**Big Mountain Resort**

**Montana**

**A picture containing text, sign

Description automatically generated**Libi Voshin

This project report will focus on :

**Which pricing strategy should Big Mountain Resort apply in the upcoming season to increase its Revenues and return its investment of $1.54 Million?**

# Problem identification

Our data science team was brought in to implement a data-driven model that would help to answer the following business questions.

Recently resort invested in a chair-lift, which has increased its operating costs by $1.54M this season. Considering this new addition, to maximize its returns relative to its position in the market, we investigated **What price should Big Mountain Resort charge for their ticket?** Resort's currently pricing strategy has been to charge a premium above the average cost of alternatives in its market segment. This report will present answers to whether it can increase the price even higher and, if so, in how much, so visitors will still be willing to pay the price.

There's a suspicion that Big Mountain is not capitalizing on its facilities as much as it could. We will be helping to determine **How to capitalize on its facilities in the best possible way**. Resort's management has presented us with four changes that they are considering. Our model tested these propositions and predicted how each scenario would influence current consumers’ willingness to pay for a ticket. I will present **Which of the considered changes will cut costs without undermining the ticket price and which change will support an even higher ticket price**.

Finally, based on the data we have gathered on the existing list of facilities on 330 resorts belonging to the same market share (See appendix B- assumptions section), we will provide a user-friendly model that would support future business decisions regarding the influence of future change to resort's facilities on ticket value, As well as, **How to adjust their ticket price based on those changes**.

# Recommendation and key findings

Since each visitor on average buying 5-day tickets and each year there are about 350,000 visitors at the resort, Our modeling suggests that charging $94.22 per ticket could be fairly supported in the marketplace by Big Mountain's facilities. Solely to cover the resort's recent investment, it should raise ticket price by $0.88 per ticket.

Features that came up as highly valued by customers (see Appendix A) include:

* Number of Fast Quads
* Number of Runs
* Snow Making area
* Vertical drop
* Skiable terrain area
* Total number of chairs

After seeing where Big mountain resort stands amongst those areas, we feel confident that existing resorts' facilities can support the higher ticket price.

For further improvements, we would recommend the following:

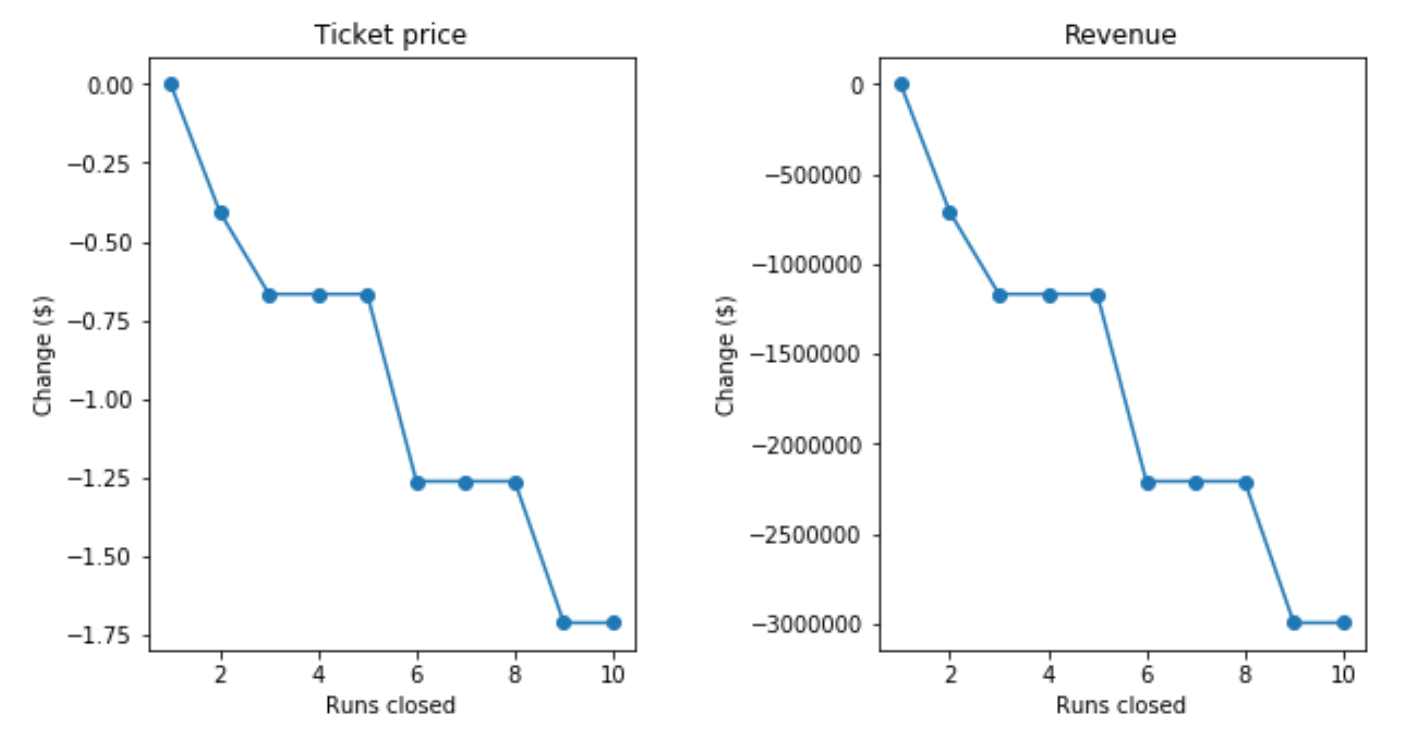
1. First, to permanently close the one least used run, as shown from the model that it is not reducing ticket value.
2. Then, our model shows that increase the vertical drop by 150 feet by adding a lower run supports a $1.99 increase in the ticket price, which over the season could be expected to amount to revenue of $3.475 M. Implementation of this scenario requires the installation of an additional chair-lift to bring skiers back up. Resort's previous experience shows that the price should be raised to $0.88 per ticket to cover additional chair-lift yearly operation costs.

Hence, It is an **extra profit of more than $2M** (not include the additional chair-lift installation cost).   
Note that only after knowing the additional chair-lifts yearly operational costs could we calculate the profit. It is also worth mentioning that adding 2 acres of snowmaking coverage makes no difference in ticket value. Therefore, it is not worth the investment.

1. Lastly, the resort should look up at its runs' operational costs and make a decision when should it permanently close its least used runs according to this decision table:

|  |  |
| --- | --- |
| **Condition**:  Yearly operational costs of.. | **Decision**: Permanently close down.. |
| 2 runs > $675K | 2 least used runs |
| 5 runs > $ 1.225 M | 5 least used runs |
| 8 runs > $ 2.2 M | 8 least used runs |

As we can see in the graph below, If Big Mountain closes down three runs, it seems they may as well close down 4 or 5 as there's no further loss in the ticket. So as 6,7 and 8.



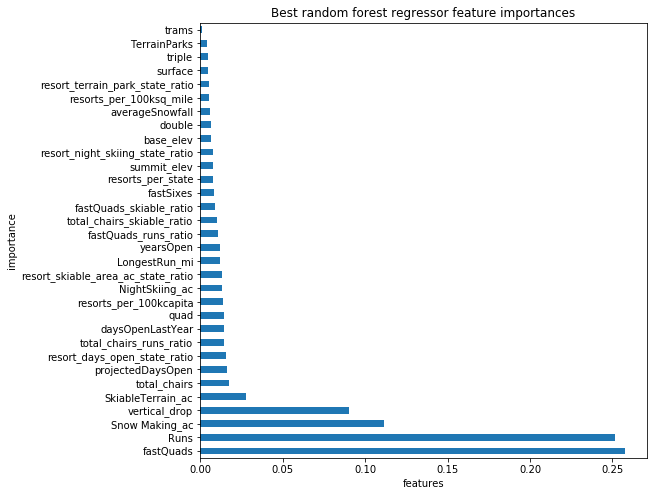
# Summary and conclusion

Currently, Big Mountain charges $81, and although state-wise, Big Mountain's ticket price is the highest, and so it sits high amongst all resorts, there are still resorts with a higher price and up to double the price. Note that this relies on the implicit assumption that all other resorts are primarily setting prices based on how much people value certain facilities. Essentially this assumes a free market sets prices.

As we saw, the ticket price is not determined by any set of parameters. The resort is free to set whatever price it likes. However, the resort operates within a market where people pay more for some facilities and less for others. Being able to sense how facilities support a given ticket price is valuable business intelligence. Thus, the utility of our model comes in.

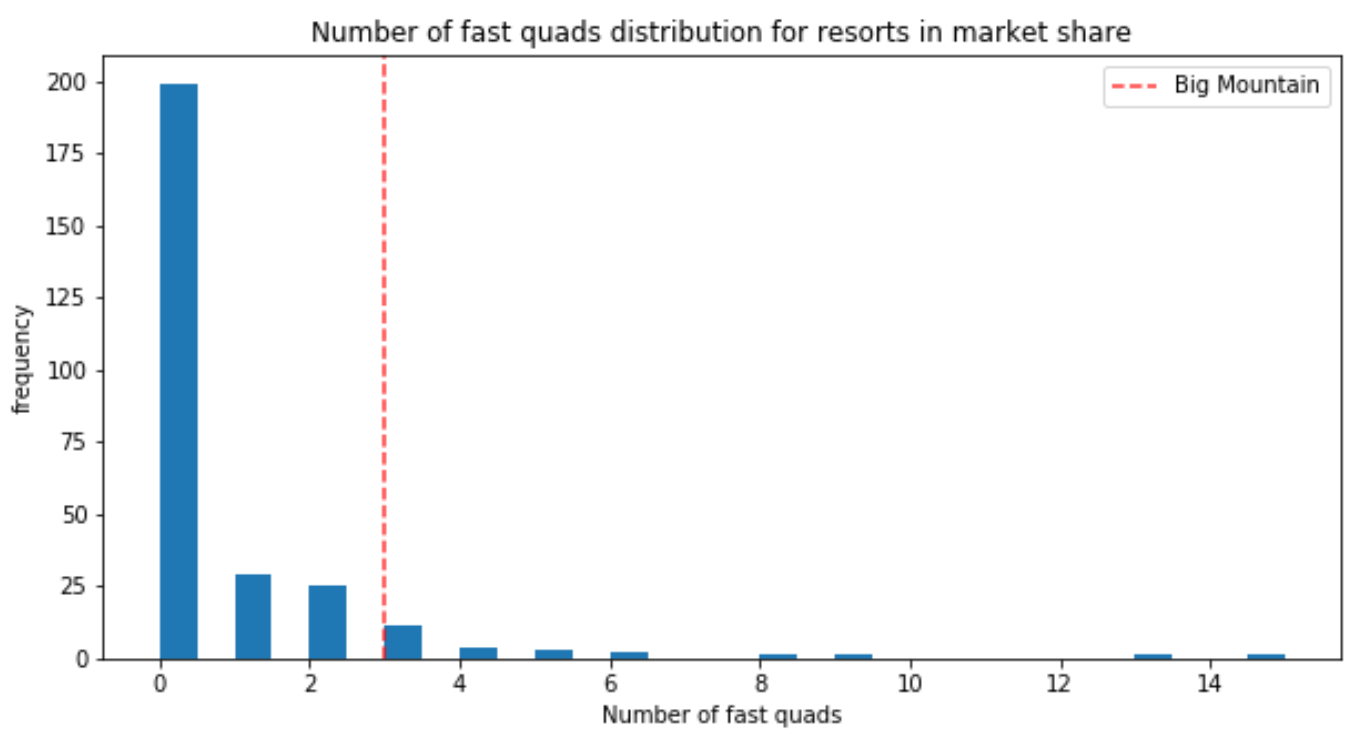
For future queries and new scenarios examinations, management can use our model to obtain the predicted increase of ticket prices from each unique scenario. all that is needed is to insert a list of all the features/facilities that will be affected by the new plan and their corresponding deltas, to the function that we supplied ("predict\_increase") to receive the predicted increase in $ (or decrease for negative results).

**Appendix A**

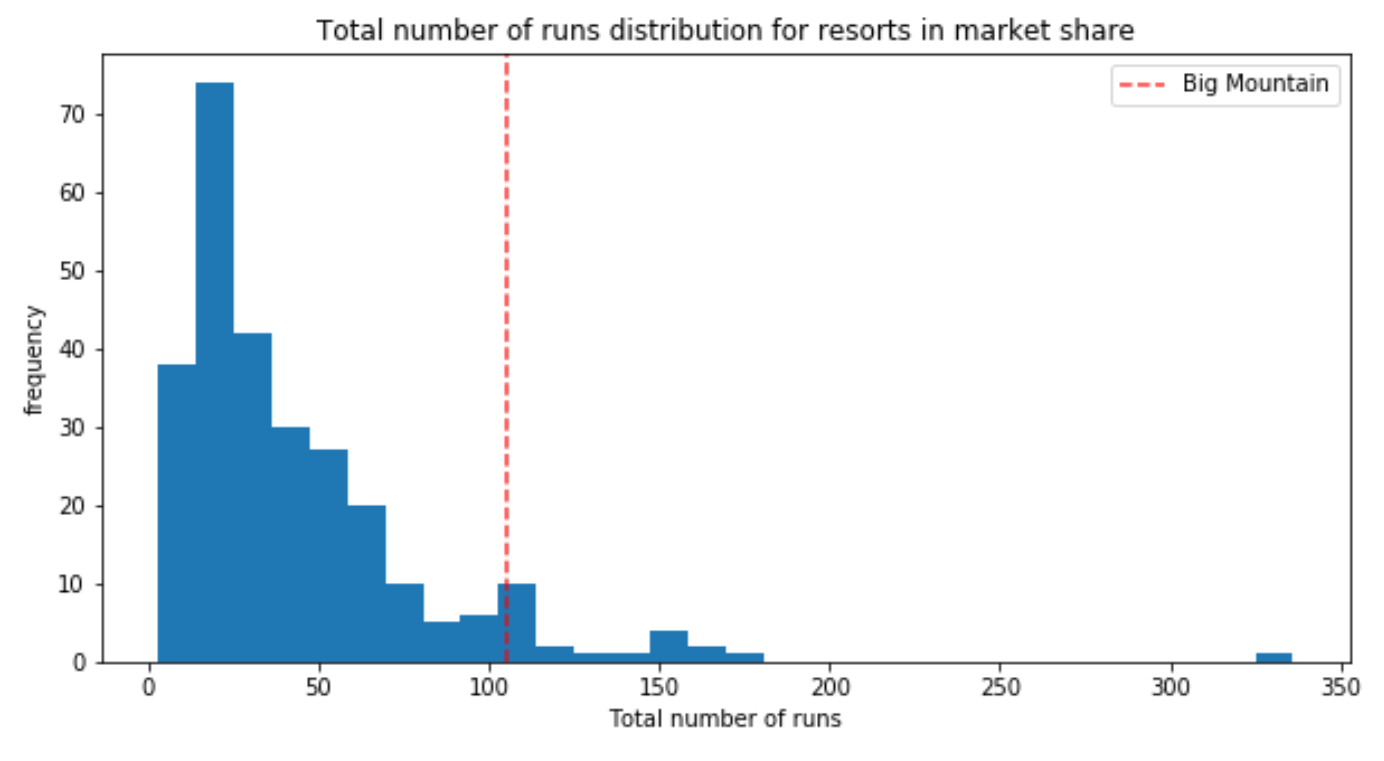


**Figure 1.** Features that came up as highly valued by customers in our best modeling

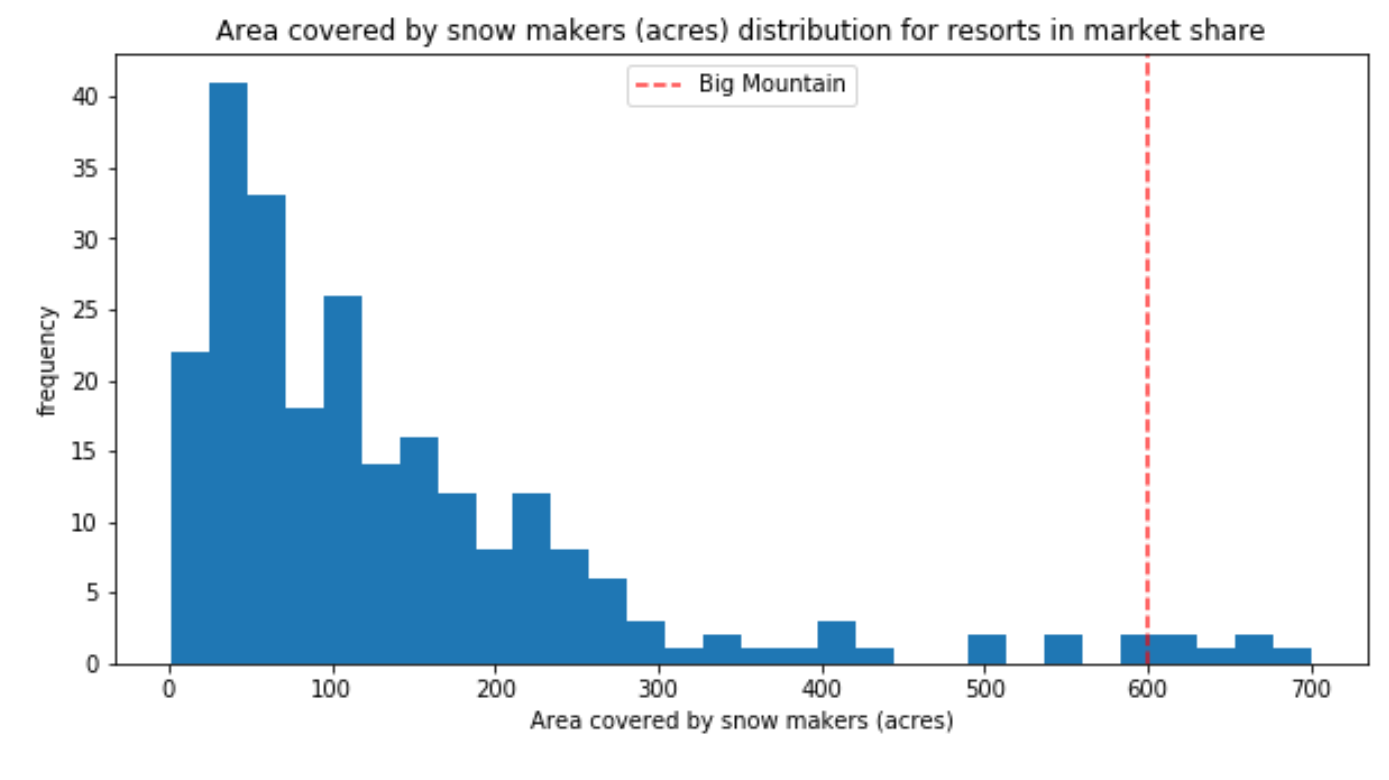
1. The number of Fast Quads - Most resorts have no fast quads. Big Mountain has 3.



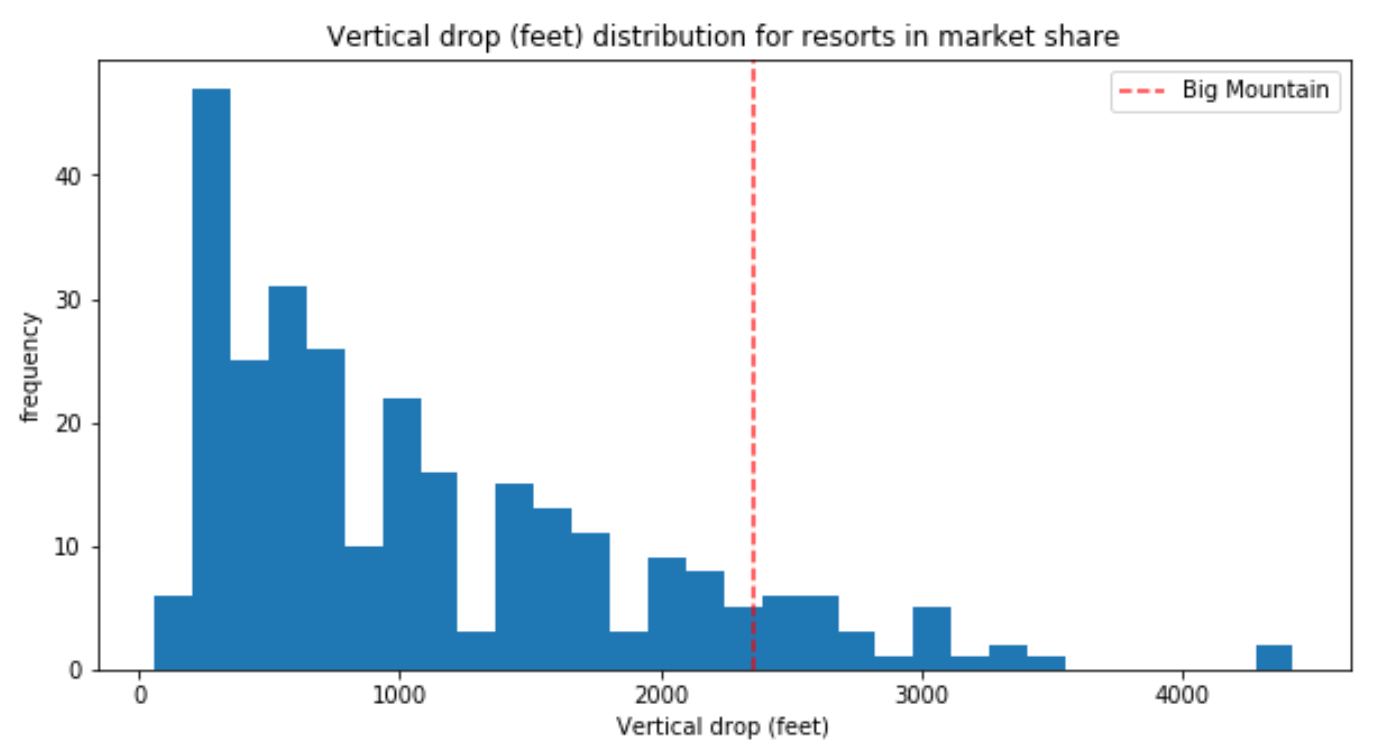
1. Runs - Big Mountain compares well for the number of runs



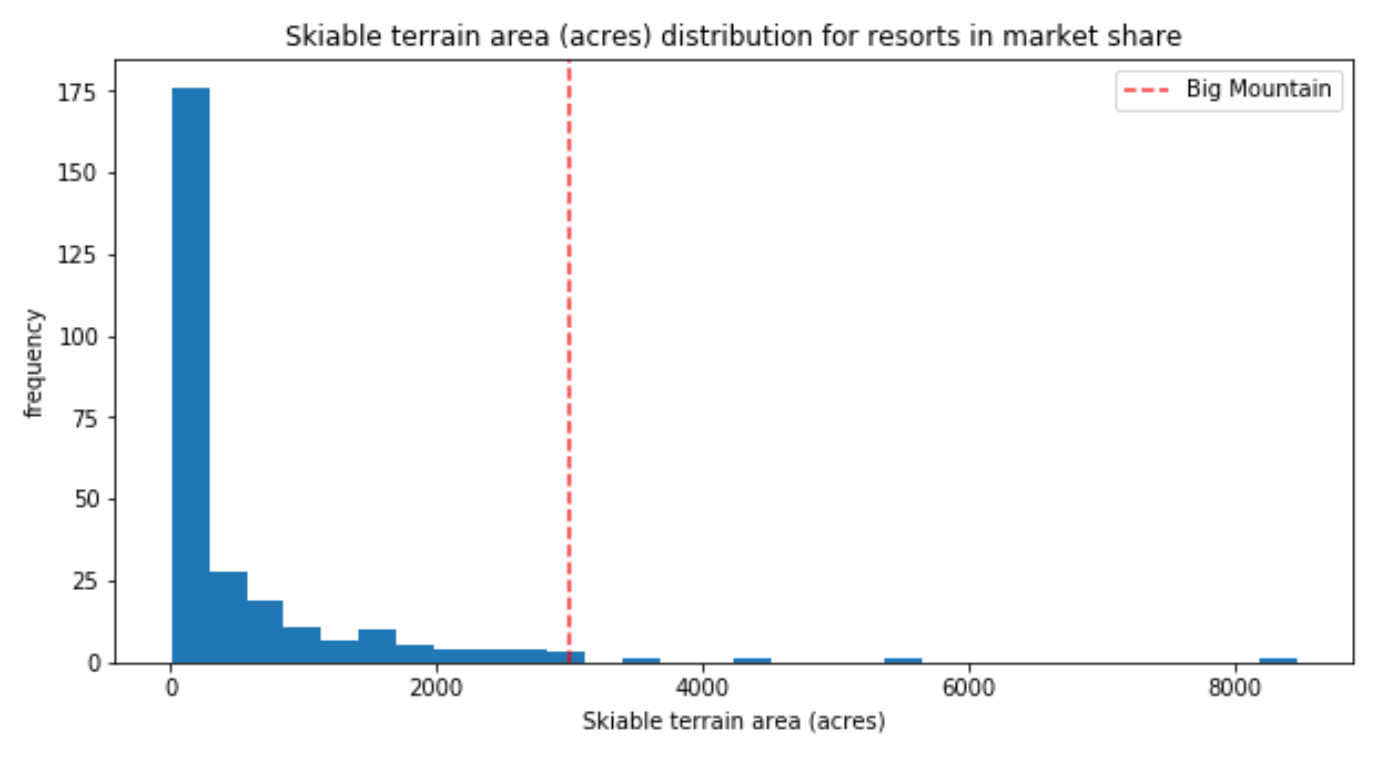
1. Snow Making area – Big Mountain resort is very high up the league table of snowmaking areas.



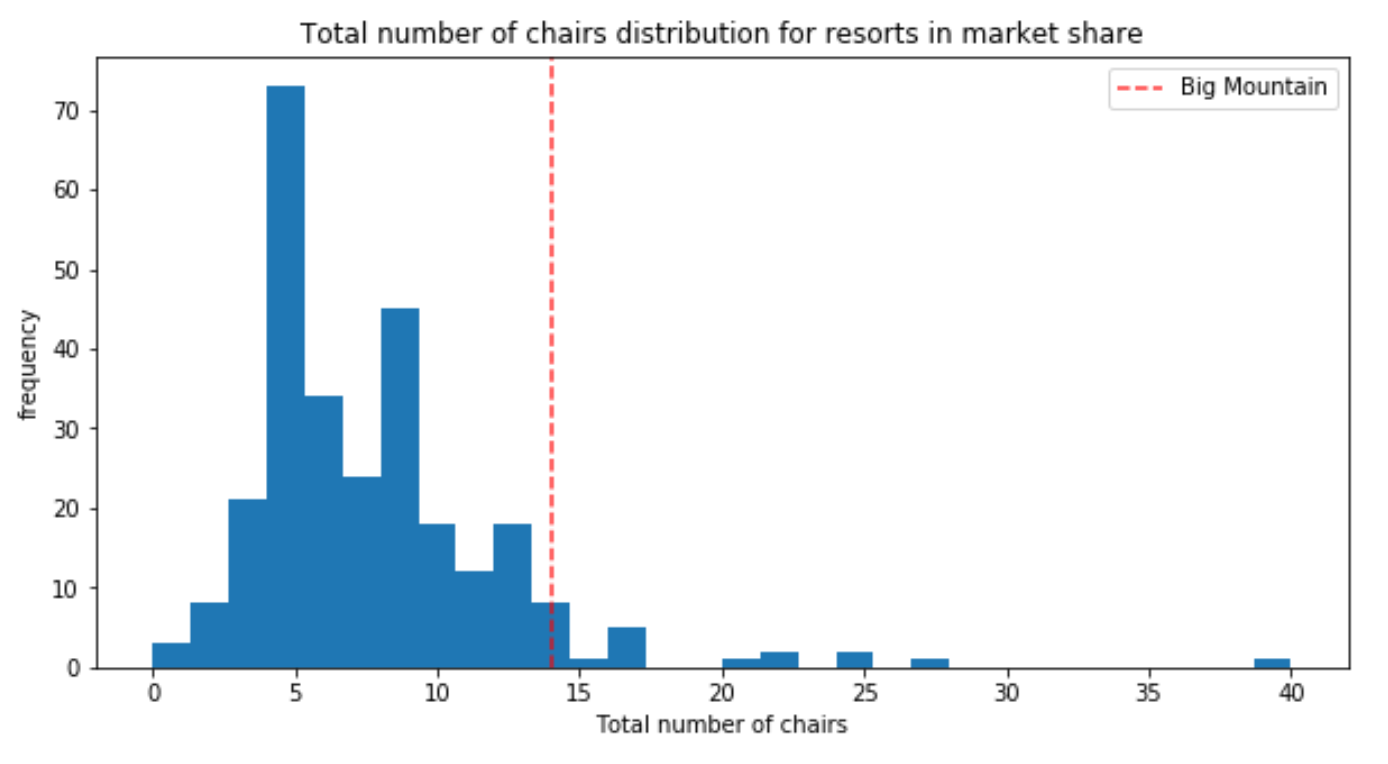
1. Vertical drop - Resort is doing well, and there are only several others with a more significant drop



1. Skiable terrain area - Big Mountain is amongst the resorts with an enormous amount of skiable terrain.



1. Total number of chairs - Big Mountain has amongst the highest total number of chairs



**Appendix B**

## Assumptions

1. Based on:
   1. The contextual insight that the provided data are for resorts all belonging to the same market share suggests that prices should be similar amongst them
   2. I haven't seen any clear grouping that would justify treating states differently

We built our model to considers all resorts within all the given states together.

1. Although typically resorts are separate between weekday and weekend prices, Big Mountain resort amongst all Montana resorts is not.
2. Weekend vs. weekday ticket:
   1. Some States have weekend prices far higher than weekday prices. In Montana, on the other hand, the data shows that it is customary to match weekend and weekday ticket prices.
   2. Weekend prices being higher than weekday prices seem restricted to sub $100 resorts
   3. After cleaning all rows that were missing pricing information, between weekday and weekend prices, Weekend prices have the least missing values of the two (four vs. seven missings)

From the mention above, we chose to focus our model on the weekend prices.

## Data quality and "data correction" steps

* **Data correction:**

**Skiable area** in ac caused concern because values were clustered down the low end. after further exploring, Silverton Mountain Resort had highly suspicious 26,819 skiable terrains in ac. The value I've looked up in other data sources was 1819 ac. This led to the data correction step: the suspect value was replaced with the new one I've obtained.

**Snow Making\_ac** forHeavenly Mountain Resort is supposedly 2880 rather than 3,379, but because it's missing ticket pricing information, it was part of the records that we dropped

* **Dropping columns:** During the data cleaning process, we removed some feature columns. these include:
  1. **fast Eight** - half the values are missing, and all but one resort having zero 'fastEight' seats. There is essentially no information in this column
  2. **Weekday** - Weekend prices have the least missing values of the two. Therefore we chose to drop the weekday prices column
* **Dropping Lines**: But not before making the most of the other available data to look for any patterns between the states
  1. "Pine Knob Ski Resort," Michigan, was removed due to its unclear Years Open data. (Years Open =2019) - no resort will have been open for 2019 years, and because we don't know when this data was gathered, we don't know whether it has been open for zero years or two years. anyway, the smallest number of years open otherwise is 6, and we decided not to consider a resort that may not have opened yet, or perhaps in its first season.  (-1 row)
  2. 'Heavenly Mountain Resort' had outlier' snow making ac' value. but because it already lacked any pricing info, I decided to drop this row instead of looking it further. (-1 row)
  3. After deriving all state-wide statistics, we dropped **all rows with missing price information**, which was about 14.33% of the rows. (-**47** **rows**)
  4. dropping the remaining four records that were missing only weekend price values (-4 rows)

The data we started with contained several missing values that led to several rows being dropped completely (52 rows) - **After data cleaning, we have 277 resort records**

## Data quantity assessment

* Is further data collection needed?



We can see how performance varies with different data set sizes, There's an initial rapid improvement in model scores, but it's essentially leveled off by around a sample size of 65-75.

70% of 276 records is 193 records (Big mountain not included), that is enough data for the model.

* Additional data that will help are:
* yearly visitors number
* facilities operational costs

## Ski Resort Features we worked on included:

By knowing that state-wide supply and demand of certain skiing facilities may factor into pricing strategies, We derived state-wide summary statistics to try and answer whether our resort dominates certain domain. state-wide derived data included:

1 resorts\_per\_state

1. state\_total\_skiable\_area\_ac
2. state\_total\_days\_open

4 state\_total\_terrain\_parks

5 state\_total\_night\_skiing\_ac

1. resorts per 100K capita (derived from ‘state\_population ‘)
2. resorts per 100L sq miles (derived from ‘state\_area (sq miles)’)

After diving into **resort-level** data:

1 resorts\_per\_state

2 resort\_skiable\_area\_ac\_state\_ratio

3 resort\_days\_open\_state\_ratio

4 resort\_terrain\_park\_state\_ratio

5 resort\_night\_skiing\_state\_ratio

6 resorts per 100K capita

7 resorts per 100L sq miles

Another feature that may be useful is how easily a resort can transport people around. from the numbers of chairs and runs. We formed the ratios of chairs to runs and skiable area. Those ratios would inform us how easily and quickly people could get to their next ski slope

* 1. total\_chairs\_runs\_ratio
  2. total\_chairs\_skiable\_ratio
  3. fastQuads\_runs\_ratio
  4. fastQuads\_skiable\_ratio

## Models' description

* **We built a best linear model and a best random forest model**
* The random forest model has lower cross-validation mean absolute error by almost $1 on the train set and a lower MAE on a test set by more than $2 (Verifying performance on the test set produces performance consistent with the cross-validation results). It also exhibits less variability.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model | Imputing missing values technic | Standard scale | Selection of K | R^2 Performance | R^2 Performance on Test set  after 5-Fold Cross Validation on Train set | | Mean Absolute Error performance  after 5-Fold Cross- Validation on Train set | |
|  |  |  |  |  | Train, Test | mean , std | range of R^2 (mean-/+2std) | Train (mean, std) | Test (mean) |
| 1 | Linear Regression | Median | Yes | - | 0.818 , 0.721 | - | - | 8.548 | 9.407 |
| 2 | Linear Regression | Mean | Yes | - | 0.817 , 0.716 | - | - |  | - |
| 3 | Linear Regression | Median | Yes | K=10 | 0.767 , 0.626 | 0.6606, 0.0657 | [0.53, 0.79] | 9.502 | 11.202 |
| 4 | Linear Regression | Median | Yes | K=15 | 0.792 , 0.638 | 0.6327, 0.0950 | [0.44, 0.82] | 9.21176 | 10.488246 |
| 5 | Linear Regression | Median | Yes | K=8 | 0.762, 0.597 | 0.6815 , 0.0459 | [0.59, 0.77] | 10.499, 1.622 | 11.7935 |
| 6 | Random Forest | Median | Yes | default | - | 0.6385, 0.1444 | [0.3497 , 0.927] |  | - |
| 7 | Random Forest | Median | None | number trees: 69 | - | 0.7082, 0.0656 | [0.5769 , 0.8395] | 9.659, 1.349 | 9.4955 |

we can also see in the scheme below, there was an initial rapid increase in the score with increase in k, followed by a slow decline after K=8. also noticeable that the variance of the results greatly increases above k=8.

Chart

Description automatically generated

As we are increasingly overfit. we can expect greater swings in performance and that can explain results that seems to be better (model 1 to 4 vs model 5)