1. Introduction

Big Mountain Resort is a premium ski resort in Montana. It recently installed an additional chair lift that increases operating costs $1.54M for the current season. The current ticket pricing model is based on the market average. This is a generic strategy that does not capitalize on Big Mountain's premium facilities. It also misses an opportunity for data collection of revenue generation by facility. Big Mountain Resort seeks to build a pricing model that takes into consideration both market factors and its individual resort features. The company aims to make more data-driven decisions in setting their ski lift ticket prices in order to maximize revenue.

Problem Statement

The goal of this project is to maximize revenue by determining the optimal price for lift tickets at Big Mountain Resort. There are 2 criteria for success:

1) After implementing new ticket price, hit breakeven within reasonable time frame (3 years)

2) Build a model Big Mountain business analysts can use without facilitation by Data Science team post project completion

2. Methodology

Data Wrangling

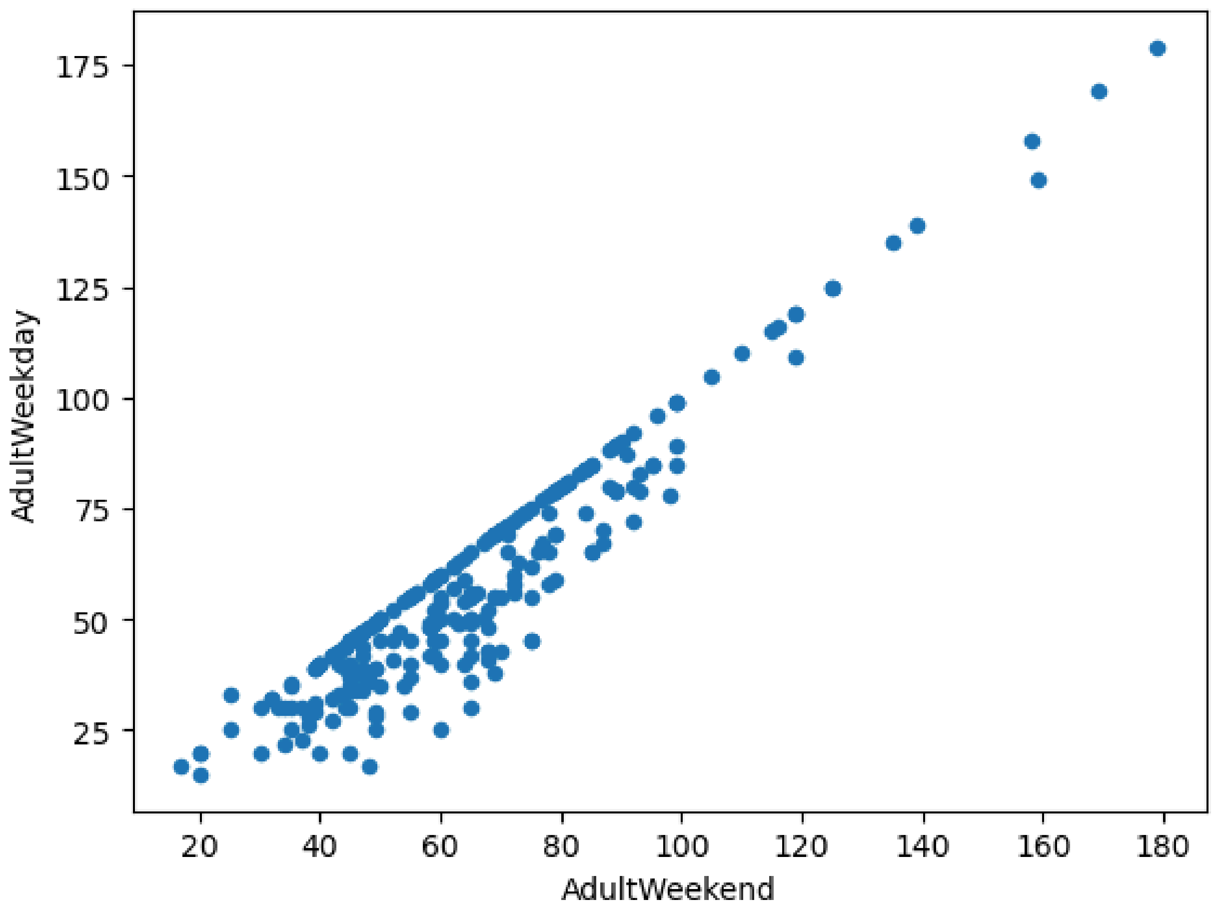
The ski_data dataset was used for this project. This dataset includes information on all ski resorts across the US, including categorical features like state region and resort name, and resort specific numerical features such as number of lifts, skiable terrain, days open, night skiing availability, and ticket prices. The data wrangling process involved collecting, auditing, and organizing the data, as well as a few data cleaning tasks. The dataset started with 330 entries, including an entry for our resort of interest, Big Mountain Resort. This entry did not have any missing values.

There were 3 main cleaning tasks performed on this dataset. The fastEight column was dropped entirely due to a lack of actual data. A record was removed due to a data entry error that made it unclear where the data was gathered from. And entries were dropped because they had no price data which was our target feature.

A separate dataset of population and area data for each US state was then pulled in. During this process we discovered a discrepancy between the list of state names and 'states' listed in the ski_data dataset. The state names with errors were then corrected.

Finally, the target feature was plotted in a scatter plot– AdultWeekday ticket prices versus AdultWeekend ticket prices. This revealed weekend prices being higher than weekday prices seems to be restricted to resorts that charge less than $100 for daily lift tickets.

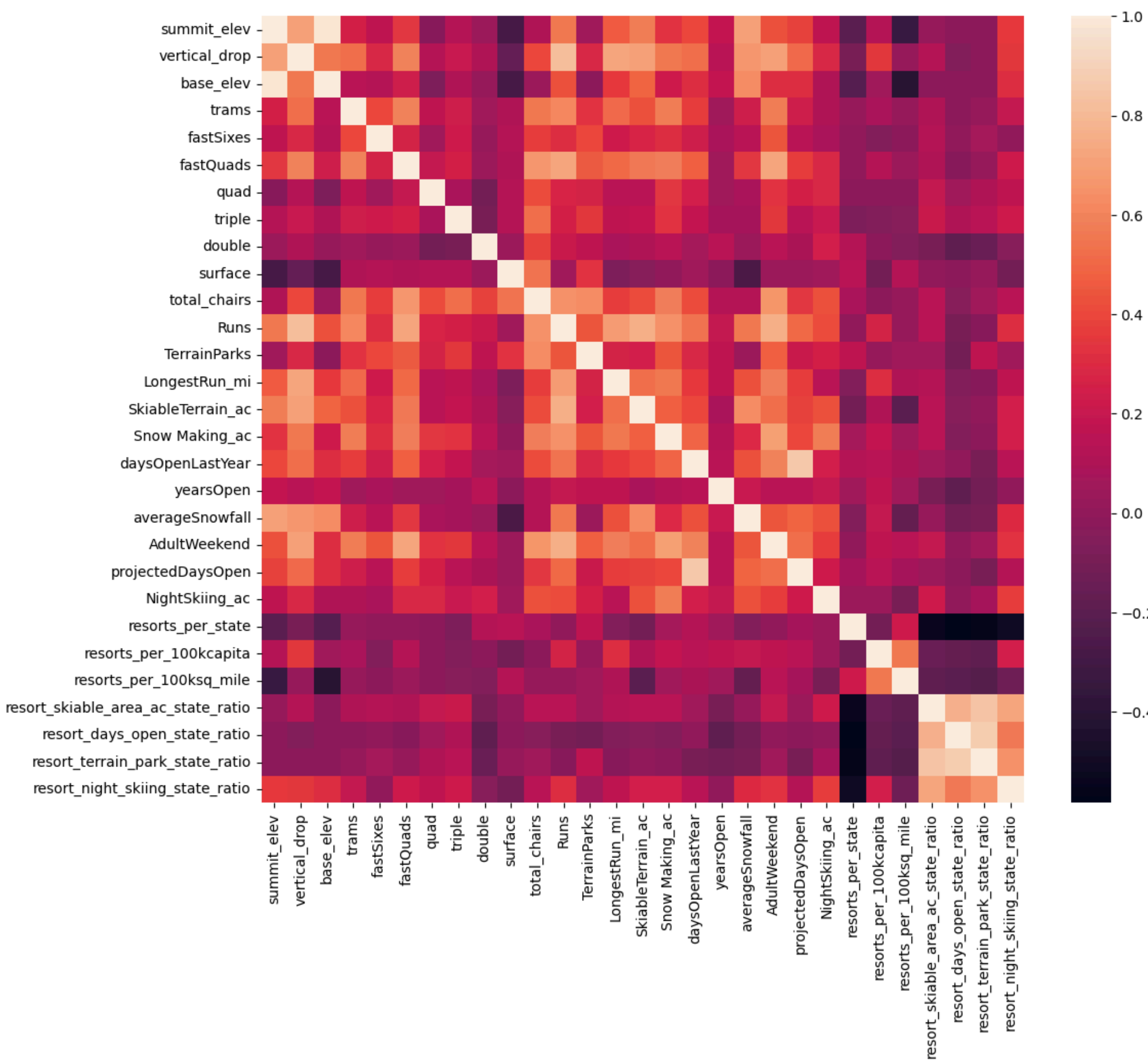
At the conclusion of the data wrangling process this dataset is left with 277 rows of data.



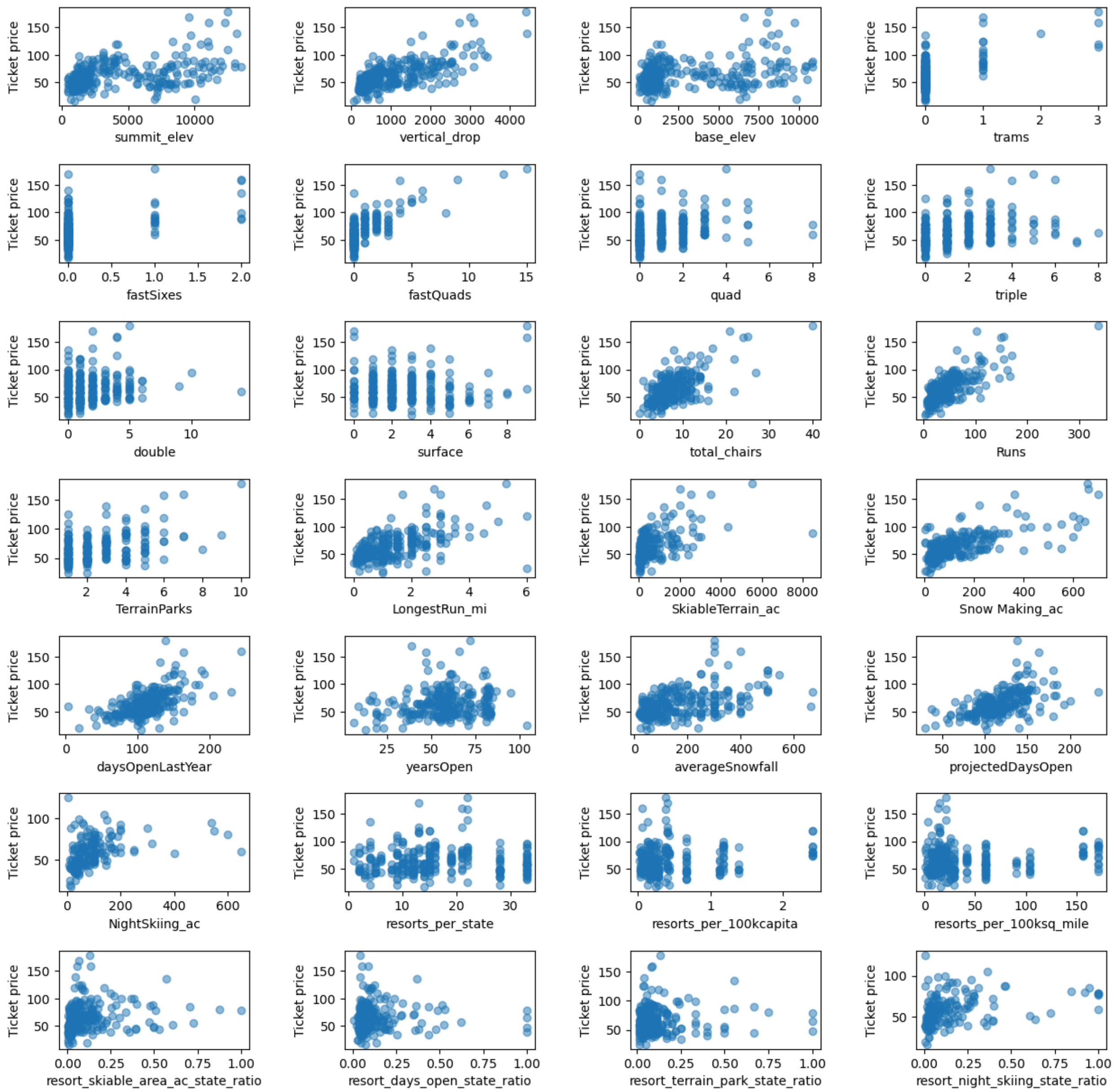
Exploratory Data Analysis

After plotting the Adult Weekend ticket prices by state, we see there is no obvious pattern that suggests a relationship between the two features. This indicates that we should build a pricing model that considers all states together, treating them all equally, since there are no clear groupings yet. Therefore focus remained on resort-level numerical data.

A heatmap was created in order to gain a high level view of relationships amongst the features. Focusing on the target feature, Adult Weekend ticket price, we see a number of correlations – vertical drop, # of fast quads, # of runs, total chairs, and night skiing acreage are most significant. These correlations suggest that more runs are well correlated with increased ticket price. One interesting thing to note here is the correlation with night skiing acreage.



Next, we created a series of scatterplots to further dive into the relationship between ticket prices and the other numeric features. The plots below appear to support the findings from the heatmap above. In addition, resorts per 100k capita also shows a strong correlation with ticket price.



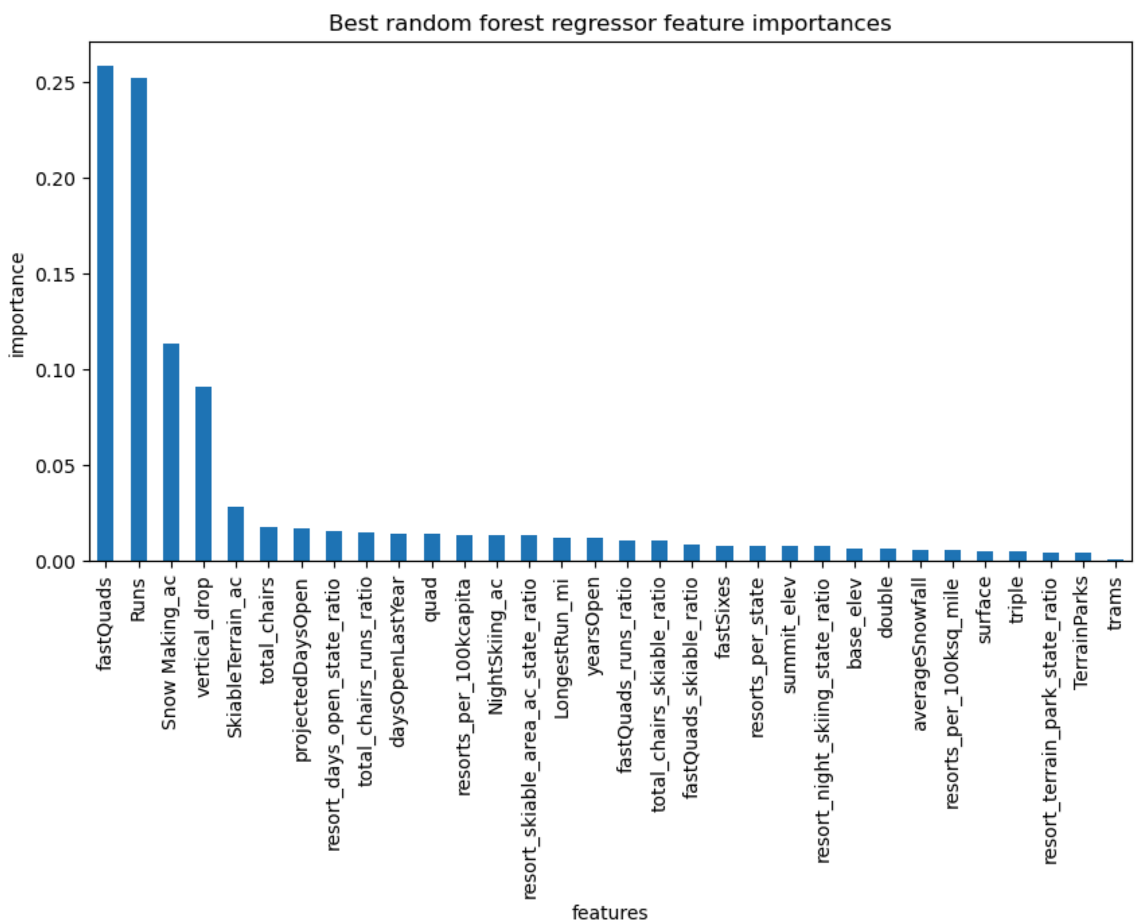
Model Pre-processing with Feature Engineering

To establish a baseline performance metric before delving into model training, we investigated the average target variable (price). To minimize bias during evaluation and validation, we first split our data into a 70/30 test/train set. This revealed an approximate average ticket price of $63.81. We then calculated the mean squared error (MSE) to be $19. Interestingly, fitting a simple dummy regressor to the training data resulted in a prediction that matched the average price to the eighth decimal place. This suggests a strong correlation between the features and the target variable

Algorithms used to build the model with evaluation metric
Two regression models were explored: a linear model and a random forest regressor.

Linear Model: The linear model achieved a mean squared error of ~$9 on the test set, suggesting some overfitting. Analyzing train/test set variance (~80% vs. 70%) confirms this. Cross-validation identified an optimal feature set size of 8. Key positive features included vertical_drop (consistent with EDA) and acreage covered by snowmaking (indicating a preference for guaranteed skiing). Interestingly, skiable terrain area displayed a negative correlation, potentially due to a constant number of lifts leading to thinner lift capacity across a larger area.

Random Forest Regressor: This model outperformed the linear model with a lower cross-validated mean absolute error by ~$1 and exhibited less variability. Feature importance analysis revealed a significant overlap with the linear model, with fastQuads.Runs, SnowMaking_ac, and vertical_drop being dominant factors.



3. Results and Findings

Model Selection

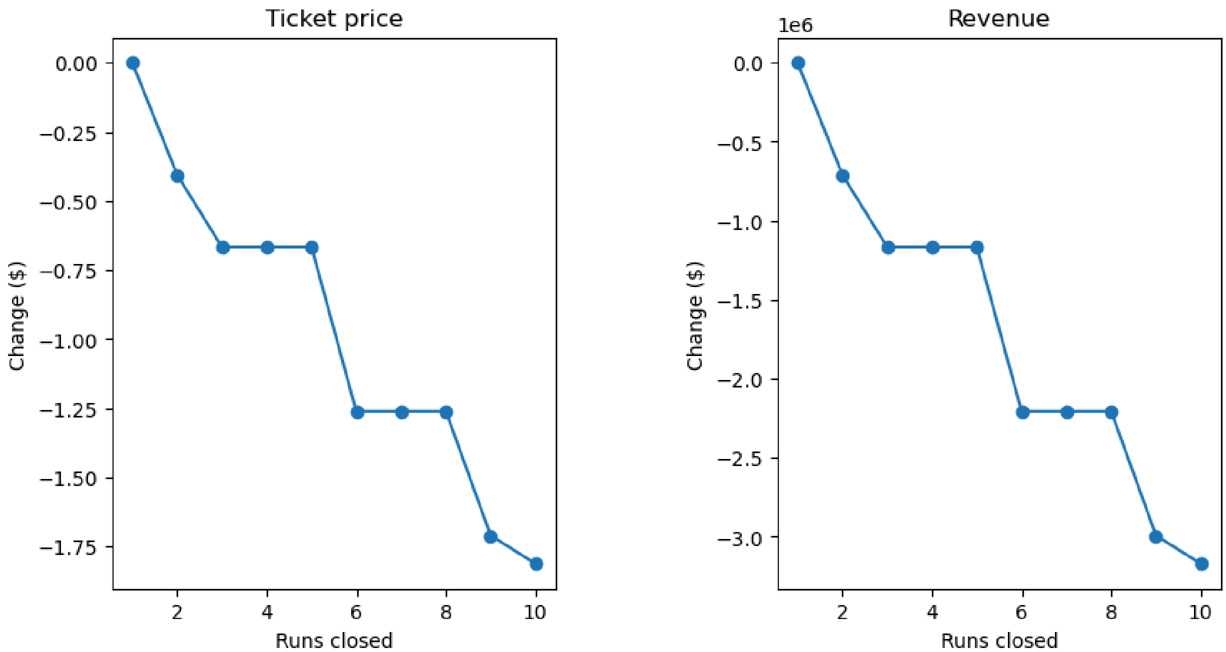
Based on these findings, the random forest regressor is recommended for moving forward. Its superior performance and lower variability make it a more reliable choice. Verifying performance on the test set confirmed consistency with cross-validation results. Additionally, the rapid improvement in model scores with increasing training set size suggests sufficient data is available. No further data collection is necessary.

Scenario Modeling

We then modeled 4 scenarios shortlisted by Big Mountain Resort for either cutting costs or increasing revenue (from ticket prices). 2 of the scenarios saw effects.

Scenario 1: Run Closure

Simulating the closure of the 10 least-used runs revealed minimal impact on ticket price until the closure of 4 or more runs. This suggests that focusing closure efforts on the least-used runs might not significantly affect revenue. However, more substantial closures (6 or more runs) would likely result in a noticeable drop in ticket price.



Scenario 2: New Ski Run and Snow Making Expansion
Modeling the addition of a new ski run with a 150-foot vertical drop increase (requiring a new chairlift) indicated a potential seasonal revenue gain of $3,474,638 due to the increased ticket price support of $2. Expanding the snowmaking area by 2 acres yielded similar projected revenue growth, suggesting minimal additional benefit from a small increase in snowmaking capacity.

4. Recommendations

Data suggests limited benefit from run closures: Closing a small number of runs (less than 4) has minimal impact on ticket prices. More significant closures (6 or more) likely decrease revenue.

New ski run and snowmaking expansion show potential: Adding a new ski run with increased vertical drop offers a projected seasonal revenue gain of $3,474,638 from the ability to support a $2 ticket price increase. Expanding the snowmaking area by 2 acres shows similar projected revenue growth.

Considering the potential revenue gains from the new ski run and snowmaking expansion, along with the limited impact of closing less-used runs, it seems more favorable to pursue the new ski run and snowmaking expansion with ticket price increase of $2.

5. **Conclusion**

Additional Considerations: Cost vs. Benefit: A crucial factor not mentioned in the data is the cost of building the new chairlift and expanding snowmaking. Leadership will need to weigh the projected revenue gain against the investment cost.

Long-term Impact: While both the new run and snowmaking could increase revenue, consider the long-term impact. Will the new run attract enough skiers over time to justify the cost?

Alternatives: Are there alternative ways to improve revenue other than change in ticket price that haven't been explored?

Next Steps/Future Scope of Work: While the data suggests new ski run and snowmaking expansion might be promising, leadership should conduct a full cost-benefit analysis before making a final decision. Additionally, exploring other revenue-generating strategies might be beneficial.