**Problem:**

Determine a pricing model for the Big Mountain Resort: How to maximize their returns relative to its market position and which facilities matter most to visitors in terms of what they are willing to pay for?

**Introduction:**

Big Mountain Resort is a ski resort in Montana with about 350,000 visitors a year. The resort recently installed an additional chair lift at a cost of 1.5 million for the current season. Current costs are based on the market average and currently charging a premium for admission.

**Data:**

Provided with single csv file from database manager, Alesha. This dataset contains information on ski resorts across the United States. This data includes but not limited to: name, summit elevation, base elevation, trams, skiable terrain acreage, etc.

Another dataset containing state population data was web scraped from wikipedia to allow for assessing the associations between population and ski ticket price per state.

**Methodology:**

1. Clean datasets
2. Data Exploration into patterns associated with pricing.
3. Predictive pricing model development
4. Testing
5. Tuning

Using the ski\_data dataset we created a train and test split dataset for model evaluation. We used linear regression and random forest models. First we processed the data manually with pandas to calculate R\_squared, Mean Absolute Error, and Mean Squared Error. Then we used sklearn's built in metrics to calculate the previously mentioned values. We then imputed the median and mean to fill in any missing values. We trained models using data with the imputed median and mean values to uncover that the results are not significant.

After training the models we used sklearn's pipeline to replace missing values with the median for each feature, scale the data to zero mean and unit variance and train a linear regression model. We trained and then refined the pipeline to deal with suspected overfitting by using SelectKBest to select the k best features. We then assessed performance using cross-validation. Then we did Hyperparameter search using GridSearchCV to determine the best value of k which we found to be 8.

Then we found the features with the largest positive coefficients to determine which features best to train on. The next step we used the random forest model to create a pipeline, fit and assess performance using cross validation, hyperparameter tuning using GridSearchCV, and completed the R\_squared, mean Absolute error and Mean Squared Error as before. We then compare the models performance and found that the random forest model has a lower cross-validation mean absolute error.

**Analysis:**

By taking all resorts to be part of the same market share, there was a pattern suggestive of a relationship between state and ticket price. The features that are correlated to price are fastQuads, Runs, Snow Making\_ac, resort\_night\_skiing\_state\_ratio, vertical\_drop and total\_chairs. We decided to focus on these numerical features for determining price. However, upon further analysis we found that fastQuads, Runs, Snow Making\_ac and vertical drop are the dominant top four features

**Results:**

We suggest the Big Mountain Leadership implement Scenario 3. In this scenario, Big Mountain adds a run, increases vertical drop by 150 feet, installs an additional chair lift, and adds 2 acres of snow making.

Big Mountain currently charges $81 per ticket. The scenario 3 Model supports a ticket price increase of $9.75 to $90.75.

**Next Steps:**

Cost about the operating expenses of the entire resort features would be very helpful. The modelled price is higher than its current price because its features are estimated at being more valuable than Big mountain is currently charging for them. I would want to know how the current business leaders determine their pricing. If the business leaders find the model useful I hope they would implement it. I think I would best serve the leadership by building multiple models based on their requirements and providing them the models which I think would provide the highest cost savings and revenue increase. Business analysts may make use of this data to find new opportunities for cost savings or capitalizing on existing features.