Big Mountain Resort located in Montana offers spectacular views of Glacier National Park and Flathead National Forest with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. It can accommodate skiers and riders of all levels and abilities. Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season. The resort's pricing strategy has been to charge a premium above the average price of resorts in its market segment.

The problem is, how should Big Mountain Resort maximize capitalization on their facilities by increasing ticket prices above average market prices and cutting operational costs by at least 20% before the end of the year?

The key data source from Big Mountain Resort is a single csv file which has 330 rows. The data contained in this file was wrangled by:

1. Filtering missing data and dropping missing columns and rows like the fastEight column which had half of its values missing and half of its values as zero and 14% of columns missing information on the ticket prices for both Adultweekday and Adultweekend prices. Inspecting the rows also revealed a row in the yearsOpen column which had 2019 as the number of years open. The value of the row was out of range and impossible and moreover the actual time of data gathering was unknown so no important conclusion could be drawn for that value. Weekday prices rows had more missing values than weekend prices and these reasons these rows were dropped.
2. Correcting wrongly entered values like the skiable area value which led to further search on the websites of Silver Mountain resort to confirm the correct value. This value was replaced with the correct one.
3. Supplementing inadequate data with USA states population and area data retrieved from safe sources from the internet, specifically Wikipedia.

Ticket prices ranged from $25 to over $100 in USA states. In Montana where Big Mountain is, ticket prices of resorts show fairly small variability and weekend and weekday ticket prices are similar. The target feature to predict ticket price was therefore weekend ticket prices because it is missing less values. The current number of rows left after the data wrangling stage was 277.

The next stage was to perform Exploratory Data Analysis. At this stage features id

1. Numerical features in the data were summit\_elev, vertical\_drop, base\_elev, trams, fastSixes, FastQuads, quads, triple, double, surface, total\_chairs, Runs, TerrainParks, longestRun\_mi, SkiableTerrain\_ac, Snow\_making, daysOpenlast year, yearsopen, average snowfall, AdultWeekend prices, projected Days open and Nightskiing.
2. The categorical features were State, Region, Quartile and Name.

It was also determined that because categorical features were variable, observations led to unreliable conclusions, so they needed to be dropped. Using the feature correlation heatmap and scatterplots of numeric features against ticket price, vertical drop and the availability of fastquads were facilities that had most impact on ticket price increase and those were the target features for future modelling.

A blue dotted graph with white text

Description automatically generated with medium confidence A graph with blue dots

Description automatically generated

After performing EDA, the next stage was to build machine learning models to predict adult weekend ticket prices. The first model built served as a baseline performance comparator for all the subsequent models built. Data for only Big Mountain was extracted and data was partitioned data 70/30 training and testing splits, then the mean is confirmed as a good predictor using average price (randomly chosen feature). The average price was determined as 63.81108808 with sklearns Dummyregressor and the assessment of the mean using three metrics methods, namely: the coefficient of determination, R^2, the Mean Absolute Error (MAE), the Mean Squared Error (MSE) from sklearn’s metrics was used. The results from these metrics show that using the average price to predict ticket price, one may be off by about $19 from the actual price.

Building the initial model involved, imputing missing values, scaling the features, training a model and calculating model performance with metrics. The results from the first model built by imputing the median and later imputing the mean were the same. This model predicted the ticket price to be about $9 from the actual price which was an improvement from the initial $19. Building other models using algorithms such as sklearn pipeline and Random Forest model form sklearn's RandomForestRegressor led to the final selection of the model built with the latter. This is because hyperparameter search (GridSearchCV) allowed missing values to be imputed with both mean and median values at the same time and the one that yields the best results of the two was selected. RF also saved time by preventing checking performance on the test split. Calculating the random forest model performance showed lower cross-validation mean absolute error by almost $1. Verifying performance on the test set produces performance consistent with the cross-validation results with less variability. The four best features be seen in the picture below. A diagram of a computer

Description automatically generated with medium confidenceA graph with blue and black text

Description automatically generated

Assuming that all other resorts are largely setting prices based on how much people value certain facilities, the actual ticket price of Big Mountain can be leveraged it to gain some insights into what price Big Mountain's facilities might support as well as explore the sensitivity of changes to various resort parameters. Using our model the actual ticket price for Big Mountain was $81. The model also predicts the ideal ticket price to be $95.89 with MAE of 10.39 suggesting there is room for further increase.

Scenario modelling suggests that the focus of increasing revenue for the business must first consider the impact of closing some facilities to cut costs because if wrongfully done this can increase the woes of the business.eg. closing more than 5 runs will affect ticket prices. The current operational cost of the chair lift may not hurt the business if coupled with increasing the vertical drop by 150 feet because it supports an increase in ticket price by $1.99 on the basis that each visitor on average buys 5-day tickets. This could account for the seasonal cumulative sum of 3,474,638.

I recommend the closure of a few facilities in numbers that will not significantly impact ticket prices eg. about four runs while expanding useful facilities like increasing the vertical drop by about 150 feet in addition to the new chair lift. Additionally, advertisements on new and expanded facilities should be made on websites together with the adjusted ticket price so that visitors will focus on buying more tickets to experience them. Doing this will channel less focus on the closed down facilities and ticket prices will appreciate, nevertheless.

Modelling would have been more efficient if the additional operating cost of some of the other important features like the skiable terrain, snow making area and fastquads were also made available. More scenarios would have been modelled and assessed for relevance to ticket price increments.

Further work is required to first fill the missing data gaps and so that more scenarios can be modelled. Additional data may also require remodeling. When the best performing model is built, it will be integrated into business systems. The performance of the model will be monitored in real world scenarios. The model will undergo maintenance by retraining the model with fresh data to keep it up to date. Alerts will also be set up for anomalies and degradation. While model is in test, documentations of inputs, functionalities, outputs and limitations will be taken. The effective use of the model will be communicated to all stakeholders and business executives.