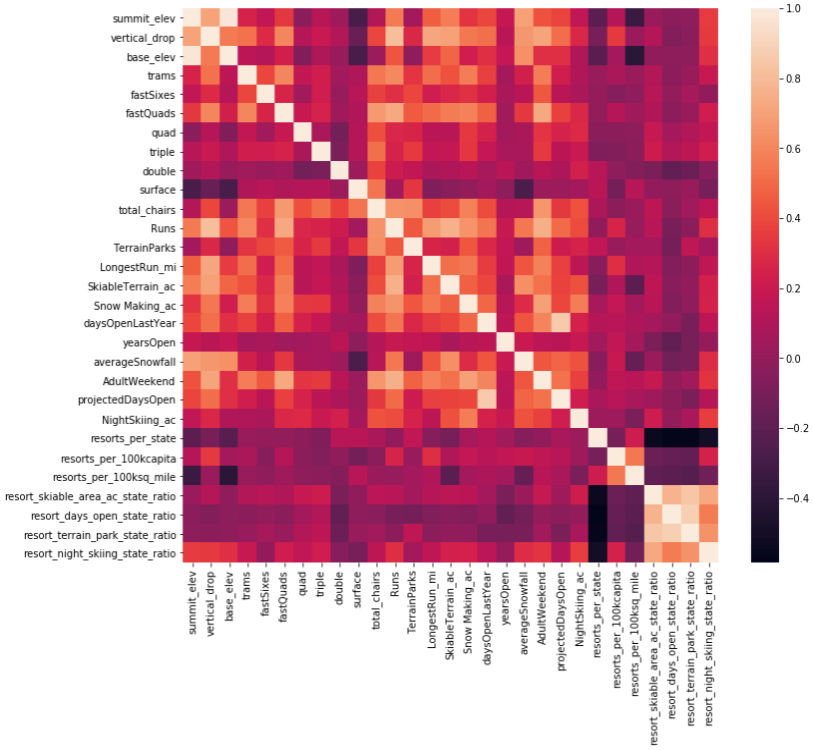
Guided Capstone Project

A report on Big Mountain Resort Analysis

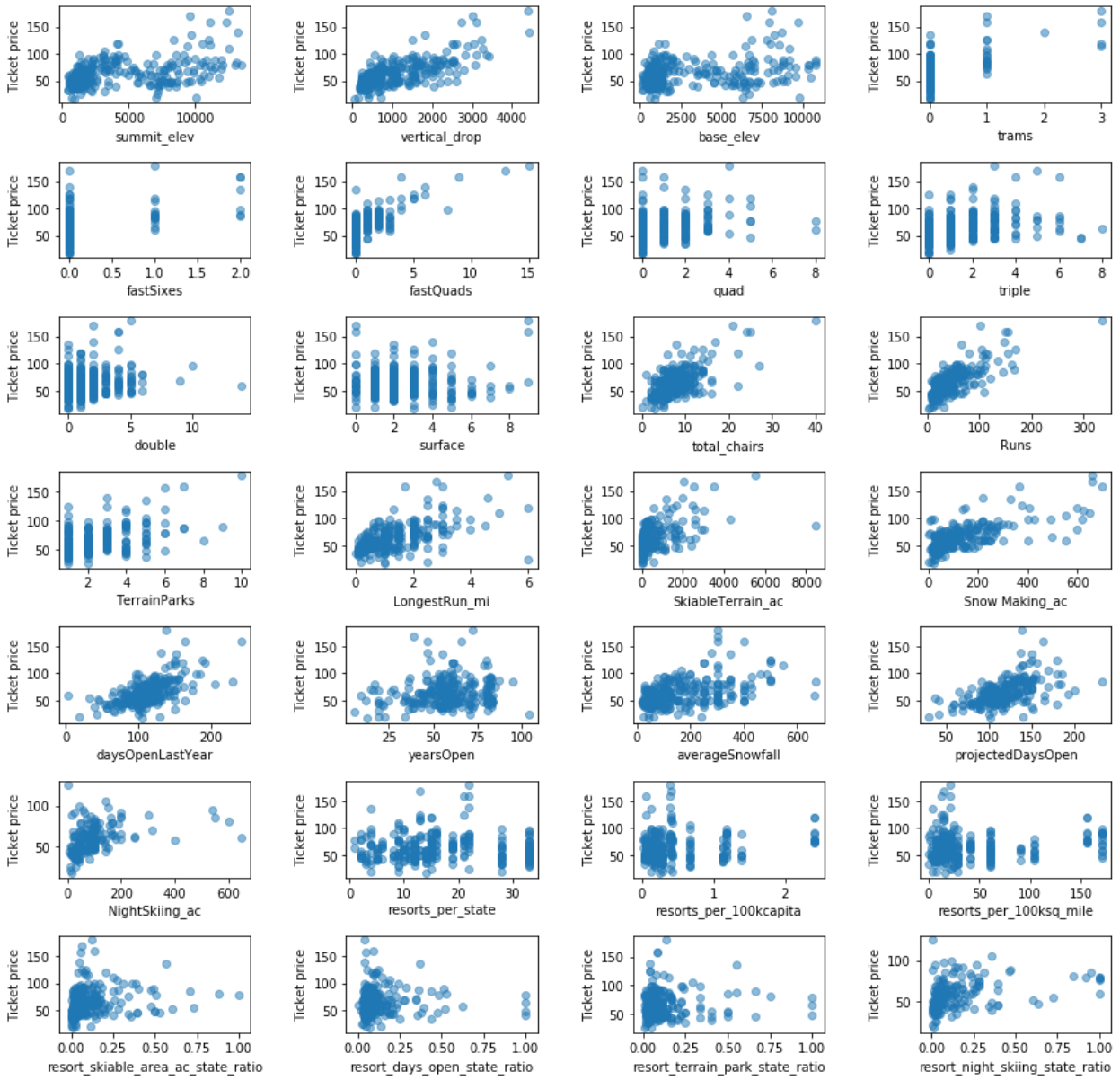
The primary objective of the project was to identify how to increase the total revenue of Big Mountain Resort. We started with cleaning the data and then performing an exploratory data analysis to understand the performance of Big Mountain Resort compared to its competitors. We explored the data and looked for missing values and duplicates in the dataset. After finding out that the data did not have any duplicates, I noticed that several columns had very skewed and suspicious entries. First off, there existed two geographic categories, 'state,' and, 'Region.' After looking into the differences between these two categories I discovered that some states are counted into multiple regions, such as California having resorts counted into the Sierra Nevada and the Northern California regions. This resulted in slightly more 'regions' than states. The bottom line however, is that I would like to refer to BMR's marketing research team to better understand how they treat categorization in regards to state and region.Further evaluation revealed that 82% of the resorts had information for both 'Adult Weekday Ticket Prices' and 'Adult Weekend Ticket Prices,' 3% had recorded only prices for one of the categories, and 14% had no records of ticket price. Before dropping the 14% with no record of ticket prices, I checked over the other categories. Seeing that the 'fast Eight,' category had almost no data, I decided to drop that column. Additionally, there was a glaring entry error for a resort in regards to it's entry in the 'years Open,' column, where it was recorded to have been open for 2,019 years. Seeing as this was impossible, this row was dropped. After evaluating the rest of the columns and their distributions, I decided they all appeared to be plausible and I dropped the entries that had no data for ticket price (14%). The last step was deciding on which price was better for moving forward. Adult Weekday Ticket Prices had more missing entries than Adult Weekend Ticket Prices. Therefore, I dropped weekday prices from the data so that I could focus only on one set of pricing data. This should matter little to BMR as their weekday and weekend ticket prices are the same. There are still some entries missing data, however, I have refrained from removing any of them until I am sure that the entries missing data aren't of any value to determining ticket pricing for BMR. As it stands, after all of this there still remains 277 rows and 25 columns for our data.

Correlation Heatmap:



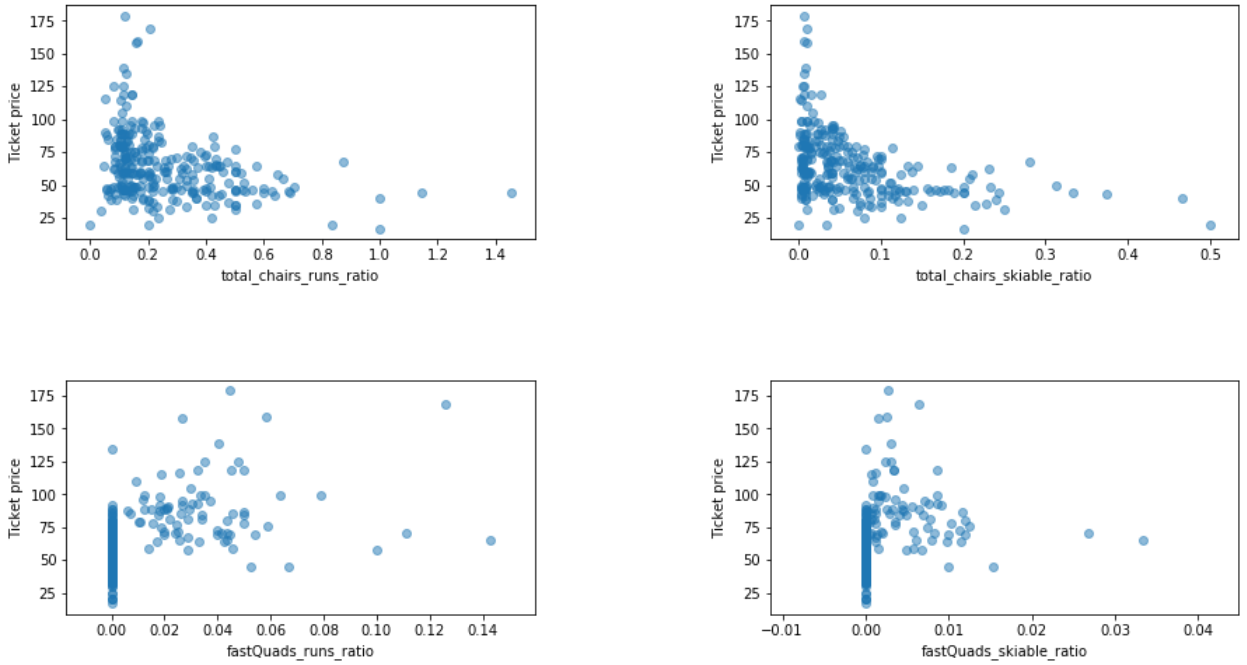
Summit and *base elevation* are quite highly correlated. An interesting observation in this region of the heatmap is that there is some positive correlation between the ratio of *night skiing area* with the *number of resorts per capita*. In other words, it seems that when resorts are more densely located with population, more night skiing is provided. Turning our attention to the target feature, *AdultWeekend* ticket price, we see quite a few reasonable correlations. *fastQuads* stands out, along with *Runs* and *Snow making ac*. Visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up. Of the new features. *Resort_night_sking_state_ratio* seems the most correlated with ticket price. If this is true, then perhaps seizing a greater share of night skiing capacity is positive for the price a resort can charge. As well as *Runs*, *total chairs* is quite well correlated with ticket price. This is plausible; the more runs you have, the more chairs you'd need to ferry people to them! Interestingly, they may count for more than the *total skiable terrain* area. For sure, the *total skiable terrain area* is not as useful as the area with snow making. People seem to put more value in guaranteed snow cover rather than more variable terrain area. The vertical drop seems to be a selling point that raises ticket prices as well.

Scatter plot analysis:

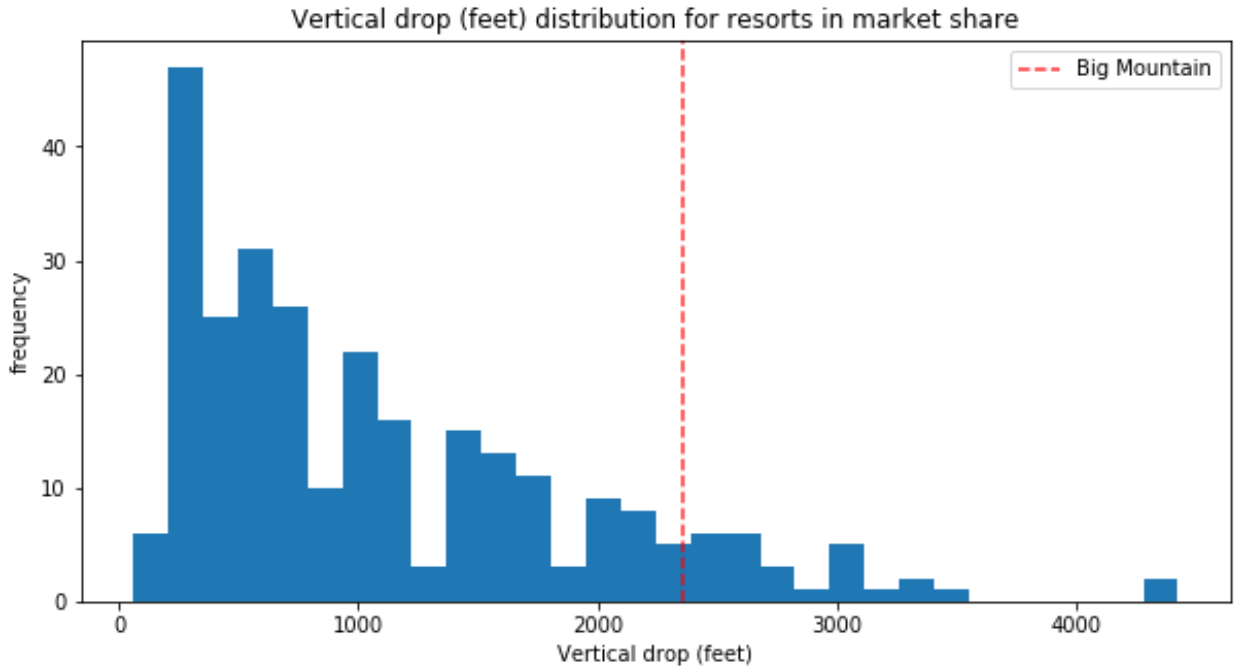


In the scatterplots we see there is a strong positive correlation with vertical drop. fastQuads seems very useful. Runs and total chairs are quite similar and also useful. resortsper100kcapita when the value is low, there is quite a variability in ticket price, although it's capable of going quite high. Ticket price may drop a little before then climbing upwards as the number of resorts per capita increases. Ticket price could climb with the number of resorts serving a population because it indicates a popular area for skiing with plenty of demand. The lower

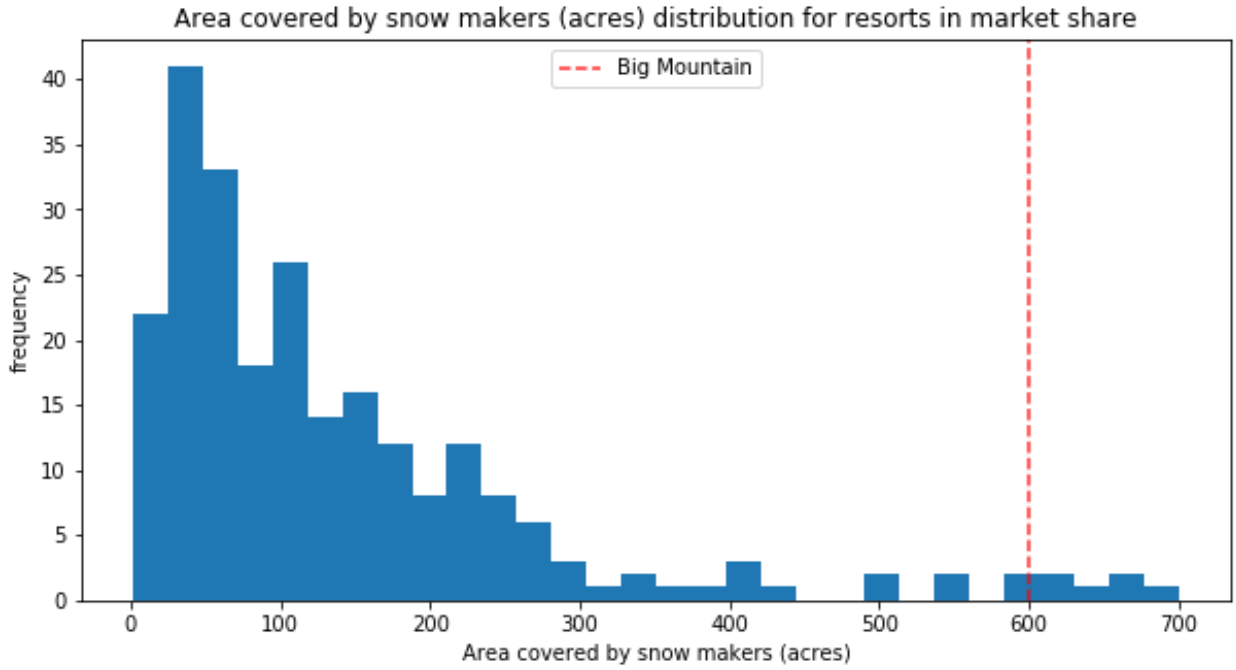
ticket price when fewer resorts serve a population may similarly be because it's a less popular state for skiing. The high price for some resorts when resorts are rare (relative to the population size) may indicate areas where a small number of resorts can benefit from a monopoly effect.



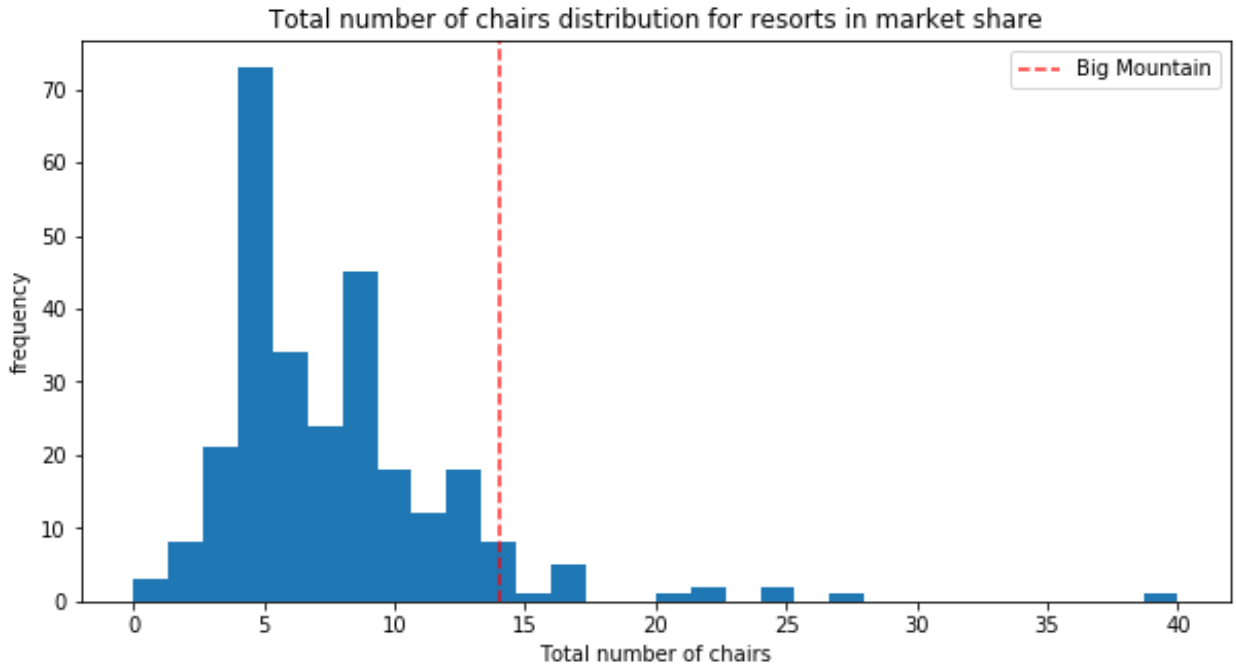
From the above scatterplots, it seems that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. What we may be seeing here is an exclusive vs. mass market resort effect; if you don't have so many chairs, you can charge more for your tickets, although with fewer chairs you're inevitably going to be able to serve fewer visitors. Hence, we need information on the number of visitors per year for Big Mountain Resort.



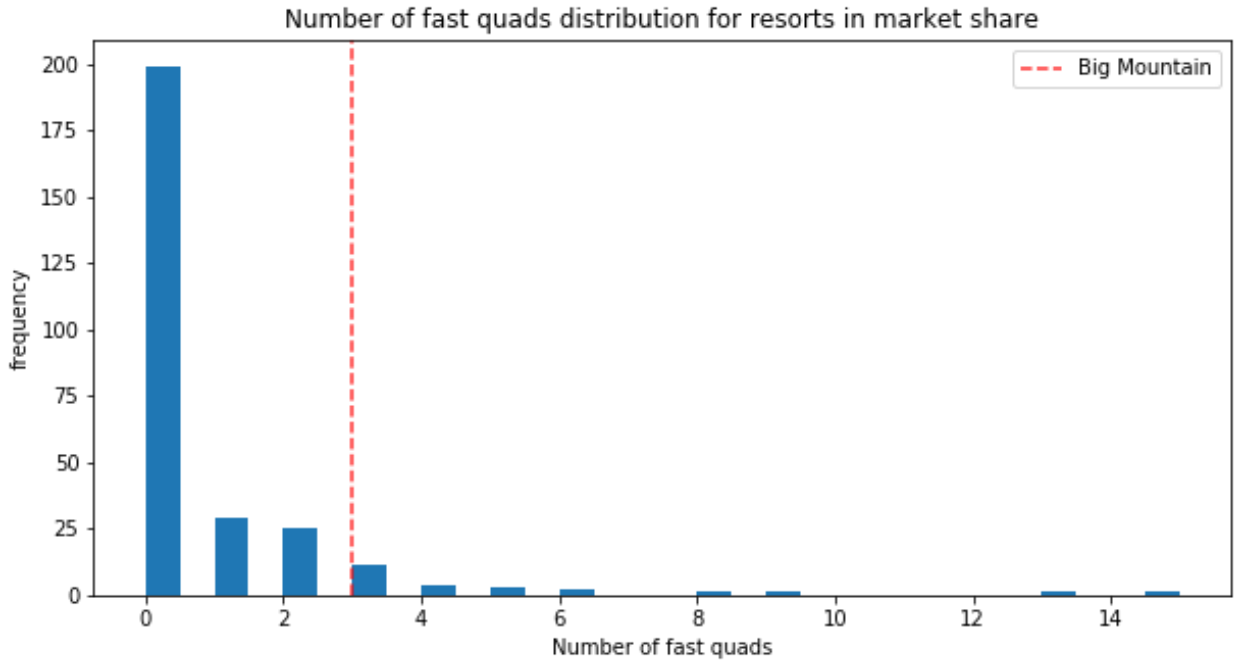
Big Mountain is doing well for vertical drop, but there are still quite a few resorts with a greater drop.



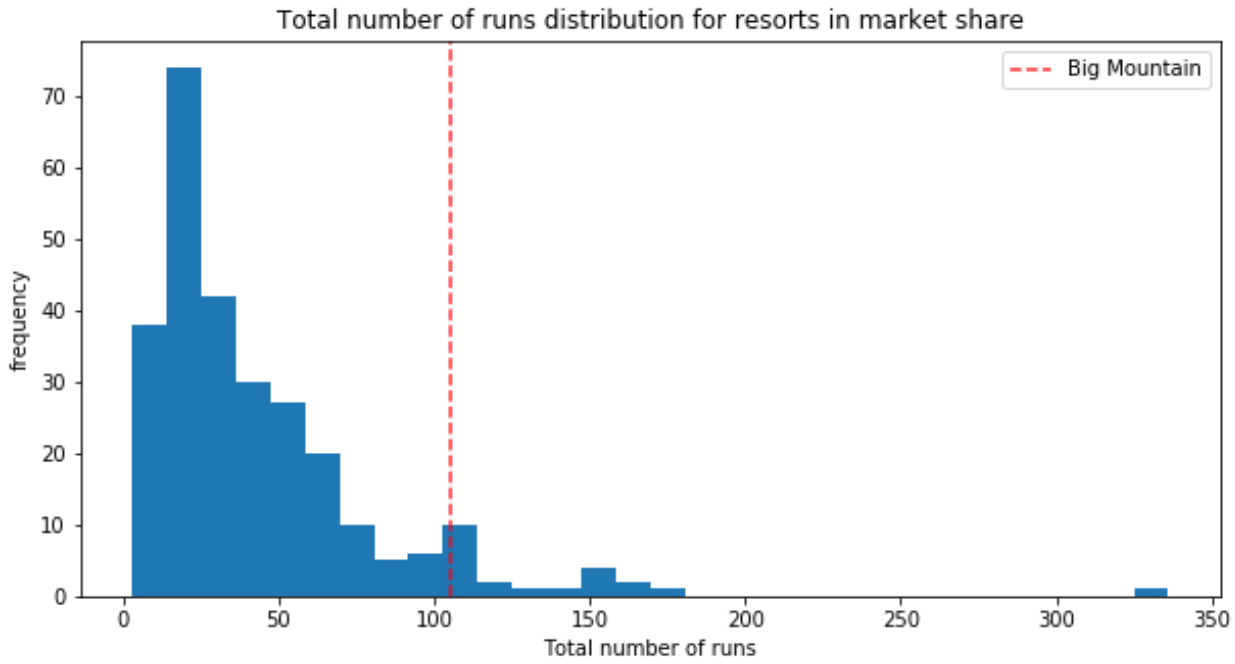
Big Mountain is very high up the league table of snow making area.



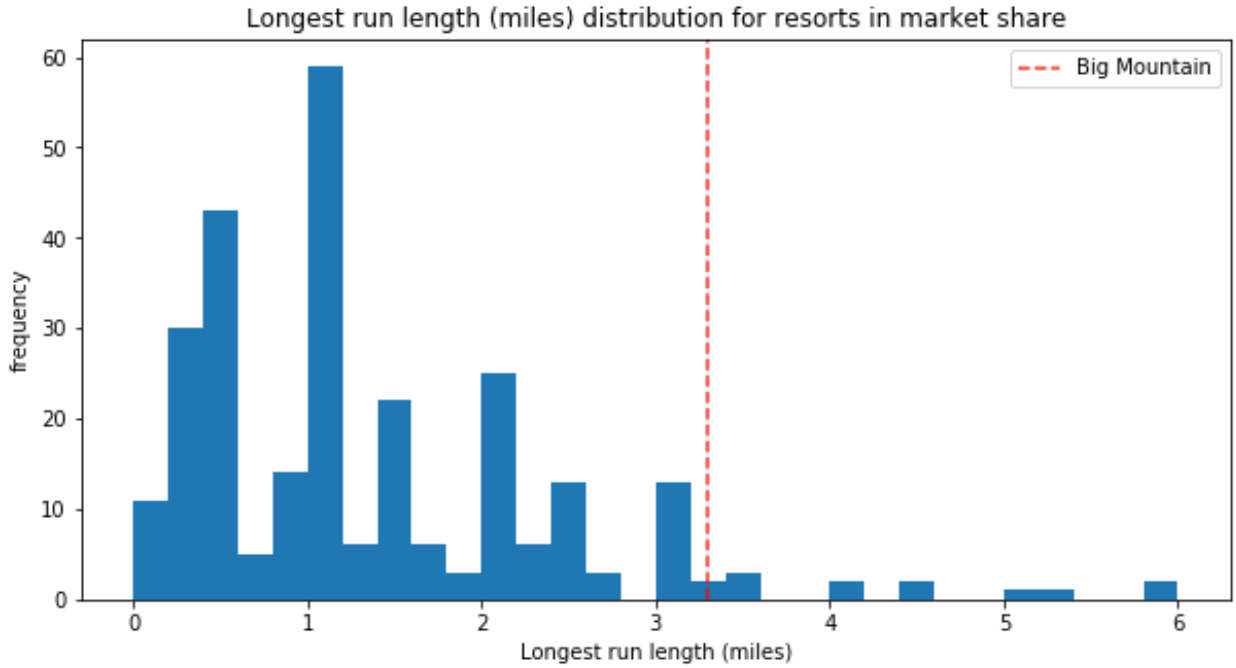
Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers.



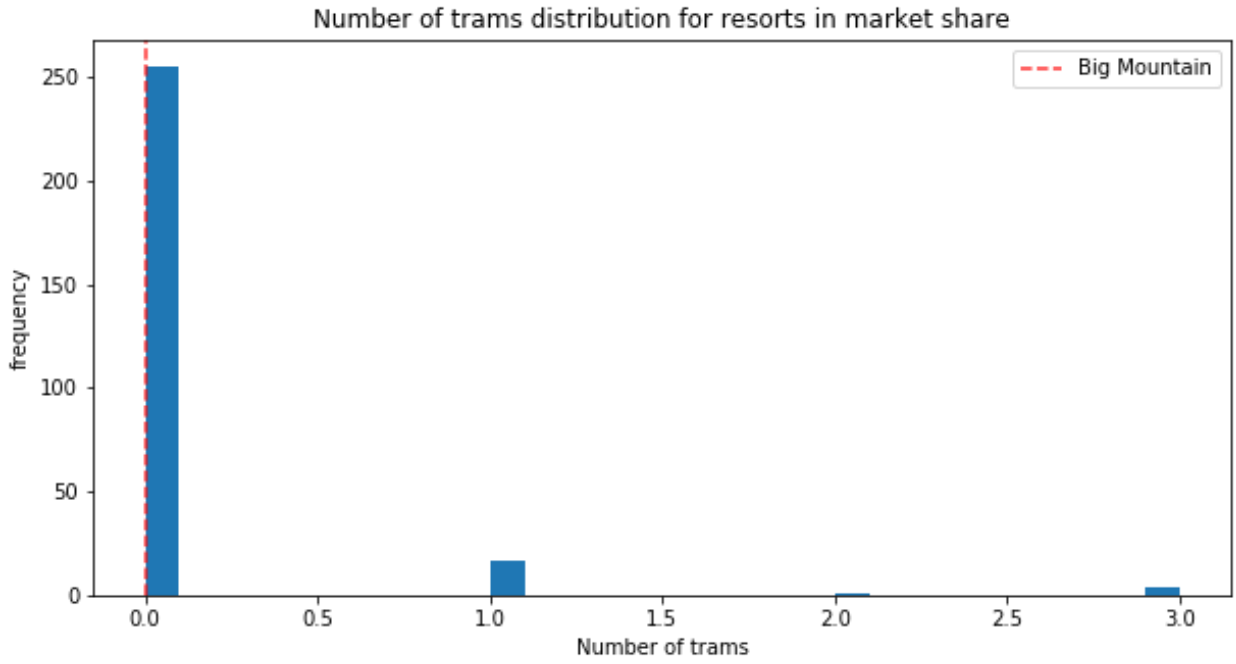
Most resorts have no fast quads. Big Mountain has 3, which puts it high up that league table. There are some values much higher, but they are rare



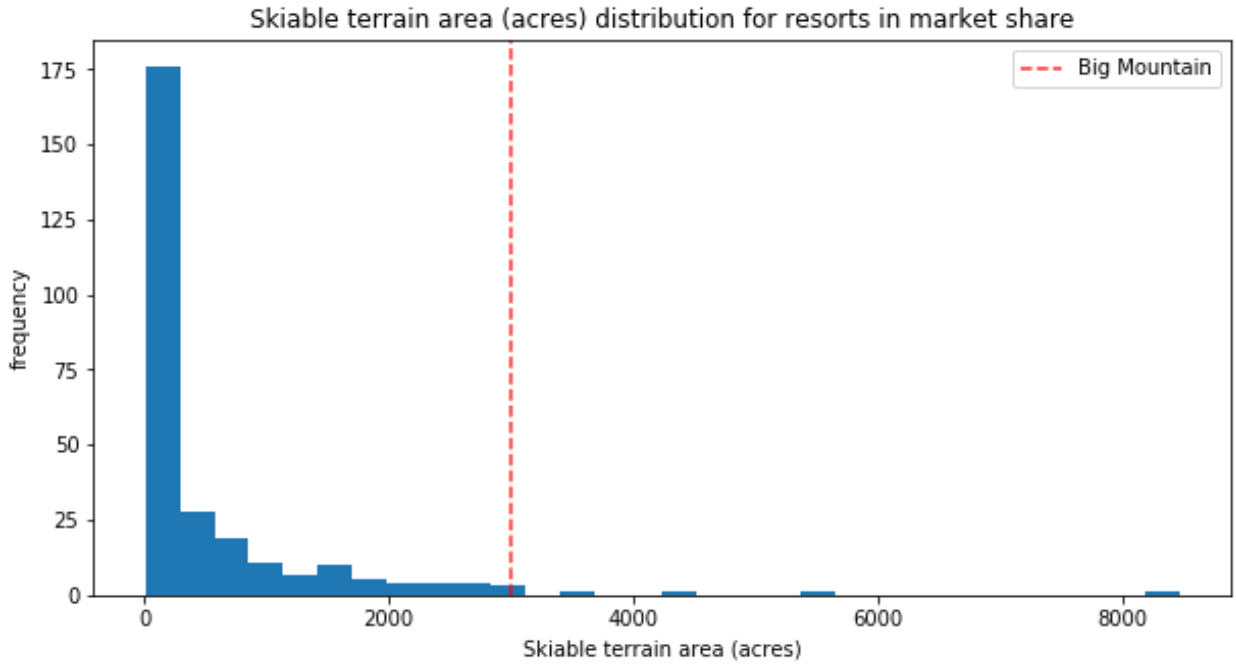
Big Mountain compares well for the number of runs. There are some resorts with more, but not many.



Big Mountain has one of the longest runs. Although it is just over half the length of the longest, the longer ones are rare.



The vast majority of resorts, such as Big Mountain, have no trams.



Big Mountain is amongst the resorts with the largest amount of skiable terrain.

Conclusion:

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. 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. If Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift, then ticket price can increase by $1.99. Over the season, this could be expected to amount to $3474638.If we add 2 acres of snow making, there occurs a very small increase in ticket price and henceforth, snow making area makes no difference. Also, increasing the longest run by 0.2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability makes no difference whatsoever. Although the longest run feature was used in the linear model, the random forest model (the one we chose because of its better performance) only has longest run way down in the feature importance list.