The business problem driving this analysis is to suggest a higher ticket price supported by the US ski market given Big Mountains existent features. The secondary goal is to evaluate potential modifications to runs and equipment which would support ticket price increases.

There are 330 rows and 27 columns in the original data. The focus resort "Big Mountain" is present in the data and is complete with no missing values. Notably absent from the data is information about ticket sales volume and length of guest visit in days.

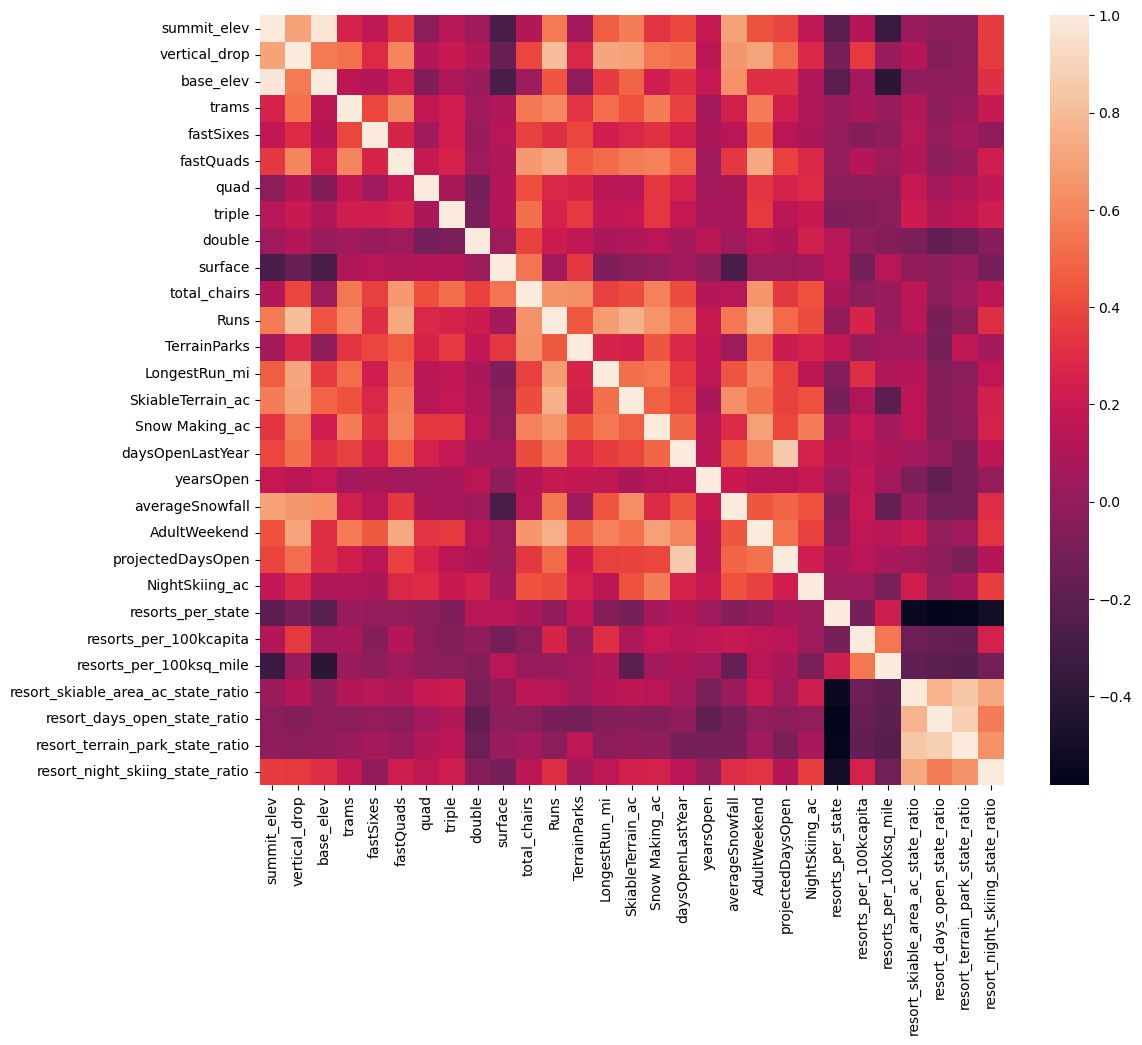
The column ‘AdultWeekend’ was used as a measure of ticket price. The column ‘AdultWeekday’ was dropped. In $100+ Montana resorts, there is no apparent discrepancy in weekday vs weekend pricing and missing values are more common in the weekend price column. The fastEight column, populated almost exclusively with null and zero values, was dropped. Two rows with ambiguous opening dates (‘yearsOpen’), were dropped. After cleaning actions were taken, 277 rows and 25 columns remain in the dataset.

Preliminary exploration was done, examining the distributions of skiable area, years open, and feature values.

The provided dataset was reviewed, wrangled and cleaned, along with preliminary exploration. The ‘AdultWeekday’ column is the target feature for predicting ticket price. The goal of predicting ticket prices is to maximize Big Mountain Resort’s returns by understanding its position in the market and identify which features guests are willing to pay more for. The state summary was created by adding population and state area columns to the ski resort area, creating two new columns 'resorts\_per\_100kcapita' and 'resorts\_per\_100ksq\_mile'.

Exploratory Data Analysis

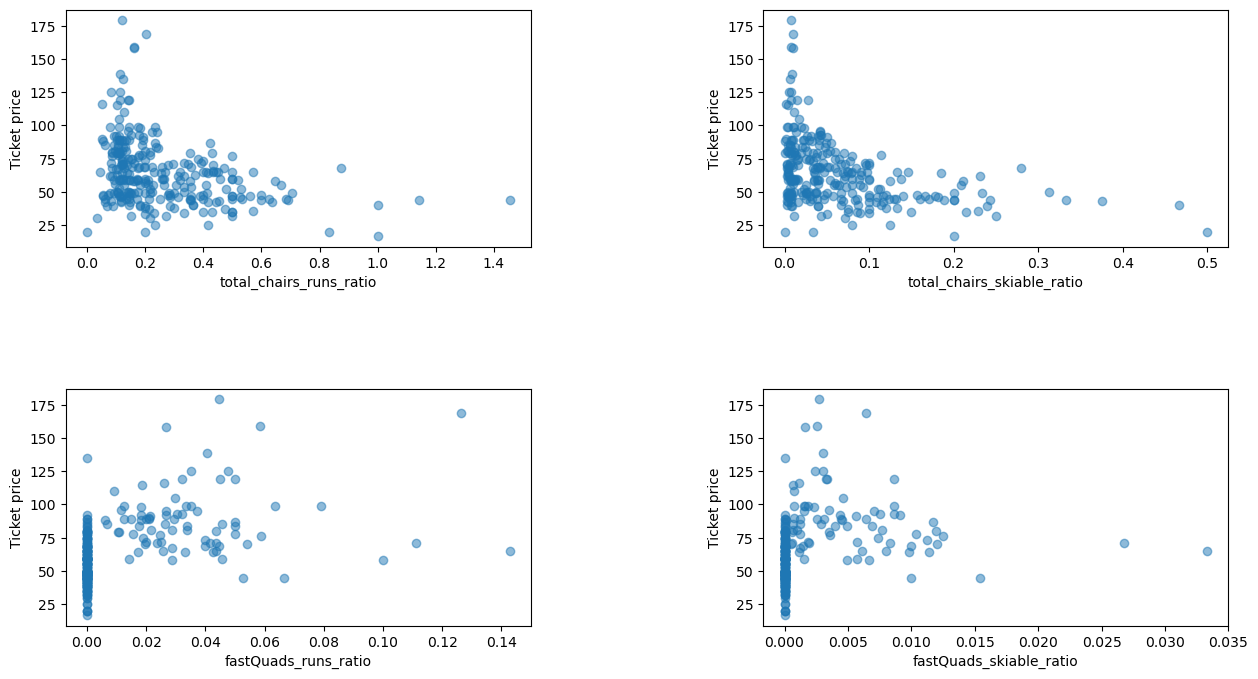
Heatmap examining correlation between features. Of particular note is the relationship between ticket price ‘Adult Weekend’ and other features.



A strong positive correlation was found between price and the following features: vertical\_drop, fastQuads, Runs, total\_chairs and SnowMaking\_ac. A more complex relationship was found between price and resorts\_per\_100kcapita.

Number of chair lifts may have a complex relationship with more lifts at resorts with lower prices serving larger numbers of visitors. Different lift types for example fastQuads may merit higher prices, but may be more or less critical depending on the amount of skiable terrain.

Scatterplots highlighting the relationship between ticket price and other features with possible correlation.

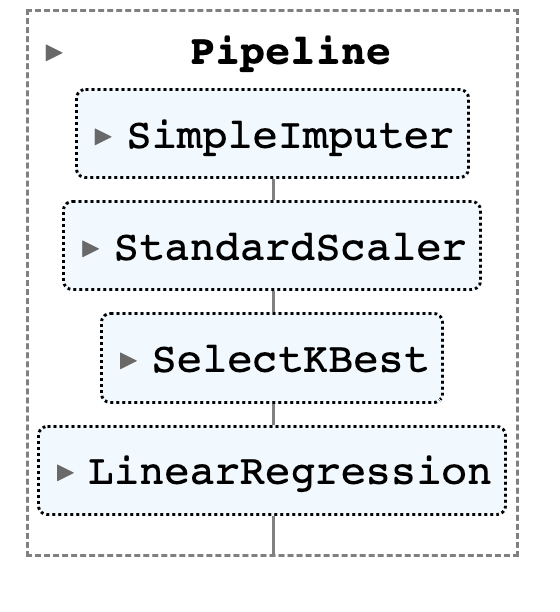


The resort level dataframe showed some skewed distributions, especially fastQuads, fastSixes and trams which are affected by small numbers of extreme values. Domain knowledge tells us these may be important. Nonlinear transformations of these features should be considered.

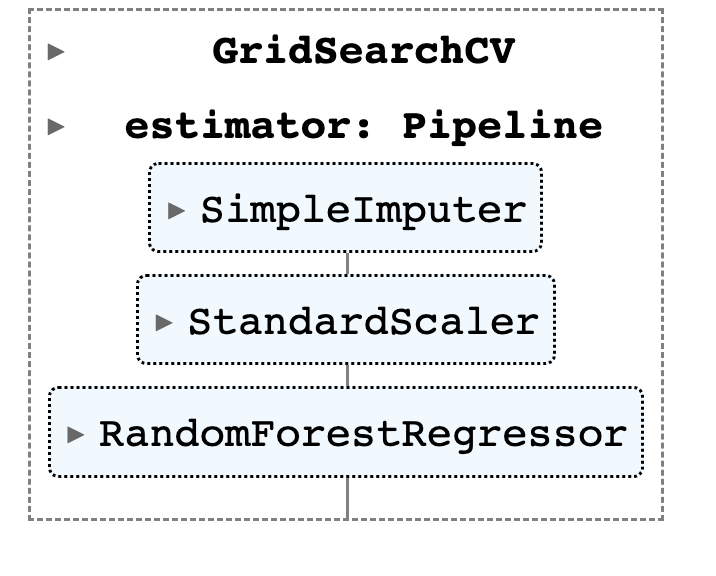
Primary Component Analysis was used to capture primary and secondary variance. Price as 'AdultWeekend' column will be used as target feature in modeling.

No clear pattern has emerged between state and ticket price. The recommended plan moving forward on handling states labels is to treat all states equally and work toward building a pricing model that considers all states without focusing on any particular state. There is no evidence yet of clear grouping, but this exploratory analysis captured potentially relevant state data in features most likely to be relevant.

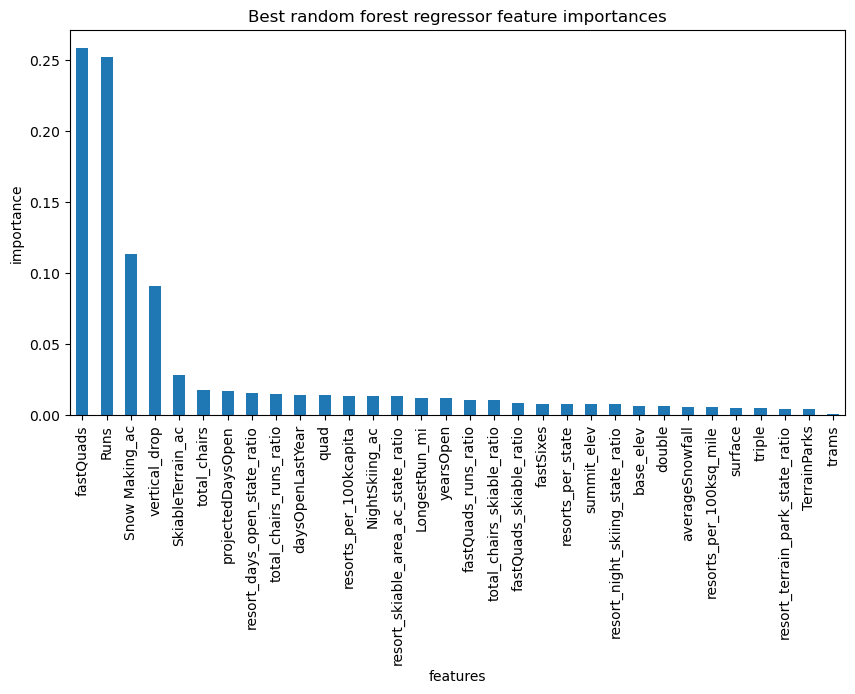
In the pre-processing and training stage, the data “ski\_data” was split into a 70/30 train/test split. A baseline check was performed using the mean as a predictor of ticket price. The MAE for this approach was about 19. The R squared value was, predictably, zero. A linear model was built after imputing median ticket price and scaling the data. It was also run using the mean ticket price for imputation, with similar results. The linear models yielded a MAE of about 9.

This was repeated using a pipeline which included Imputation (median), scaling, and linear regression. It yielded the same results as the manual approach.

The model was suspected to be overfitting, so the pipeline was redefined using different k values. Using best\_params\_, k= 8 was found to be ideal.

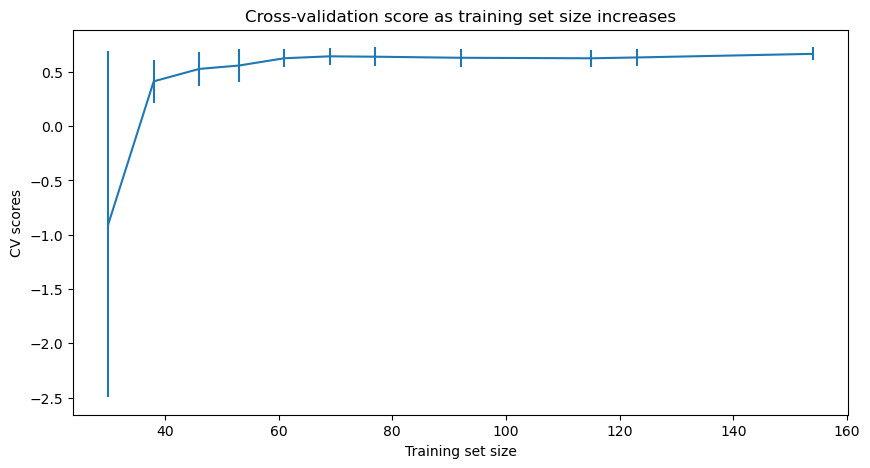
Finally, a Random Forest Model was added to the pipeline and successfully identified the important features as “fastQuads, Runs, Snow Making\_ac and vertical\_drop.” SkiableTerrain\_ac was the next important feature after the first four, but unlike the others, it fell far below .05 in importance and actually had an inverse relationship with ticket sales, perhaps due to unexplored reasons involving total ticket sales.

Important features: vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, Runs, LongestRun\_mi, trams, SkiableTerrain\_ac

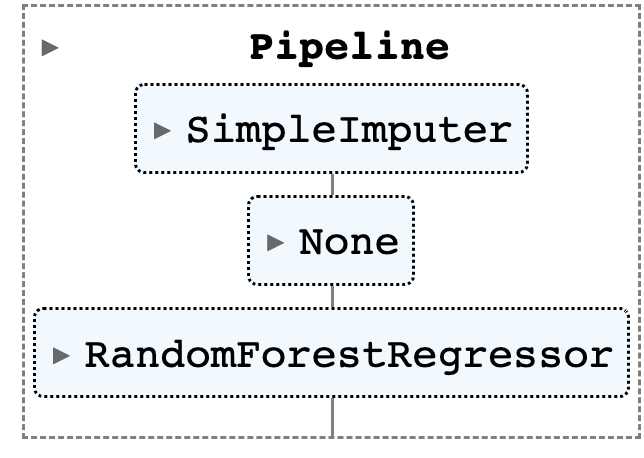


The performance for the random forest regression model mean absolute error mean is 9.64 and the mean absolute error standard deviation is 1.35, a lower cross-validation MAE. The test set produces performance consistent with cross-validation results.

Additional data collection is not recommended. The cross validations score as training set size increases validated our training set size as CV scores leveled out around sample-size 40-50. Hyperparameters could be further tuned and the number of features considered by the model could be reduced.



The model was refitted on all available data (excluding Big Mountain).



The current adult weekend ticket price charged by Big Mountain is $81. With no changes to Big Mountain (and including the new run), modeling suggests increasing ticket price. The predicted ticket price $95.87 with a mean absolute error of $10.39

According to this model, further price increase could be supported in the marketplace with feature modification at Big Mountain. Here is the shortlist of provided options with the model’s prediction of the effect on market-value ticket price.

* Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so 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.

* Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage.

This scenario increases support for ticket price by $1.99

Over the season, this could be expected to amount to $3,474,638

Cost of run expansion and chair lift installation, maintenance, and staffing is not known, but would have to be found and considered. Adding 2 acres of snow making cover to this combination had no effect.

* Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

No data was provided on other resorts’ total tickets sold or how many consecutive days their visit lasted.

The model shows potential for ticket price increase, and further market-supported increase with modifications. Each provided shortlist scenario was examined in this analysis.

If future predictions of new scenarios are required, the Jupyter notebook is available for use.