Big Mountain Resort Data Science Project Report

Submitted by:
Dennis Dang

Executive Summary

Big Mountain Resort aims to optimize ski ticket prices to increase revenue. This report details the data-driven methodology used to explore, model, and recommend pricing strategies. A comprehensive analysis of resort data reveals opportunities to adjust pricing in line with value-added services, which is anticipated to boost profitability while maintaining customer satisfaction.

Table of Contents

Introduction

Problem Statement

Big Mountain Resort charges premium ticket prices within their market segment but may not be fully capitalizing on their facilities. The Big Mountain Resort is aiming to increase its revenue for the 2024 snow season. To achieve this, we hypothesize that a thorough analysis of ticket pricing factors across the industry, combined with an optimization of our own pricing strategy, will unlock new growth opportunities. Success hinges on our ability to:

- Accurately identify the critical elements that drive ticket pricing.
- Develop strategic recommendations rooted in robust data analysis to inform pricing adjustments.

The scope of this initiative centers on dissecting industry-wide data from comparable ski resorts to pinpoint the determinants of ticket pricing and to understand how different pricing strategies might influence our revenue trajectory. We face the challenge of relying on the availability of precise and comprehensive industry and financial data.

Key insights into operational constraints and data management will be sought from Jimmy Blackburn (Director of Operations) and Alesha Eisen (Database Manager). The primary data source is a CSV file detailing attributes and pricing of 330 ski resorts, augmented by Big Mountain Resort's historical and projected revenue figures. This data-driven approach will guide our pursuit of our revenue increase target for the 2024 season.

Methodology

The methodology for optimizing ticket pricing at Big Mountain Resort is organized into four key stages: Data-Wrangling, Exploratory Data Analysis, Pre-processing and Training Data Development, and Modeling.

Data-Wrangling

The provided CSV file contained resort characteristics and ticket pricing data of 330 ski resorts across the United States. A thorough examination was conducted to identify and rectify any inconsistencies, such as missing values or anomalies. Related information, like geographic data, was aggregated to provide a clearer picture of regional factors that may influence pricing strategies. Records missing our primary objective, 'AdultWeekend' ticket price were removed. The initial dataset contained 330 entries. After cleaning, we refined this number to 277 high-quality records. As demonstrated in Figure 1, we visualized the distributions of each variable to detect outliers or anomalies. By understanding what these variables represent, we can make informed decisions about how to handle irregular data.

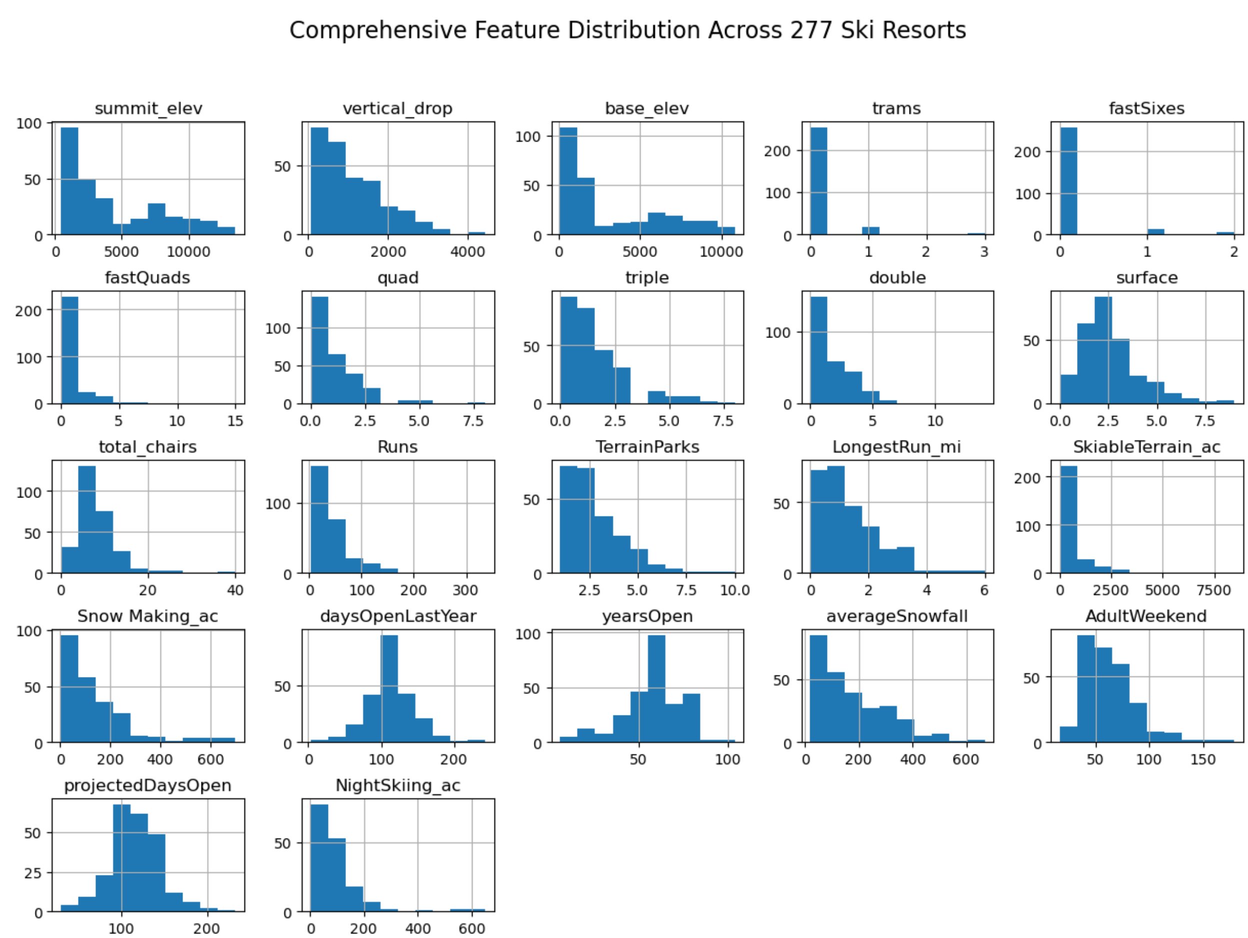


Figure 1: Distributions of 22 Numerical Features for 277 Ski Resorts representing the range and frequency of each feature.

Exploratory Data Analysis (EDA)

Our EDA focused on a thorough investigation of both numerical and categorical variables within a dataset of 277 ski resorts. The aim was to uncover underlying patterns and relationships that could inform the development of a predictive pricing model.

Numerical Features and Feature Generation

Numerical features such as resort elevation, number of ski runs, and area available for night skiing were thoroughly examined. To augment our analysis, we generated insightful state-level features by aggregating state population data and state area in square miles. From these aggregates, we created nuanced features reflecting the density and availability of ski resort amenities relative to state statistics:

- resorts_per_100k_capita: The number of ski resorts per 100,000 people in the state.
- resorts_per_100ksq_mile: The number of ski resorts per 100,000 square miles of state land.

Additionally, we engineered several ratios to capture the proportionate scale of resort features within their respective states:

- resort_skiable_area_ac_state_ratio: The ratio of a resort's skiable area to the total skiable area available in the state.
- resort_terrain_park_state_ratio: The ratio of the number of terrain parks at a resort to the total number in the state.
- resort_night_skiing_state_ratio: The ratio of a resort's night skiing area to the total night skiing area available in the state.
- resort_days_open_state_ratio: The ratio of the number of days a resort is open to the total number of operational days for all resorts in the state.

These ratios provide a relative scale that contextualizes each resort's features within the broader state-level market. The engineered features also provide a nuanced view of market saturation and potential access to ski facilities for the population.

Categorical Features

The categorical data was leveraged through geographical identifiers such as resort names, states, and regions. These labels facilitated an analysis of location-based pricing differences, a dimension commonly regarded as significant in market analysis.

Principal Component Analysis (PCA)

The PCA focused on identifying which numerical and categorical features play a significant role in determining resort ticket prices. Contrary to initial assumptions, the PCA revealed that the state-level differences, reflected in categorical features, did not exhibit a strong correlation with resort ticket prices. The state labels did not show a clear influence across the principal components extracted during the PCA. This suggests that while geographical location is an intuitive factor for pricing, it does not significantly drive the differences in ticket prices when considering the dataset as a whole. To visually illustrate these findings, refer to the following PCA scatter plot of ski resorts by state:

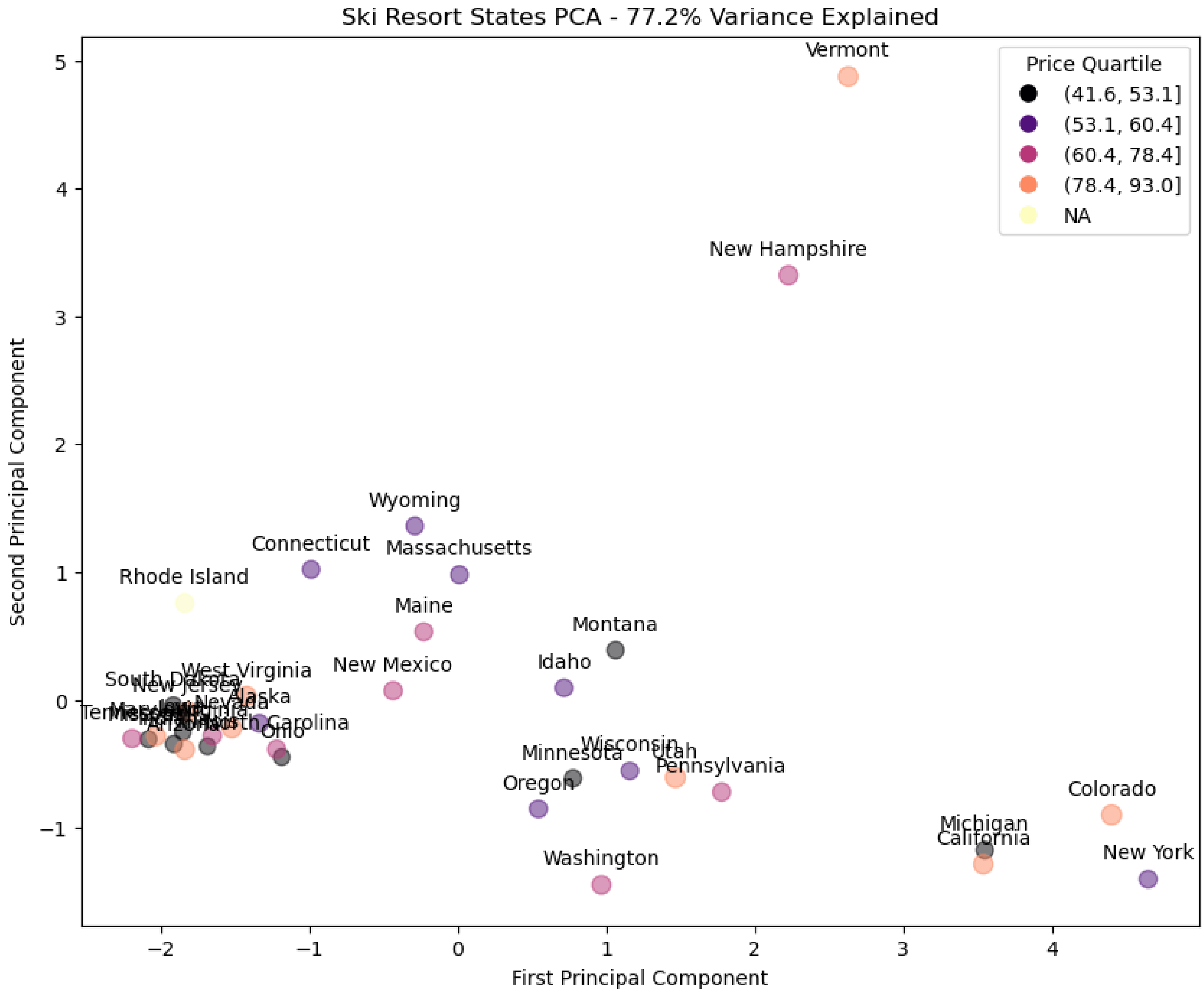


Figure 2: PCA scatter plot highlighting the dispersion of states across principal components, color-coded by ticket price quartiles. This visualization demonstrates the lack of a clear pattern correlating states with ticket prices.

The lack of a dominant pattern between state labels and ticket prices has led us to a strategic decision in our modeling approach—we will not prioritize state labels over other variables. Instead, we will treat geographic data equally alongside other features. This is a deliberate choice that sets the direction for our modeling strategy.

Correlation Analysis

The correlation analysis, shown in Figure 3, highlighted a stronger link between ticket prices and certain features like 'fastQuads', 'Runs', and 'Snow Making_ac'. These findings informed our feature selection process, emphasizing variables with a more direct relationship to the ticket prices.

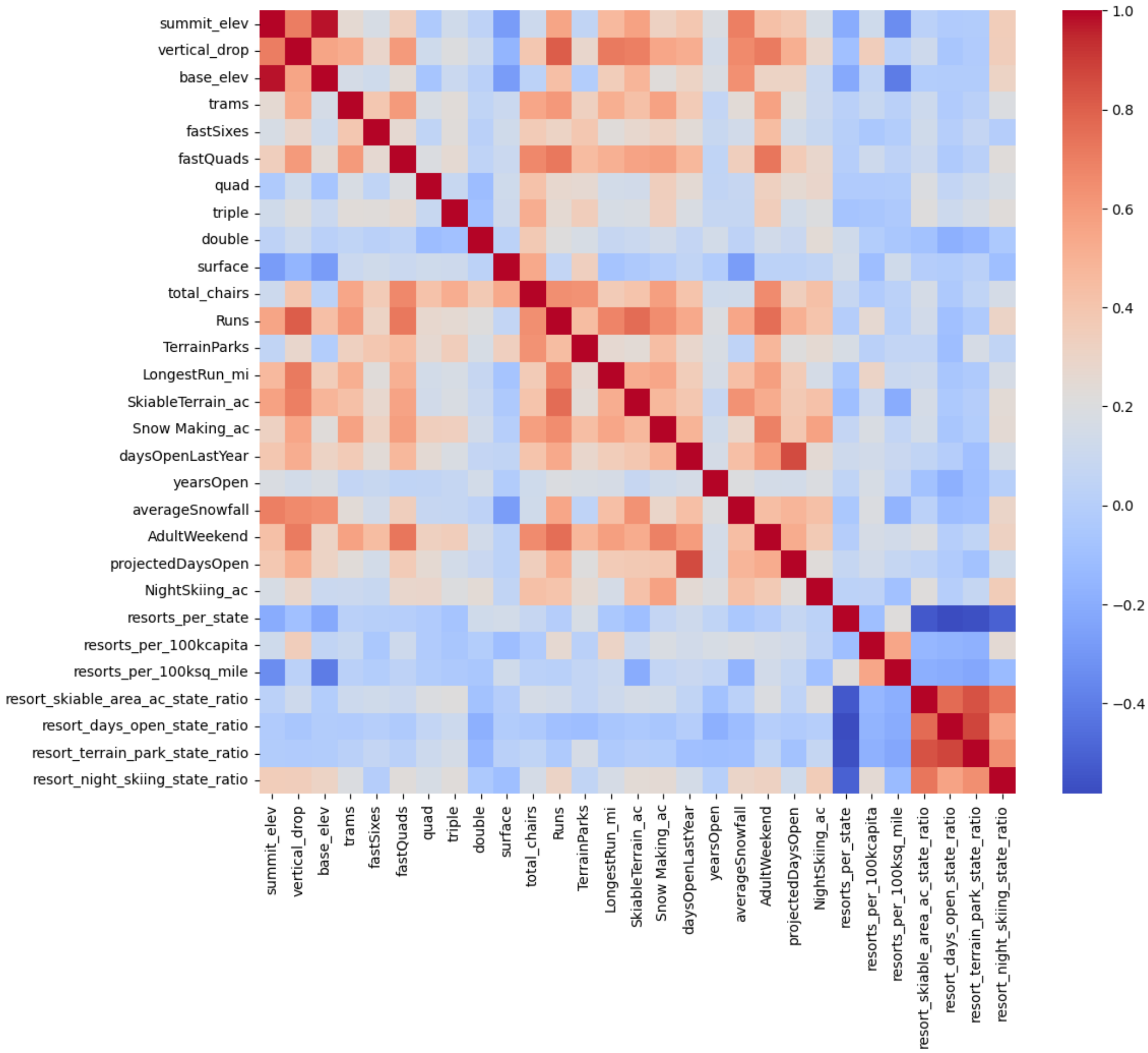


Figure 3: Heatmap of Correlation Coefficients Among Resort Features. This plot displays the strength and direction of relationships between different resort variables, with a focus on those most strongly associated with ticket prices. Key insights include the notable correlations between 'AdultWeekend' ticket prices and features such as 'fastQuads', 'Runs', and 'Snow Making_ac'.

Pre-Processing and Training Data Development

We partitioned the data into training and test sets, which is crucial for an objective evaluation of our predictive models. To deal with missing values, we opted for median imputation – a reliable approach due to its resistance to outliers. For our Linear Regression Model, we also implemented feature scaling to ensure each variable contributed equally to the model.

Algorithms Used and Evaluation Metrics

We compare the accuracy of our Linear Model and Random Forest Model to the baseline model that uses the mean as a constant prediction for ticket prices. We employed 5-fold cross-validation to scrutinize the linear model's performance, comparing both Mean Absolute Error (MAE) and Mean Squared Error (MSE) across the folds. This process allows for a more nuanced evaluation of the model's predictive power and helps to identify potential overfitting, as evidenced by a lower training MAE. The cross-validation approach verified the model's consistency and reliability before deployment.

Linear Modeling and Feature Selection

Refinement of our linear model involved optimizing the number of features (K) used. Through an iterative process, we determined that the best K was 8. The selected features, shown in Table 1 alongside their feature coefficients, led to an improved linear model performance and demonstrated lower MAE than models burdened with excessive features.

	Coefficient
vertical_drop	10.767857
Snow Making_ac	6.290074
total_chairs	5.794156
fastQuads	5.745626
Runs	5.370555
LongestRun_mi	0.181814
trams	-4.142024
SkiableTerrain_ac	-5.249780

Table 1: Feature Coefficients from Linear Regression.

Random Forest Model and Feature Importance

A hyperparameter-tuned Random Forest model suggested that median imputation was beneficial, while feature scaling was not. The model highlighted the significance of fastQuads, Runs, Snow Making_ac, and vertical_drop — features also prominent in the linear model's selection. The Random Forest's test performance modestly exceeded the cross-validation estimates, showcasing its ability to generalize.

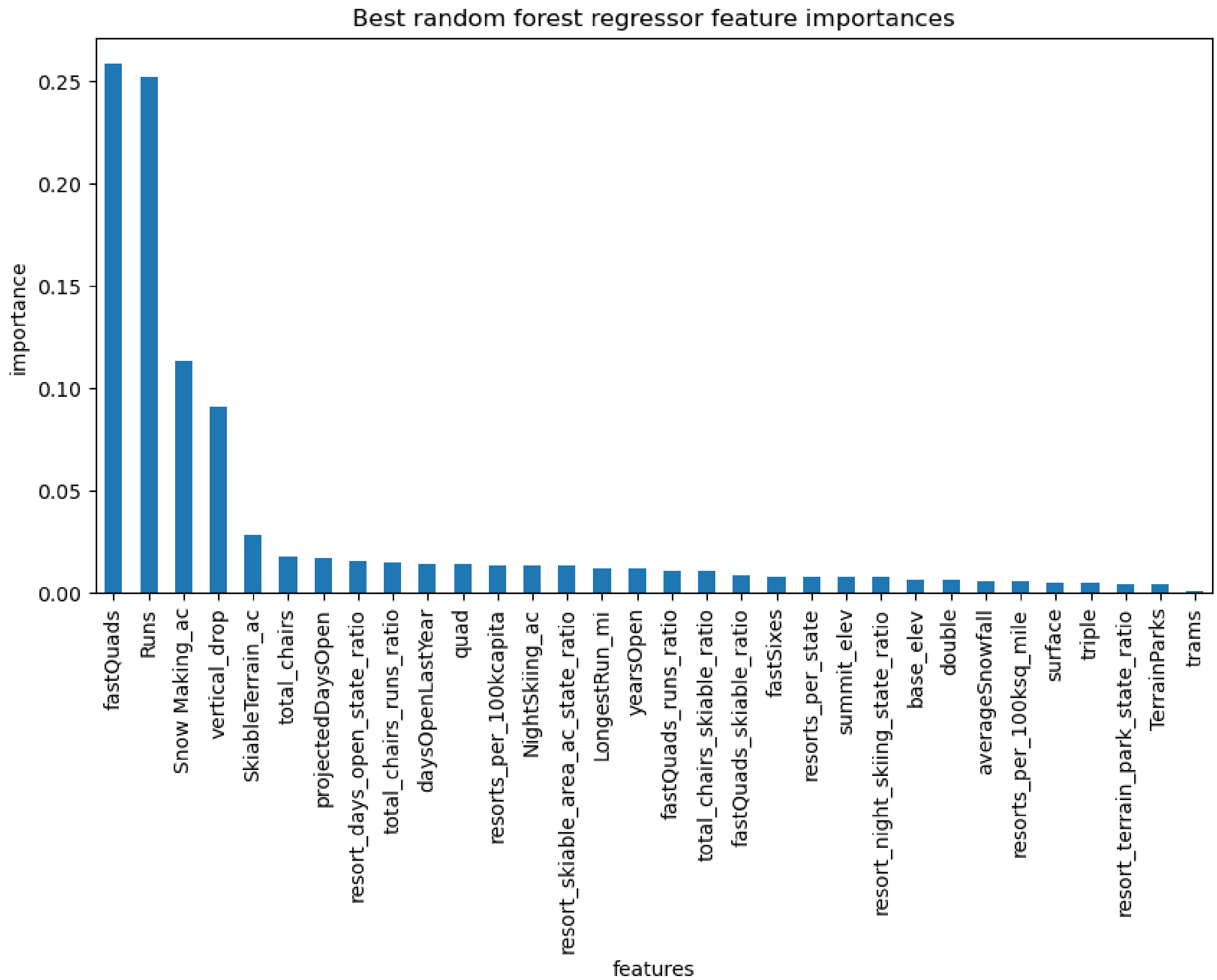


Figure 4: Feature Importances Derived from the Optimal Random Forest Regressor. The bar chart visualizes the relative importance of each feature in predicting ski resort ticket prices. It underscores the significance of 'fastQuads', 'Runs', 'Snow Making_ac', and 'vertical_drop' as the most influential predictors, which corroborates the findings from our linear model. This alignment between models reinforces the robustness of these features as critical factors in determining pricing.

Findings and Recommendations

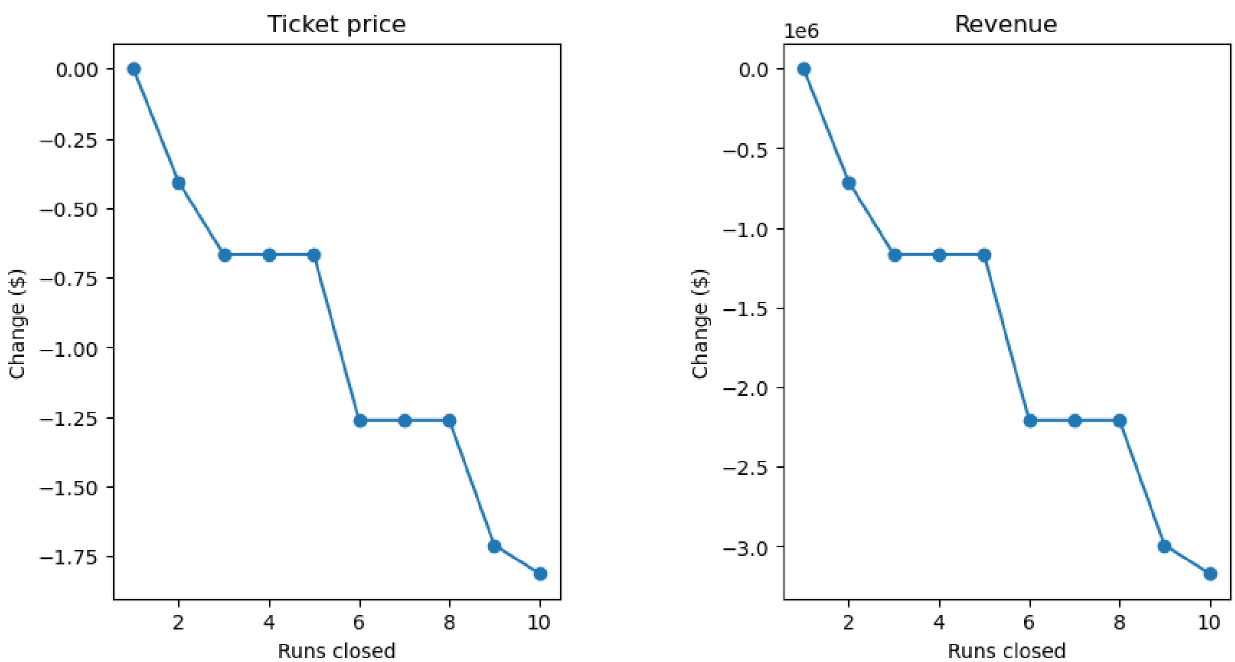
Winning Model

The Random Forest Regressor was ultimately chosen as the preferred model. It stood out for its lower MAE and the consistency of its cross-validation performance, demonstrating an advantage over the baseline and linear models.

Scenario Modeling

Our data-driven analysis of Big Mountain Resort's pricing strategy has revealed several scenarios that could influence the optimal ticket price. Current pricing stands at $81 while our model suggests a market-supported price of $95.87.

Scenario 1 - Closing Runs:
Closing a single run shows negligible impact on ticket pricing support. However, closing two to three runs marginally decreases price support, and shuttering six or more significantly undermines it.



Scenario 2 - Adding Facilities:
Enhancements including an additional run, a 150-foot increase in vertical drop, and a new chair lift can justify a ticket price increase of $1.99. This scenario forecasts a potential revenue boost of approximately $3,474,638 over the season.

Scenario 3 - Incremental Snow Making:

Adding 2 acres of snow-making capabilities to the developments in Scenario 2 does not provide a substantial increase in ticket price support or revenue.

Scenario 4 - Enhancing Longest Run and Snow Making:
The expansion of the longest run by 0.2 miles coupled with an additional 4 acres of snow making does not affect ticket price support, likely because these features are not highly valued in the predictive model.

Pricing Recommendation

Given these scenarios, the most strategic recommendation for Big Mountain Resort is to pursue the developments outlined in Scenario 2. This scenario enhances the resort's value proposition without diluting ticket price support. It’s essential to provide a detailed cost-benefit analysis to ensure that the projected revenue increase will offset the investment costs.

Conclusion

Our data-driven analysis suggests that Big Mountain Resort can increase its ticket prices for the 2024 season. Key features like the number of fast quad ski lifts, number of ski runs, and area of land covered by snow makers are highly correlated with ticket pricing, indicating that enhancing these aspects could potentially justify higher prices. Our recommended ticket price of $95.87 is based on a robust Random Forest Regressor model. The most promising strategy involves adding a new run, increasing the vertical drop by 150 feet, and installing a new chairlift, as outlined in Scenario 2. This approach is projected to bring in $3474638 in revenue over the season and aligns with the resort's upscale market position.

It is important for Big Mountain Resort to balance these changes with market conditions and customer satisfaction. Implementing the recommended pricing strategy should be done with careful consideration to maintain a value proposition that meets guest expectations. If managed well, these changes could lead to a profitable and successful 2024 season for the resort.

Future Scope of Work

Operational Testing
Before implementing any permanent changes, a pilot phase is advisable. This could entail temporary run closures or facility enhancements during off-peak times to gauge customer satisfaction and observe the financial ramifications without fully committing to the changes.

Comprehensive Cost Analysis
To refine our recommendations, we need a more detailed understanding of operational costs. Incorporating maintenance, labor, and capital expenditure data would allow for a more nuanced financial analysis.

Model Integration and Tool Development
Integrating our predictive model into the resort's strategic planning process can guide future investment decisions. Developing a user-friendly dashboard that allows for real-time scenario analysis would be an asset to the business team.