Problem Statement

Big Mountain Resort (BMR) has recently invested $1,540,000 for an additional lift chair. The problem to address in this document is how we can recoup this investment. Two avenues explored are:

1. Increasing our ticket prices for the season
2. Introducing some changes to cut costs elsewhere

After running a comparative study between states, as well as within our target state of Montana, we **recommend** a price increase of **$1.99** (from current $81 to $82.99) while **adding an additional run** and **increasing the vertical drop by 150 feet**.

Data Wrangling

We use two data sources for this project:

i) Country-wide resort data of size 330 by 27, meaning we have 330 entries for various resorts within the country and 26 features for each of them.
ii) State population data pulled from the web

Starting with the first dataset, we did some basic cleaning; i.e. remove null values, duplicates, and outliers. We also dropped 2 columns/features due to significant size of missing data or their irrelevance. Finally, we added extra information about states, i.e. their area and population, from the second dataset to the first one.

This left us with 277 rows and 32 columns (or 29 features in total __ 3 columns are categorical with state, region and resort names).

Exploratory Data Analysis

To begin with, and assuming state data can give us some direction, we ran comparative analyses across states. However, they do not show a clear pattern that can be used to deduce a reasonable ticket price based on prices in other states. More specifically, while we see that Montana is the 3rd biggest state in term of area, it falls in the 12th place in terms of number of resorts.

We calculated resort density per capita and per square miles by dividing the number of resorts by the state population and area, respectively. On the one hand, Montana is 4th in resort density per 100K capita. But it is not in top 20 states when it comes to resort density per 100K square miles.

This told us we need to consider other factors. For that reason, we created a new dataset by keeping the density information from the second dataset and adding information pertaining states from the first dataset; i.e. *TerrainParks*, *SkiableTerrain_ac*, *daysOpenLastYear*, and *NightSkiing_ac*.

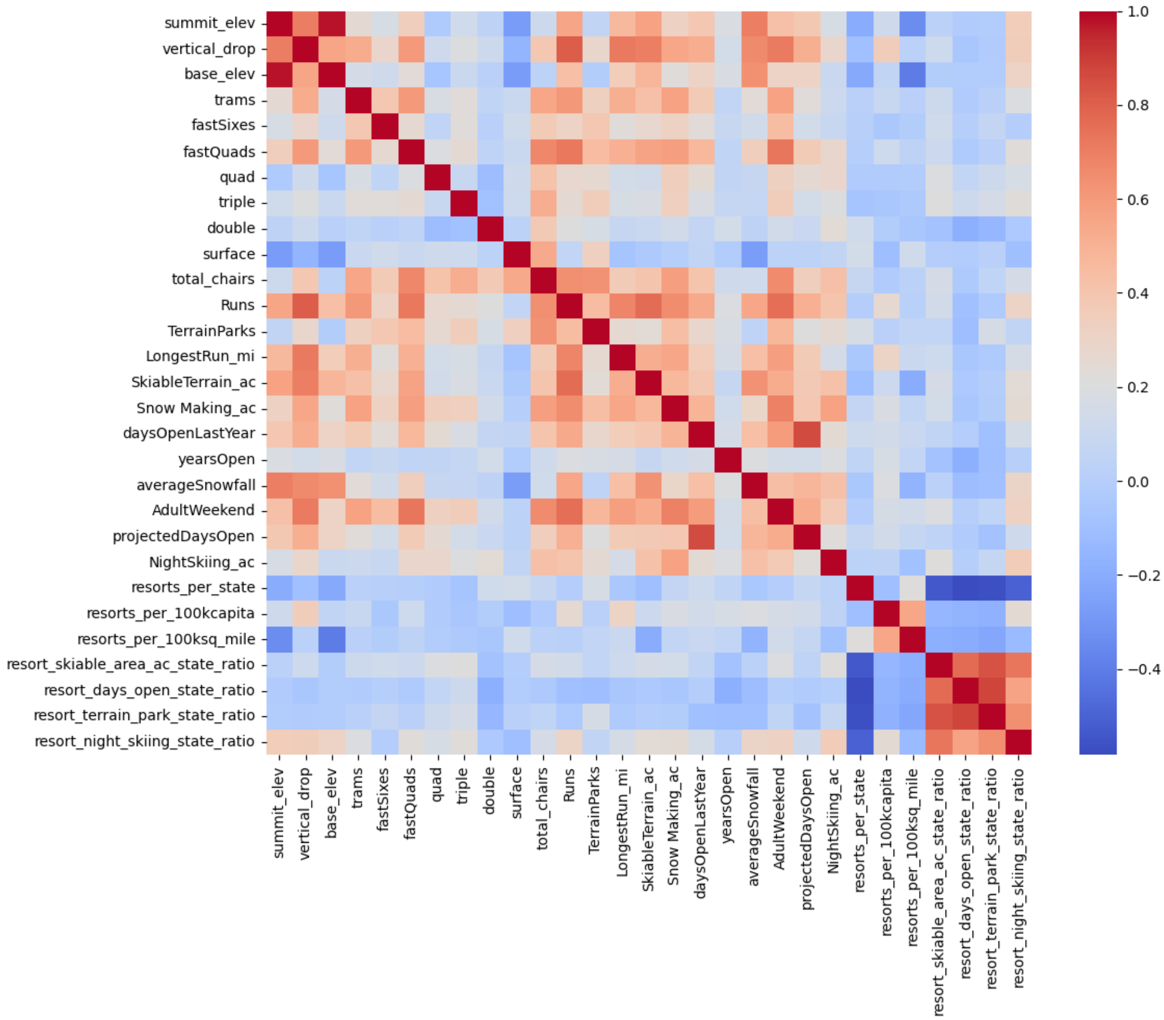
Model Preprocessing with feature engineering

In order to do modeling, we first did dimensionality reduction using PCA (Principle Component Analysis) method on our 7 numerical features, on the custom dataset we created. No clear pattern was observed between ticket prices and states on this custom dataset either.

This led us to conclude that ticket prices are independent of states and that we can treat all states equally. Therefore, we returned to our original data and doing some feature engineering, we calculated 4 ratios:

i) ratio of resort skiable area to total state skiable area,
ii) ratio of resort days open to total state days open,
iii) ratio of resort terrain park count to total state terrain park count, and
iv) ratio of resort night skiing area to total state night skiing area

Plotting a heatmap (figure below), as well as scatter plots for the adult weekend ticket price as opposed to all other columns showed some strong correlation between the prices and features, such as *vertical_drop*, *total_chairs*, and *longest_run*.



Algorithms used to build the model with evaluation metric

After splitting the data to training and test sets, including 70% and 30% of the data respectively, we ran baseline models using mean and median as our predictive methods. The model with median method had a smaller MAE (Mean Absolute Error) compared to the one using mean ($9 to $19). So, we decided to use median as our method for the actual model.

The first model we tried was a **Linear Regression (LR)** model, which did not show much difference from the baseline model based on median.

We started with 29 features, but it is not clear we needed them all. Therefore, we ran some cross_validation to find the optimal number of features as well as what they are. The results showed that the optimal number of features to consider is 8. Among these, *vertical_drop*, *Snow Making_ac*, *total_chairs* and *fastQuads* were the top 4 relevant features.

We then validated this finding by using **Random Forest (RF)** model. This model concluded with *fastQuads*, *Runs*, *Snow Making_ac*, and *vertical_drop* as 4 top features.

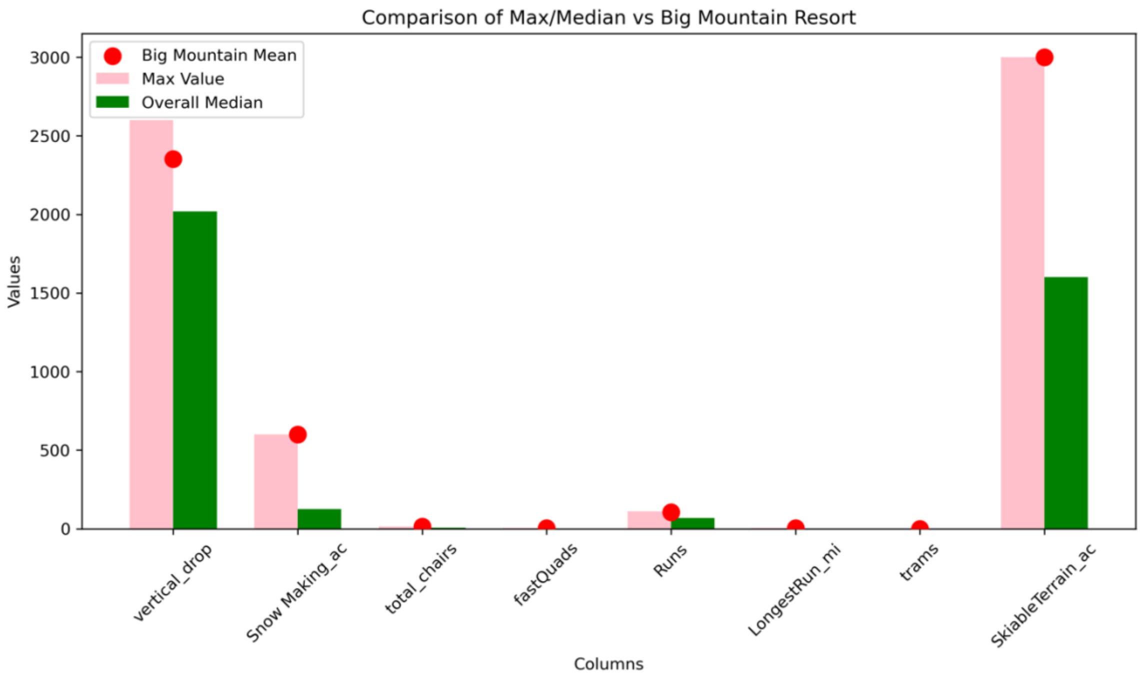
Winning model and scenario modelling

In order to make a selection between LR and RF, we checked the MAE for these two models and with RF showing a lower cross-validation MAE by almost $1, it seemed to be a better fit for our dataset.

The current ticket price for adults on weekends in the resort is $81.0.

Our random forest model predicts an increase of over $15 and up to $95.87 with a mean absolute error of $10.39 for our new ticket prices is justified.

Comparing BMR's facilities to other resorts in Montana, we see that it generally sits at the high end of providing various services, with the exception of *vertical drop*, for which it is outnumbered by other facilities. However, given the services provided and the margin of error ($10.39), there is room for ticket price increase.



This could go hand-in-hand or in addition to cost cutting strategies explored here. We explored:

1. **Closing Runs:** Closing up to 10 of the least used runs. The analysis shows that closing down runs does not support increase in ticket prices. On the contrary, closing more than 1 run results in a support for decreasing ticket prices.

2. **Adding Runs, Increasing Vertical Drop by 150 feet and Adding a Chait Lift:** This strategy supports increasing ticket prices by $1.99, which could result in $3,474,638 revenue increase over the season.

3. **#2 + Adding 2 Acres of Snow Making:** This shows no additional gain compared to #2.

4. **Increasing the Longest Run by .2 miles + adding 4 acres of snow making capability:** Our modeling predicts that this strategy will have no effect on the ticket pricing.

Pricing recommendation

Based on the analysis done, scenario #2 is suggested. We recommend adding 1 run, increasing vertical drop by 150 feet and adding one lift chair. Following these strategies, we recommend increasing the ticket price by $1.99 to $82.99 for this season

Conclusion

We started with an issue of having to balance out our $1,540,000 investment in a new lift chair. While our cross-state comparative analysis did not bear reliable results, deriving from ticket prices versus services and amenities within state, and using Random Forest model as the one fitting better with our data, we conclude that we can safely increase the weekend ticket prices by $1.99, while adding 1 additional run and increasing the vertical drop by 150 feet.

Future scope of work

The modeling and analysis done here has been based on limited available data. For instance, the number of visitors per skiing season could play a role in our predictions, but is missing from the current data.

- The current available data also was not unproblematic. There were missing values for ticket prices for various resorts, and we needed to impute such data using mean and/or median. But, these are just best guesses and might not reflect reality.

- Another limitation or concern is the fact that we are going through this practice for our resort based on the data from previous year. What if other resorts are going through the same practice now and might increase their ticket prices in the upcoming season?

- On the other hand, even though the suggestion here is to add 1 run and increase the vertical drop, we have no data to show us how much they will cost and what the trade-off is. We know that adding one chair costs $1,540,000. So, with our prediction of $3,474,638 increase in revenue, we are looking at net profit of $1,934,638. But, is that really so? We need to deduct the costs for adding a run and increasing vertical drop to our expenditure number.

- One surprising observation is that the current ticket price for BMR is the highest among the resorts in Montana. It is true that it offers great services when it comes to the 8 most relevant features our model predicted. However, a predicted price of $95+ seems too ambitious given these facts. This begs questions like: "Why is the current price the highest in the state? Is there useful information we are missing here that can shed light on it?"

There are still (combination of) scenarios that we have not explored here. In order to make the analysis scalable, we need to deploy our model on a company owned platform, with enough information as how to run it with minimal help, like a README file. A better strategy would be to make it streamlined by writing scripts to combine multiple steps and minimalize the burden on other users. For instance, to reduce it to one command, which asks for input parameters, and in the background, runs the model, replaces our parameters with the user's parameters and outputs predictions.