Guided Capstone Project Report

Big Mountain Resort is currently basing their ticket pricing on the market average and adding a premium. An additional chair lift was installed and it will increase their operating costs by $1,540,000 this season. There is speculation that the resort is not capitalizing on its facilities as much as it could. The resort wants a data driven business strategy to guide ticket pricing.

Alesha Eisen, the database manager, has delivered to me a CSV file containing ski resort data for 330 resorts across the United States. Wrangling the data brought the data set down from 330 rows to 277 rows, all while being sure to preserve Montana's Big Mountain Resort data. The reduction of rows are a result of removing resorts with missing ticket pricing and the removal of a resort with an erroneous entry for resort age. A key insight was gained during the data wrangling process – the weekend ticket pricing strongly correlated with the weekday ticket pricing, as shown in **Figure 1**, below:

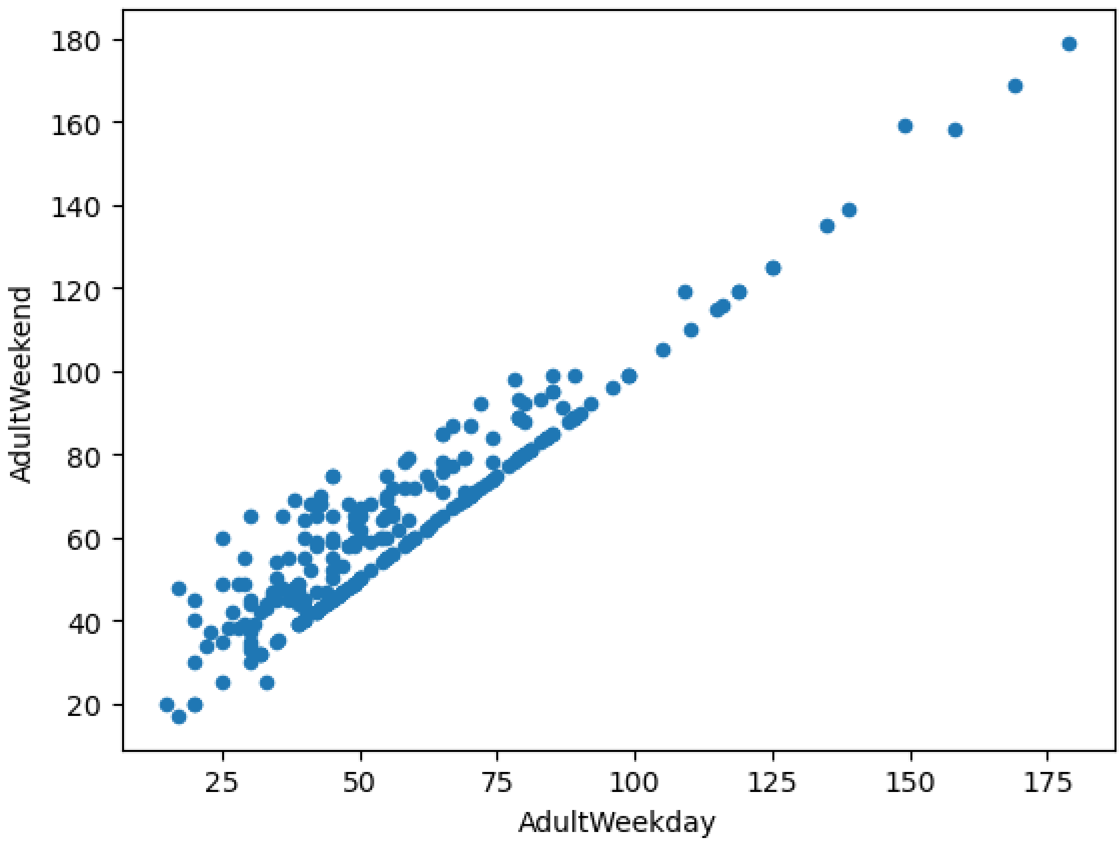

Figure 1: Correlation between weekend and weekday ticket pricing

Due to this strong correlation, and the fact that more data existed for adult weekend ticket pricing than the adult weekday pricing, the adult weekday ticket pricing column was dropped from the dataset.

To look for a simpler way to approach analyzing this high dimensional dataset, a Principal Component Analysis (PCA) was done to find the principal components that accounted for the majority of the variability in the mean AdultWeekend ticket prices by state. The first two derived components from this analysis (which are linear combinations of the variables from within the dataset) were found to influence 77.2% of the variability in the ticket price. **Figure 2**, below, will show this:

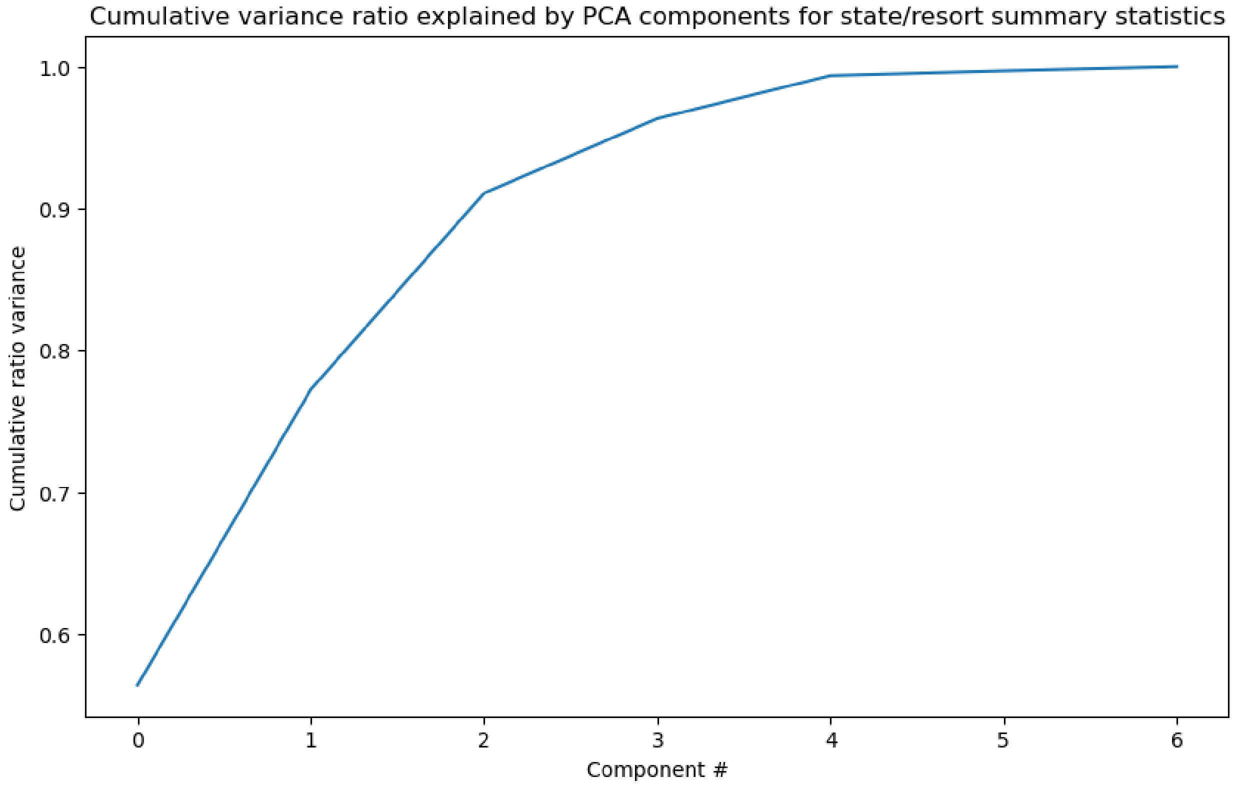

Figure 2: Cumulative variance ratio explained by PCA components for state/resort summary statistics

Now, when these first two PCA components are used to create a scatter plot including state and their respective ticket prices, we get the following plot in **Figure 3**, below:

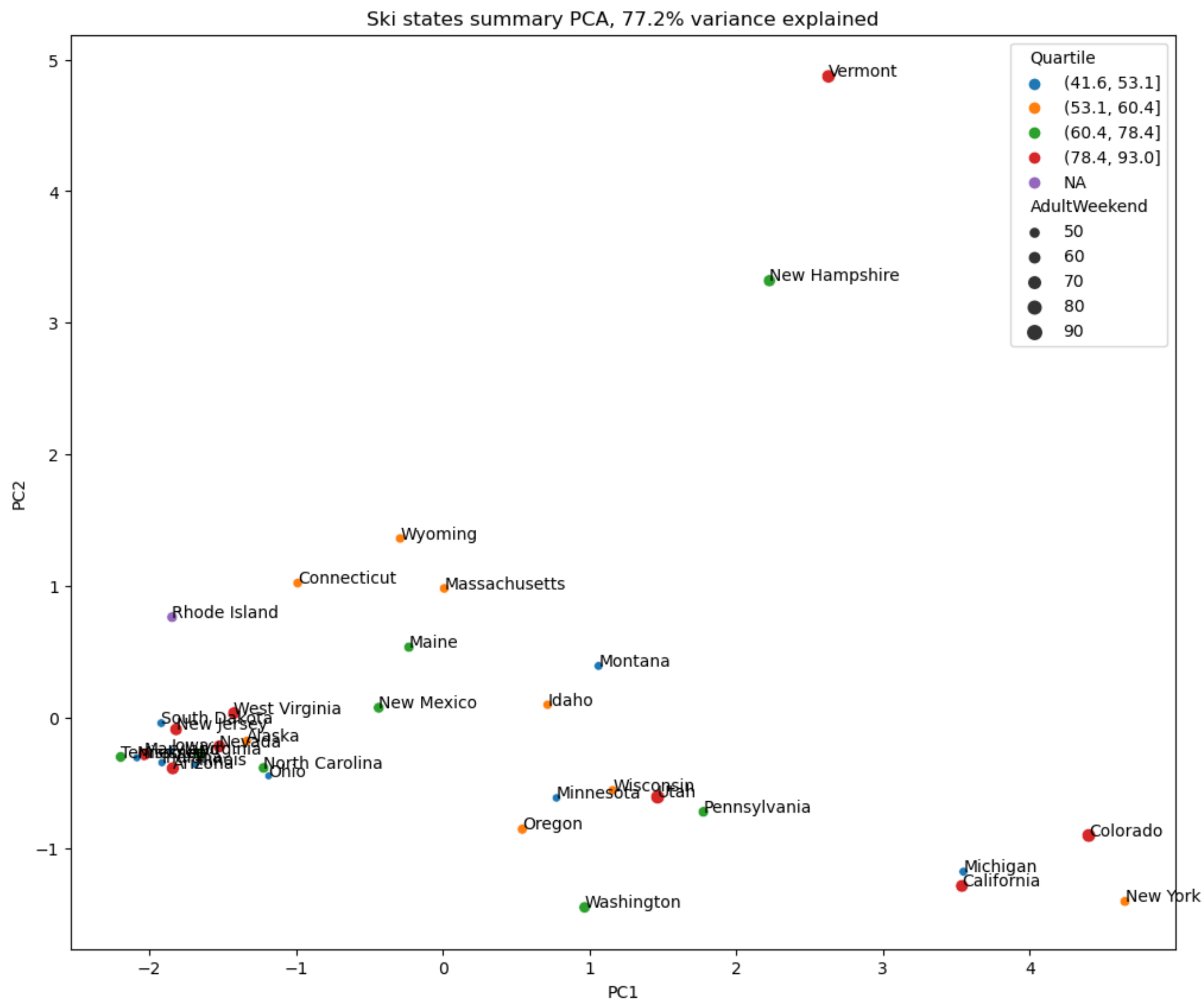

Figure 3: Ski states plotted using their PCA component 1 and 2 scores

From the plot above, it can be seen that there are no clear groupings or correlations between ticket price and state.

Looking for stronger correlations between resort features and their ticket prices, scatterplots were created for each variable in the dataset. The strongest correlations were found to be: vertical_drop, fastQuads, Runs, and total_chairs. **Figure 4**, below, shows these top four positive correlations:

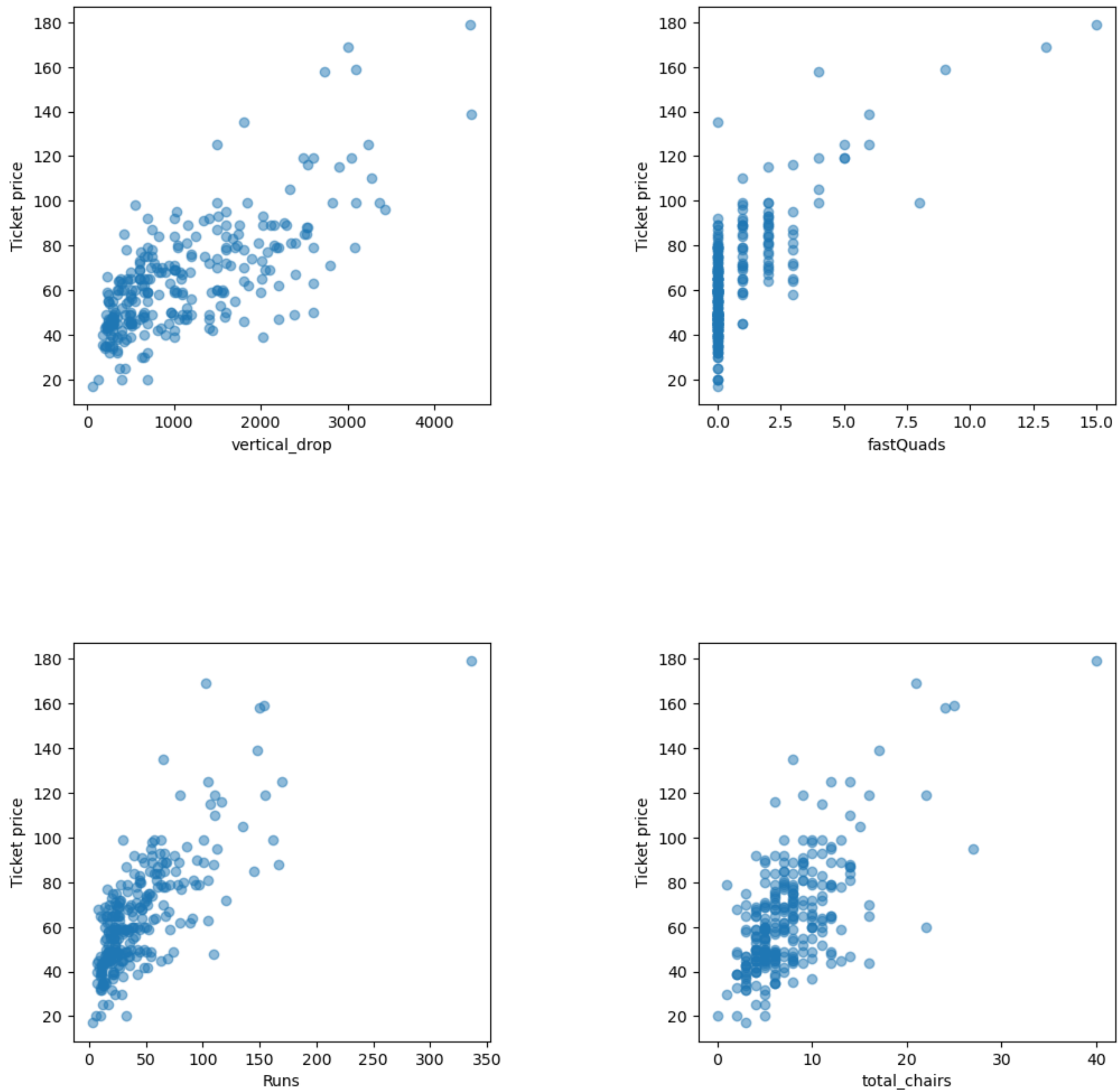

Figure 4: Resort feature correlations to ticket price

When it comes to modeling, the following different models were tested and their performance compared by their mean absolute error:

1. Average price (not an actual model)
    a. The mean absolute error had a value of 19.136 (in other words, $19).

2. Linear model
    a. To avoid tuning the model to the arbitrary test set, cross-validation was used. The results of the linear regression cross validation deemed 8 features to be used in the model, to maximize accuracy while minimizing the variance in the results. The linear model determined the top 8 feature importances (and their respective scoring) to be:
        i.

vertical_drop	10.767857
Snow Making_ac	6.290074
total_chairs	5.794156
fastQuads	5.745626
Runs	5.370555
LongestRun_mi	0.181814
tram	-4.142024
SkiableTerrain_ac	-5.249780

   b. Note that tram and SkiableTerrain_ac have negative scores associated with the features. This means that these features have a negative correlation to ticket price.

   c. On the training set, the linear model had a mean average error of $10.499 with a standard deviation of $1.622. On the test set, the linear model had a mean average error of $11.793.

3. Random forest
   a. A hyperparameter search was done using GridSearchCV to find the best parameter settings to use with this random forest model. The parameters are as follows:
      i. randomforestregressor__n_estimators: 69
      ii. simpleimputer__strategy: 'median'
      iii. standardscaler: None

   b. The random forest regressor determined the top 4 feature importances to be: fastQuads, Runs, Snow Making_ac, and vertical_drop, which agree with the results of the linear model feature selection.

   c. On the training set, the random forest model had a mean average error of $9.645 with a standard deviation of $1.353. On the test set, the random forest model had a mean average error of $9.538.

The random forest model was chosen because it has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

Lastly, a quick study was done to determine whether collecting more data would benefit the accuracy of our model using sklearn's learning_curve function, as seen in **Figure 5**, below:

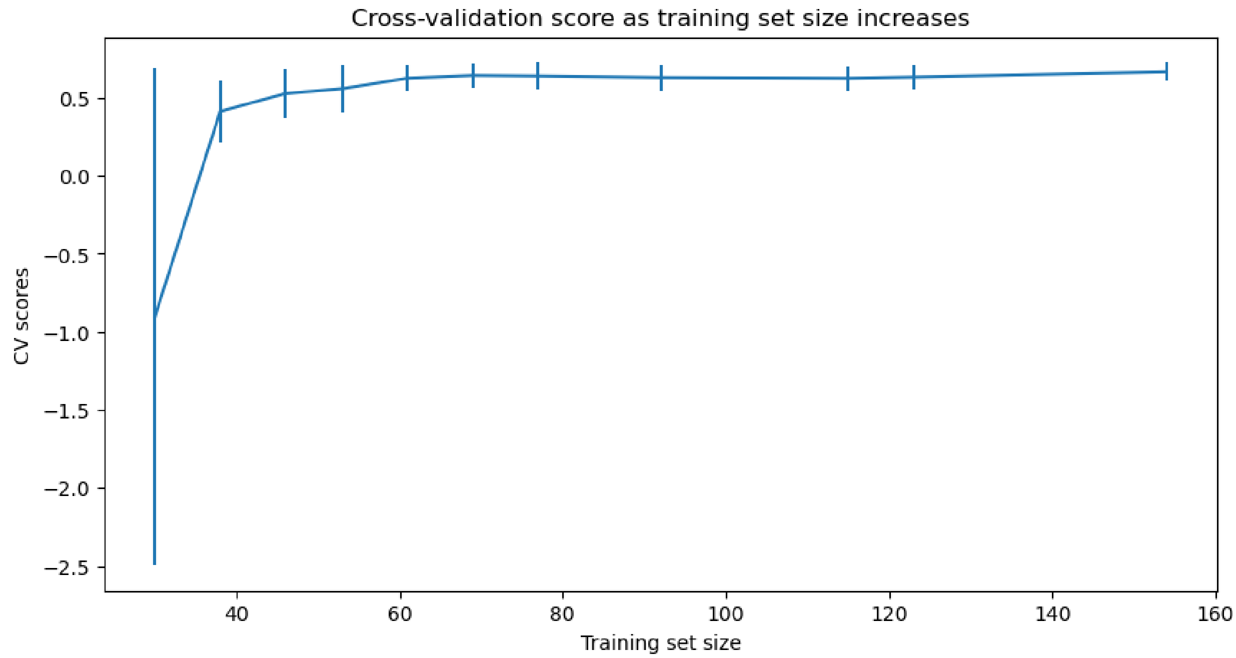

Figure 5: Cross-validation scores as training set size increases

The results show that we seem to have plenty of data. Increasing from a fraction to our full training dataset, there's an initial rapid improvement in model scores as one would expect, but it's essentially leveled off by around a sample size of 40-50% percent of our training dataset.

It is important to understand where Big Mountain Resort lies within market context. The following league charts show how Big Mountain compares to other resorts.

Figure 6, below, shows that Big Mountain currently prices their tickets on the right tail of the market share distribution of adult weekend ticket pricing:

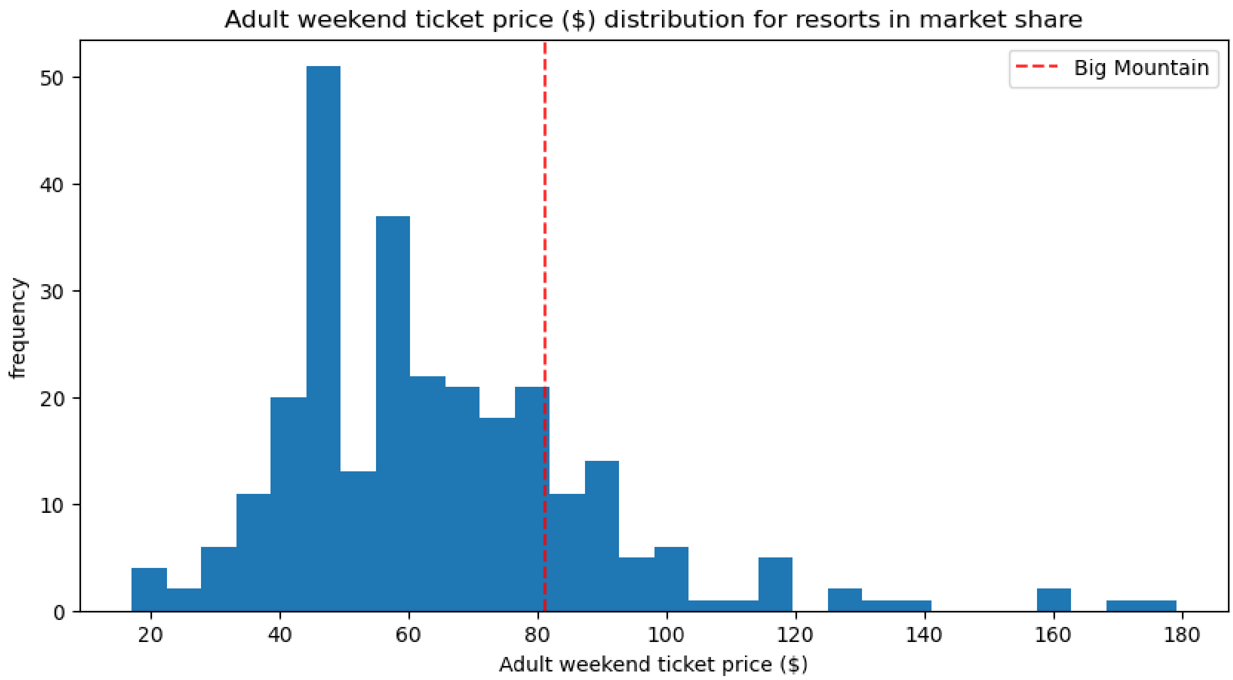

Figure 6: Adult weekend ticket price distribution

Figure 7, below, shows that Big Mountain is currently the most expensive adult weekend lift ticket in the state of Montana:

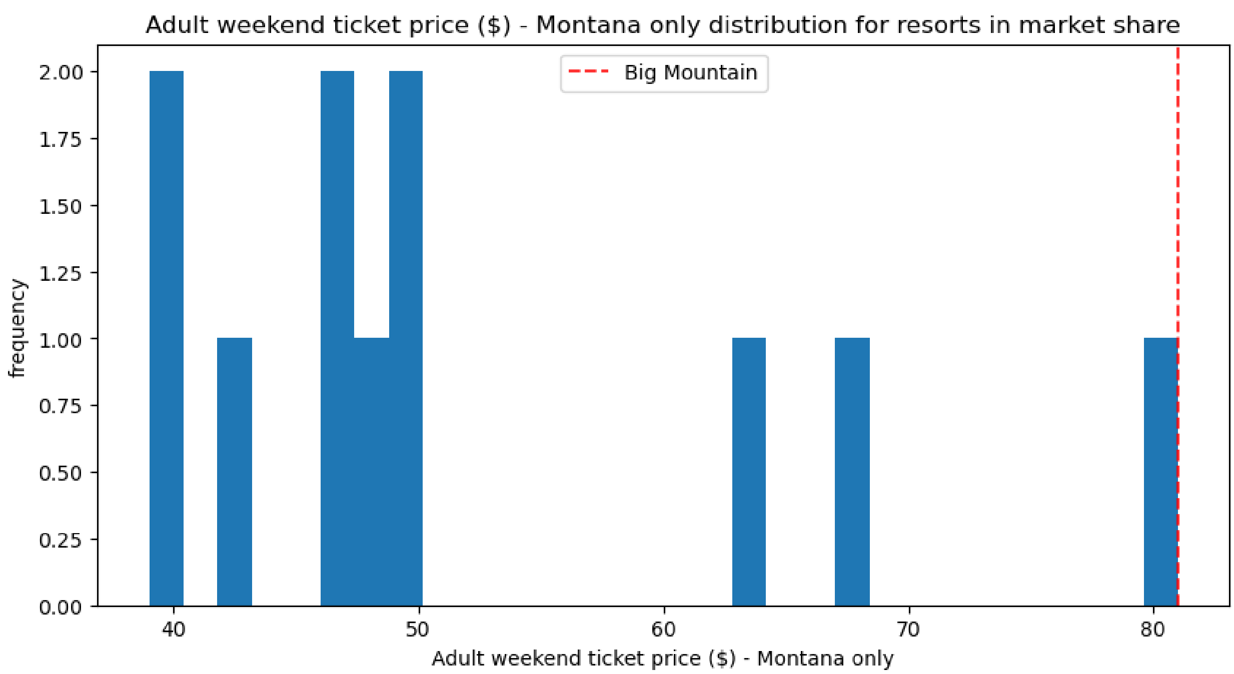

Figure 7: Adult weekend ticket price distribution (Montana Only)

Figure 8, below, shows that Big Mountain is doing well for vertical drop, but there are still quite a few resorts with a greater drop:

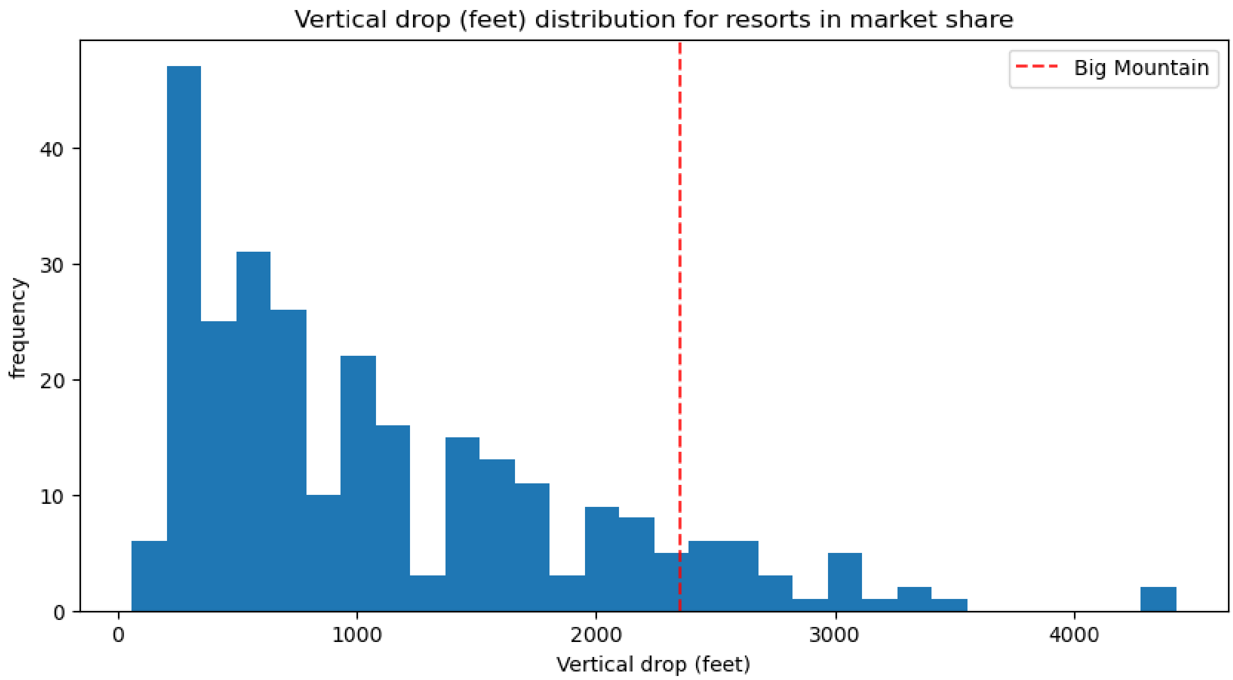

Figure 8: Vertical drop distribution

Figure 9, below, shows that Big Mountain is very high up the league table of snowmaking area:

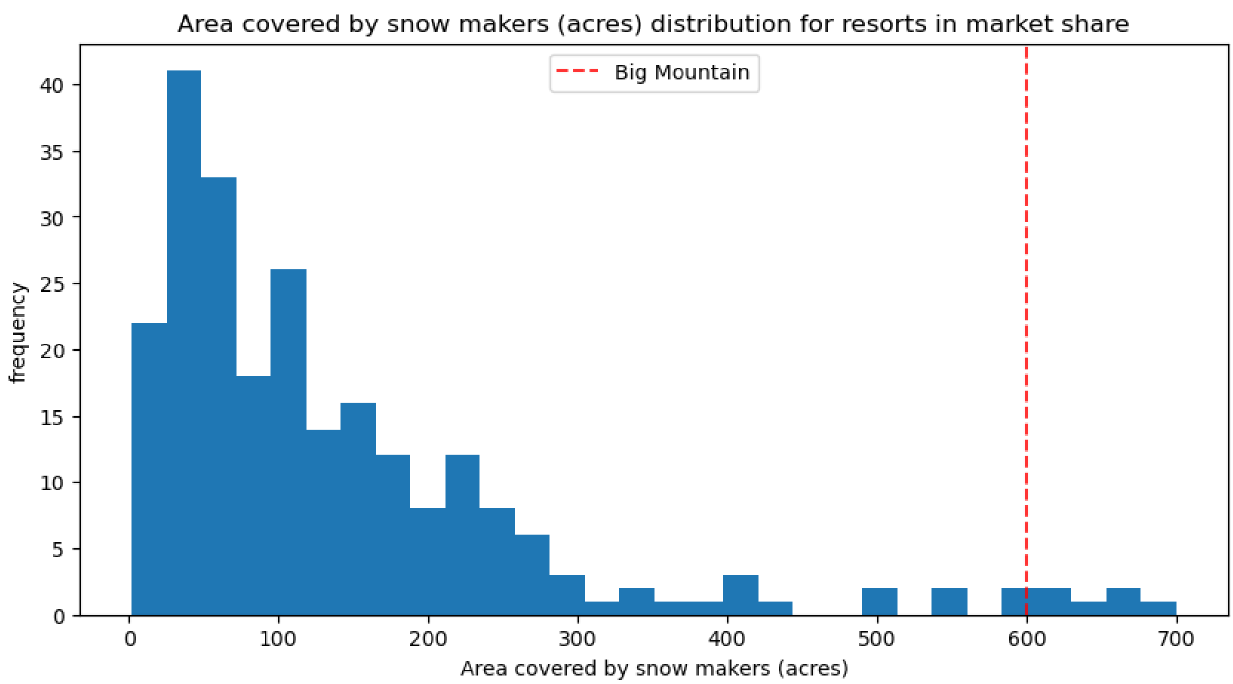

Figure 9: Snowmaker cover distribution

Figure 10, below, shows that Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers:

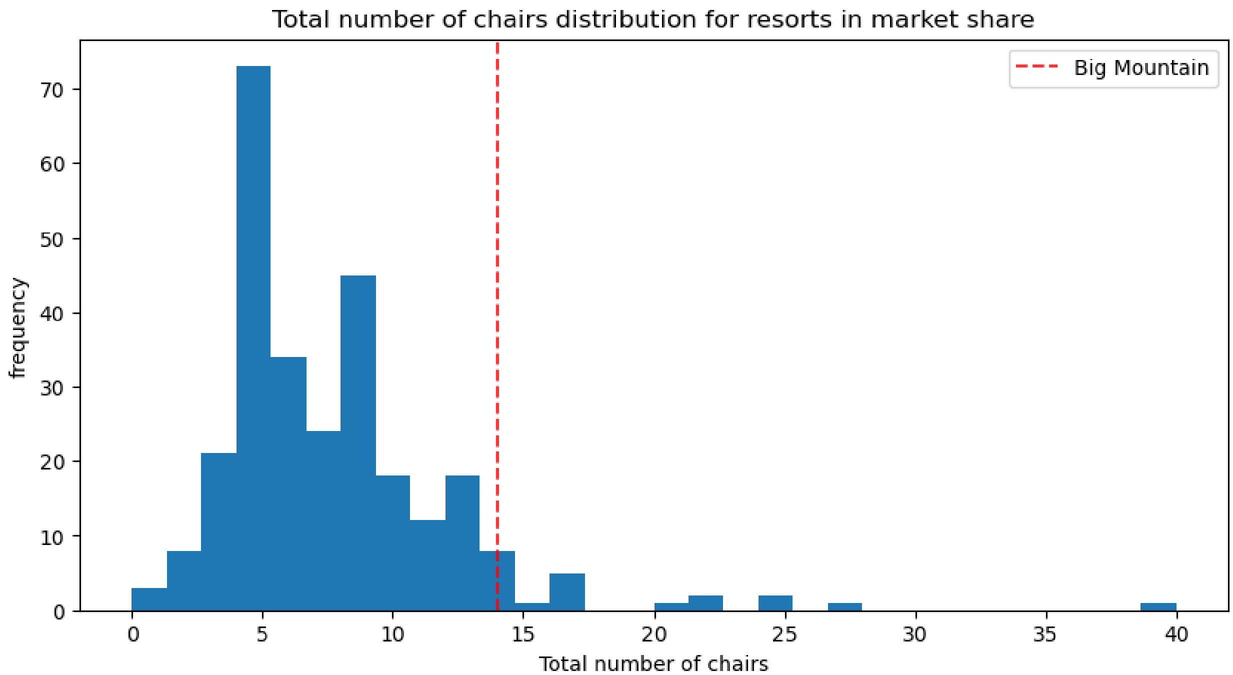

Figure 10: Total number of chairs distribution

Figure 11, below, shows that most resorts have no fast quads. Big Mountain has 3, which puts it high up that league table. There are some values much higher, but they are rare:

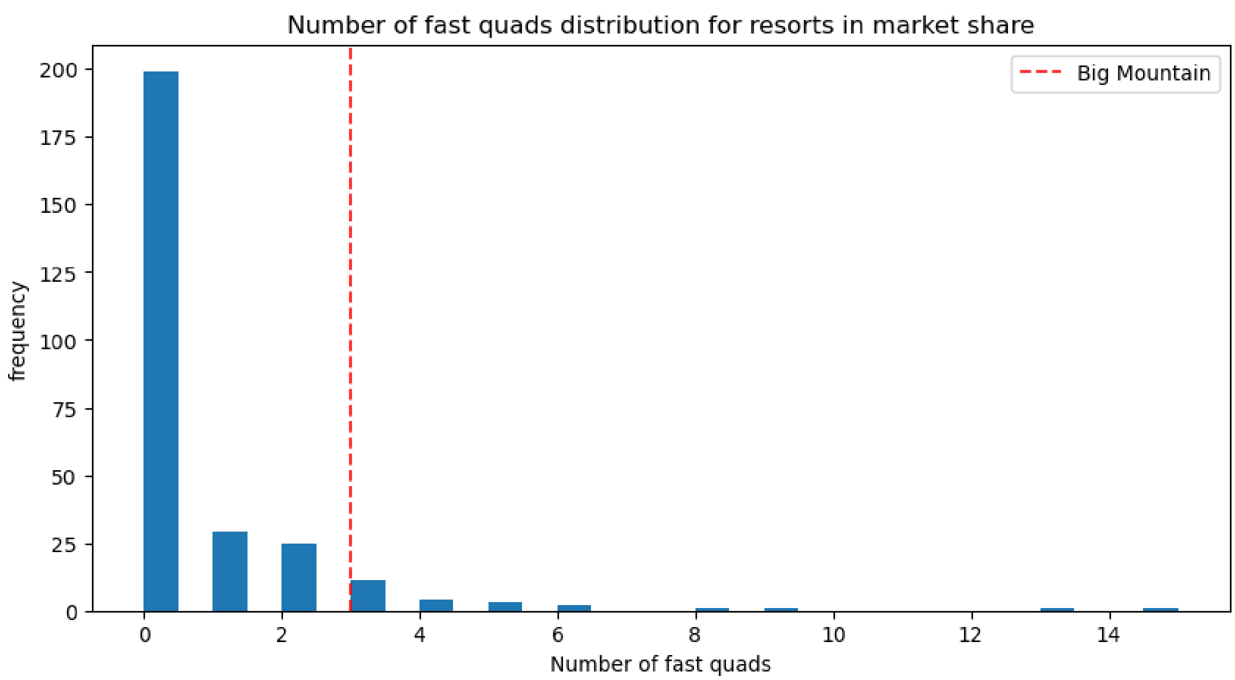

Figure 11: Number of fast quads distribution

Figure 12, below, shows that Big Mountain compares well for the number of runs. There are some resorts with more, but not many:

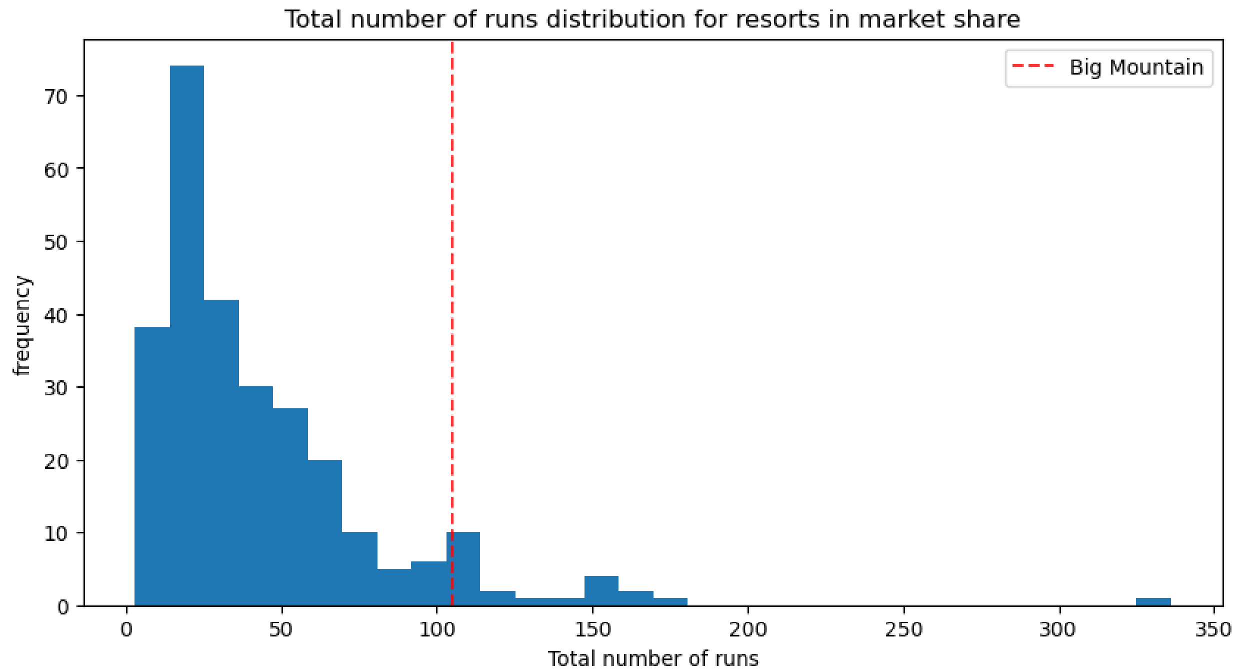

Figure 12: Total number of runs distribution

Figure 13, below, shows that Big Mountain has one of the longest runs. Although it is just over half the length of the longest, the longer ones are rare:

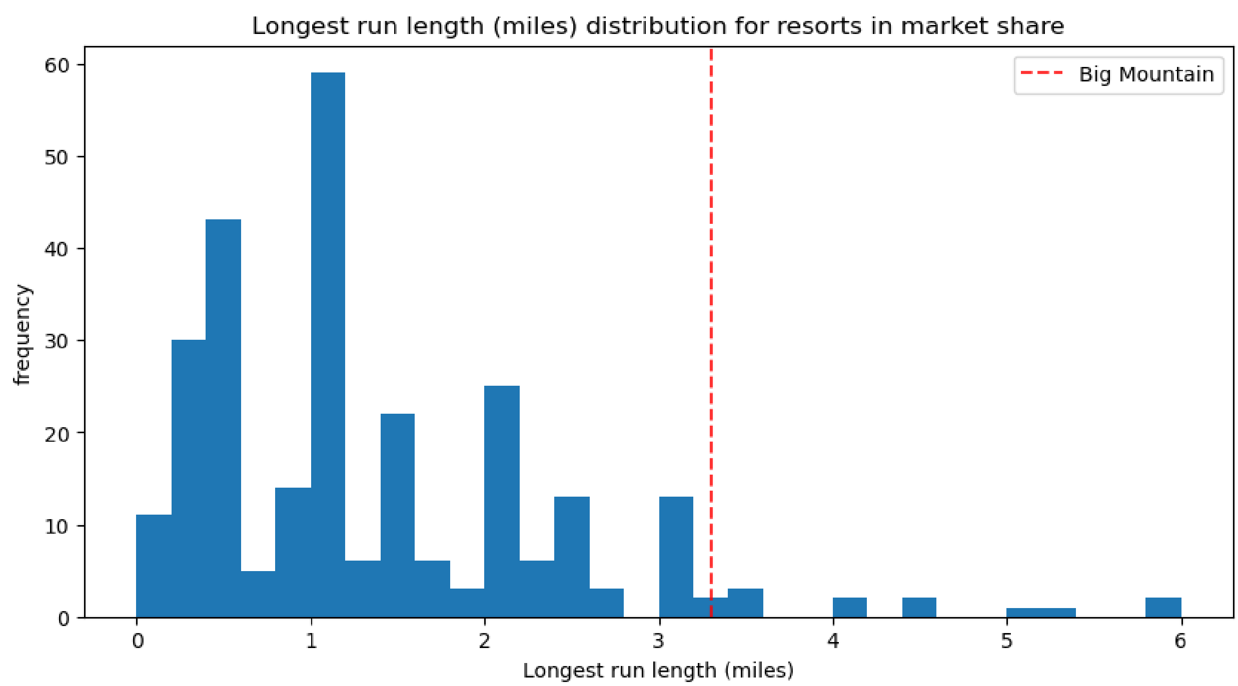

Figure 13: Longest run length distribution

Figure 14, below, shows that the vast majority of resorts, such as Big Mountain, have no trams:

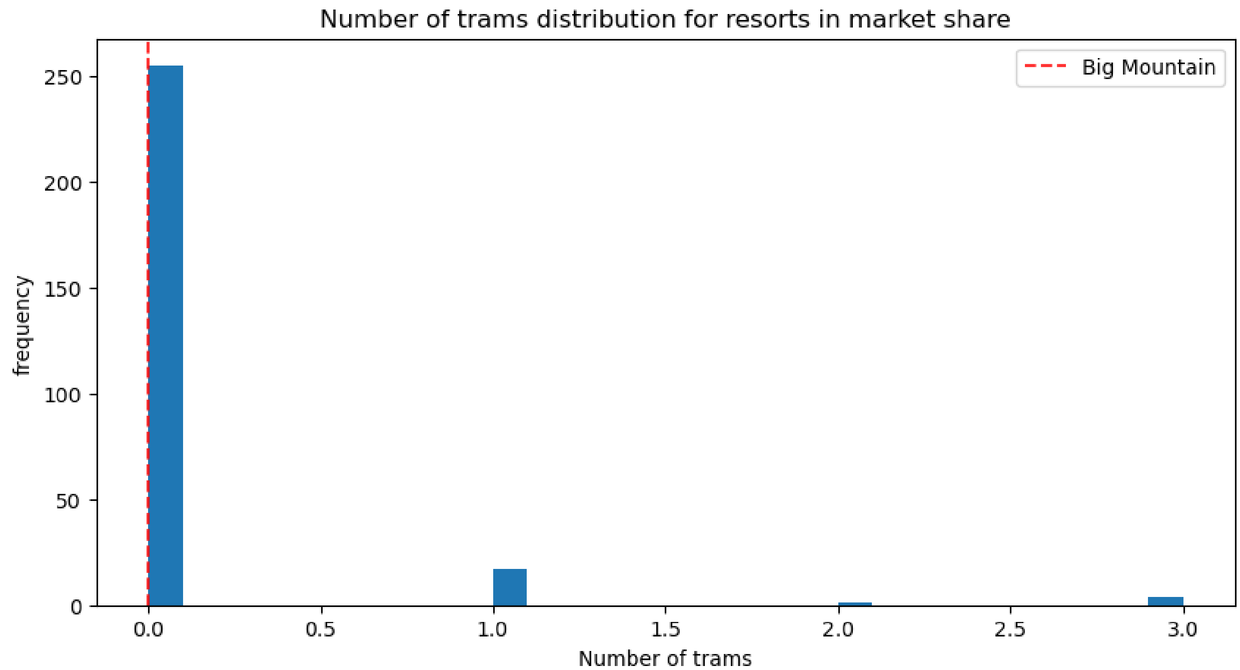

Figure 14: Number of trams distribution

Figure 15, below, shows that Big Mountain is amongst the resorts with the largest amount of skiable terrain:

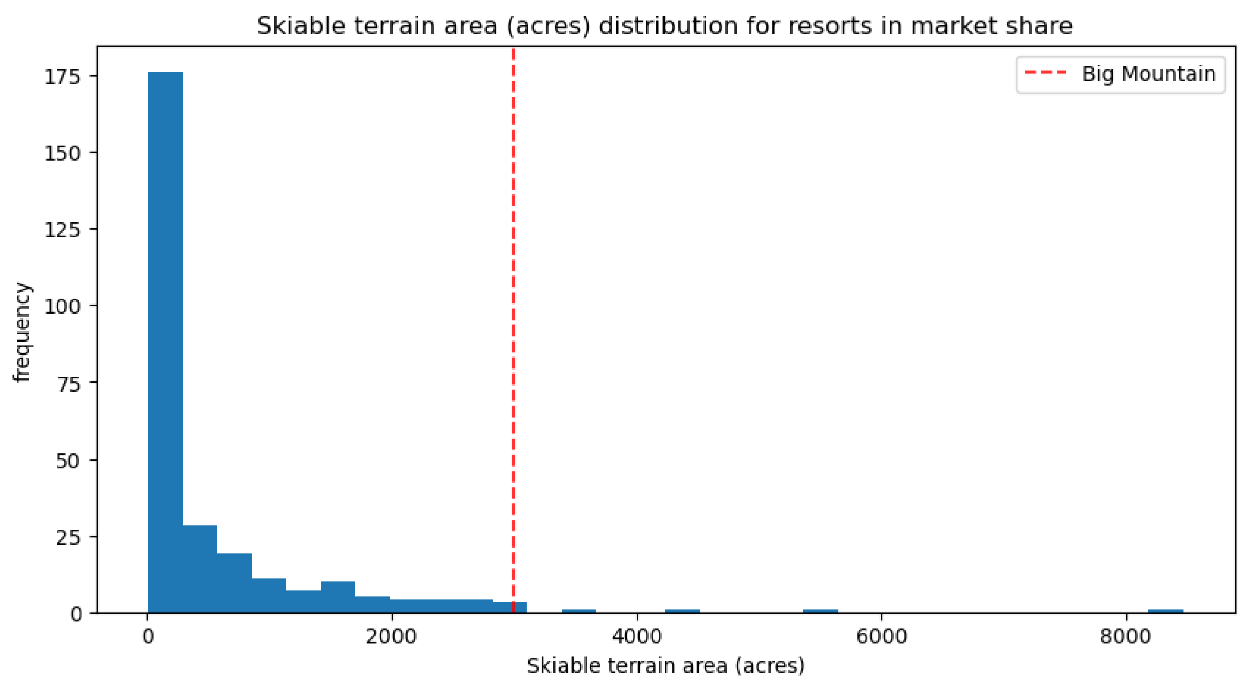

Figure 15: Skiable terrain area distribution

Big Mountain Resort currently charges $81.00 for their adult weekend tickets, while the newly modeled price is $95.87 (calculated from the chosen random forest model). Even with the expected mean absolute error of $10.39, this suggests there is room for an increase. The modeled price is based on the current facilities Big Mountain offers and the rest of the country's resorts.

Taking a look at specifically Montana, Big Mountain Resort already has the highest priced adult weekend ticket in the state. If the public's response to a ticket price increase is of concern to the business leadership, a less aggressive price increase can be taken and a new study conducted to measure a change in business, if any.

In response to the options that the business has shortlisted:

1. If the resort is willing to close up to 10 of the least used runs, then it is important to note that this model predicts the following price changes: The model says closing one run makes no difference to ticket price. Closing 2 and 3 successively reduces support for ticket price by $0.41 and $0.67, respectively and therefore decreases business revenue. These decreases in ticket prices would result in a loss of revenue to the amount of $710,145 and $1,166,667, respectively, assuming the resort continued to receive 350,000 ticket holders that purchased 5 tickets on average. 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 in ticket price support.

2. If the resort increases the vertical drop by adding a run to a point 150 feet lower, but requires the installation of an additional chair lift to bring skiers back up, then the following should be considered: the model predicts an increased support for ticket price by $1.99. Over the season, this could be expected to amount to $3,474,638, with the assumption that the resort continued to receive 350,000 visitors that purchase 5 tickets on average over the course of the season. However, this additional revenue does not take into account the cost of installing or operating an additional ski lift.

3. If the resort took the same actions as option 2, but also added 2 additional acres of snow making cover, then the model predicts no additional support for ticket price.

4. Lastly, if the resort increased the longest run by 0.2 miles to boast a new 3.5 mile longest run and added an additional snow making coverage of 4 acres, then the model also predicts no additional support for ticket price.

Now, for the business's consideration, the resort can experiment with closing unpopular runs if they wish, however, the model predicts that after closing 2 or more runs, there will be a lower support for ticket price which may result in a change of business.

In conclusion, if Big Mountain Resort wishes to offset their new chairlift’s operational cost of $1,540,000 this upcoming season, then perhaps, proceeding with option 2 may be of interest.

The data available for the ski resorts did not include the number of ticket holders per season or number of tickets sold per season. This additional data would provide insight on how ticket price and tickets sold contribute to a resort's revenue. To leave no stone unturned, I would advocate for presenting the league charts in addition to the modeled price to the business executives to openly discuss what information, if any, comes as a surprise.

If the business found this model useful, the model can be packaged up into a well documented function for another business analyst to use for testing any new combination of parameters.