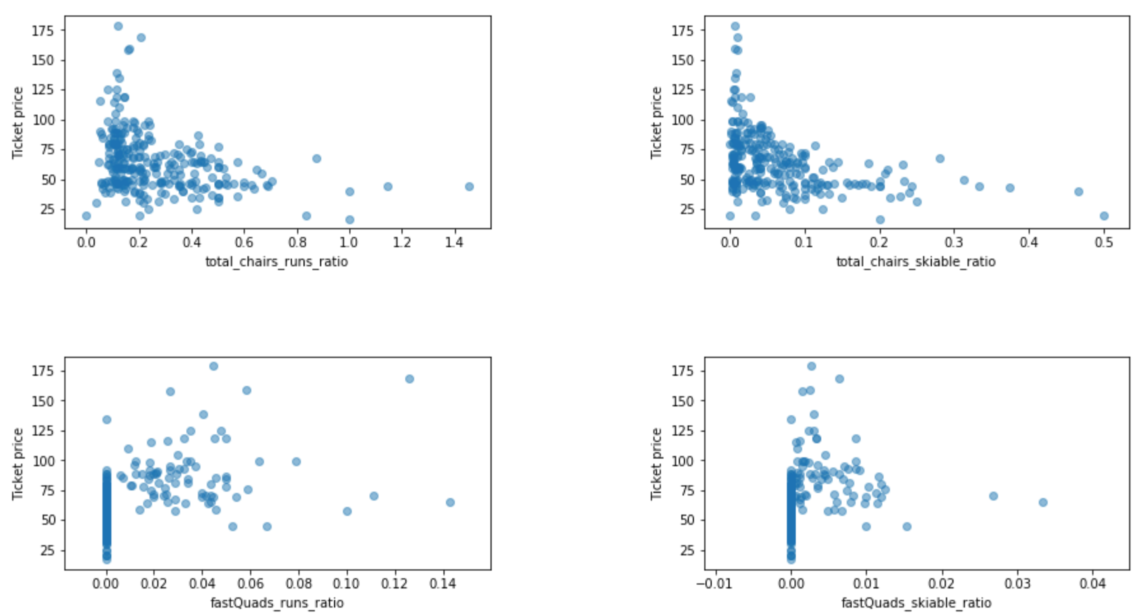
The Project Summary

The Big Mountian ski resort in Montana offers various facilities to skiers or riders of all levels. The ticket prices are set based on the market average plus a premium, but this may not be a fair calculation. Furthermore, the business suspects the prices could undervalue some of their facilities. Therefore, the resort would like to identify a suitable value for the ticket prices based on the services they provide compared to others so that they can change the ticket prices or cut operational costs. Therefore, the problem we are trying to solve is: Can we increase the weekend ticket prices by 20% to increase the revenue by the next quarter?

The database manager provided a CSV file with 330 rows and 27 columns. The first step of this work was cleaning the data set. We noticed that data related to the Big Mountain resort were all present without missing values. First, we counted the missing values for each feature and noticed that 50% of fastEight data was missing. We ended up dropping the entire column. We also saw that weekday and weekend ticket prices are similar and are between $25 and $100, except for some outliers, such as California and Utah. Montana, on the other hand, shows little to no variability in ticket prices.

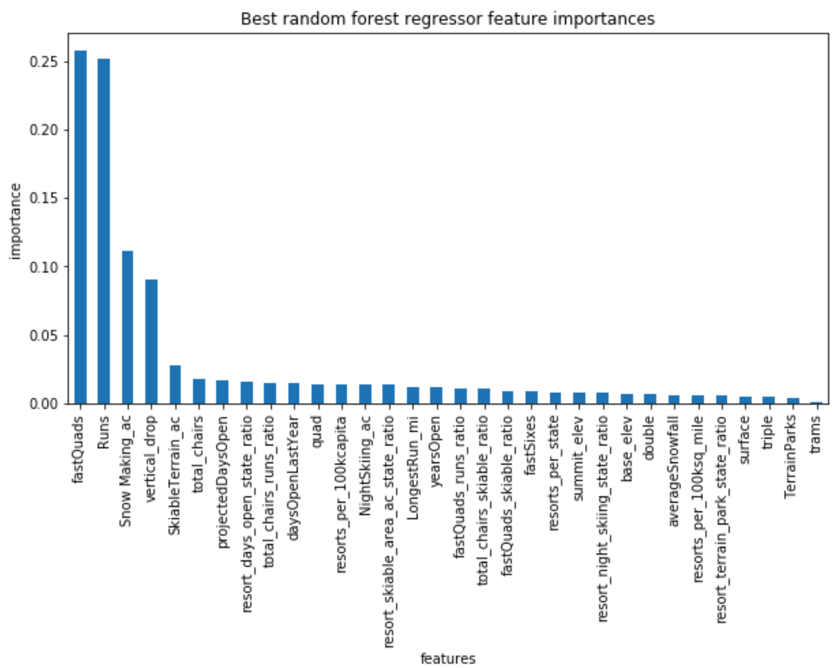
Furthermore, we found that skiable terrain and snowmaking data had unusual values for certain resorts, and we corrected them using the information on the Internet. Fast sixes and tram features were excluded from further exploration as we do not intend to use them in our analysis. 14% of the rows have no price data at all, so we dropped all these as well. Next, we used Wikipedia to create another data file with states and their populations to supplement future analysis. Since weekend and weekday prices are similar and weekday prices have more missing values, we dropped the weekday prices column. Additionally, we dropped the rows with no weekend prices. Finally, the data set came to 277 rows and 25 columns. The adult weekend ticket price was selected as the target feature. We also obtain a new data set from Wikipedia containing information about each state and its population.

During our exploratory data analysis step, we performed a principal component analysis. Through this analysis, we conclude that all states can be treated equally. Then, we calculated the correlation coefficients among all features. Focusing on the target feature, the adult weekend price, we saw positive correlations with fast quads, vertical drop, runs, total chairs, and snowmaking capacity. We made scatterplots (figure 1) with some of these features.



Notice a strong positive correlation between vertical drop, fast quads, and ticket prices. Runs and total chairs display some correlation as well.

The next stage of the project was preprocessing and training the data. First, we deleted the data related to our ski resort to exclude bias. The data set was split into 70% and 30% for training and testing. Three models were trained and tested for our problem: A baseline model that just predicts the average ticket price, a linear regressor, and a random forest regressor. We imputed the missing values with the median of the relevant data as the median is a better statistical representation of some skewed data. Using measurements such as mean absolute error and R- squared, we noticed that both regressors perform better than the baseline model, and the random forest regressor slightly outperforms the linear regressor. Both linear and random forest models confirmed that the most important features are fast quads, runs, snow-making capacity, and vertical drop. Since we selected a random forest model to go forward in our analysis, here's how the model ranked the features based on importance.



Currently, the resort charges $81 per adult weekend ticket, and our modeling predicted the price could be roughly $94. Considering our mean absolute error is $10, we have room to increase the ticket price. The resort's current ticket price sits on the higher end of the ticket prices for resorts in Montana. However, the resort is among the largest snow-making areas with the total number of chairs, fast quads, number of runs, the longest run, and skiable terrain. The resort is also close to the higher end of resorts with high vertical drops. Therefore, we recommend increasing the ticket price as the analysis shows we are undervaluing our facilities.

Next, we explored possibilities related to cutting operational costs or extending facilities. The assumptions we made for this study were that the expected number of visitors during the season was 350000, and each person bought five tickets. First we explored permanently closing down up to 10 least-used runs. Through our model, we noticed while closing one run doesn't make a difference, closing 2 and 3 runs reduces the ticket price and revenue. Furthermore, closing six or more runs would lead to a significant revenue drop. Then, we looked at the possibility of increasing the vertical drop, which requires an additional chair lift. This would support an increase of the ticket by roughly $2, which could result in an extra $3474638. However, this number must be compared with the additional maintenance cost of the chair lift. Next, we looked at adding 2 acres of snow to the previous changes and found no difference. Finally, we explored the effect of increasing the longest run by 0.2 miles, which also requires adding 4 acres of snow. This change would also have no impact on the price or revenue.

The modeling supports a ticket price increase of $3 (considering mean absolute error). We recommend increasing the vertical drop and adding a chair lift to support an additional ticket price hike of $2. However, further analysis of this recommendation is required with data such as the number of visitors, ticket sales, and maintenance cost increase for the abovementioned changes. From further analysis, if the maintenance cost of a run is significant, we could also remove 1 or 2 runs with minimal or no change to the revenue.