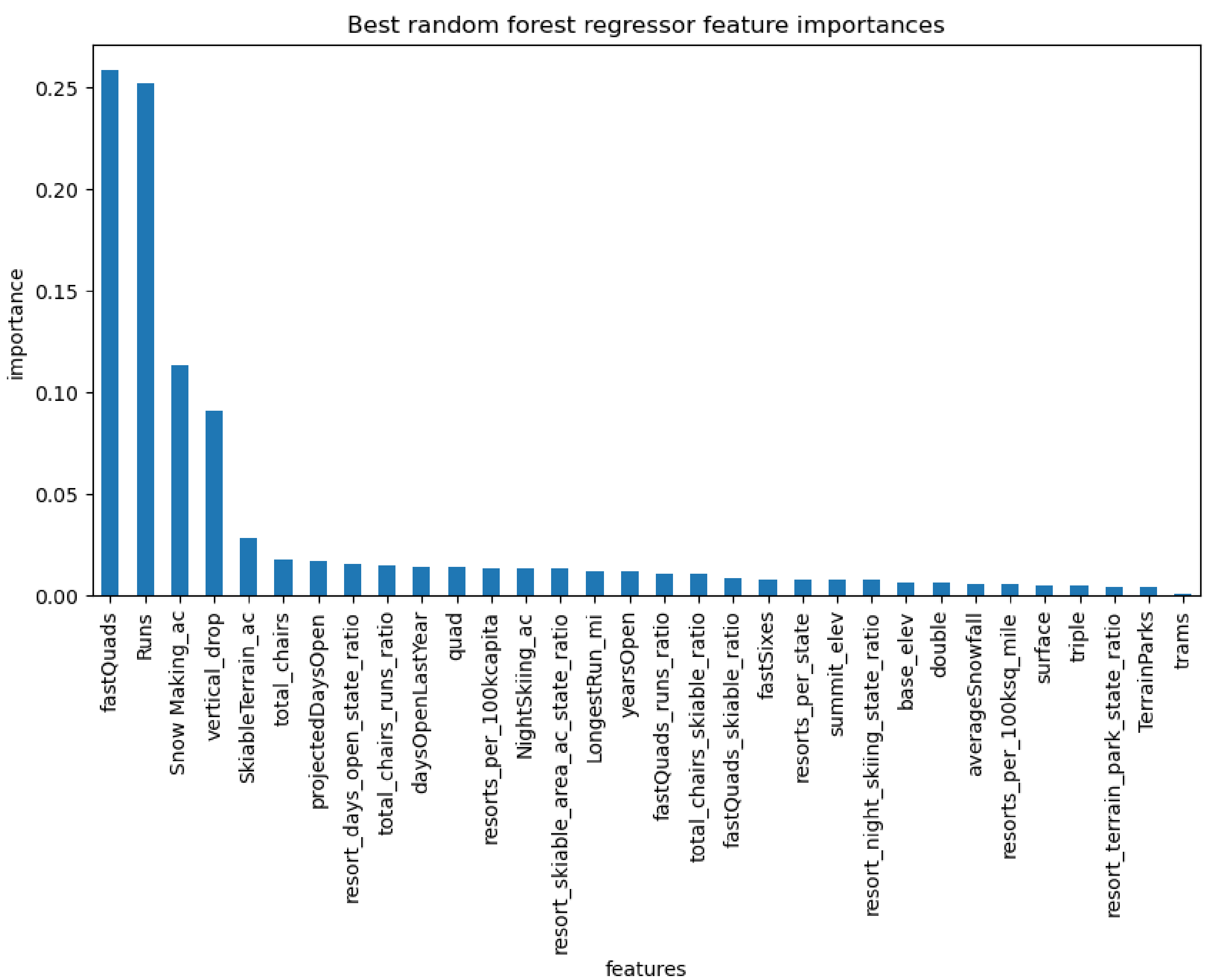
Hyperparameter search showed a best k of 8 with respective model results indicating vertical drop is the biggest positive feature, consistent with EDA conclusions, followed by snow making equipment, total chairs, fast quads and number of runs. Trams and skiable terrain showed negative results for this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

Additional random forest model assessed using cross-validation showed a mean score of 0.6976 with a standard deviation of 0.0709. Exploration of hyper parameters showed imputing median helps but scaling the features doesn’t. Final scores were .7097, 0.0645. The top four features in common with linear model are fast quads, number of runs, snow making equipment and vertical drop.

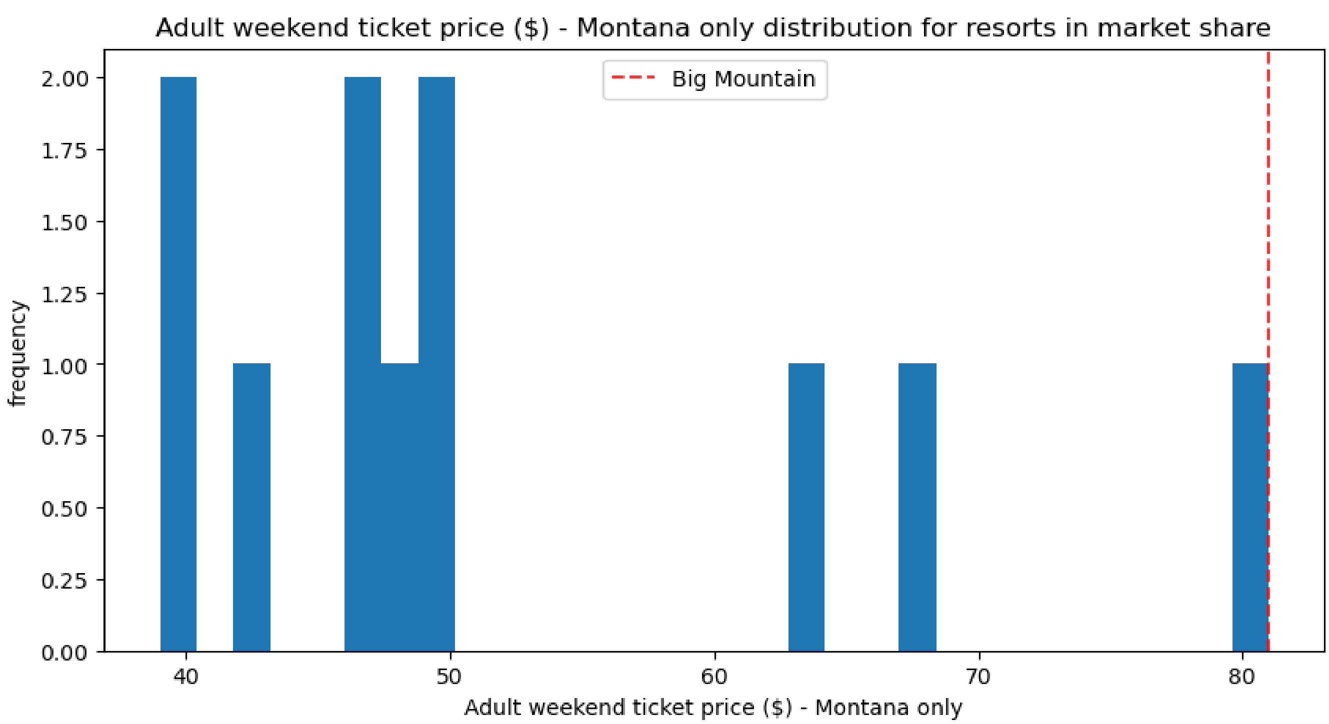


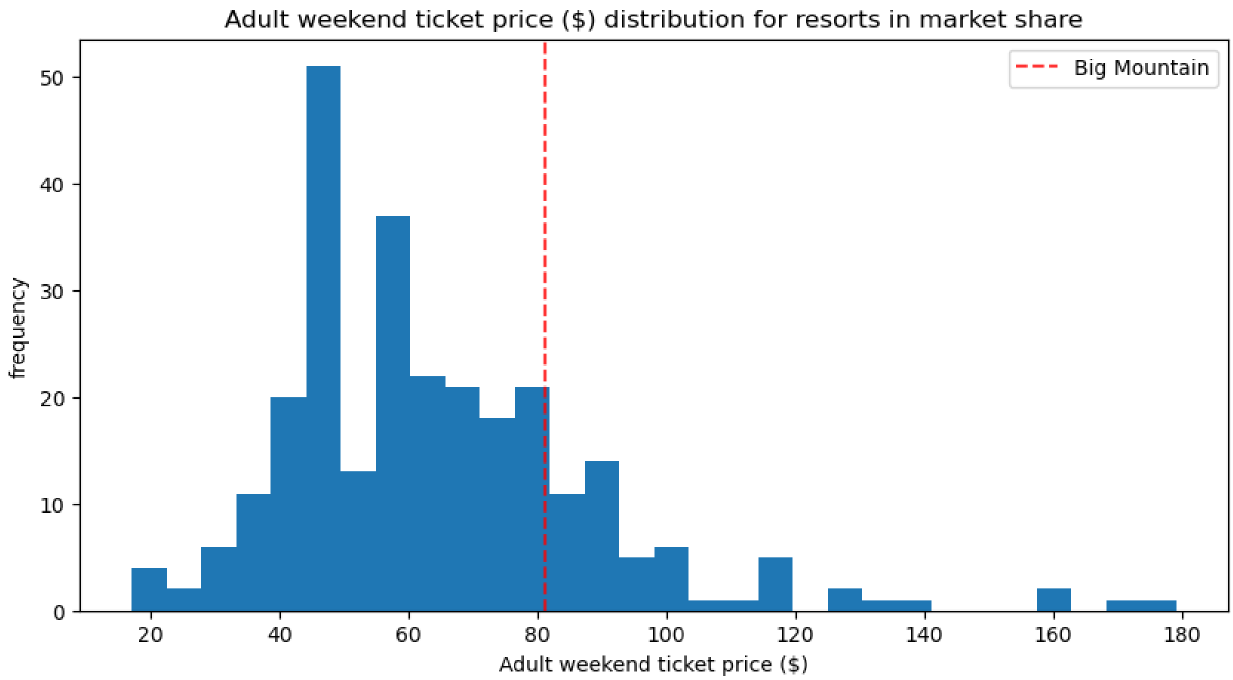
After calculating the mean absolute error for both regression models, linear = 11.7047 and random forest = 9.5377, the random forest model has a lower cross-validation mean absolute error by almost $1 while exhibiting less variability. Verifying performance on the test set produces performance consistent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

2.4 Modeling

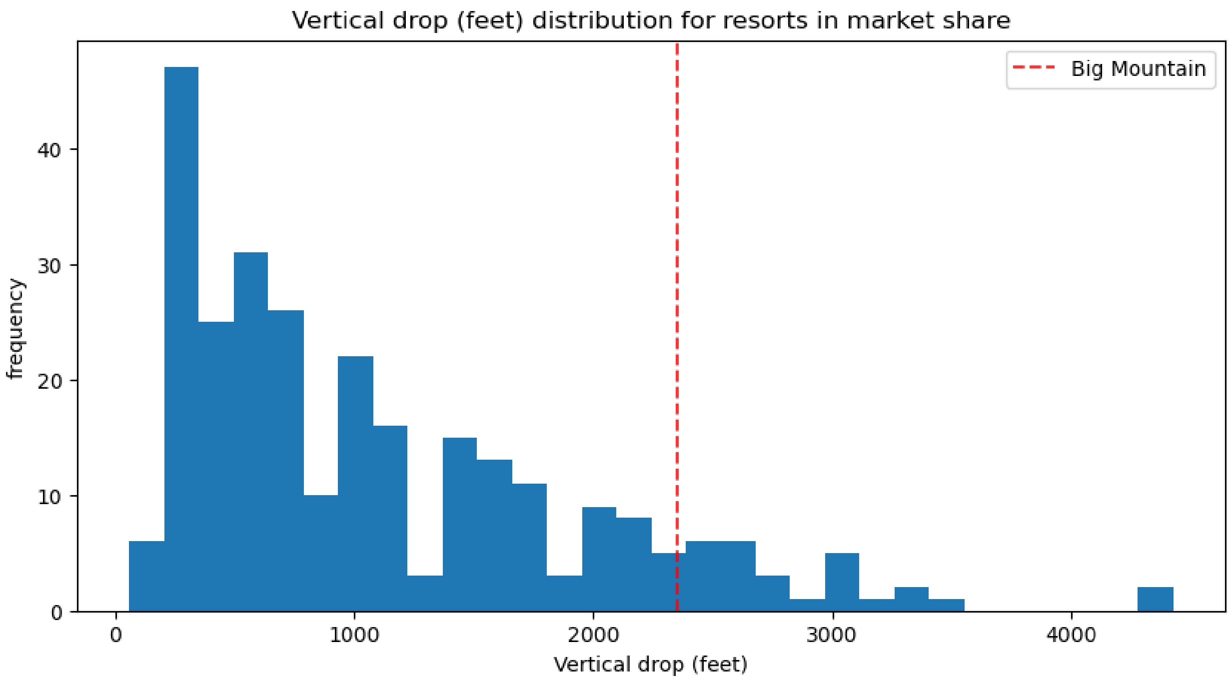
Big Mountain Resort modeled price is $95.87, compared to actual current price of $81.00. Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

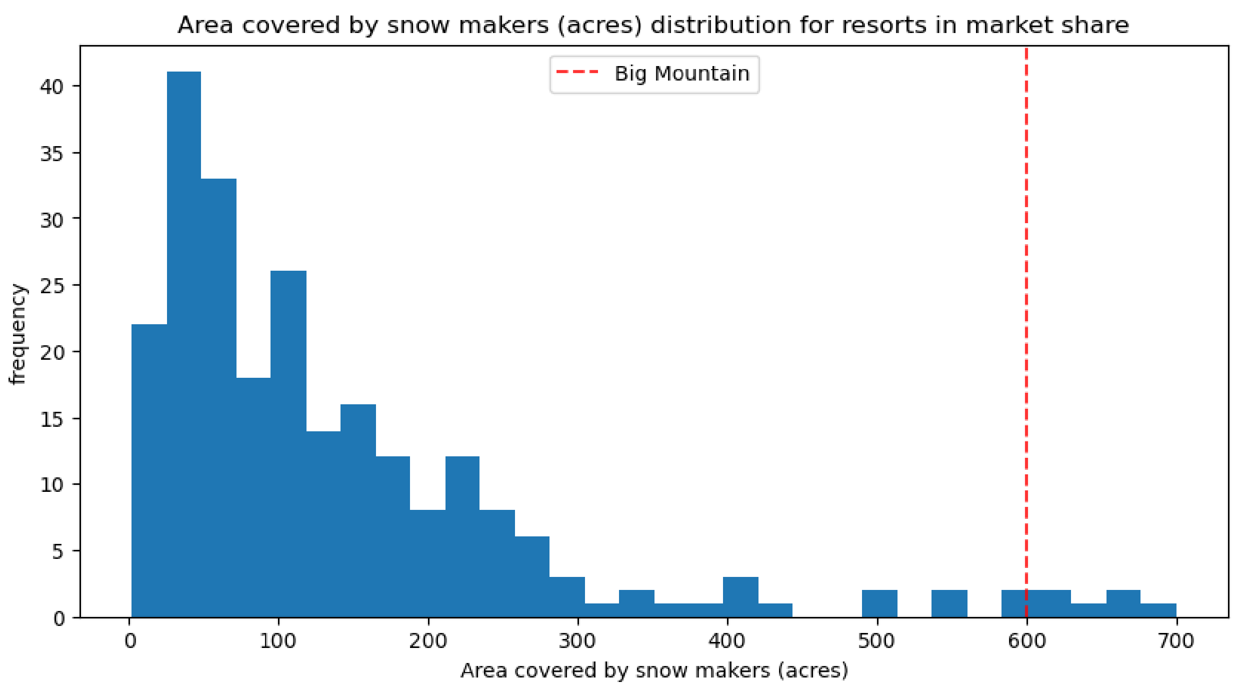
Looking at Montana only, Big Mountain Resort has the highest ticket price with two distrant price competitors. From market-perspective, however, increasing the ticket price is supported in relation to competitors.

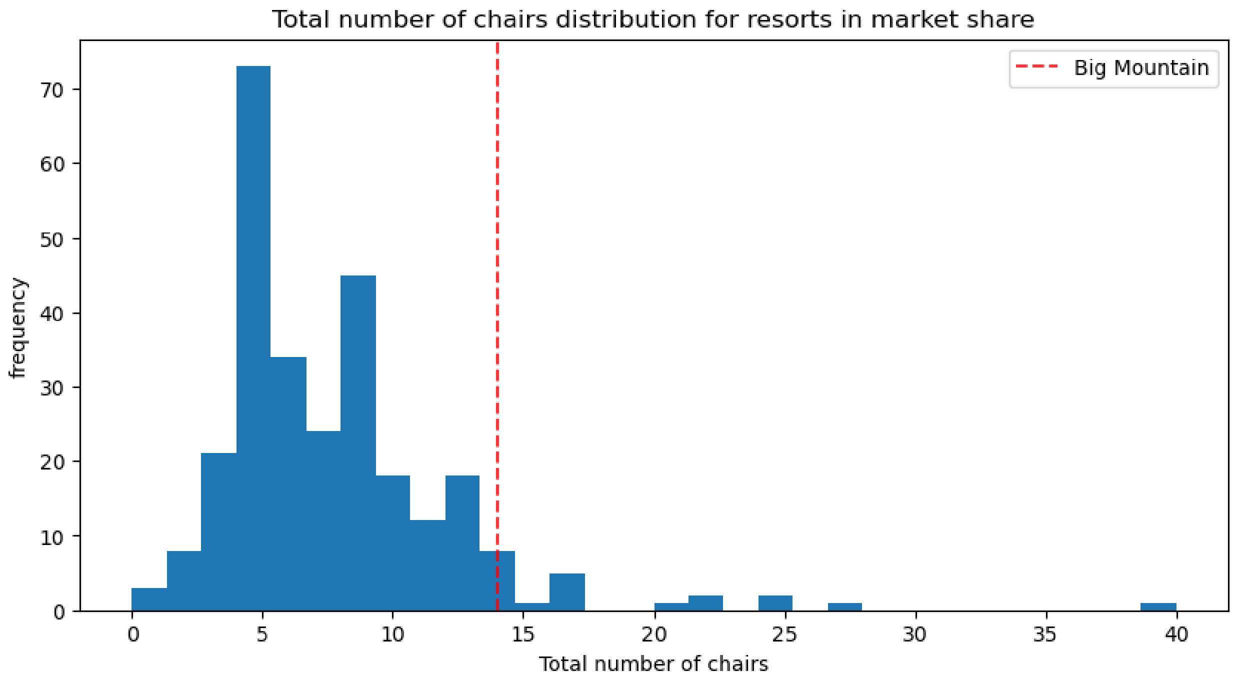


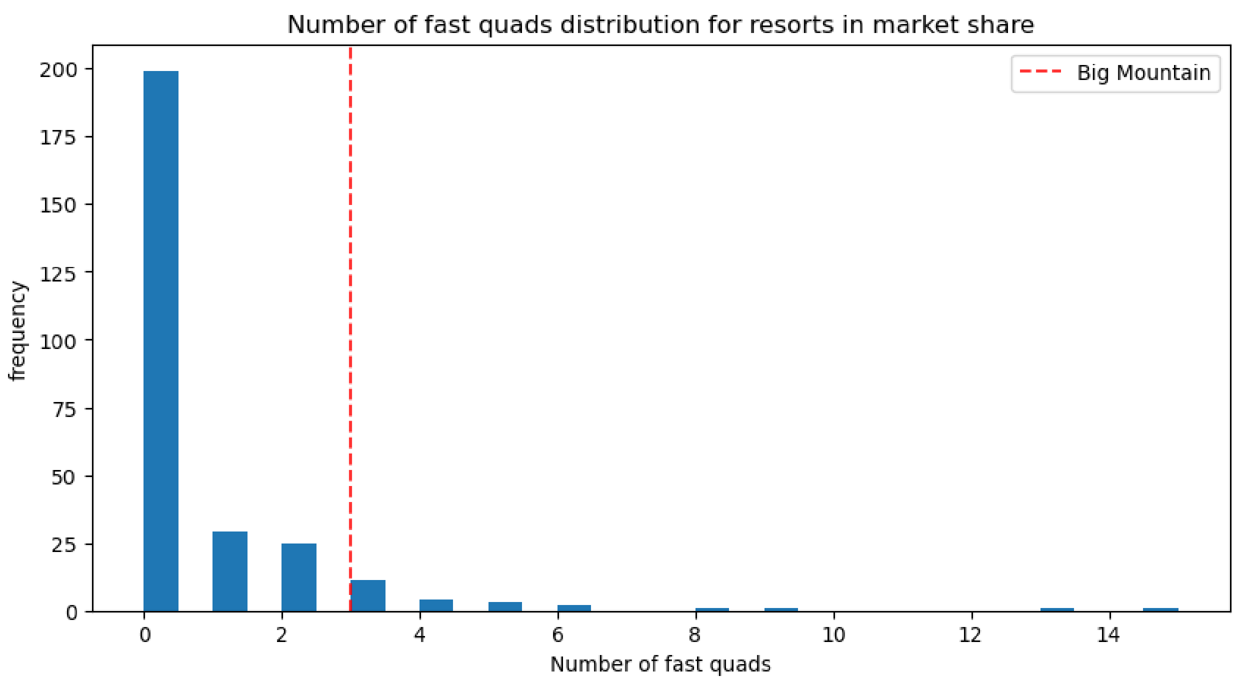


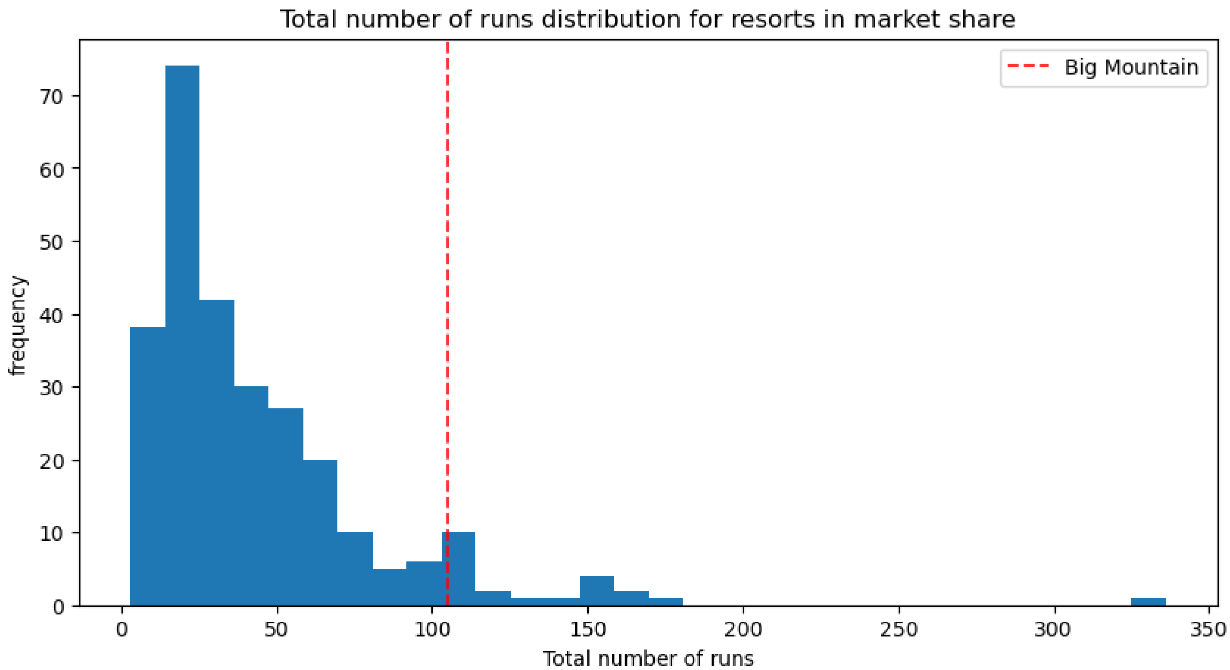
Big Mountain is very high up on the league table of snow making area. Big Mountain has amongst the highest number of total chairs. Resorts with more appear to be outliers. Most resorts have no fast quads. Big Mountain has 3, which puts it high up on the league table. There are some values much higher, but they are rare. Big Mountain compares well for the number of runs. There are some resorts with more, but not many. Big Mountain has one of the longest runs. Although it is just over half the length of the longest; the longer ones are rare. Big Mountain is among the resorts with the largest amount of skiable terrain.

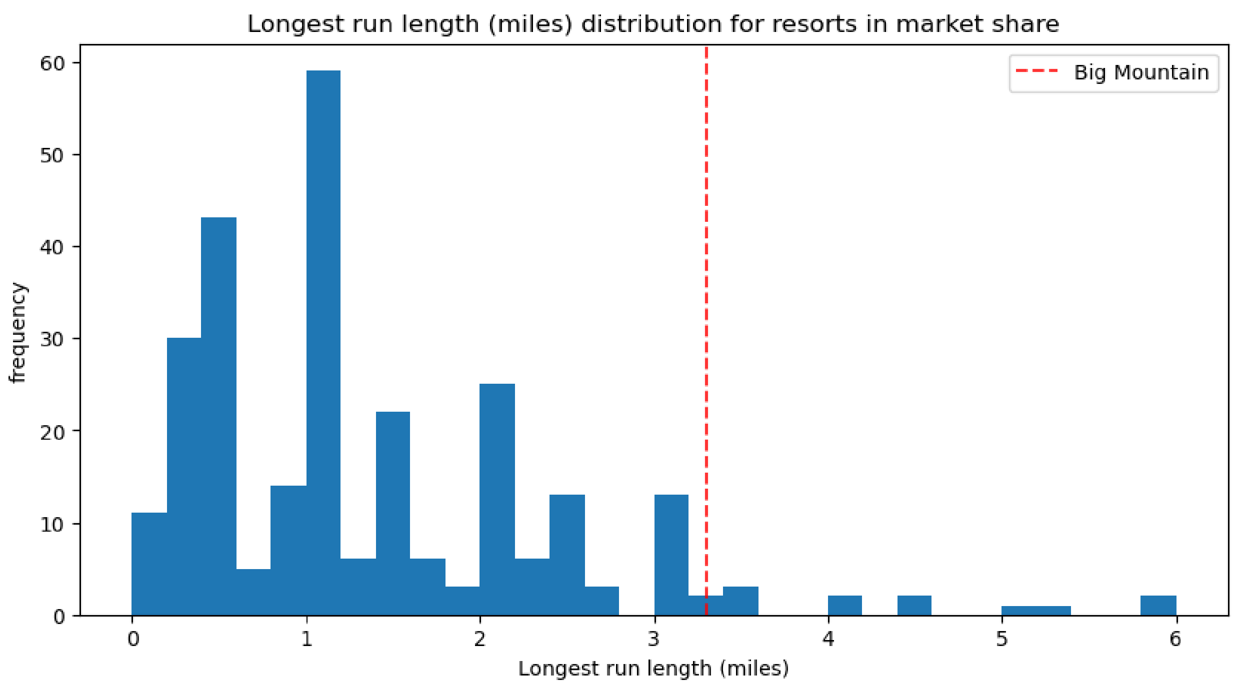

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
0
1000
2000
3000
4000
Vertical drop (feet)

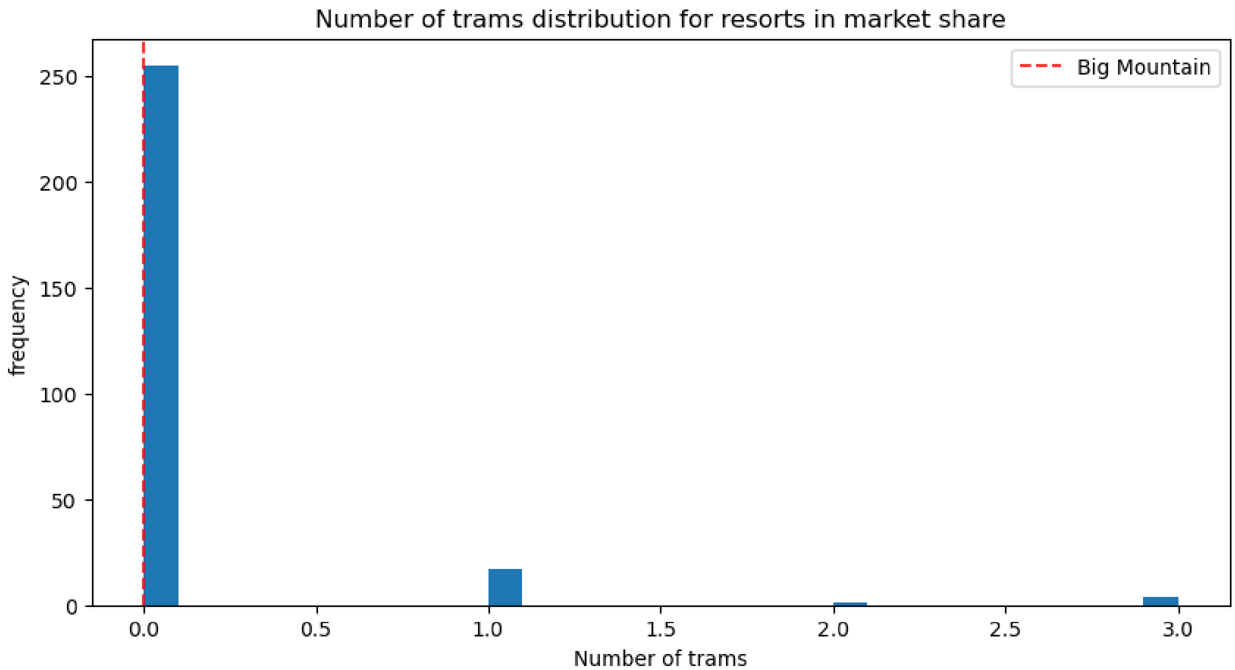

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
0
5
10
15
20
25
30
35
40
0
100
200
300
400
500
600
700
Area covered by snow makers (acres)

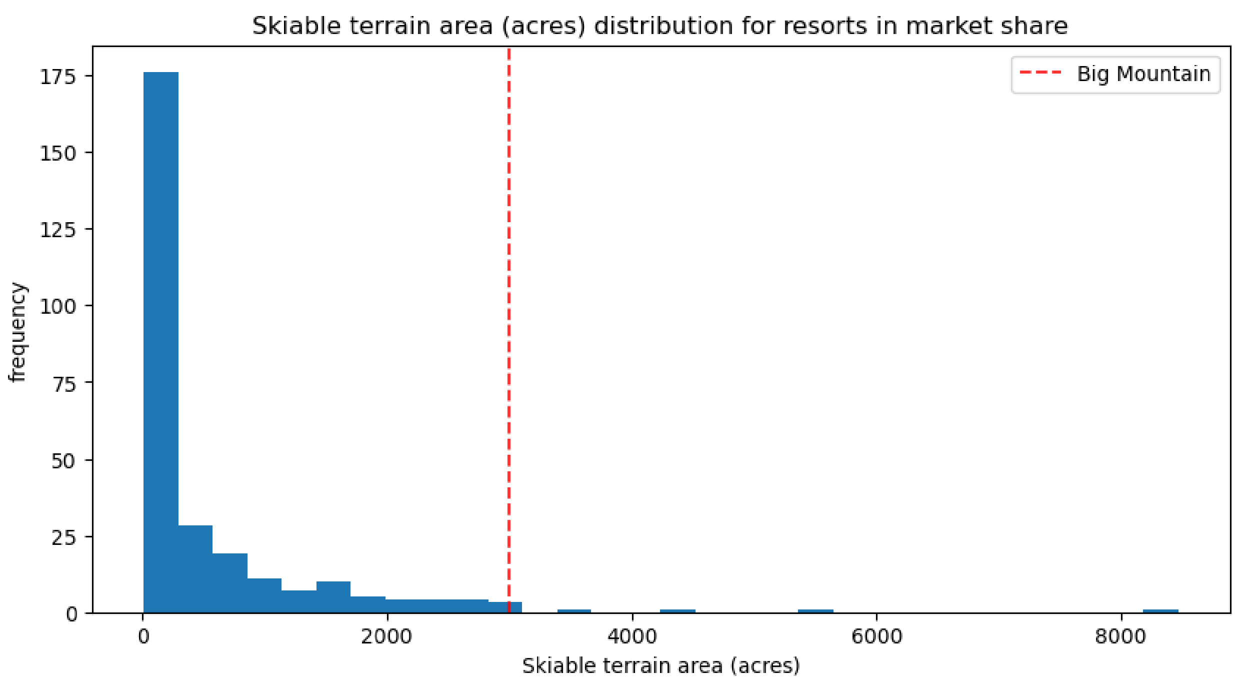

Total number of chairs distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
50
60
70
0
5
10
15
20
25
30
35
40
Total number of chairs


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


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


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


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


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

Scenario Modeling + Recommendations

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

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so 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.

<u>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</u>

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

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

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

<u>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</u>

Model results show no impact

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, creating additional revenue to cover the new operating costs, it also represents an opportunity to increase competitive advantage.

If 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.

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 EDA, 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.