Recommendation Report for Big Mountain Resort

1. Problem Statement

Big Mountain Resort is a ski resort in Montana with access to 105 trails and is visited by about 350000 people annually. Big Mountain has recently installed an additional chair lift, increasing their operating costs by $1.5M. Historically, its pricing strategy was to set ticket prices just above the average market rate for comparable resorts. However, management suspects this simplistic approach does not accurately reflect how facilities (e.g., number of lifts, terrain parks, snowmaking acreage, and skiable acreage influence perceived value. The management is considering changes to either cut costs or support a higher ticket price. As a part of the data science team, the aim of this project is to implement a data driven prediction model to help the management make decisions.

2. Data Wrangling

Our dataset contained 330 US ski resorts, each described by ticket prices, facility counts, and terrain-related statistics. We calculated percentage of missing values for each feature, analyzed distribution of features by state, corrected some incorrect values, removed columns with large number of missing values and merged state-level data (population, total state area) to our data.
We finally decided to use the weekend price as our target variable because it had fewer missing values than the weekday price.

3. Exploratory Data Analysis

We grouped resorts by state and explored state wide summary data. We also calculated data for features per 100,000 population and per 100,00 sq. miles area for each state, and performed PCA and visualized high dimensional data. However, we couldn't conclusively say if state could truly influence the price on its own, so we decided to treat all state labels equally. After performing exploratory data analysis, we found that useful features having high positive correlation with the ticket price were `vertical_drop`, `fastQuads`, `Runs` and `total_chairs`.

4. Model Preprocessing

We excluded Big Mountain Resort and split the model into training(70%) and testing(30%) sets. Then, to set a baseline, we trained a DummyRegressor model always predicting the average ticket price, which gave us an R^2 of 0 and a high MAE of $19.14.
From there, we trained a Linear Regression model. We addressed missing values (through imputation) and scaled features, then used cross-validation for a better measure of generalization performance. Next, we trained and tuned a Random Forest model using similar preprocessing steps (imputation, scaling) and GridSearchCV to explore hyperparameters. Its cross-validation performance improved upon Linear Regression's results, and on the test set it achieved a lower MAE of about $9.54, showing a better overall fit.
Both models highlighted similar key features (`vertical_drop`, `Snow Making_ac`, `total_chairs`, `fastQuads, SkiableTerrian_ac` and `Runs`) as the most important features.

	CV MAE	Test MAE
Linear Reg	$10.50	$11.79
Random Forest	$9.64	$9.54

Based on these findings, we decided to move forward with Random Forest as our final model. It consistently performed well in cross-validation, generalized effectively to the test set, and provided better accuracy than Linear Regression.

We also evaluated whether collecting more data would improve model performance using learning curves. The results showed that the model's performance plateaued, suggesting that additional data is unlikely to yield significant gains.

5. Winning Model & Scenario Modeling

Having selected the Random Forest, we retrained it with all resorts except Big Mountain's row. This final model predicted Big Mountain's AdultWeekend price to be $95–$96, distinctly higher than the current $81. Even allowing for a $10 margin of error, the model strongly suggests a potential underpricing.

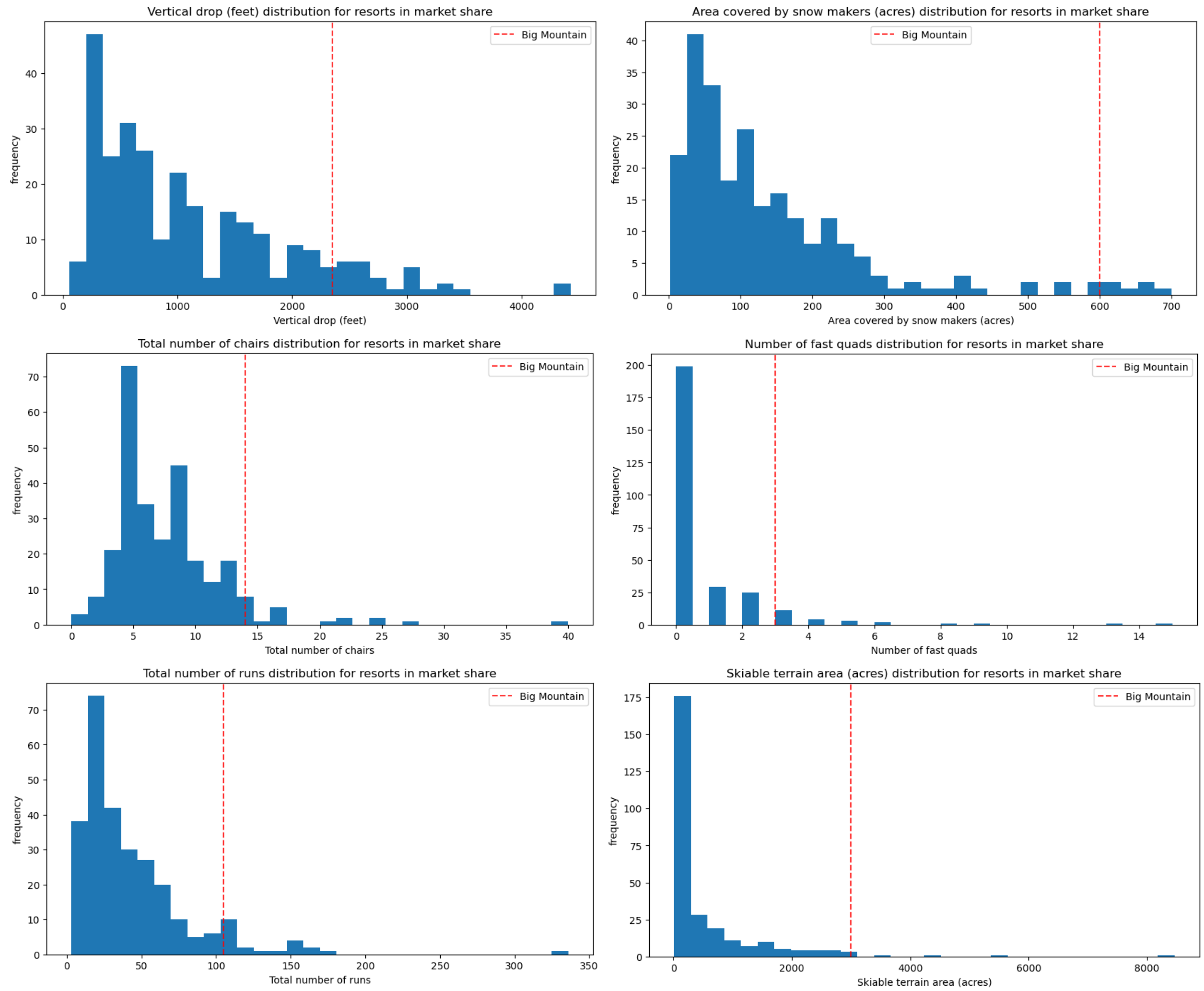
Using our Random Forest model, we analyzed the following scenarios:

1. Closing one run had no impact on ticket price, while closing 2 or 3 slightly reduced it. Closing 4 or 5 runs showed minimal effect, but closing 6 or more caused a significant drop, indicating a threshold where value perception is affected.
2. Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift suggested we could increase the price by $1.99, resulting in increased revenue of $3.5M.
3. Increasing the longest run by 0.2 miles and increasing area covered under snow making by 4 acres did not suggest any increase in price.

6. Pricing Recommendation

Analyzing the position of Big Mountain in the distributions of the most important features, we see that Big Mountain's facilities are among the best in the country, much better than most other ski resorts (see image below).

1. Our model suggests that management is currently undervaluing the ticket price and suggests raising it to $96 based on the current resort facilities. However, if executives are unwilling to increase price by $15, they should increase the price by at least $5 keeping the $10 margin of error of our model in mind.
2. Since they are installing a new chair lift, our model suggests that just adding a chair lift supports a price increase of just $0.29, but if it accompanied by an extra run and vertical drop increased by 150 feet, this would support a price increase of $2, leading to increased revenue by $3.5M, enough to sustain the cost of the chair and generate profits.
3. Our model has also predicted how much would be the dip in revenue if they closed down up to 10 runs. However, we currently do not have information about how much it costs to operate each run. Using the operation cost data, executives should evaluate closing down a few of the least used runs if the cost to operate them is much higher than the reduction in revenue.



7. Conclusion & Future Scope

Our model clearly suggests that ticket price for Big Mountain is undervalued, and executives should seriously consider raising the price of the ticket based on the high value of their facilities. While these insights are highly promising, they do not account for Big Mountain's detailed operating costs: costs of maintaining runs, lifts and chairs. We recommend collecting more granular data including visitor demographics, length of stay, and cost breakdowns. More data can help us refine this model further and to potentially explore dynamic pricing based on seasonal or real-time demand.