Big Mountain Resort, a ski resort located in Montana, needs guidance on how to select a better value for their ticket price. Its current strategy is to charge a premium above the average price of resorts in its market segment. However, there are limitations to this approach, and there is also a possibility that Big Mountain is not capitalizing on its facilities as much as it could. The purpose of this project is to develop a predictive pricing model for ski resort tickets in Big Mountain’s market segment based on a number of facilities.

The dataset includes information for 330 total ski resorts, including Big Mountain. There are 27 columns in the dataset, including columns for the adult weekday and weekend ticket prices. However, there were a number of missing values, which led to one column being dropped completely, several data errors corrected, and some other rows dropped. Additionally, as there were more weekday prices missing, we kept only the rows with weekend price. Therefore, our model will be based on only adult weekend ticket prices. After cleaning the data, there were 277 rows and 25 columns. Additionally, we obtained additional US state population and size data to augment the cleaned ski data.

As part of the exploratory data analysis, we conducted a PCA transformation to derive features explaining the variance in the data and visualize the data in a lower dimension. We were able to determine that there is not a clear pattern between the state the resort is in and the adult weekend ticket price, as can be seen from the plot below. As such, we decided to treat all states equally.

![A picture containing chart

Description automatically generated]()

The next step was building machine learning models to predict the adult weekend ticket price. The baseline model, which simply predicts the ticket price as the average of all ticket prices, had an absolute mean error of $19. Next, we conducted a simple linear regression model using cross-validation and a random forest regression using cross-validation. The random forest model had a lower cross-validation mean absolute error ($9.60) by about $1 and exhibited less variability. Therefore, we decided to use the random forest regression model going forward.

We were also able to determine the most important features based on the random forest regression. The following plot shows that the top four features were fastQuads, Runs, Snow Making\_ac, and vertical\_drop.

![Chart, histogram

Description automatically generated]()

Big Mountain currently charges $81 for an adult weekend ticket. Using the random forest regression model, a ticket price that could be supported by Big Mountain’s facilities is $95.87. Even with the expected mean absolute error of $10.39, this suggests there is room for an increase in ticket price.

We then explored four different scenarios for increasing revenue from ticket prices and cutting costs. The first scenario was closing down up to 10 of the least used runs. Based on the plot below, we see that closing one run makes no difference but closing 2 and 3 successively reduces support for ticket price and so revenue. Closing down 4 or 5 results in no further loss in ticket price than closing 3. Closing down 6 or more leads to a large drop.

![Chart, line chart

Description automatically generated]()

The second scenario was increasing vertical drop by 150 ftby adding a run and an additional chair lift. This scenario increased support for ticket price by $1.99, which amounts to about $3.5 million in revenue, well over the additional operating costs of $1.5 million listed in the project overview. Scenario 3 was the same as Scenario 2 but adding 2 acres of snow making cover. The increase in snow making area made no difference. Scenario 4 was increasing the longest run by 0.2 miles and adding 4 acres of snow making cover. This made no difference as well.

Based on the four scenarios, I would recommend Scenario 2 to the business leadership, as it allows for an increase in ticket price and results in an increased revenue that is more than enough to cover the operating costs of the additional chairlift.

Note that this model relies on the implicit assumption that all other resorts are setting prices based on how much people value certain facilities. Additionally, other resorts could be mispricing tickets, which would skew the results.