Big Mountain Resort Price Analysis Report

Billie Kim

1 Introduction

Big Mountain Resort (BMR) is a ski resort located in Montana. It has a base elevation of 4,464 ft with a summit of 6,817 ft and a vertical drop of 2,353 ft. The resort is serviced by 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. The Hellfire is BMR's longest run at 3.3 miles in length. Recently, BMR has installed an additional chair lift to increase the distribution of visitors across the mountain, which has increased operating costs by $1.54m this season. Big Mountain Resort would like to adopt a data-driven business strategy to select a better value for their ticket price based on the importance of their facilities.

2 Problem Identification

How can Big Mountain Resort increase profits to offset costs by $1.54m this season through modeling ticket prices based on key features?

3 Data Wrangling

3.1 Data Exploration

We obtained the Ski Resort dataset in a CSV file from the database manager. It contained information on 330 resorts, including BMR, along with 27 features—these include categorical features such as state and region as well as numerical features such as elevation and total skiable area. We also obtained population and total area data for the US states from Wikipedia and added this to our state summary statistics data in order to create business relevant features such as the ratio of resorts serving a given population or a given area.

3.2 Data Cleaning

Before we move on to performing any type of analysis, we needed to address some house-keeping issues in order to make sure that we had good data to work with. This involves identifying missing, incomplete or inaccurate data and either completely removing them from our data set or correcting them if a viable or alternative solution existed.

Initially, we found several problems from exploring our data. We found that over 50% of the values in the fastEight column were missing while the other half were denoted with the value zero. Therefore, we dropped the fastEight column because we would not necessarily be able to produce any type of actionable insights from this column's information. Also, when we filtered the yearsOpen column to see if there were any irregularities, we did find a row that contained a value of 2019. Since it is highly unlikely that a resort had been opened for over two thousand years, we went ahead and dropped this row. We also found that about 14% of rows have no pricing data in both the AdultWeekend and AdultWeekday columns. Thus, we dropped the rows that did not contain any price information since ticket price is going to be the key feature that we will be using to build our model.

3.3 AdultWeekend vs. AdultWeekday

Our data set contained two types of ticket prices: AdultWeekend (weekend ticket price) and AdultWeekday (weekday ticket price). One of these two ticket prices will serve as our target feature price to build our predictive pricing model.

In order to get a better understanding of the relationship between AdultWeekday and AdultWeekend, we created a scatterplot (Figure 1). There is a clear line where weekend and weekday prices are equal, meaning that there were, in fact, resorts that charged the same price for weekend and weekday tickets. Montana, where Big Mountain Resort is located, is also a point on this line whereas it charges the same price for both weekend and weekday tickets. This information helps us to determine our preference for selecting our target feature. In addition, we also found that AdultWeekend had fewer missing data than AdultWeekday. Thus, we dropped the weekday prices and kept the data that contained weekend prices.

We determined that AdultWeekend will be our target feature going forward for modeling ticket price. At this point, our cleaned data contained 277 rows and 25 columns.

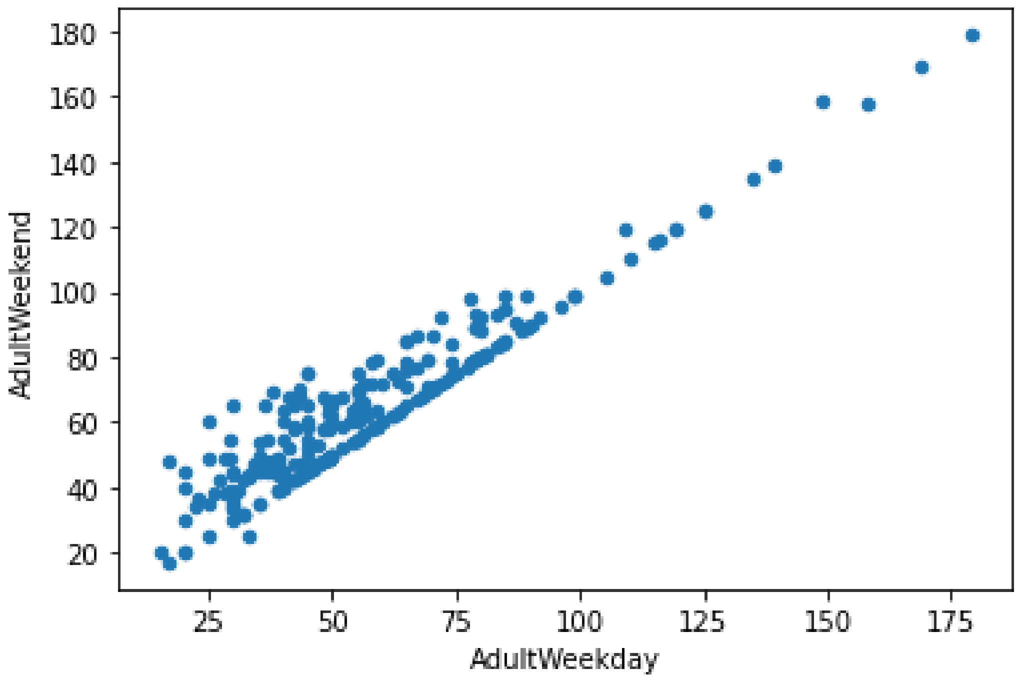

FIGURE 1: AdultWeekend vs. AdultWeekday

4 Exploratory Data Analysis

4.1 Summary Statistics

From our state summary data that we obtained from Wikipedia earlier on, we wanted to explore resort density in terms of the ratio of resorts serving a given population and the ratio of resorts a given area.

4.2 Principle Components Analysis (PCA)

Our state summary data consisted of 7 features. Using PCA, we visualized the data in order to figure out how much variance the representation explains. This allowed us to identify and extract the features from 7 dimensions down to 2 dimensions that are most important, while taking into account all 7 features.

When we fit the PCA transformation using our scaled data, we found that the first two components account for over 75% of the cumulative variance, and the first four account for over 95%. We plotted the first two principle components and labeled each point with the name of the state (Figure 2).This helped us to visualize the relationships between states based on features such as the total skiable terraiin area. We also included information about average ticket price by state to see if there were any notable patterns.

There is a spread of states across the first component. Vermont and New Hamshire are considered extreme in the second dimension while New York and Colorado could be extreme in the first dimension. We did not find any clear groupings or patterns in terms of price, which leans us towards the idea of treating all states equally and aim towards building a pricing model that considers all states together.

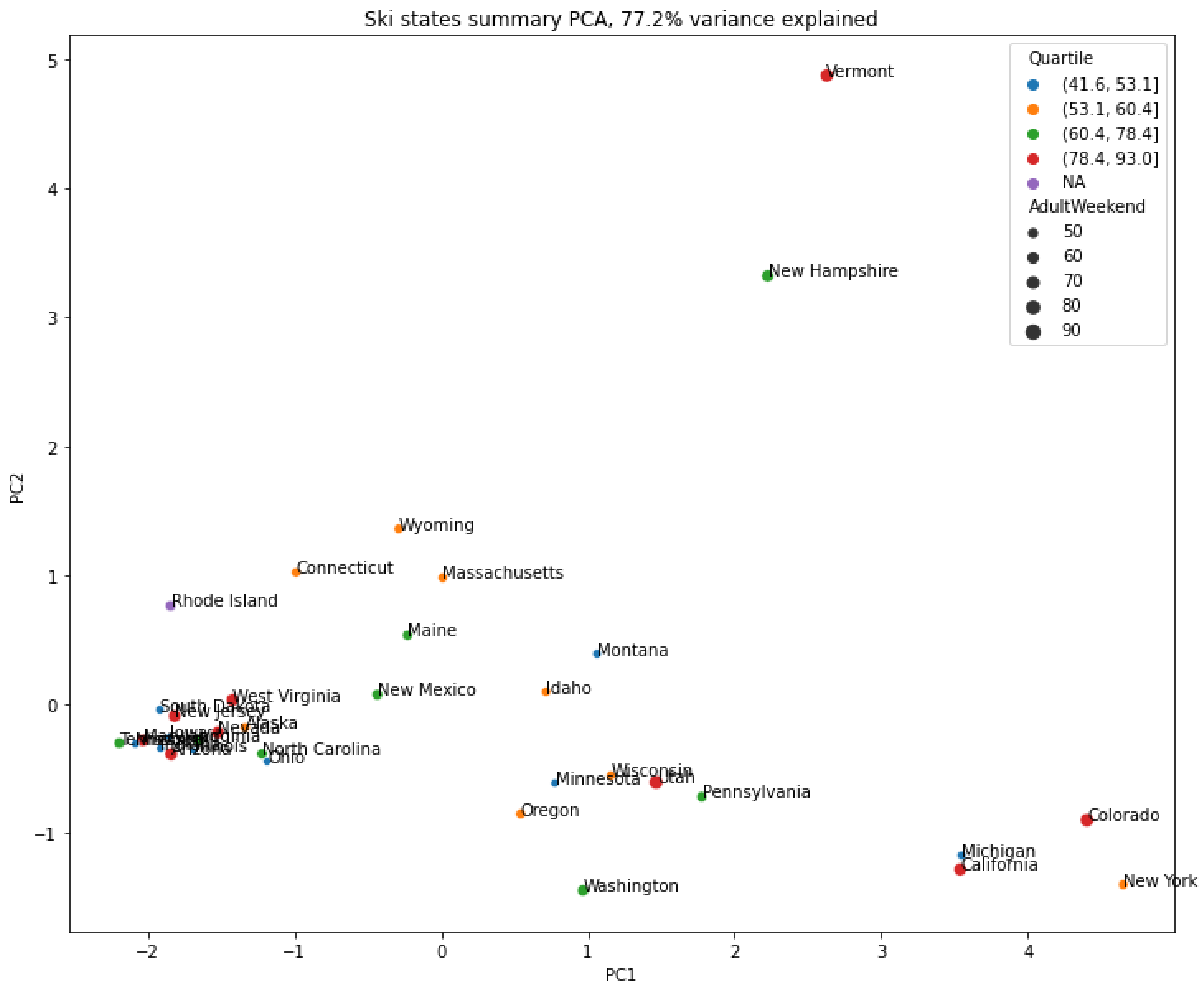

FIGURE 2: PCA2 vs. PCA1

4.3 Feature Engineering

Now, we wanted to explore resort-level data in more detail. We engineered a resort's share of the supply for a given state and created "state resort competition features"— these are ratios pertaining to resort skiable area to total state skiable area, resort days open to total state days open, resort terrain park count to total state terrain park count, and resort night skiking area to total state night skiing area.

In order to gain a high level view of relationships among our features, we created a feature correlation heatmap to identify patterns (Figure 3).

We found positive correlations between our target feature, AdultWeekend ticket price, and fastQuads, Runs, Snow Making_ac, total_chairs, vertical_drop, and resort_night_skiing_state_ratio.

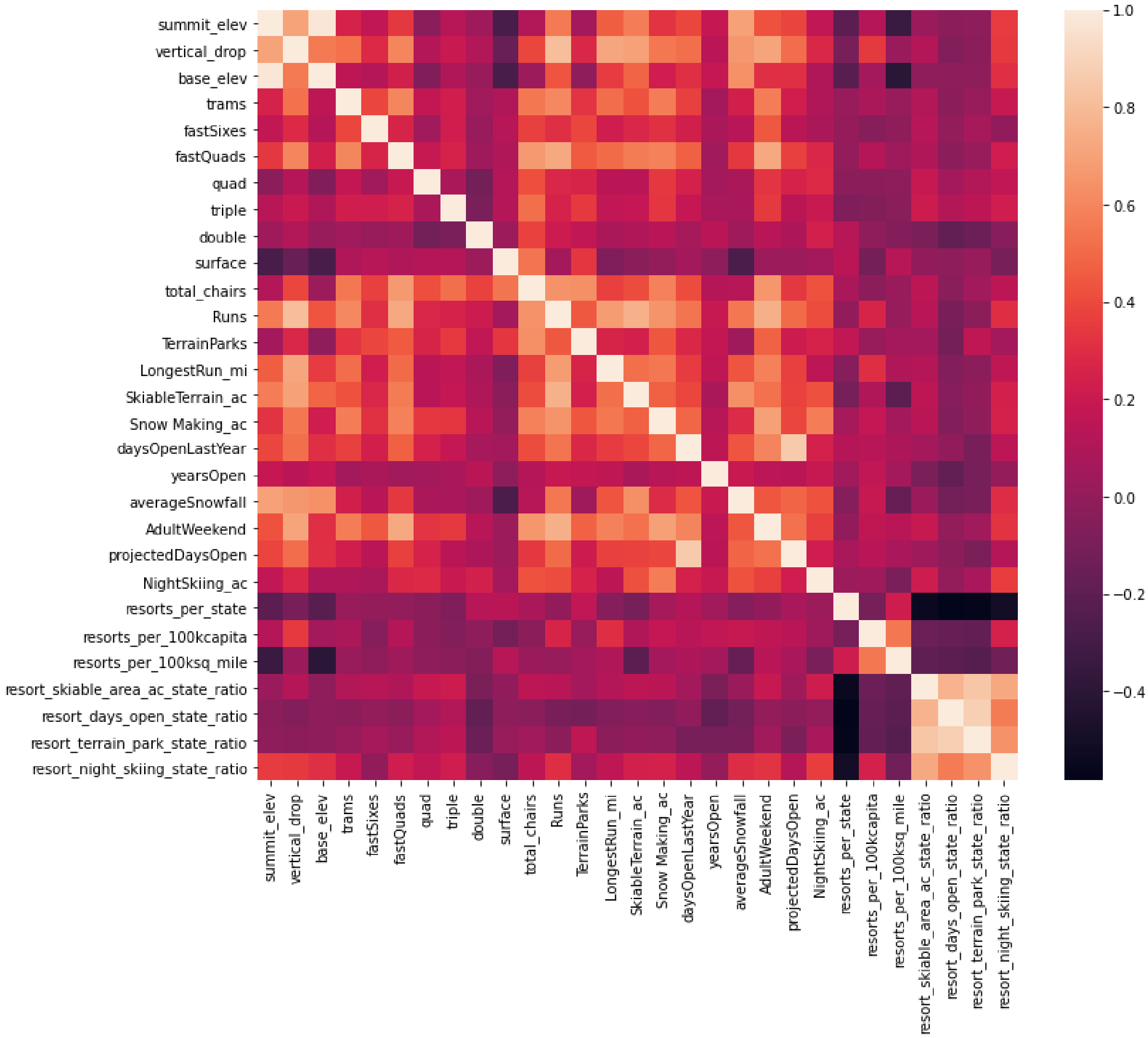

FIGURE 3: Feature Correlation Heatmap

5 Pre-Processing and Training Data

5.1 Mean as a predictor

We gained a baseline idea of performance by taking the average as the best guess for price. With a mean of 63.811 as a predictor, we calculated an r2 of -0.003 on our testing set. We also found a mean absolute error (MAE) of 19 which tells us that, on average, we might expect to be off by around $19 if we just guessed ticket price based on an average of known values.

5.2 Model with imputed median values

We built a linear model using median values to impute missing feature values. We chose to use the median rather than the mean to fill missing values because many of our predictor feature distributions were skewed. We found an r2 of 0.818 for our train set and 0.721 for our test set— which means that our simple linear regression model explains about 80% of the variance on the train set and about 70% on the test set. We calculated an MAE of 10.489— which means that by using this model with imputed median values, on average we'd expect to estimate a ticket price around $10 of the real price. This is a huge improvement from just using the average.

We created a pipeline and used cross-validation to select the best features to use that gives the best performance. We found that the best number of features to use was 8 features. We found that vertical_drop had the biggest positive feature associated with ticket price. We also found other strong positive features such as Snow Making_ax, total_chairs, fastQuads, and Runs. Interestingly, we found that both trams and SkiableTerrain_ac were negatively associated with ticket price for this model.

5.3 Random Forest Model

We created a random forest model and found that the dominant top four features were fastQuads, Runs, Snow Making_ac, and vertical_drop (Figure 4). These features were also consistent in the feature selection of our linear model.

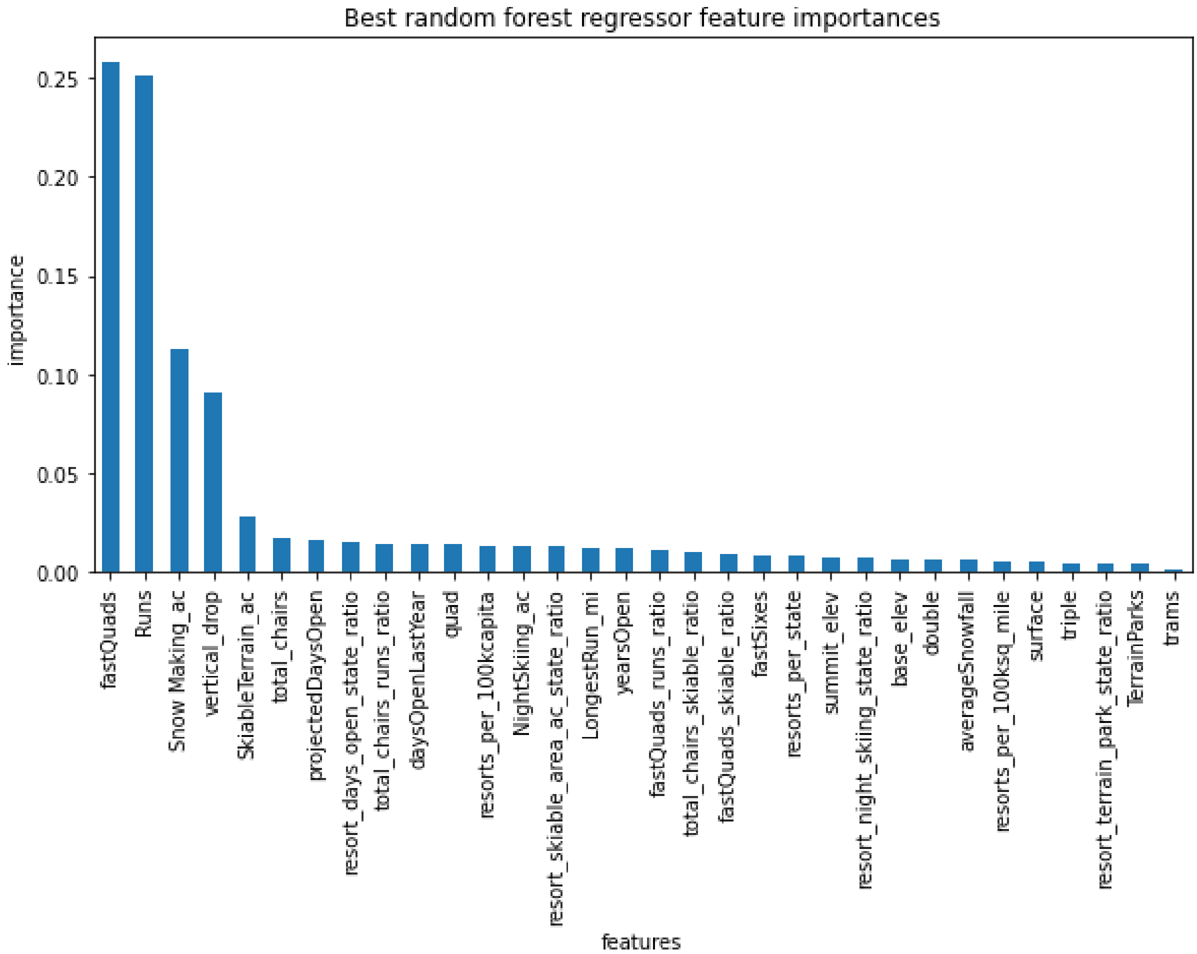

FIGURE 4: Barplot of the Random Forest's Feature Importances

Going forward, we chose to use the random forest regression model. The random forest model had a lower MAE of 9.645 than the linear regression model which had MAE of 10.499— the difference being almost a whole $1. In other words, this means that, on average, using the random forest model to estimate ticket price may be off by around $9 or so, while the linear regression model may be off by around $10 or so. The random forest model also exhibited less variability and verifying the performance on the test set produced consistent performance with the cross-validation results.

6 Modeling

6.1 Big Mountain Resort's Expected Ticket Price

We calculated the expected ticket price using our random forest model and determined it to be $95.87 while the current price is $81. Even with the expected mean absolute error of $10.39, this suggests that there is room for an increase.

We found that the features that came up as important in the modeling included vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams, and SkiableTerrain_ac. We examined where Big Mountain Resort sits amongst all resorts for price regarding these features and marked BMR's position with a red dotted line (Figures 6-15). Big Mountain Resort competes very well among all resorts and ranks high in all of the key features that determines price— with the exception of trams where majority of the resorts, including BMR, does not offer any trams.

6.2 Modeling Scenarios

At BMR's request, we modeled several potential scenarios for either cutting costs and/or increasing revenue— under the assumption that the expected number of visitors over the season is 350,000 visitors and, on average, visitors ski for five days.

1. Permanently closing down up to 10 of the least used runs
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
3. Same as number 2, but adding 2 acres of snow making cover
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

For scenario 1, the model predicts that closing down one run makes no difference in ticket price or revenue (Figure 5). However, support for ticket price and revenue drops successively from closing down 2 to 3 runs. The largest drop on ticket price is from closing down 6 or more runs. For scenario 2, our model estimates that this scenario increases support for ticket price by $1.99, which could amount to $3,474,638 over the season. We found that scenario 3 and scenario 4 will make no difference in ticket price.

7 Conclusion

Big Mountain Resort operates within a market where visitors pay more for certain facilities. Our analysis shows that BMR is a resort that places high up the league table in terms of the features that we found to be the important determinants of price.

Our pricing model suggests that BMR should increase its ticket price to around $95, which would increase profits by $5,204,500. This leaves BMR with an additional $3,664,500 in profit after taking into account the cost of the new chairlift.

BMR could also consider implementing scenario 2— which would add 150ft to its vertical drop with an extra run, but would also require an additional chair lift. Our model suggests that the combined effect of these two features would support an increase in ticket price by $1.99, with an expected revenue of $3,474,638 per season. This possible solution would take into account adding an additional chair lift on top of the one BMR recently added. The revenue generated from increasing the ticket price by $1.99 would be more than enough to cover the cost of 2 chair lifts with an excess profit of $394,638.

In addition, we would like to recommend BMR to explore the idea of closing down 1 run in order to decrease operating costs, since our model says that there would be no change in ticket price. However, we cannot provide any actionable insights on this area since we did not have any data on the operating costs of resort facilities.

The options that we've laid out, from utilizing our modeled price to implementing the modeled scenarios, could not only offset the cost of BMR's new chair lift, but also increase BMR's future earnings.

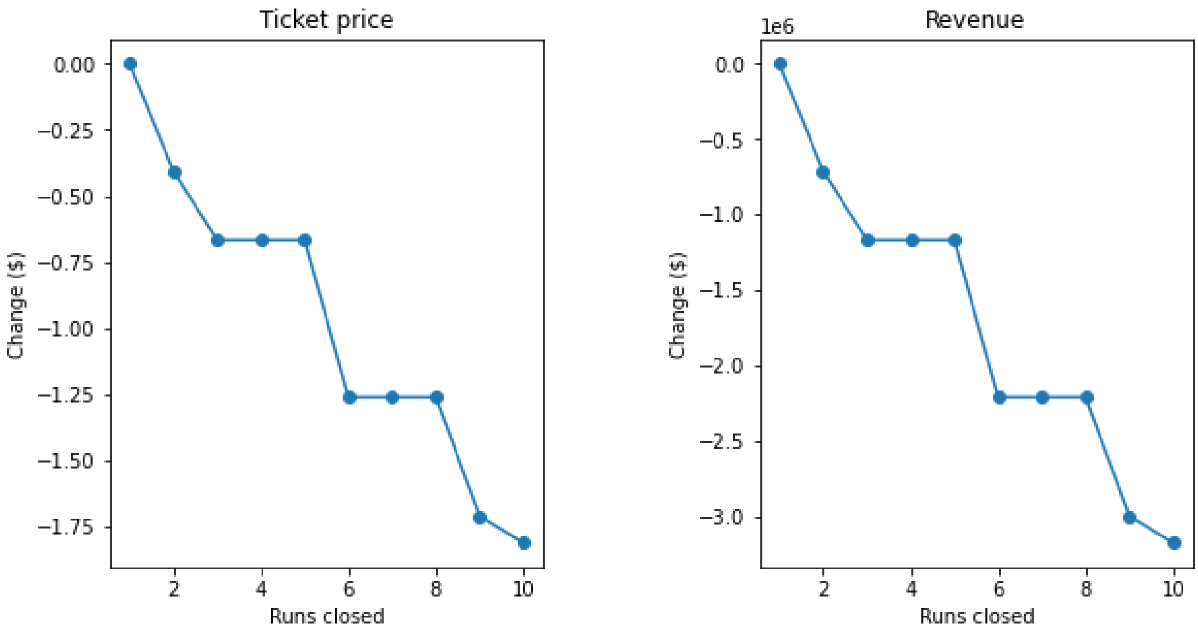

FIGURE 5: Predicted ticket price change for each number of runs closed

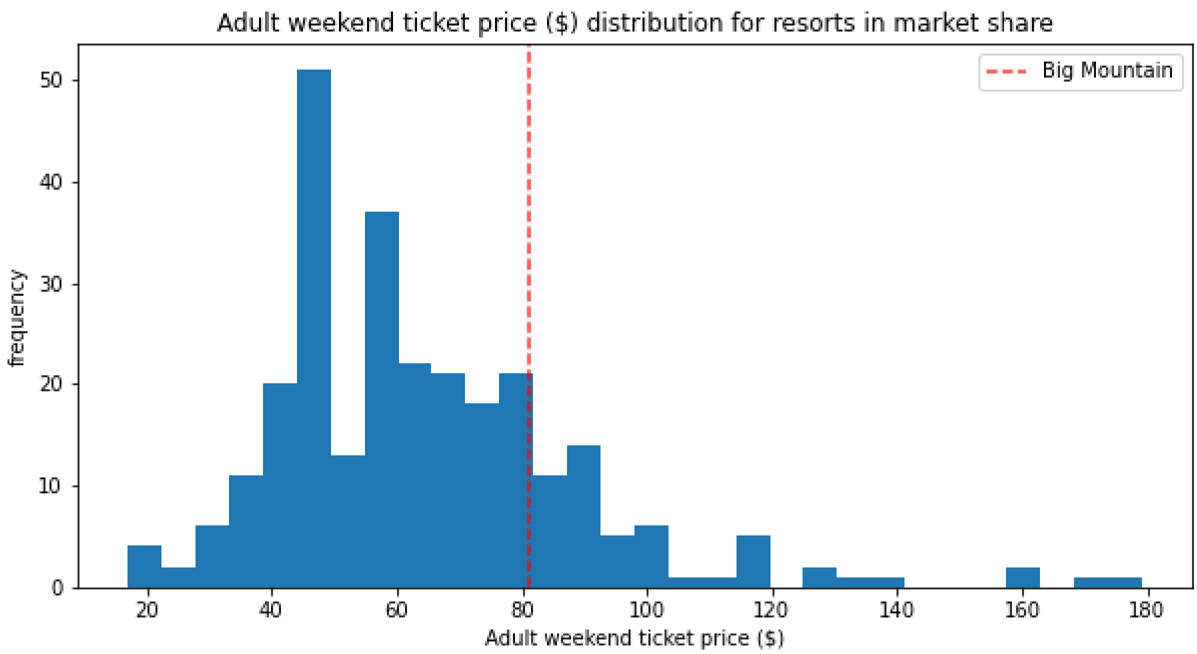

FIGURE 6: BMR's position for price among all resorts

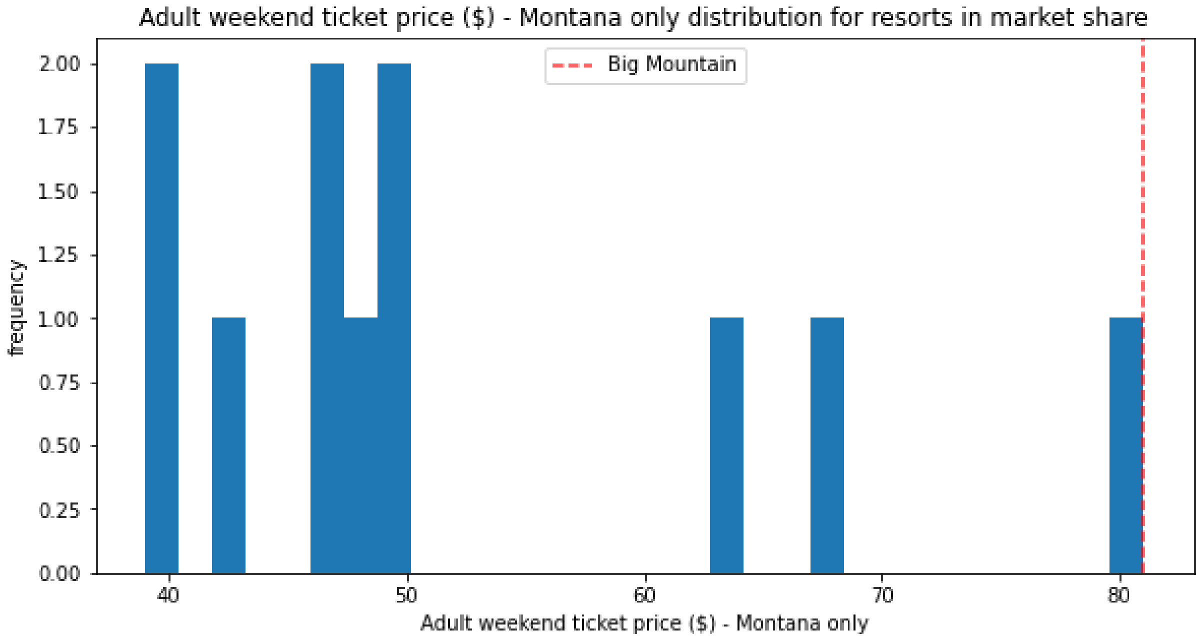

FIGURE 7: BMR's position for price for resorts in Montana

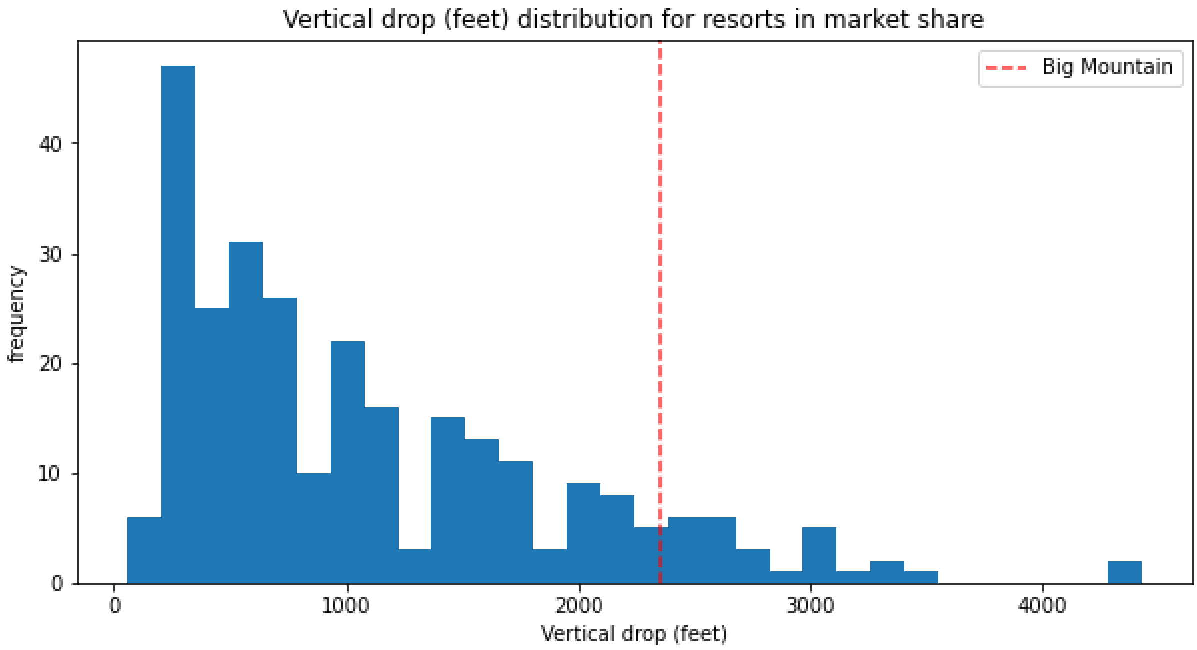

FIGURE 8: Vertical drop

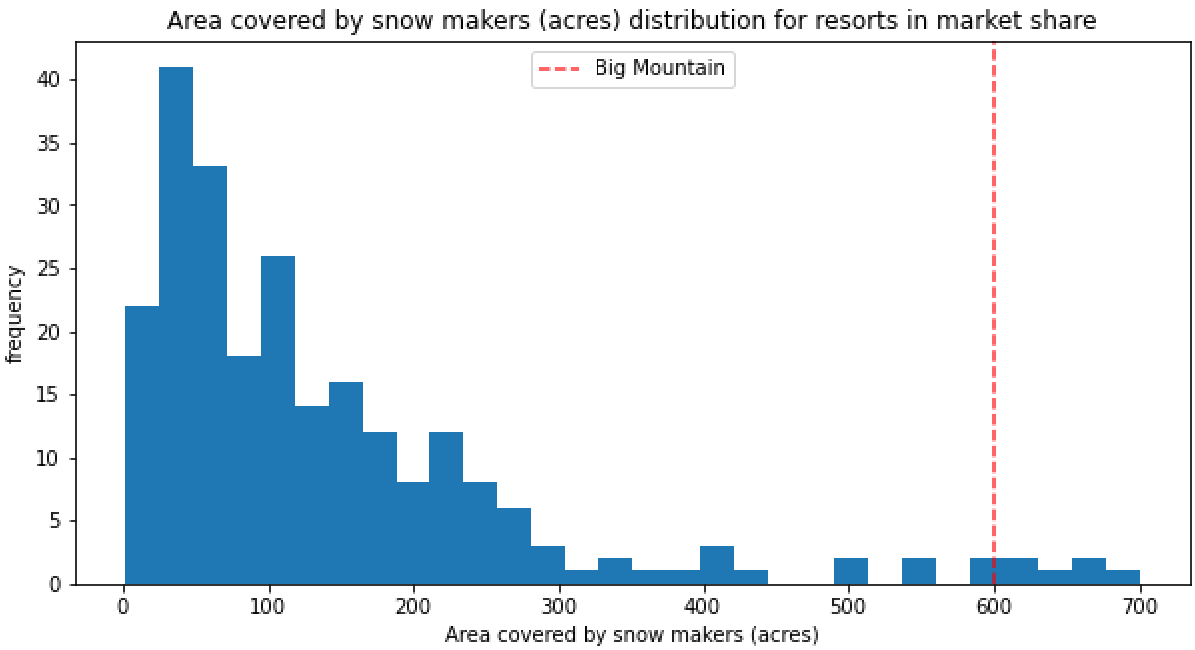

FIGURE 9: Snow making area

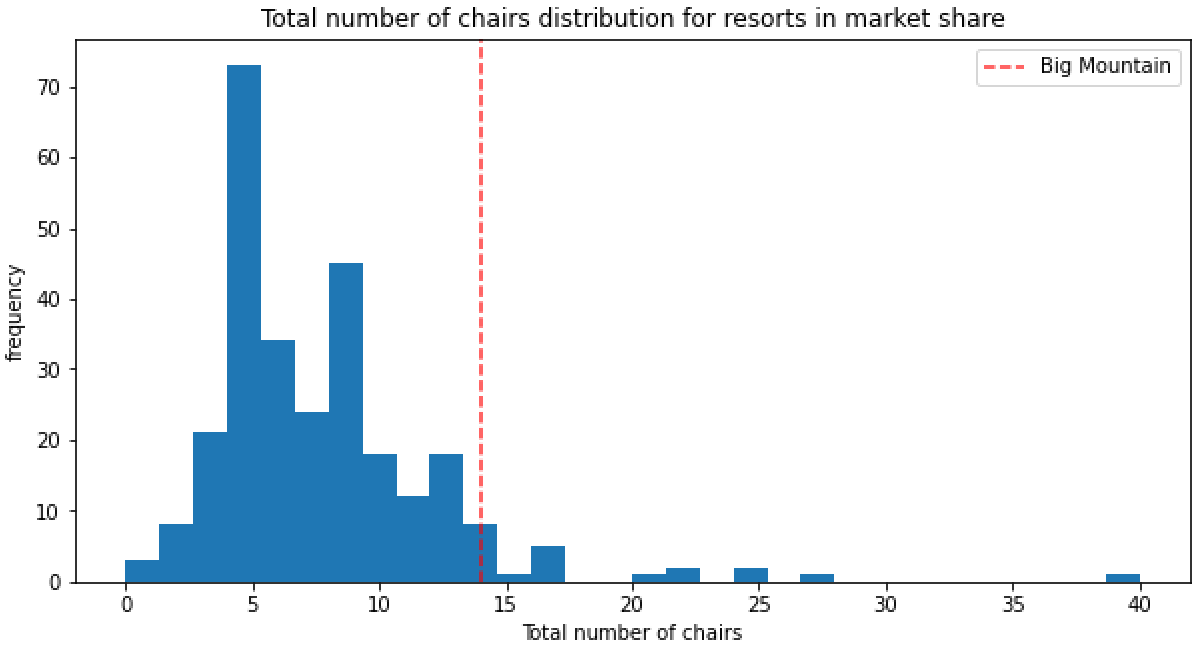

FIGURE 10: Total number of chairs

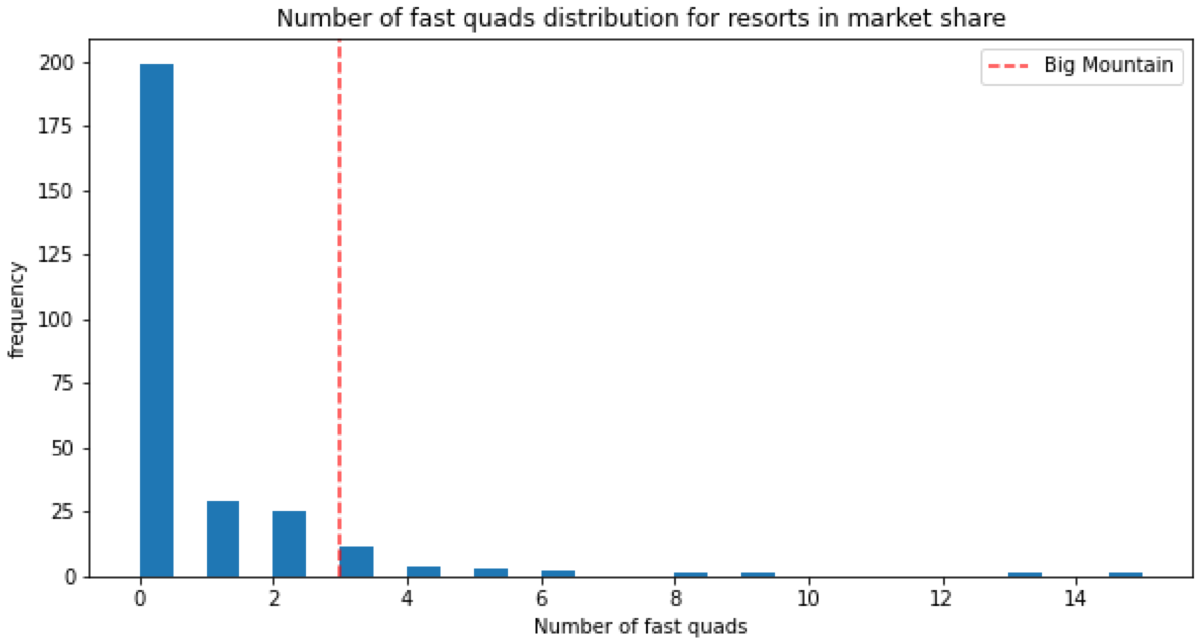

FIGURE 11: Fast quads

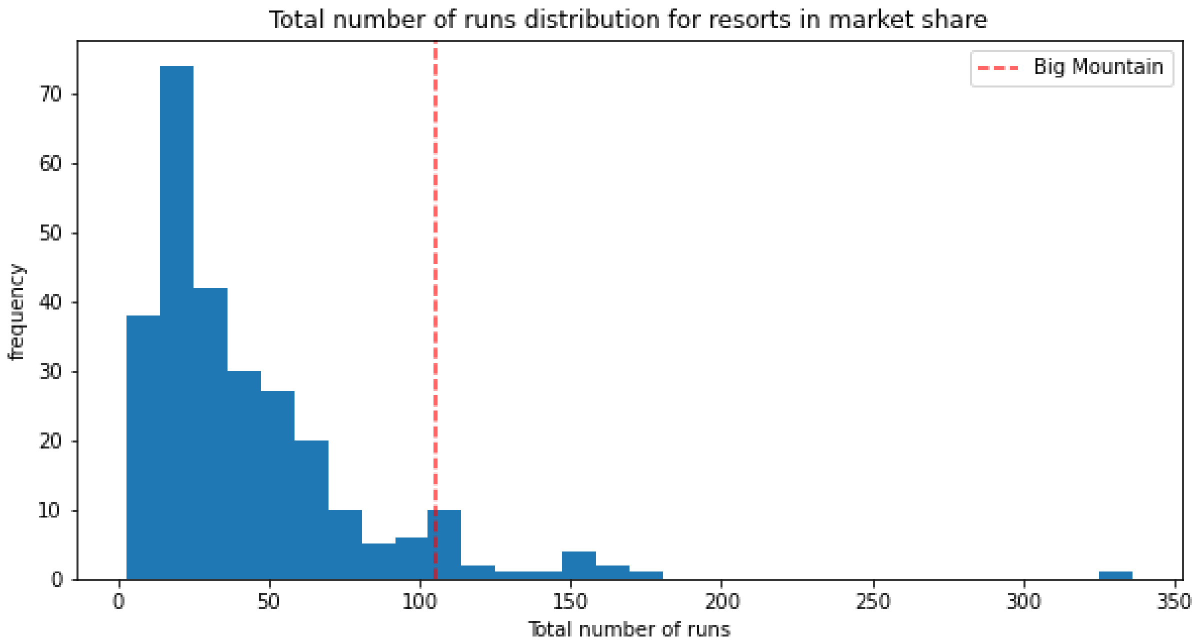

FIGURE 12: Runs

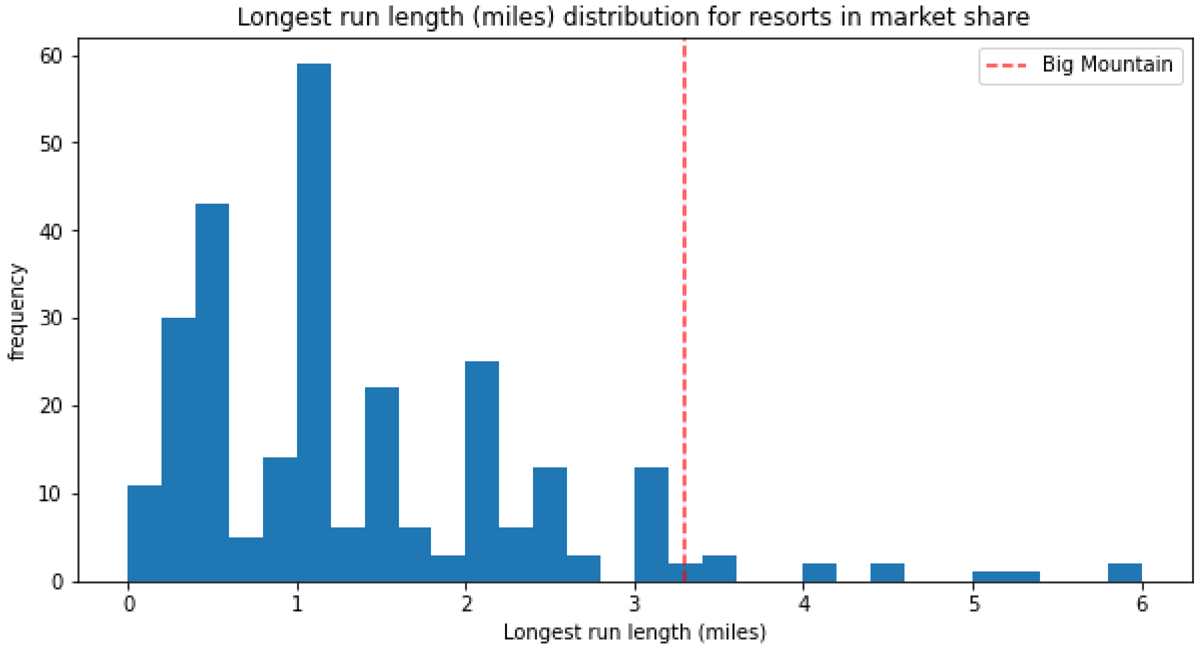

FIGURE 13: Longest run

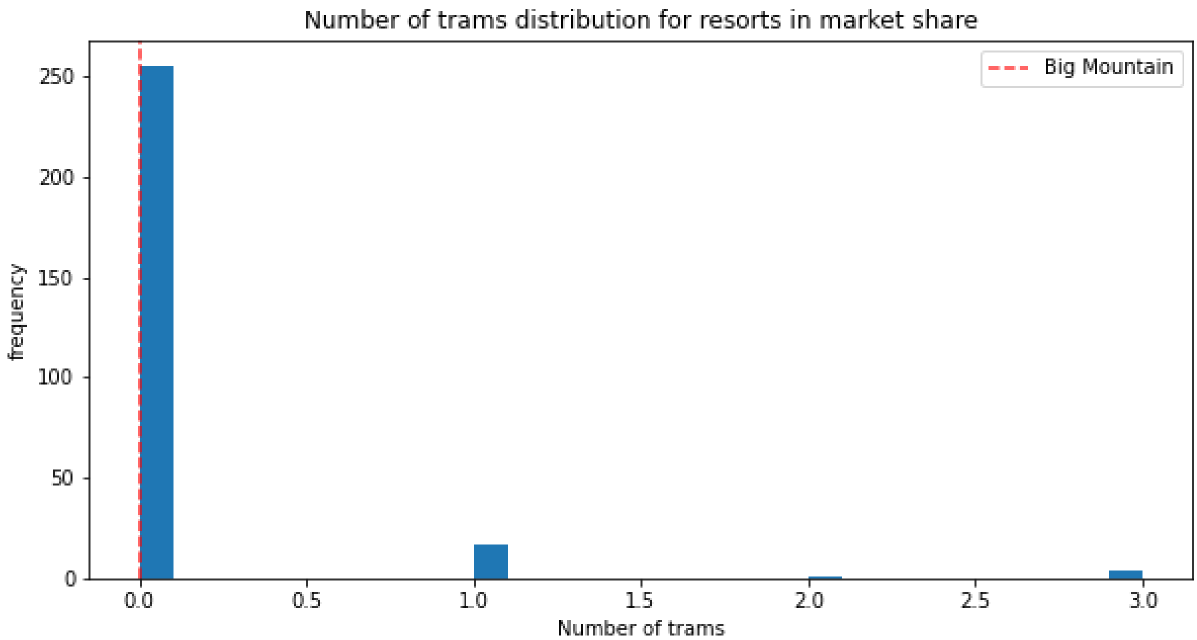

FIGURE 14: Trams

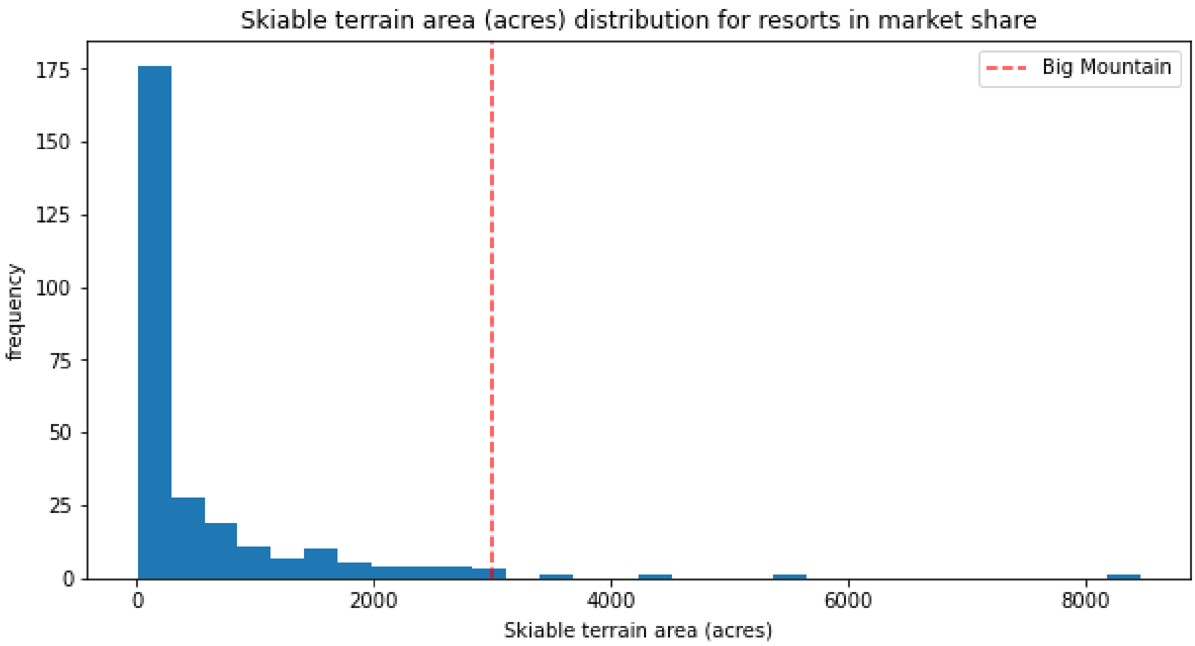

FIGURE 15: Skiable terrain area