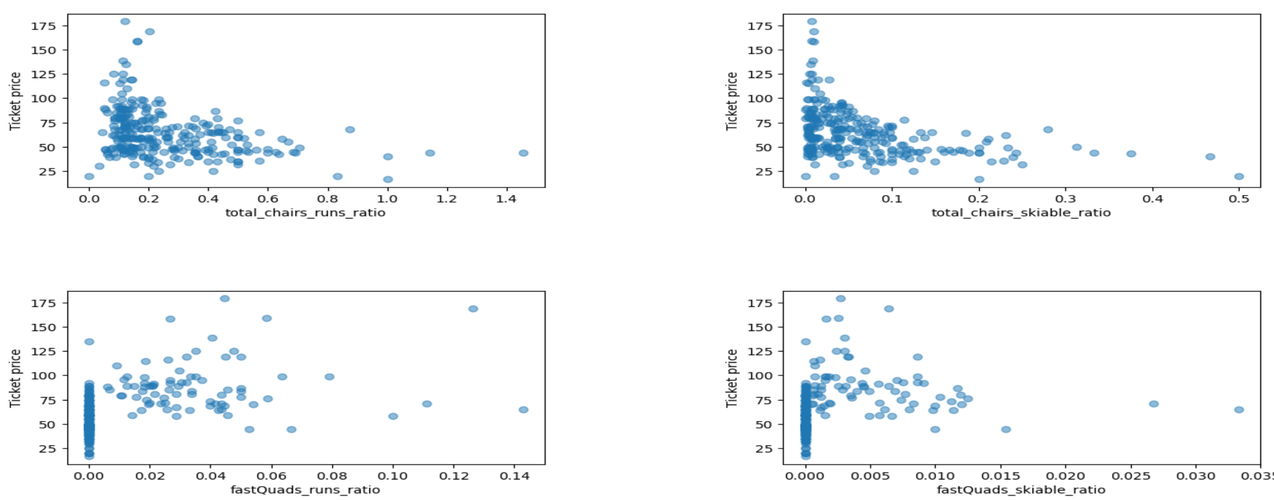
How can Big Mountain Resort extract better value from their tickets by at least $1,540,000 through cutting costs without undermining ticket value or increasing ticket prices by the next skiing season?

I began by exploring the dataset, which originally had 330 rows and 27 columns, prominently featuring "Big Mountain Resort." To improve data quality, I removed the "fastEight" column due to its absence in many rows and constant value of 0 in most cases. Rows without price data (our target value) were also removed. I addressed null values and corrected outliers, such as the "SkiableTerrain_ac" value for "Silverton Mountain" and an inaccurate "yearsOpen" value of 2009. Various exploratory analyses were conducted, including creating bar graphs for average ticket prices by state, box plots, descriptive statistics, and histograms.

I then created the "state_summary" table by grouping data by state and aggregating key metrics. This table was merged with the "usa_states_sub" table, which I sourced from the internet and contained state-related information. We chose "AdultWeekend" as the target feature due to fewer missing values compared to "AdultWeekday." As a result, I removed the "AdultWeekday" column and eliminated remaining null rows in "AdultWeekend." This process resulted in a dataset with 277 rows and 25 columns.

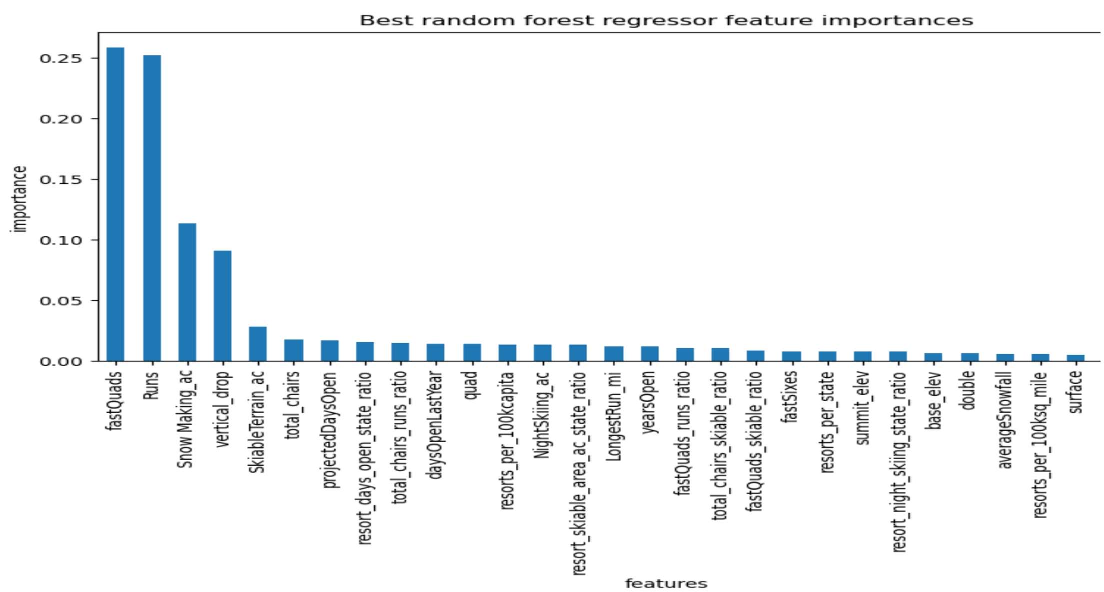
Our exploratory data analysis unveiled correlations between ticket prices and several factors, including 'fastQuads,' 'Runs,' 'Snow Making_ac,' 'total_chairs,' and 'vertical_drop.' Notably, the new ratio, 'resort_night_skiing_state_ratio,' displayed the strongest correlation with ticket prices. Intriguingly, a scatter plot highlighted a counterintuitive trend: an increase in the number of chairs per run seemed to result in lower ticket prices. This may stem from exclusive pricing for a limited number of chairs as opposed to mass pricing for a larger quantity. To gain a more comprehensive understanding, an analysis of total visitor numbers to each resort is required, although this data is presently unavailable. Fast quads emerged as a significant factor affecting ticket pricing; resorts with these facilities tend to command higher prices. While numerous variables exhibit correlations with ticket prices, a modeling approach was needed to identify the most influential variables and their relationships with ticket pricing.



For data preprocessing and training, we constructed a linear regression pipeline, including imputation, scaling, and linear regression steps. This linear regression model showed a significant improvement over the basic average method. Cross-validation indicated a variability range of [0.44, 0.82]. We performed hyperparameter tuning and feature selection via grid search and found that k=8 was the optimal number of features. Notably, 'vertical_drop' was the top positive predictor, followed by 'Snow Making_ac,' 'total_chairs,' 'fastQuads,' and 'Runs.'

A similar pipeline was applied to a random forest regressor, using RandomForestRegressor(). Cross-validation resulted in a mean R2 of 0.698 and a standard deviation of 0.071. We performed hyperparameter tuning for n_estimators, StandardScaler(), and imputation method (mean or median).

Comparing both regressors, the random forest approach had a mean absolute error of 9.5, outperforming linear regression's 11.8. The random forest also showed lower variability, with a standard deviation of 1.35 compared to the linear regression's 1.62. Consequently, we selected the random forest model due to its lower error and reduced variability.



For modeling, we fitted our model and conducted cross-validation, resulting in an average mean absolute error of 10.39 with a standard deviation of 1.47. Using our model to predict the price for Big Mountain Resort, we arrived at a figure of 95.87, significantly higher than the original price of 81.00, suggesting potential for a price increase.We compared Big Mountain Resort to other resorts using a histogram of key predictive features, and it performed well in all areas except the number of trams, where it lagged behind only a few others.

We explored various scenarios, starting with an assumption of 350,000 visitors buying five tickets each. Closing one run had no apparent impact, but closing 2 or 3 runs gradually reduced support for ticket prices and revenue. Surprisingly, closing 4 or 5 additional runs after closing 3 had no further effect on ticket prices, but closing 6 or more runs led to a significant drop in revenue. In the second scenario, adding a run, increasing vertical drop by 150 feet, and installing an extra chair lift raised ticket prices by 1.99, resulting in an additional revenue of 3,474,638, offsetting the chair lift cost of 1,540,000. Other tried scenarios did not change ticket prices.

In conclusion, we recommend that Big Mountain Resort consider adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift to potentially increase revenue.