Big Mountain Resort Report

Big Mountain Resort is a ski resort located in Whitefish, Montana. They are building a new lift and as a result are expecting an increase of 1.5 million dollars in operating costs. Also their management suspects they aren't properly pricing their tickets with respect to their facilities and market position. Combine these two factors and they are looking to possibly change their ticket prices, ideally increase them in this case, to cover the increased costs and result in more revenue. Our goal is to take a data driven approach, using a dataset of 330 resorts across the US including Big Mountain, to compare and build a model to choose how to best position Big Mountain's ticket prices in their market segment.

We went about our goal by going through all the steps of the Data Science Method from examining the data, wrangling it, EDA, cleaning it, to finally pre-processing and trying several machine learning models. The task being to make a clean, clear dataset with all the important representative features for the resorts so as to be as useful as possible when modeling and thus predicting what the target feature (AdultWeekend ticket prices) should be.

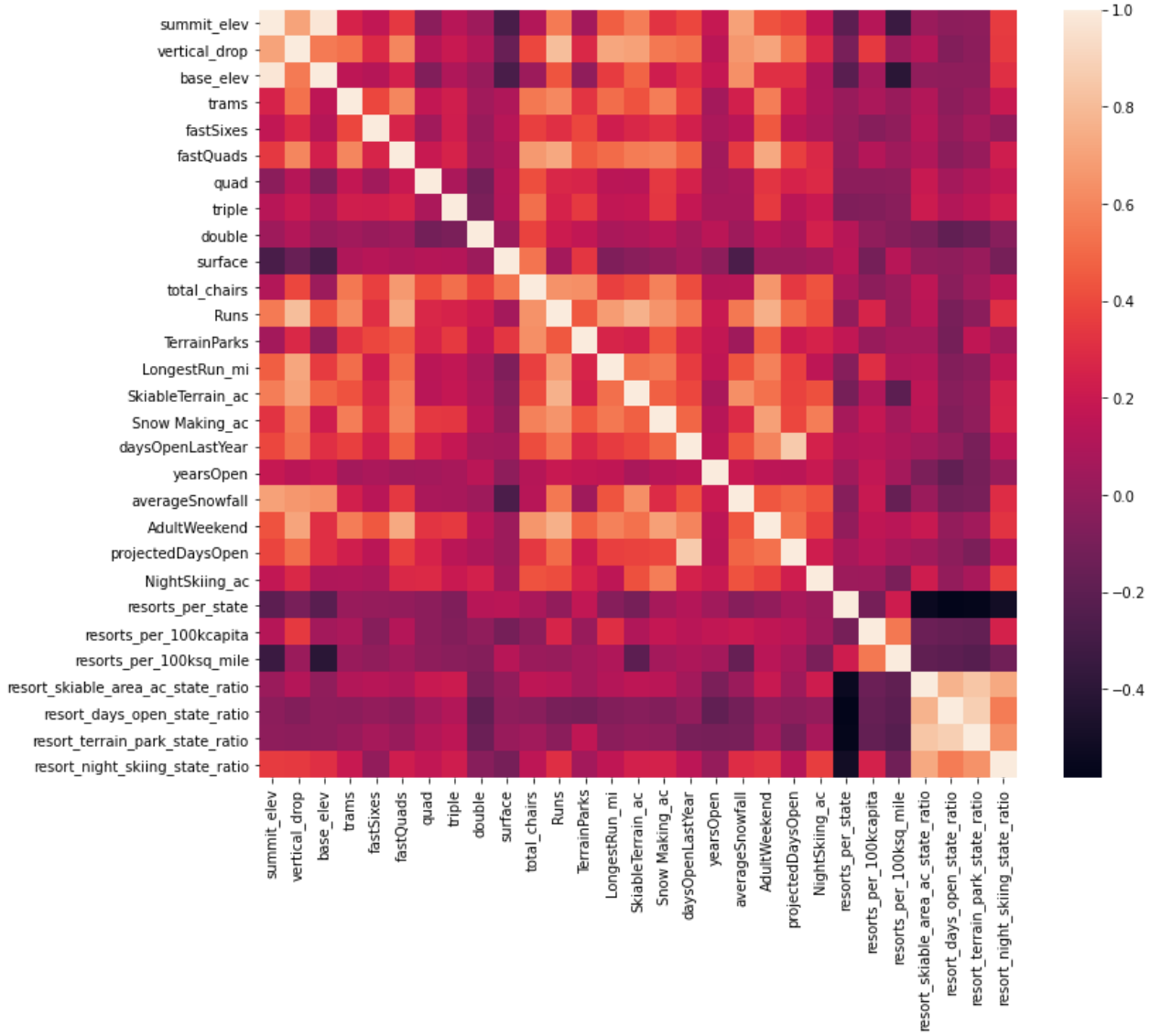


Image 1 : A correlation heatmap of the features in the Ski Resort Dataset, lighter colors correspond to higher positive correlations. Of particular interest is AdultWeekend seeing as it is our target feature.

An example of some of the work done in the EDA of the data is shown in the correlation heatmap of the features above. We can quickly and clearly see the strong positive correlations of some features (such as vertical drop, fast quads, total chairs and runs) or the negative correlations of others (doubles, surface, years open and resorts per state for example) with our target feature 'AdultWeekend' (ticket prices). This shows how important EDA is to quickly gain insights about our data.

Now having identified our target feature and our dataset being explored and cleaned we moved on to pre-processing our data and building a few machine learning models. Seeing as our target feature is a continuous, numerical feature we used regression models. We tried a couple different models, with the primary two being Linear Regression and Random Forest Regressor. Both of which were iterated upon and finally a GridSearch cross validation was done to search for the optimal hyperparameters. Then we compared the finalized model of each using a few scoring metrics (particularly the mean absolute error and its standard deviation, along with the coefficient of determination or R^2). The Random Forest model outperformed the Linear Regression model in every metric and thus was chosen to be the final model to make our prediction.

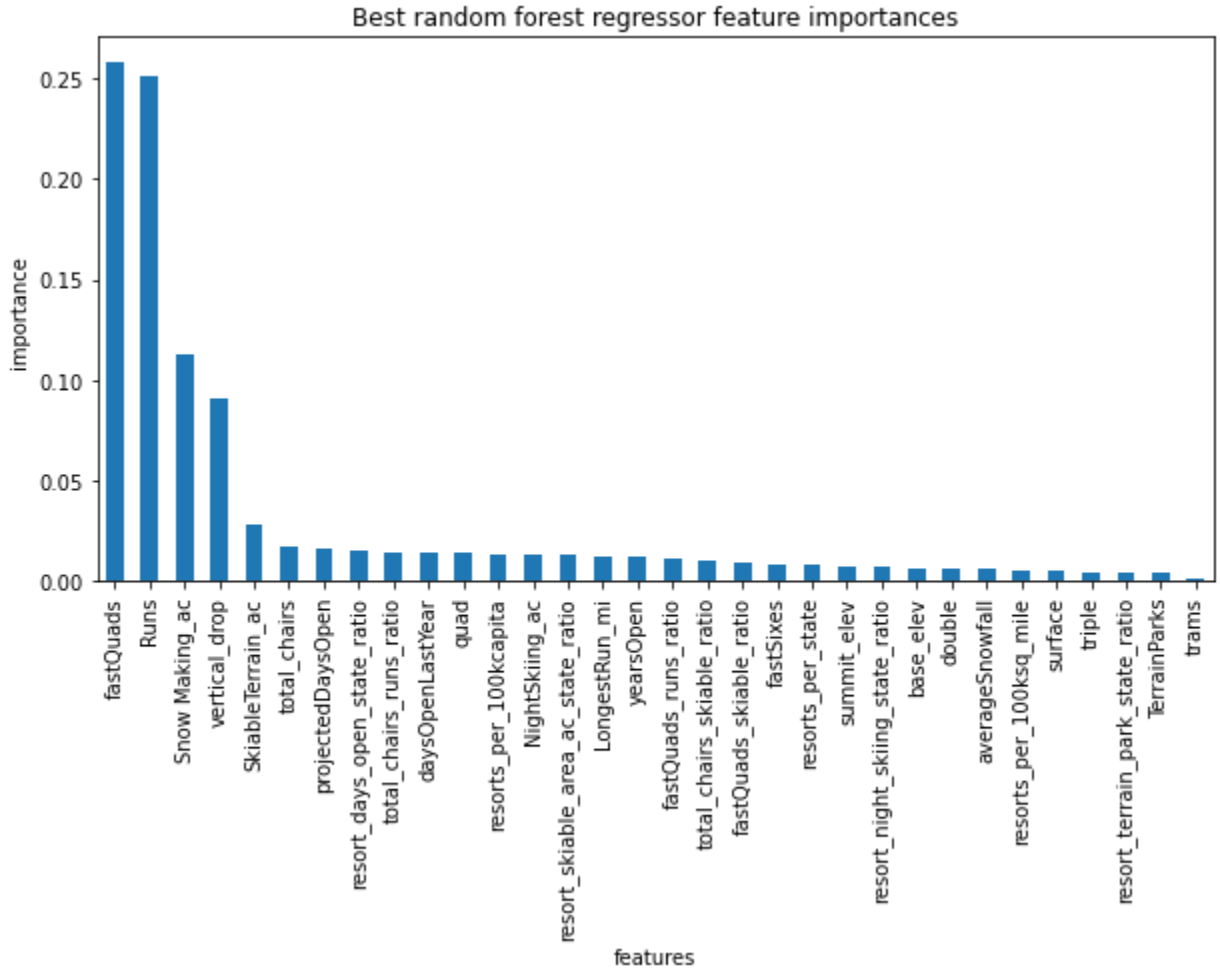
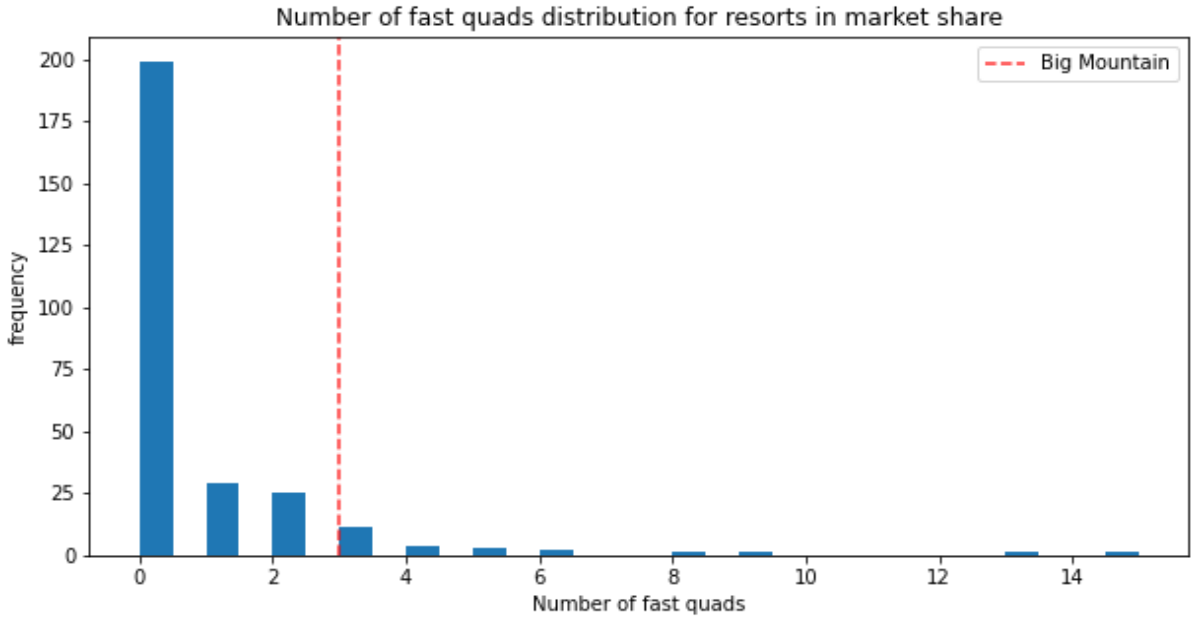
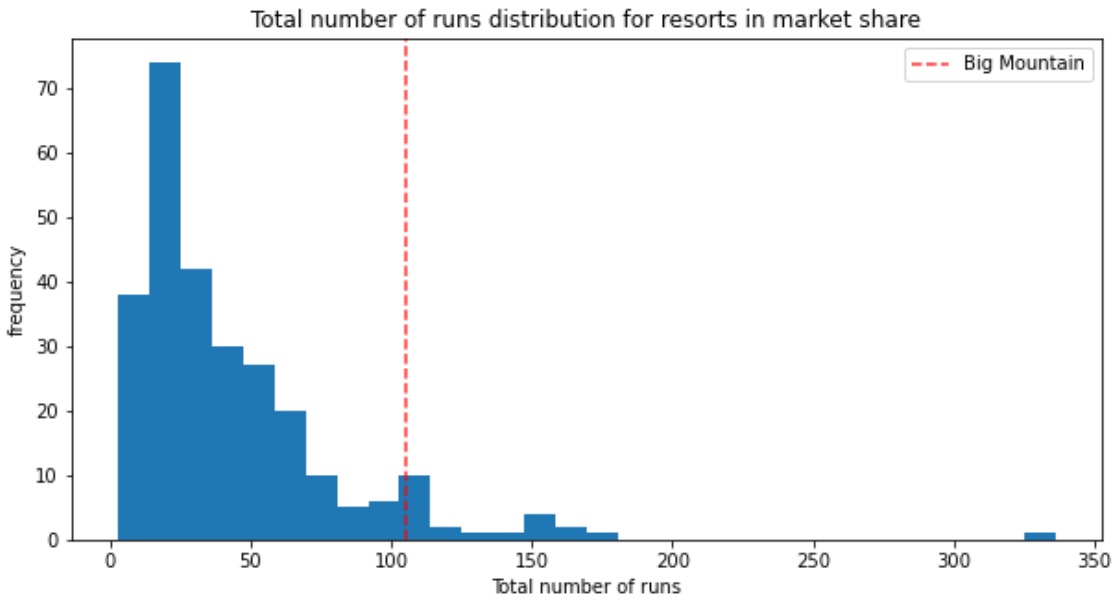


Image 2: The Image on the previous page shows the feature importance in the final model, the Random Forest Regressor. We see that a few features were especially important for predicting ticket prices and that they line up well with our correlation heatmap.

So now having chosen the best model we can finally use it to see how Big Mountain's current ticket prices compare to the it's predictions. The current price is $81 dollars for a weekend ticket. The model predicted the price to be $95.87, which even accounting for the mean absolute error of $10.39 in the worst scenario points to potential for a price increase. Obviously there are many other unknown factors such as how well do the other resorts price their tickets (which the model is based on) and are we maybe lacking key data. Having made that qualification, it still seems that Big Mountain should strongly consider increasing their ticket prices based on this model, depending on how aggressive or conservative they want to be.





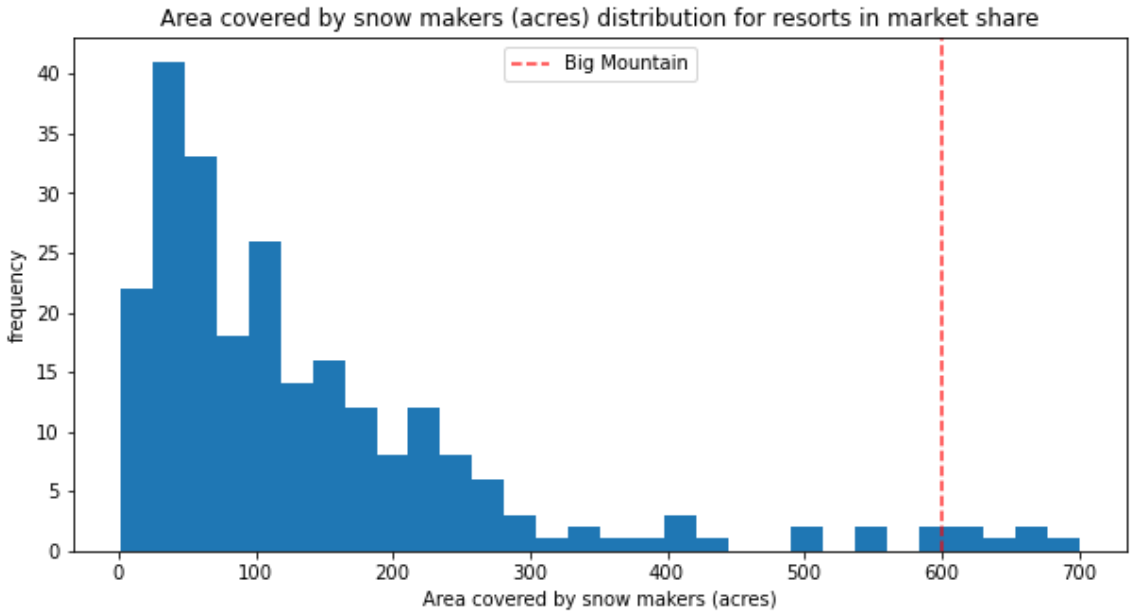


Image 3, 4 and 5: These figures highlight the three most important features in our Random Forest Regressor model. The red-dotted line shows how Big Mountain lands on the histogram for each. Big Mountain consistently is on the higher end in all the most important features, it is no surprise then that our model predicts a relatively high ticket price.