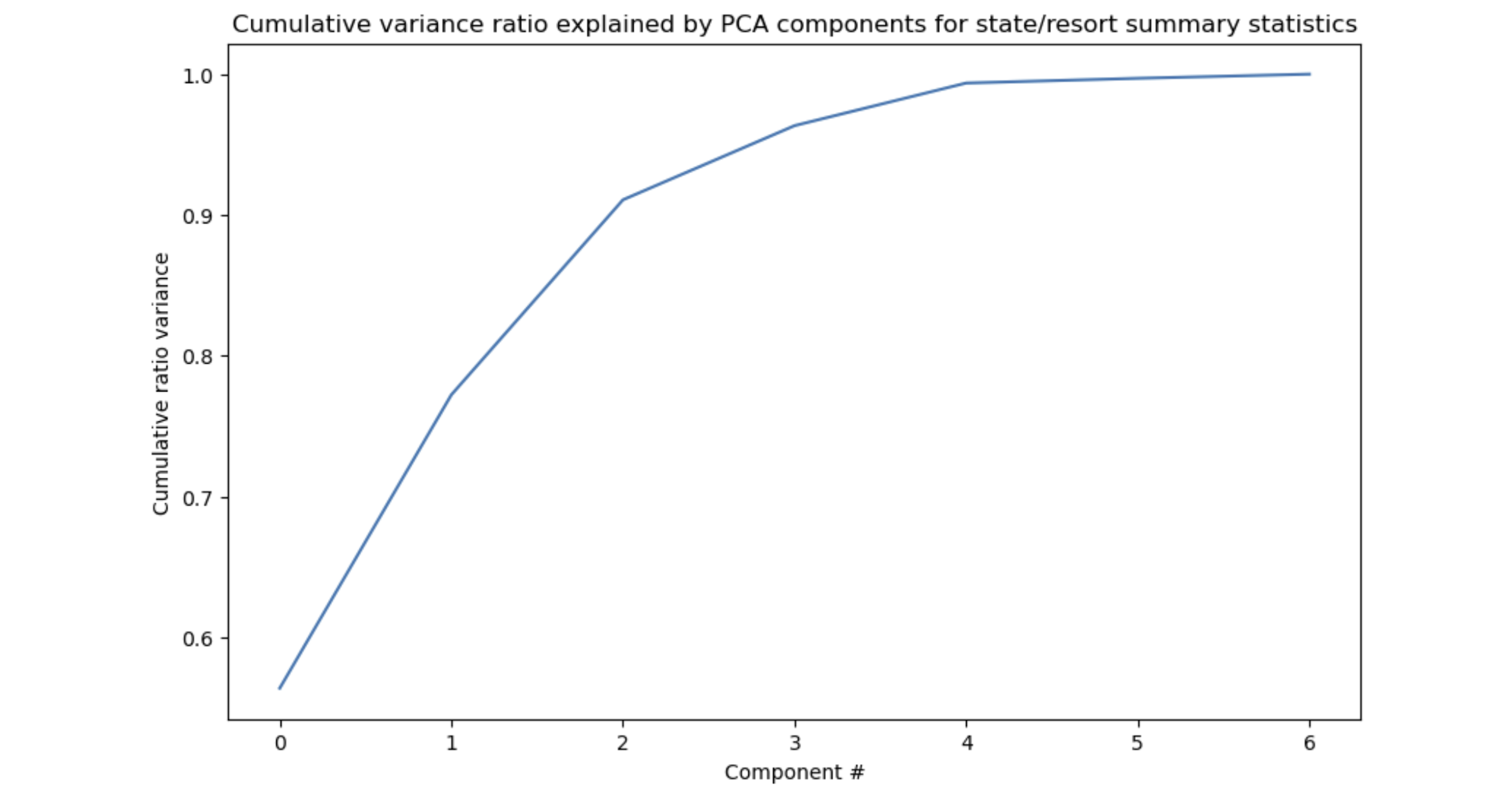
What opportunities exist for Big Mountain Resort to reduce costs or increase revenue by an amount equal to or greater than the added operating costs of an additional chair lift? This was the question asked by Big Mountain Resort, a ski resort located in Montana hosting 350,000 people annually who recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season. While exploring such opportunities, the resort was also curious if a better pricing strategy existed, one that gives the company a clear sense of what facilities are more important than others. This would make future investment plans easier to navigate. This is where our journey begins.

We loaded in a CSV file given to us by the database manager containing data for various features of multiple resorts in the market, including Big Mountain Resort. This dataset had numerous missing values, several repeated values, and a few extreme values that needed proper attention before advancing. Special notice was given to columns containing ticket prices. Some resorts were removed from the dataset completely for not possessing any ticket price information. It was determined that the target feature when modeling ticket price would be weekend ticket prices, so weekday prices were dropped all together. The data was further cleaned and prepared for exploratory data analysis (EDA).

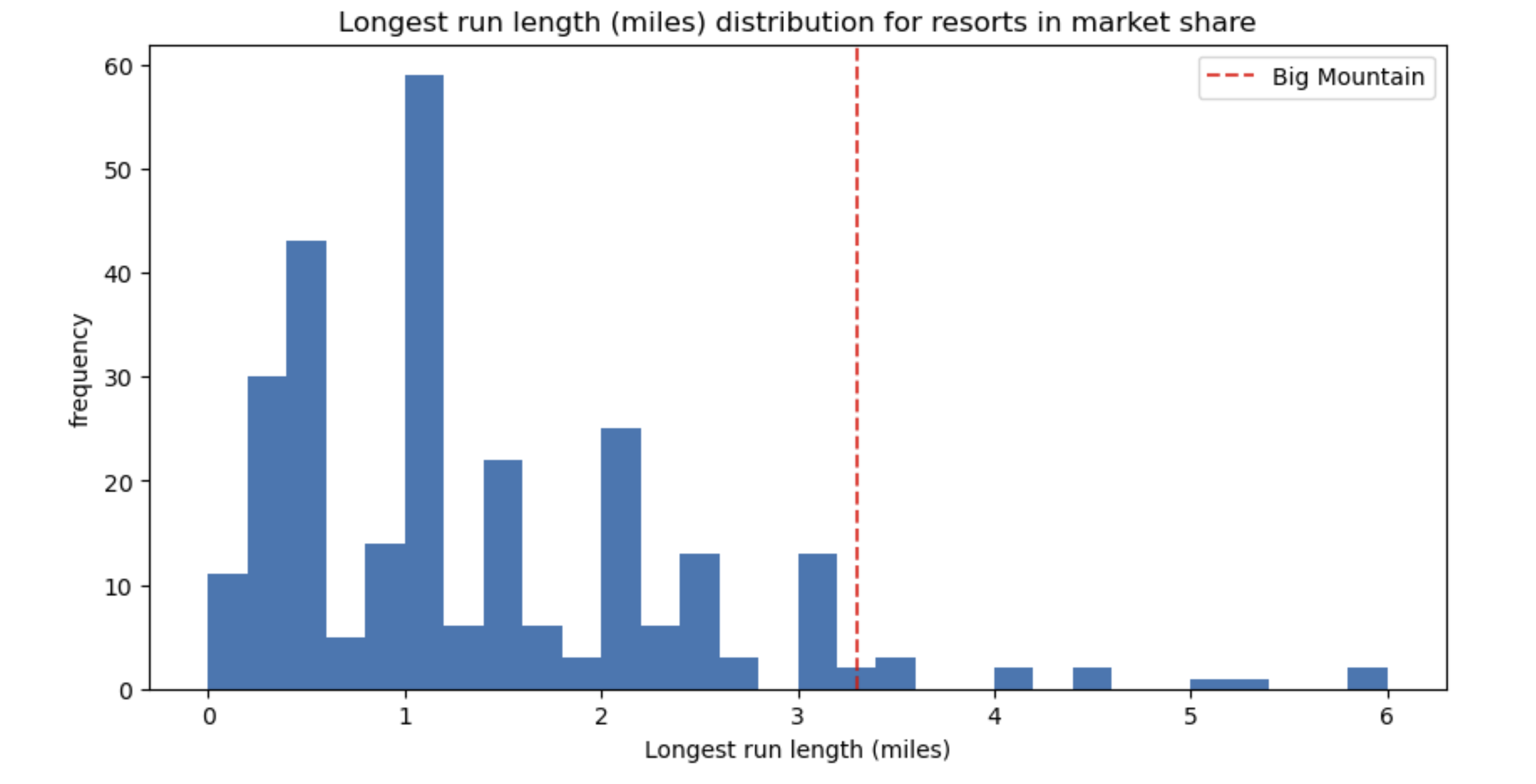
We used principal component analysis (PCA) during EDA to find how much of the variance in the data could be explained across different components for each state in the market. The first two components accounted for over 75% of the variance (Figure 1). We subset by these two components and used a scatterplot as a visualization. This visualization did not yield any conclusive results. We then merged our original dataset with summary statistics, creating new columns from the result. A heatmap was created to identify correlation between the variables in the resulting merge. Scatterplots were also utilized to get a better idea of correlation between variables and ticket price specifically. The features vertical\_drop, Runs, and total\_chairs displayed strong, positive correlations.



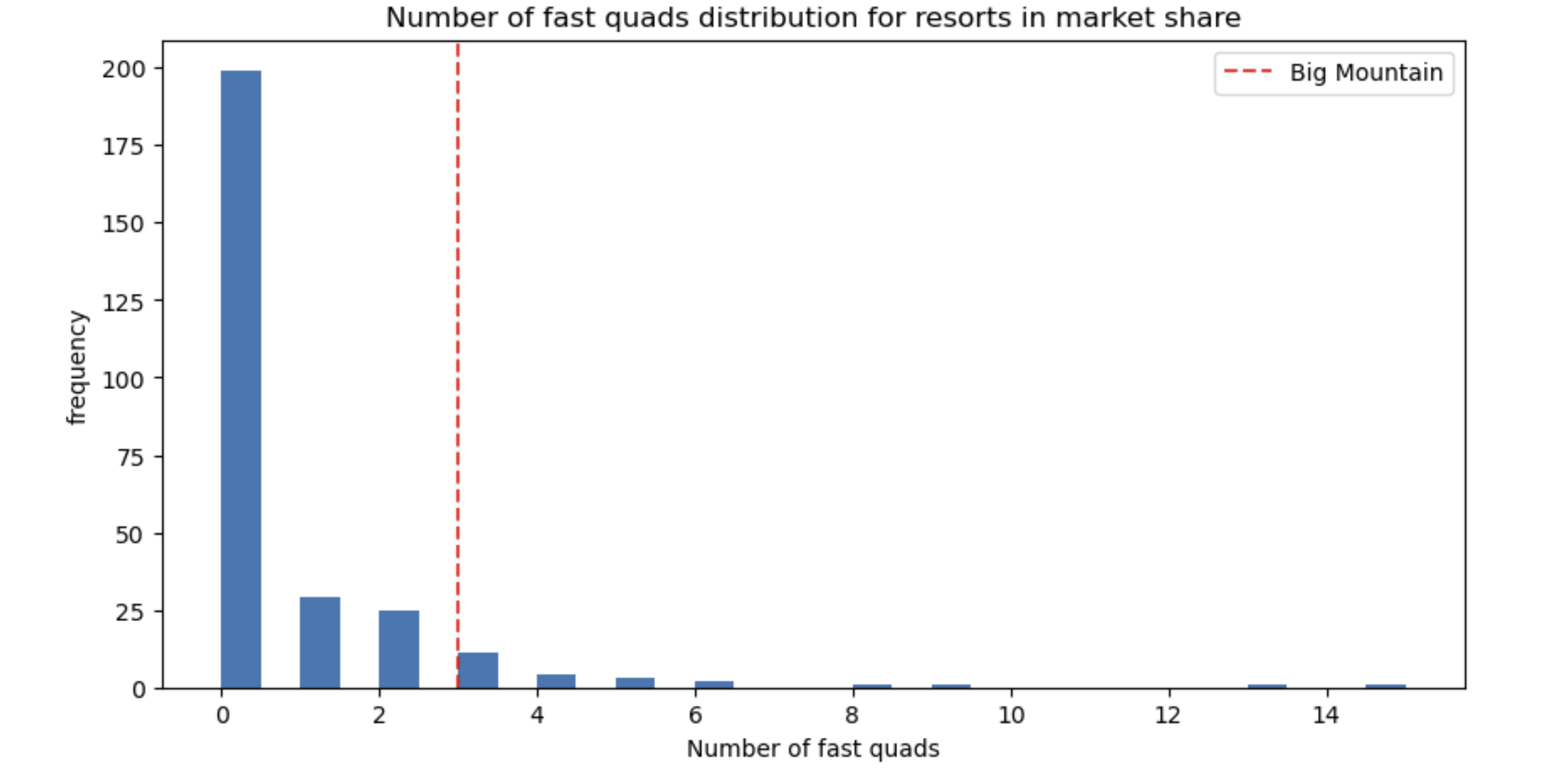
**Figure 1**

For the preprocessing and training portion, we initialized a 70/30 train/test split needed to evaluate a model's performance. We created an initial model that used median imputation. The data was then scaled, and linear regression was applied to predict y\_train and y\_test values. The model's performance was assessed by R-squared, Mean Absolute Error, and Mean Squared Error metrics. We created an identical model except mean imputation was used. The result was no different than the previous model. To make the entire process more efficient, the sklearn pipeline class was introduced. We refined our initial model by determining the k best features to use so we could exclude some features from our analysis. We used the default value of k=10, but this made our model worse. We increased k to 15 and the result was better. However, we needed another process aside from steadily increasing our value of k because we would only be improving our model for the given train data against the test data. This is when we decided to implement cross-validation. This process divides our train set into folds and leaves one fold aside so that we have something to assess our model with instead of our test set. We used the GridSearchCV function to utilize cross-validation and determine what is the best k to use (which was 8). Thanks to attributes provided to us in the GridSearchCV function, we are also able to identify the names of the k best features. vertical\_drop was at the top of the list as well as Snow Making\_ac. The same process was repeated using a random forest model. The same top features present in our linear model were present in the random forest model. With a mean absolute error of roughly 9.54, the random forest regression model was deemed better than the linear regression model which had a mean absolute error around 1.79.

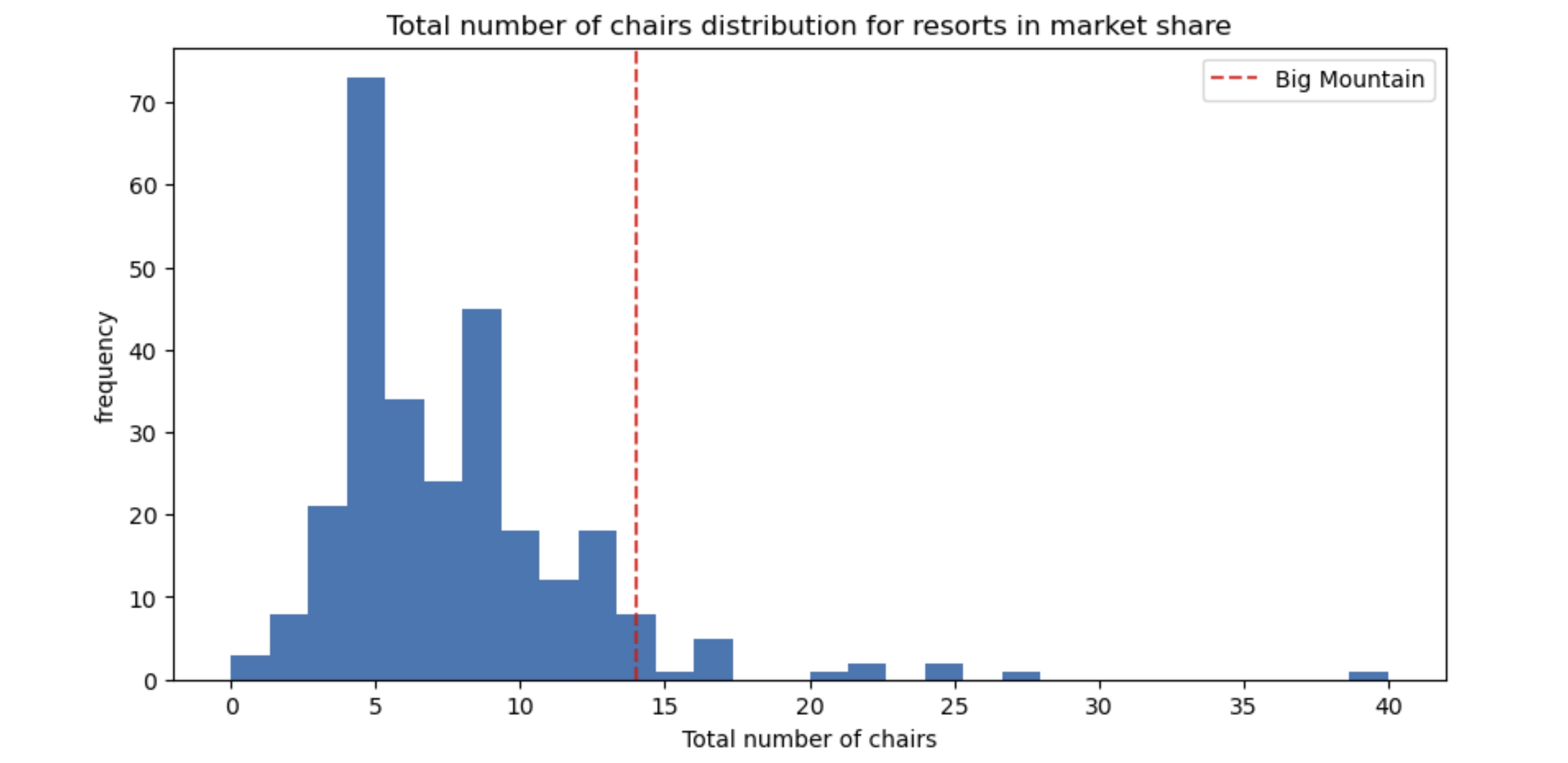
The mean absolute error for the market excluding Big Mountain Resort was about $10.39. The model was then fit on data for Big Mountain Resort exclusively. The modeled price for the company was $95.87, but the actual price is $81.00. This modeled price is well above the limits set by the mean absolute error and should be referenced when suggesting an increase in ticket price. From a list of features that appeared frequently in our EDA and preprocessing stage, we plotted distributions containing frequencies of these features across the entire market. Big Mountain Resort was in the higher percentile for almost every feature (Figures 2-7). Potential scenarios for cutting costs or increasing revenue were suggested by the business. Each scenario revealed interesting information, particularly scenario 2. Scenario 2 increases the vertical drop by adding a run to a point 150 feet lower down but requires 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 $3,474,638.00. This would be the modeled scenario I suggest. In terms of run closures, since the business was curious, the data suggests that one run closure will allow for ticket price to remain constant. Any other additional run closures will require a lower ticket price.



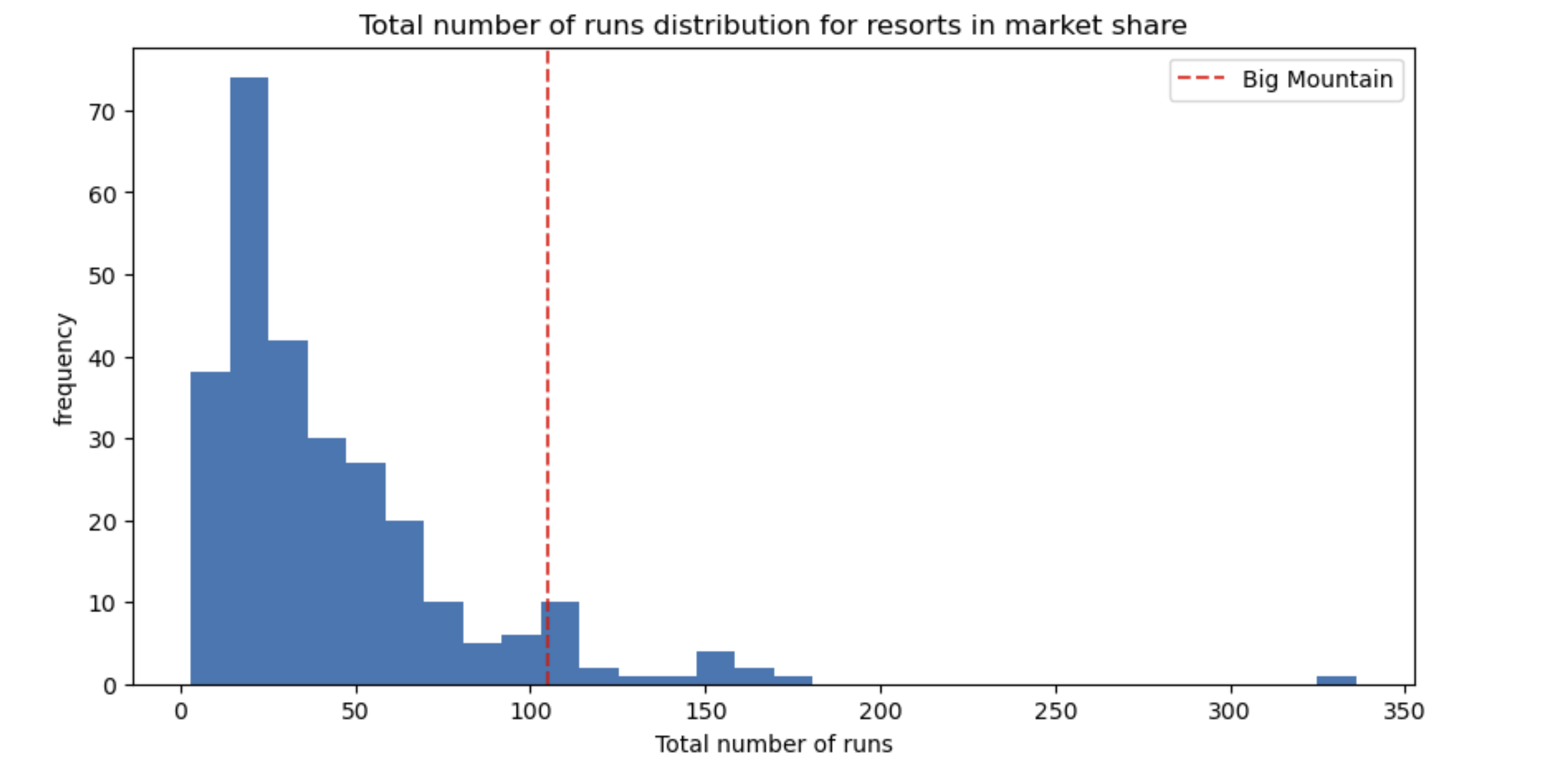
**Figure 7**



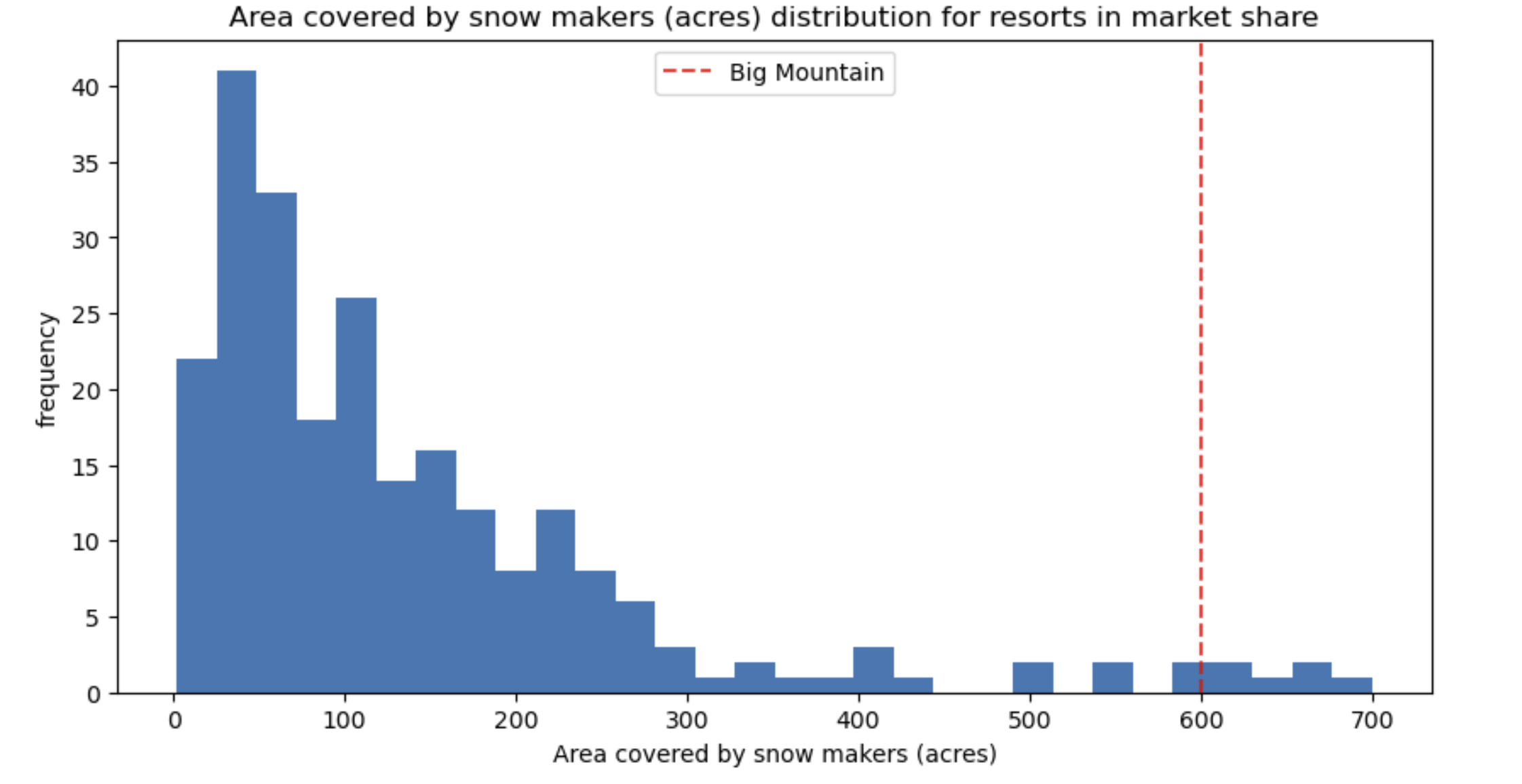
**Figure 5**



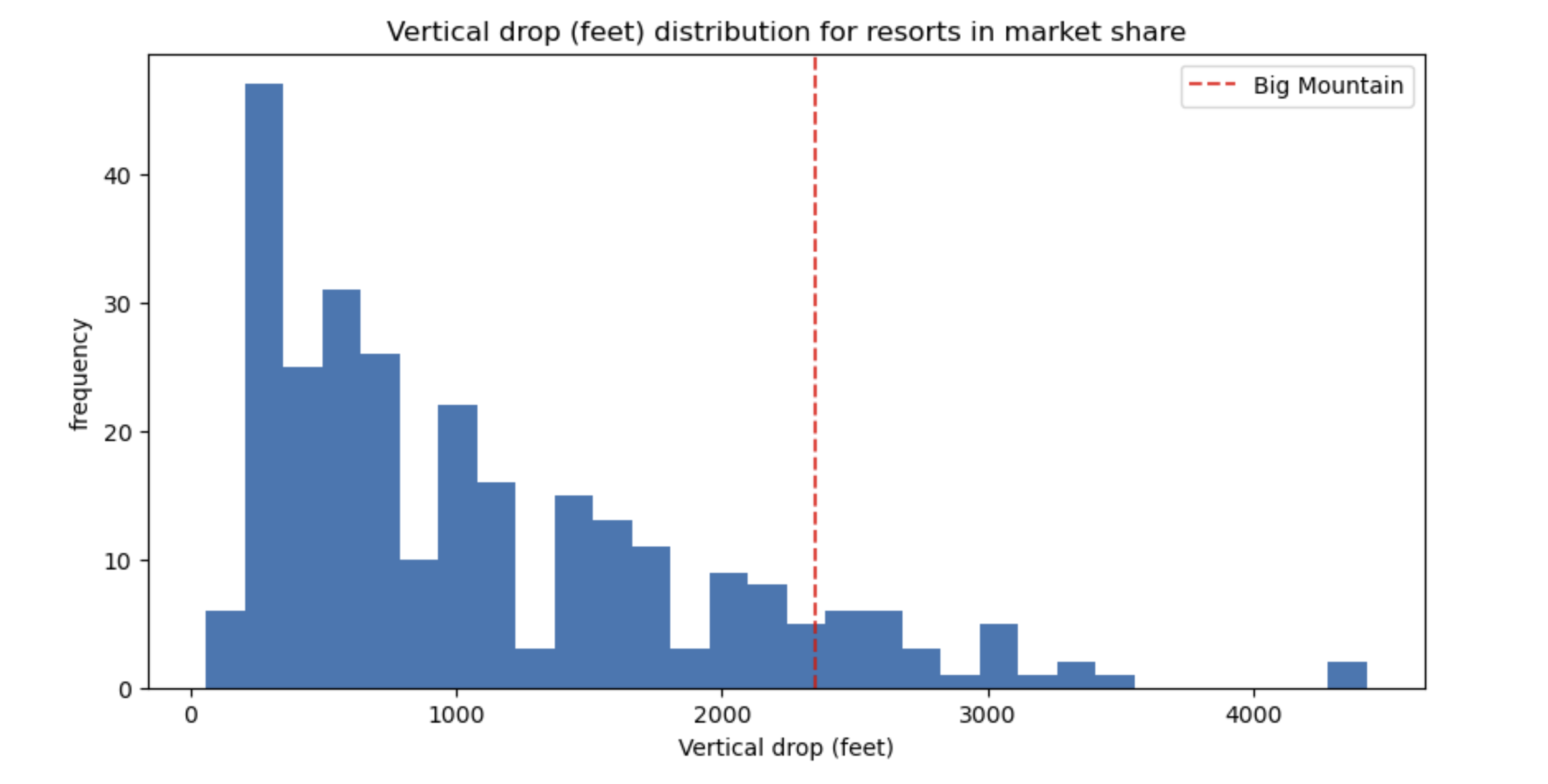
**Figure 4**



**Figure 6**



**Figure 3**



**Figure 2**

To conclude, even though Big Mountain Resort ranked highly across many features, their current price does not match the modeled price. The resort must not have realized how well they stack against others in the market when prices were set. This revelation will not come as a surprise because the execs assumed this from the beginning.

The business could make use of the given model by adjusting price slightly and then modifying their resort data to gain even further insights. Maybe once a more reasonable price is set by the resort, the mean absolute error will adjust slightly and provide justification for future price increases. They could also use the model when they are brainstorming other scenarios and trying to determine whether the scenario supports an increase or decrease in ticket price. This model is saved and is readily available for others to use.