* + **Problem statement**

What opportunities exist for Big Mountain Resort to optimize its pricing strategy based on its available amenities by the end of the season?

Big Mountain Resort has installed an additional lift with a seasonal operation cost of approx. $1.5 million USD this season. Historically, the business has used a broad spectrum approach to pricing tickets ( premium). However, leadership is looking for more of a data-driven pricing strategy.

* + **Data Wrangling**

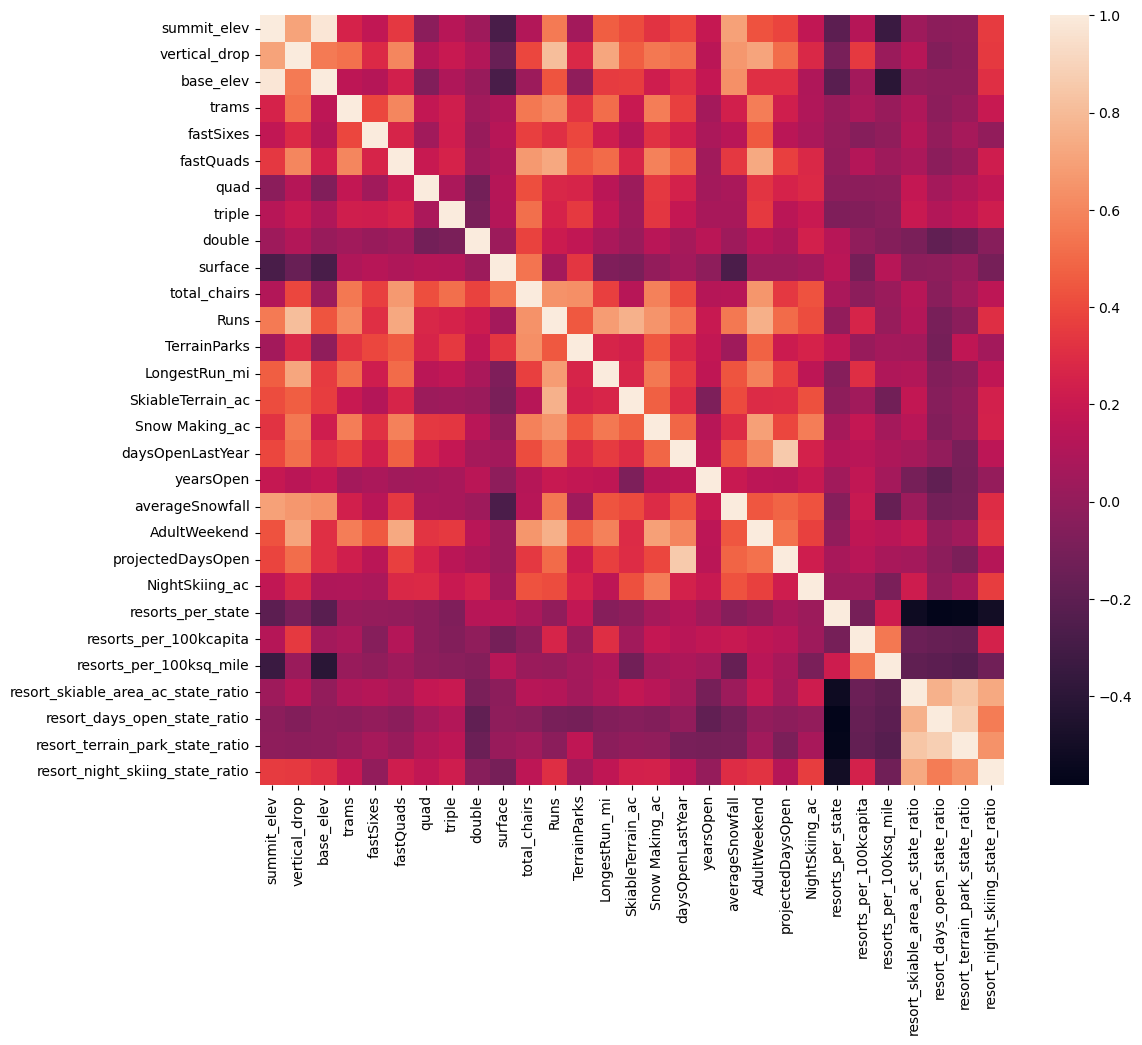
I first needed to examine the ski data set to check for completeness by first loading it into a dataframe and transposing it. Next I looked at locating missing data and duplicates to clean the data frame. I dropped 16% of the dataset as a result and plotted various features to identify other possible issues in the data.

* + **Exploratory data Analysis**

The analysis began by exploring the relationship between state and ticket prices, initially finding no significant correlation. Attempts to categorize resorts based on quartiles and variance did not reveal clear patterns for a pricing model. However, a heatmap of ski data features uncovered a positive association between adult weekend pricing and snowmaking ability. Further investigation through scatter plots highlighted strong correlations of pricing with vertical drops, Fastquads, runs, and total chairs. Examining ratios such as total chairs to runs and fastQuads to runs revealed a quadratic relationship with prices, indicating an optimal point for price increase before diminishing returns. Initially treating all states similarly in labeling is plausible, but revisiting this approach based on features may yield advantages, especially for states resembling Montana.

* + **Model Preprocessing with feature engineering**

Some of feature engineering i did involved creating a heatmap of various feature correlations showing some possible features we could use in subsequent modeling.



The analysis reveals several key observations. Summit and base elevation show a high correlation, as expected. The introduction of new ratio features creates multicollinearity, negatively correlated with the number of resorts in each state. Notably, there's a positive correlation between the ratio of night skiing area and the number of resorts per capita, suggesting more night skiing in densely populated areas. For the target feature, AdultWeekend ticket price, notable correlations include fastQuads, Runs, and Snow Making\_ac, indicating the value visitors place on guaranteed snow. Among the new features, resort\_night\_skiing\_state\_ratio emerges as highly correlated with ticket price. Additionally, total\_chairs and vertical drop are well-correlated with ticket prices, emphasizing their influence on pricing.

* + **Algorithms used to build the model with evaluation metric**

The analysis began with a baseline using the average price, resulting in an R-squared value of 0.7 to 0.8, indicating useful predictions based on the mean. Subsequently, a linear regression model was developed with imputation, scaling, and cross-validation, demonstrating consistent performance on the test set. An alternative random forest model was then created, exhibiting improved accuracy with a final MAE of 9.5 on the test set. The decision to favor the random forest regressor was based on its superior performance in both cross-validation and test set scenarios, highlighting its predictive accuracy.

* + **Winning model and scenario modeling**

In conclusion, the carefully tuned random forest regressor will be employed in addressing the next steps of the business problem

* + **Pricing recommendation**

The ski data modeling, excluding Big Mountain for prediction, yielded an estimated ticket price of $81 with a MAE of $10. A comparison of Big Mountain against other resorts highlighted its potential to justify a higher price due to superior features. Exploration of scenarios, including closing runs and enhancing vertical drop with snowmaking, revealed that closing underperforming runs (Scenario 1) could lead to potential savings without impacting ticket prices.

* + **Conclusion**

Considering closing 2 or 3 runs initially and tracking revenue quarterly would allow for a measured approach to assess the ROI of this operational adjustment. We can also consider raising the weekend pricing upwards of $1.99 per ticket.

* + **Future scope of work**

Outside of operational cost data for chair lifts, I think taking a look at demographic and marketing spend data could be useful. One could create another model to include this to figure out who is the target demographic(s) and where marketing spend can be optimized to get the best ROI. I do think the model ticket price vs actual should come as a surprise to executives. Big Mountain seems to be a premium resort due to its pricing being significantly greater than other resorts in Montana.