Big Mountain Resort Ticket Pricing Recommendations from Market Analysis

Big Mountain Resort (BMR) is a premier ski resort located in the Montana Rockies, accommodating skiers and riders of all levels and abilities. Located near many trails in large parks, every year about 350,000 people ski or snowboard at BMR. After installing a new chair-lift, operating costs have soared by $1.54 million for this season.

As a result of this, BMR determined to make changes that will either cut costs without undermining the ticket price, or support higher ticket prices.

Taking into consideration possible loss of customers and/or customer backlash resulting from a price hike, we set out to use data from market competitors to figure out a ticket price increase acceptable to customers, such that a ticket sales decrease is more than offset by the price increase. We endeavored to find areas of the business which can be made cost efficient without decreasing the quality of the property.

With the interests of you, our key stakeholders, at the forefront, we took market metadata from your database manager and performed our analysis.

Our initial step was performing data wrangling.

We observed that the original data included 330 rows and 27 columns of information. After identifying a duplicate resort name in the dataset, we verified that the state was different for the 2 resorts. We also investigated the "State" vs "Region" difference, finding that some regions encompassed multiple states.

After visualizing the ticket pricing for adult weekday and adult weekend, we saw that and disparities were in the direction of higher weekend pricing. We eliminated the 14.33% of rows with no pricing data, and also discovered that there were 7 rows with missing weekday values, compared to 4 rows with missing weekend values in the dataset, so we excluded the 'AdultWeekday' column, as well as eliminating those 4 rows which contained no weekend price. The only pricing that cared to use is the Adult Weekend pricing. Due to the number of missing values and 0 values, we eliminated the 'fastEight' column.

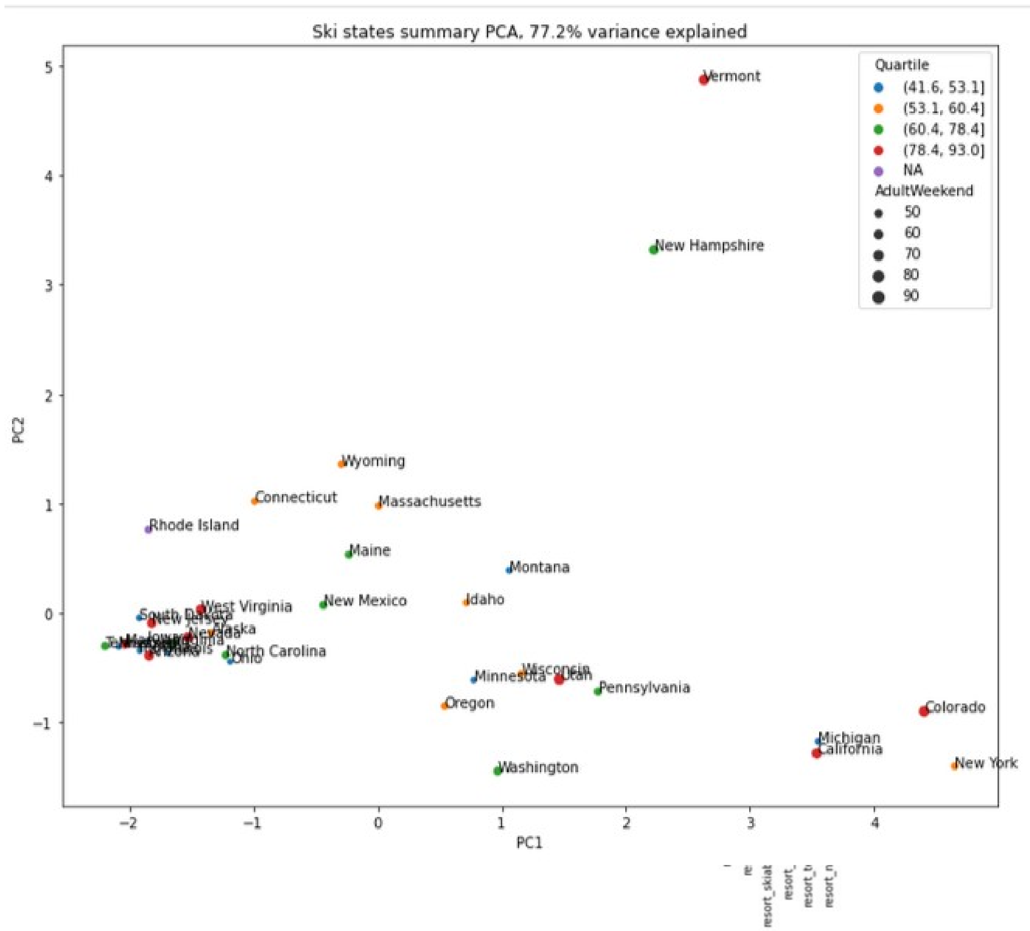
To assess extreme outliers, we plotted histograms for the values of each category. One resort had an incorrect value for Skiable terrain, which we corrected after a quick search. One resort was eliminated due to an incorrect "years open". The focus features of each resort were decided as: 'TerrainParks', 'SkiableTerrain_ac', 'dayOpenLastYear', and 'NightSkiing_ac'. Using wikipedia as a source, we merged population and area to each state in our dataset. This collected and cleaned data was the basis for our analysis.

One segment of our data consists of state names and the features of interest which may have the most effect on ticket pricing; namely: ‘resorts_per_state’, ‘state_total_skiable_area_ac’, ‘state_total_days_open_last_year’, ‘state_total_terrain_parks’, ‘state_total_night_skiing_ac’, ‘resorts_per_100kcapita’, and ‘resorts_per_100ksq_mile’.

After potting the cumulative variance ratio explained by these 7 PCA components, we determined that the largest 2 components explained greater than 75% of the variance, and decided to simplify our task by reducing the number of components down to 2.

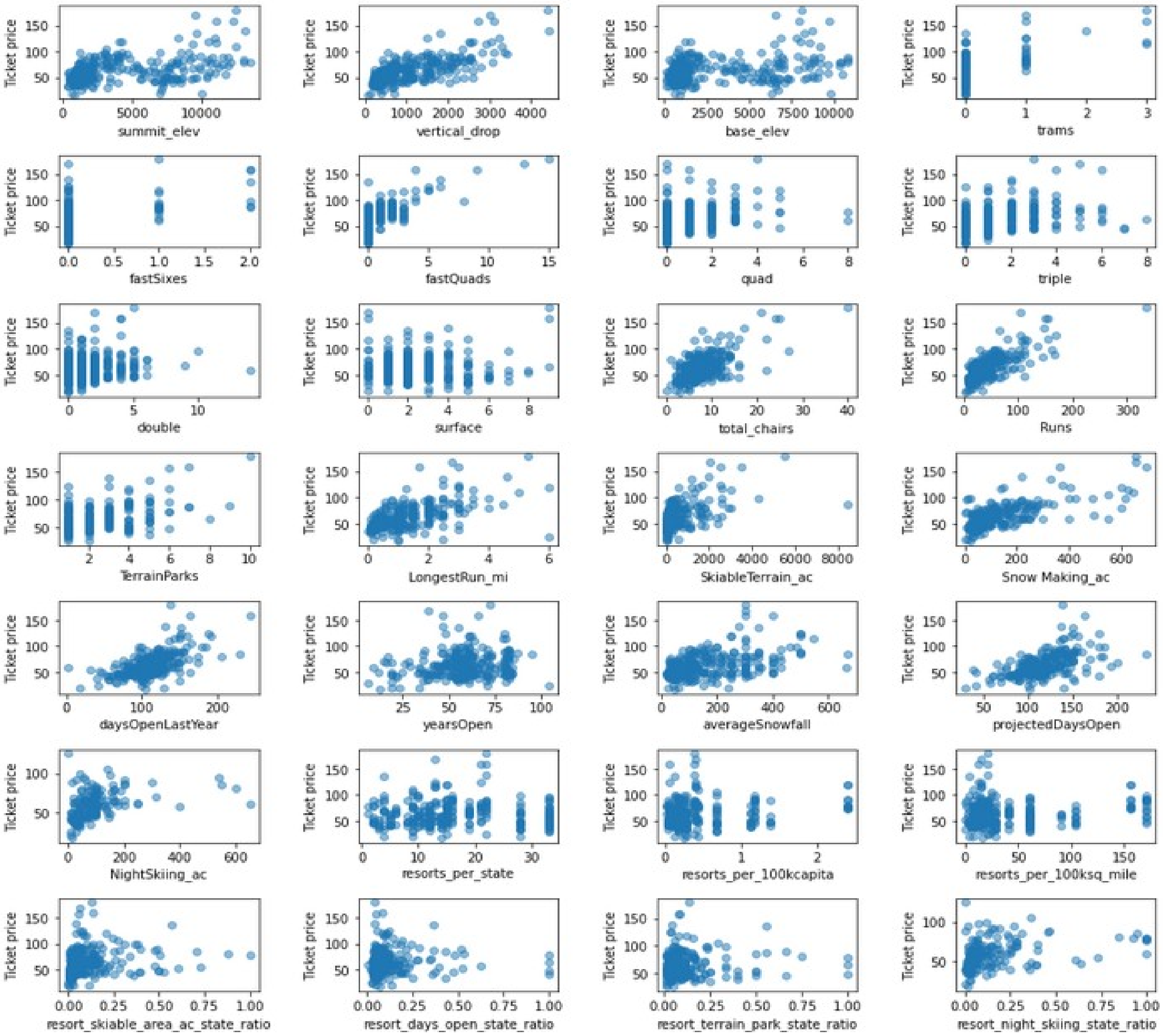
There was no clear state grouping when charting the 2 PCA components and ticket price.

We then used a heat map to visualize the correlation between each of the resort variables in our dataset. The results indicate that AdultWeekend ticket price is highly correlated with: vertical drop fastQuads total chairs runs Snow making acres Subsequent modeling can focus on these particular features.



We were mindful of exclusive vs. mass market resort effect on our numbers when choosing the features which we based our modeling on.

Visualizing scatter plots of various features vs. ticket price elucidated a few obvious positive correlations: vertical drop, fast quads, runs, and total chairs.



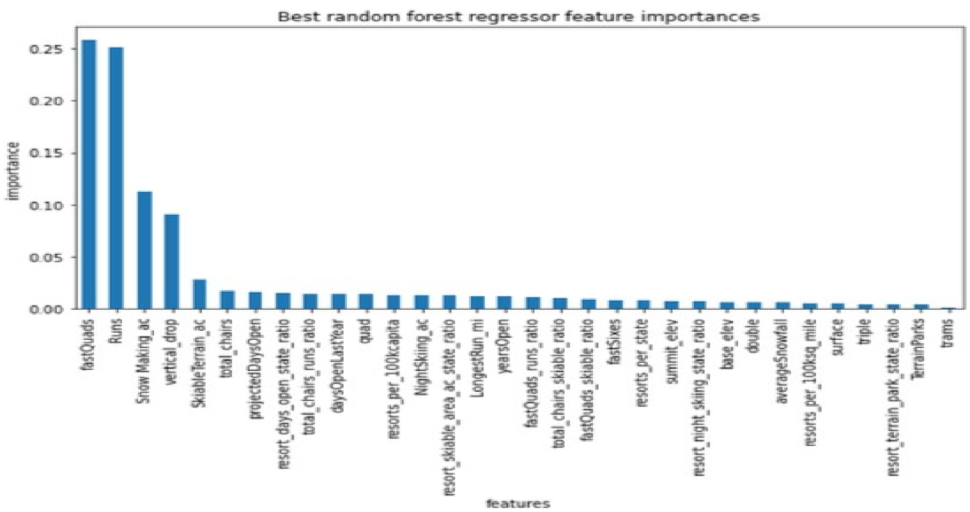
This exploratory data analysis narrowed our data, which allowed us to form our models on the most predictive aspects of our data.

We completed the step of pre-processing our cleaned data, and training different models on a test set of the data for use in modeling. We partitioned the data into a training set and a test set. Initially, we used the mean price of the training set to get a baseline understanding of performance, finding that the mean absolute error was ~$19.13 using the mean as the training/testing model.

Next, we built 2 parallel linear regression models. We imputed, using the median in the first and the mean in the second, as a fill-in for any missing values. Before using the data, which vary in size, spread, and magnitude, we used ‘StandardScaler()’ to scale each of the values to a mean/variance of 0. Assessing the 'median imputing' linear regression model performance, we saw that the r-squared of training and testing sets explains 81% of the variance on the train set and 72% on the test set. The mean absolute error of the test set was ~$9.41, which means that we expect to estimate a ticket price within $9.41 of the real ticket price. This was $10 better than the mean price model.

Assessing the 'mean imputing' linear regression model performance, we saw that the r-squared of training and testing sets explains 81% of the variance on the train set and 71% on the test set. The mean absolute error of the test set was ~$9.42, which means that we expect to estimate a ticket price within $9.42 of the real ticket price. This was $10 better than the mean price model, but not any different than the other linear regression model. Using pipelines, we tried to adjust the linear regression models for over-fitting issues, using sklearn's 'SelectKBest' feature use the best data feature subsets to build the model from. We used a for loop to create a parameter grid of all the possible variables in our dataset. Using cross-validation feature 'GridSearchCV', we looped through all of the possible number of parameters that could be used in the model, and found that 8 was the optimal number of parameters. The features most useful to the model (highest coefficients) were discovered to be 'vertical_drop', 'Snow_Making_ac','total_chairs','fastQuads', and 'runs'.

Next, we built a random forest model. Using a five-fold cross-validation, we trained and tested the model, and determined the mean cross-validation score to be ~0.698. GridSearchCV was used again to assess best parameters, and after printing the best parameter attribute, we saw that the best n_estimators number was 69. Using the best estimator in the model marginally improved our cross-validation results: mean of ~0.709. Plotting the top features for this model was consistent with the linear regression model. The 'fastQuads','runs','Snow_making_ac', and 'vertical_drop' features held the most importance.



We evaluated the models and found that linear regression tested with a mean absolute error of 11.79, while the random forest tested with a mean absolute error of 9.54. We moved forward using the more accurate random forest model, since its cross-validation MAE was over $1 more accurate, and it also had a smaller variability. We feel strongly about the model given that a training set of only 60 would still give a strong cross-validation score.

Our final findings:

Big Mountain currently charges $81 for Adult Weekend tickets. Out random forest model results suggests that, based on certain features which the market values, Big Mountain could justifiably increase their ticket prices to $95.87, a delta of $14.87. Hypothetically changing certain aspects of Big Mountain resort provide multiple options for the executives to consider. Closing 5 of the least used runs would save Big Mountain operating costs and only decrease expected ticket pricing by $0.66, equalling ~$1.2 million. This would be more than offset by 2 options:

1) Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift.
This scenario increases support for ticket price by $8.61. Over the season, this could be expected to amount to $15.0 million in increased revenue.

2) Option 1, plus adding 2 acres of snow making.
This scenario increases support for ticket price by $9.90. Over the season, this could be expected to amount to $17.3 million in increased revenue. This may not be worth the operational costs compared to option 1, due to the insignificant ticket price increase.

A third explored option had no modeled effect on ticket pricing:

3) Increasing the longest run by 0.2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability.
Even with the additional costs of the planned chair lift, for $1.54 million, the per ticket lift cost at the current price of $81 is only ~$0.011.

Regarding future improvements, **Option 1 is the most straightforward path to safeguarding your bottom line.**

Regarding run closures, Big Mountain could collect data on run use during the season to see which runs are most frequented. This would help us determine the best runs to close. Closing the least used runs, then analyzing the ticket sales afterward, could give insights into whether the closed runs were, in fact, the least popular.

Other information which would be helpful to our analysis would be operating costs associated with each feature of the resort, such as snow making, run building, vertical height increasing, etc. Our model suggests that Big Mountain has been an undervalued asset (based on its desirable features), even before recent and suggested modifications. If you business leaders feel that this model is useful, you could brainstorm property modifications to BMR and use this model to assess the fair ticket pricing change.

Thank you.