Big Mountain Resort - Ticket Pricing Model

A Montana-based client ski resort, Big Mountain Resort, seeks to improve their market-average-based ticket pricing by evaluating their facilities. Accordingly, they desire facility-based metrics to consider investments or divestments in their amenities. Facility metrics and ticket prices were considered from over 200 resorts in North America to create a ticket-pricing model. The model indicated that the client should increase their ticket price by fifteen dollars to match their provided facilities, which would add ~$26 million in revenue over the year. Additionally, the model was used to simulate a few scenarios proposed by the client. I recommended one of the investment options —add a run and chairlift, increase vertical drop, increase snowmaking area—, as the combination of factors further increased the ticket price suggested by the model. If the increased ticket prices do not affect the amount of expected visitors or the duration of their visits, then the increased revenue resulting from the investments will more than cover their costs.

The client provided a spreadsheet containing data for over 300 resorts to support the modeling effort, including their state/region and various numerical metrics. The resort attributes were compared against ticket-price ($), data for resort visitors was not provided. Weekday and weekend prices were provided in the dataset, but only the weekend prices were used as less values were missing. Weekend prices were typically the same or higher for most resorts. Resorts without any pricing data were not considered in the model. Additionally a resort feature describing “fast eight person chairs”, fastEight, was removed as data was mostly missing or equal to zero. Clearly false outlier values for other features were checked and fixed or removed.

Resorts were grouped by state and analyzed for trends with ticket price. No clear patterns emerged, and therefore localities were not used in the final pricing model. As an exercise, principal component analysis was performed on the state-summarized data and its first two derived features explained 77% of the dataset’s variance. The states and ticket price did not show any clear clustering within these dimensions.

Prior to modeling, the distributions and correlations of the various numeric features were explored. Attention was paid to features that had strong correlations with ticket price, AdultWeekend. A resort’s “importance” within its state was also briefly examined by calculating its share of important features within its state. These plots are not included in the report.

Facility-based metrics were used to predict ticket-price with a number of different models. A 70/30 split was used for training/testing, and cross-validation ensured that a number of different training sets were used. Missing values were imputed by simply using a feature’s median value. For a linear regression mode, no difference was found using mean or median as an imputation strategy. Data was scaled (remove mean, transform to unit variance) for most modeling efforts prior to training. Models were evaluated by their R^2, Mean Squared Error, and Mean Absolute Error.

The amount of features included in the linear regression models was optimized using a hyperparameter search. Selected best features were varied (scored via f_regression) and models were scored by cross-validation. The search suggested 8 dimensions would be best to include in a linear model (plot 3) and also provided the most impactful resort attributes .

A Random Forest model was also used to predict ticket price. Again, a hyperparameter search was used to refine the model, focusing on the number of estimators used. The exercise yielded a best performance model by median-imputation, no data scaling, and using 69 estimators. Again, the importance of resort features was explored (plot 4). This model was chosen over the optimized linear model due to slightly better performance. The cross-validation test scores showed its mean absolute error was about one dollar less, and had less variability than the linear model.

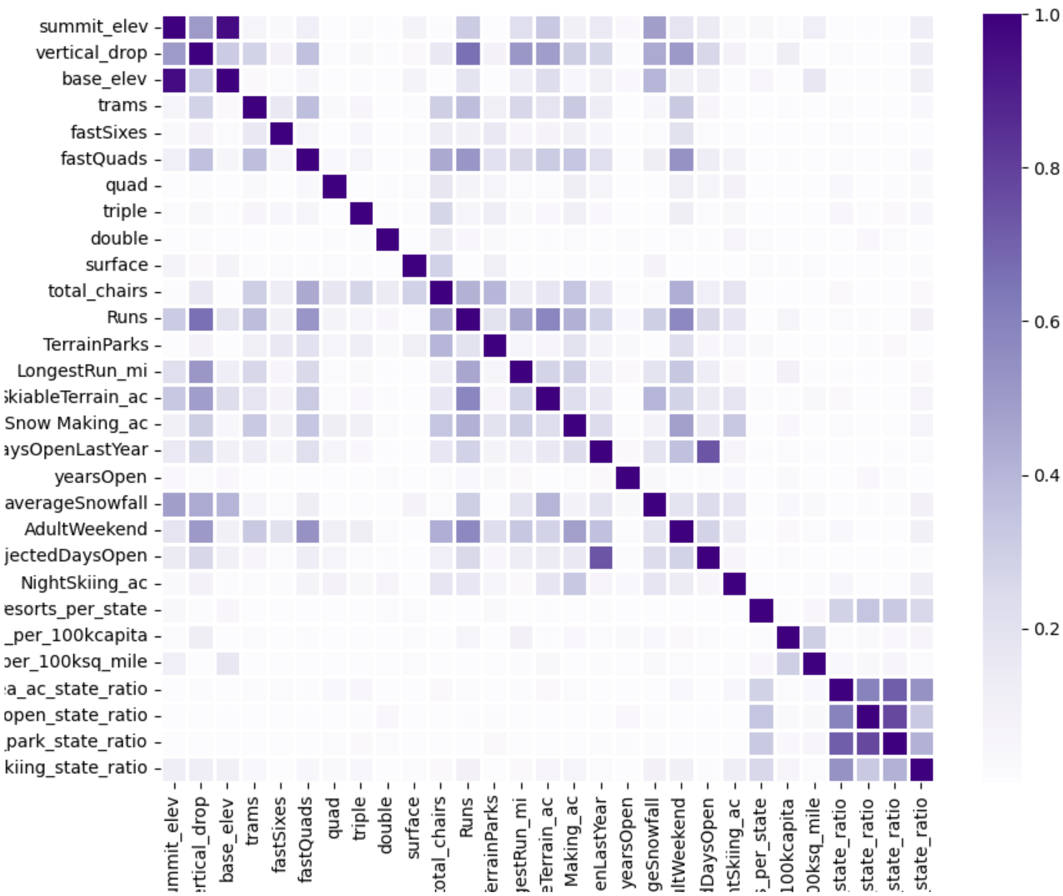
Finally the Random Forest model was used to predict a new ticket price for Big Mountain Resort, and to evaluate the implication of a few different scenarios. The model's ticket price was ~$96, which is $15 more than the current price of $81. Even when considering the possible error of the model (~$10), Big Mountain's ticket price has room for increase. Given the resort's projected 350k visitors and their average five day stay, pricing according to the model will add $26.25 million in revenue to Big Mountain. This more than covers their recent $1.54 million investment in a new chair lift. One possibility proposed by the client involved cutting costs by shutting down their least used runs. The model suggested minimal devaluation in ticket price, with plateaus from 3-5 and 6-8 runs closed. The runs could be closed without decreasing their current ticket price, therefore this option is viable.

I proposed the client follow their investment options to further increase their ticket price. An option to increase the resort's longest run and increase its overall snow coverage did not yield ticket-price value. But an option to add a run, chairlift, and increase vertical drop (with or without adding snow-making) would add value to the ticket. The model is sensitive to increasing the resort's vertical drop and adding to the number of total chairs together. If the client follows through with their second scenario, they could increase ticket price by $17, and therefore expect adding $29.75 in revenue over the year! I expect this to also outstrap the investment costs, given the information in the briefing.

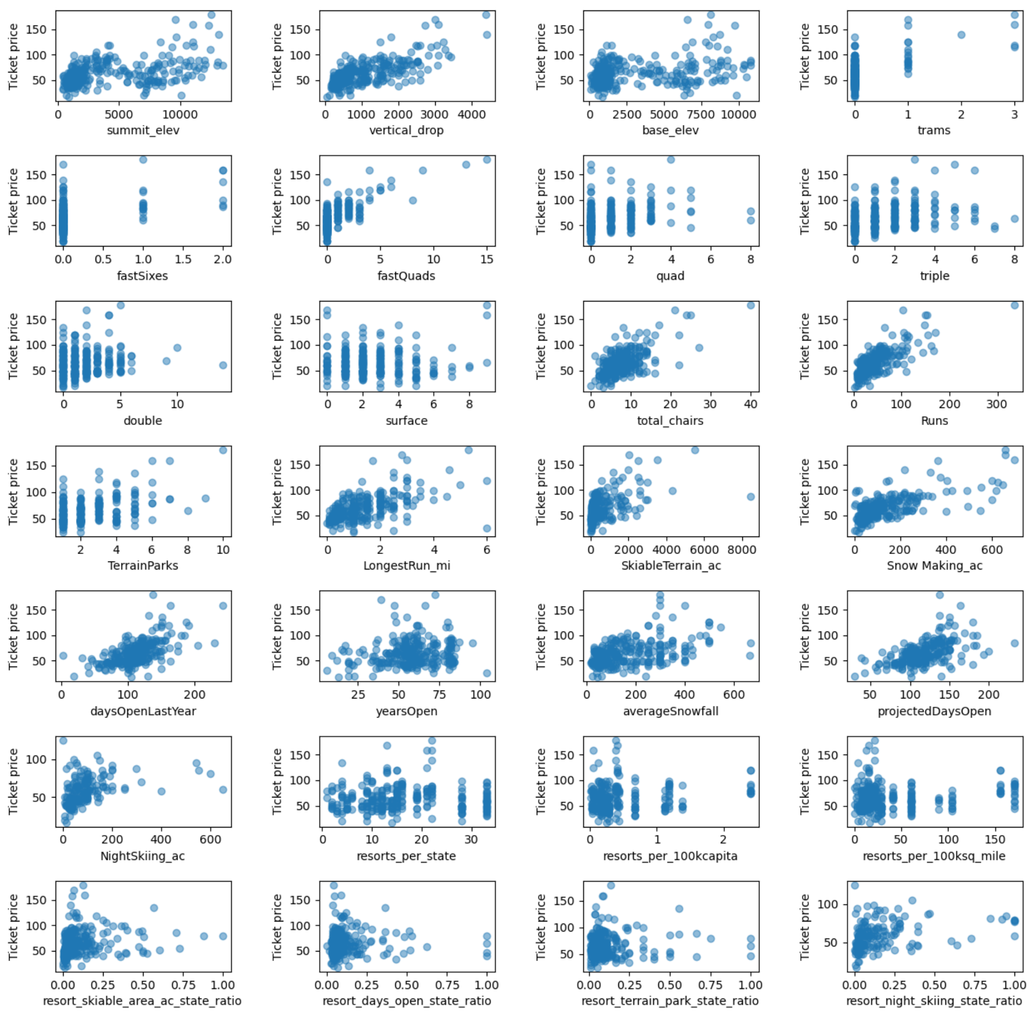
The model will be finalized into a form that the business executives at Big Mountain Resort can use, allowing them to plug and play various scenarios. If they would like to continue our work together and improve the model, then I would recommend them to add visitor information to their dataset. A brief data quantity assessment indicated that the amount of resorts in the dataset is sufficient. A separate table of possible investments and their cost would allow even more model improvement, in that direct scenarios could be recommended.

Rough plots are provided on the following pages. They have not been styled for final presentation. *Report Link*

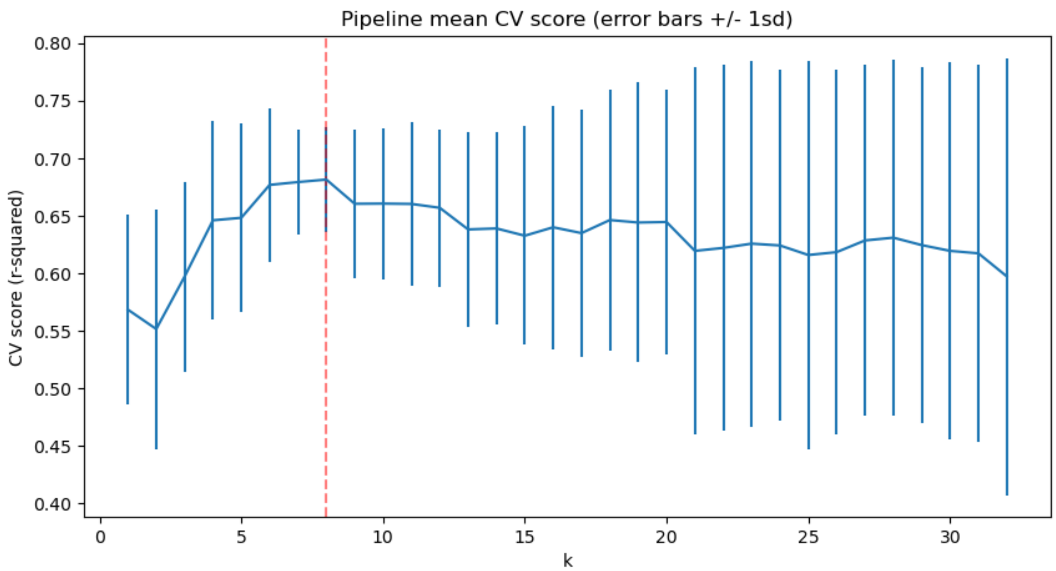
Plot 1 - Feature Correlation Heatmap



Plot 2 - Ticket price vs feature distributions



Plot 3 - Cross-validation scoring for Linear Regression, number of features to select



Plot 4 - Feature importance for Random Forest model