Guided Capstone Project Report

August Perez
Springboard Data Science career track
October2023

Steps

1. Write a 1-2 page report summarizing your recommendations for Big Mountain Resort. Be sure to include the figures you created to back up your recommendations. See what should be included in [this rubric](#).
2. Title this document "Guided Capstone Project Report"
3. Add your project report to your GitHub repo Guided Capstone folder
4. Submit a link to the folder below

Big Mountain Resort is looking to adjust ticket prices to more accurately reflect the facilities and services they offer. The current pricing practice is to charge a premium on the market average with the market being all ski resorts the USA (around 300). BMR is also considering changes such as cost-cutting (without undermining ticket price) or additions to justify a higher ticket price. A new pricing strategy would more accurately reflect what BMR offers and also give insight into priorities for investment strategies.

What changes to ticket price and total night skiing area can Big Mountain Resort make before the month preceding next year's season opening to increase revenue by 20%?

Data Wrangling was performed to: ensure needed data to solve the problem is available, determine which feature in the data will represent the target to model, determine the data features to be used for determining the target in the model, and address missing data with imputation, removal of the feature, or removal of the source of that data (a ski resort).
Data from 330 ski resorts was gathered, organized, & filtered. The target feature was determined to be "adult weekend ticket price", removing "adult weekday ticket price" due to weekend prices having less null values for resorts. Data from 277 resorts remained after filtering for too little data. Two data features were removed: "8-person chairlifts" due to only one resort reporting them, and the "adult weekday tickets". Since metrics tied to location could have an effect on ticket price, state population metrics were gathered and merged into the ski resort dataset.

Exploratory data analysis was performed to identify themes, trends, and to summarize and visualize the characteristics of the data. The relationship between the state a ski resort is located in and ticket price was explored, determining to not use a resorts state but usefulness of ratios of a resorts features to state population compared to other resorts was noted for further study. Features used to create ratio metrics were skiable area, days open, terrain park count, & night skiing area. Ratio metrics are not the final answer because states with small resort counts would have more non-uniform distribution since small changes in a single resort could greatly change the ratio. Though there were states that on average had higher prices that others, it was determined that no full pattern or grouping existed (Fig.1, Table 1). Correlation of resort features to price was also explored (Fig.2, Fig.3).

With appropriate data selected and processed and EDA performed for insight, initial modeling was performed to determine the best model suited for the job of suggesting a new ticket price for BMR. The best model was determined to be a random forest regressor from lowest mean-absolute-error and lowest standard deviation. Training was performed by using resort features as input and comparing calculated price to actual price. Data was split into groups so the models were not exposed to the entire dataset all at once in order to be able to test the models and allow for retraining if any changes made. Models explored include mean-prices (mean-absolute-error around $19 from actual price), linear regression (MAE: about $12, standard deviation: about 1.62), random forest regressor with median imputation strategy (MAE: about $9.50, st.dev: 1.35).

The RF regressor model was then fully trained on all the data, excluding BMR. Using the now trained model, a price was predicted for BMR based on its current features. Changes BMR could make to cut costs or justify even higher ticket prices were explored through 4 scenarios by changing BMR feature values and running through the model. Scenario 1 being closure of up to 10 least used ski runs (Figure 4). Scenario 2 being adding a run, increasing vertical drop by 150ft, and adding a chairlift. Scenario 3 being the same as scenario 2 and adding 2 acres of snow making. Scenario 4 being increasing the longest runs distance by 0.2 miles and adding 4 acres of snow making. Revenue was estimated using assumptions of 350,000 visitors per year and 5 days worth of tickets purchased per visitor. Exploration was also done to compare BMR's features to the other resorts in the dataset to get a sense of where BMR fell. Discovering that BMR is a top contender for almost all of the major ski resort features (Fig.5-11) so a higher price would likely not be negatively viewed by visitors. Scenario 1 would be the suggested course of action because the other scenarios add operating costs rather than any reductions.

<u>Pricing recommendation:</u>
Current conditions: Price: $81.00 Est.Revenue: $141,750,000.00
No changes: Price: $95.87 (MAE of $10.39) Est.Revenue: $167,772,500 ($26,022,500 increase)

In summary: In order to provide BMR with a ticket price that more accurately reflects its facilities and services a model was created that predicts price based on a resort's facilities and services. Multiple scenarios and their effect on ticket price were also explored.

Future work would include obtaining more data about resorts such as operating costs, pricing practices of other resorts, and making the model available to analysts to simulate other possible changes BMR could make.

Figures and Tables

Figure 1: Plotting of states using PCA analysis to visualize possible groupings (dot color represents ticket price range, in dollars)

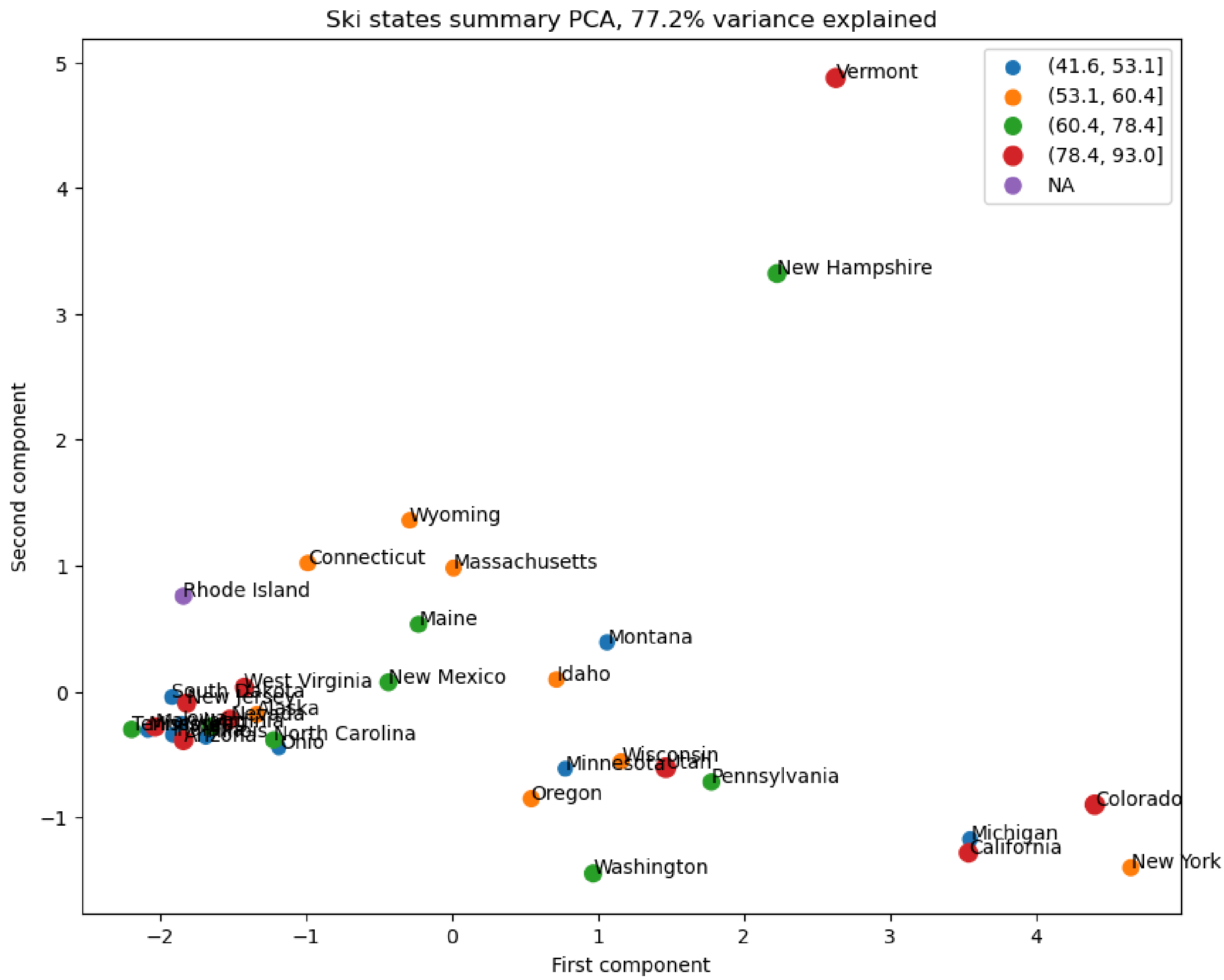


Table 1: Compared states (ranked by average ticket price) with aggregated metrics

Rank	state	Adult Weekend	resorts_per_state	state_total_skiable_area_ac	state_total_days_open	state_total_terrain_parks	state_total_nightskiing_ac	resorts_per_100kcapita	resorts_per_100ksq_mile
1	Utah	93.00	13	30508.0	1544.0	26.0	642.0	0.405495	15.312673
2	Colorado	90.71	22	43682.0	3258.0	74.0	428.0	0.382028	21.134744
3	Vermont	87.90	15	7239.0	1777.0	50.0	50.0	2.403889	155.990017
4	Arizona	83.50	2	1577.0	237.0	6.0	80.0	0.027477	1.754540

5	California	81.41	21	25948.0	2738.0	81.0	587.0	0.053148	12.828736
6	Nevada	81.00	4	2110.0	415.0	9.0	0.0	0.129864	3.617552
7	New Jersey	79.99	2	190.0	170.0	4.0	181.0	0.022517	22.927892
8	West Virginia	79.75	4	542.0	342.0	9.0	187.0	0.223196	16.508461
9	Maryland	79.00	1	172.0	121.0	3.0	118.0	0.016541	8.060616
10	New Hampshire	76.50	16	3427.0	1847.0	43.0	376.0	1.176721	171.141299
11	Washington	70.14	10	15330.0	1022.0	21.0	1997.0	0.131322	14.025639
12	Virginia	68.00	4	269.0	366.0	4.0	135.0	0.046863	9.351257
13	New Mexico	65.66	9	5223.0	966.0	18.0	50.0	0.429220	7.401925
14	Tennessee	65.00	1	0.0	83.0	1.0	0.0	0.014643	2.372817
15	North Carolina	64.16	6	370.0	506.0	9.0	335.0	0.057208	11.148479
16	Rhode Island	64.12	1	30.0	100.0	1.0	30.0	0.094397	64.724919
17	Pennsylvania	63.68	19	1888.0	1404.0	47.0	1528.0	0.148414	41.255917
18	Maine	61.00	9	3216.0	865.0	17.0	388.0	0.669537	25.438101
19	Oregon	59.85	10	11774.0	1180.0	22.0	1127.0	0.237094	10.164771
20	New York	58.94	33	5514.0	2384.0	72.0	2836.0	0.169635	60.489414
21	Alaska	57.33	3	2280.0	345.0	4.0	580.0	0.410091	0.450867
22	Massachusetts	57.20	11	1166.0	671.0	18.0	583.0	0.159594	104.225886
23	Connecticut	56.80	5	358.0	353.0	10.0	256.0	0.140242	90.203861
24	Wyoming	56.16	8	6523.0	716.0	14.0	110.0	1.382268	8.178872

25	Idaho	55.90	12	16396.0	1136.0	27.0	415.0	0.671492	14.359392
26	Wisconsin	55.21	15	1750.0	1519.0	40.0	1065.0	0.257624	22.902162
27	Michigan	52.40	28	4406.0	2389.0	63.0	1946.0	0.280368	28.951341
28	Montana	51.90	12	21410.0	951.0	27.0	710.0	1.122778	8.161045
29	South Dakota	51.50	2	950.0	183.0	3.0	0.0	0.226076	2.593496
30	Minnesota	49.66	14	1560.0	1490.0	29.0	1020.0	0.248243	16.103800
31	Indiana	48.50	2	165.0	157.0	4.0	165.0	0.029708	5.491488
32	Missouri	48.00	2	60.0	69.0	2.0	47.0	0.032587	2.869152
33	Ohio	45.40	5	421.0	489.0	12.0	421.0	0.042775	11.154241
34	Illinois	43.33	4	191.0	221.0	6.0	191.0	0.031566	6.906793
35	Iowa	41.66	3	140.0	100.0	5.0	140.0	0.095085	5.331153

Figure 2: Heatmap to show correlation between resort features and price

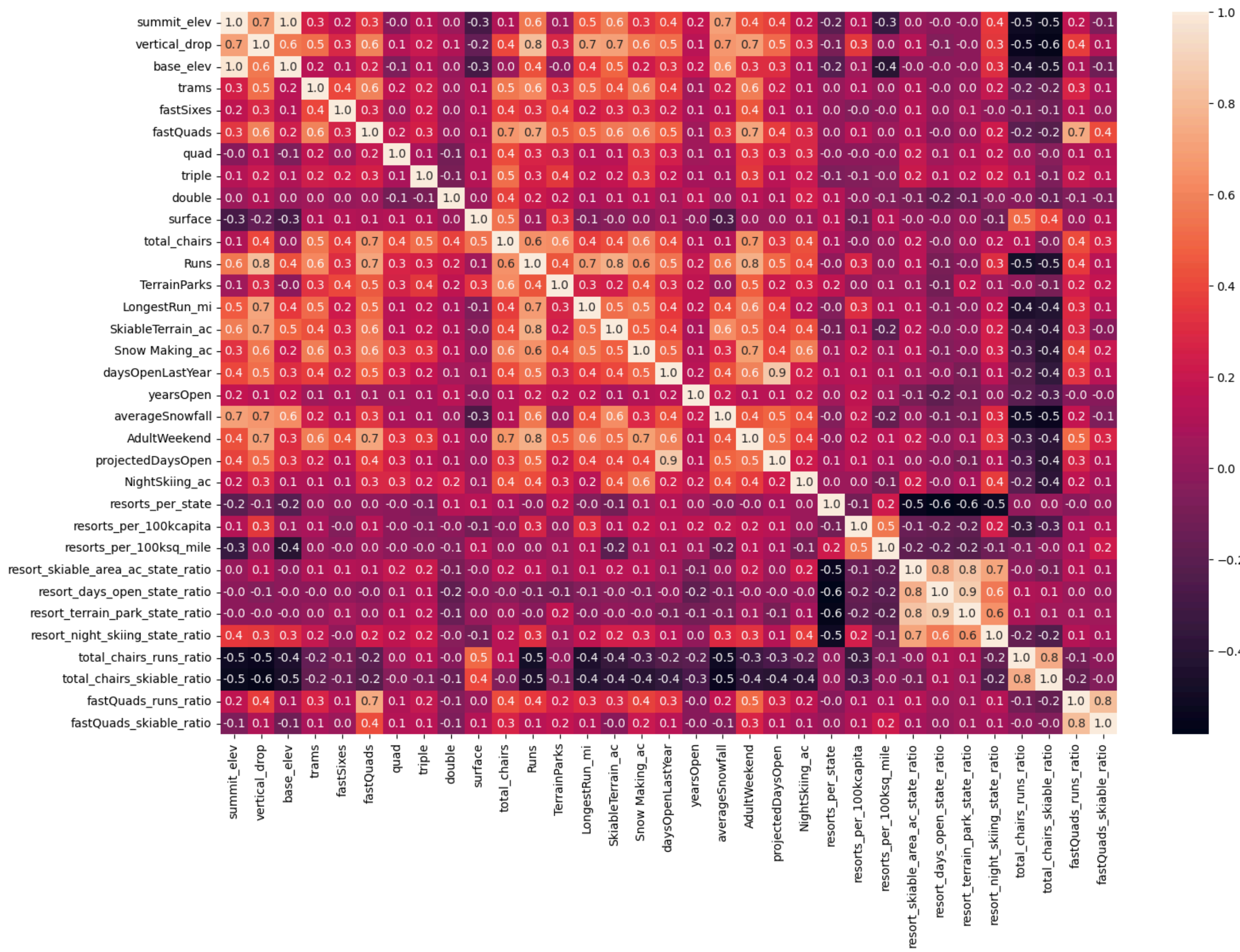


Figure 3: Scatterplots of ticket price vs resort_state_ratio features

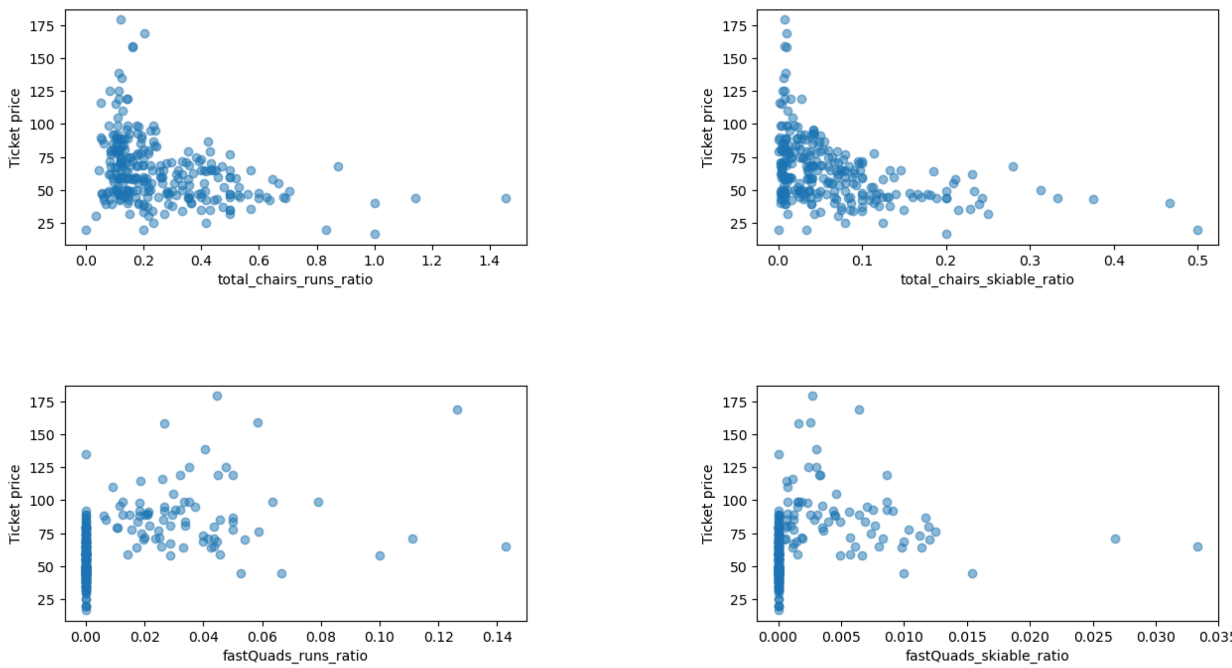


Figure 4: Ticket price and revenue changes per number of runs closed.

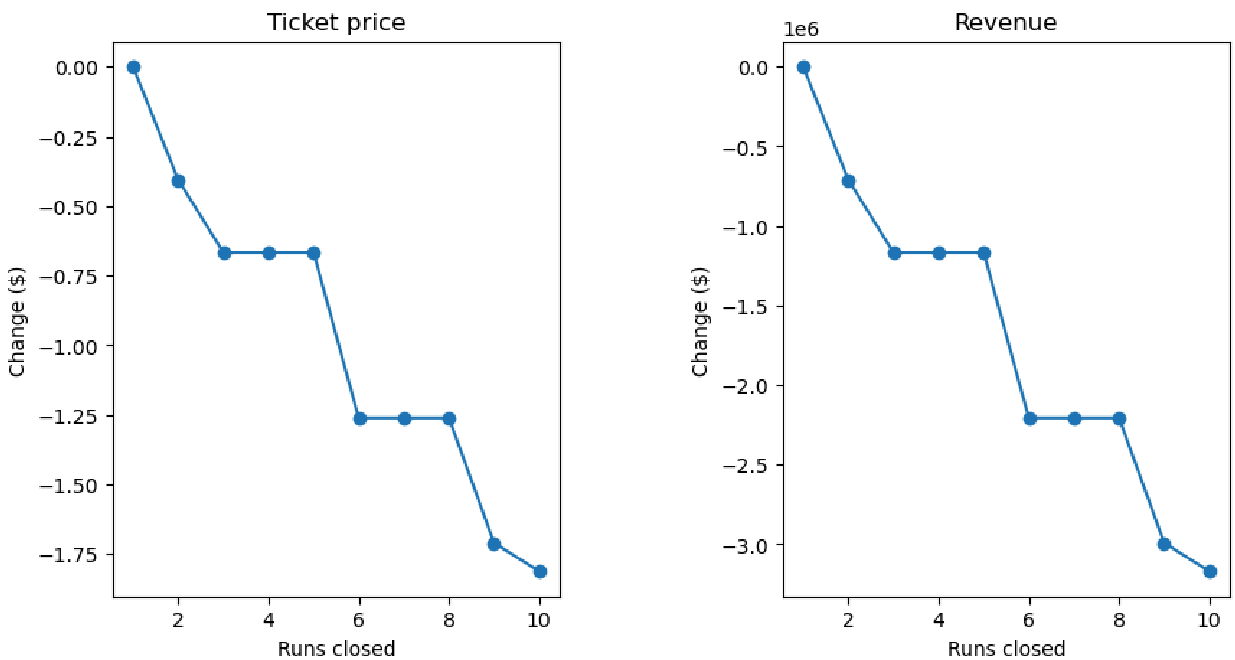


Figure 5:

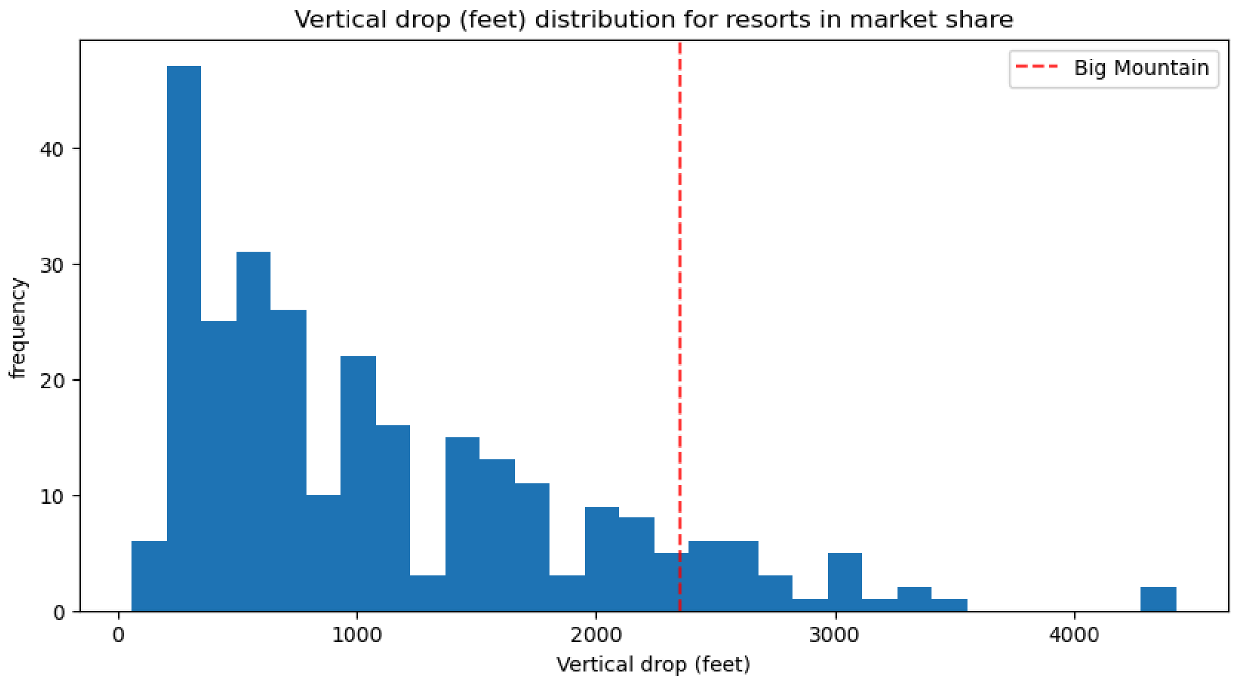


Figure 6:

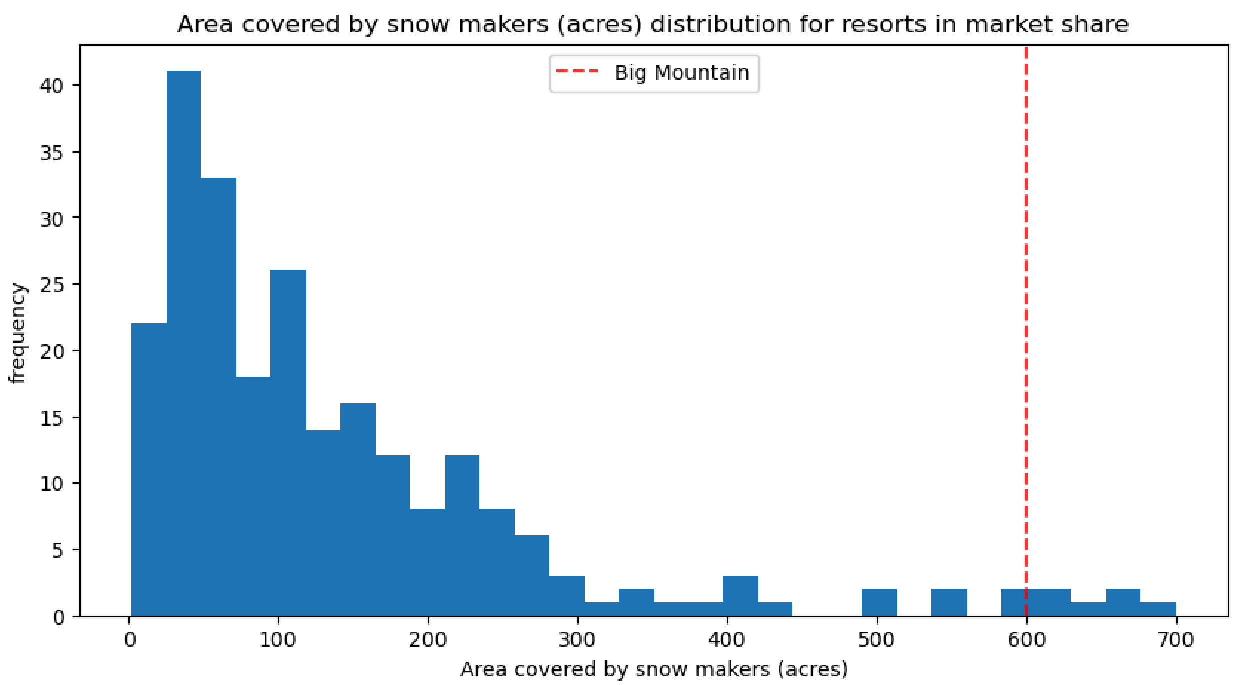


Figure 7:

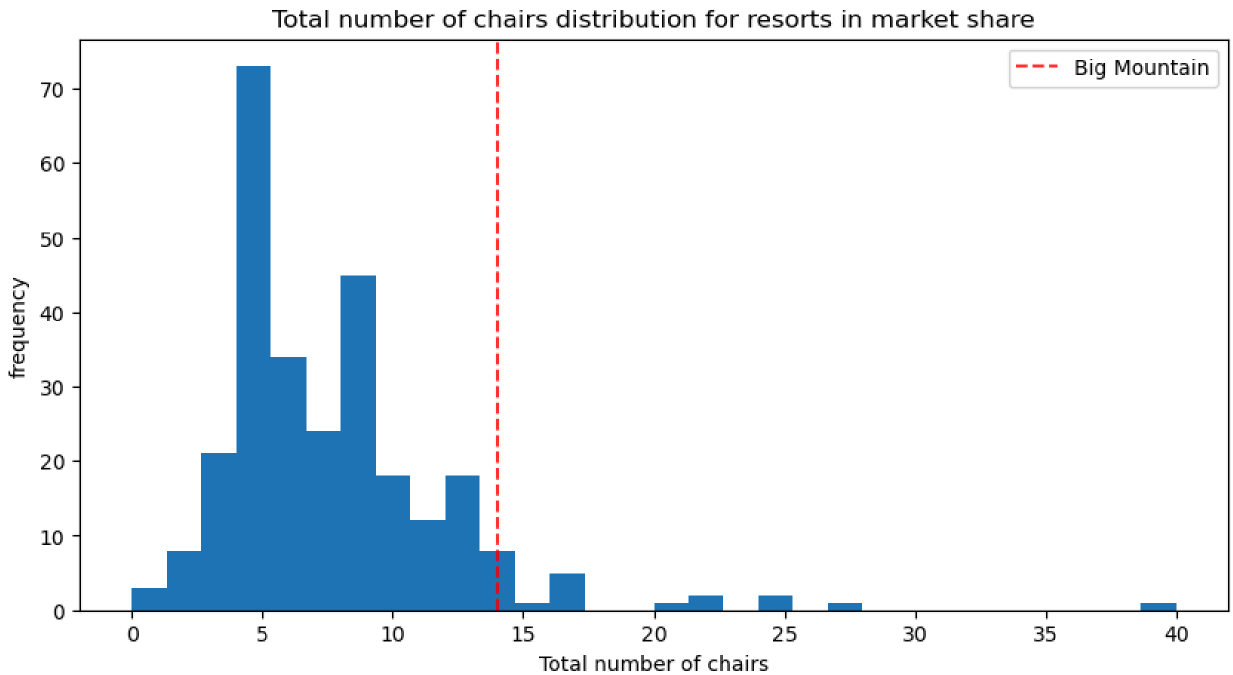


Figure 8:

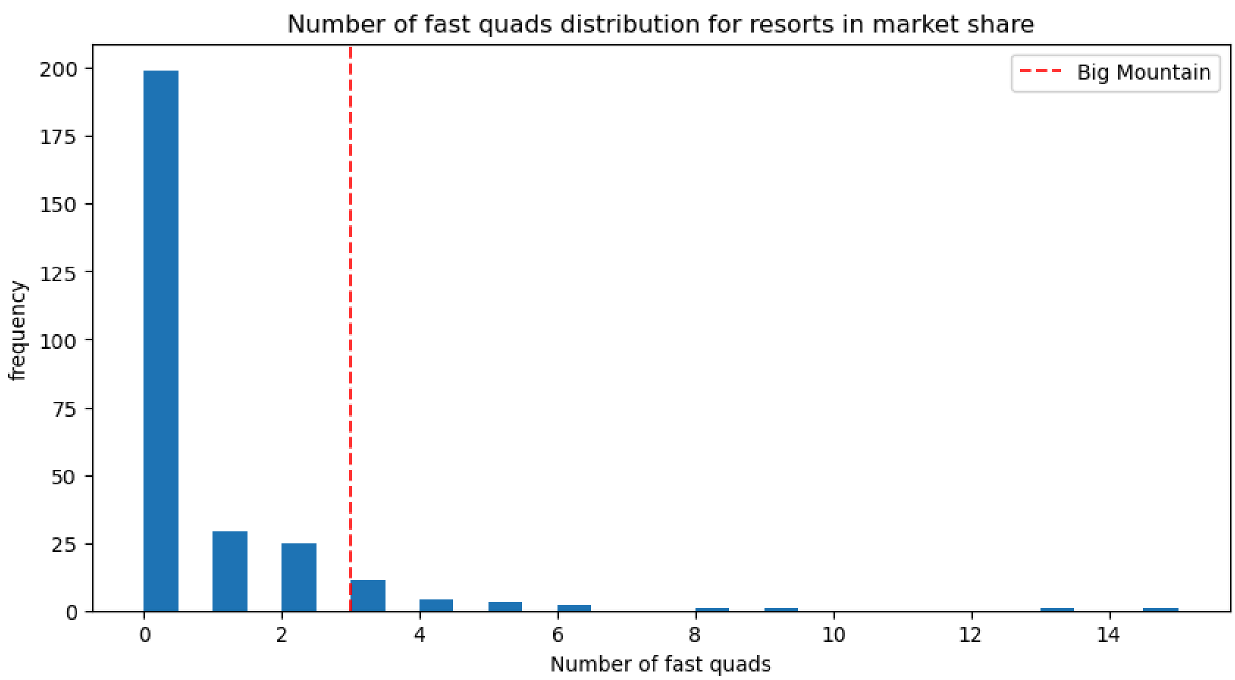


Figure 9:

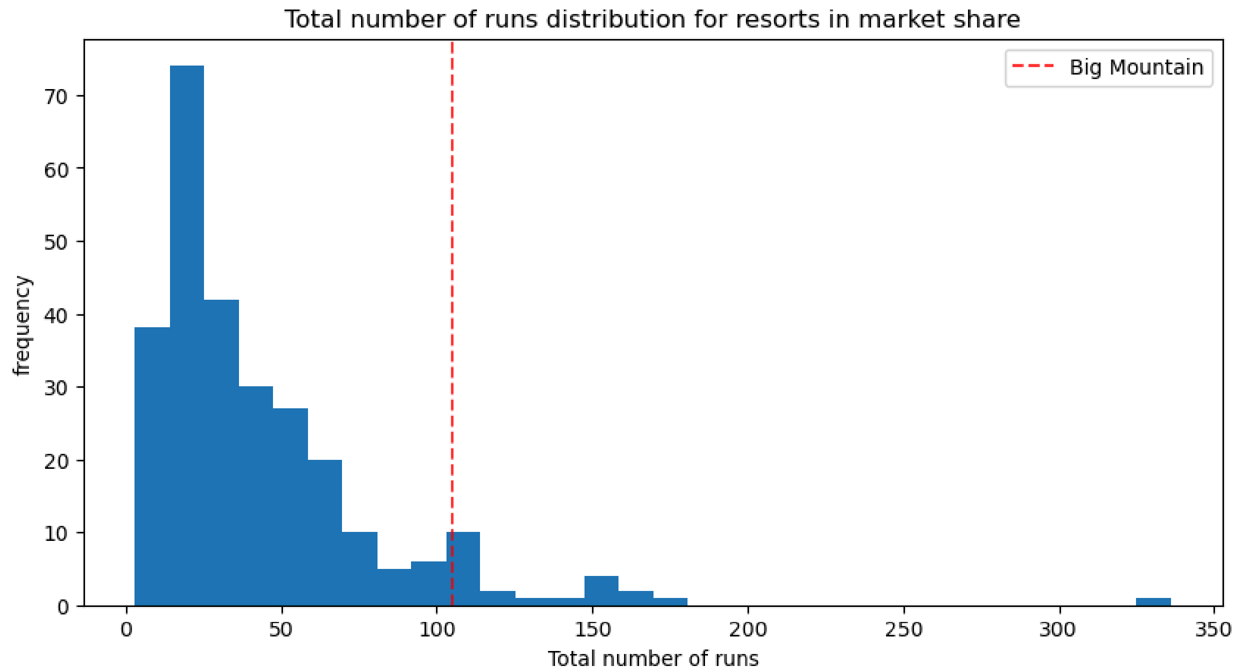


Figure 10:

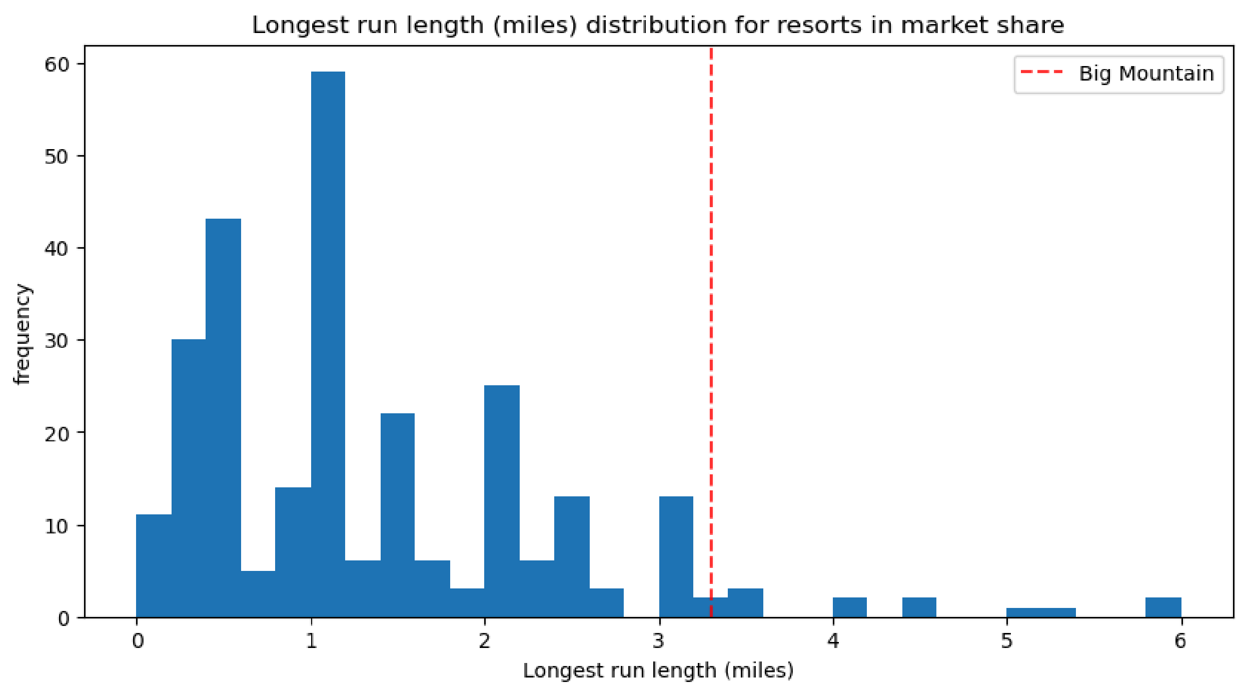
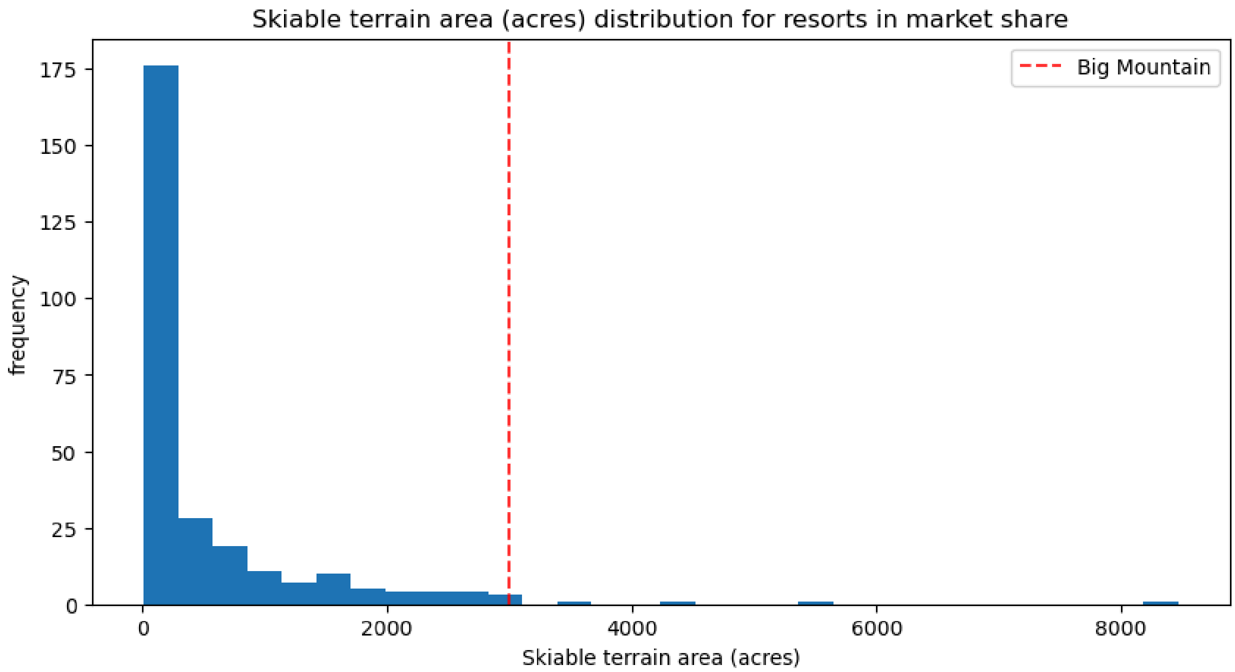


Figure 11:


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)