Before COVID-19

Springboard

Data Science Career Track

During COVID-19

Guided Capstone Project Report: Big Mountain Resort Business Problem

Flor A. Espinoza Hidalgo

Ph.D. in Applied Mathematics

Research and Teaching Consultant

Email: fespinoza.science@gmail.com

Albuquerque, New Mexico, USA

October 2020

Big Mountain Resort Problem Identification

Big Mountain Resort suspects it may not be maximizing its returns relative to its market position. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more.

This project aims to build a predictive model for ticket prices based on resorts feature facilities. Then, the model will be used to guide Big Mountain's pricing and future facility investment plans.

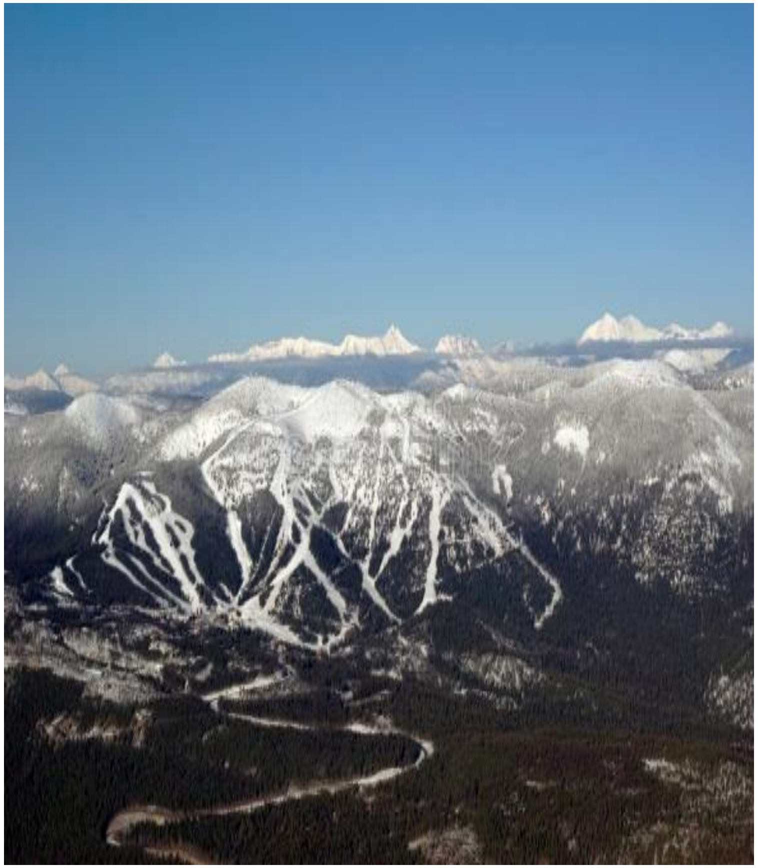
Big Mountain Resort Information

Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest.

The resort originally opened in 1947 with an annual snowfall of 333 inches and 3,000 acres of skier and rider accessible terrain.

Big Mountain Resort offers access to 105 named trails and vast bowl and tree skiing. All these are serviced by 11 lifts, 2 T-bars and 1 magic carpet for novice skiers.

The longest run is named Hellfire and is 3.3 miles in length. The base elevation is 4,464 feet, and the summit is 6,817 feet with a vertical drop of 2,353 feet.



Ref: Photo taken from the internet

Data Analysis

The ski resorts data was obtained from Springboard.

We assessed data quality by examining the raw data observations and features distributions to identify any issues such as outliers and missing feature values.

We dropped records with missing price data. We examined any patterns between the states, but since we did not observe any apparent relationship between state and ticket price, we decided to treat all states the same when building the price modeling.

We determined that predicting the adult weekend ticket price was our primary aim in our model. The target feature to predict ticket price is the resorts features that matter most to visitors, and they are most likely to pay more for them

Data Modeling

The data was split into training and testing sets. We employed machine learning to build three ski price models.

- The baseline model using the training set's average ticket price to predict ticket prices.
- The linear regression model addressing the missing values imputing them with the median and mean values.
- Random forest model exploring several parameters.

Model's performance was evaluated on the test set by computing the R squared (R2), mean absolute error (MAE), and mean squared error values.

Data Modeling Results

Baseline Model (average ticket price): The mean absolute error (MAE) result showed that, on average, we might expect to be off by $19.00 if we guessed ticket prices based on the average of the known values.

Linear Regression Model: The MAE from the two imputing cases for missing values were very similar. These results showed that the linear model could predict prices within $9.00 of the accurate average prices.

The preliminary results of the linear model outperformed the baseline model without removing any features.

Data Modeling Results

Linear Regression Model: We also performed 5-fold cross-validation on features space exploration to identify the predictors that contribute the most to the ticket prices.

The results showed that out of 32 features, 8 are the most significant contributors (**vertical_drop, snow making_ac, total chairs, fastQuads, Runs, LongestRun_mi, trams, and SkiableTerrain_ac**) to make price predictions in the linear model. We also observed that the performance results' variance increases for the number of features higher than 8 (Fig. 1).

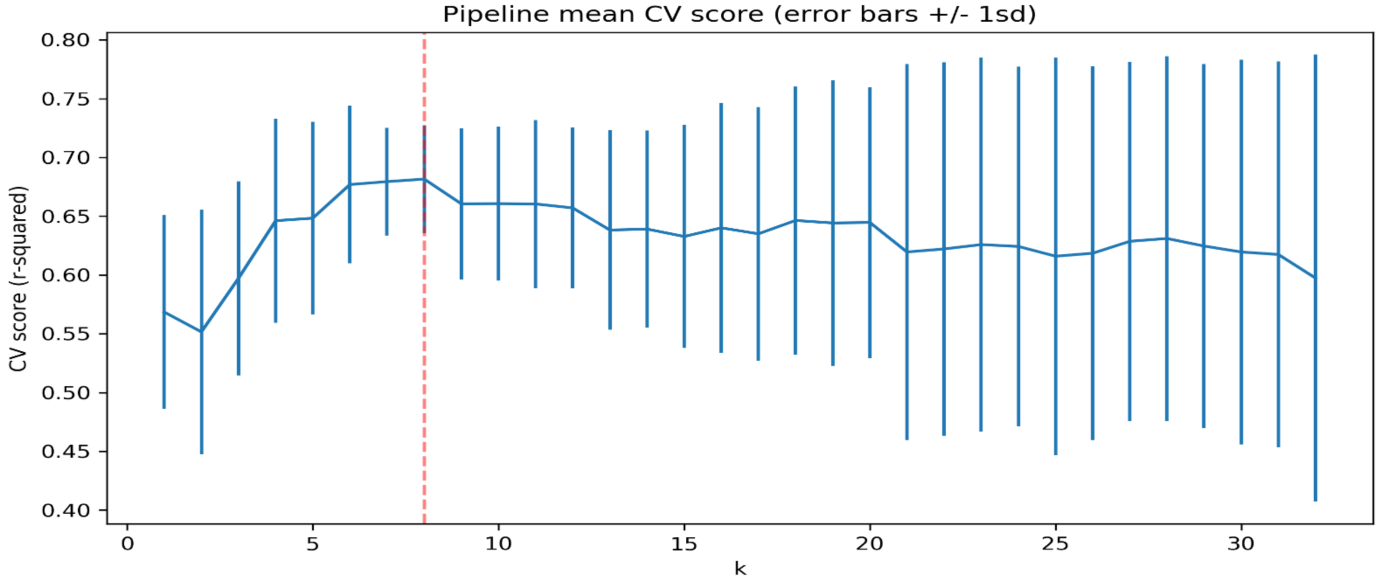


Figure 1: Linear model features space mean cross-validation (CV) score results shown in blue and the optimum number of feature k=8 shown in a vertical red dashed line. The k = 1-32 features are shown on the x-axis, and the r-square values are shown on the y-axis.

Data Modeling Results

Random Forest Model: We focused on with and without scaling, imputing missing values using the mean and median values, and employing 5-fold cross-validation to evaluate this performance on the training set.

The results showed that a random forest model without scaling and using the median for imputing missing values gives the best performance. We also obtained the contribution scores of each of the 32 features to the model.

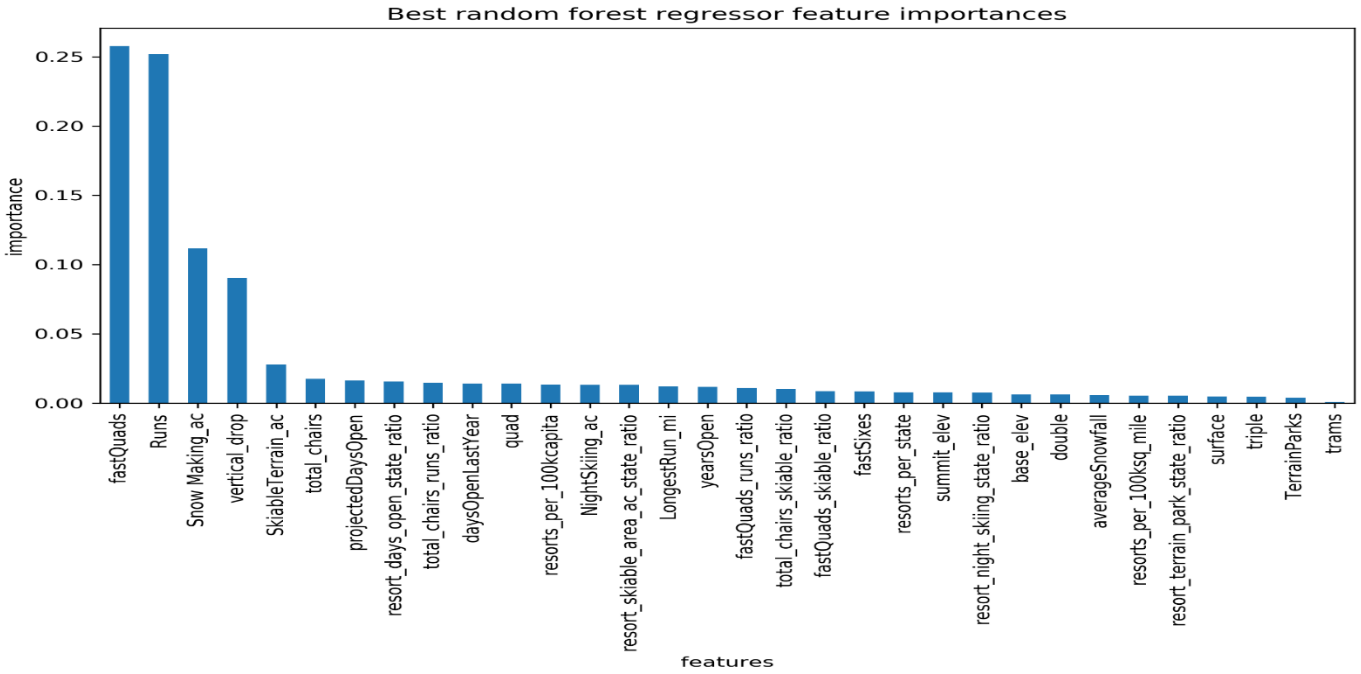


Figure 2: Random forest model feature importance distribution.

The top contributors to the selected random forest model (Fig. 2) are: **fastQuads, Runs, Snow Making_ac,** and **vetical_drop,** which are similar top contributors in the linear model.

Data Modeling Results

When comparing the linear model's cross-validation MAE values to the random forest model, we found that the random forest model outperforms the linear model by almost $1.00. It also exhibits less variability than the linear model.

Therefore, we applied the random forest model to predict Big Mountain resort prices. The random forest model prediction price for Big Mountain Resort is $94.22. With an expected mean absolute error of $10.39, this result gives Big Mountain Resort room to increase its current price of $81.00.

Next, we explored four potential scenarios for Big Mountain resort for either cutting costs or increasing revenue from ticket prices.

1) Permanently closing down up to 10 of the least used runs.
2) Increase the vertical drop by adding a run to a point 150 feet lower down but requiring installing an additional chair lift to bring skiers back up, without additional snowmaking coverage.
3) Same as number 2, but adding 2 acres of snowmaking cover.
4) Increase the longest run by 0.2 miles to boast 3.5 miles length, requiring additional snowmaking coverage of 4 acres.

Data Modeling Results

Considering an expected number of visitors over the season of 350,000 who ski an overage of five days and assuming that the provided data includes the additional lift chair that Big Mountain Resort recently installed, the results from scenarios 1-4 are:

Scenario 1: Closing one run makes no difference in ticket price from the predicted price. However, Closing 2 and 3 successively reduces support for ticket price and 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.

Scenario 2: By adding a run to a point 150 feet and adding an additional chair lift without additional snow making coverage, this results showed support for an increase of ticket price by $1.99. Over the season, this ticket adjustment can increase the revenue amount to $3,474,638.

Data Modeling Results

Scenario 3: By repeating scenario 2 and adding 2 acres of snow making, this scenario increases support for ticket price by $1.99. As we can noticed, such a small increase in the snow making area makes no difference from the results of the previous scenario.

Scenario 4: By increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability, the results showed that this scenario does not provide support for ticket price increase. Therefore, this scenario shows no difference whatsoever.

Based on the results from scenarios 1-4, our suggestion to the business leadership is to go with scenario 2 to set up a new ticket price.

As we recall, by adding an additional chair lift, Big Mountain's increases their operating costs by $1,540,000. Therefore, by increasing the price by $1.99 provides an increase of $3,474,638 in revenue, which covers the operational cost of the installed chair.

Summary and Conclusions

We have shown that Big Mountain Resort should not base their ticket pricing on just the market average. This approach does not provide the business with a good sense of how valued some feature facilities are than others.

The linear model has shown eight features as the most significant contributors to ticket prices. In contrast, the random forest price model retaining all features provides a smaller MAE than the baseline and linear models.

When comparing the linear model's cross-validation MAE values to the random forest model, we found that the random forest model outperforms the linear model by almost $1.00. It also exhibits less variability than the linear model.

Therefore, we selected the random forest as the best ticket price model. We delivered this robust model to the Big Mountain Resort business decision team to make future price explorations and adjustments.

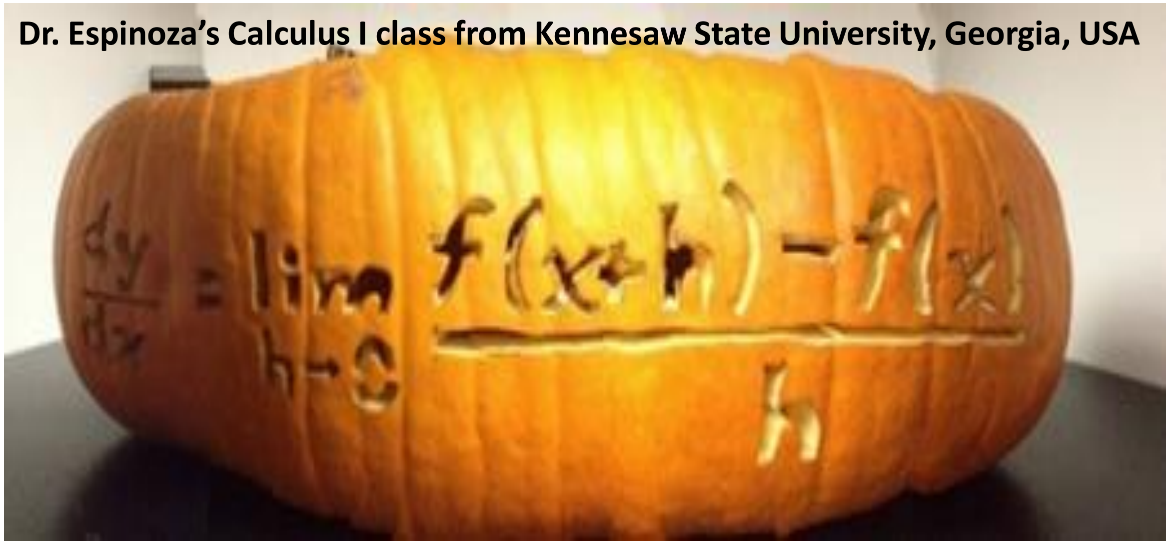
End of the presentation

Thanks for your attention

Please feel free to ask questions, share your comments, and provide feedback.

International Balloon Fiesta, Albuquerque, New Mexico, USA

Learning feeds my brain,
teaching feeds my soul,
research feeds my curiosity,
and love feeds my heart.
Flor A. Espinoza.



Dr. Espinoza’s Calculus I class from Kennesaw State University, Georgia, USA