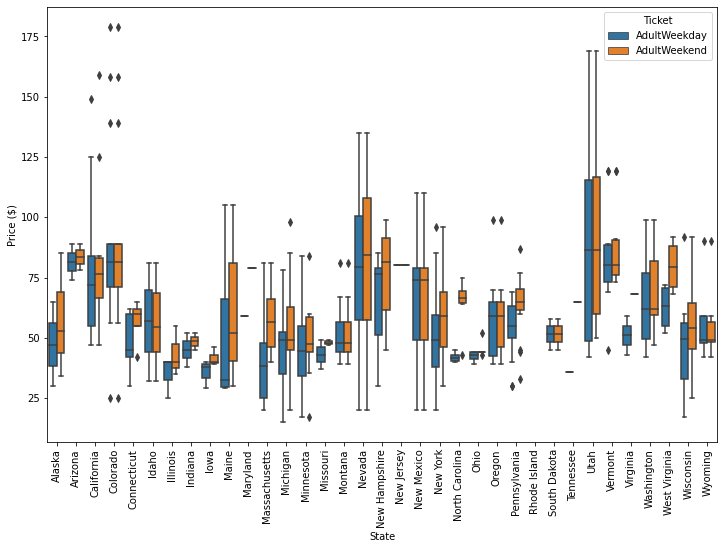
**Guided Capstone Project Report**

For this guided capstone, I was asked to take a look at data for Big Mountain Resort, discover if their ticket prices were in line with other resorts and the amenities they offered, then make a recommendation on how they could increase profits by adjusting their ticket prices, and discovering which amenities were the most profit driving. For the problem statement for this project, I chose to set a goal of increasing profits by 5% over the next ski season, by focusing solely on finding the optimal ticket price for the resort, while acknowledging that without more detailed information on operating costs, it would be difficult to do any analysis on possible cost saving measures.

The next step in the project was data wrangling, where I went over the data in a general sense in order to look for any obvious problems and find what target feature I wanted to use. As only one resort had any fast eight person lifts, and much of the data in that column was missing anyways, the fastEight column was dropped entirely. Some other problems that stood out was the fact that large outliers were found in the skiable terrain and years open columns, and these were investigated further and both corrected with the proper data. As for target feature selection, information on both weekday and weekend prices was given, but only one of these could be chosen for the final analysis. As shown below; Montana, where our resort is located, has identical weekend and weekday prices, as do several other states. This suggests that for our analysis, which ticket price is chosen is not very important for predicting the ticket price of our resort, and as the weekend price column has more complete data, it was chosen to be our target feature and the weekday prices column dropped.



*Figure 1: Boxplots showing the weekday and weekend ticket prices for each state*

The next step in the project was doing exploratory data analysis. One of the major questions I attempted to resolve in this section was how to treat the different states, if they should all be considered equally or if including all of them would throw off the final analysis. By doing a principal component analysis (PCA) I determined that while some states like Vermont and New York accounted for a lot of the variance in the data, nothing indicated that they would affect the analysis later. In addition, our contact at Big Mountain Resort told us to treat all of the states as if they were part of the same market, so I felt justified in keeping all of the states and regions for the final analysis. Some state-based derived features were also created at this stage, such as the ratio of total skiable area in the state the resort had, or the number of resorts per 100k square miles, among others. A large correlation matrix was made with both the old and derived features in order to see any interesting relationships between potential features and our Adult Weekend price. This confirmed that none of the “state-based” derived features correlated strongly with he ticket price, confirming the decision to treat all states equally.

A picture containing text

Description automatically generated

*Figure 2: A correlation matrix/heatmap of all of the features to be used in the final analysis*

Finally, the next step was model preprocessing and training. I began by establishing a standard baseline for model performance, which simply predicted the mean value for ticket price each time, the “DummyRegressor” model. The evaluation metrics used were the coefficient of determination (R2) and the mean absolute error (MAE). These metrics, when applied to the dummy regressor model, gave an R2 value of 0 and an MAE of ~$19. Next, a linear regression model was created. After doing some feature selection with “SelectKBest” to avoid overfitting, a model with an mae of ~$10.5 with a standard deviation of $1.6 was made. It was also discovered that vertical drop and snow making capability were some of the most important features. A random forest regressor model was also tried; it was found that a random forest of 69 "trees", with an imputer strategy of 'median', and no feature scaling produced the best results. It had an estimated mae of $9.6, with a standard deviation of $1.35 on cross validation on the training set. This agrees with its performance on the test set of $9.53 mae. The lower mean absolute error on the random forest classifier as well as its lower standard deviation meant that it was chosen as the model to use going forward.

In conclusion, Big Mountain Resort currently charges $81 for its weekend ticket, while the model suggests the price should be $95.87. The mean absolute error for this model is $9.53, so clearly some level of price increase seems to be justified, perhaps ~$10. This seems to be because Big Mountain has high numbers in features, such as skiable terrain and fast quads, that predict a high ticket price. In addition, some scenarios were modeled in order to predict how they would affect profitability at the resort. It was found that for the resort closing between 1 and 10 runs, closing just one made no difference to the expected ticket price, and closing 3 runs was the same ticket price wise as closing up to 5 runs. It was also found that adding a run and a new chair lift to service it could justify a ticket price increase alone of $1.99, but adding small amounts of snow making capability to the resort made no real difference. In terms of the future scope of this work, it would have been useful to have more detailed cost information on this and ideally other parks. It was difficult to predict real profit increasing measures without a good idea of costs throughout the park. It would also be potentially useful to set up the model in such a way that analysts at the resort could run scenarios on their own without my help, assuming they found this model to be useful.

A picture containing shape

Description automatically generated

*Figure 3: One feature where Big Mountain scored highly compared to other resorts was total skiable terrain, which was also a feature that was positively correlated with high ticket prices.*