

Big Mountain Resort Price Model Report

Introduction

Big Mountain ski Resort in Montana offers spectacular views of Glacier National Park and Flathead National Forest, and its facility provides access to 105 trails entertaining 350,000 skiers and riders every year. Recently, the Resort has installed an additional chair lift to help increase the distribution of visitors. This however, increases the operating costs by \$1.5M this season. The resort needs to find out how to offset this cost, maximize capitalization in its market segment and project the revenue for the upcoming season.

Problem statement

How can Big Mountain Resort reduce its operational cost of \$1.5M caused by a new installment of additional chair lift this season through implementing new pricing strategy.

Data Wrangling

The following key process steps have been undertaken during data wrangling process for Big Mountain resort data:

1. Originally, the dataset had 330 number of rows with 27 columns.
2. The rows 'AdultWeekend' and 'AdultWeekday' have been identified as the target rows and the remaining records as potential features.
3. AdultWeekday had significant amount of missing values than Weekday prices. So the dataset has been updated to keep only rows that contain weekend price.
4. The data for Silverton Mountain, has been replaced with the accurate values obtained from the web.
5. The 'yearsOpen' record was excluded from the dataset because the record was found to be inaccurate.
7. The columns 'fastEight' and 'AdultWeekend' were dropped because they had the most missing values.
8. It has been verified that the dataset contains only unique record in each row.
9. Identified the following features as relevant because their values can be summed up helping to identify the state-wide market size for each state:
 1. TerrainParks
 2. SkiableTerrain_ac
 3. daysOpenLastYear
 4. NightSkiing_ac
10. Population and area data for the US states have been extracted from Wikipedia and joined as two columns ('state_population' and 'state_area_sq_miles') with the ski-data frame

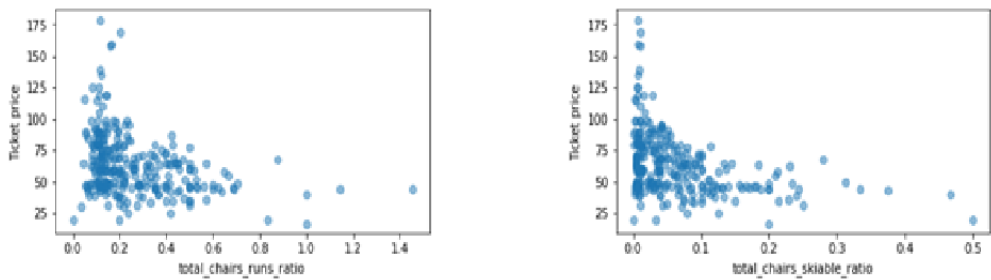
Exploratory data Analysis

The dataset Ski_data has 277 rows and 36 features. The features Region and states are categorical features. It has been concluded to treat all states equally when working on building a pricing model in subsequent modeling. The pricing model shall consider all states because there is no association between state and ticket price. In addition, there were high correlations identified between some features and ticket price. The following features demonstrated strong positive correlation; thus, they will be used in subsequent modeling:

• Vertical_drop • FastQuads • Runs • Total_chairs, • Days open last year • Projected days open

The scatter plot below shows a negative correlation between total number of chairs and pricing. The more chairs a resort offers, the cost of ticket price goes down. This leads to the assumption of exclusive vs. mass market resort effect. However, since the data does not

include number of visitors per year, this assumption cannot be further investigated nor confirmed.



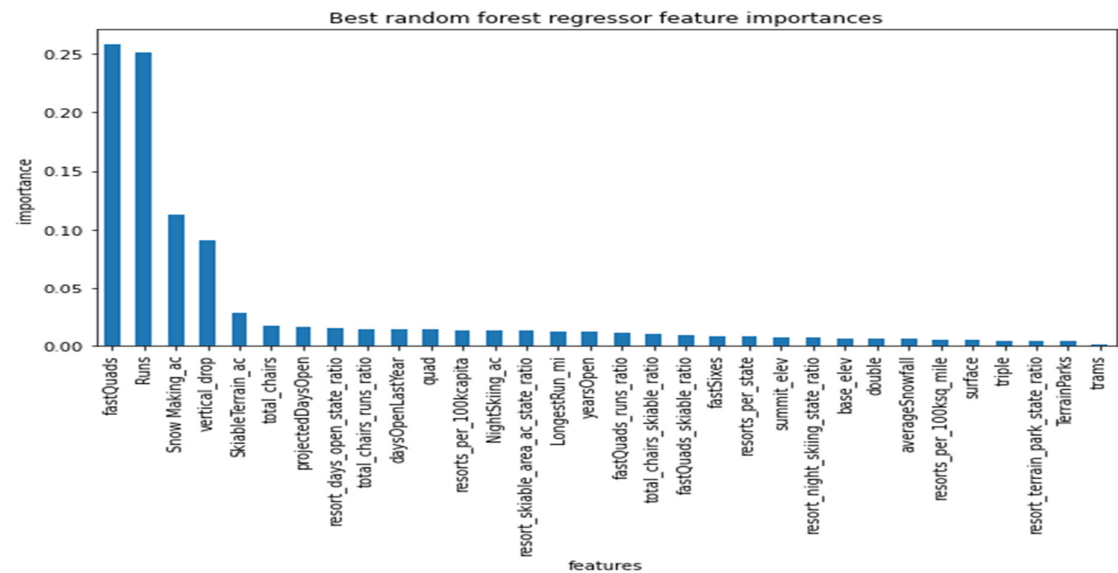
Model Preprocessing with feature engineering

The average price was calculated and determined to be \$63 which was used as the baseline for comparison.

To further investigate, a simple linear regression model was developed. We imputed missing values using the median and mean values in each column and generated predicated values and assessed the model's performance. After running a MAE test, it was clear that this model demonstrates a better prediction; the ticket price is off only by about \$9.

On the other hand, this model was overfitting. To adjust, the cross-validation technique was applied, and the value of K was reduced to only 8 features. It has been determined that the vertical drop is one of the biggest positive features along with 7 features. These results are consistent with the outcomes during our previous EDA work.

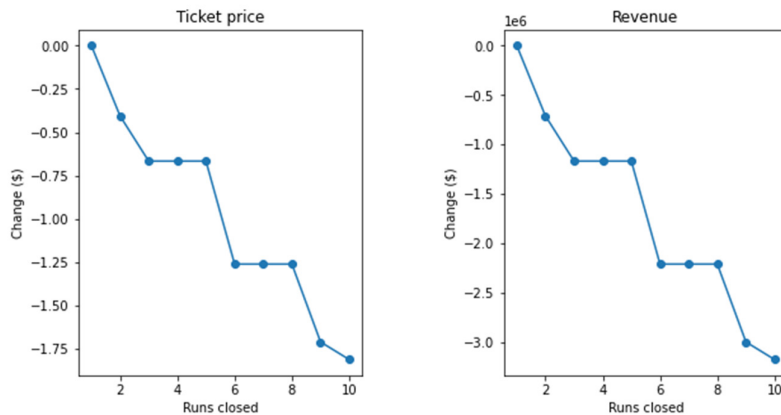
Next, a random forest model was developed which demonstrated that the top four features are fastQuads ,Runs, Snow Making_ac and vertical_drop consistent with the outcome of our linear model. The random forest model had a lower mean absolute error by almost \$1 and had low variability.



Conclusion and price recommendation

It was concluded that the random forest model is a better fit for Big Mountain Resort's price model. It demonstrated consistent results with cross-validation outcomes and the mean absolute error was the lowest in comparison to the linear model.

Currently, Big mountain charges 81 dollars and according to the model built, it is suggested that Big Mountain should charge 95.87. According to the model, big mountain can reduce the number of runs without affecting its position in the marketplace. They can close 3-6 runs and reduce their operating costs while still maintaining their market position locally and nationwide.



It is also advisable for Big Mountain to increase their price per ticket considering the number of chair lifts. Installing additional chair lift increases the value of the resort which will justify the price increase by about 8 dollars assuming visitors will buy tickets for an average of 5 days. This will amount to an average revenue of \$1.5 m which can offset the operational cost.

As future scope of work, to improve the model it would be beneficial to source more data pertaining to number of visitors, operating costs per feature and overall operating cost. Lastly it would help to have data from visitors stating the price they would be willing to pay for premium features.