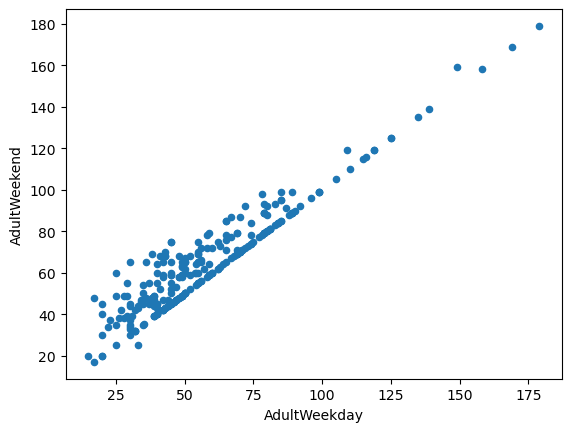
Big Mountain Resort, located in Montana, is a popular ski resort that attracts approximately 350,000 visitors annually. It boasts a range of facilities, including 105 trails, 11 lifts, and a recent addition of an extra chair lift to improve visitor distribution. The resort has traditionally adopted a premium pricing strategy but is facing concerns about its pricing and overall operational efficiency.

The initial dataset comprised 330 rows and 27 columns, with problems such as missing values, duplicate rows and irrelevant columns. Specific approaches were used to deal with missing values, remove redundant columns and convert the 'state' column to a categorical type. US population data were integrated and cleaned, merging with ski resort data. The two variables that could be used as target variables were 'AdultWeekend' and 'AdultWeekday'. So A scatter plot was made to visualize the relationship between these two variables.

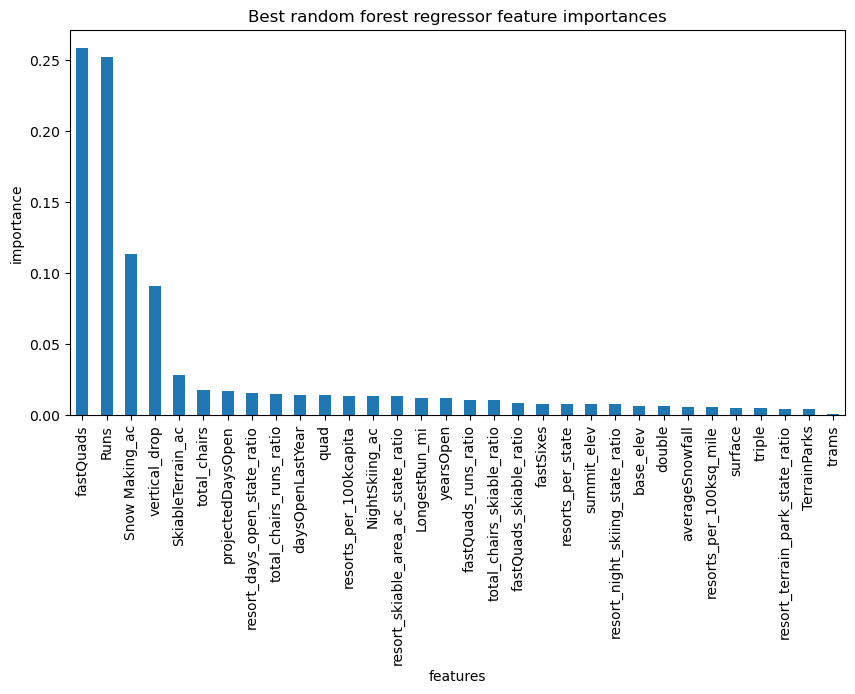


Both variables can be used as target variables, but the choice was finally made for the 'AdultWeekend' variable, mainly because it contains fewer missing variables and could therefore provide more information.

Despite an initial exploration of the relationship between states and ticket prices, no discernible patterns emerged, leading to the decision to treat all states equally in subsequent modeling efforts. Principal Component Analysis (PCA) was applied to extract relevant state-level information and reduce dimensionality, focusing on features like the number of resorts per state, skiable area, days open, and terrain parks. Throughout the analysis, caution was advised regarding multicollinearity introduced by new ratio features, and the counterintuitive relationships observed, such as the ratio of chairs to runs affecting ticket prices. The decision was made to emphasize equality among states in the modeling process while capturing state-level nuances through carefully selected features.

We also produced a heatmap of correlations to verify the relationships between the target variable and the other variables.

A baseline model that predicted prices based on the average was created, providing a simple benchmark for comparison. Subsequently, a linear regression model was built, revealing key features that influence ski resort pricing. Cross-validation was employed to estimate the model's performance, showing a mean absolute error of approximately 10.50 with a standard deviation of 1.62. The model was then evaluated on a test set, producing a mean absolute error of 11.79, consistent with the cross-validation estimate. Additionally, a random forest regressor was explored, incorporating preprocessing steps such as imputing missing values with the median and omitting feature scaling. Cross-validation indicated a mean absolute error of approximately 9.64 with a standard deviation of 1.35. Evaluation on the test set demonstrated a mean absolute error of 9.54, aligning well with the cross-validation results. After comparing the two models, the random forest regressor was selected for further development due to its lower mean absolute error and consistent performance across cross-validation and the test set.



The company selected four options, which were used as possible scenarios: permanently closing up to 10 of the least-used trails, increasing the vertical drop by adding a trail at a point 150 feet lower, adding 2 acres of snowmaking coverage, or increasing the longest trail by 0.2 mile to reach a length of 3.5 miles, requiring an additional 4 acres of snowmaking coverage.

In its current state, Big Mountain Resort offers a weekend adult ticket at $81.00. However, our model suggests that an increase to $95.87 could be considered, given the resort's facilities and market conditions, indicating potential room for a tariff increase. Regarding the modeled scenarios, the closure of up to five runs does not seem to significantly influence ticket prices, while closing six runs or more leads to a noticeable decrease. Conversely, adding a run, increasing the vertical drop by 150 feet, and installing a new chair lift could support a ticket price increase of $1.99, potentially generating around $3,474,638 in additional revenue over the season. it would be advisable to increase the ticket price to $95.87 to better reflect the value of Big Mountain's facilities in the market. Additionally, to offset the additional operating costs of the new chair lift, it would be prudent to calculate the cost per ticket by dividing the total cost by the expected number of visitors purchasing 5-day passes.

The analysis encountered limitations primarily related to data deficiencies. The absence of detailed cost information, beyond ticket prices and the new chair lift's operating cost, hinders a comprehensive understanding of the financial implications of different scenarios. To address this, additional data on maintenance, labor, and operational expenses would be invaluable. The modeled price, significantly higher than the current price, prompts questions about the model's accuracy and potential missing variables. This discrepancy, potentially surprising to business executives, underscores the need for a thorough examination and clarification of assumptions.

If the model proves useful to business leaders, integrating it into decision-making processes could be facilitated through a user-friendly interface or tool. This would empower business analysts to independently explore different parameter combinations and scenarios without frequent engagement with data scientists. By making the model more accessible and adaptable, the business can embrace a more agile, data-driven decision-making approach.