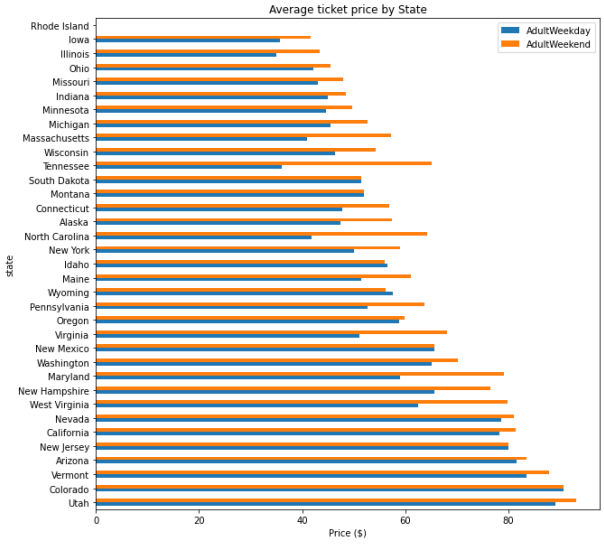
Big Mountain Ski Resort Data Analysis for Business Optimization

Problem:

What is the best business strategy for determining fair value ski resort ticket prices and offsetting the $1,540,000 upgrade so that Big Mountain Resort can increase their return on revenue?

Data Wrangling and Exploratory Data Analysis:

Raw data collected and loaded as csv file. Exploring the categorical features, we decided there was very little difference between region and state and so went forth with finding the distribution of adult weekday and adult weekend ticket price by state. The result is as shown:

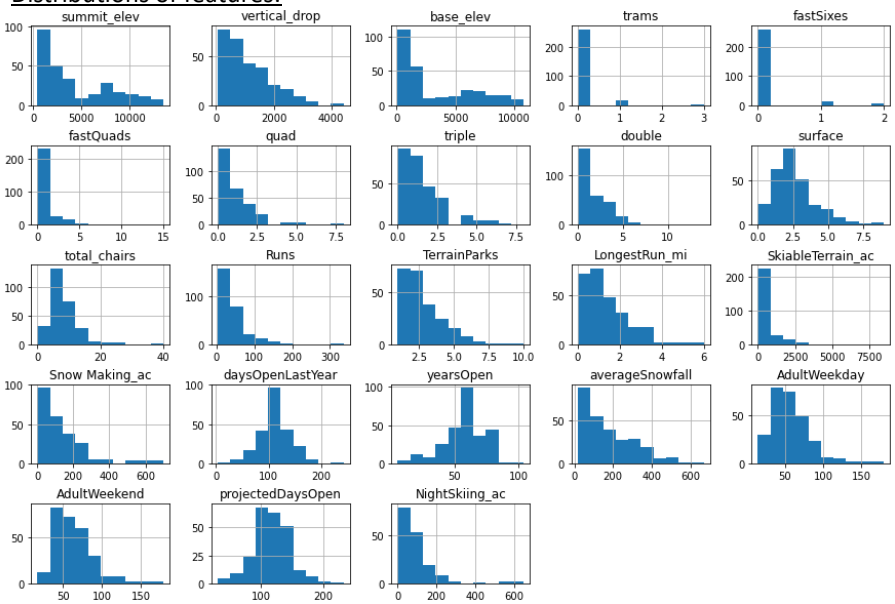


Exploring the numerical data, we used transpose method to reveal the stat columns, found out that 3% are missing one value of ticket price and 14% are missing both values (but we did not drop these missing values just yet), and then went ahead to finding the distribution of feature values, keeping in mind possible outliers that will result. We found that one resort claims to have been open for 2019 years, which is certainly incorrect and by logic we assume it opened in 2019 instead. We then used these features to derive state-wide summary statistics for our market segment via named aggregation and histograms, ensuring to omit the 14% of rows with no price data this time around:

<u>Features of interest:</u>

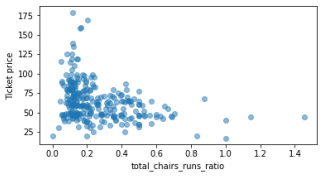
- TerrainParks
- SkiableTerrain_ac
- daysOpenLastYear
- NightSkiing_ac

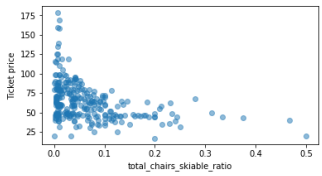
<u>Distributions of features:</u>

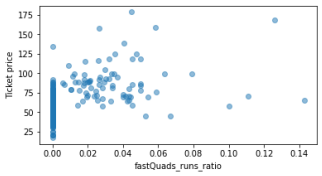


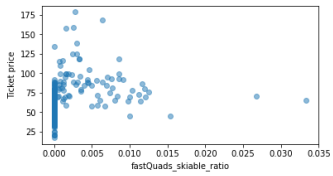
In addition, we found that adult weekend prices were higher than weekday prices sub $100 resorts and weekend prices had less missing values, so adult weekend prices was our value of interest. This was our attempt to predict adult weekend ticket price for ski resorts. After some more exploration of data, we used principle components analysis (PCA) to find linear combinations of the initial features that were uncorrelated with each other and organize them by variance. We found Montana to fall within the range of the majority of states. A feature correlation heatmap revealed fastQuads, Runs, and Snow Making_ac would increase ticket price.

A scatterplot of numeric features against ticket price, showed high correlation between vertical_drop, fastQuads, Runs, and total_chairs. Scatterplots for ratio of chairs to runs and skiable terrain, as shown below, predicted a decrease in ticket price as more chairs are available:





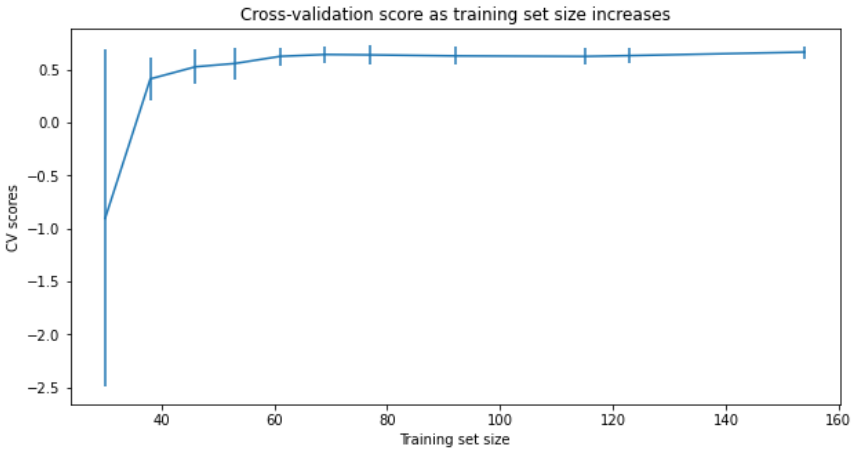


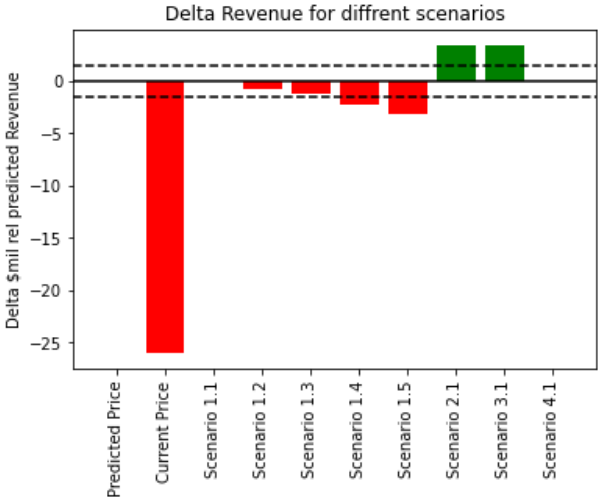


This may infer an exclusive vs. mass market effect, however the missing data of number of visitors per year weakens this inference.

Modeling:

Random forest model was chosen as the product model because it exhibited the least variability and had lower cv mean absolute error (MAE=1.0). According to the model, scores rapidly improve but eventually level off at sample size 40-50, indicating data available is sufficient. However, operation costs and number of visitors/year are highly useful data that should be collected for more accurate results. The final random forest model used the median to impute missing values but no feature scaling was necessary. The mean absolute cross-validation error was $9.54 and standard deviation $1.35.





Recommendations:

The current ticket price for Big Mountain Ski Resort is $81.00. When we excluded Big Mountain Resort data when training the model to predict ticket price (based data on all other resorts), we obtained a mean absolute error of 10.39 and a standard deviation of 1.47. This training data set had less bias than the previous, and so this suggests an optimal ticket price of $95.87. Business management team shortlisted options, and when running each scenario delta related to predicted price we can suggest that scenario 2.1 (adding a run, increasing vertical drop by 150 feet, and installing an additional chairlift) is a good business plan to take up as it provides higher ticket price value and increased revenue. It seems that having higher ticket prices could circumvent the $1.54M operating cost for a season. However, it is not a solid business strategy until this model is run again with the operating costs that management should provide. Other useful knowledge to request for would be a valid number of average visitors for each season for each resort. On our end, we should find a way to look for valid ticket price value for all resorts, as this was another limitation.