Big Mountain Resort Pricing Model

1. PROBLEM STATEMENT

How can Big Mountain Resort increase revenues by revising their ticket pricing strategy to boost their facility investments and compensate for the recent $1.540,000 in operating costs this season?

2. DATASET DESCRIPTION

The original data includes entries for 330 resorts from 38 regions and 35 states with 27 features:

- 3 categorical features: Name, Region, and state
- 24 numerical features
- Target features: *Adultweekday* and *Adultweekend* are ticket prices for weekdays and weekends respectively.

```
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Name               330 non-null    object
 1   Region             330 non-null    object
 2   state              330 non-null    object
 3   summit_elev        330 non-null    int64
 4   vertical_drop      330 non-null    int64
 5   base_elev          330 non-null    int64
 6   trams              330 non-null    int64
 7   fastEight          164 non-null    float64
 8   fastSixes          330 non-null    int64
 9   fastQuads          330 non-null    int64
 10  quad               330 non-null    int64
 11  triple             330 non-null    int64
 12  double             330 non-null    int64
 13  surface            330 non-null    int64
 14  total_chairs       330 non-null    int64
 15  Runs               326 non-null    float64
 16  TerrainParks       279 non-null    float64
 17  LongestRun_mi      325 non-null    float64
 18  SkiableTerrain_ac  327 non-null    float64
 19  Snow Making_ac     284 non-null    float64
 20  daysOpenLastYear   279 non-null    float64
 21  yearsOpen          329 non-null    float64
 22  averageSnowfall    316 non-null    float64
 23  AdultWeekday       276 non-null    float64
 24  AdultWeekend       279 non-null    float64
 25  projectedDaysOpen  283 non-null    float64
 26  NightSkiing_ac     187 non-null    float64
```

3. DATA WRANGLING

In this section, we describe activities performed on the data to organize it and ensure it's well-defined. The following actions are taken to change the data:

1. The skiable terrain area of the Silverton Mountain resort in Colorado was updated
2. The *fastEight* column with many missing values was dropped in its entirety
3. Resorts having a number of years open *yearsOpen* > 1000 were removed
4. Rows having the two prices missing were dropped
5. The weekday prices column was dropped (since it has the most missing values)
6. Rows with a missing weekend price were dropped.
7. 0 duplicate was removed

4. EXPLORATORY DATA ANALYSIS

Summarizing data by state and exploring the top states by order revealed that:

- California dominates the state population figures despite coming in second, far behind Alaska in size.
- The resort's state of Montana was in the top five for size, but it is less densely populated (not among the five most populous states). It is in the top five of skiable areas
- New York leads the way in terms of the number of resorts in the market, but it is absent in the top five of skiable areas. It dominates the area of skiing available at night.
- The ranking of states by total number of days open is similar to that for the number of resorts. New Hampshire is in the top five for total days open, despite being a small state and absent from the top five stations by state.

⇒ New York has more, smaller resorts, while Montana has fewer, larger resorts. Colorado seems to have a reputation for skiing; it's in the top 5 for resorts and in first place for total skiable area.

⇒The northern states are prominent in the top five of our summary statistics

4.1. FEATURE ENGINEERING

The Ski data was updated by:

- Adding new features representing aggregate state information:
    - *resorts_per_state*: number of resorts per state
    - *state_total_skiable_area_ac: total of skiable area in the state*

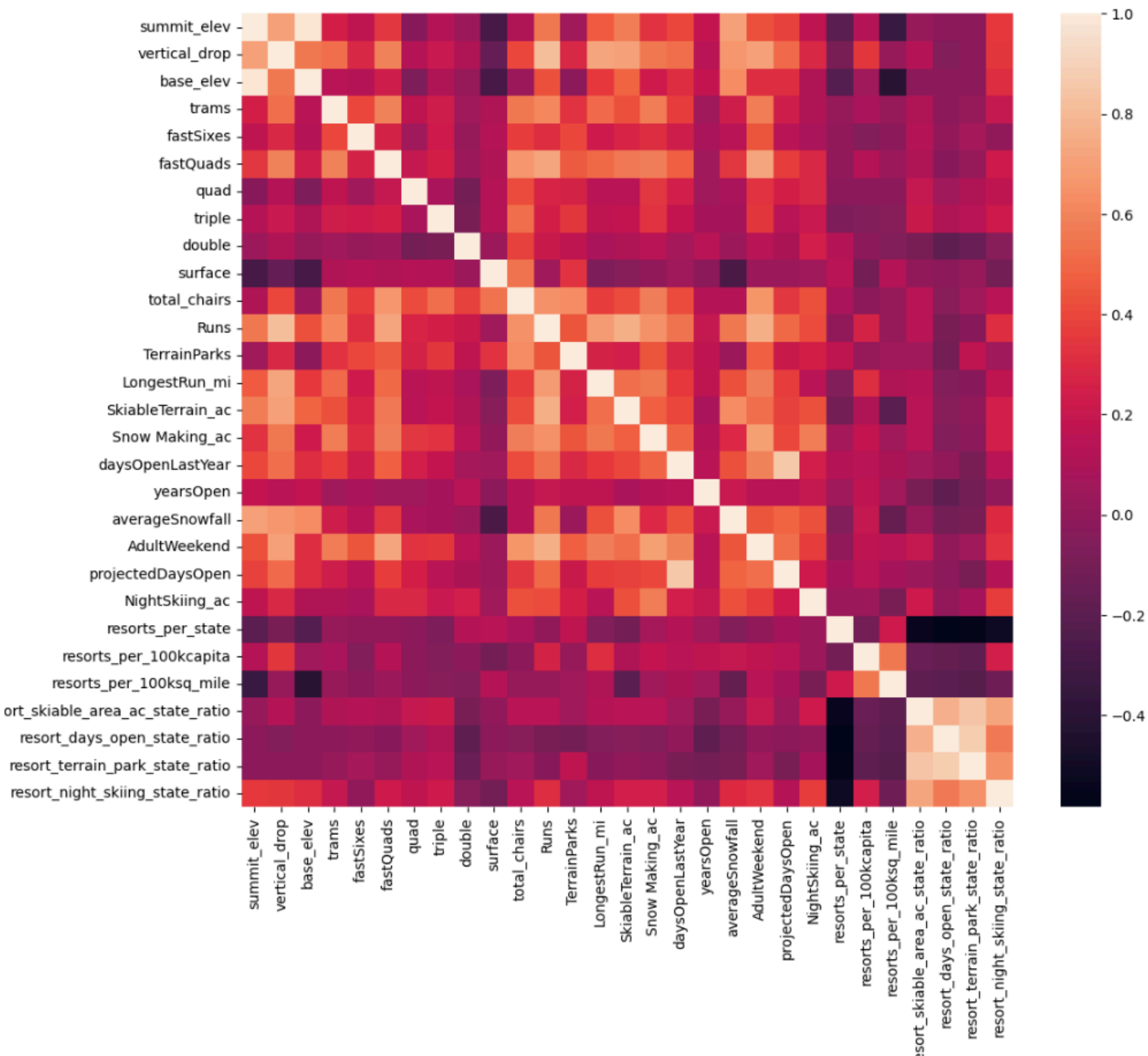
    - *state_total_days_open: total open days at state resorts*
    - *state_total_terrain_parks: total terrain parks per state*
    - *state_total_nightskiing_ac: total of night skiing per state*
    - *resorts_per_100kcapita: number of resorts per 100K capita*
    - *resorts_per_100ksq_mile: number of resorts per 100 sq mile*
- Adding 4 new ratio features allows putting each resort within the context of its state
    - ratio of resort skiable area to total state skiable area
    - ratio of resort days open to total state days open
    - ratio of resort terrain park count to total state terrain park count
    - ratio of resort night skiing area to total state night skiing area
- Dropping the state-relative data
- Adding 4 other features representing:
    - the ratio of chairs over *runs*
    - the ratio of chairs over skiable terrain
    - the ratio of *fastQuads* over *runs*
    - the ratio of *fastQuads* over skiable terrain

4.2 FEATURE CORRELATION

To gain a high-level view of relationships among the features, we present below the feature correlation heatmap. We can observe that:

- There is a strong positive correlation with vertical_drop
- fastQuads seems very useful. Runs and total_chairs appear quite similar and also useful
- When the value of resorts_per_100kcapita is low, the variability of the ticket price is quite significant, ticket price may drop a little before then climbing upwards as the number of resorts per capita increases
    - 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. It's not a clear picture



We next study the relationship between price and the rest of the numerical features, here are our findings:

- There's a strong positive correlation with *vertical_drop*
- *fastQuads* seems very useful. Runs and total_chairs appear quite similar and also useful
- When the value of *resorts_per_100kcapita* is low, the variability of the ticket price is quite significant, ticket price may drop a little before then climbing upwards as the number of resorts per capita increases

  - 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. It's not a clear picture
- The more chairs a resort has to move people around, relative to the number of runs, the faster ticket prices drop and stay low.
- If there are fewer chairs, ticket price could be increased, but with fewer chairs, inevitably fewer visitors will be served. the price per visitor is high, but the number of visitors may be low.
- Having no fast quads may limit the ticket price, but if the resort covers a wide area then getting a small number of fast quads may be beneficial to the ticket price.

5. METHODS AND ALGORITHMS

In this section, we present the tested models to predict the ticket price of the Big Mountain resort.

5.1 PREPROCESSING AND TRAINING

The dataset was split into 70% for training and 30% for testing. The training and test sets were standardized. The main metrics used for validation are R^2, mean absolute error (MAE) and the root mean square error (RMSE).

The current pricing strategy of the Big Mountain resort is based on the mean of other resorts' prices. Using the mean as a predictor, we obtain the following results:

- the R^2 value is zero on the training set and negative on the test set
- The mean absolute error (MAE) is around 19.14, i.e. On average, we can expect an error of around $19 if the ticket price is fixed based on an average of known values.
- The root mean square error (RMSE) is equal to 24.11

Two predictive models were applied namely Linear Regression and Random Forest.

5.1.1. LINEAR REGRESSION

The first predictive model applied was Linear regression. To overcome the overfitting issue, a feature selection method is applied which selects features according to the k-highest scores, and a 5-fold cross-validation was performed to select the optimum value of k. The following results were obtained:

- The results of the different metrics for all folds are uniform, it can be concluded that the model can generalize
- The linear regression model explains over 60% of the variance on the train set
- The MAE is around 11
- The evaluation of model performance is inherently open to variability. We can get different results depending on the quirks of which points are in which fold, enabling us to calculate an estimate of the variability or uncertainty of the performance estimate.
- The 8 most useful features are:

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

- Vertical drop is the biggest positive feature, area covered by snow making equipment is a strong positive too.
- The skiable terrain area is negatively associated with ticket price, people will pay less for larger resorts. It could be an effect whereby larger resorts can host more visitors at any one time and so can charge less per ticket

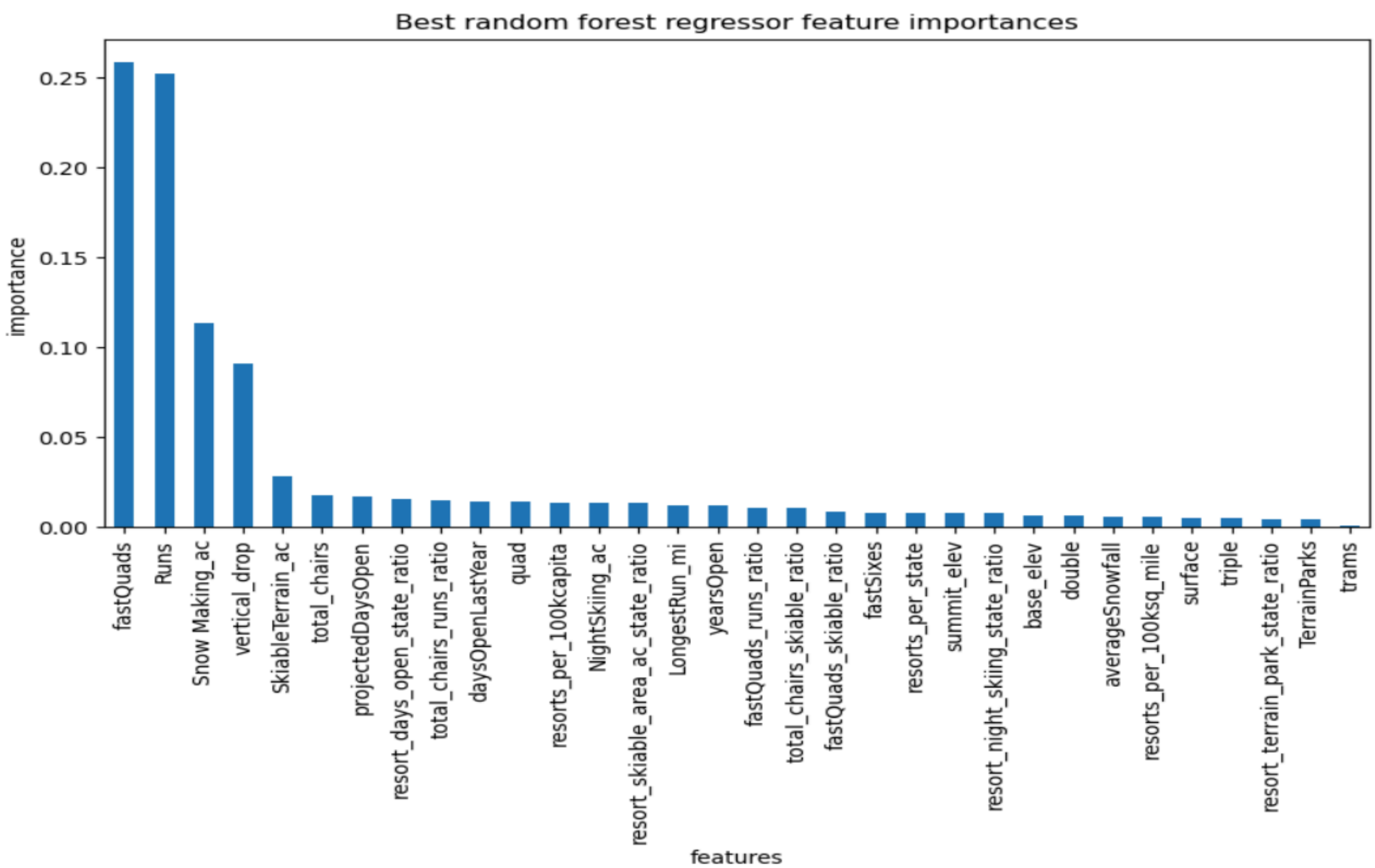
5.1.2. RANDOM FOREST

Random Forest was applied as a second model for regression, to predict the ticket price:

- Performance was assessed using cross-validation as well and it gives a value of about 0.7 for r squared with a standard deviation of 0.071
- A hyperparameter search was conducted to determine the best ones, three parameters were considered:
  - number of trees: a value in [10, 12, 16, 20, 26, 33, 42, 54, 69, 88, 112, 143, 183, 233, 297, 379, 483, 615, 784, 1000]
  - feature scaling: with or without

    ○ strategies for imputing missing values: mean or median
- The best parameters are 69 for the number of trees, the median for missing values imputing, and none for scaling. These parameters give slightly better results than the previous ones (using the model with its default parameters), the r squared is around 0.7 with a standard deviation of 0.06.

Studying the best random forest regressor feature importances shows that the dominant top four features are *fastQuads*, *Runs*, *Snow Making_ac*, *vertical_drop*, those features are in common with the linear model.



5.1.3. MODEL SELECTION

The table below compares the two considered models

Model	Cross validation MAE	Std of cross validation MAE	MAE on test set
Linear regression	10.5	1.62	11.8
Random Forest	9.65	1.35	9.53

The random forest model 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.

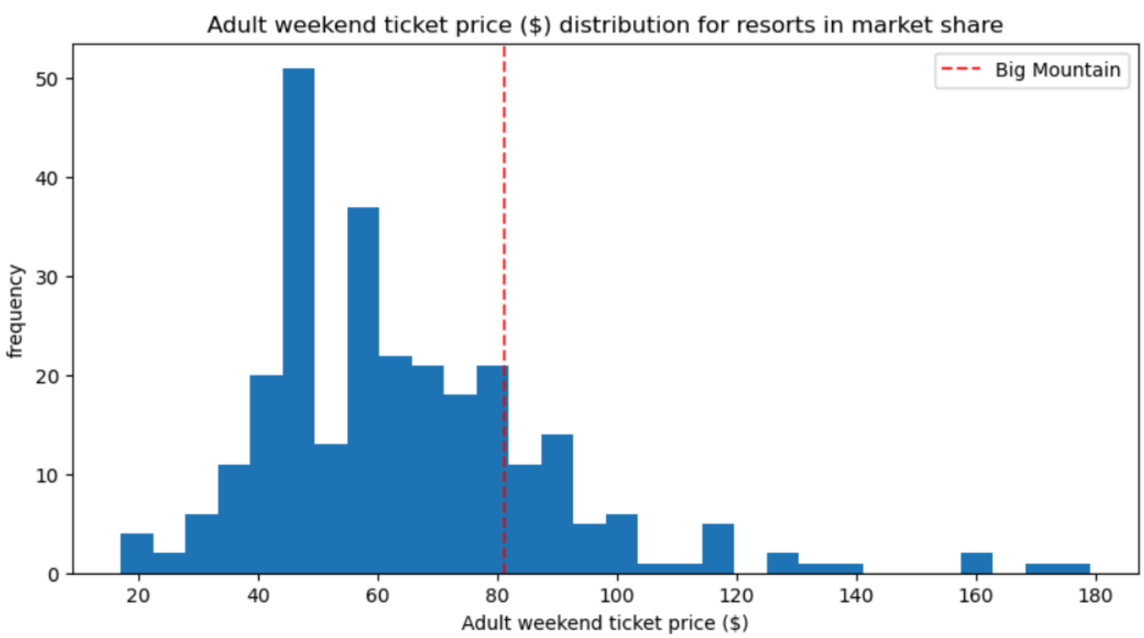
5.2. TICKET PRICE PREDICTION USING RANDOM FOREST

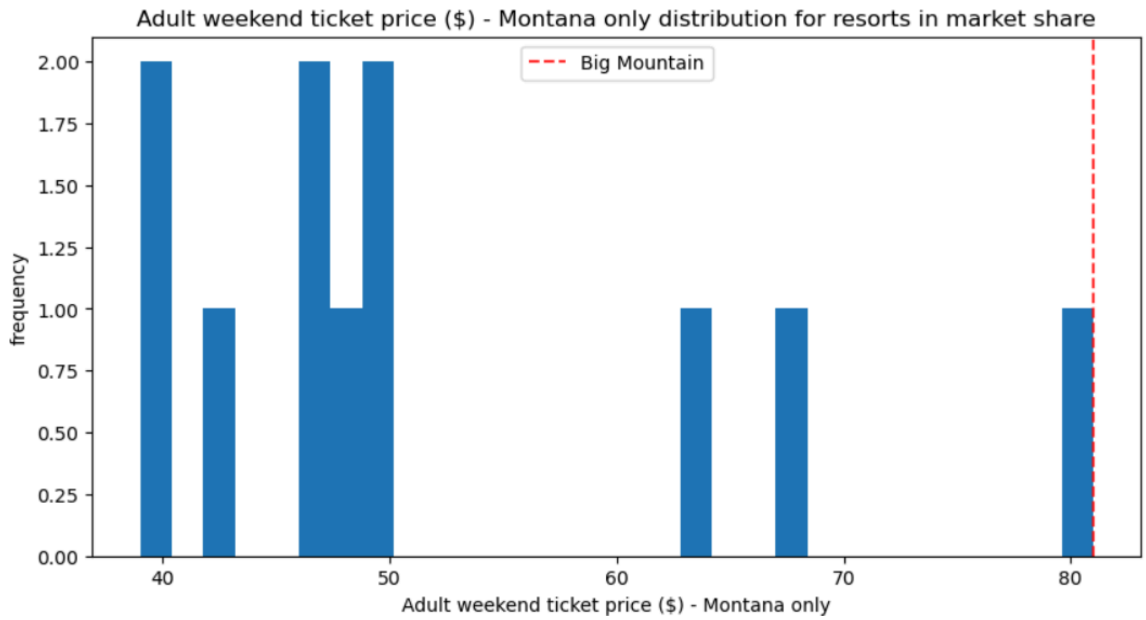
Model prediction suggests Big Mountain resort might be undercharging

- The actual price is **$81.00**, Big Mountain Resort's predicted price is **$95.87.**
- Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

5.3. BIG MOUNTAIN RESORT IN MARKET CONTEXT

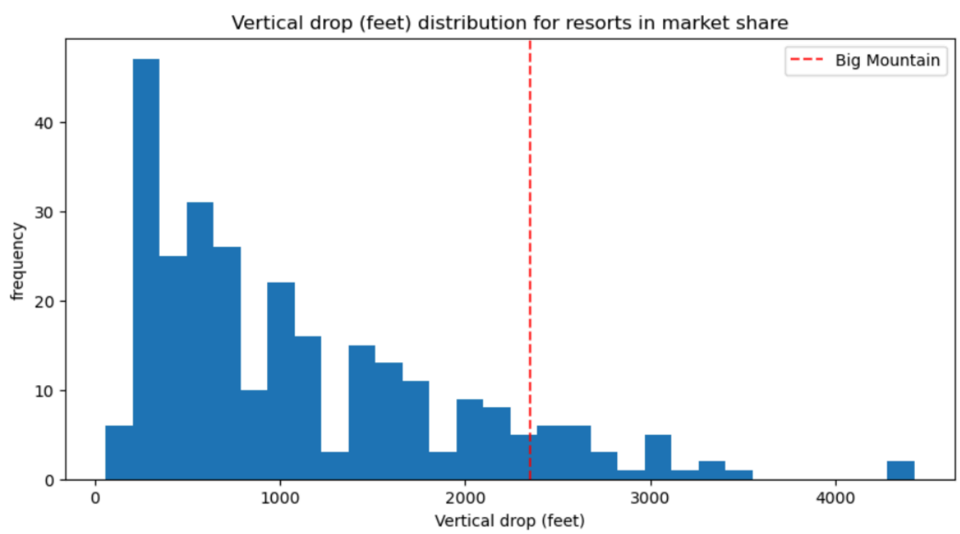
To compare Big Mountain to other resorts on the market, we start by examining Big Mountain's position among all resorts and other Montana resorts in terms of price.



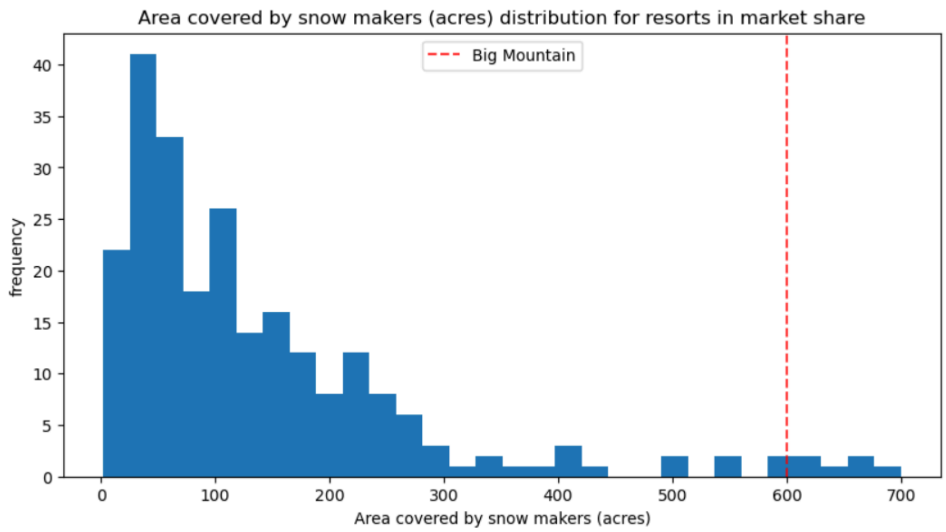


The current Big Mountain ticket price is around the mean of all resort prices and is the highest of all Montana resorts.

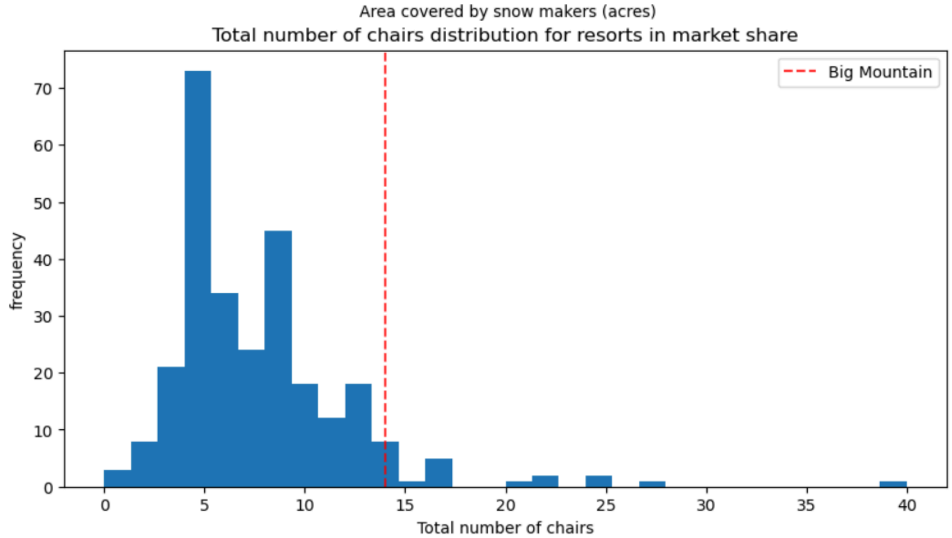
We consider next the 8 most important features to compare the current place of Big Mountain resort to the other resorts.



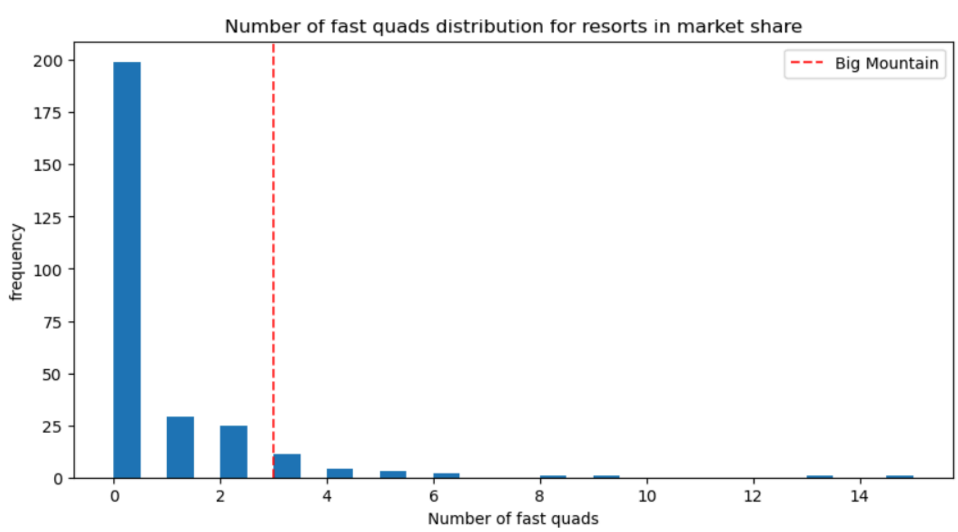
Big Mountain does well in terms of vertical drop, but a few resorts still have a higher vertical drop



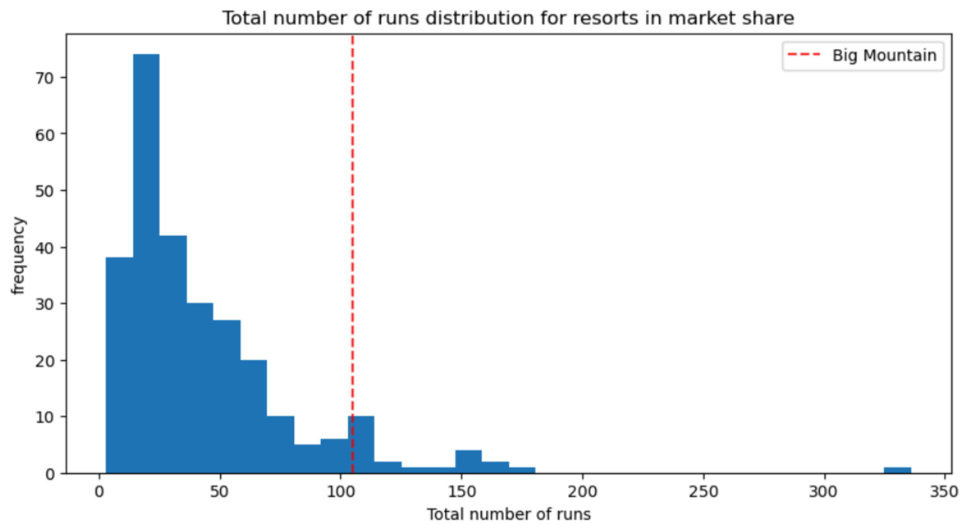
Big Mountain is very well placed in terms of snow-making area.



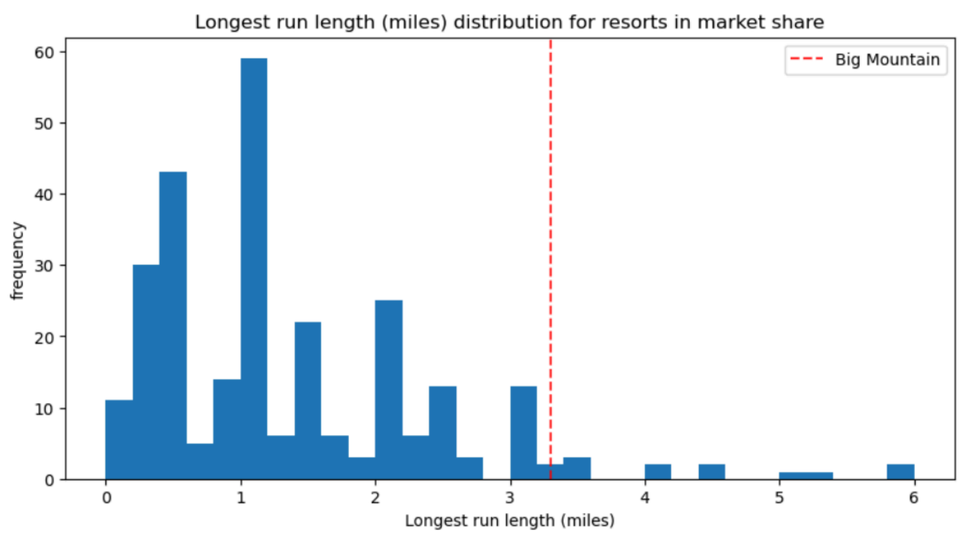
Big Mountain is one of the resorts with the highest number of chairs; resorts with more chairs seem to be outliers.



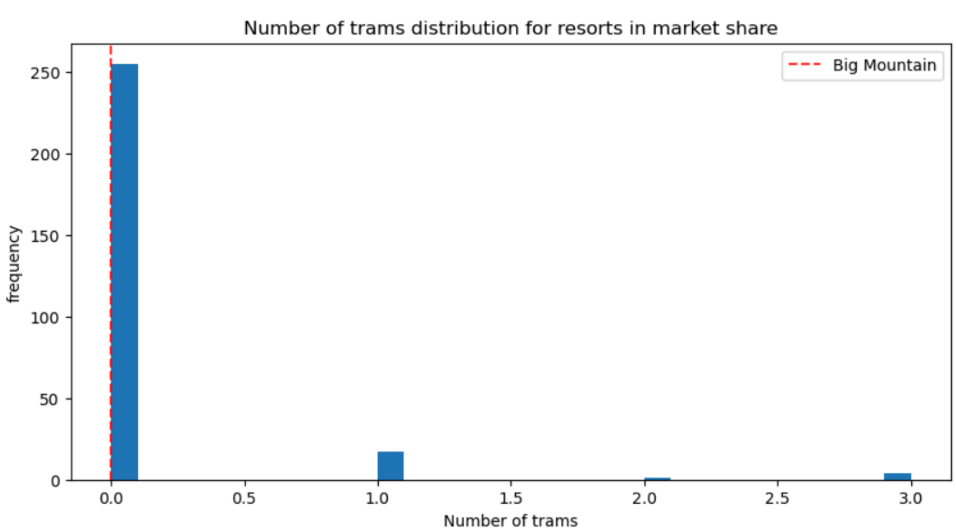
Big Mountain is one of the best resorts with three of them, as the vast majority of other resorts have none. There are some values much higher, but they are rare.



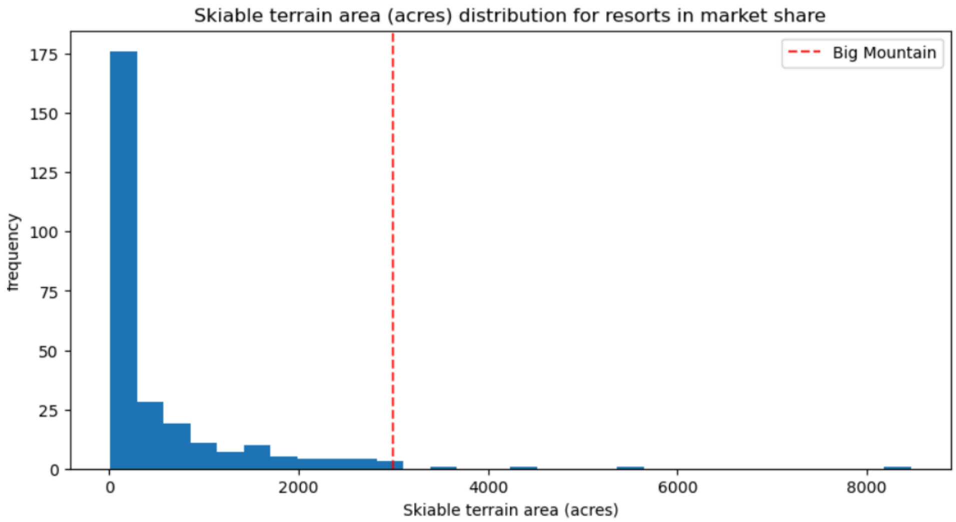
Big Mountain compares well for the number of runs. There are some resorts with more, but not many.



Big Mountain has one of the longest runs. Although it's just over half the length of the longest, the longest trails are rare.



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



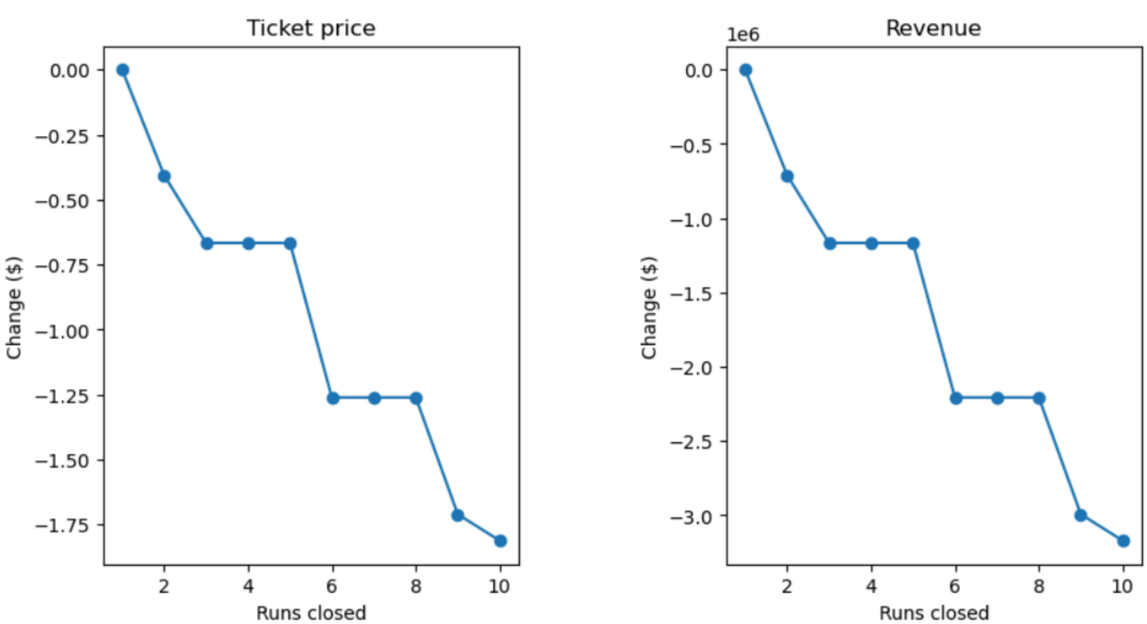
Big Mountain is one of the resorts with the largest skiable terrain.

5.4. MODELING SCENARIOS

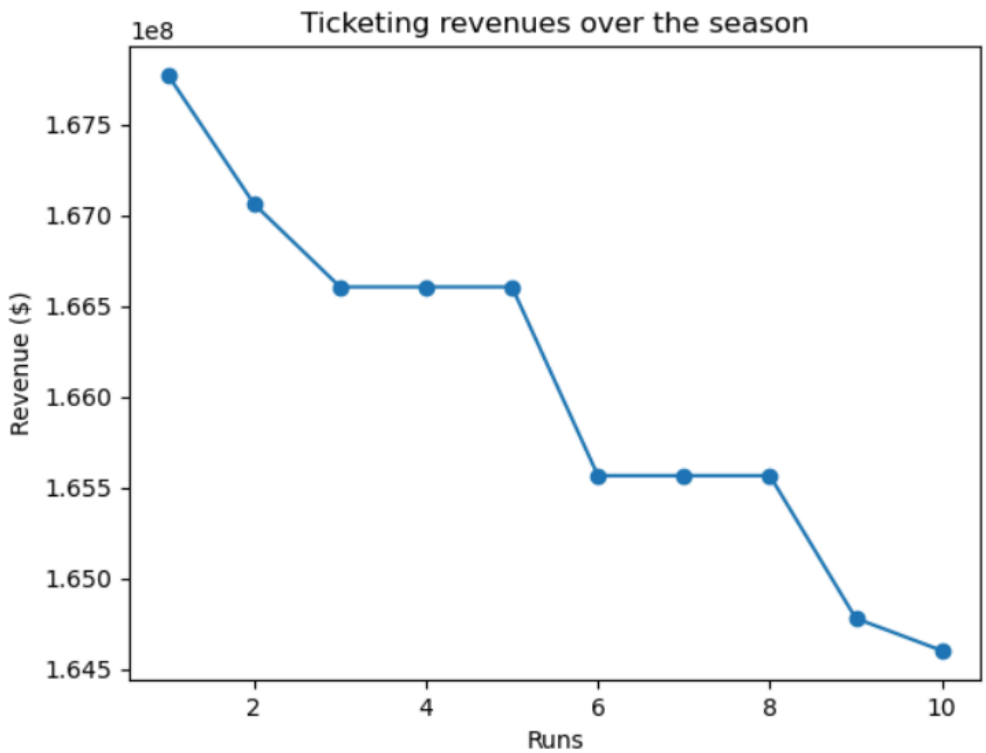
The ability to know how establishments support a given ticket price is valuable business information. Four scenarios were proposed to assess with the assumption that The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days.

5.4.1. SCENARIO 1

This scenario suggests ***closing up to 10 of the least used runs. The number of runs is the only parameter varying.*** We calculate the ticket price increase and revenue increase based on the number of closed runs. The results are given in the figure below.



The ticketing revenues over the season are presented in the figure below.



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.

5.4.2. SCENARIO 2

This scenario stipulates ***adding a run, increasing the vertical drop by 150 feet, and installing an additional chairlift.*** This scenario increases support for ticket price by **$1.99** Over the season, this could be expected to amount to **$3,474,638**

Taking into account the expenses of the additional chair which increases the operating costs by $1,540,000 this season, this scenario can guarantee a profit of **$1,934,638** from the revenue increase, i.e. around **$1,11** per ticket.

5.4.3. SCENARIO 3

The third scenario is **Identical to the previous one** but with the **addition of 2 acres of artificial snow**. Results are Similar to scenario 2, the small increase in the snow-making area makes no difference. Although additional artificial snow does not improve ticket prices, it can increase operating costs (costs of the snow-making machines and/or their depreciation).

5.4.4. SCENARIO 4

In this scenario, we propose to ***Increase the longest run by 0.2 miles and guarantee its snow coverage by adding 4 acres of snow-making capability.*** The predicted increase was null so the proposed adjustments make no difference. This could be explained by the fact that the longest run is among the less important features for the Random Forest model. The proposed changes in this scenario are likely to increase costs without any benefits.

6. CONCLUSION

In this section, we compare the previously presented scenarios and we provide our recommendations.

6.1. SCENARIO COMPARISON

We compare here the different proposed scenarios considering:

- for the first scenario, 1 run is closed
- the predicted ticket price is the new price for Big Mountain which is $95.87.
- The additional operating cost of the new chairlift per ticket is $0.88

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Ticket increase ($)	0.0	1.99	1.99	0.0
Revenue increase over the season ($)	0.0	3,474,638	3,474,638	0.0
costs	No cost	$1,540,000 + C1	$1,540,000 + C1 + 2*C2	C3 + 4*C2
Ticketing revenues over the season (costs excluded*) ($)	167,771,486	169,706,123	169,706,123	167,771,486

- C1: The cost of increasing the vertical drop by 150 feet
- C2: Snow-making cost for 1 acre
- C3: the cost of increasing the longest run

*We only include the cost of the chairlift which is estimated to be $1,540,000

6.2. RECOMMENDATIONS

As a conclusion to our study, here are our recommendations

- Scenarios 2 and 3 seem to give the most profit, but we recommend scenario 2 as it has fewer costs than scenario 3.

- If the cost of increasing the vertical drop is below $1,934,637, we recommend scenario 2 otherwise we recommend scenario 1 with 1 run closed.
- scenario 1 could be considered with more closed runs if the cost of increasing the vertical drop is much higher than $1,934,637.
    - We can close another run after the first one, and test the result.
    - Then close 3 runs and test
    - Then close 3 other runs and test
    - The last two runs could be closed together, as there is a loss of $177,536 between these two last closings if the cost of maintaining the last run is greater than this amount. Otherwise, consider closing each run apart and test.

7. FUTURE SCOPE

Big Mountain is well placed in terms of the most important features for the trained model. If we assume that the prices of the other resorts are set according to the value of their installations, we expect Big Mountain's ticket price to be among the highest, but, given its current ticket price, it figures around the mean of all resort prices, thus the great difference between the predicted price and the current one. In future work, some missing data should be taken into account, especially cost data such as the cost of increasing the vertical drop by 100 feet, snow-making cost per acre, costs of increasing the longest run, and costs of operating a run.

An application with a user interface could be delivered to simulate different scenarios by tuning the values of the important features (a control field, such as text boxes and sliders, can be provided to modify each of these features) and then calculate the profit/losses of these adjustments after prediction.