**Big Mountain Resort Report**

**Introduction:**

This report will describe the client and why the reason and purpose(s) behind this analysis, carefully specify the questions addressed by the analysis methods, describe the analysis techniques, describe all the instruments and tools used during the analysis, spell out the findings, and present recommendations based upon the results.

**Background:** Big Mountain Resort, the client, is a ski resort located in Montana that offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails that serves about 350,000 people every year. They have recently invested in their facilities which increased their operating costs by $1,540,000 this season.

**Current Strategy**: The resort is current pricing strategy is to charge a premium above the average price of Resorts in its market segment, and the client suspects they are not capitalizing on its facilities as much as possible.

**Problem Statement:** Are there any possibilities concerning Big Mountain Resort to create a more efficient pricing strategy that maximally capitalizes on their facilities and offsets their current operating cost?

**Instrumentation & Methods:**

**Pre-Processing and Training Data:**

* **Linear Model**: In building the linear model, we used median and mean values in place of missing values. Predicting ticket prices using the linear model would be off by about $9. However, the initial linear model was overfitting and needed to be adjusted by the number of features. Through cross-validation, the value of k was set to eight features to focus on: vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, Runs, LongestRun\_mi, trams, and SkiableTerrain\_ac. These features fit our initial assumptions from EDA.
* **Random Forest Model**: Like the linear model, we used median and mean values in place of missing values and was not helpful to scale the features. The random forest model revealed that the top four features to consider are fastQuads, Runs, Snow Making\_ac, and vertical\_drop.
* **Final Model Selection**: After testing both the linear model and random forest model, the project will be moving forward with the forest regression model. Comparison of the two demonstrated that performance on the test set was consistent cross-validation results. Additionally, the cross-validation means the absolute error was lower using the random forest regressor.

**Feature Engineering:**

**Feature Correlation heatmap**: A great way to gain a high level view of relationships amongst the features like the ratio of resorts serving a given population and the ratio of resorts in a given area, the number of resorts per state, total skiable area, and days of skiing. Relationships we found were:

* summit and base elevation are quite highly correlated.
* We also see that if you increase the number of resorts in a state, the share of all the other state features will drop for each.
* When resorts are more densely located with population, more night skiing is provided.
* AdultWeekend ticket pric has a few reasonable correlations. fastQuads stands out, along with Runs and Snow Making\_ac.
* Visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up.
* As well as Runs, total\_chairs is quite well correlated with ticket price. This is plausible; the more runs you have, the more chairs you'd need to ferry people to them! Interestingly, they may count for more than the total skiable terrain area. For sure, the total skiable terrain area is not as useful as the area with snow making. People seem to put more value in guaranteed snow cover rather than more variable terrain area.

Chart

Description automatically generated

* **Feature Correlation Scatterplots**: Correlations, particularly viewing them together as a heatmap, can be a great first pass at identifying patterns. But correlation can mask relationships between two variables
  + There's a strong positive correlation with vertical\_drop.
  + fastQuads seems very useful. Runs and total\_chairs appear quite similar and also useful.

A picture containing diagram

Description automatically generated

**Limitations:**

**The main data source is missing some important information like:**

**1. Weekdays ticket price**

**2. Operating costs for most of the resorts features**

**3. exact number of visitors per year**

**Result:**

**Potential scenarios:**

Big Mountain Resort has been reviewing potential scenarios for either cutting costs or increasing revenue (from ticket prices). Ticket price is not determined by any set of parameters; the resort is free to set whatever price it likes. However, the resort operates within a market where people pay more for certain facilities, and less for others. Being able to sense how facilities support a given ticket price is valuable business intelligence. This is where the utility of our model comes in.

The business has shortlisted some options:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
3. Same as number 2, but adding 2 acres of snow making cover
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days. Assume the provided data includes the additional lift that Big Mountain recently installed.

Our Model suggests that Mountain Resort’s ticket price is lower than the predicted model by $10.39 and the resort have many potential scenarios for either cutting costs by closing runs or increasing ticket price by increasing vertical drop, adding acres snow making or increasing the longest run.

* When it comes to closing up to 10 used Runs, our Model predicted the following:
  + The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue from between ~450,000 to ~700,000. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 loosing ~750,000 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop ~1.5 millon.
* Increasing the vertical drop by 150 ft would increase the ticket price by $1.99 resulting in a revenue increase of $3,474,638
* Adding 2 acres of snow making would result ticket price staying at $1.99 resulting in a revenue increase of $3,474,638 which is such a small increase in the snow making area and makes no difference.
* When it comes to closing increasing run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres there is 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 longest run way down in the feature importance list.

Chart, histogram

Description automatically generatedChart, line chart

Description automatically generated

**Recommendations:**

After applying our Model for ski resort ticket price and leverage it to explore Big Mountain Resort’s potential scenarios for increasing revenue, we can conclude that:

* The best scenario where we managed to gain the highest revenue increase possible was by increasing the vertical drop by 150 ft, this scenario has increased ticket price by $1.99 from $81 to $82.99, resulting in a bottom-line increase by $1,934,638 (After deducting operating costs = $1.54M).
* Due to lack of data in regards of operating cost per used run and weekdays ticket price, our model cannot recommend closing down used runs or implementing a dynamic ticket pricing.
* Cannot recommend more chairs because the effects may be attributed to exclusive vs. mass market resort effect and the data does not have the number of visitors per year.