**Problem statement**

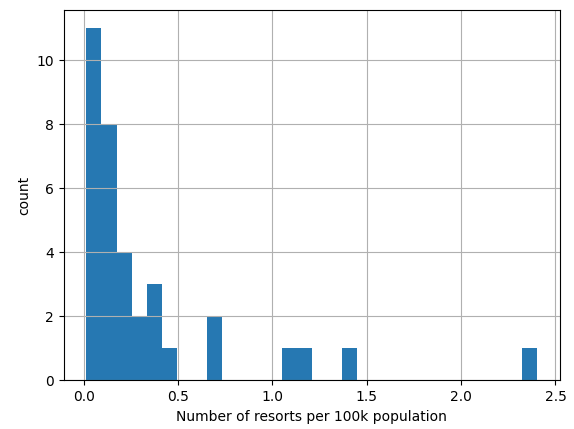
What are ways Big Mountain Resort can ensure they are charging a fair (or slight premium) price for their tickets and/or making changes to increase profit margin or increase value of tickets by considering if a $1,540,000 investment in lifts this season is worth it specifically and increasing lifts and/or run capacity generally is worth it.

**Data Wrangling**

Data wrangling involved dropping any features for which we had little or no data, verifying we had no duplicates, correcting any errors in features like area etc. Also determining our target, weekend ticket prices as we had less nan values.

**Exploratory data Analysis**

We started by looking at several features (# of resorts, ticket price etc.) as a function of various state attributes like population, area, etc. See the below plot showing the number of resorts per 100k population. Ultimately we chose to treat each state as equals when moving on to the feature engineering stage.



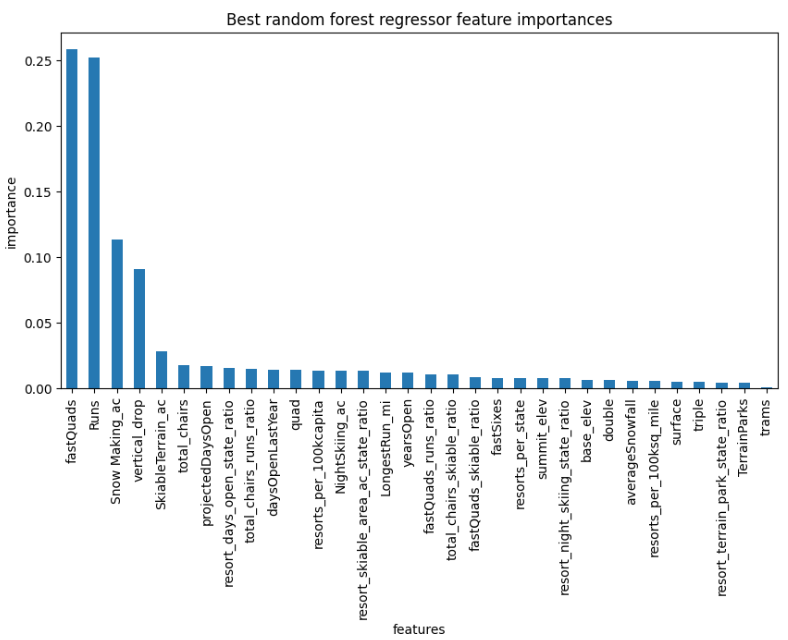
**Model Preprocessing with feature engineering**

Feature engineering mainly involved normalizing our data so that features from any state regardless of population or number of resorts could be more easily compared, e.g. ratio of resort skiable area to total state skiable area.

Model selection started with a prediction of the average ticket price on a training set and tested this prediction on a test set to get a baseline. We then moved to a linear model and ultimately a random forest (RF) regressor, more on that next. The last pre-processing before fitting our model involved imputing missing values with the median (we also tried scaling to 0 mean and unit variance but opted against in the end).

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

We evaluated the models using R2 (which states the % of variation explained), mean absolute error (MAE) and mean squared error (and its derivative, root mean squared error, which is basically 1 standard deviation of our prediction). The cross-validated MAE was $9.6 with a standard deviation of $1.35 and $9.5 on the test set meaning this model would be within ~$10 for the average predicted ticket price. The below plot ranks the features by importance according to the RF model.



**Winning model and scenario modeling**

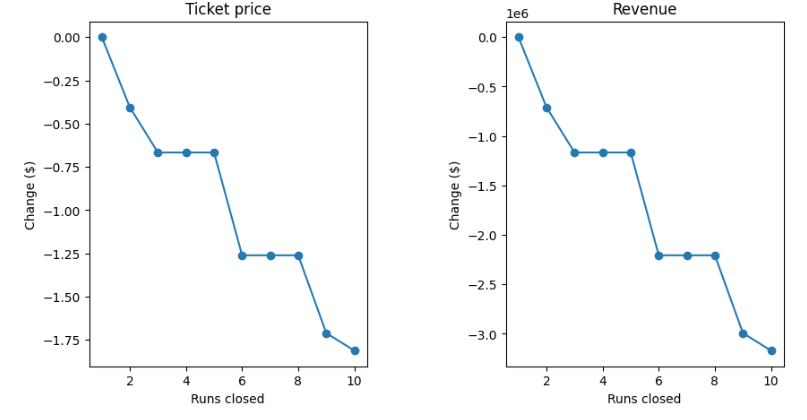
The winning model was a random forest regressor which

**Pricing recommendation**

The modeled price per ticket was $95.87 with an expected mean absolute error of $10.39 and actual price set at $81, suggesting there’s room for an increase in price per ticket.

**Conclusion**

The scenario which adds a vertical drop by adding a run to a point 150 feet down, adds an additional chair lift without additional snow making coverage seems to make the most sense. This will increase the value per ticket by $1.99 and with expected 350k visitors will bring in ~$3.5 million in additional revenue assuming we can sell the same number of tickets. We get no added value in increasing the snow making area. With a cost of ~$1.5 million for the added lift, this will mean ~$2 million in profit after a season and assuming the operating costs are significantly less than the installation cost, this will be even higher in subsequent years. Also since we can close 1 run and seemingly lose no revenue, after adding the run that comes with this added lift, we could potentially close two other unrelated runs somewhere else, ideally ones that aren’t very popular and have relatively high operating costs, and cut costs even more. See the below graphs that show change in value per ticket and revenue as a function of runs closed.



**Future scope of work**

The next step would be deploying this model into a production environment for repeated use by analysts. They could plug in various combinations of adding/subtracting features and finding a sweet spot or costs and revenue.

Also getting more data which has information on the number of visitors per unit time. Also more cost data on each and every feature, e.g. runs, snow making area, etc. This would potentially help build a better model and more accurately model what affects ticket price/revenue.