Report for Big Mountain Resort

Problem statement

The purpose of this data science project is to come up with a pricing model for ski resort tickets in our market segment. Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for. This project aims to build a predictive model for ticket prices based on several facilities, or properties, boasted by resorts. This model will be used to guide Big Mountain's pricing and future facility investment plans.

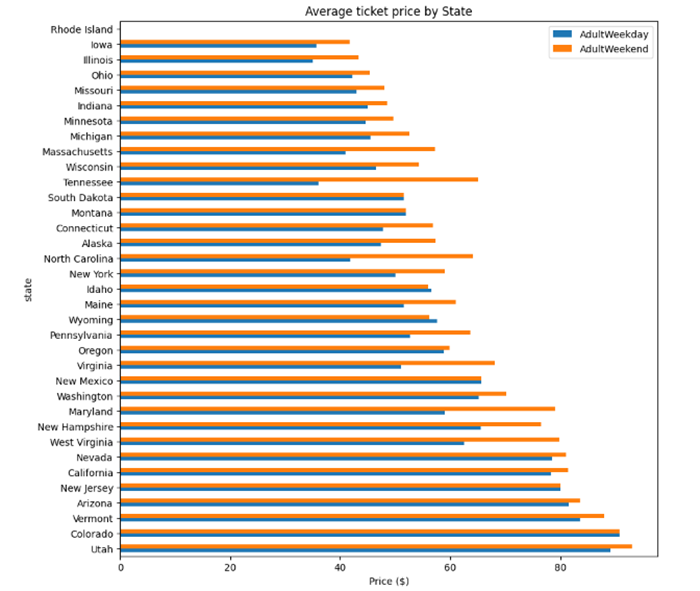
Data Wrangling

The dataset initially contained 330 entries including Big Mountain Resort among other resorts in different states. It also contained two types of pricing, one for the weekend and the other for the weekdays. After cleaning and removing unnecessary rows and columns, the dataset was reduced to 277 entries. This refined dataset is more focused and relevant for the goal of predicting ski resort ticket prices. Some actions taken in this stage are as follows:

|  |
| --- |
| **Data Cleaning and Feature Removal:** Columns like 'fastEight' with high missing values or little variance were dropped. data entry errors in 'yearsOpen' (values like 2019) and 'SkiableTerrain\_ac' were corrected. For 'SkiableTerrain\_ac', an anomalously high value for Silverton Mountain was adjusted from 26819 to 1819 acres based on external verification. The 'AdultWeekday' ticket price column was dropped due to higher missing values compared to 'AdultWeekend', making 'AdultWeekend' the primary target for the pricing model. |
| **Handling Missing Data:** Rows with no ticket pricing information were removed, as the target variable (ticket price) is essential for the model. Other missing values were left intact for further analysis, pending a deeper understanding of their impact on the modeling. |
| **State-wide Summary Statistics:** Additional features like the total skiable area, days open last year, and night skiing area by state were aggregated to enrich the dataset and provide a broader market context. |

A comparison of a number of states

Description automatically generatedA graph of different colored columns

Description automatically generated with medium confidence

A graph of blue dots

Description automatically generated

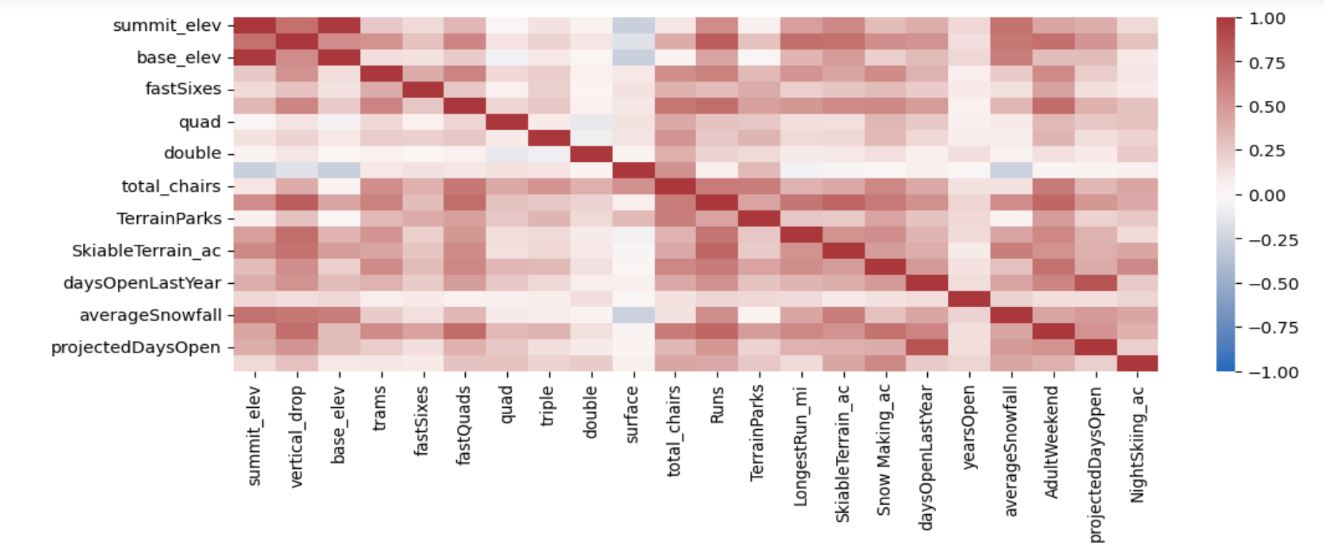
Exploratory Data Analysis

In this phase, we examined various features of the data set we built in the previous stage including total state area, population, number of resorts, skiable area, night skiing area, and days open, as well as resort density per capita and square mile. This exploration revealed diverse trends, with some states ranking higher in certain aspects than in others, and noticeable correlations between certain features. However, this initial analysis raised more questions than it answered. In this stage, we yield that:

1. There is no clear relationship between states and ticket prices.

A graph with colorful dots and text

Description automatically generated

2. The heatmap revealed strong correlations between ticket price and features such as fastQuads, runs, and snow-making capacity, with the resort\_night\_skiing\_state\_ratio appearing most correlated with ticket prices. Total chairs also showed a significant correlation with ticket prices.  
  


3. The ticket prices might initially decrease before increasing as the number of resorts per capita rises, indicating either popular skiing areas with high demand or less popular states for skiing. High ticket prices in areas with fewer resorts might reflect a monopoly effect in those regions.

4. We conducted scatterplot analyses excluding 'Name', 'Region', 'state', and 'AdultWeekend' to explore these correlations further. These plots confirmed strong positive correlations with vertical drop and fastQuads, among others. Scatterplots revealed complex relationships between ticket price and resort density measures, suggesting possible monopoly effects or exclusivity in certain states.

This analysis provided a comprehensive understanding of how various features, especially those derived from state data and resort-specific characteristics, could influence the adult weekend ticket price. These insights will guide the feature selection and modeling strategy, focusing on relative measures over absolute state data and considering resort-specific features that directly impact visitor experience and pricing.

Model Preprocessing with feature engineering,

Algorithms used to build the model with evaluation metrics

Winning model and scenario modeling

we created a baseline model using the average ticket price as a starting point for comparison, resulting in an R squared score of zero.

Initial linear models were created by imputing missing values (using median and mean) and scaling the data. The models were refined through feature selection, with SelectKBest used to identify the most significant features. Linear regression models showed improvement over the baseline, with the best model explaining over 70% of the variance in the test set.

* The result of linear regression suggests that vertical drop is our biggest positive feature.
* Also, We saw the area covered by snow-making equipment is a strong positive as well.
* The skiable terrain area is negatively associated with the ticket price.

Then we proceeded with building a Random Forest model. We initially used default parameters and then proceeded with hyperparameter tuning (particularly the number of trees) to optimize performance. The best Random Forest model was determined using cross-validation and feature importance was evaluated.

Encouragingly, the dominant top four features are in common with your linear model:

* fastQuads
* Runs
* Snow Making\_ac
* vertical\_drop

A graph with blue and white text

Description automatically generated

Finally, we compared our linear regression model and random forest model. Both models were evaluated using cross-validation to estimate their performance, with the Random Forest model showing slightly better performance and consistency.

The random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

Pricing recommendation

Big Mountain Resort's modeled price is $97.85, actual price is $81.00. Even with the expected mean absolute error of $10.17, this suggests there is room for an increase.

Eventually, we want to use our model to help us decide which proposed scenario supports Big Mountain’s goal of cutting costs or increasing revenue.

Senario1:

Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics. As shown below, the model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.

A graph of a ticket price

Description automatically generated with medium confidence

Scenario2:

In this scenario, Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. This scenario increases support for ticket price by $2.36. Over the season, this could be expected to amount to $4136364

Scenario3:

In this scenario, you are repeating the previous one but adding 2 acres of snowmaking. This scenario increases support for ticket prices by $2.36

Over the season, this could be expected to amount to $4136364

Such a small increase in the snow-making area makes no difference!

Scenario4:

This scenario calls for increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow-making capability. No difference whatsoever. Although the longest run feature was used in the linear model, the random forest model (the one we chose because of its better performance) only has the longest run way down in the feature importance list.

Conclusion

We concluded that Big Mountain is amongst the resorts with the largest amount of skiable terrain. The vast majority of resorts, such as Big Mountain, have no trams. Big Mountain has one of the longest runs. Although it is just over half the length of the longest, the longer ones are rare. Big Mountain compares well for the number of runs. There are some resorts with more, but not many. Most resorts have no fast quads. Big Mountain has 3, which puts it high up that league table. There are some values much higher, but they are rare. Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers. Big Mountain is very high up the league table of snow-making areas. Big Mountain is doing well for vertical drop, but there are still quite a few resorts with a greater drop.

Then, we explored various scenarios to achieve our main objective, which is to either reduce costs or boost revenue (through ticket prices).

Future scope of work

lacking broader financial information regarding operation cost and solily working with ticket price potentially will limit the analysis. Incorporating data on operating costs, maintenance expenses, and investment in facilities would provide a more comprehensive view for pricing strategies. The model's suggestion of a higher price than currently charged by Big Mountain could be due to a variety of factors, including market positioning, customer demographics, and possibly overlooked market forces. This discrepancy could indeed be surprising to business executives. Further investigation through market research, customer feedback analysis, and competitor strategy review would be necessary to understand this mismatch. Rather than relying on data scientists for every scenario analysis, potentially developing a user-friendly interface where business analysts can input different parameters and receive instant predictions would be more efficient. This tool would allow for agile decision-making and ongoing refinement of Big Mountain Resort's pricing strategy, making the model a dynamic and integral part of their business analysis toolkit.