**Big Mountain Resort Report**

Big Mountain Resort is a ski resort located in Montana that has access to 105 trails and an annual customer base of around 350,000 people. The 11 lifts, 2 T-bars, and 1 magic carpet comprise the resort’s facilities. The resort has also recently installed a new chair lift that added around $1,540,000 in additional expenses per season. The company suspects that they are not capitalizing enough on their facilities as they should, which is why this modeling project looks at selecting a better ticket price for Big Mountain.

Basing the model of the data in the CSV file received form the Database Manager, the problem statement this model tries to combat has to do with the opportunities that exist for Big Mountain to increase their annual revenue by ~20% within the next financial year through selecting a better ticket value for their ticket prices by comparing market comparable facilities.

The data that was read in from the csv file was the ski resort data consisting of the comparable ski resorts. Inspecting the dataframe led to the removal of several null or irrelevant values that made columns and rows useless for further analysis. After plotting the weekday and weekend prices on a scatter plot, it was very clear that there is a line where prices between the two are equal. Because of this obvious relationship, the weekday price column was removed due to having more null values than the weekend price column. Because of this, the rows for the resorts that had missing weekend price values were further rendered useless to analysis and removed.

The numerical features of the data are any of the columns that have to do with quantifying an amount or distance with rows such as TerrainParks and SkiableTerrain\_ac respectively. The categorical features of the data are rows that display qualitative data such as labels, with 'state' being the perfect example in this dataframe. When analyzing the relationship between state and ticket price after calculating the mean weekend ticket price for each state, it became clear through principal component analysis that certain variables had more influence over the deviation of certain states away from the clustered PC values. After further exploratory analysis and heat mapping (Figure 1) the numerical variables from ski\_data, it became visibly simple to observe the correlation of these values between each other. Paying special attention to the correlations with the weekend ticket prices, it became clear that the variables with which AdultWeekend had the most significant absolute correlation values are features such as Runs, surface, fastQuads, trams, and vertical\_drop. What this means in terms of moving forward into feature selection for modeling is that now there is a better picture for which variables are the most influential on ski resort ticket prices. The target feature for the modeling will be the AdultWeekend feature since the significantly correlated features can be used to justify another AdultWeekend value for Big Mountain Resort in Montana.

By starting the preprocessing stage by splitting the data 70/30 training and testing respectively, and taking the mean average price as the predictor then calculating the mean absolute error, a baseline idea of performance was given with a value of around $19. Afterwards, a linear model was built to compare the results of the R-squared calculation between mean and median linear models. This showed that the linear model that used the median average had an R-squared value of 0.8178 for the training set and 0.7209. These values give the idea that there may be overfitting and this was expected, and using the median confirmed that the direction of the preprocessing is moving towards explaining more variance of the ticket price. Random forest regression was extremely important for identifying the best regressor feature importances. From Figure 2, it is visible that the top four dominant features that explain target feature variance are in line with the linear regression model. These four features are fastQuads, Runs, Snow Making\_ac, and vertical\_drop. Building on the cross-validation results, the random forest model has a lower mean absolute error for its cross-validation by almost a dollar (9.645 compared to 10.499), and a lower variability. The model that was used going forwards is the random forest model due to its more accurate predictive ability from validating the performance on the test set and getting an MAE (9.538) consistent with cross-validation.

Big Mountain currently charges $81 as their ticket price. The random forest regression model gave a modeled price of $95.87. This gives a general idea of how much further the ticket price can be supported in the marketplace by the facilities. With an expected mean absolute error of $10.39, the lower bound for the ticket price that falls within the model's MAE purview would be $85.48. Suggesting a ticket price of 89.99 is a good example of selecting a ticket price supported by the model that is lower than the model price but substantially higher than their current price. The additional operating cost of the new chair lift also should ideally be covered by the increased price. If the new chair lift added 1,540,000 of operating costs during the season, the difference between the suggested ticket price of 89.99 and the current price of 81 (8.99) multiplied by the average number of day tickets bought per annum should be greater than the additional operating costs. Every year about 350,000 people go to Big Mountain to snowboard or ski and on average, each visitor buys 5 day tickets. The suggested ticket price would increase revenues by 15,732,000 which would absorb the 1,540,000 of expenses from the new chair lift, putting the business at a more profitable position. It is important to note that estimating the revenue while generalizing the 350,000 yearly visitors as adult ticket prices means that there is likely overestimation. Big Mountain also put forth four potential scenarios. These scenarios were individually modeled to establish a predicted change in supported ticket price. In scenario one, which is closing up to 10 of the least used runs, creating two plots (Figure 3) to predict the change (delta) in ticket price and revenue. These plots uncovered that it would make no difference if one run were closed, and also conveyed that there would be no difference between closing 3 and 5 runs, as well as 6 and 8 runs. This means that if Big Mountain would like to reduce their expenses without sacrificing ***any*** support for ticket price, they should close their least used run. As for the second scenario, adding a run, increasing the vertical drop, and installing another chairlift, using a python function gave us an increased ticket value by $1.99. Over the season, this could be expected to amount to $3,474,638. If Big Mountain goes this route they should look closely at whether the benefitted ticket value will support the additional recurring expenses resulting from the added facilities. For the other two scenarios, no increase in price was predicted therefore it would not be in the company's best interest to add snow making or lengthen runs.

**Appendix**

Figure 1

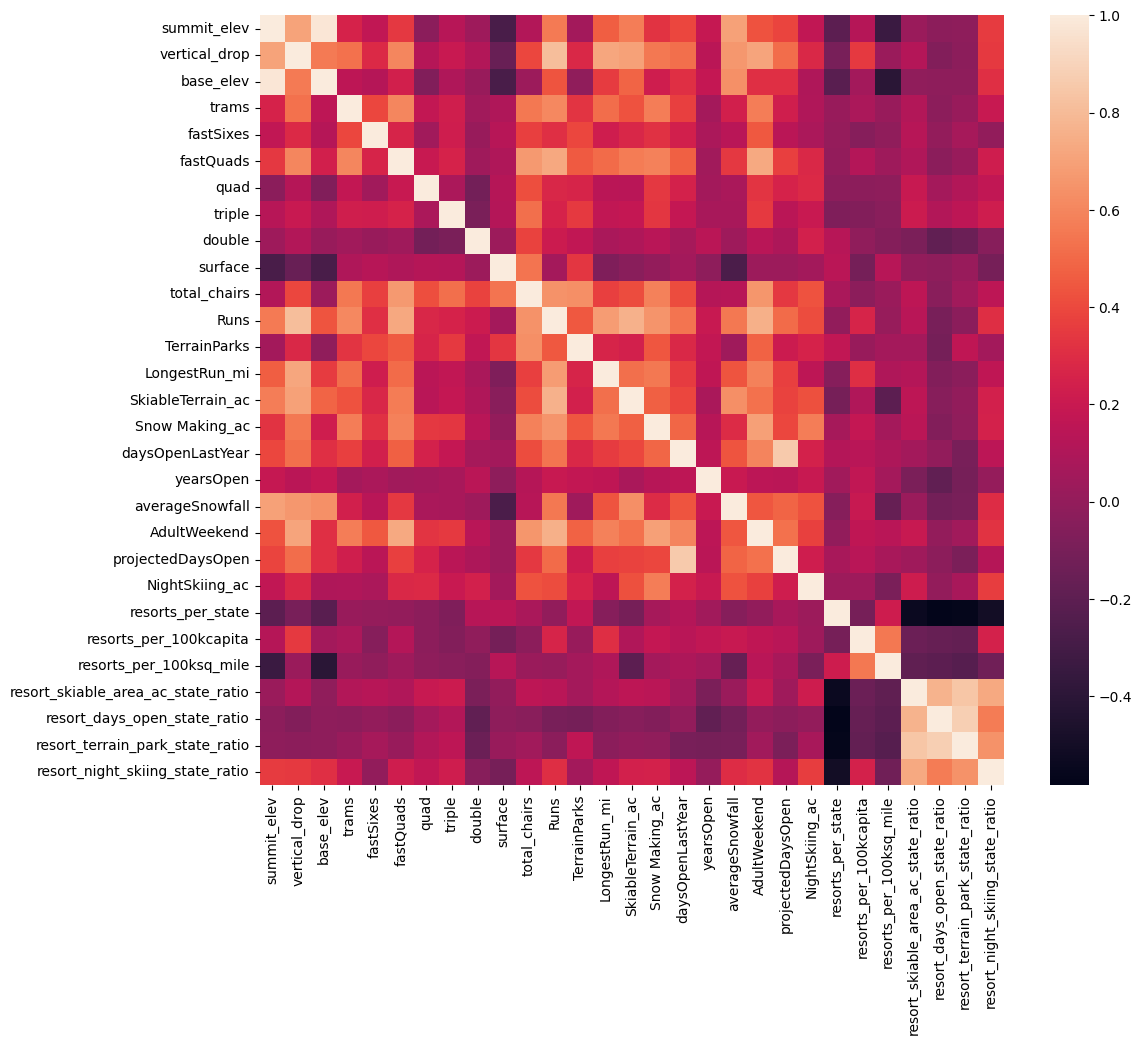


Figure 2

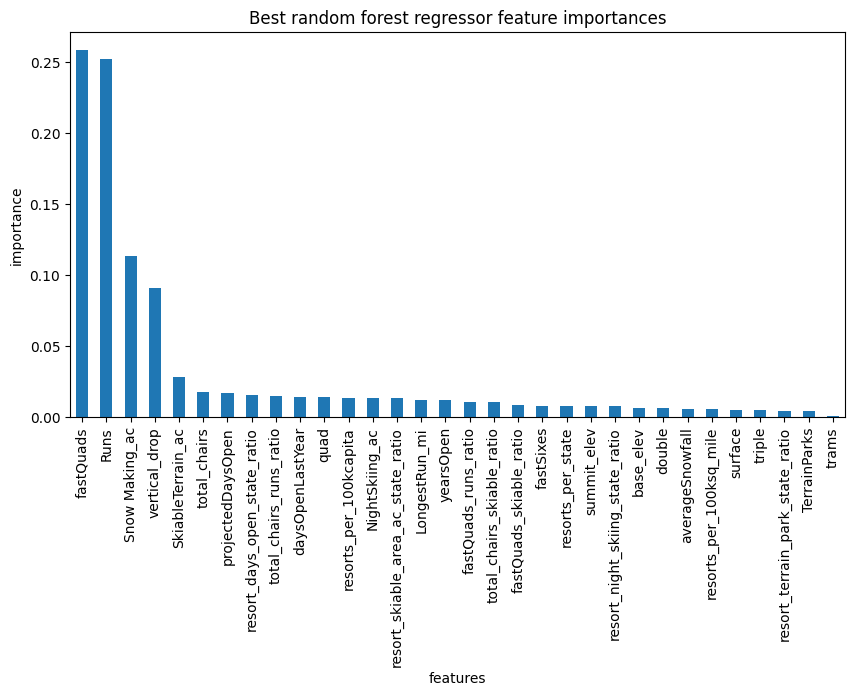


Figure 3

