Big Mountain Resort Project

Author | Rand Sobczak

Table of Contents

1.0 Problem Identification Overview

Big Mountain Resort (BMR) is a ski resort in Montana. BMR recently installed a new chair lift to help increase the distribution of visitors across the mountain. The new chair lift increased the operating costs by $1,540,000 per annum. BMR sees about 350,000 visitors ski or snowboard every year.

In summary, the problem we are addressing is, **how can we increase profits, starting this year, by a minimum of $1,540,000 per annum to offset the aforementioned costs with the new chair lift**?

We used BMR's available dataset with information about BMR as well as 330 other ski resorts across the United States in 35 states and representing ~119,000 acres of skiable terrain. This list was eventually brought down to 302 ski resorts for reasons stipulated below in #3 Data Cleaning.

Additional information and coordination with relevant BMR personnel is always welcome. One additional piece of information I believe would be helpful would be the proximity of the resorts to the closest airport respectively.

2.0 Generated Deliverables

Three (3) deliverables were generated:

1. The source code for the modeling developed to analyze the aforementioned problem
2. This document outlining the process
3. A PDF presentation with our advice to management to address the problem

3.0 Data Pre-Processing Steps

Data Cleaning:

1. **I received a single CSV file** provided by the Database Manager. Details can be found in Appendix I.
2. I amended SkiableTerrain_ac for a resort named Silver Mountain in Colorado as the information was noticeably off kilter in the dataset and also did not comply with the information on their website; the website information was used
3. I removed the fastEight column as 50% of the resorts had no values (Appendix II)
4. I also removed ~14% of the resorts as they did not have both AdultWeekday and AdultWeekend information; thus they weren't useful considering our target is price (Appendix III)
5. **Population and state size data was pulled from Wikipedia** (link) to setup per capita and acre metrics
6. Data for one of the resort's yearsOpen column was 2019. This was removed as that data is unlikely to be true as ski resorts aren't open for 2019 years; it probably opened in 2019.

Exploratory Data Analysis:

Two datasets were pulled from the previous step; namely:

- ski_resort_data: provided by the Database Manager on information which was cleaned down to 277 resorts across 34 states; and
- state_summary: an assembled dataset from Wikipedia which provided high level population / state size information.

The **categorical features** are the resort names, the respective region (which did not appear useful), & the state in which the resort is located in. **Numerical features** inside the 2 datasets are plentiful & additional variables were added. Among them, I believe the resort_night_skiing_state_ratio, Snow Making_ac, Runs & the vertical_drop provided possibly valuable insights to focus on higher prices; notably, people's preference for night skiing, their value for some guarantee of snow, their value for different run options, & desire for a steeper drop respectively. These **observations in the numerical features were identified using a Feature Correlation Heatmap** (Appendix IV). **These items appear to be of higher importance to achieve our desired goal of higher prices; thus, should take priority**.

Issues of the data & it's visualization of it that I addressed modeling include but are not limited to:

- Filling in the missing numbers with averages of others to handle missing data which may throw off scatter plots & quartile numbers
- Utilizing seaborn rather than relying on matplotlib for scatter plots which may not be the ideal display candidate
- Scaling the numbers when required

Pre-processing:

1. The first step in pre-processing was to **split the data into a training and testing splits**, 70% and 30% respectively, and also removing any object types so that the dataset is only numeric types.
2. The next step, was to determine **the "Best Guess" number** which, with the help of a standard mean and the DummyRegressor functions, was determined to be 63.81.
3. Using **the mean absolute error**, I determined that on average the metric may be off by ~$19.
4. **I then scaled the data** to ensure that the data can work better together.
5. Once scaled, **the R squared for the train and test datasets became ~81% and 72% respectively**. Additionally, **the mean absolute errors dropped down to ~$9**; much better than the $19 from just guessing.
6. I also replaced any missing data & the results weren't much different.
7. I then **ran the training and testing splits through the Pipeline function**. In conjunction with cross-validation, **I determined that the best k value was 8** (Appendix V).
8. Using the linear coefficient numbers for each item versus AdultWeekend price, the most positively correlated item was vertical_drop (10.76) & most negatively correlated was SkiableTerrain_ac (-5.25) (Appendix VI)
9. **Moving from a linear model to a Random Forest Generator marginally improved upon our cross-validation results** and showed the top four dominate features were similar to those found in the linear model (Appendix VII)
10. As **the Random Forest Model had a lower cross-validation mean absolute error by almost $1 and exhibits less variability, it was chosen for our modeling**. To note, the verifying performance on the test set produced performance consistent with the cross-validation results.

4.0 Model Description

Due to its better performance, **the model I used was the Random Forest Model with the goal of determining what price BMR's facilities support as well as sensitivity of changes to various resort parameters**. This model relies on the implicit assumption that all other resorts are largely setting prices based on how the market values certain facilities; essentially prices are set by the free market.

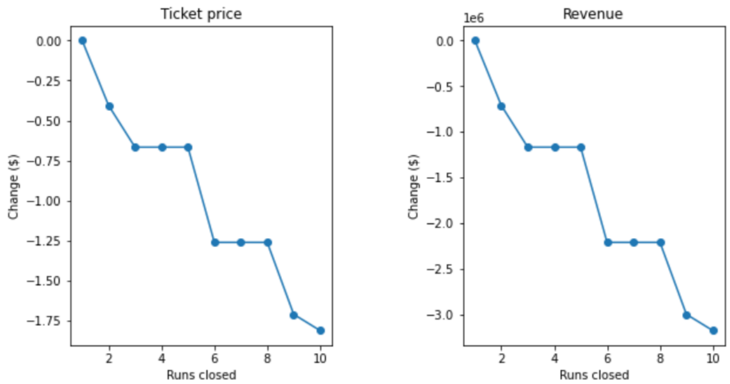
5.0 Model Findings

As the business sought to review potential scenarios for either cutting costs or increasing revenue from ticket prices, **I modelled four scenarios to ascertain how facilities support a given ticket price** as per below:

1. Permanently closing down up to 10 of the least used runs
2. Increasing the vertical drop to a point 150 feet lower but requiring the installation of an additional chair lift without additional snow making coverage
3. Same as #2 above albeit adding 2 acres of snow making cover
4. Increasing the longest run by 0.2 miles to 3.5 miles in length and adding additional snow making cover of 4 acres to cover it

As previously mentioned, BMR sees 350,000 visitors per season. In conjunction with this, we added an assumption that each visitor stays for 5 days.

In **Scenario 1**, the model shows that closing 1 run makes no difference to either Ticket Prices or Revenue. The impact begins to be seen when 2 runs are closed. If 3 runs are closed, BMR could also close 4 or 5 as they have the same impact on Ticket prices and Revenue. Closing 6 or more is when the model indicates a substantially negative impact would occur.



To reiterate, in **Scenario 2** I modeled increasing the vertical drop to a point 150 feet lower but requiring the installation of an additional chair lift without additional snow making coverage. In this scenario, the model increases support for ticket prices by $1.99 to $82.99 which could increase revenue by $3,474,638 over the season.

In **Scenario 3** I modeled the same as Scenario 2 but added 2 additional acres of snow making. The model's output was identical as Scenario 2 as the model increases support for ticket prices by $1.99 to $82.99 and may increase revenue by $3,474,638 over the season.

In our final scenario, **Scenario 4,** I increased the longest run by 0.2 miles and added 4 acres of additional snow making cover. The output showed no difference from the current. I believe that since the Random Forest Model placed the longest run low on the importance list, this was our output.

6.0 Next Steps

I would recommend that Scenario 2 be taken under consideration as the model presented a possible solution to our problem. While Scenario 3 presents identical findings, I believe it may incorporate additional cost(s) above those in Scenario 2.

Additional analysis can be undertaken with different departments; notably:

- Database Department | Review the sourcing methodology of the information; particularly those that were removed
- Accounting & Finance Department | Further details and or structure of operational expenses
- Operations Department | The perceived benefit to operations by an additional chair lift

Other data that may paint an even clearer picture may be proximity to the respective airports. This may not be relevant but if BMR's distance to an airport is comparatively higher than the competition in the dataset, this may make it less competitive and negatively impact our models.

Appendix I

Name	The name of the resort.
Region	The region within the United States where the resort is located.
state	The state within the United States where the resort is located.
summit_elev	The elevation in feet of the summit mountain at the resort.
vertical_drop	The vertical change in elevation from the summit to the base in feet.
base_elev	Elevation in feet at the base of the resort.
trams	The number of operational trams.
fastEight	The number of fast eight chairs.
fastSixes	The number of fast six chairs.
fastQuads	The number of fast four chairs.
quad	The number of regular speed four person chairs.
triple	The number of regular speed three person chairs.
double	The number of regular speed two person chairs.
surface	The number of regular speed single person chairs.
total_chairs	The number of chairlifts at the resort
Runs	The total number of runs at the resort.
TerrainParks	The total number of terrain parks at the resort.
LongestRun_mi	The length of the longest run at the resort in miles.
SkiableTerrain_ac	The total amount of skiable acres at the resort.
Snow Making_ac	The total number of acres covered by snow making machines.
daysOpenLastYear	The total number of days open last year.
yearsOpen	The total number of years the resort has been open.
averageSnowfall	Average annual snow fall at the resort in inches.
AdultWeekday	Cost of an adult weekday chairlift ticket
AdultWeekend	Cost of an adult weekend chairlift ticket
projectedDaysOpen	The projected number of days open in the upcoming season.
NightSkiing_ac	Total number of acres covered by light for night skiing.

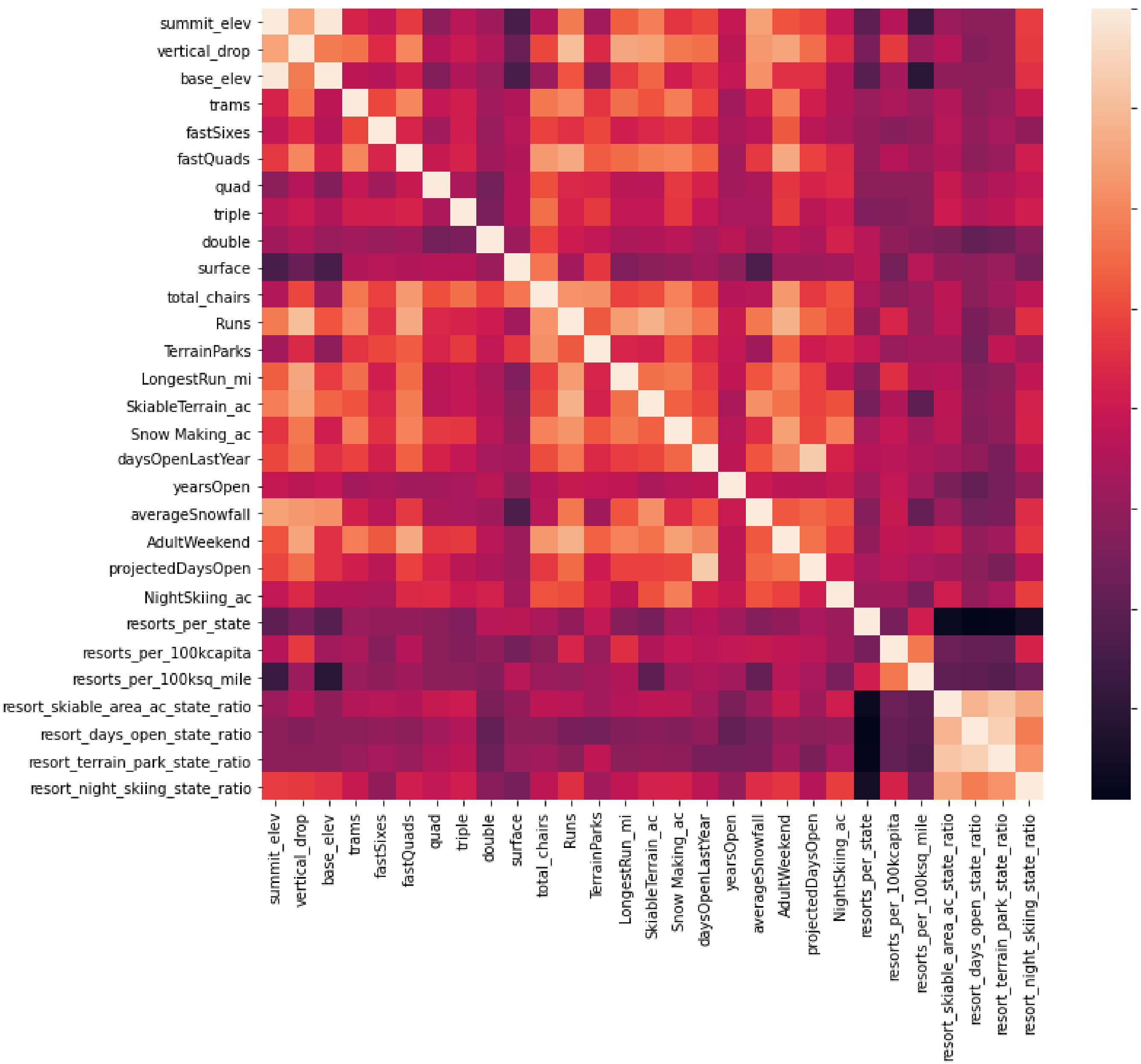
Appendix II

	count	%
fastEight	166	50.303030
NightSkiing_ac	143	43.333333
AdultWeekday	54	16.363636
AdultWeekend	51	15.454545
daysOpenLastYear	51	15.454545
TerrainParks	51	15.454545
projectedDaysOpen	47	14.242424
Snow Making_ac	46	13.939394
averageSnowfall	14	4.242424
LongestRun_mi	5	1.515152
Runs	4	1.212121
SkiableTerrain_ac	3	0.909091
yearsOpen	1	0.303030
total_chairs	0	0.000000
Name	0	0.000000
Region	0	0.000000
double	0	0.000000
triple	0	0.000000
quad	0	0.000000
fastQuads	0	0.000000
fastSixes	0	0.000000
trams	0	0.000000
base_elev	0	0.000000
vertical_drop	0	0.000000
summit_elev	0	0.000000
state	0	0.000000
surface	0	0.000000

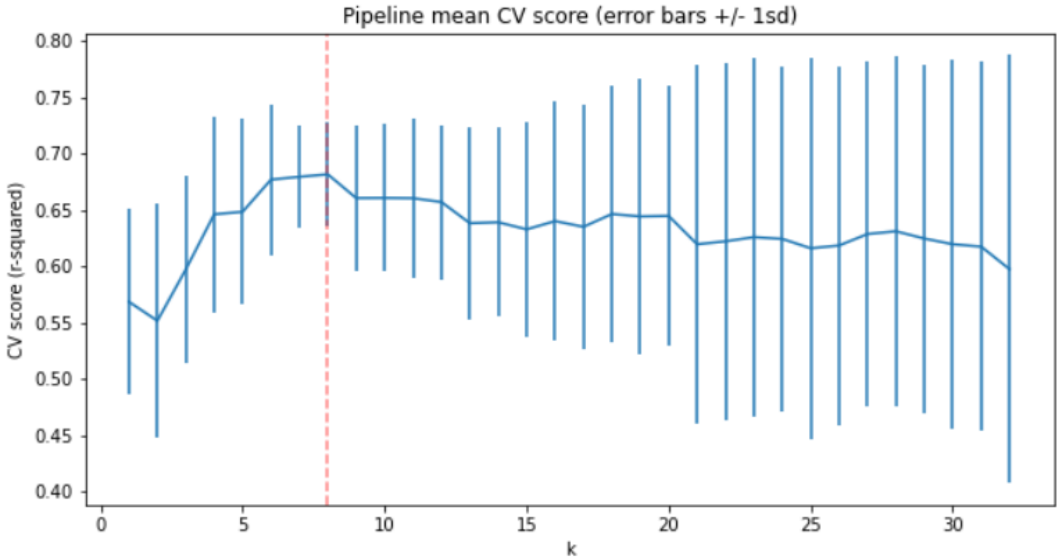
Appendix III

```
0    82.317073
2    14.329268
1     3.353659
dtype: float64
```

Appendix IV (Feature Correlation Heatmap)



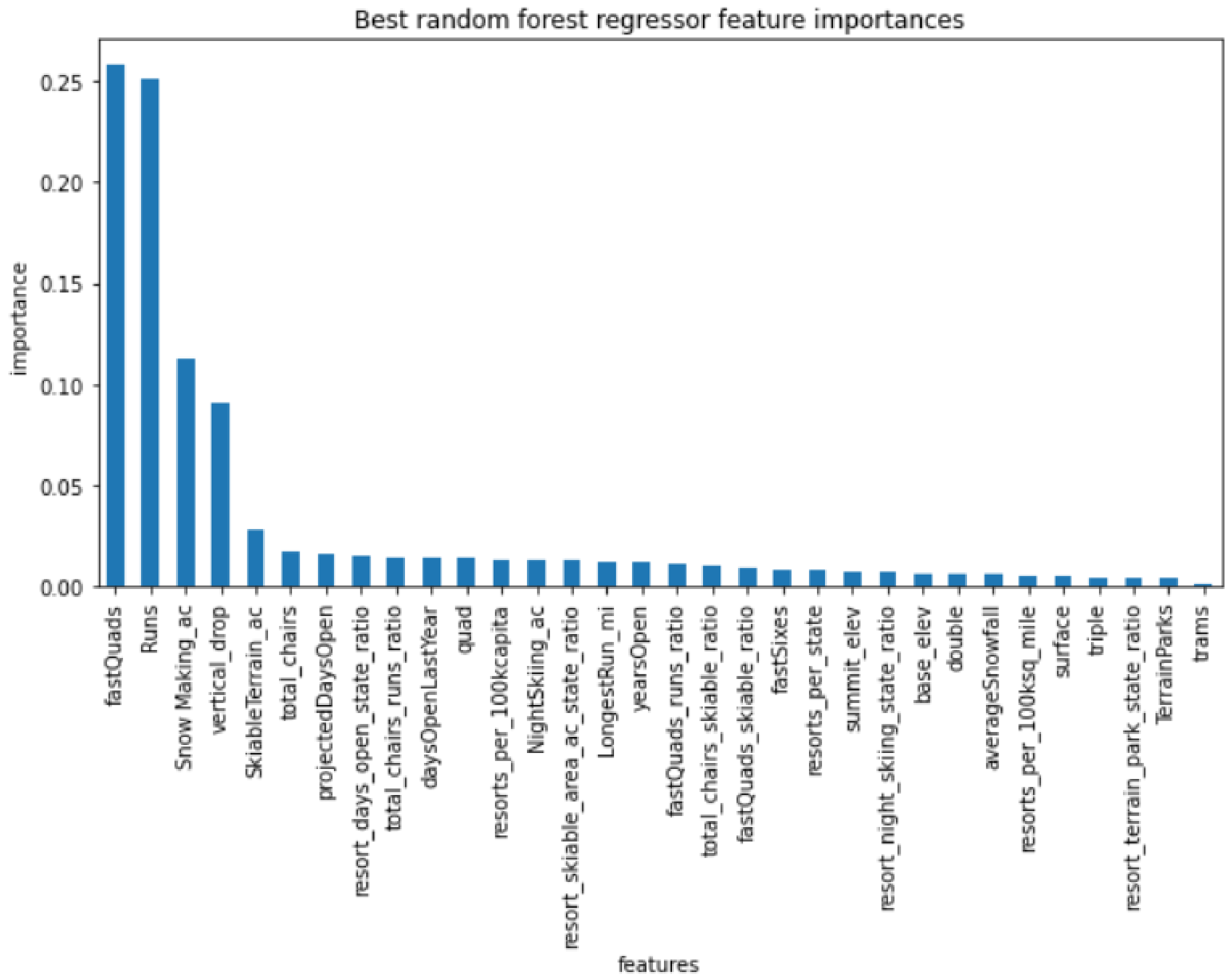
Appendix V



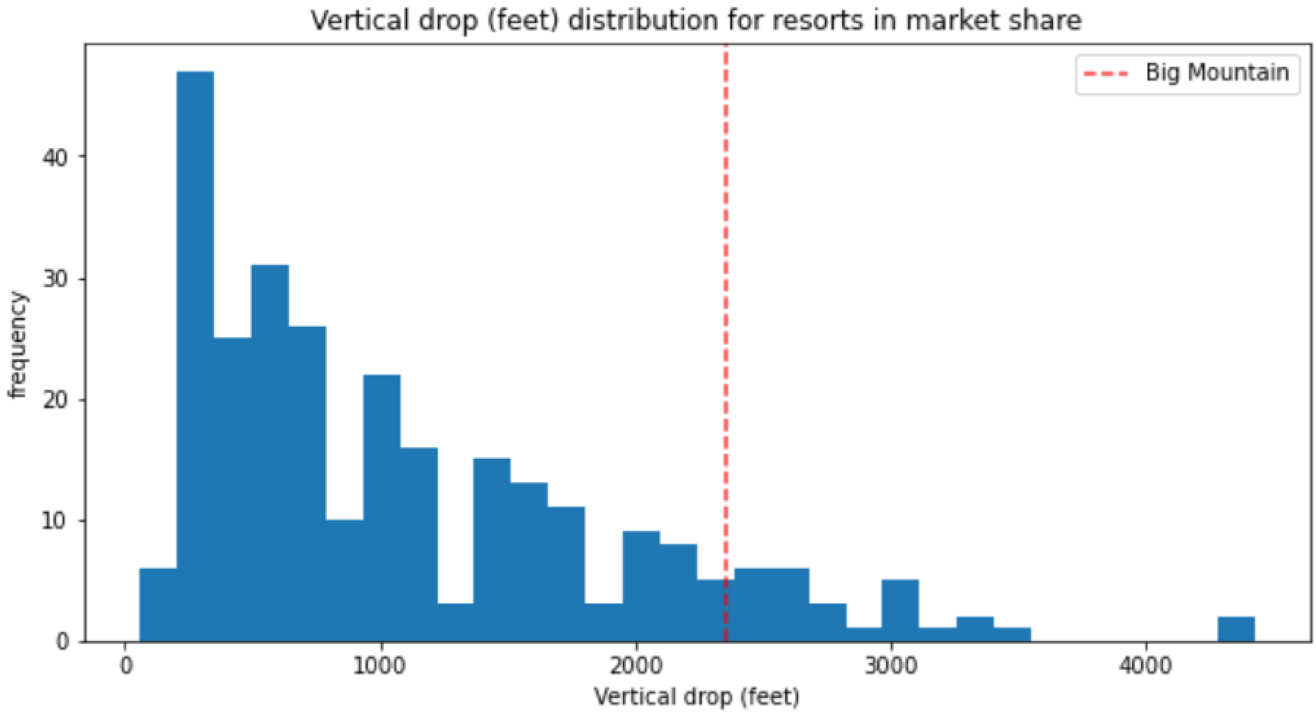
Appendix VI

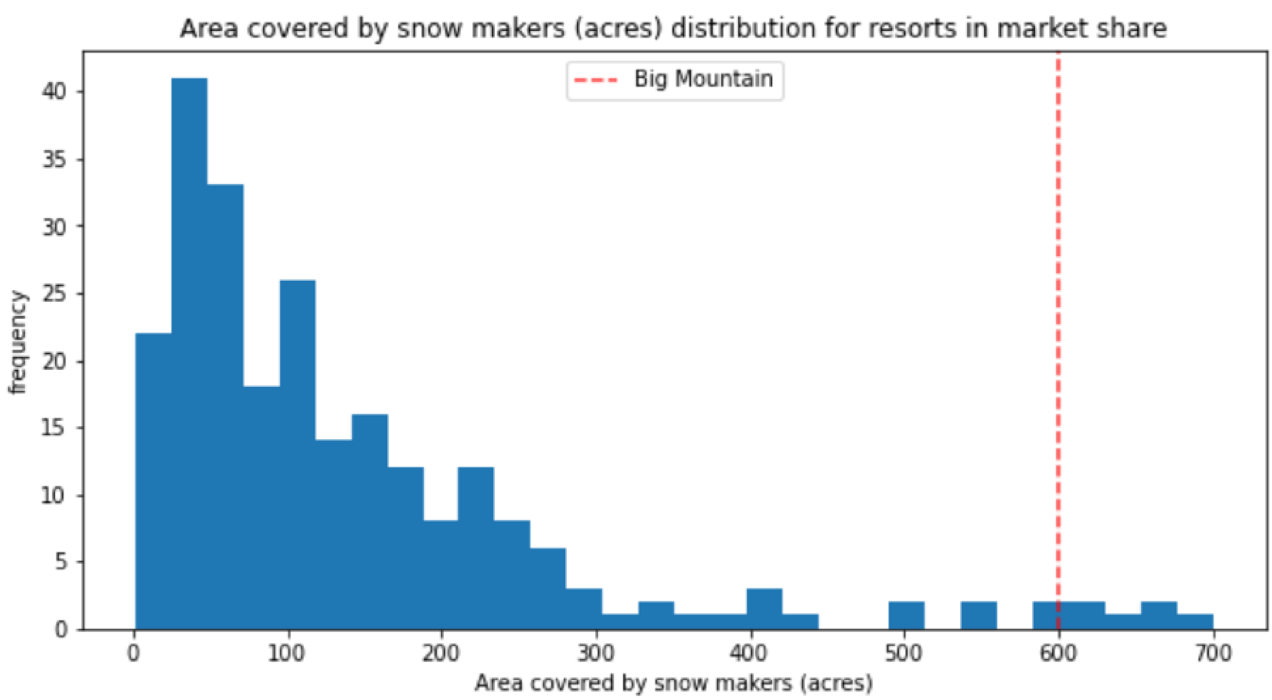
```
vertical_drop         10.767857
Snow Making_ac         6.290074
total_chairs           5.794156
fastQuads              5.745626
Runs                   5.370555
LongestRun_mi          0.181814
trams                 -4.142024
SkiableTerrain_ac     -5.249780
dtype: float64
```

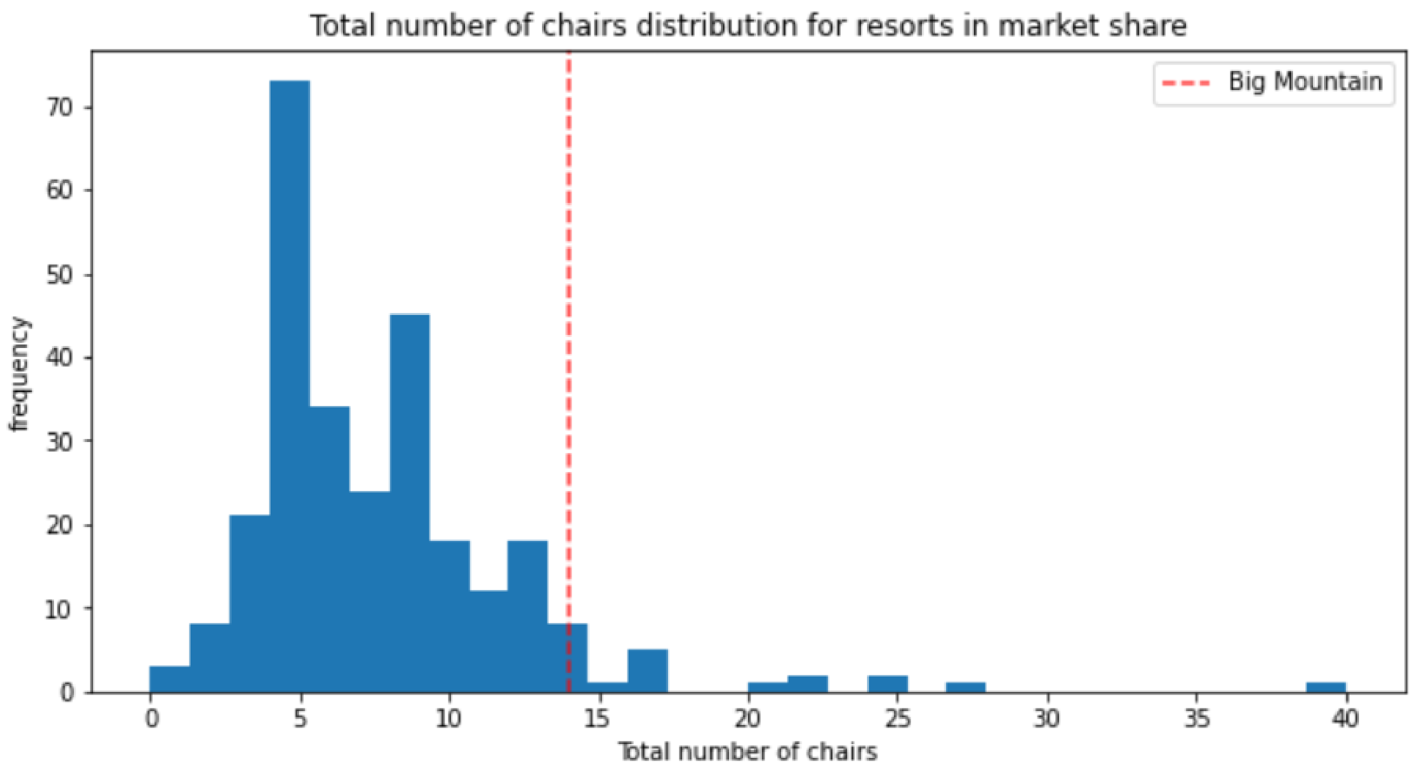
Appendix VII

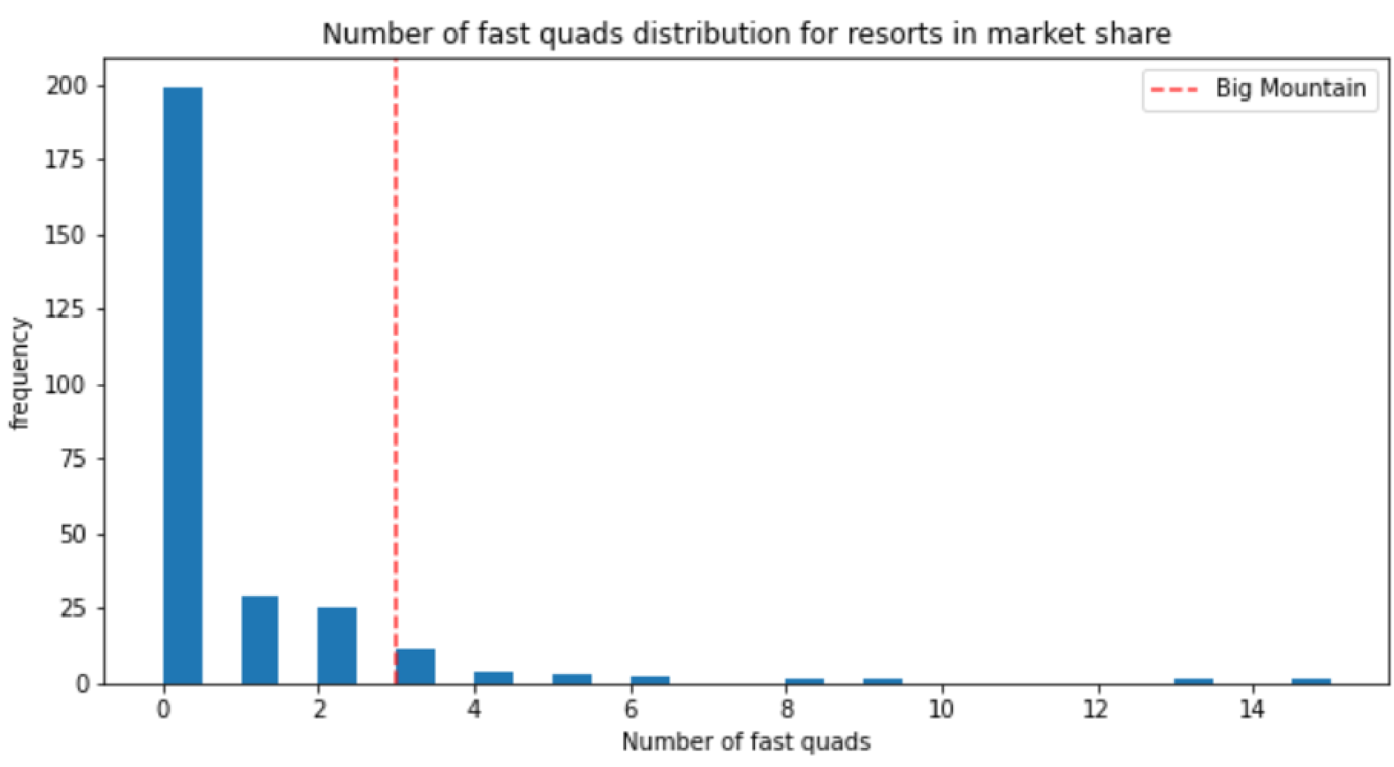


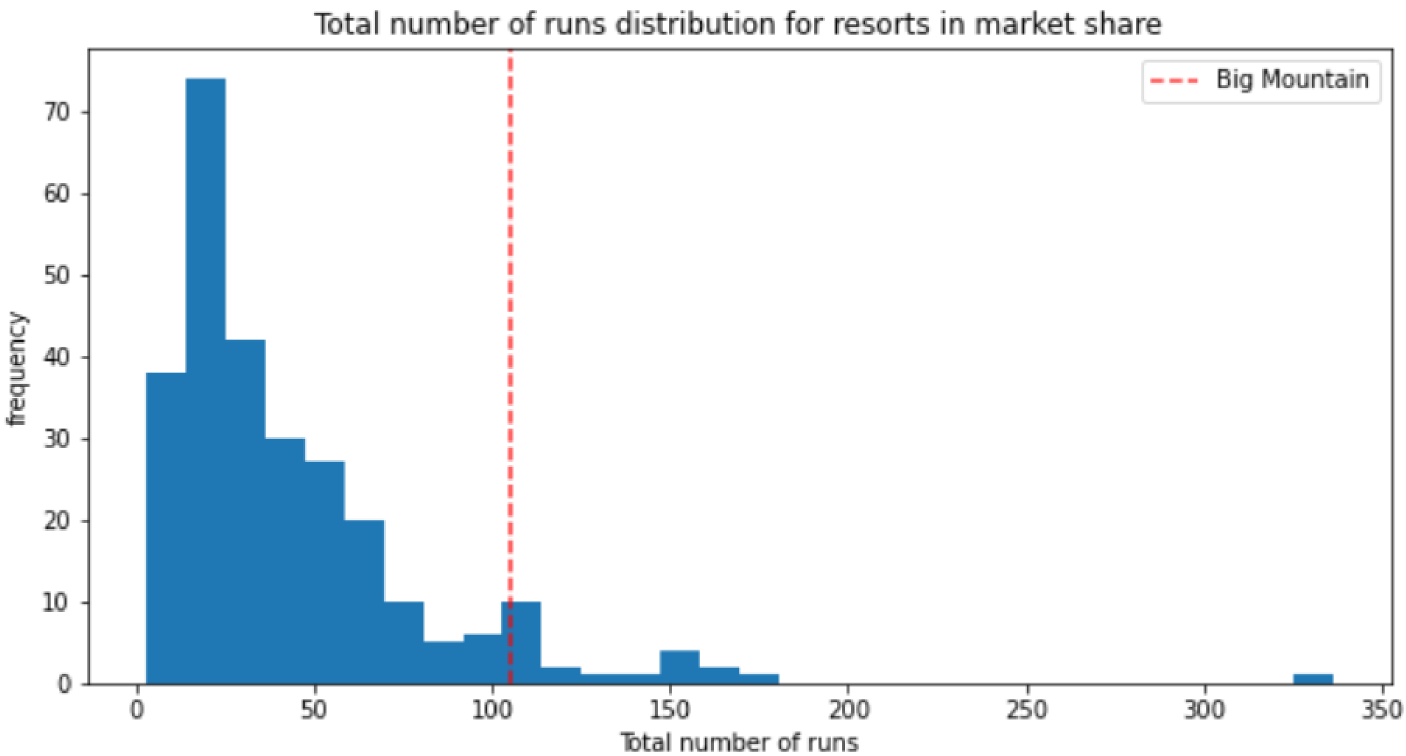
Appendix VIII

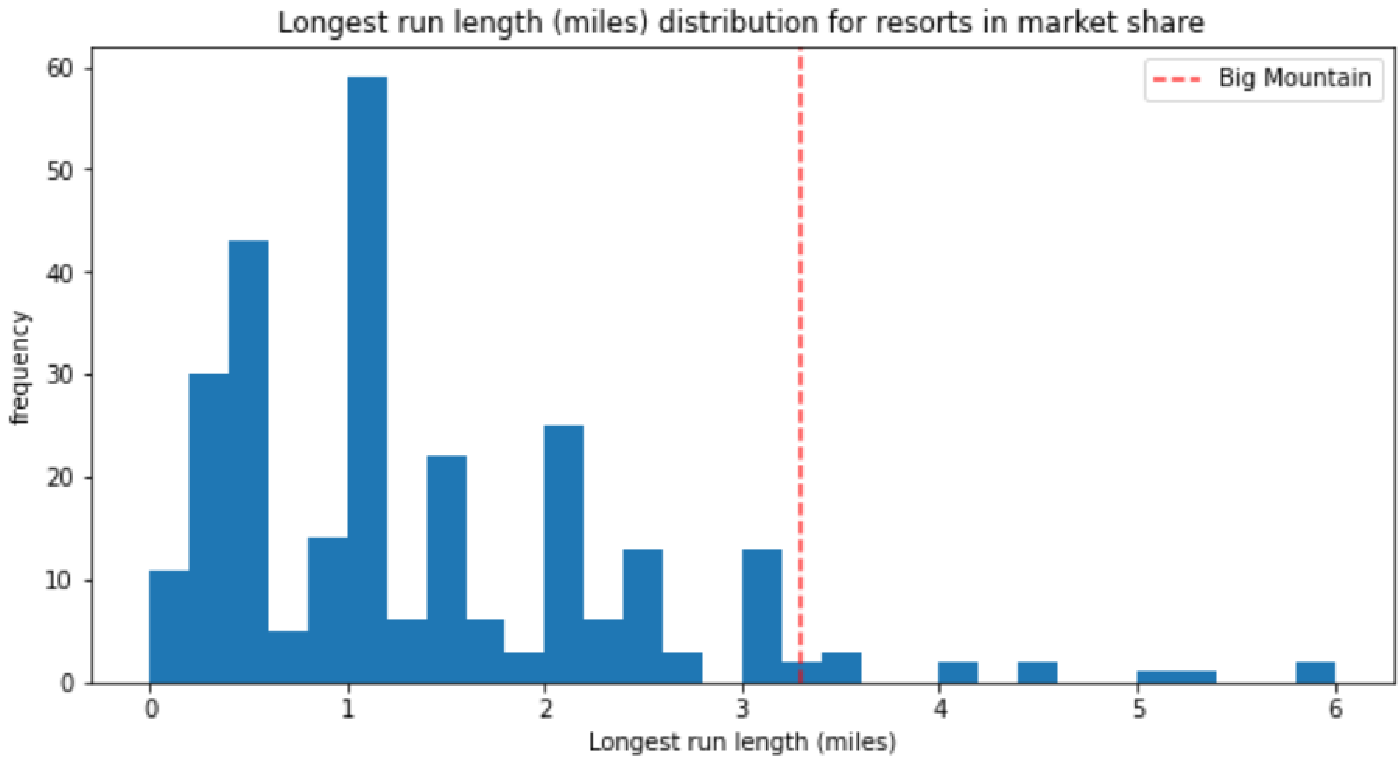


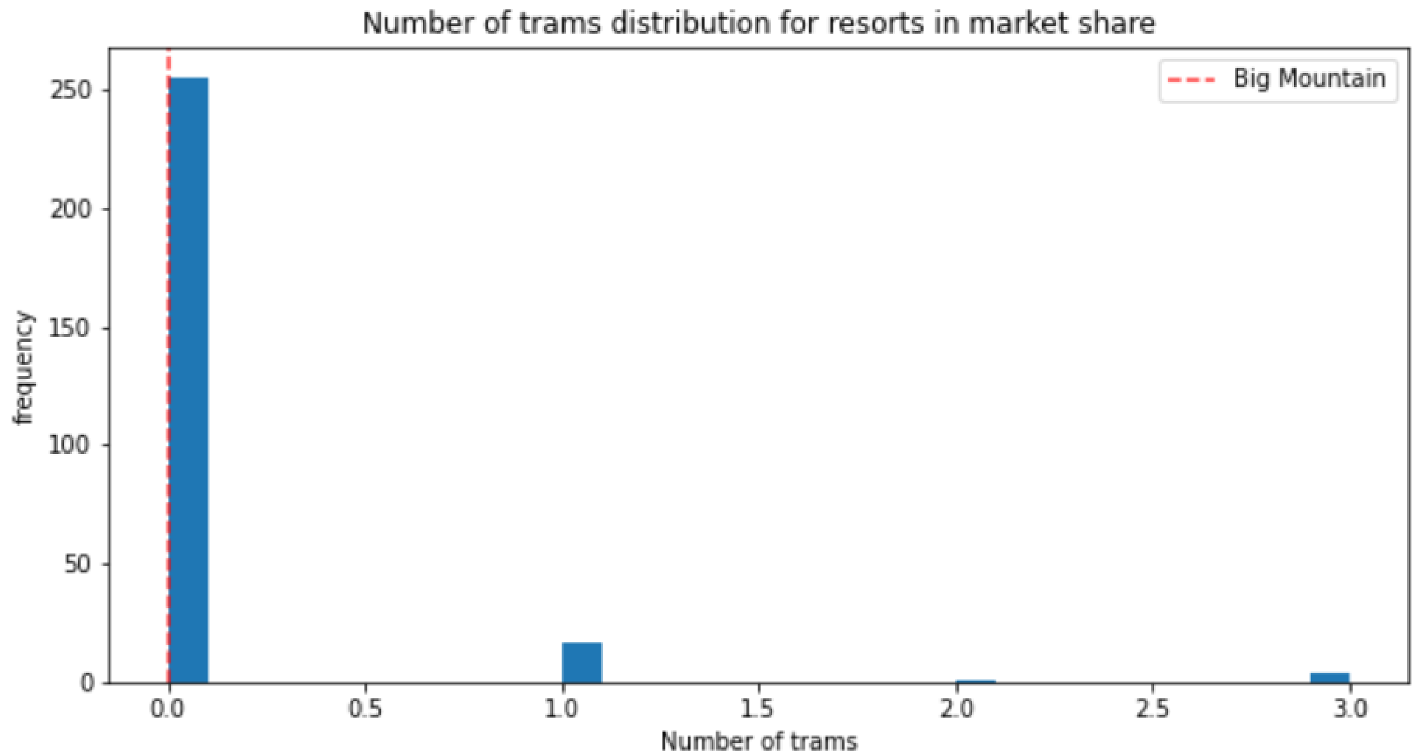


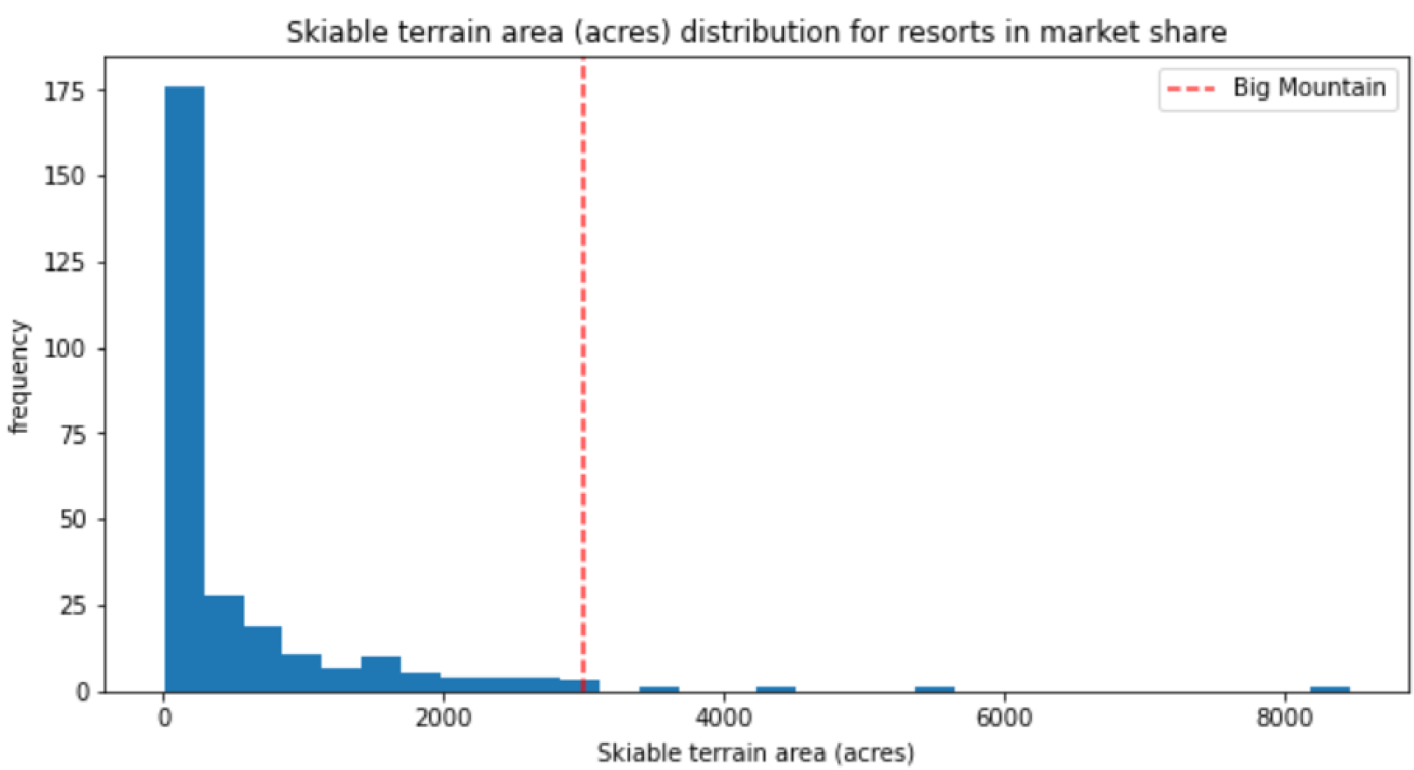

Total number of chairs distribution for resorts in market share
Big Mountain
frequency
Total number of chairs


Number of fast quads distribution for resorts in market share
Big Mountain
frequency
Number of fast quads


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


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
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)