Big Mountain Resort is wanting some guidance on how to select a better value for their ticket price. They are considering a number of changes that they hope will either cut costs without undermining the ticket price or will support an even higher ticket price. With the installation of an additional chair lift costing over $1.5 million, the company wants to implement a more data driven business strategy.

There were multiple processes used in gathering information. Many different pandas, matplotlib, and seaborn methods were used frequently. Originally there were 58 rows of data. By the end of data wrangling, there were 67. The ‘weekday prices’ as well as the ‘fastEight’ columns were dropped. Other rows were dropped because they were missing the desired target ticket price.The percentage of missing values per row appeared in multiples of 4, as if they were artificially removed. Values were sorted and many columns were not dropped for the main purpose of the importance of their missing data.

In further exploration of data analysis, I found there were many numerical data features, such as: Total state area, Total state population, Resorts per state, Total skiable area, Total night skiing area, Total days open, Resort density, Average ticket price by state and an Average ticket price scatterplot. I did not find as many categorical data features, however, I was able to find a couple; Top states by order of each of the summary statistics and Top states by resort density. I could not find a specific pattern suggesting a relationship between state and ticket price. I was led to the conclusion that seaborn is the most comprehensive feature to use regarding subsequent modeling. I've found to always remain wary of the following aspects when performing feature selection: Multicollinearity, Irrelevant features, Overfitting, Feature scaling, Data leakage, Missing values, Feature interaction, Domain knowledge, and Dimensionality.

During preprocessing and training, a baseline idea of performance was gained by simply taking the average ticket price. However that prediction was found to be within $19 of the real ticket price. Hmm. To get even closer to the real ticket price, a linear regression model was used and that model explains over 80% of the variance on the train set as well as over 70% on the test set. Using this model, on average, you would expect to estimate a ticket price within approximately $9 of the real ticket price. Testing performance using the test/split method, as expected, did not hold up consistently. The next model used is the Random Forest Model. This model has an even lower cross-validation estimate, to the real price, by almost $1 and also tests consistent estimates with the various performance results. With all of this data, I have chosen to use the Random Forest Model, basing my choice off of the consistency of the model's results and the ability to use the estimate on various areas of data for additional proactive solutions or predictions for conflict resolution.

Big Mountain Resort currently charges $81 average price per ticket. The price suggested per ticket, from modeling the data, indicates an average ticket price of $82.53 with a mean absolute error of roughly $14.31. This model/estimate surely indicates room for a ticket price increase. The modeling also indicates that adding a new chair lift can build support for ticket price increase by about $2.25, which can be expected to amount to about $3,931,729 over the season. Modeling also indicates that an increase in snow making area makes no difference. It seems the model also indicates that closing one run makes no difference. Successfully closing 2 or 3 runs reduces support for ticket price increase and, of course, revenue. Closing 4 or 5 indicates no further loss/gain in ticket price. Any amount of closures after 6 indicates a large drop in support for ticket price increase. I would recommend the 2nd modeled scenario, which is an increase in the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage.

For further exploration/work, the ‘Runs’ data not being found in the DataFrame hampered/limited the findings in this assignment. That information and cost of each specific Run would be useful information for data understanding. The modeled price estimating so high compared to the actual price could be because of the possibility that some of the competing resorts are overpriced and Big Mountain Resort could be underpricing. Based on the data and comparison of what is offered at Big Mountain and what is offered at other resorts, as well as the estimated revenue increase for Big Mountain, I think that the business executives would be surprised and pleased with this information. They could make use of this information by saving the file and altering the algorithm based on what findings/information is being requested.