Project Link:
https://docs.google.com/document/d/1o-Doa784H3Sov2BWAy5Lm3IcCPDcDXjU1z3DsWvY4aA/edit?usp=sharing

Big Mountain Resort: Guided Capstone Project Report

***Problem Statement*:**
What changes can be made this season to offset the $1,540,000 cost of the new lift by cutting operational costs without undermining ticket pricing, or by supporting a higher ticket price?

***Conclusion*:**
Big Mountain Resort provided four possible options to accomplish offsetting the $1,540,000 cost of the new lift. Of the four possible options provided by Big Mountain Resort, the second option of adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift is the best choice. Doing this allows a $1.99 ticket price incrase from $81 to $82.99, and brings in $3,474,638 in additional revenue. This additional revenue is more than double the original $1,540,000 Big Mountain needed to offset the cost of the new lift.

***Process*:**
In the initial data wrangling stage, the data for Big Mountain Resort was visualized, and the number of missing values for the entire data set were calculated per observation. The fast eight observation was 50% missing, and adult weekday ticket pricing had more missing values than the adult weekend ticket pricing information. These two observations were dropped from our data set. Several state-wide comparisons were done to determine if the state had any affect on our target feature, ticket pricing. Though there were some patterns in regards to population density and resort density, no definitive affect on ticket pricing was observable by measure of the state alone. Next, each individual feature was visualized in histograms displaying the frequency of each feature in all states. These histograms indicated that there may be problems using the fast eights, fast sixes, snow making acres, skiable terrain, trams, and years open due to low clustering indiciating it is not an important feature, or an unreasonably large entries indicating a recording error. Further exploration and research of Colorado's Silverton Resort website indicated that it's unreasonably large skiing terrain was an error. The error was corrected in our data set according to the website's information. Similarly, there was an error recording the snow making area at Heavenly Mountain Resort. The incorrect data value was replaced by calculating 60% of the total terrain, as indicated in the resort's website. To correct the probable years open data recording error, any resort that had reportedly been open for more than 1,000 years was excluded from the data set. The resulting data set's summary was modeled in histograms. Deciding that population may affect our marketshare comparisons, each state's demographic information was imported from Wikipedia. The state name, total population, and total area were extracted from the Wikipedia data set and added it to our original data set.

Next, ticket pricing and individual features were more closely explored. Resort density was explored by comparing the number of resorts per 100,00 capita in each state, as well as the number of resorts per 100,00 square miles in each state. The resort density had a varied effect on ticket pricing though, so the high dimensional data was explored by scaling and comparing data in a principal components analysis (PCA). The results indicated that the first two components seem to account for over 75% of the variance, and the first four for over 95%. Modeling this in a scatterplot, Vermont and New Hampshire were definite outliers. Next the average ticket price per state was calculated and incorporated into the PCA scatterplot color coded by price quartile, with size also indicating a price increase. This again showed Vermont and New Hampshire to be outliers in respect to the first component. Both states had high resort density in regards to resorts per 100,000 acres, and Vermont also had a high resort density in regards to resorts per 100,000 capita. Combining all information from our original set of resort features with our resort density calculations, it was observed that the ratio of resort skiable area to total state skiable area, the ratio of resort days open to total state days open, the ratio of resort terrain park count to total state terrain park count, and the ratio of resort night skiing area to total state night skiing area were the primary competitive state features for resorts. A heatmap comparing individual resort features and adult weekend ticket pricing was created. The heatmap indicated that fast quads, runs, snow making acres, total chairs, and night skiing, and vertical drop were most correlated with ticket pricing. Scatterplots comparing each feature and ticket pricing were created. The scatterplots revealed a negative association between the ratio of total number of chairs to total runs and ticket pricing. They also indicate that having no fast quads may limit ticket pricing, but a small number of fast quads is sufficient if the resort has a wide area to cover.

Linear regression and forest regression were both explored with cross-validation. The forest regression model proved to be the most accurate method while using 8 best features on the test/train split. The top four features were observed to be fast quads, runs, snow making area, and vertical drop. In the final steps of our project, the model was reset over all the data, excluding Big Mountain. This indicated that, according to statewide comparisons, Big Mountain could increase ticket pricing from the current $81 per ticket to $95.87 per ticket. However, analyzing Big Mountain's ranking among the key features compared to state placed Big Mountain on the upper end, but not the maximum ranking of any category. Therefore, more specific cost reduction or feature improvements were analyzed to determin a course of action to offset the cost of the new lift. The business leaders at Big Mountain provided the possible options of (1) permanently closing down up to 10 of the least used runs, (2) increasing 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, (3) same as number 2, but adding 2 acres of snow making cover, and (4) increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres. Modeling option (1), it was apparent that closing one run had no affect on ticket pricing, but closing more than one run would decrease ticket pricing, with the sharpest price decrease associated with closing 6 or more runs. As this does not increase revenue, the other otions were modeled. Option (2) indicated that the improvements would support increasing ticket pricing by $1.99 per ticket, raising individual ticket prices from $81 to $82.99, yielding a revenue increase of $3,474,638 this season. Modeling option (3) did not support any additional ticket price increase, and modeling option (4) showed no affect on the current ticket pricing at all. Therefore, it is likely that option (2) is the best option to pursue.

Further Investigation:

Before proceeding with scenario, the cost of creating the new trail needs to be explored to ensure it does not outweigh the increased revenue of the new trail. Business leaders may also want to create additonal options combining an operating cost reduction with increased feature attractions. In order to do this using this model, they could compare the amount saved by decreasing operational costs while closing runs, with the increased revenue made by increasing top features. However, additional models projecting overall affect on revenue per feature closed to decrease operating cost or per feature improved to increase revenue may be more helpful.