**Introduction**

Big Mountain resort has requested a data drive solution to reduce operating costs and increase revenues. This report provides an overview of the data analysis, the model built to predict a ticket price to achieve the solution and a recommended pricing strategy.

**Problem statement**

This problem statement questioned if there were opportunities for Big Mountain Resort to reduce operating costs by $1.54 M within a year by increasing revenue through the adjustments to ticket pricing strategies based on the value and utilization of different resort facilities.

**Data Wrangling**

The dataset for this project came from a CSV file provided by the client which contained information on ski resorts across the U.S. Big Mountain Resort, the target resort, was included in the data provided. The initial dataset included 330 rows and 27 columns of data but after cleaning the data for usability and format, was reduced to 277 rows and 25 columns of data. Initial analysis of the resultant dataset indicated that a likely target feature that could be used for predicting a pricing strategy was the Adult Weekend ticket price. The Adult Weekend ticket price was compared against the Adult Weekday prices and found that in the target state of Montana, the AdultWeekend and AdultWeekday prices were the same. Either one could have been used as the primary feature; however, Adult Weekend prices had fewer missing data points and was chosen because of this.

**Exploratory data Analysis**

Exploratory Data Analysis focused on reviewing data of the chosen feature for Adult Weekend ticket prices. Analysis focused on looking at various state and resort features. Initially, State populations were used along with the size of the state, numbers of resorts, skiable area and days open. These features were used to gain a broad scope of the competitive skiing landscape of each state. In looking at the visualizations of state features against ticket prices, there did not appear to be a relationship between state and ticket prices. Therefore, the state labels were not used in the data since they did not provide meaningful data or distinctions. This discovery led to the creation of state-level features by combining the state summary features with ski resort data. The heatmap and histograms below were used to identify features that could be used as a basis for increasing ticket prices.

A screenshot of a computer screen

AI-generated content may be incorrect.

Fig. 1 Heatmap identifying patterns between listed variables.

A group of blue dots

AI-generated content may be incorrect.

Fig. 2 Scatterplots of numeric features against ticket prices.

After correlation, the features that were most notable to further pursue were fast quads, number of runs and total chairs, snow-making capability, vertical drops, and resort night skiing.

**Model Preprocessing with feature engineering**

A training and testing grouping was created from the data that was used in the exploratory data analysis notebook. The split was 70% of the data for training and 30% of the data for testing. The data from Big Mountain Ski Resort was excluded for this. Only numeric features were used. A baseline of performance was created by using the mean as a predictor. The metrics used to assess the predictor was R-squared, Mean absolute error, and Mean squared error. The notebook cycled through creating a function to create a metrics and calling on a pre-built function to perform the task. Using mean absolute error as the metrics, calculated values indicated that if just the mean was used to predict ticket prices, the ticket prices would likely be off by about $19.

**Algorithms used to build the model with evaluation metric**

Two models were built and then compared. The first model was a linear regression model and the second was a random forest regression model.

The first model was built with a pipeline that used SimpleImputer, StandardScaler, SelectKBest and LinearRegression. This linear regression model used GridSearchCV cross-validation to estimate performance on the train and test data and found the test data consistent with the training data. The first model identified the top 8 features that were best for use in the model with the top 5 being vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, and Runs in that order. Using mean absolute error as the metrics, calculated values indicated that if the linear regression model was used to predict ticket prices, the ticket prices would likely be off by about $11.

A graph with blue squares

AI-generated content may be incorrect.

The second model was built with a pipeline that used SimpleImputer, StandardScaler, and RandomForestRegressor. The random forest regressor also used GridSearchCV cross-validation to estimate performance with the test result being consistent with the training data. This model identified the top 4 features that would be useful that were also identified in the linear model; however the order of precedence was slightly different with fastQuads topping the list followed by Runs, Snow Making\_ac, and vertical\_drop. Using mean absolute error as the metrics, calculated values indicated that if the linear regression model was used to predict ticket prices, the ticket prices would likely be off by about $9.

A graph with blue and white text

AI-generated content may be incorrect.

**Winning model and scenario modelling**

The model selected for use was the random forest regressor. It had a lower absolute error, produced the least variability and the test set was consistent with the cross-validation results.

Four scenarios were run to determine a recommendation to increase revenue based on features. For each scenario, the number of visitors to the resort over the season with an average visit time of five days was 350,000.

* Scenario 1: This scenario focused on the possibility of closing Runs to decrease cost and increase revenues. This scenario determined that closing 1-5 runs would reduce the support for increased ticket prices while closing 6 or more greatly reduced the support for a price increase.

A graph of a price

AI-generated content may be incorrect.

* Scenario 2: This scenario added a run while increasing the vertical drop by 150 feet and installing an extra chair lift. This scenario increased support for ticket prices by $1.99 resulting in an increase of revenues of $3.47M over the season.
* Scenario 3: This scenario added 2 acres of snowmaking to Scenario 2 without a discernible increase in revenue.
* Scenario 4: This scenario increased the longest run by just under a quarter of a mile which also resulted in no discernible increase in revenue.

**Pricing recommendation**

Big Mountain Ski Resort currently charges $81.00 per Adult Weekend ticket. Modeling indicated that this price could be increased to $95.87 based on cross-validation and available data. The result infers that an increase of the Adult Ticket price could be feasible even with a mean absolute error of $10.40.

**Conclusion and Future scope of work**

The price of $95.87 is recommended to achieve Big Mountain Ski Resort’s objective of increasing revenue. It is a data driven solution based on modeling of available data and is best supported by the scenario of adding a run and increasing vertical drop at the resort. Further modeling and testing should be performed to look at run closures or other potential changes.