Problem statement

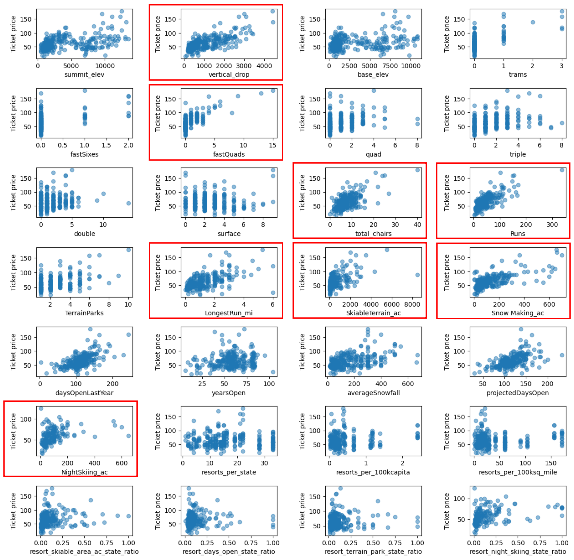
How can Big Mountain Resort increase its profit by at least $2 million next season by adopting a more advanced pricing policy that leverages the resort's key facilities as competitive advantages?

Data Wrangling

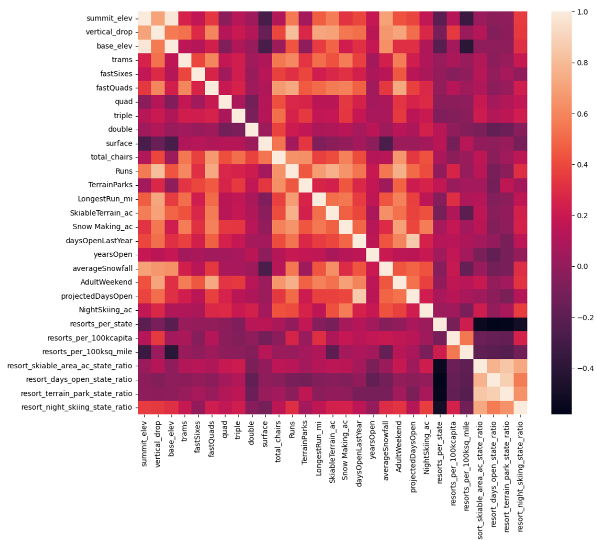
The initial dataset had 330 resorts and 27 features. After removing rows and columns with missing or redundant data, and correcting some values, the cleaned dataset contains 277 rows and 25 columns. The data was also aggregated by state, with population and area added, creating a new state-level dataset. **Initial analysis suggested weekend ticket prices as a potential target feature for modelling.**

Exploratory Data Analysis (EDA)

The analysis made didn't reveal any obvious relationship between states and ticket prices, so ***there is no need to handle the state label in modelling***.



Graph 1: the scatter plots of various features vs. ticket price in red boxes represent the positive correlation

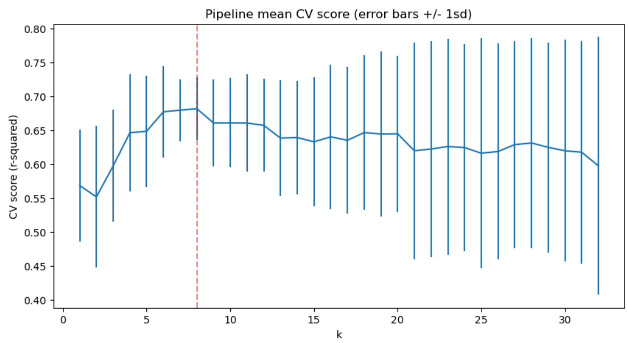


Graph 2: the heat map proves the correlation the same features vs. ticket price

Correlation analysis showed that ticket price positively correlates with several features: *Total Chairs, Fast Quads, Runs, Vertical Drop, Longest Run, Skiable Terrain, Snowmaking, and Night Skiing*. **These correlations support using ticket prices as the target feature for modelling.**

Pre-Processing and Training Data

A hyperparameter search with cross-validation on **the Linear Regression model** was conducted to determine the optimal number of features for predicting ticket prices. The best results (Graph 3), measured by how well the model explains the variability in ticket prices (R^2 metric), were achieved using 8 features: *Vertical Drop, Snowmaking Area, Total Chairs, Fast Quads, Runs, Longest Run (mi), Trams, and Skiable Terrain*. These features largely align with those revealed in the earlier correlation analysis.

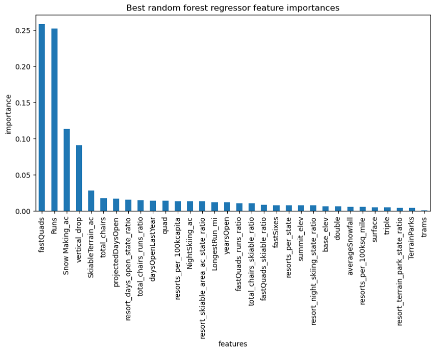


Graph 3: Model performance vs. number of features

Another model used was the **Random Forest**. A hyperparameter search with cross-validation showed that the best results, measured by how well the model explains the variability in ticket prices (R^2 metric), were achieved with median imputation (rather than mean) and without scaling. The most important features for the Random Forest model, in descending order of importance (Graph 4), were *Fast Quads, Runs, Snowmaking,* and *Vertical Drop*. These features also overlap with those identified in the correlation analysis and the Linear Regression model.

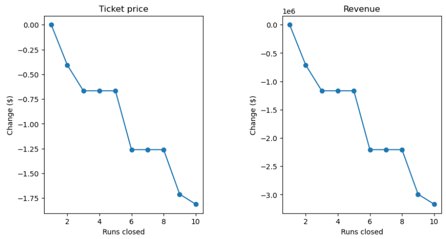
Winning model and scenario modelling

The Random Forest model (RFM) was chosen over the Linear Regression model due to its better performance on both the training and test sets. The average difference between the model's predictions and actual values (mean absolute error) was 9.6 for the training set and 9.5 for the test set in the Random Forest model, compared to 10.5 and 11.8 for the Linear Regression model.



Graph 4: Best features for Random Forest model

Scenario Analysis:



Graph 5: Influence of number of Runs closed on Revenue and supported ticket price

Scenario 1: Closing up to 10 of the least-used Runs

The model shows (Graph 5) that closing one run has no impact, while closing 2 or 3 reduces ticket prices and revenue. After 3 closures, closing 4 or 5 runs has no further effect, but closing 6 or more causes a significant drop. This revenue decrease should be weighed against potential cost savings.

Scenario 2: Increasing the Vertical Drop by adding a run 150 feet lower (requiring an additional chair lift but no extra snowmaking) increases the supported ticket prices by $1.99, leading to $3.5 million in revenue. However, this gain must be weighed against the required investments and operational costs.

Scenario 3: Adding 2 more acres of snowmaking doesn't increase supported ticket prices or revenue compared to Scenario 2, but adds costs, making it an impractical option.

Scenario 4: Increasing the Longest Run by 0.2 miles with adding 4 acres of snowmaking coverage has no effect on the ticket price or revenue.

Pricing recommendation

The suggested by RFM **ticket price that could be supported by Big Mountain's facilities is $95.87** per day compared to current price $81.00. For comparison the recently installed chair lift generates an additional cost $0.88 per ticket.

Conclusion

A Random Forest-based ticket price prediction model was developed, enabling advanced pricing strategies and evaluation of business scenarios. Results show that Big Mountain Resort has room to raise ticket prices to cover costs from the new chair lift. Among the management's proposed scenarios, scenarios 1 and 2 are viable with further assessment. I recommend closing one least-used run, which would reduce costs without affecting revenue.

Future Scope of Work

To improve the analysis, additional data such as customer satisfaction, demand elasticity, and the number of visitors per resort would be valuable. Understanding capital and operational costs is crucial for evaluating development scenarios, as focusing only on revenue without considering expenses can be misleading.

Further research into competition and audience behaviour is needed to refine pricing strategies and identify true competitors.

The model can evolve into a scenario calculator, incorporating a user-friendly interface with features like facilities inventory, cost-revenue statistics, and scenario definitions, allowing better simulation of business development and optimization strategies.