Key processes in this notebook included:

1. Identifying areas of missing data
2. Investigating where the resort of interest fits into the overall data
3. Reviewing distributions of columns with histograms to identify anything unusual

Original row count was 330, which was reduced down to 277 primarily due to lack of price data. One row was removed due to the resort being "2019" years old. The fast eights column was removed due to lack of data (and lack of >0 entries where data was present). One entry, Silverton Mountain had unusually enormous skiable terrain, which was assumed to be a typo and was updated from doing a google search. Futher, state population and state area was incorporated into the original dataset. A handful of target features, such as skiable terrain and number of runs, have been identified for helping to predict ticket price.

In this data, categorical features included things such as region and state, whereas numerical features included parameters such as summit elevation, number of runs, and night skiing acreage. State did not suggest a relationship with ticket price, so some feature engineering was conducted and a correlation heatmap was produced. Some features added for investigation included:

* ratio of resort skiable area to total state skiable area
* ratio of resort days open to total state days open
* ratio of resort terrain park count to total state terrain park count
* ratio of resort night skiing area to total state night skiing area

One aspect to be wary of with relationaship between features are confounding variables from multicollinearity - while something may be correlated with ticket price, it may not be an actual driver of ticket price. Further, while state does not seem to heavily contribute to ticket price, we should still remain wary of ignoring it for modeling.

Three models were evaluated in this notebook - the mean model, a linear model, and a random forest model. The mean provided a baseline comparison, coming in with a MAE of ~19.

The linear model was evaluated with both median and mean imputed values, with MAE being improved upon the mean model to ~9 for both impute approaches. Refining the linear model using k best features identied 8 features performed the best, and this brought the MAE to ~12 with the following features and their respecitve coeffiecients:

* vertical\_drop 10.767857
* Snow Making\_ac 6.290074
* total\_chairs 5.794156
* fastQuads 5.745626
* Runs 5.370555
* LongestRun\_mi 0.181814
* trams -4.142024
* SkiableTerrain\_ac -5.249780

Further, cross-validation provides CV values in line with this linear model and suggests that the uncertainty falls to under a dollar with high confidence (am I interpreting that correctly?).

The random forest model was evaluated using median imputed values and standard scaling, and its MAE came in at ~9. Cross-validation provides consistent CV values with a low standard deviation. It also identified fast quads, number of runs, snow making acreage, and vertical drop as important features. Because the random forest model has a lower cross-validation mean absolute error by almost 1 and exhibits less variability, it has been chosen as the model moving forward.

Currently, Big Mountain charges \$81 per ticket. While this is on the high end for Montana, Big Mountain does justify it by also providing some of the best facilities in the state. In fact, the model suggests the market can support a ticket price of \$96. I would approach business leadership with this suggestion by 1) reviewing the need for covering the new operating costs of the new chair, 2) what the model considers a fair ticket price - with the mean absolute error in mind - with the current facilities, including how vertical rise, total skiable acreage, and snow making acreage contribute to Big Mountain being among the best in Montana, and 3) a couple features that could be improved (or ignored) for further justifying a higher ticket price. Every dollar in ticket price increase is expected to contribute ~\$1.75 million for operating costs of the new chair lift and the bottom line. With respect to the potential scenarios for improving the bottom line, the model suggests the second of potential scenarios - increasing vertical drop with the installation of an additional chair lift - is worth further consideration. Since the model suggests that closing the first run has no impact on ticket price, the business could also consider closing the least used run to start.

More operating costs would have been useful for analysis, especially in considerations with what to change in the upcoming season. For example, what's the cost of operating a new chair lift vs. another 100 acres of snow making? It would have been useful for focusing on ROI instead of just focusing on the cost of the ticket. While Big Mountain already asks a high price for a ticket, it also among the best in many features that all tend to influence ticket price, such as snow making acreage, total skiable terrain, fast quads, vertical rise, and longest run length. Depending on their familiarity with the market, it could come as a surprise to the business executives. To find out, I would provide a quick few powerpoint slides demonstrating the outcome of the model. Further, I would provide access to the model via some user friendly UI so that they could tinker with the parameters if they were interested.