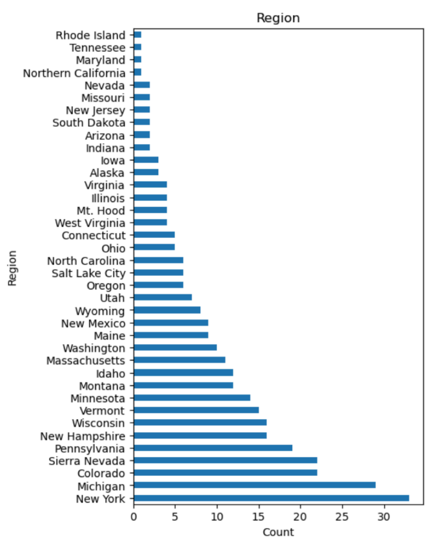
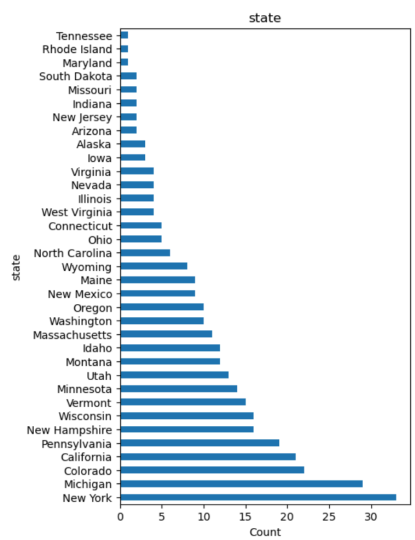
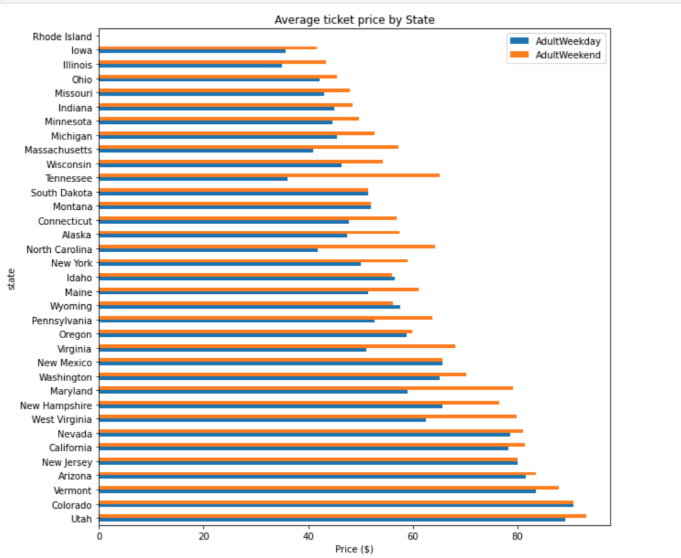
Big Mountain Resort is a ski resort that provides great views and services for skiers and riders of all levels. To improve the even spread of visitors on the mountain, they added another chair lift which increased operating costs by $1.5 million this season. To accommodate for this increase and to ensure continuous and significant profit, Big Mountain Resort must find a way to capitalize on its facilities in order to support their increase in ticket prices.

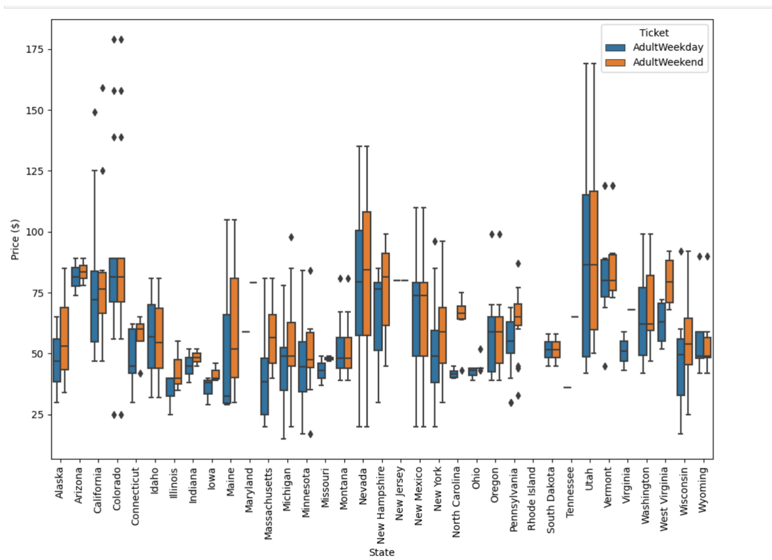
To ensure that Big Mountain Resort is maximizing its returns based on where it stands in the market, we needed to come up with a pricing model for ski resort tickets in the market segment. This pricing model will be built based on which facilities are favored most by visitors and the number of facilities at each resort.

To begin, we collected data on resorts all around the United States, organized the data, and ensured that it is well defined. The data was read from a .csv data file, holding information about different ski resorts around the United States. Our resort of interest, Big Mountain Resort in Montana, was included in the data that was given, containing no missing values. We identified that the state with the most number of ski resorts was New York. Montana was in 13th place. We also identified that most ticket prices ranged from $25 to $100 using box plots. Since one of the most important information given in the data set is the ticket price, we decided to drop the ski resorts that did not have data on their ticket price for both weekend and weekday tickets. This was about 14% of the ski resorts. In Montana we identified that weekend and weekday prices were equal, for the resorts that had both weekend and weekday prices. Since there was more missing data for weekday prices than weekend prices, we decided to drop weekday prices for Montana.



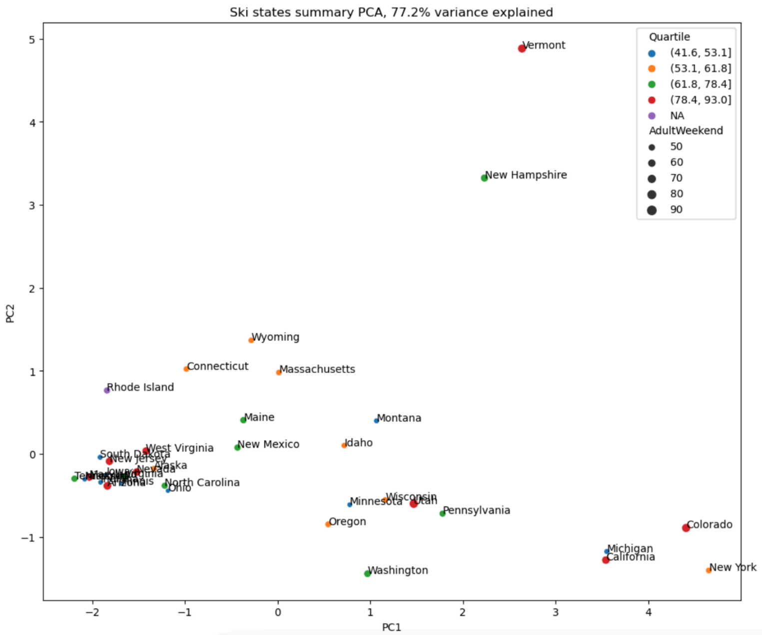


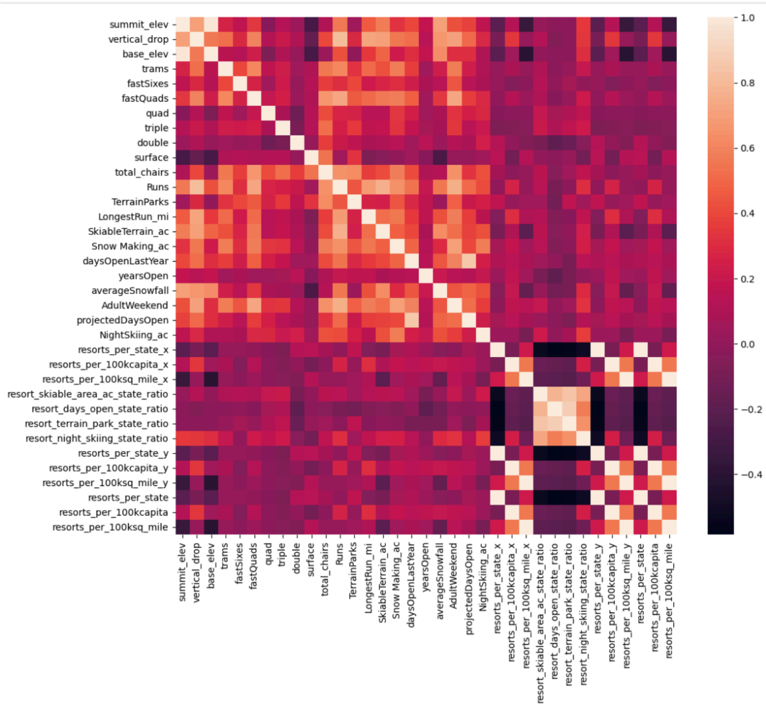


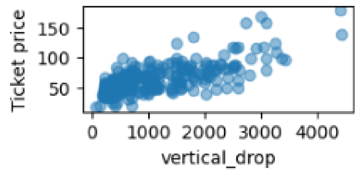


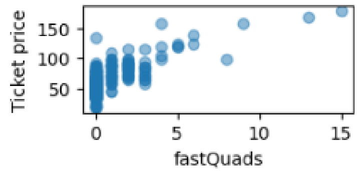
The data science problem identified is to predict the adult weekend ticket price for ski resorts. Montana state area is the third largest and is less densely populated. New York has the greatest number of resorts in the market but does not make it into the top skiable areas. Montana on the other hand does make it into the top five skiable areas. New York dominates the area of skiing available at night and the top five states are generally in the more northerly states. Principle components analysis (PCA) helps to disentangle the interconnected web of relationships. Heat maps show correlation amongst the features. Correlations with adultWeekend ticket prices include fastQuads, runs, snow making acres, resort night skiing state ratio, runs, and total chairs. Visitors value more in guaranteed snow cover rather than more variable terrain area. Vertical drop also seems like a selling point that raises ticket prices as well. A scatterplot was used to show correlation between features and ticket price. Vertical drop showed a strong positive correlation with ticket price. Additionally, fast quads,

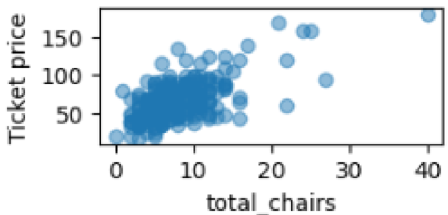
runs, and total chairs did as well. These features are going to be addressed for the modeling up ahead.

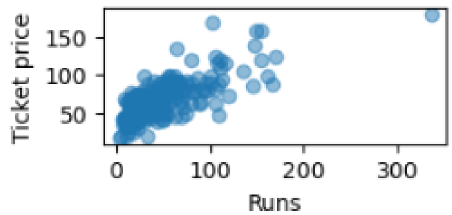


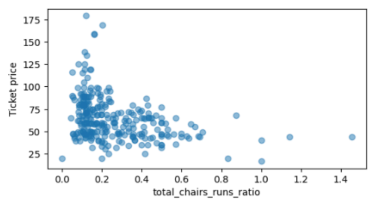


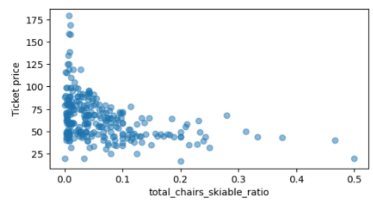


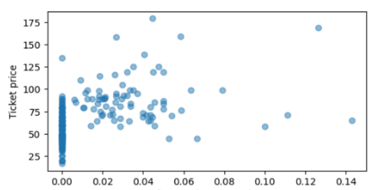


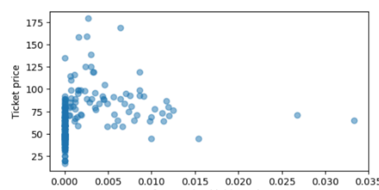






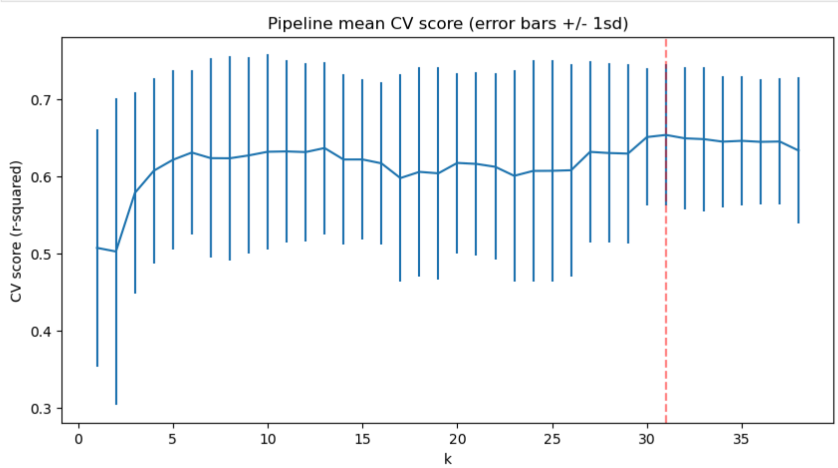


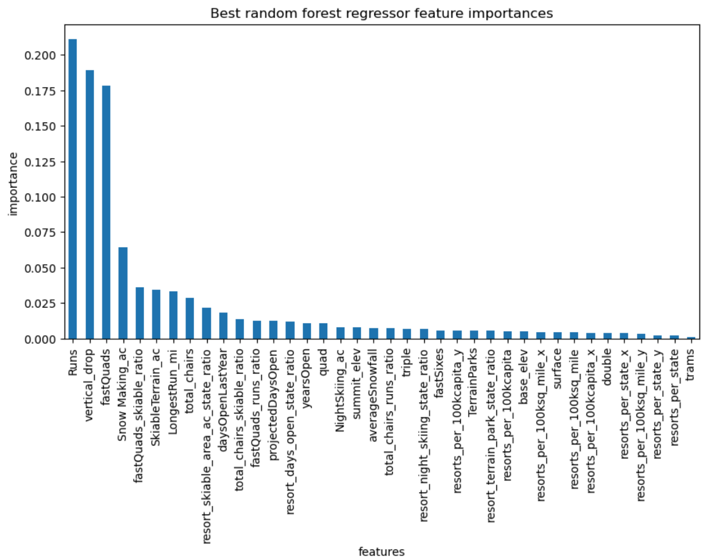




Next, I focused on building machine learning models and considered how useful the mean value is as a predictor. First, I partitioned the data into training and testing splits so that I can have an independent assessment of how the model might perform in the future. After, to determine how good the mean is as a predictor, I found the average price on the training split. To determine how good the mean is, I used R^2, coefficient of determination, which measures the proportion of variance in the dependent variable (ticket price) that is predicted by the "model". The test set resulted in R^2 = -0.02469 and it is common to see performance slightly worse on a test set versus on a training set. Mean absolute error and mean squared error summarizes the difference between predicted and actual values. On average, we might expect to be off by around 19 dollars if you guessed ticket price based on an average of known values. The mean squared error converts us back to our measurement space by

taking the square root of the mean absolute error. This can all be done using Sklearn metrics so that we don't have to define functions every time we want to assess performance. Rather than using the mean, I experimented with the median to see if we could get a better mean absolute error. To do this, we filled missing predictor values with the median ticket prices, scaled the data using StandardScaler which scales each feature to zero mean and unit variance, trained the model on the train split, made predictions using the model on both train and test splits, assessed model performance, and used skilean to find the mean absolute error. Using the median gave us an estimate ticket price within 9 dollars or so of the real price. Compared to the 19 dollars difference we got using the mean, this is a much better result. To refine the linear model, SelectKBest can be used. This selects the k best features, but this can take some trial and error. Instead, we can use cross validation which partitions the training set into k fold, trains the model on k-1 of those folds, and calculates performance on the fold not used in training. With cross validation, we discovered a good value for k is 8. Random forest model is a model that can work very well in a lot of cases, as it stops bad practice of repeatedly checking performance on the test split. This allows us to go straight from defining the pipeline to assessing performance using cross validation. Through this model, we identified that the dominant top four features in common with our linear model are fastQuads, Runs, SnowMaking_ac, and vertical_drop. In the end, we identified that the random forest model has a lower cross validation mean absolute error by almost $1 and also exhibits less variability in comparition to the linear model. Therefore, our final model selection was the random forest model. Lastly, we did a data quantity assessment to see if more or less data is useful. The learning_curve function assesses the trade off of more/less data by seeing how performance varies with differing data set sizes. We saw that there is an initial rapid improvement in model scores but around sample size of 40-50, it levels off.

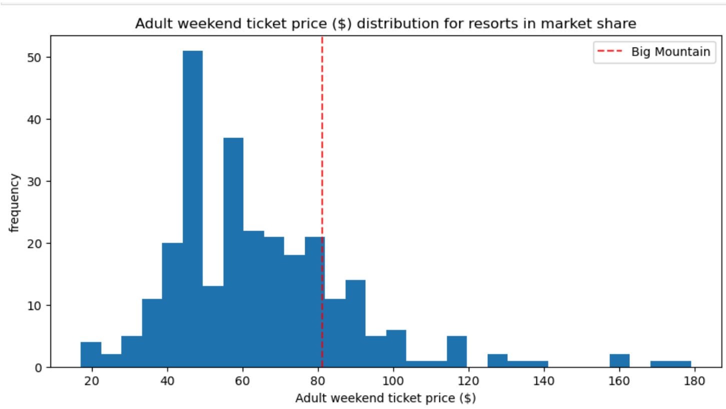


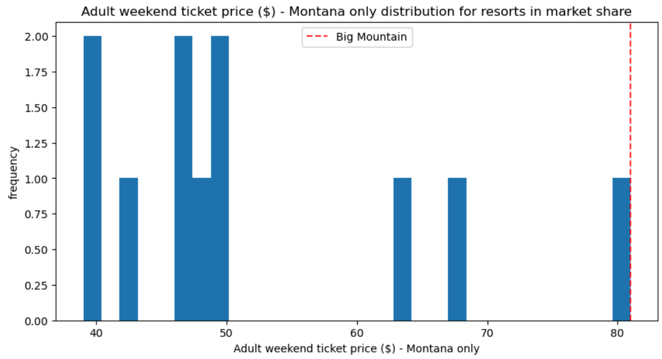


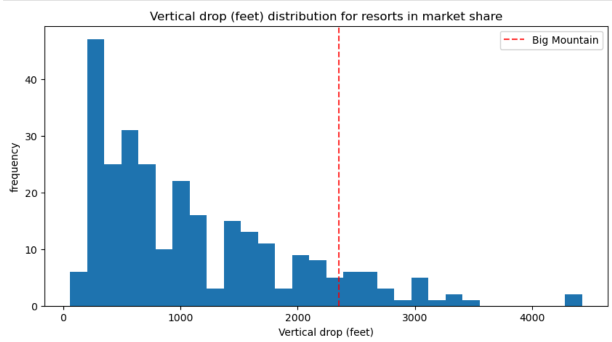
Currently, Big Mountain Resort currently charges 81 dollars for their adult weekend ticket prices. The model suggests that the ticket price should be 92.29 dollars, showing room for an increase from the current actual ticket price. In comparison to all other resorts, Big Mountain ticket prices are higher than the majority of resorts, but there are still a good number of other resorts that carry higher ticket prices. In comparison to other resorts in Montana, Big Mountain ticket prices is shown to have the highest. Important features in market context to consider include vertical drop, snow making acres, total number of chairs, number of fast quads, number of runs, longest run, number of trams, and skiable terrain acres. Vertical drops are doing well at Big Mountain but there are still quite a few resorts with greater drops. Snow making area, total number of chairs, fast quads, longest run, and skiable terrain area for Big Mountain resort all show as the highest compared to all other resorts. For future improvements, closing down 2-3 runs does reduce support for ticket price and revenue slightly, but it is not a detrimental decrease. Additionally, closing down 4 to 5 runs makes no difference in ticket prices and revenue in comparison to closing down 3 runs. Yes, closing down 5 runs can affect the ticket price and revenue slightly, it can help reduce other

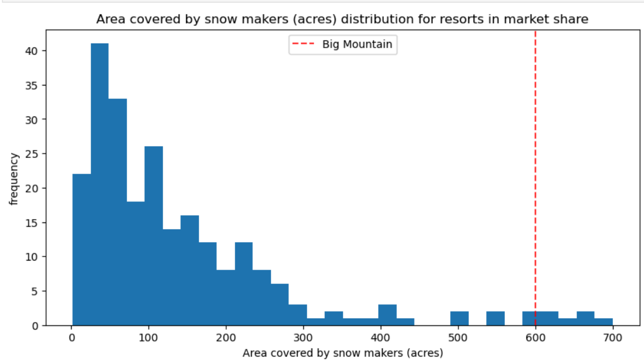
operating costs because there is less to maintain, therefore this could ultimately benefit the resort financially. To test that these predictions are true, Big Mountain can test run closures by temporarily closing 1-2 runs at a time to see how this affects operating costs and if it significantly impacts ticket price and revenue. This can be a low risk approach but can take more time to actually determine and implement any permanent changes.

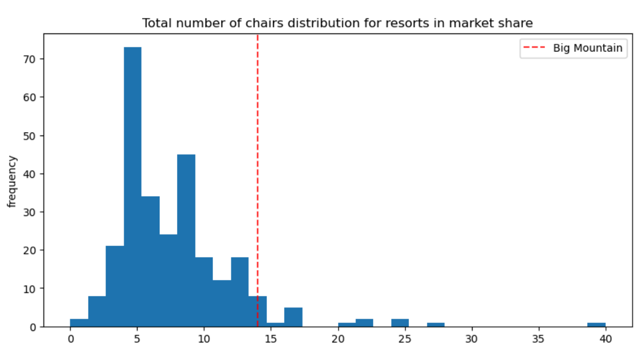
From the model, we don't necessarily know if some resorts charge much less than what is predicted, resulting in the Big Mountain resort undercharging rather than overcharging. Because we don't know anything about the other resorts' operating costs it is hard to tell if some resorts are "overpriced" or "underpriced". Additional to the operatin costs of the new chair lift, I think it would be helpful to know the operating costs for maintaining the runs because if Big Mountain were to go with the modeling scenario involving the shut down of 3-5 runs, having an idea of the operating costs for maintaining these runs could allow us to see how shutting down these runs helps the overall business financially while still maintaining a reasonable ticket price and revenue. A reason for Big Mountain's modeled price being so much higher than its current price might be because its prices were maybe only compared with the other resorts in Montana rather than all the resorts in the U.S.. In comparison to other resorts in Montana, Big Mountain has the highest price, but in comparison to other resorts all over the United States, Big Mountain is on the higher end of ticket prices, but there are a great number of resorts with even higher prices. This may suprise the business executives and they may want more data being compared between Big Mountain and all the resorts in the U.S. rather than just the ones in Montana. The business would take its next steps by taking the model and slowly incorporating it into the resort, temporarily closing some runs in increments and seeing how this affects ticket prices, operating costs, and revenue. If the business leaders wanted to test new combinations of parameters in a scenario, we can provide them the function where they only need to impute the necessary parameters to assess if these parameters will or will not increase the ticket price and revenue.

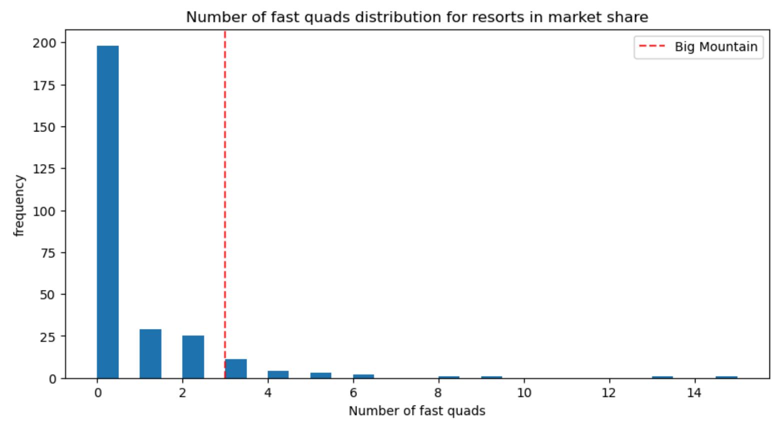


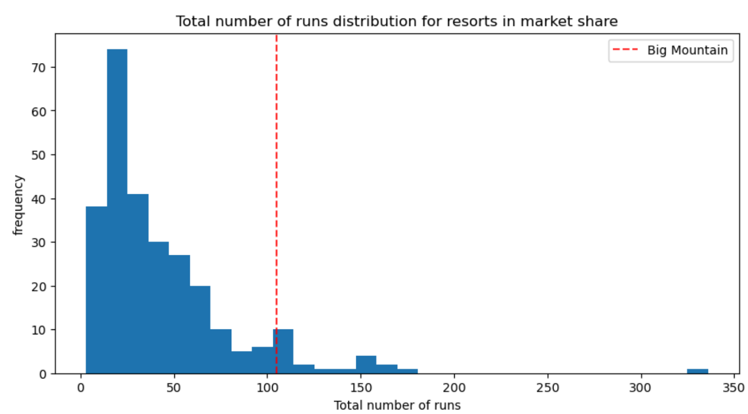

Adult weekend ticket price ($) - Montana only distribution for resorts in market share
Big Mountain
frequency
Adult weekend ticket price ($) - Montana only

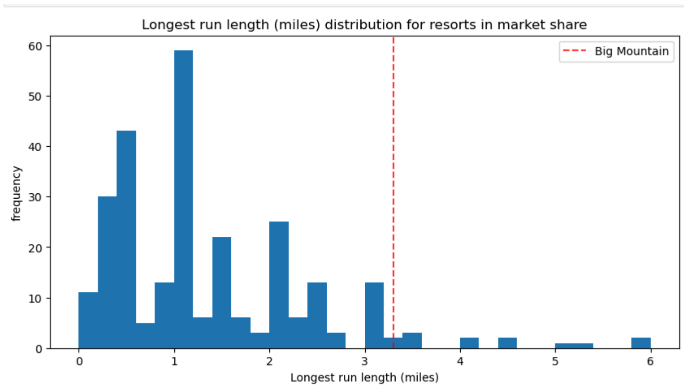

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet)

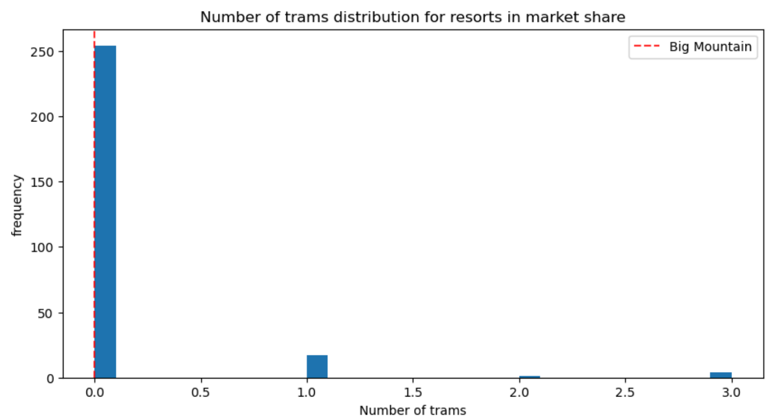

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
Area covered by snow makers (acres)

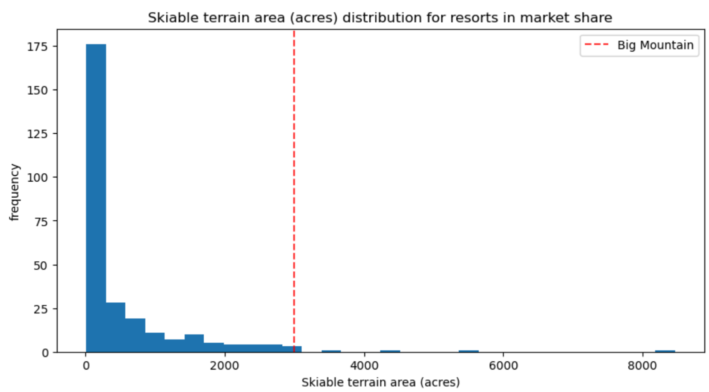

Total number of chairs distribution for resorts in market share
Big Mountain
frequency
Total number of chairs

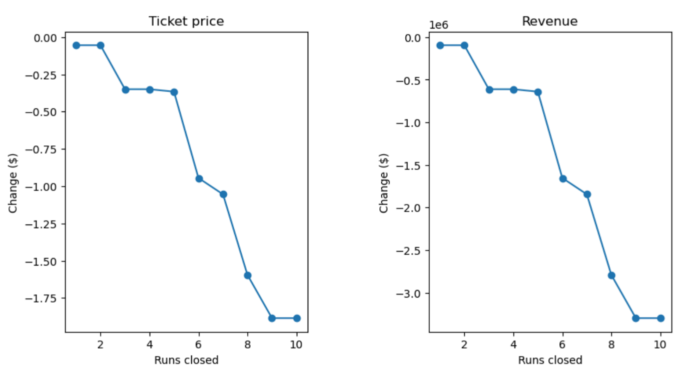

Number of fast quads distribution for resorts in market share
Big Mountain
frequency
Number of fast quads


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


Longest run length (miles) distribution for resorts in market share
Big Mountain
frequency
Longest run length (miles)


Number of trams distribution for resorts in market share
Big Mountain
frequency
Number of trams


Skiable terrain area (acres) distribution for resorts in market share
Big Mountain
frequency
Skiable terrain area (acres)


Ticket price
Change ($)
Runs closed
Revenue
1e6
Change ($)
Runs closed